CHAPTER 4

Logic and Propositional Calculus

4.1 INTRODUCTION

Many algorithms and proofs use logical expressions such as:

"IF p THEN q" or "If p_1 AND p_2, THEN q_1 OR q_2"

Therefore it is necessary to know the cases in which these expressions are TRUE or FALSE, that is, to know the "truth value" of such expressions. We discuss these issues in this chapter.

We also investigate the truth value of quantified statements, which are statements which use the logical quantifiers "for every" and "there exist."

4.2 PROPOSITIONS AND COMPOUND STATEMENTS

A *proposition* (or *statement*) is a declarative statement which is true or false, but not both. Consider, for example, the following six sentences:

(i) Ice floats in water. (iii) $2 + 2 = 4$ (v) Where are you going?

(ii) China is in Europe. (iv) $2 + 2 = 5$ (vi) Do your homework.

The first four are propositions, the last two are not. Also, (i) and (iii) are true, but (ii) and (iv) are false.

Compound Propositions

Many propositions are *composite*, that is, composed of *subpropositions* and various connectives discussed subsequently. Such composite propositions are called *compound propositions*. A proposition is said to be *primitive* if it cannot be broken down into simpler propositions, that is, if it is not composite.

For example, the above propositions (i) through (iv) are primitive propositions. On the other hand, the following two propositions are composite:

"Roses are red and violets are blue." and "John is smart or he studies every night."

The fundamental property of a compound proposition is that its truth value is completely determined by the truth values of its subpropositions together with the way in which they are connected to form the compound propositions. The next section studies some of these connectives.

4.3 BASIC LOGICAL OPERATIONS

This section discusses the three basic logical operations of conjunction, disjunction, and negation which correspond, respectively, to the English words "and," "or," and "not."

Conjunction, $p \wedge q$

Any two propositions can be combined by the word "and" to form a compound proposition called the *conjunction* of the original propositions. Symbolically,

$$p \wedge q$$

read "p and q," denotes the conjunction of p and q. Since $p \wedge q$ is a proposition it has a truth value, and this truth value depends only on the truth values of p and q. Specifically:

Definition 4.1: If p and q are true, then $p \wedge q$ is true; otherwise $p \wedge q$ is false.

The truth value of $p \wedge q$ may be defined equivalently by the table in Fig. 4-1(a). Here, the first line is a short way of saying that if p is true and q is true, then $p \wedge q$ is true. The second line says that if p is true and q is false, then $p \wedge q$ is false. And so on. Observe that there are four lines corresponding to the four possible combinations of T and F for the two subpropositions p and q. Note that $p \wedge q$ is true only when both p and q are true.

p	q	$p \wedge q$
T	T	T
T	F	F
F	T	F
F	F	F

(a) "p and q"

p	q	$p \vee q$
T	T	T
T	F	T
F	T	T
F	F	F

(b) "p or q"

p	$\neg p$
T	F
F	T

(c) "not p"

Fig. 4-1

EXAMPLE 4.1 Consider the following four statements:

(i) Ice floats in water and $2 + 2 = 4$. (iii) China is in Europe and $2 + 2 = 4$.

(ii) Ice floats in water and $2 + 2 = 5$. (iv) China is in Europe and $2 + 2 = 5$.

Only the first statement is true. Each of the others is false since at least one of its substatements is false.

Disjunction, $p \vee q$

Any two propositions can be combined by the word "or" to form a compound proposition called the *disjunction* of the original propositions. Symbolically,

$$p \vee q$$

read "p or q," denotes the disjunction of p and q. The truth value of $p \vee q$ depends only on the truth values of p and q as follows.

Definition 4.2: If p and q are false, then $p \vee q$ is false; otherwise $p \vee q$ is true.

The truth value of $p \vee q$ may be defined equivalently by the table in Fig. 4-1(b). Observe that $p \vee q$ is false only in the fourth case when both p and q are false.

EXAMPLE 4.2 Consider the following four statements:

(i) Ice floats in water or $2 + 2 = 4$. (iii) China is in Europe or $2 + 2 = 4$.

(ii) Ice floats in water or $2 + 2 = 5$. (iv) China is in Europe or $2 + 2 = 5$.

Only the last statement (iv) is false. Each of the others is true since at least one of its sub-statements is true.

Remark: The English word "or" is commonly used in two distinct ways. Sometimes it is used in the sense of "p or q or both," i.e., at least one of the two alternatives occurs, as above, and sometimes it is used in the sense of "p or q but not both," i.e., exactly one of the two alternatives occurs. For example, the sentence "He will go to Harvard or to Yale" uses "or" in the latter sense, called the *exclusive disjunction*. Unless otherwise stated, "or" shall be used in the former sense. This discussion points out the precision we gain from our symbolic language: $p \vee q$ is defined by its truth table and *always* means "p and/or q."

Negation, $\neg p$

Given any proposition p, another proposition, called the *negation* of p, can be formed by writing "It is not true that ..." or "It is false that ..." before p or, if possible, by inserting in p the word "not." Symbolically, the negation of p, read "not p," is denoted by

$$\neg p$$

The truth value of $\neg p$ depends on the truth value of p as follows:

Definition 4.3: If p is true, then $\neg p$ is false; and if p is false, then $\neg p$ is true.

The truth value of $\neg p$ may be defined equivalently by the table in Fig. 4-1(c). Thus the truth value of the negation of p is always the opposite of the truth value of p.

EXAMPLE 4.3 Consider the following six statements:

(a_1) Ice floats in water. (a_2) It is false that ice floats in water. (a_3) Ice does not float in water.

(b_1) $2 + 2 = 5$ (b_2) It is false that $2 + 2 = 5$. (b_3) $2 + 2 \neq 5$

Then (a_2) and (a_3) are each the negation of (a_1); and (b_2) and (b_3) are each the negation of (b_1). Since (a_1) is true, (a_2) and (a_3) are false; and since (b_1) is false, (b_2) and (b_3) are true.

Remark: The logical notation for the connectives "and," "or," and "not" is not completely standardized. For example, some texts use:

$$\begin{array}{ccc} p \,\&\, q,\ p \cdot q \text{ or } pq & \text{for} & p \wedge q \\ p + q & \text{for} & p \vee q \\ p',\ \bar{p} \text{ or } \sim p & \text{for} & \neg p \end{array}$$

4.4 PROPOSITIONS AND TRUTH TABLES

Let $P(p, q, \ldots)$ denote an expression constructed from logical variables $p, q, \ldots$, which take on the value TRUE (T) or FALSE (F), and the logical connectives $\wedge$, $\vee$, and $\neg$ (and others discussed subsequently). Such an expression $P(p, q, \ldots)$ will be called a *proposition*.

The main property of a proposition $P(p, q, \ldots)$ is that its truth value depends exclusively upon the truth values of its variables, that is, the truth value of a proposition is known once the truth value of each of its variables is known. A simple concise way to show this relationship is through a *truth table*. We describe a way to obtain such a truth table below.

Consider, for example, the proposition $\neg(p \wedge \neg q)$. Figure 4-2(*a*) indicates how the truth table of $\neg(p \wedge \neg q)$ is constructed. Observe that the first columns of the table are for the variables $p, q, \ldots$ and that there are enough rows in the table, to allow for all possible combinations of *T* and *F* for these *variables*. (For 2 variables, as above, 4 rows are necessary; for 3 variables, 8 rows are necessary; and, in general, for n variables, 2^n rows are required.) There is then a column for each "elementary" stage of the construction of the proposition, the truth value at each step being determined from the previous stages by the definitions of the connectives $\wedge$, $\vee$, $\neg$. Finally we obtain the truth value of the proposition, which appears in the last column.

The actual truth table of the proposition $\neg(p \wedge \neg q)$ is shown in Fig. 4-2(*b*). It consists precisely of the columns in Fig. 4-2(*a*) which appear under the variables and under the proposition; the other columns were merely used in the construction of the truth table.

p	q	$\neg q$	$p \wedge \neg q$	$\neg(p \wedge \neg q)$
T	T	F	F	T
T	F	T	T	F
F	T	F	F	T
F	F	T	F	T

(*a*)

p	q	$\neg(p \wedge \neg q)$
T	T	T
T	F	F
F	T	T
F	F	T

(*b*)

Fig. 4-2

Remark: In order to avoid an excessive number of parentheses, we sometimes adopt an order of precedence for the logical connectives. Specifically,

$$\neg \text{ has precedence over } \wedge \text{ which has precedence over } \vee$$

For example, $\neg p \wedge q$ means $(\neg p) \wedge q$ and not $\neg(p \wedge q)$.

Alternate Method for Constructing a Truth Table

Another way to construct the truth table for $\neg(p \wedge \neg q)$ follows:

(a) First we construct the truth table shown in Fig. 4-3. That is, first we list all the variables and the combinations of their truth values. Also there is a final row labeled "step." Next the proposition is written on the top row to the right of its variables with sufficient space so there is a column under each variable and under each logical operation in the proposition. Lastly (Step 1), the truth values of the variables are entered in the table under the variables in the proposition.

(b) Now additional truth values are entered into the truth table column by column under each logical operation as shown in Fig. 4-4. We also indicate the step in which each column of truth values is entered in the table.

The truth table of the proposition then consists of the original columns under the variables and the last step, that is, the last column is entered into the table.

p	q	$\neg$	$(p$	$\wedge$	$\neg$	$q)$
T	T		T			T
T	F		T			F
F	T		F			T
F	F		F			F
Step						

Fig. 4-3

p	q	$\neg$	$(p$	$\wedge$	$\neg$	$q)$
T	T		T		F	T
T	F		T		T	F
F	T		F		F	T
F	F		F		T	F
Step			1		2	1

(*a*)

p	q	$\neg$	$(p$	$\wedge$	$\neg$	$q)$
T	T		T	F	F	T
T	F		T	T	T	F
F	T		F	F	F	T
F	F		F	F	T	F
Step			1	3	2	1

(*b*)

p	q	$\neg$	$(p$	$\wedge$	$\neg$	$q)$
T	T	T	T	F	F	T
F	T	F	T	T	T	F
F	F	T	F	F	F	T
F	F	T	F	F	T	F
Step		4	1	3	2	1

(*c*)

Fig. 4-4

4.5 TAUTOLOGIES AND CONTRADICTIONS

Some propositions $P(p, q, \ldots)$ contain only T in the last column of their truth tables or, in other words, they are true for any truth values of their variables. Such propositions are called *tautologies*. Analogously, a proposition $P(p, q, \ldots)$ is called a *contradiction* if it contains only F in the last column of its truth table or, in other words, if it is false for any truth values of its variables. For example, the proposition "p or not p," that is, $p \vee \neg p$, is a tautology, and the proposition "p and not p," that is, $p \wedge \neg p$, is a contradiction. This is verified by looking at their truth tables in Fig. 4-5. (The truth tables have only two rows since each proposition has only the one variable p.)

p	$\neg p$	$p \vee \neg p$
T	F	T
F	T	T

(*a*) $p \vee \neg p$

p	$\neg p$	$p \wedge \neg p$
T	F	F
F	T	F

(*b*) $p \wedge \neg p$

Fig. 4-5

Note that the negation of a tautology is a contradiction since it is always false, and the negation of a contradiction is a tautology since it is always true.

Now let $P(p, q, \ldots)$ be a tautology, and let $P_1(p, q, \ldots)$, $P_2(p, q, \ldots), \ldots$ be any propositions. Since $P(p, q, \ldots)$ does not depend upon the particular truth values of its variables $p, q, \ldots$, we can substitute P_1 for p, P_2 for $q, \ldots$ in the tautology $P(p, q, \ldots)$ and still have a tautology. In other words:

Theorem 4.1 (Principle of Substitution): If $P(p, q, \ldots)$ is a tautology, then $P(P_1, P_2, \ldots)$ is a tautology for any propositions $P_1, P_2, \ldots$.

4.6 LOGICAL EQUIVALENCE

Two propositions $P(p, q, \ldots)$ and $Q(p, q, \ldots)$ are said to be *logically equivalent*, or simply *equivalent* or *equal*, denoted by

$$P(p,\ q,\ \ldots) \equiv Q(p,\ q,\ \ldots)$$

if they have identical truth tables. Consider, for example, the truth tables of $\neg(p \wedge q)$ and $\neg p \vee \neg q$ appearing in Fig. 4-6. Observe that both truth tables are the same, that is, both propositions are false in the first case and true in the other three cases. Accordingly, we can write

$$\neg(p \wedge q) \equiv \neg p \vee \neg q$$

In other words, the propositions are logically equivalent.

Remark: Let p be "Roses are red" and q be "Violets are blue." Let S be the statement:

"It is not true that roses are red and violets are blue."

Then S can be written in the form $\neg(p \wedge q)$. However, as noted above, $\neg(p \wedge q) \equiv \neg p \vee \neg q$. Accordingly, S has the same meaning as the statement:

"Roses are not red, or violets are not blue."

p	q	$p \wedge q$	$\neg(p \wedge q)$
T	T	T	F
T	F	F	T
F	T	F	T
F	F	F	T

(a) $\neg(p \wedge q)$

p	q	$\neg p$	$\neg q$	$\neg p \vee \neg q$
T	T	F	F	F
T	F	F	T	T
F	T	T	F	T
F	F	T	T	T

(b) $\neg p \vee \neg q$

Fig. 4-6

4.7 ALGEBRA OF PROPOSITIONS

Propositions satisfy various laws which are listed in Table 4-1. (In this table, T and F are restricted to the truth values "True" and "False," respectively.) We state this result formally.

Theorem 4.2: Propositions satisfy the laws of Table 4-1.

(Observe the similarity between this Table 4-1 and Table 1-1 on sets.)

Table 4-1 Laws of the algebra of propositions

Idempotent laws:	(1a) $p \vee p \equiv p$	(1b) $p \wedge p \equiv p$
Associative laws:	(2a) $(p \vee q) \vee r \equiv p \vee (q \vee r)$	(2b) $(p \wedge q) \wedge r \equiv p \wedge (q \wedge r)$
Commutative laws:	(3a) $p \vee q \equiv q \vee p$	(3b) $p \wedge q \equiv q \wedge p$
Distributive laws:	(4a) $p \vee (q \wedge r) \equiv (p \vee q) \wedge (p \vee r)$	(4b) $p \wedge (q \vee r) \equiv (p \wedge q) \vee (p \wedge r)$
Identity laws:	(5a) $p \vee F \equiv p$ (6a) $p \vee T \equiv T$	(5b) $p \wedge T \equiv p$ (6b) $p \wedge F \equiv F$
Involution law:	(7) $\neg\neg p \equiv p$	
Complement laws:	(8a) $p \vee \neg p \equiv T$ (9a) $\neg T \equiv F$	(8b) $p \wedge \neg p \equiv T$ (9b) $\neg F \equiv T$
DeMorgan's laws:	(10a) $\neg(p \vee q) \equiv \neg p \wedge \neg q$	(10b) $\neg(p \wedge q) \equiv \neg p \vee \neg q$

4.8 CONDITIONAL AND BICONDITIONAL STATEMENTS

Many statements, particularly in mathematics, are of the form "If p then q." Such statements are called *conditional* statements and are denoted by

$$p \rightarrow q$$

The conditional $p \rightarrow q$ is frequently read "p implies q" or "p only if q."

Another common statement is of the form "p if and only if q." Such statements are called *biconditional* statements and are denoted by

$$p \leftrightarrow q$$

The truth values of $p \rightarrow q$ and $p \leftrightarrow q$ are defined by the tables in Fig. 4-7(a) and (b). Observe that:

(a) The conditional $p \rightarrow q$ is false only when the first part p is true and the second part q is false. Accordingly, when p is false, the conditional $p \rightarrow q$ is true regardless of the truth value of q.

(b) The biconditional $p \leftrightarrow q$ is true whenever p and q have the same truth values and false otherwise.

The truth table of $\neg p \wedge q$ appears in Fig. 4-7(c). Note that the truth table of $\neg p \vee q$ and $p \rightarrow q$ are identical, that is, they are both false only in the second case. Accordingly, $p \rightarrow q$ is logically equivalent to $\neg p \vee q$; that is,

$$p \rightarrow q \equiv \neg p \vee q$$

In other words, the conditional statement "If p then q" is logically equivalent to the statement "Not p or q" which only involves the connectives $\vee$ and $\neg$ and thus was already a part of our language. We may regard $p \to q$ as an abbreviation for an oft-recurring statement.

p	q	$p \to q$
T	T	T
T	F	F
F	T	T
F	F	T

(*a*) $p \to q$

p	q	$p \leftrightarrow q$
T	T	T
T	F	F
F	T	F
F	F	T

(*b*) $p \leftrightarrow q$

p	q	$\neg p$	$\neg p \vee q$
T	T	F	T
T	F	F	F
F	T	T	T
F	F	T	T

(*c*) $\neg p \vee q$

Fig. 4-7

4.9 ARGUMENTS

An *argument* is an assertion that a given set of propositions $P_1, P_2, \ldots, P_n$, called *premises*, yields (has a consequence) another proposition Q, called the *conclusion*. Such an argument is denoted by

$$P_1,\ P_2,\ \ldots,\ P_n \vdash Q$$

The notion of a "logical argument" or "valid argument" is formalized as follows:

Definition 4.4: An argument $P_1,\ P_2,\ \ldots,\ P_n \vdash Q$ is said to be *valid* if Q is true whenever all the premises $P_1, P_2, \ldots, P_n$ are true.

An argument which is not valid is called *fallacy*.

EXAMPLE 4.4

(a) The following argument is valid:

$$p,\ p \to q \vdash q \quad (\textbf{\textit{Law of Detachment}})$$

The proof of this rule follows from the truth table in Fig. 4-7(*a*). Specifically, p and $p \to q$ are true simultaneously only in Case (row) 1, and in this case q is true.

(b) The following argument is a fallacy:

$$p \to q,\ q \vdash p$$

For $p \to q$ and q are both true in Case (row) 3 in the truth table in Fig. 4-7(*a*), but in this case p is false.

Now the propositions $P_1, P_2, \ldots, P_n$ are true simultaneously if and only if the proposition $P_1 \wedge P_2 \wedge \ldots P_n$ is true. Thus the argument $P_1, P_2, \ldots, P_n \vdash Q$ is valid if and only if Q is true whenever $P_1 \wedge P_2 \wedge \ldots \wedge P_n$ is true or, equivalently, if the proposition $(P_1 \wedge P_2 \wedge \ldots \wedge P_n) \to Q$ is a tautology. We state this result formally.

Theorem 4.3: The argument $P_1,\ P_2,\ \ldots,\ P_n \vdash Q$ is valid if and only if the proposition $(P_1 \wedge P_2 \ldots \wedge P_n) \to Q$ is a tautology.

We apply this theorem in the next example.

EXAMPLE 4.5

A fundamental principle of logical reasoning states:

"If p implies q and q implies r, then p implies r"

p	q	r	$[(p$	$\rightarrow$	$q)$	$\wedge$	$(q$	$\rightarrow$	$r)]$	$\rightarrow$	$(p$	$\rightarrow$	$r)$
T	T	T	T	T	T	T	T	T	T	T	T	T	T
T	T	F	T	T	T	F	T	F	F	T	T	F	F
T	F	T	T	F	F	F	F	T	T	T	T	T	T
T	F	F	T	F	F	F	F	T	F	T	T	F	F
F	T	T	F	T	T	T	T	T	T	T	F	T	T
F	T	F	F	T	T	F	T	F	F	T	F	T	F
F	F	T	F	T	F	T	F	T	T	T	F	T	T
F	F	F	F	T	F	T	F	T	F	T	F	T	F
Step			1	2	1	3	1	2	1	4	1	2	1

Fig. 4-8

That is, the following argument is valid:

$$p \rightarrow q,\ q \rightarrow r \vdash p \rightarrow r \quad (\textbf{\textit{Law of Syllogism}})$$

This fact is verified by the truth table in Fig. 4-8 which shows that the following proposition is a tautology:

$$[(p \rightarrow q) \wedge (q \rightarrow r)] \rightarrow (p \rightarrow r)$$

Equivalently, the argument is valid since the premises $p \rightarrow q$ and $q \rightarrow r$ are true simultaneously only in Cases (rows) 1, 5, 7, and 8, and in these cases the conclusion $p \rightarrow r$ is also true. (Observe that the truth table required $2^3 = 8$ lines since there are three variables p, q, and r.)

We now apply the above theory to arguments involving specific statements. We emphasize that the validity of an argument does not depend upon the truth values nor the content of the statements appearing in the argument, but upon the particular form of the argument. This is illustrated in the following example.

EXAMPLE 4.6 Consider the following argument:

S_1 : If a man is a bachelor, he is unhappy.
S_2 : If a man is unhappy, he dies young.

S : Bachelors die young

Here the statement S below the line denotes the conclusion of the argument, and the statements S_1 and S_2 above the line denote the premises. We claim that the argument $S_1,\ S_2 \vdash S$ is valid. For the argument is of the form

$$p \rightarrow q,\ q \rightarrow r \vdash p \rightarrow r$$

where p is "He is a bachelor," q is "He is unhappy" and r is "He dies young;" and by Example 4.5 this argument (Law of Syllogism) is valid.

4.10 PROPOSITIONAL FUNCTIONS, QUANTIFIERS

Let A be a given set. *A propositional function* (or an *open sentence* or *condition*) defined on A is an expression

$$p(x)$$

which has the property that $p(a)$ is true or false for each $a \in A$. That is, $p(x)$ becomes a statement (with a truth value) whenever any element $a \in A$ is substituted for the variable x. The set A is called the *domain* of $p(x)$, and the set T_p of all elements of A for which $p(a)$ is true is called the *truth set* of $p(x)$. In other words,

$$T_p = \{x \mid x \in A,\ p(x) \text{ is true}\} \quad \text{or} \quad T_p = \{x \mid p(x)\}$$

Frequently, when A is some set of numbers, the condition $p(x)$ has the form of an equation or inequality involving the variable x.

EXAMPLE 4.7 Find the truth set for each propositional function $p(x)$ defined on the set $\mathbf{N}$ of positive integers.

(a) Let $p(x)$ be "$x + 2 > 7$." Its truth set is $\{6, 7, 8, \ldots\}$ consisting of all integers greater than 5.

(b) Let $p(x)$ be "$x + 5 < 3$." Its truth set is the empty set $\varnothing$. That is, $p(x)$ is not true for any integer in $\mathbf{N}$.

(c) Let $p(x)$ be "$x + 5 > 1$." Its truth set is $\mathbf{N}$. That is, $p(x)$ is true for every element in $\mathbf{N}$.

Remark: The above example shows that if $p(x)$ is a propositional function defined on a set A then $p(x)$ could be true for all $x \in A$, for some $x \in A$, or for no $x \in A$. The next two subsections discuss quantifiers related to such propositional functions.

Universal Quantifier

Let $p(x)$ be a propositional function defined on a set A. Consider the expression

$$(\forall x \in A)\, p(x) \quad \text{or} \quad \forall x\ p(x) \tag{4.1}$$

which reads "For every x in A, $p(x)$ is a true statement" or, simply, "For all x, $p(x)$." The symbol

$$\forall$$

which reads "for all" or "for every" is called the *universal quantifier*. The statement (*4.1*) is equivalent to the statement

$$T_p = \{x \mid x \in A,\ p(x)\} = A \tag{4.2}$$

that is, that the truth set of $p(x)$ is the entire set A.

The expression $p(x)$ by itself is an open sentence or condition and therefore has no truth value. However, $\forall x\ p(x)$, that is $p(x)$ preceded by the quantifier $\forall$, does have a truth value which follows from the equivalence of (*4.1*) and (*4.2*). Specifically:

Q_1: If $\{x \mid x \in A,\ p(x)\} = A$ then $\forall x\ p(x)$ is true; otherwise, $\forall x\ p(x)$ is false.

EXAMPLE 4.8

(a) The proposition $(\forall n \in \mathbf{N})(n + 4 > 3)$ is true since $\{n \mid n + 4 > 3\} = \{1, 2, 3, \ldots\} = \mathbf{N}$.

(b) The proposition $(\forall n \in \mathbf{N})(n + 2 > 8)$ is false since $\{n \mid n + 2 > 8\} = \{7, 8, \ldots\} \neq \mathbf{N}$.

(c) The symbol $\forall$ can be used to define the intersection of an indexed collection $\{A_i \mid i \in I\}$ of sets A_i as follows:

$$\cap(A_i \mid i \in I) = \{x \mid \forall_i \in I,\ x \in A_i\}$$

Existential Quantifier

Let $p(x)$ be a propositional function defined on a set A. Consider the expression

$$(\exists x \in A)\, p(x) \quad \text{or} \quad \exists x,\ p(x) \tag{4.3}$$

which reads "There exists an x in A such that $p(x)$ is a true statement" or, simply, "For some x, $p(x)$." The symbol

$$\exists$$

which reads "there exists" or "for some" or "for at least one" is called the *existential quantifier*. Statement (*4.3*) is equivalent to the statement

$$T_p = \{x \mid x \in A,\ p(x)\} \neq \varnothing \tag{4.4}$$

i.e., that the truth set of $p(x)$ is not empty. Accordingly, $\exists x\ p(x)$, that is, $p(x)$ preceded by the quantifier $\exists$, does have a truth value. Specifically:

Q_2: If $\{x \mid p(x)\} \neq \varnothing$ then $\exists x\ p(x)$ is true; otherwise, $\exists x\ p(x)$ is false.

EXAMPLE 4.9

(a) The proposition $(\exists n \in N)(n + 4 < 7)$ is true since $\{n \mid n + 4 < 7\} = \{1,\ 2\} \neq \varnothing$.

(b) The proposition $(\exists n \in N)(n + 6 < 4)$ is false since $\{n \mid n + 6 < 4\} = \varnothing$.

(c) The symbol $\exists$ can be used to define the union of an indexed collection $\{A_i \mid i \in I\}$ of sets A_i as follows:

$$\cup(A_i \mid i \in I) = \{x \mid \exists\, i \in I,\ x \mid \in A_i\}$$

4.11 NEGATION OF QUANTIFIED STATEMENTS

Consider the statement: "All math majors are male." Its negation reads:

> "It is not the case that all math majors are male" or, equivalently, "There exists at least one math major who is a female (not male)"

Symbolically, using M to denote the set of math majors, the above can be written as

$$\neg(\forall x \in M)(x \text{ is male}) \equiv (\exists\, x \in M)\ (x \text{ is not male})$$

or, when $p(x)$ denotes "x is male,"

$$\neg(\forall x \in M)p(x) \equiv (\exists\, x \in M)\neg p(x) \quad \text{or} \quad \neg\forall x p(x) \equiv \exists x \neg p(x)$$

The above is true for any proposition $p(x)$. That is:

Theorem 4.4 (DeMorgan): $\neg(\forall x \in A)p(x) \equiv (\exists\, x \in A)\neg p(x)$.

In other words, the following two statements are equivalent:
(1) It is not true that, for all $a \in A$, $p(a)$ is true. (2) There exists an $a \in A$ such that $p(a)$ is false.

There is an analogous theorem for the negation of a proposition which contains the existential quantifier.

Theorem 4.5 (DeMorgan): $\neg(\exists x \in A)p(x) \equiv (\forall x \in A)\neg p(x)$.

That is, the following two statements are equivalent:
(1) It is not true that for some $a \in A$, $p(a)$ is true. (2) For all $a \in A$, $p(a)$ is false.

EXAMPLE 4.10

(a) The following statements are negatives of each other:

"For all positive integers n we have $n + 2 > 8$"
"There exists a positive integer n such that $n + 2 \not> 8$"

(b) The following statements are also negatives of each other:

"There exists a (living) person who is 150 years old"
"Every living person is not 150 years old"

Remark: The expression $\neg p(x)$ has the obvious meaning:

"The statement $\neg p(a)$ is true when $p(a)$ is false, and vice versa"

Previously, $\neg$ was used as an operation on statements; here $\neg$ is used as an operation on propositional functions. Similarly, $p(x) \wedge q(x)$, read "$p(x)$ and $q(x)$," is defined by:

"The statement $p(a) \wedge q(a)$ is true when $p(a)$ and $q(a)$ are true"

Similarly, $p(x) \vee q(x)$, read "$p(x)$ or $q(x)$," is defined by:

"The statement $p(a) \vee q(a)$ is true when $p(a)$ or $q(a)$ is true"

Thus in terms of truth sets:

(i) $\neg p(x)$ is the complement of $p(x)$.
(ii) $p(x) \wedge q(x)$ is the intersection of $p(x)$ and $q(x)$.
(iii) $p(x) \vee q(x)$ is the union of $p(x)$ and $q(x)$.

One can also show that the laws for propositions also hold for propositional functions. For example, we have DeMorgan's laws:

$$\neg(p(x) \wedge q(x)) \equiv \neg p(x) \vee \neg q(x) \quad \text{and} \quad \neg(p(x) \vee q(x)) \equiv \neg p(x) \wedge \neg q(x)$$

Counterexample

Theorem 4.6 tells us that to show that a statement $\forall x,\ p(x)$ is false, it is equivalent to show that $\exists\, x \neg p(x)$ is true or, in other words, that there is an element x_0 with the property that $p(x_0)$ is false. Such an element x_0 is called a *counterexample* to the statement $\forall x,\ p(x)$.

EXAMPLE 4.11

(a) Consider the statement $\forall x \in \mathbf{R},\ |x| \neq 0$. The statement is false since 0 is a counterexample, that is, $|0| \neq 0$ is not true.

(b) Consider the statement $\forall x \in \mathbf{R},\ x^2 \geq x$. The statement is not true since, for example, $\frac{1}{2}$ is a counterexample. Specifically, $(\frac{1}{2})^2 \geq \frac{1}{2}$ is not true, that is, $(\frac{1}{2})^2 < \frac{1}{2}$.

(c) Consider the statement $\forall x \in \mathbf{N}, x^2 \geq x$. This statement is true where $\mathbf{N}$ is the set of positive integers. In other words, there does not exist a positive integer n for which $n^2 < n$.

Propositional Functions with more than One Variable

A propositional function (of n variables) defined over a product set $A = A_1 \times \cdots \times A_n$ is an expression

$$p(x_1, x_2, \ldots, x_n)$$

which has the property that $p(a_1, a_2, \ldots, a_n)$ is true or false for any n-tuple $(a_1, \ldots a_n)$ in A. For example,

$$x + 2y + 3z < 18$$

is a propositional function on $\mathbf{N}^3 = \mathbf{N} \times \mathbf{N} \times \mathbf{N}$. Such a propositional function has no truth value. However, we do have the following:

Basic Principle: A propositional function preceded by a quantifier for each variable, for example,

$$\forall x \exists y,\ p(x, y) \quad \text{or} \quad \exists x\, \forall y\, \exists z,\ p(x, y, z)$$

denotes a statement and has a truth value.

EXAMPLE 4.12 Let $B = \{1, 2, 3, \ldots, 9\}$ and let $p(x, y)$ denote "$x + y = 10$." Then $p(x, y)$ is a propositional function on $A = B^2 = B \times B$.

(a) The following is a statement since there is a quantifier for each variable:

$$\forall x \exists y,\ p(x, y), \quad \text{that is,} \quad \text{"For every } x, \text{ there exists a } y \text{ such that } x + y = 10\text{"}$$

This statement is true. For example, if $x = 1$, let $y = 9$; if $x = 2$, let $y = 8$, and so on.

(b) The following is also a statement:

$$\exists y \forall x,\ p(x, y), \quad \text{that is,} \quad \text{"There exists a } y \text{ such that, for every } x, \text{ we have } x + y = 10\text{"}$$

No such y exists; hence this statement is false.

Note that the only difference between (a) and (b) is the order of the quantifiers. Thus a different ordering of the quantifiers may yield a different statement. We note that, when translating such quantified statements into English, the expression "such that" frequently follows "there exists."

Negating Quantified Statements with more than One Variable

Quantified statements with more than one variable may be negated by successively applying Theorems 4.5 and 4.6. Thus each $\forall$ is changed to $\exists$ and each $\exists$ is changed to $\forall$ as the negation symbol $\neg$ passes through the statement from left to right. For example,

$$\begin{aligned} \neg[\forall x \exists y \exists z,\ p(x, y, z)] &\equiv \exists x \neg[\exists y \exists z,\ p(x, y, z)] \equiv \neg \exists z \forall y [\exists z,\ p(x, y, z) \\ &\equiv \exists x \forall y \forall z,\ \neg p(x, y, z) \end{aligned}$$

Naturally, we do not put in all the steps when negating such quantified statements.

EXAMPLE 4.13

(a) Consider the quantified statement:

"Every student has at least one course where the lecturer is a teaching assistant."

Its negation is the statement:

"There is a student such that in every course the lecturer is not a teaching assistant."

(b) The formal definition that L is the limit of a sequence $a_1, a_2, \ldots$ follows:

$$\forall \in > 0, \exists\, n_0 \in \mathrm{N},\ \forall n > n_0 \text{ we have } |a_n - L| < \in$$

Thus L is not the limit of the sequence $a_1, a_2, \ldots$ when:

$$\exists \in > 0, \forall n_0 \in \mathrm{N},\ \exists n > n_0 \text{ such that } |a_n - L| \geq \in$$

Solved Problems

PROPOSITIONS AND TRUTH TABLES

4.1. Let p be "It is cold" and let q be "It is raining". Give a simple verbal sentence which describes each of the following statements: (*a*) $\neg p$; (*b*) $p \wedge q$; (*c*) $p \vee q$; (*d*) $q \vee \neg p$.

In each case, translate $\wedge$, $\vee$, and $\sim$ to read "and," "or," and "It is false that" or "not," respectively, and then simplify the English sentence.

(*a*) It is not cold.

(*b*) It is cold and raining.

(*c*) It is cold or it is raining.

(*d*) It is raining or it is not cold.

4.2. Find the truth table of $\neg p \wedge q$.

Construct the truth table of $\neg p \wedge q$ as in Fig. 4-9(*a*).

p	q	$\neg p$	$\neg p \wedge q$
T	T	F	F
T	F	F	F
F	T	T	T
F	F	T	F

(*a*) $\neg p \wedge q$

p	q	$p \wedge q$	$\neg (p \wedge q)$	$p \vee \neg (p \wedge q)$
T	T	T	F	T
T	F	F	T	T
F	T	F	T	T
F	F	F	T	T

(*b*) $p \vee \neg (p \wedge q)$

Fig. 4-9

4.3. Verify that the proposition $p \vee \neg(p \wedge q)$ is a tautology.

Construct the truth table of $p \vee \neg(p \wedge q)$ as shown in Fig. 4-9(*b*). Since the truth value of $p \vee \neg(p \wedge q)$ is T for all values of p and q, the proposition is a tautology.

4.4. Show that the propositions $\neg(p \wedge q)$ and $\neg p \vee \neg q$ are logically equivalent.

Construct the truth tables for $\neg(p \wedge q)$ and $\neg p \vee \neg q$ as in Fig. 4-10. Since the truth tables are the same (both propositions are false in the first case and true in the other three cases), the propositions $\neg(p \wedge q)$ and $\neg p \vee \neg q$ are logically equivalent and we can write

$$\neg(p \wedge q) \equiv \neg p \vee \neg q.$$

p	q	$p \wedge q$	$\neg (p \wedge q)$
T	T	T	F
T	F	F	T
F	T	F	T
F	F	F	T

(*a*) $\neg (p \wedge q)$

p	q	$\neg p$	$\neg q$	$\neg p \vee \neg q$
T	T	F	F	F
T	F	F	T	T
F	T	T	F	T
F	F	T	T	T

(*b*) $\neg p \vee \neg q$

Fig. 4-10

4.5. Use the laws in Table 4-1 to show that $\neg(p \wedge q) \vee (\neg p \wedge q) \equiv \neg p$.

Statement		Reason
(1) $\neg(p \vee q) \vee (\neg p \wedge q)$	$\equiv (\neg p \wedge \neg q) \vee (\neg p \wedge q)$	DeMorgan's law
(2)	$\equiv \neg p \wedge (\neg q \vee q)$	Distributive law
(3)	$\equiv \neg p \wedge T$	Complement law
(4)	$\equiv \neg p$	Identity law

CONDITIONAL STATEMENTS

4.6. Rewrite the following statements without using the conditional:

(*a*) If it is cold, he wears a hat.

(*b*) If productivity increases, then wages rise.

Recall that "If p then q" is equivalent to "Not p or q;" that is, $p \to q \equiv \neg p \vee q$. Hence,

(*a*) It is not cold or he wears a hat.

(*b*) Productivity does not increase or wages rise.

4.7. Consider the conditional proposition $p \to q$. The simple propositions $q \to p$, $\neg p \to \neg q$ and $\neg q \to \neg p$ are called, respectively, the *converse*, *inverse*, and *contrapositive* of the conditional $p \to q$. Which if any of these propositions are logically equivalent to $p \to q$?

Construct their truth tables as in Fig. 4-11. Only the contrapositive $\neg q \to \neg p$ is logically equivalent to the original conditional proposition $p \to q$.

p	q	$\neg p$	$\neg q$	Conditional $p \to q$	Converse $q \to p$	Inverse $\neg p \to \neg q$	Contrapositive $\neg q \to \neg p$
T	T	F	F	T	T	T	T
T	F	F	T	F	T	T	F
F	T	T	F	T	F	F	T
F	F	T	T	T	T	T	T

Fig. 4-11

4.8. Determine the contrapositive of each statement:

(*a*) If Erik is a poet, then he is poor.

(*b*) Only if Marc studies will he pass the test.

(*a*) The contrapositive of $p \to q$ is $\neg q \to \neg p$. Hence the contrapositive follows:

If Erik is not poor, then he is not a poet.

(*b*) The statement is equivalent to: "If Marc passes the test, then he studied." Thus its contrapositive is:

If Marc does not study, then he will not pass the test.

4.9. Write the negation of each statement as simply as possible:

(*a*) If she works, she will earn money.

(*b*) He swims if and only if the water is warm.

(*c*) If it snows, then they do not drive the car.

(*a*) Note that $\neg(p \to q) \equiv p \wedge \neg q$; hence the negation of the statement is:

She works or she will not earn money.

(*b*) Note that $\neg(p \leftrightarrow q) \equiv p \leftrightarrow \neg q \equiv \neg p \leftrightarrow q$; hence the negation of the statement is either of the following:

He swims if and only if the water is not warm.
He does not swim if and only if the water is warm.

(*c*) Note that $\neg(p \rightarrow \neg q) \equiv p \wedge \neg\neg q \equiv p \wedge q$. Hence the negation of the statement is:

It snows and they drive the car.

ARGUMENTS

4.10. Show that the following argument is a fallacy: $p \rightarrow q, \neg p \vdash \neg q$.

Construct the truth table for $[(p \rightarrow q) \wedge \neg p] \rightarrow \neg q$ as in Fig. 4-12. Since the proposition $[(p \rightarrow q) \wedge \neg p] \rightarrow \neg q$ is not a tautology, the argument is a fallacy. Equivalently, the argument is a fallacy since in the third line of the truth table $p \rightarrow q$ and $\neg p$ are true but $\neg q$ is false.

p	q	$p \rightarrow q$	$\neg p$	$(p \rightarrow q) \wedge \neg p$	$\neg q$	$[(p \rightarrow q) \wedge \neg p] \rightarrow \neg p$
T	T	T	F	F	F	T
T	F	F	F	F	T	T
F	T	T	T	T	F	F
F	F	T	T	T	T	T

Fig. 4-12

4.11. Determine the validity of the following argument: $p \rightarrow q, \neg p \vdash \neg p$.

Construct the truth table for $[(p \rightarrow q) \wedge \neg q] \rightarrow \neg p$ as in Fig. 4-13. Since the proposition $[(p \rightarrow q) \wedge \neg q] \rightarrow \neg p$ is a tautology, the argument is valid.

p	q	$[(p$	$\rightarrow$	$q)$	$\wedge$	$\neg$	$q]$	$\rightarrow$	$\neg$	p
T	T	T	T	T	F	F	T	T	F	T
T	F	T	F	F	F	T	F	T	F	T
F	T	F	T	T	F	F	T	T	T	F
F	F	F	T	F	T	T	F	T	T	F
Step		1	2	1	3	2	1	4	2	1

Fig. 4-13

4.12. Prove the following argument is valid: $p \rightarrow \neg q$, $r \rightarrow q$, $r \vdash \neg p$.

Construct the truth table of the premises and conclusions as in Fig. 4-14(*a*). Now, $p \rightarrow \neg q$, $r \rightarrow q$, and r are true simultaneously only in the fifth row of the table, where $\neg p$ is also true. Hence the argument is valid.

	p	q	r	$p \rightarrow \neg q$	$r \rightarrow q$	$\neg q$
1	T	T	T	F	T	F
2	T	T	F	F	T	F
3	T	F	T	T	F	F
4	T	F	F	T	T	F
5	F	T	T	T	T	T
6	F	T	F	T	T	T
7	F	F	T	T	F	T
8	F	F	F	T	T	T

(*a*)

p	q	$\neg q$	$p \rightarrow \neg q$	$\neg p$
T	T	F	F	F
T	F	T	T	F
F	T	F	T	T
F	F	T	T	T

(*b*)

Fig. 4-14

4.13. Determine the validity of the following argument:

If 7 is less than 4, then 7 is not a prime number.
7 is not less than 4.

---

7 is a prime number.

First translate the argument into symbolic form. Let p be "7 is less than 4" and q be "7 is a prime number." Then the argument is of the form

$$p \to \neg q,\ \neg q \vdash q$$

Now, we construct a truth table as shown in Fig. 4-14(b). The above argument is shown to be a fallacy since, in the fourth line of the truth table, the premises $p \to \neg q$ and $\neg p$ are true, but the conclusion q is false.

Remark: The fact that the conclusion of the argument happens to be a true statement is irrelevant to the fact that the argument presented is a fallacy.

4.14. Test the validity of the following argument:

If two sides of a triangle are equal, then the opposite angles are equal.
Two sides of a triangle are not equal.

The opposite angles are not equal.

First translate the argument into the symbolic form $p \to q$, $\neg p \vdash \neg q$, where p is "Two sides of a triangle are equal" and q is "The opposite angles are equal." By Problem 4.10, this argument is a fallacy.

Remark: Although the conclusion *does* follow from the second premise and axioms of Euclidean geometry, the above argument does not constitute such a proof since the argument is a fallacy.

QUANTIFIERS AND PROPOSITIONAL FUNCTIONS

4.15. Let $A = \{1, 2, 3, 4, 5\}$. Determine the truth value of each of the following statements:

(a) $(\exists x \in A)(x + 3 = 10)$ (c) $(\exists x \in A)(x + 3 < 5)$

(b) $(\forall x \in A)(x + 3 < 10)$ (d) $(\forall x \in A)(x + 3 \leq 7)$

(a) False. For no number in A is a solution to $x + 3 = 10$.

(b) True. For every number in A satisfies $x + 3 < 10$.

(c) True. For if $x_0 = 1$, then $x_0 + 3 < 5$, i.e., 1 is a solution.

(d) False. For if $x_0 = 5$, then $x_0 + 3$ is not less than or equal 7. In other words, 5 is not a solution to the given condition.

4.16. Determine the truth value of each of the following statements where $\mathbf{U} = \{1, 2, 3\}$ is the universal set: (a) $\exists x \forall y, x^2 < y + 1$; ($b$) $\forall x \exists y, x^2 + y^2 < 12$; ($c$) $\forall x \forall y, x^2 + y^2 < 12$.

(a) True. For if $x = 1$, then 1, 2, and 3 are all solutions to $1 < y + 1$.

(b) True. For each x_0, let $y = 1$; then $x_0^2 + 1 < 12$ is a true statement.

(c) False. For if $x_0 = 2$ and $y_0 = 3$, then $x_0^2 + y_0^2 < 12$ is not a true statement.

4.17. Negate each of the following statements:

(a) $\exists x\ \forall y,\ p(x, y)$; ($b$) $\exists x\ \forall y,\ p(x, y)$; ($c$) $\exists y\ \exists x\ \forall z,\ p(x, y, z)$.

Use $\neg \forall x\ p(x) \equiv \exists x \neg p(x)$ and $\neg \exists x\ p(x) \equiv \forall x \neg p(x)$:

(a) $\neg(\exists x \forall y,\ p(x, y)) \equiv \forall x \exists y \neg p(x, y)$

(b) $\neg(\forall x \forall y,\ p(x, y)) \equiv \exists x \exists y \neg p(x, y)$

(c) $\neg(\exists y\ \exists x\ \forall z,\ p(x, y, z)) \equiv \forall y\ \forall x\ \exists z \neg p(x, y, z)$

4.18. Let $p(x)$ denote the sentence "$x + 2 > 5$." State whether or not $p(x)$ is a propositional function on each of the following sets: (*a*) **N**, the set of positive integers; (*b*) $M = \{-1, -2, -3, \ldots\}$; (*c*) C, the set of complex numbers.

(*a*) Yes.

(*b*) Although $p(x)$ is false for every element in M, $p(x)$ is still a propositional function on M.

(*c*) No. Note that $2i + 2 > 5$ does not have any meaning. In other words, inequalities are not defined for complex numbers.

4.19. Negate each of the following statements: (*a*) All students live in the dormitories. (*b*) All mathematics majors are males. (*c*) Some students are 25 years old or older.

Use Theorem 4.4 to negate the quantifiers.

(*a*) At least one student does not live in the dormitories. (Some students do not live in the dormitories.)

(*b*) At least one mathematics major is female. (Some mathematics majors are female.)

(*c*) None of the students is 25 years old or older. (All the students are under 25.)

Supplementary Problems

PROPOSITIONS AND TRUTH TABLES

4.20. Let p denote "He is rich" and let q denote "He is happy." Write each statement in symbolic form using p and q. Note that "He is poor" and "He is unhappy" are equivalent to $\neg p$ and $\neg q$, respectively.

(a) If he is rich, then he is unhappy.

(b) He is neither rich nor happy.

(c) It is necessary to be poor in order to be happy.

(d) To be poor is to be unhappy.

4.21. Find the truth tables for. (a) $p \vee \neg q$; (b) $\neg p \wedge \neg q$.

4.22. Verify that the proposition $(p \wedge q) \wedge \neg(p \vee q)$ is a contradiction.

ARGUMENTS

4.23. Test the validity of each argument:

(a)
If it rains, Erik will be sick.
It did not rain.

Erik was not sick.

(b)
If it rains, Erik will be sick.
Erik was not sick.

It did not rain.

4.24. Test the validity of the following argument:

If I study, then I will not fail mathematics.
If I do not play basketball, then I will study.
But I failed mathematics.

Therefore I must have played basketball.

QUANTIFIERS

4.25. Let $A = \{1, 2, \ldots, 9, 10\}$. Consider each of the following sentences. If it is a statement, then determine its truth value. If it is a propositional function, determine its truth set.

(a) $(\forall x \in A)(\exists y \in A)(x + y < 14)$

(b) $(\forall y \in A)(x + y < 14)$

(c) $(\forall x \in A)(\forall y \in A)(x + y < 14)$

(d) $(\exists y \in A)(x + y < 14)$

4.26. Negate each of the following statements:

(a) If the teacher is absent, then some students do not complete their homework.

(b) All the students completed their homework and the teacher is present.

(c) Some of the students did not complete their homework or the teacher is absent.

4.27. Negate each statement in Problem 4.15.

4.28. Find a counterexample for each statement were $\mathbf{U} = \{3, 5, 7, 9\}$ is the universal set:

(a) $\forall x, x + 3 \geq 7$, (b) $\forall x, x$ is odd, (c) $\forall x, x$ is prime, (d) $\forall x, |x| = x$

Answers to Supplementary Problems

4.20. (a) $p \rightarrow \neg q$; (b) $\neg p \wedge \neg q$; (c) $q \rightarrow \neg p$; (d) $\neg p \rightarrow \neg q$.

4.21. (a) T, T, F, T; (b) F, F, F, T.

4.22. Construct its truth table. It is a contradiction since its truth table is false for all values of p and q.

4.23. First translate the arguments into symbolic form: p for "It rains," and q for "Erik is sick:"

(a) $p \rightarrow q,\ \neg p \vdash \neg q$ (b) $p \rightarrow q,\ \neg q \vdash \neg p$

By Problem 4.10, (a) is a fallacy. By Problem 4.11, (b) is valid.

4.24. Let p be "I study," q be "I failed mathematics," and r be "I play basketball." The argument has the form:

$$p \rightarrow \neg q,\ \neg r \rightarrow p, q \vdash r$$

Construct the truth tables as in Fig. 4-15, where the premises $p \rightarrow \neg q$, $\neg r \rightarrow p$, and q are true simultaneously only in the fifth line of the table, and in that case the conclusion r is also true. Hence the argument is valid.

p	q	r	$\neg q$	$p \rightarrow \neg q$	$\neg r$	$\neg r \rightarrow p$
T	T	T	F	F	F	T
T	T	F	F	F	T	T
T	F	T	T	T	F	T
T	F	F	T	T	T	T
F	T	T	F	T	F	T
F	T	F	F	T	T	F
F	F	T	T	T	F	T
F	F	F	T	T	T	F

Fig. 4-15

4.25. (a) The open sentence in two variables is preceded by two quantifiers; hence it is a statement. Moreover, the statement is true.

(b) The open sentence is preceded by one quantifier; hence it is a propositional function of the other variable. Note that for every $y \in A$, $x_0 + y < 14$ if and only if $x_0 = 1, 2$, or 3. Hence the truth set is $\{1, 2, 3\}$.

(c) It is a statement and it is false: if $x_0 = 8$ and $y_0 = 9$, then $x_0 + y_0 < 14$ is not true.

(d) It is an open sentence in x. The truth set is A itself.

4.26. (a) The teacher is absent and all the students completed their homework.

(b) Some of the students did not complete their homework or the teacher is absent.

(c) All the students completed their homework and the teacher is present.

4.27. (a) $(\forall x \in A)(x + 3 \neq 10)$ (c) $(\forall x \in A)(x + 3 \geq 5)$

(b) $(\exists x \in A)(x + 3 \geq 10)$ (d) $(\exists x \in A)(x + 3 > 7)$

4.28. (a) Here 3 is a counterexample.

(b) The statement is true; hence no counterexample exists.

(c) Here 9 is the only counterexample.

(d) The statement is true; hence there is no counterexample.

CHAPTER 2

Relations

2.1 INTRODUCTION

The reader is familiar with many relations such as "less than," "is parallel to," "is a subset of," and so on. In a certain sense, these relations consider the existence or nonexistence of a certain connection between pairs of objects taken in a definite order. Formally, we define a relation in terms of these "ordered pairs."

An *ordered pair* of elements a and b, where a is designated as the first element and b as the second element, is denoted by (a, b). In particular,

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

if and only if $a = c$ and $b = d$. Thus $(a, b) \neq (b, a)$ unless $a = b$. This contrasts with sets where the order of elements is irrelevant; for example, $\{3, 5\} = \{5, 3\}$.

2.2 PRODUCT SETS

Consider two arbitrary sets A and B. The set of all ordered pairs (a, b) where $a \in A$ and $b \in B$ is called the *product*, or *Cartesian product*, of A and B. A short designation of this product is $A \times B$, which is read "A cross B." By definition,

$$A \times B = \{(a, b) \mid a \in A \text{ and } b \in B\}$$

One frequently writes A^2 instead of $A \times A$.

EXAMPLE 2.1 $\mathbf{R}$ denotes the set of real numbers and so $\mathbf{R}^2 = \mathbf{R} \times \mathbf{R}$ is the set of ordered pairs of real numbers. The reader is familiar with the geometrical representation of $\mathbf{R}^2$ as points in the plane as in Fig. 2-1. Here each point P represents an ordered pair (a, b) of real numbers and vice versa; the vertical line through P meets the x-axis at a, and the horizontal line through P meets the y-axis at b. $\mathbf{R}^2$ is frequently called the *Cartesian plane*.

EXAMPLE 2.2 Let $A = \{1, 2\}$ and $B = \{a, b, c\}$. Then

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

Also, $A \times A = \{(1, 1), (1, 2), (2, 1), (2, 2)\}$

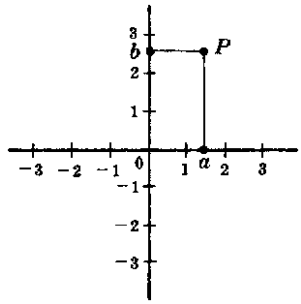


Fig. 2-1

There are two things worth noting in the above examples. First of all $A \times B \neq B \times A$. The Cartesian product deals with ordered pairs, so naturally the order in which the sets are considered is important. Secondly, using $n(S)$ for the number of elements in a set S, we have:

$$n(A \times B) = 6 = 2(3) = n(A)n(B)$$

In fact, $n(A \times B) = n(A)n(B)$ for any finite sets A and B. This follows from the observation that, for an ordered pair (a, b) in $A \times B$, there are $n(A)$ possibilities for a, and for each of these there are $n(B)$ possibilities for b.

The idea of a product of sets can be extended to any finite number of sets. For any sets $A_1, A_2, \ldots, A_n$, the set of all ordered n-tuples $(a_1, a_2, \ldots, a_n)$ where $a_1 \in A_1, a_2 \in A_2, \ldots, a_n \in A_n$ is called the *product* of the sets $A_1, \ldots, A_n$ and is denoted by

$$A_1 \times A_2 \times \cdots \times A_n \quad \text{or} \quad \prod_{i=1}^{n} A_1$$

Just as we write A^2 instead of $A \times A$, so we write A^n instead of $A \times A \times \cdots \times A$, where there are n factors all equal to A. For example, $\mathbf{R}^3 = \mathbf{R} \times \mathbf{R} \times \mathbf{R}$ denotes the usual three-dimensional space.

2.3 RELATIONS

We begin with a definition.

Definition 2.1: Let A and B be sets. A *binary relation* or, simply, *relation* from A to B is a subset of $A \times B$.

Suppose R is a relation from A to B. Then R is a set of ordered pairs where each first element comes from A and each second element comes from B. That is, for each pair $a \in A$ and $b \in B$, exactly one of the following is true:

(i) $(a, b) \in R$; we then say "a is R-related to b", written aRb.

(ii) $(a, b) \notin R$; we then say "a is not R-related to b", written $a \not R b$.

If R is a relation from a set A to itself, that is, if R is a subset of $A^2 = A \times A$, then we say that R is a relation *on* A.

The *domain* of a relation R is the set of all first elements of the ordered pairs which belong to R, and the *range* is the set of second elements.

Although n-ary relations, which involve ordered n-tuples, are introduced in Section 2.10, the term relation shall then mean binary relation unless otherwise stated or implied.

EXAMPLE 2.3

(a) $A = (1, 2, 3)$ and $B = \{x, y, z\}$, and let $R = \{(1, y), (1, z), (3, y)\}$. Then R is a relation from A to B since R is a subset of $A \times B$. With respect to this relation,

$$1Ry, 1Rz, 3Ry, \quad \text{but} \quad 1\not Rx, 2\not Rx, 2\not Ry, 2\not Rz, 3\not Rx, 3\not Rz$$

The domain of R is $\{1, 3\}$ and the range is $\{y, z\}$.

(b) Set inclusion $\subseteq$ is a relation on any collection of sets. For, given any pair of set A and B, either $A \subseteq B$ or $A \nsubseteq B$.

(c) A familiar relation on the set $\mathbf{Z}$ of integers is "m divides n." A common notation for this relation is to write $m \mid n$ when m divides n. Thus $6 \mid 30$ but $7 \nmid 25$.

(d) Consider the set L of lines in the plane. Perpendicularity, written "$\perp$," is a relation on L. That is, given any pair of lines a and b, either $a \perp b$ or $a \not\perp b$. Similarly, "is parallel to," written "$\parallel$," is a relation on L since either $a \parallel b$ or $a \nparallel b$.

(e) Let A be any set. An important relation on A is that of *equality*,

$$\{(a, a) \mid a \in A\}$$

which is usually denoted by "$=$." This relation is also called the *identity* or *diagonal* relation on A and it will also be denoted by Δ_A or simply Δ.

(f) Let A be any set. Then $A \times A$ and $\emptyset$ are subsets of $A \times A$ and hence are relations on A called the *universal relation* and *empty relation*, respectively.

Inverse Relation

Let R be any relation from a set A to a set B. The *inverse* of R, denoted by R^{-1}, is the relation from B to A which consists of those ordered pairs which, when reversed, belong to R; that is,

$$R^{-1} = \{(b, a) \mid (a, b) \in R\}$$

For example, let $A = \{1, 2, 3\}$ and $B = \{x, y, z\}$. Then the inverse of

$$R = \{(1, y), (1, z), (3, y)\} \quad \text{is} \quad R^{-1} = \{(y, 1), (z, 1), (y, 3)\}$$

Clearly, if R is any relation, then $(R^{-1})^{-1} = R$. Also, the domain and range of R^{-1} are equal, respectively, to the range and domain of R. Moreover, if R is a relation on A, then R^{-1} is also a relation on A.

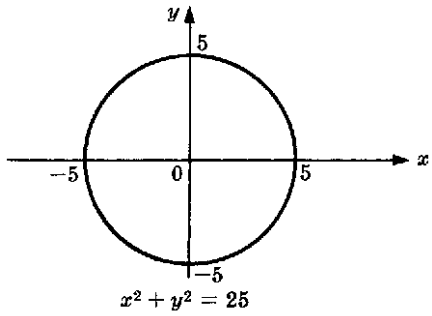
2.4 PICTORIAL REPRESENTATIVES OF RELATIONS

There are various ways of picturing relations.

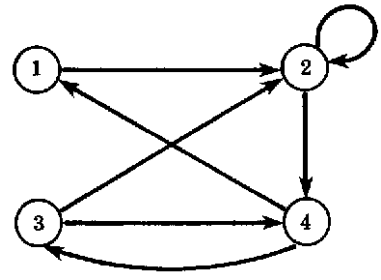
Relations on R

Let S be a relation on the set $\mathbf{R}$ of real numbers; that is, S is a subset of $\mathbf{R}^2 = \mathbf{R} \times \mathbf{R}$. Frequently, S consists of all ordered pairs of real numbers which satisfy some given equation $E(x, y) = 0$ (such as $x^2 + y^2 = 25$).

Since $\mathbf{R}^2$ can be represented by the set of points in the plane, we can picture S by emphasizing those points in the plane which belong to S. The pictorial representation of the relation is sometimes called the *graph* of the relation. For example, the graph of the relation $x^2 + y^2 = 25$ is a circle having its center at the origin and radius 5. See Fig. 2-2(a).



(a)



(b)

Fig. 2-2

Directed Graphs of Relations on Sets

There is an important way of picturing a relation R on a finite set. First we write down the elements of the set, and then we draw an arrow from each element x to each element y whenever x is related to y. This diagram is called the *directed graph* of the relation. Figure 2-2(b), for example, shows the directed graph of the following relation R on the set $A = \{1, 2, 3, 4\}$:

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

Observe that there is an arrow from 2 to itself, since 2 is related to 2 under R.

These directed graphs will be studied in detail as a separate subject in Chapter 8. We mention it here mainly for completeness.

Pictures of Relations on Finite Sets

Suppose A and B are finite sets. There are two ways of picturing a relation R from A to B.

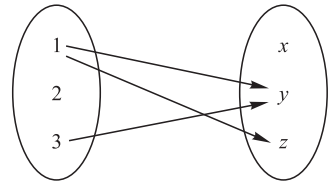
(i) Form a rectangular array (matrix) whose rows are labeled by the elements of A and whose columns are labeled by the elements of B. Put a 1 or 0 in each position of the array according as $a \in A$ is or is not related to $b \in B$. This array is called the *matrix of the relation*.

(ii) Write down the elements of A and the elements of B in two disjoint disks, and then draw an arrow from $a \in A$ to $b \in B$ whenever a is related to b. This picture will be called the *arrow diagram* of the relation.

Figure 2-3 pictures the relation R in Example 2.3(a) by the above two ways.

	x	y	z
1	0	1	1
2	0	0	0
3	0	1	0

(i)



(ii)

$R = \{(1, y),\ (1, z),\ (3, y)\}$

Fig. 2-3

2.5 COMPOSITION OF RELATIONS

Let A, B and C be sets, and let R be a relation from A to B and let S be a relation from B to C. That is, R is a subset of $A \times B$ and S is a subset of $B \times C$. Then R and S give rise to a relation from A to C denoted by $R \circ S$ and defined by:

$$a(R \circ S)c \text{ if for some } b \in B \text{ we have } aRb \text{ and } bSc.$$

That is ,

$$R \circ S = \{(a, c) \mid \text{there exists } b \in B \text{ for which } (a, b) \in R \text{ and } (b, c) \in S\}$$

The relation $R \circ S$ is called the *composition* of R and S; it is sometimes denoted simply by RS.

Suppose R is a relation on a set A, that is, R is a relation from a set A to itself. Then $R \circ R$, the composition of R with itself, is always defined. Also, $R \circ R$ is sometimes denoted by R^2. Similarly, $R^3 = R^2 \circ R = R \circ R \circ R$, and so on. Thus R^n is defined for all positive n.

Warning: Many texts denote the composition of relations R and S by $S \circ R$ rather than $R \circ S$. This is done in order to conform with the usual use of $g \circ f$ to denote the composition of f and g where f and g are functions. Thus the reader may have to adjust this notation when using this text as a supplement with another text. However, when a relation R is composed with itself, then the meaning of $R \circ R$ is unambiguous.

EXAMPLE 2.4 Let $A = \{1, 2, 3, 4\}$, $B = \{a, b, c, d\}$, $C = \{x, y, z\}$ and let

$$R = \{(1, a), (2, d), (3, a), (3, b), (3, d)\} \quad \text{and} \quad S = \{(b, x), (b, z), (c, y), (d, z)\}$$

Consider the arrow diagrams of R and S as in Fig. 2-4. Observe that there is an arrow from 2 to d which is followed by an arrow from d to z. We can view these two arrows as a "path" which "connects" the element $2 \in A$ to the element $z \in C$. Thus:

$$2(R \circ S)z \quad \text{since } 2Rd \text{ and } dSz$$

Similarly there is a path from 3 to x and a path from 3 to z. Hence

$$3(R \circ S)x \quad \text{and} \quad 3(R \circ S)z$$

No other element of A is connected to an element of C. Accordingly,

$$R \circ S = \{(2, z), (3, x), (3, z)\}$$

Our first theorem tells us that composition of relations is associative.

Theorem 2.1: Let A, B, C and D be sets. Suppose R is a relation from A to B, S is a relation from B to C, and T is a relation from C to D. Then

$$(R \circ S) \circ T = R \circ (S \circ T)$$

We prove this theorem in Problem 2.8.

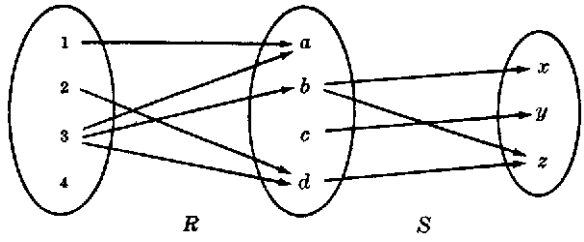


Fig. 2-4

Composition of Relations and Matrices

There is another way of finding $R \circ S$. Let M_R and M_S denote respectively the matrix representations of the relations R and S. Then

$$M_R = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{cccc} a & b & c & d \end{array} \\ \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \end{array} \quad \text{and} \quad M_S = \begin{array}{c} \\ a \\ b \\ c \\ d \end{array} \begin{array}{c} \begin{array}{ccc} x & y & z \end{array} \\ \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \end{array}$$

Multiplying M_R and M_S we obtain the matrix

$$M = M_R M_S = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{ccc} x & y & z \end{array} \\ \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 2 \\ 0 & 0 & 0 \end{bmatrix} \end{array}$$

The nonzero entries in this matrix tell us which elements are related by $R \circ S$. Thus $M = M_R M_S$ and $M_{R \circ S}$ have the same nonzero entries.

2.6 TYPES OF RELATIONS

This section discusses a number of important types of relations defined on a set A.

Reflexive Relations

A relation R on a set A is *reflexive* if aRa for every $a \in A$, that is, if $(a, a) \in R$ for every $a \in A$. Thus R is not reflexive if there exists $a \in A$ such that $(a, a) \notin R$.

EXAMPLE 2.5 Consider the following five relations on the set $A = \{1, 2, 3, 4\}$:

$$R_1 = \{(1, 1), (1, 2), (2, 3), (1, 3), (4, 4)\}$$
$$R_2 = \{(1, 1)(1, 2), (2, 1), (2, 2), (3, 3), (4, 4)\}$$
$$R_3 = \{(1, 3), (2, 1)\}$$
$$R_4 = \emptyset, \text{ the empty relation}$$
$$R_5 = A \times A, \text{ the universal relation}$$

Determine which of the relations are reflexive.

Since A contains the four elements 1, 2, 3, and 4, a relation R on A is reflexive if it contains the four pairs $(1, 1)$, $(2, 2)$, $(3, 3)$, and $(4, 4)$. Thus only R_2 and the universal relation $R_5 = A \times A$ are reflexive. Note that R_1, R_3, and R_4 are not reflexive since, for example, $(2, 2)$ does not belong to any of them.

EXAMPLE 2.6 Consider the following five relations:

(1) Relation $\leq$ (less than or equal) on the set $\mathbf{Z}$ of integers.

(2) Set inclusion $\subseteq$ on a collection C of sets.

(3) Relation $\perp$ (perpendicular) on the set L of lines in the plane.

(4) Relation $\parallel$ (parallel) on the set L of lines in the plane.

(5) Relation $|$ of divisibility on the set $\mathbf{N}$ of positive integers. (Recall $x \mid y$ if there exists z such that $xz = y$.)

Determine which of the relations are reflexive.

The relation (3) is not reflexive since no line is perpendicular to itself. Also (4) is not reflexive since no line is parallel to itself. The other relations are reflexive; that is, $x \preceq x$ for every $x \in \mathbf{Z}$, $A \subseteq A$ for any set $A \in C$, and $n \mid n$ for every positive integer $n \in \mathbf{N}$.

Symmetric and Antisymmetric Relations

A relation R on a set A is *symmetric* if whenever aRb then bRa, that is, if whenever $(a, b) \in R$ then $(b, a) \in R$. Thus R is not symmetric if there exists $a, b \in A$ such that $(a, b) \in R$ but $(b, a) \notin R$.

EXAMPLE 2.7

(a) Determine which of the relations in Example 2.5 are symmetric.

R_1 is not symmetric since $(1, 2) \in R_1$ but $(2, 1) \notin R_1$. R_3 is not symmetric since $(1, 3) \in R_3$ but $(3, 1) \notin R_3$. The other relations are symmetric.

(b) Determine which of the relations in Example 2.6 are symmetric.

The relation $\perp$ is symmetric since if line a is perpendicular to line b then b is perpendicular to a. Also, $\parallel$ is symmetric since if line a is parallel to line b then b is parallel to line a. The other relations are not symmetric. For example:

$$3 \leq 4 \text{ but } 4 \not\leq 3; \quad \{1, 2\} \subseteq \{1, 2, 3\} \text{ but } \{1, 2, 3\} \not\subseteq \{1, 2\}; \quad \text{and} \quad 2 \mid 6 \text{ but } 6 \nmid 2.$$

A relation R on a set A is *antisymmetric* if whenever aRb and bRa then $a = b$, that is, if $a \neq b$ and aRb then $b \not R a$. Thus R is not antisymmetric if there exist distinct elements a and b in A such that aRb and bRa.

EXAMPLE 2.8

(a) Determine which of the relations in Example 2.5 are antisymmetric.

R_2 is not antisymmetric since $(1, 2)$ and $(2, 1)$ belong to R_2, but $1 \neq 2$. Similarly, the universal relation R_3 is not antisymmetric. All the other relations are antisymmetric.

(b) Determine which of the relations in Example 2.6 are antisymmetric.

The relation $\leq$ is antisymmetric since whenever $a \leq b$ and $b \leq a$ then $a = b$. Set inclusion $\subseteq$ is antisymmetric since whenever $A \subseteq B$ and $B \subseteq A$ then $A = B$. Also, divisibility on $\mathbf{N}$ is antisymmetric since whenever $m \mid n$ and $n \mid m$ then $m = n$. (Note that divisibility on $\mathbf{Z}$ is not antisymmetric since $3 \mid -3$ and $-3 \mid 3$ but $3 \neq -3$.) The relations $\perp$ and $\parallel$ are not antisymmetric.

Remark: The properties of being symmetric and being antisymmetric are not negatives of each other. For example, the relation $R = \{(1, 3), (3, 1), (2, 3)\}$ is neither symmetric nor antisymmetric. On the other hand, the relation $R' = \{(1, 1), (2, 2)\}$ is both symmetric and antisymmetric.

Transitive Relations

A relation R on a set A is *transitive* if whenever aRb and bRc then aRc, that is, if whenever $(a, b), (b, c) \in R$ then $(a, c) \in R$. Thus R is not transitive if there exist $a, b, c \in R$ such that $(a, b), (b, c) \in R$ but $(a, c) \notin R$.

EXAMPLE 2.9

(a) Determine which of the relations in Example 2.5 are transitive.

The relation R_3 is not transitive since $(2, 1), (1, 3) \in R_3$ but $(2, 3) \notin R_3$. All the other relations are transitive.

(b) Determine which of the relations in Example 2.6 are transitive.

The relations $\leq$, $\subseteq$, and $|$ are transitive, but certainly not $\perp$. Also, since no line is parallel to itself, we can have $a \parallel b$ and $b \parallel a$, but $a \parallel a$. Thus $\parallel$ is not transitive. (We note that the relation "is parallel or equal to" is a transitive relation on the set L of lines in the plane.)

The property of transitivity can also be expressed in terms of the composition of relations. For a relation R on A we did define $R^2 = R \circ R$ and, more generally, $R^n = R^{n-1} \circ R$. Then we have the following result:

Theorem 2.2: A relation R is transitive if and only if, for every $n \geq 1$, we have $R^n \subseteq R$.

2.7 CLOSURE PROPERTIES

Consider a given set A and the collection of all relations on A. Let P be a property of such relations, such as being symmetric or being transitive. A relation with property P will be called a P-relation. The P-closure of an arbitrary relation R on A, written $P(R)$, is a P-relation such that

$$R \subseteq P(R) \subseteq S$$

for every P-relation S containing R. We will write

$$\text{reflexive}(R), \quad \text{symmetric}(R), \quad \text{and} \quad \text{transitive}(R)$$

for the reflexive, symmetric, and transitive closures of R.

Generally speaking, $P(R)$ need not exist. However, there is a general situation where $P(R)$ will always exist. Suppose P is a property such that there is at least one P-relation containing R and that the intersection of any P-relations is again a P-relation. Then one can prove (Problem 2.16) that

$$P(R) = \cap(S \mid S \text{ is a } P\text{-relation and } R \subseteq S)$$

Thus one can obtain $P(R)$ from the "top-down," that is, as the intersection of relations. However, one usually wants to find $P(R)$ from the "bottom-up," that is, by adjoining elements to R to obtain $P(R)$. This we do below.

Reflexive and Symmetric Closures

The next theorem tells us how to obtain easily the reflexive and symmetric closures of a relation. Here $\Delta_A = \{(a, a) \mid a \in A\}$ is the diagonal or equality relation on A.

Theorem 2.3: Let R be a relation on a set A. Then:

(i) $R \cup \Delta_A$ is the reflexive closure of R.

(ii) $R \cup R^{-1}$ is the symmetric closure of R.

In other words, reflexive(R) is obtained by simply adding to R those elements (a, a) in the diagonal which do not already belong to R, and symmetric(R) is obtained by adding to R all pairs (b, a) whenever (a, b) belongs to R.

EXAMPLE 2.10 Consider the relation $R = \{(1, 1), (1, 3), (2, 4), (3, 1), (3, 3), (4, 3)\}$ on the set $A = \{1, 2, 3, 4\}$. Then

$$\text{reflexive}(R) = R \cup \{(2, 2), (4, 4)\} \quad \text{and} \quad \text{symmetric}(R) = R \cup \{(4, 2), (3, 4)\}$$

Transitive Closure

Let R be a relation on a set A. Recall that $R^2 = R \circ R$ and $R^n = R^{n-1} \circ R$. We define

$$R^* = \bigcup_{i=1}^{\infty} R^i$$

The following theorem applies:

Theorem 2.4: R^* is the transitive closure of R.

Suppose A is a finite set with n elements. We show in Chapter 8 on graphs that

$$R^* = R \cup R^2 \cup \ldots \cup R^n$$

This gives us the following theorem:

Theorem 2.5: Let R be a relation on a set A with n elements. Then

$$\text{transitive}\,(R) = R \cup R^2 \cup \ldots \cup R^n$$

EXAMPLE 2.11 Consider the relation $R = \{(1, 2), (2, 3), (3, 3)\}$ on $A = \{1, 2, 3\}$. Then:

$$R^2 = R \circ R = \{(1, 3), (2, 3), (3, 3)\} \quad \text{and} \quad R^3 = R^2 \circ R = \{(1, 3), (2, 3), (3, 3)\}$$

Accordingly,

$$\text{transitive}\,(R) = \{(1, 2), (2, 3), (3, 3), (1, 3)\}$$

2.8 EQUIVALENCE RELATIONS

Consider a nonempty set S. A relation R on S is an *equivalence relation* if R is reflexive, symmetric, and transitive. That is, R is an equivalence relation on S if it has the following three properties:

(1) For every $a \in S$, aRa. (2) If aRb, then bRa. (3) If aRb and bRc, then aRc.

The general idea behind an equivalence relation is that it is a classification of objects which are in some way "alike." In fact, the relation "=" of equality on any set S is an equivalence relation; that is:

(1) $a = a$ for every $a \in S$. (2) If $a = b$, then $b = a$. (3) If $a = b$, $b = c$, then $a = c$.

Other equivalence relations follow.

EXAMPLE 2.12

(a) Let L be the set of lines and let T be the set of triangles in the Euclidean plane.

(i) The relation "is parallel to or identical to" is an equivalence relation on L.

(ii) The relations of congruence and similarity are equivalence relations on T.

(b) The relation $\subseteq$ of set inclusion is not an equivalence relation. It is reflexive and transitive, but it is not symmetric since $A \subseteq B$ does not imply $B \subseteq A$.

(c) Let m be a fixed positive integer. Two integers a and b are said to be *congruent modulo* m, written

$$a \equiv b \pmod{m}$$

if m divides $a - b$. For example, for the modulus $m = 4$, we have

$$11 \equiv 3 \pmod{4} \quad \text{and} \quad 22 \equiv 6 \pmod{4}$$

since 4 divides $11 - 3 = 8$ and 4 divides $22 - 6 = 16$. This relation of congruence modulo m is an important equivalence relation.

Equivalence Relations and Partitions

This subsection explores the relationship between equivalence relations and partitions on a non-empty set S. Recall first that a partition P of S is a collection $\{A_i\}$ of nonempty subsets of S with the following two properties:

(1) Each $a \in S$ belongs to some A_i.
(2) If $A_i \neq A_j$ then $A_i \cap A_j = \emptyset$.

In other words, a partition P of S is a subdivision of S into disjoint nonempty sets. (See Section 1.7.)

Suppose R is an equivalence relation on a set S. For each $a \in S$, let [a] denote the set of elements of S to which a is related under R; that is:

$$[a] = \{x \mid (a, x) \in R\}$$

We call $[a]$ the *equivalence class* of a in S; any $b \in [a]$ is called a *representative* of the equivalence class.

The collection of all equivalence classes of elements of S under an equivalence relation R is denoted by S/R, that is,

$$S/R = \{[a] \mid a \in S\}$$

It is called the *quotient set* of S by R. The fundamental property of a quotient set is contained in the following theorem.

Theorem 2.6: Let R be an equivalence relation on a set S. Then S/R is a partition of S. Specifically:

(i) For each a in S, we have $a \in [a]$.
(ii) $[a] = [b]$ if and only if $(a, b) \in R$.
(iii) If $[a] \neq [b]$, then $[a]$ and $[b]$ are disjoint.

Conversely, given a partition $\{A_i\}$ of the set S, there is an equivalence relation R on S such that the sets A_i are the equivalence classes.

This important theorem will be proved in Problem 2.17.

EXAMPLE 2.13

(a) Consider the relation $R = \{(1, 1), (1, 2), (2, 1), (2, 2), (3, 3)\}$ on $S = \{1, 2, 3\}$.

One can show that R is reflexive, symmetric, and transitive, that is, that R is an equivalence relation. Also:

$$[1] = \{1, 2\}, [2] = \{1, 2\}, [3] = \{3\}$$

Observe that $[1] = [2]$ and that $S/R = \{[1], [3]\}$ is a partition of S. One can choose either $\{1, 3\}$ or $\{2, 3\}$ as a set of representatives of the equivalence classes.

(b) Let R_5 be the relation of congruence modulo 5 on the set $\mathbf{Z}$ of integers denoted by

$$x \equiv y \pmod 5$$

This means that the difference $x - y$ is divisible by 5. Then R_5 is an equivalence relation on $\mathbf{Z}$. The quotient set $\mathbf{Z}/R_5$ contains the following five equivalence classes:

$$\begin{aligned} A_0 &= \{\ldots, -10, -5, 0, 5, 10, \ldots\} \\ A_1 &= \{\ldots, -9, -4, 1, 6, 11, \ldots\} \\ A_2 &= \{\ldots, -8, -3, 2, 7, 12, \ldots\} \\ A_3 &= \{\ldots, -7, -2, 3, 8, 13, \ldots\} \\ A_4 &= \{\ldots, -6, -1, 4, 9, 14, \ldots\} \end{aligned}$$

Any integer x, uniquely expressed in the form $x = 5q + r$ where $0 \leq r < 5$, is a member of the equivalence class A_r, where r is the remainder. As expected, $\mathbf{Z}$ is the disjoint union of equivalence classes A_1, A_2, A_3, A_4. Usually one chooses $\{0, 1, 2, 3, 4\}$ or $\{-2, -1, 0, 1, 2\}$ as a set of representatives of the equivalence classes.

2.9 PARTIAL ORDERING RELATIONS

A relation R on a set S is called a *partial ordering* or a *partial order* of S if R is reflexive, antisymmetric, and transitive. A set S together with a partial ordering R is called a *partially ordered set* or *poset*. Partially ordered sets will be studied in more detail in Chapter 14, so here we simply give some examples.

EXAMPLE 2.14

(a) The relation $\subseteq$ of set inclusion is a partial ordering on any collection of sets since set inclusion has the three desired properties. That is,

(1) $A \subseteq A$ for any set A.
(2) If $A \subseteq B$ and $B \subseteq A$, then $A = B$.
(3) If $A \subseteq B$ and $B \subseteq C$, then $A \subseteq C$.

(b) The relation $\leq$ on the set $\mathbf{R}$ of real numbers is reflexive, antisymmetric, and transitive. Thus $\leq$ is a partial ordering on $\mathbf{R}$.

(c) The relation "a divides b," written $a \mid b$, is a partial ordering on the set $\mathbf{N}$ of positive integers. However, "a divides b" is not a partial ordering on the set $\mathbf{Z}$ of integers since $a \mid b$ and $b \mid a$ need not imply $a = b$. For example, $3 \mid -3$ and $-3 \mid 3$ but $3 \neq -3$.

2.10 n-ARY RELATIONS

All the relations discussed above were binary relations. By an *n-ary relation*, we mean a set of ordered n-tuples. For any set S, a subset of the product set S^n is called an n-ary relation on S. In particular, a subset of S^3 is called a *ternary relation* on S.

EXAMPLE 2.15

(a) Let L be a line in the plane. Then "betweenness" is a ternary relation R on the points of L; that is, $(a, b, c) \in R$ if b lies between a and c on L.

(b) The equation $x^2 + y^2 + z^2 = 1$ determines a ternary relation T on the set $\mathbf{R}$ of real numbers. That is, a triple (x, y, z) belongs to T if (x, y, z) satisfies the equation, which means (x, y, z) is the coordinates of a point in $\mathbf{R}^3$ on the sphere S with radius 1 and center at the origin $O = (0, 0, 0)$.

Solved Problems

PRODUCT SETS

2.1. Given: $A = \{1, 2\}$, $B = \{x, y, z\}$, and $C = \{3, 4\}$. Find: $A \times B \times C$.

$A \times B \times C$ consists of all ordered triplets (a, b, c) where $a \in A$, $b \in B$, $c \in C$. These elements of $A \times B \times C$ can be systematically obtained by a so-called tree diagram (Fig. 2-5). The elements of $A \times B \times C$ are precisely the 12 ordered triplets to the right of the tree diagram.

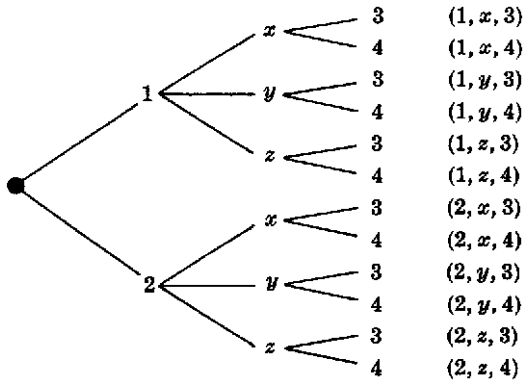


Fig. 2-5

Observe that $n(A) = 2$, $n(B) = 3$, and $n(C) = 2$ and, as expected,

$$n(A \times B \times C) = 12 = n(A) \cdot n(B) \cdot n(C)$$

2.2. Find x and y given $(2x, x + y) = (6, 2)$.

Two ordered pairs are equal if and only if the corresponding components are equal. Hence we obtain the equations

$$2x = 6 \quad \text{and} \quad x + y = 2$$

from which we derive the answers $x = 3$ and $y = -1$.

RELATIONS AND THEIR GRAPHS

2.3. Find the number of relations from $A = \{a, b, c\}$ to $B = \{1, 2\}$.

There are $3(2) = 6$ elements in $A \times B$, and hence there are $m = 2^6 = 64$ subsets of $A \times B$. Thus there are $m = 64$ relations from A to B.

2.4. Given $A = \{1, 2, 3, 4\}$ and $B = \{x, y, z\}$. Let R be the following relation from A to B:

$$R = \{(1, y), (1, z), (3, y), (4, x), (4, z)\}$$

(a) Determine the matrix of the relation.

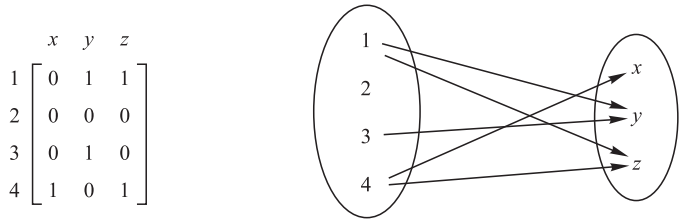
(b) Draw the arrow diagram of R.

(c) Find the inverse relation R^{-1} of R.

(d) Determine the domain and range of R.

(a) See Fig. 2-6(a) Observe that the rows of the matrix are labeled by the elements of A and the columns by the elements of B. Also observe that the entry in the matrix corresponding to $a \in A$ and $b \in B$ is 1 if a is related to b and 0 otherwise.

(b) See Fig. 2.6(b) Observe that there is an arrow from $a \in A$ to $b \in B$ iff a is related to b, i.e., iff $(a, b) \in R$.



(*a*) (*b*)

Fig. 2-6

(*c*) Reverse the ordered pairs of R to obtain R^{-1}:

$$R^{-1} = \{(y, 1), (z, 1), (y, 3), (x, 4), (z, 4)\}$$

Observe that by reversing the arrows in Fig. 2.6(*b*), we obtain the arrow diagram of R^{-1}.

(*d*) The domain of R, $\mathrm{Dom}(R)$, consists of the first elements of the ordered pairs of R, and the range of R, $\mathrm{Ran}(R)$, consists of the second elements. Thus,

$$\mathrm{Dom}(R) = \{1, 3, 4\} \quad \text{and} \quad \mathrm{Ran}(R) = \{x, y, z\}$$

2.5. Let $A = \{1, 2, 3\}$, $B = \{a, b, c\}$, and $C = \{x, y, z\}$. Consider the following relations R and S from A to B and from B to C, respectively.

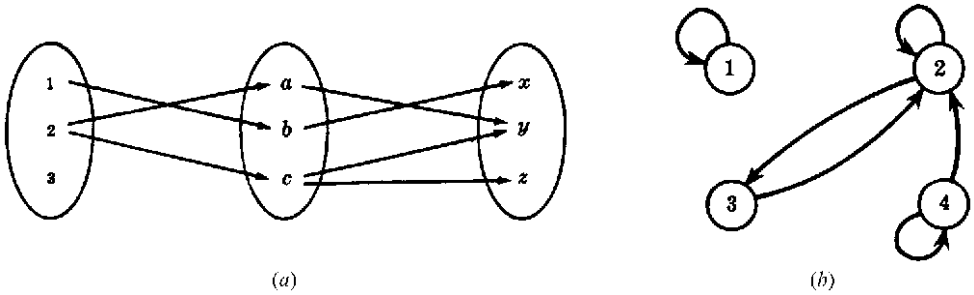
$$R = \{(1, b), (2, a), (2, c)\} \quad \text{and} \quad S = \{(a, y), (b, x), (c, y), (c, z)\}$$

(*a*) Find the composition relation $R \circ S$.

(*b*) Find the matrices M_R, M_S, and $M_{R \circ S}$ of the respective relations R, S, and $R \circ S$, and compare $M_{R \circ S}$ to the product $M_R M_S$.

(*a*) Draw the arrow diagram of the relations R and S as in Fig. 2-7(*a*). Observe that 1 in A is "connected" to x in C by the path $1 \to b \to x$; hence $(1, x)$ belongs to $R \circ S$. Similarly, $(2, y)$ and $(2, z)$ belong to $R \circ S$.
We have

$$R \circ S = \{(1, x), (2, y), (2, z)\}$$



(*a*) (*b*)

Fig. 2-7

(*b*) The matrices of M_R, M_S, and $M_{R\circ S}$ follow:

$$M_R = \begin{array}{c} \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{ccc} a & b & c \end{array} \\ \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \end{array} \quad M_S = \begin{array}{c} \\ a \\ b \\ c \end{array} \begin{array}{c} \begin{array}{ccc} x & y & z \end{array} \\ \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 1 \end{bmatrix} \end{array} \quad M_{R\circ S} = \begin{array}{c} \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{ccc} x & y & z \end{array} \\ \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix} \end{array}$$

Multiplying M_R and M_S we obtain

$$M_R M_S = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

Observe that $M_{R\circ S}$ and $M_R M_S$ have the same zero entries.

2.6. Consider the relation $R = \{(1, 1), (2, 2), (2, 3), (3, 2), (4, 2), (4, 4)\}$ on $A = \{1, 2, 3, 4\}$.

(*a*) Draw its directed graph. (*b*) Find $R^2 = R\circ R$.

(*a*) For each $(a, b) \in R$, draw an arrow from a to b as in Fig. 2-7(*b*).

(*b*) For each pair $(a, b) \in R$, find all $(b, c) \in R$. Then $(a, c) \in R^2$. Thus

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

2.7. Let R and S be the following relations on $A = \{1, 2, 3\}$:

$$R = \{(1, 1), (1, 2), (2, 3), (3, 1), (3, 3)\}, \quad S = \{(1, 2), (1, 3), (2, 1), (3, 3)\}$$

Find (*a*) $R \cup S$, $R \cap S$, R^{C}; (*b*) $R\circ S$; (*c*) $S^2 = S\circ S$.

(*a*) Treat R and S simply as sets, and take the usual intersection and union. For R^{C}, use the fact that $A \times A$ is the universal relation on A.

$$\begin{aligned} R \cap S &= \{(1, 2), (3, 3)\} \\ R \cup S &= \{(1, 1), (1, 2), (1, 3), (2, 1), (2, 3), (3, 1), (3, 3)\} \\ R^{\mathrm{C}} &= \{(1, 3), (2, 1), (2, 2), (3, 2)\} \end{aligned}$$

(*b*) For each pair $(a, b) \in R$, find all pairs $(b, c) \in S$. Then $(a, c) \in R\circ S$. For example, $(1, 1) \in R$ and $(1, 2)$, $(1, 3) \in S$; hence $(1, 2)$ and $(1, 3)$ belong to $R\circ S$. Thus,

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

(*c*) Following the algorithm in (*b*), we get

$$S^2 = S\circ S = \{(1, 1), (1, 3), (2, 2), (2, 3), (3, 3)\}$$

2.8. Prove Theorem 2.1: Let A, B, C and D be sets. Suppose R is a relation from A to B, S is a relation from B to C and T is a relation from C to D. Then $(R\circ S)\circ T = R\circ(S\circ T)$.

We need to show that each ordered pair in $(R\circ S)\circ T$ belongs to $R\circ(S\circ T)$, and vice versa.

Suppose (a, d) belongs to $(R\circ S)\circ T$. Then there exists $c \in C$ such that $(a, c) \in R \circ S$ and $(c, d) \in T$. Since $(a, c) \in R\circ S$, there exists $b \in B$ such that $(a, b) \in R$ and $(b, c) \in S$. Since $(b, c) \in S$ and $(c, d) \in T$, we have $(b, d) \in S\circ T$; and since $(a, b) \in R$ and $(b, d) \in S\circ T$, we have $(a, d) \in R \circ (S\circ T)$. Therefore, $(R\circ S)\circ T \subseteq R\circ(S\circ T)$. Similarly $R\circ(S\circ T) \subseteq (R\circ S)\circ T$. Both inclusion relations prove $(R\circ S)\circ T = R\circ(S\circ T)$.

TYPES OF RELATIONS AND CLOSURE PROPERTIES

2.9. Consider the following five relations on the set $A = \{1, 2, 3\}$:

$$R = \{(1, 1), (1, 2), (1, 3), (3, 3)\}, \qquad \emptyset = \text{empty relation}$$
$$S = \{(1, 1)(1, 2), (2, 1)(2, 2), (3, 3)\}, \qquad A \times A = \text{universal relation}$$
$$T = \{(1, 1), (1, 2), (2, 2), (2, 3)\}$$

Determine whether or not each of the above relations on A is: (a) reflexive; (b) symmetric; (c) transitive; (d) antisymmetric.

(a) R is not reflexive since $2 \in A$ but $(2, 2) \notin R$. T is not reflexive since $(3, 3) \notin T$ and, similarly, $\emptyset$ is not reflexive. S and $A \times A$ are reflexive.

(b) R is not symmetric since $(1, 2) \in R$ but $(2, 1) \notin R$, and similarly T is not symmetric. S, $\emptyset$, and $A \times A$ are symmetric.

(c) T is not transitive since $(1, 2)$ and $(2, 3)$ belong to T, but $(1, 3)$ does not belong to T. The other four relations are transitive.

(d) S is not antisymmetric since $1 \neq 2$, and $(1, 2)$ and $(2, 1)$ both belong to S. Similarly, $A \times A$ is not antisymmetric. The other three relations are antisymmetric.

2.10. Give an example of a relation R on $A = \{1, 2, 3\}$ such that:

(a) R is both symmetric and antisymmetric.

(b) R is neither symmetric nor antisymmetric.

(c) R is transitive but $R \cup R^{-1}$ is not transitive.

There are several such examples. One possible set of examples follows:

(a) $R = \{(1, 1), (2, 2)\}$; (b) $R = \{(1, 2), (2, 3)\}$; (c) $R = \{(1, 2)\}$.

2.11. Suppose C is a collection of relations S on a set A, and let T be the intersection of the relations S in C, that is, $T = \cap(S \mid S \in C)$ Prove:

(a) If every S is symmetric, then T is symmetric.

(b) If every S is transitive, then T is transitive.

(a) Suppose $(a, b) \in T$. Then $(a, b) \in S$ for every S. Since each S is symmetric, $(b, a) \in S$ for every S. Hence $(b, a) \in T$ and T is symmetric.

(b) Suppose (a, b) and (b, c) belong to T. Then (a, b) and (b, c) belong to S for every S. Since each S is transitive, (a, c) belongs to S for every S. Hence, $(a, c) \in T$ and T is transitive.

2.12. Let R be a relation on a set A, and let P be a property of relations, such as symmetry and transitivity. Then P will be called *R-closable* if P satisfies the following two conditions:

(1) There is a P-relation S containing R.

(2) The intersection of P-relations is a P-relation.

(a) Show that symmetry and transitivity are R-closable for any relation R.

(b) Suppose P is R-closable. Then $P(R)$, the P-closure of R, is the intersection of all P-relations S containing R, that is,

$$P(R) = \cap(S \mid S \text{ is a } P\text{-relation and } R \subseteq S)$$

(a) The universal relation $A \times A$ is symmetric and transitive and $A \times A$ contains any relation R on A. Thus (1) is satisfied. By Problem 2.11, symmetry and transitivity satisfy (2). Thus symmetry and transitivity are R-closable for any relation R.

(*b*) Let $T = \cap(S \mid S$ is a P-relation and $R \subseteq S)$. Since P is R-closable, T is nonempty by (1) and T is a P-relation by (2). Since each relation S contains R, the intersection T contains R. Thus, T is a P-relation containing R. By definition, $P(R)$ is the smallest P-relation containing R; hence $P(R) \subseteq T$. On the other hand, $P(R)$ is one of the sets S defining T, that is, $P(R)$ is a P-relation and if $R \subseteq P(R)$. Therefore, $T \subseteq P(R)$. Accordingly, $P(R) = T$.

2.13. Consider the relation $R = \{(a, a), (a, b), (b, c), (c, c)\}$ on the set $A = \{a, b, c\}$. Find: (*a*) reflexive(R); (*b*) symmetric(R); (*c*) transitive(R).

(*a*) The reflexive closure on R is obtained by adding all diagonal pairs of $A \times A$ to R which are not currently in R. Hence,

$$\text{reflexive}(R) = R \cup \{(b, b)\} = \{(a, a), (a, b), (b, b), (b, c), (c, c)\}$$

(*b*) The symmetric closure on R is obtained by adding all the pairs in R^{-1} to R which are not currently in R. Hence,

$$\text{symmetric}(R) = R \cup \{(b, a), (c, b)\} = \{(a, a), (a, b), (b, a), (b, c), (c, b), (c, c)\}$$

(*c*) The transitive closure on R, since A has three elements, is obtained by taking the union of R with $R^2 = R \circ R$ and $R^3 = R \circ R \circ R$. Note that

$$R^2 = R \circ R = \{(a, a), (a, b), (a, c), (b, c), (c, c)\}$$
$$R^3 = R \circ R \circ R = \{(a, a), (a, b), (a, c), (b, c), (c, c)\}$$

Hence

$$\text{transitive}(R) = R \cup R^2 \cup R^3 = \{(a, a), (a, b), (a, c), (b, c), (c, c)\}$$

EQUIVALENCE RELATIONS AND PARTITIONS

2.14. Consider the $\mathbf{Z}$ of integers and an integer $m > 1$. We say that x is congruent to y modulo m, written

$$x \equiv y \pmod{m}$$

if $x - y$ is divisible by m. Show that this defines an equivalence relation on $\mathbf{Z}$.

We must show that the relation is reflexive, symmetric, and transitive.

(i) For any x in Z we have $x \equiv x \pmod{m}$ because $x - x = 0$ is divisible by m. Hence the relation is reflexive.

(ii) Suppose $x \equiv y \pmod{m}$, so $x - y$ is divisible by m. Then $-(x - y) = y - x$ is also divisible by m, so $y \equiv x \pmod{m}$. Thus the relation is symmetric.

(iii) Now suppose $x \equiv y \pmod{m}$ and $y \equiv z \pmod{m}$, so $x - y$ and $y - z$ are each divisible by m. Then the sum

$$(x - y) + (y - z) = x - z$$

is also divisible by m; hence $x \equiv z \pmod{m}$. Thus the relation is transitive.

Accordingly, the relation of congruence modulo m on $\mathbf{Z}$ is an equivalence relation.

2.15. Let A be a set of nonzero integers and let $\approx$ be the relation on $A \times A$ defined by

$$(a, b) \approx (c, d) \quad \text{whenever} \quad ad = bc$$

Prove that $\approx$ is an equivalence relation.

We must show that $\approx$ is reflexive, symmetric, and transitive.

(i) *Reflexivity*: We have $(a, b) \approx (a, b)$ since $ab = ba$. Hence $\approx$ is reflexive.

(ii) *Symmetry*: Suppose $(a, b) \approx (c, d)$. Then $ad = bc$. Accordingly, $cb = da$ and hence $(c, d) = (a, b)$. Thus, $\approx$ is symmetric.

(iii) *Transitivity*: Suppose $(a, b) \approx (c, d)$ and $(c, d) \approx (e, f)$. Then $ad = bc$ and $cf = de$. Multiplying corresponding terms of the equations gives $(ad)(cf) = (bc)(de)$. Canceling $c \neq 0$ and $d \neq 0$ from both sides of the equation yields $af = be$, and hence $(a, b) \approx (e, f)$. Thus $\approx$ is transitive. Accordingly, $\approx$ is an equivalence relation.

2.16. Let R be the following equivalence relation on the set $A = \{1, 2, 3, 4, 5, 6\}$:

$$R = \{(1, 1), (1, 5), (2, 2), (2, 3), (2, 6), (3, 2), (3, 3), (3, 6), (4, 4), (5, 1), (5, 5), (6, 2), (6, 3), (6, 6)\}$$

Find the partition of A induced by R, i.e., find the equivalence classes of R.

Those elements related to 1 are 1 and 5 hence

$$[1] = \{1, 5\}$$

We pick an element which does not belong to [1], say 2. Those elements related to 2 are 2, 3, and 6, hence

$$[2] = \{2, 3, 6\}$$

The only element which does not belong to [1] or [2] is 4. The only element related to 4 is 4. Thus

$$[4] = \{4\}$$

Accordingly, the following is the partition of A induced by R:

$$[\{1, 5\}, \{2, 3, 6\}, \{4\}]$$

2.17. Prove Theorem 2.6: Let R be an equivalence relation in a set A. Then the quotient set A/R is a partition of A. Specifically,

(i) $a \in [a]$, for every $a \in A$.

(ii) $[a] = [b]$ if and only if $(a, b) \in R$.

(iii) If $[a] \neq [b]$, then $[a]$ and $[b]$ are disjoint.

(*a*) *Proof of (i)*: Since R is reflexive, $(a, a) \in R$ for every $a \in A$ and therefore $a \in [a]$.

(*b*) *Proof of (ii)*: Suppose $(a, b) \in R$. We want to show that $[a] = [b]$. Let $x \in [b]$; then $(b, x) \in R$. But by hypothesis $(a, a) \in R$ and so, by transitivity, $(a, x) \in R$. Accordingly $x \in [a]$. Thus $[b] \subseteq [a]$. To prove that $[a] \subseteq [b]$ we observe that $(a, b) \in R$ implies, by symmetry, that $(b, a) \in R$. Then, by a similar argument, we obtain $[a] \subseteq [b]$. Consequently, $[a] = [b]$.

On the other hand, if $[a] = [b]$, then, by (i), $b \in [b] = [a]$; hence $(a, b) \in R$.

(*c*) *Proof of (iii)*: We prove the equivalent contrapositive statement:

$$\text{If } [a] \cap [b] \neq \emptyset \quad \text{then} \quad [a] = [b]$$

If $[a] \cap [b] \neq \emptyset$, then there exists an element $x \in A$ with $x \in [a] \cap [b]$. Hence $(a, x) \in R$ and $(b, x) \in R$. By symmetry, $(x, b) \in R$ and by transitivity, $(a, b) \in R$. Consequently by (ii), $[a] = [b]$.

PARTIAL ORDERINGS

2.18. Let ℓ be any collection of sets. Is the relation of set inclusion $\subseteq$ a partial order on ℓ?

Yes, since set inclusion is reflexive, antisymmetric, and transitive. That is, for any sets A, B, C in ℓ we have: (i) $A \subseteq A$; (ii) if $A \subseteq B$ and $B \subseteq A$, then $A = B$; (iii) if $A \subseteq B$ and $B \subseteq C$, then $A \subseteq C$.

2.19. Consider the set $\mathbf{Z}$ of integers. Define aRb by $b = a^r$ for some positive integer r. Show that R is a partial order on $\mathbf{Z}$, that is, show that R is: (*a*) reflexive; (*b*) antisymmetric; (*c*) transitive.

(*a*) R is reflexive since $a = a^1$.

(*b*) Suppose aRb and bRa, say $b = a^r$ and $a = b^s$. Then $a = (a^r)^s = a^{rs}$. There are three possibilities: (i) $rs = 1$, (ii) $a = 1$, and (iii) $a = -1$. If $rs = 1$ then $r = 1$ and $s = 1$ and so $a = b$. If $a = 1$ then $b = 1^r = 1 = a$, and, similarly, if $b = 1$ then $a = 1$. Lastly, if $a = -1$ then $b = -1$ (since $b \neq 1$) and $a = b$. In all three cases, $a = b$. Thus R is antisymmetric.

(*c*) Suppose aRb and bRc say $b = a^r$ and $c = b^s$. Then $c = (a^r)^s = a^{rs}$ and, therefore, aRc. Hence R is transitive.

Accordingly, R is a partial order on $\mathbf{Z}$.

Supplementary Problems

RELATIONS

2.20. Let $S = \{a, b, c\}$, $T = \{b, c, d\}$, and $W = \{a, d\}$. Find $S \times T \times W$.

2.21. Find x and y where: (a) $(x + 2, 4) = (5, 2x + y)$; (b) $(y - 2, 2x + 1) = (x - 1, y + 2)$.

2.22. Prove: (a) $A \times (B \cap C) = (A \times B) \cap (A \times C)$; (b) $A \times (B \cup C) = (A \times B) \cup (A \times C)$.

2.23. Consider the relation $R = \{(1, 3), (1, 4), (3, 2), (3, 3), (3, 4)\}$ on $A = \{1, 2, 3, 4\}$.

(a) Find the matrix M_R of R.

(b) Find the domain and range of R.

(c) Find R^{-1}.

(d) Draw the directed graph of R.

(e) Find the composition relation $R \circ R$.

(f) Find $R \circ R^{-1}$ and $R^{-1} \circ R$.

2.24. Let $A = \{1, 2, 3, 4\}$, $B = \{a, b, c\}$, $C = \{x, y, z\}$. Consider the relations R from A to B and S from B to C as follows:

$$R = \{(1, b), (3, a), (3, b), (4, c)\} \quad \text{and} \quad S = \{(a, y), (c, x), (a, z)\}$$

(a) Draw the diagrams of R and S.

(b) Find the matrix of each relation R, S (composition) $R \circ S$.

(c) Write R^{-1} and the composition $R \circ S$ as sets of ordered pairs.

2.25. Let R and S be the following relations on $B = \{a, b, c, d\}$:

$$R = \{(a, a), (a, c), (c, b), (c, d), (d, b)\} \quad \text{and} \quad S = \{(b, a), (c, c), (c, d), (d, a)\}$$

Find the following composition relations: (a) $R \circ S$; (b) $S \circ R$; (c) $R \circ R$; (d) $S \circ S$.

2.26. Let R be the relation on $\mathbf{N}$ defined by $x + 3y = 12$, i.e. $R = \{(x, y) \mid x + 3y = 12\}$.

(a) Write R as a set of ordered pairs.

(b) Find the domain and range of R.

(c) Find R^{-1}.

(d) Find the composition relation $R \circ R$.

PROPERTIES OF RELATIONS

2.27. Each of the following defines a relation on the positive integers $\mathbf{N}$:

(1) "x is greater than y."

(2) "xy is the square of an integer."

(3) $x + y = 10$

(4) $x + 4y = 10$.

Determine which of the relations are: (a) reflexive; (b) symmetric; (c) antisymmetric; (d) transitive.

2.28. Let R and S be relations on a set A. Assuming A has at least three elements, state whether each of the following statements is true or false. If it is false, give a counterexample on the set $A = \{1, 2, 3\}$:

(a) If R and S are symmetric then $R \cap S$ is symmetric.

(b) If R and S are symmetric then $R \cup S$ is symmetric.

(c) If R and S are reflexive then $R \cap S$ is reflexive.

(d) If R and S are reflexive then $R \cup S$ is reflexive.

(e) If R and S are transitive then $R \cup S$ is transitive.

(f) If R and S are antisymmetric then $R \cup S$ is antisymmetric.

(g) If R is antisymmetric, then R^{-1} is antisymmetric.

(h) If R is reflexive then $R \cap R^{-1}$ is not empty.

(i) If R is symmetric then $R \cap R^{-1}$ is not empty.

2.29. Suppose R and S are relations on a set A, and R is antisymmetric. Prove that $R \cap S$ is antisymmetric.

EQUIVALENCE RELATIONS

2.30. Prove that if R is an equivalence relation on a set A, than R^{-1} is also an equivalence relation on A.

2.31. Let $S = \{1, 2, 3, \ldots, 18, 19\}$. Let R be the relation on S defined by "xy is a square," (a) Prove R is an equivalence relation. (b) Find the equivalence class [1]. (c) List all equivalence classes with more than one element.

2.32. Let $S = \{1, 2, 3, \ldots, 14, 15\}$. Let R be the equivalence relation on S defined by $x \equiv y \pmod 5$, that is, $x - y$ is divisible by 5. Find the partition of S induced by R, i.e. the quotient set S/R.

2.33. Let $S = \{1, 2, 3, \ldots, 9\}$, and let $\sim$ be the relation on $A \times A$ defined by

$$(a, b) \sim (c, d) \quad \text{whenever} \quad a + d = b + c.$$

(a) Prove that $\sim$ is an equivalence relation.

(b) Find $[(2, 5)]$, that is, the equivalence class of $(2, 5)$.

Answers to Supplementary Problems

2.20. $\{(a, b, a), (a, b, d), (a, c, a), (a, c, d), (a, d, a), (a, d, d), (b, b, a), (b, b, d), (b, c, a), (b, c, d), (b, d, a), (b, d, d), (c, b, a), (c, b, d), (c, c, a), (c, c, d), (c, d, a), (c, d, d)\}$

2.21. (a) $x = 3$, $y = -2$; (b) $x = 2$, $y = 3$.

2.23. (a) $M_R = [0, 0, 1, 1;\ 0, 0, 0, 0;\ 0, 1, 1, 1;\ 0, 0, 0, 0]$;

(b) Domain $= \{1, 3\}$, range $= \{2, 3, 4\}$;

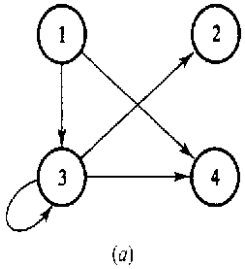
(c) $R^{-1} = \{(3, 1), (4, 1), (2, 3), (3, 3), (4, 3)\}$;

(d) See Fig. 2-8(a);

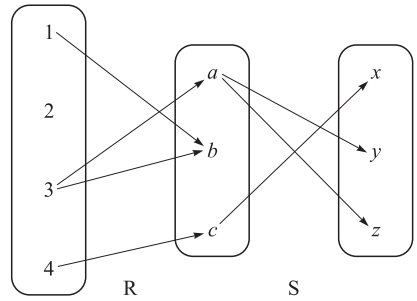
(e) $R \circ R = \{(1, 2), (1, 3), (1, 4), (3, 2), (3, 3), (3, 4)\}$.

2.24. (a) See Fig. 2-8(b);
(b) $R = [0, 1, 0;\ 0, 0, 0;\ 1, 1, 0;\ 0, 0, 1]$,
$S = [0, 1, 1;\ 0, 0, 0;\ 1, 0, 0]$,
$R \circ S = [0, 0, 0;\ 0, 0, 0;\ 0, 1, 1;\ 1, 0, 0]$;
(c) $\{(b, 1), (a, 3), (b, 3), (c, 4)\}$, $\{(3, y), (3, z), (4, x)\}$.

2.25. (a) $R \circ S = \{(a, c), (a, d), (c, a), (d, a)\}$
(b) $S \circ R = \{(b, a), (b, c), (c, b), (c, d), (d, a), (d, c)\}$
(c) $R \circ R = \{(a, a), (a, b), (a, c), (a, d), (c, b)\}$
(d) $S \circ S = \{(c, c), (c, a), (c, d)\}$



(a)



(b)

Fig. 2-8

2.26. (a) $\{(9, 1), (6, 2), (3, 3)\}$; (b) (i) $\{9, 6, 3\}$, (ii) $\{1, 2, 3\}$, (iii) $\{(1, 9), (2, 6), (3, 3)\}$; (c) $\{(3, 3)\}$.

2.27. (a) None; (b) (2) and (3); (c) (1) and (4); (d) all except (3).

2.28. All are true except: (e) $R = \{(1, 2)\}$, $S = \{(2, 3)\}$; (f) $R = \{(1, 2)\}$, $S = \{(2, 1)\}$.

2.31. (b) $\{1, 4, 9, 16\}$; (c) $\{1, 4, 9, 16\}$, $\{2, 8, 18\}$, $\{3, 12\}$.

2.32. [$\{1, 6, 11\}$, $\{2, 7, 12\}$, $\{3, 8, 133\}$, $\{4, 9, 14\}$, $\{5, 10, 15\}$]

2.33. (b) $\{(1, 4), (2, 5), (3, 6), (4, 7), (5, 8), (6, 9)\}$.

CHAPTER 3

Functions and Algorithms

3.1 INTRODUCTION

One of the most important concepts in mathematics is that of a function. The terms "map," "mapping," "transformation," and many others mean the same thing; the choice of which word to use in a given situation is usually determined by tradition and the mathematical background of the person using the term.

Related to the notion of a function is that of an algorithm. The notation for presenting an algorithm and a discussion of its complexity is also covered in this chapter.

3.2 FUNCTIONS

Suppose that to each element of a set A we assign a unique element of a set B; the collection of such assignments is called *a function* from A into B. The set A is called the *domain* of the function, and the set B is called the *target set* or *codomain*.

Functions are ordinarily denoted by symbols. For example, let f denote a function from A into B. Then we write

$$f\colon A \to B$$

which is read: "f is a function from A into B," or "f takes (or maps) A into B." If $a \in A$, then $f(a)$ (read: "f of a") denotes the unique element of B which f assigns to a; it is called the *image* of a under f, or the *value* of f at a. The set of all image values is called the *range* or *image* of f. The image of $f\colon A \to B$ is denoted by $\mathrm{Ran}(f)$, $\mathrm{Im}(f)$ or $f(A)$.

Frequently, a function can be expressed by means of a mathematical formula. For example, consider the function which sends each real number into its square. We may describe this function by writing

$$f(x) = x^2 \quad \text{or} \quad x \mapsto x^2 \quad \text{or} \quad y = x^2$$

In the first notation, x is called a *variable* and the letter f denotes the function. In the second notation, the barred arrow $\mapsto$ is read "goes into." In the last notation, x is called the *independent variable* and y is called the *dependent variable* since the value of y will depend on the value of x.

Remark: Whenever a function is given by a formula in terms of a variable x, we assume, unless it is otherwise stated, that the domain of the function is $\mathbf{R}$ (or the largest subset of $\mathbf{R}$ for which the formula has meaning) and the codomain is $\mathbf{R}$.

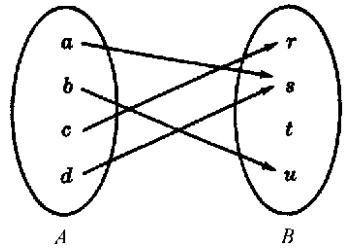


Fig. 3-1

EXAMPLE 3.1

(a) Consider the function $f(x) = x^3$, i.e., f assigns to each real number its cube. Then the image of 2 is 8, and so we may write $f(2) = 8$.

(b) Figure 3-1 defines a function f from $A = \{a, b, c, d\}$ into $B = \{r, s, t, u\}$ in the obvious way. Here

$$f(a) = s, \quad f(b) = u, \quad f(c) = r, \quad f(d) = s$$

The image of f is the set of image values, $\{r, s, u\}$. Note that t does not belong to the image of f because t is not the image of any element under f.

(c) Let A be any set. The function from A into A which assigns to each element in A the element itself is called the *identity function* on A and it is usually denoted by 1_A, or simply 1. In other words, for every $a \in A$,

$$1_A(a) = a.$$

(d) Suppose S is a subset of A, that is, suppose $S \subseteq A$. The *inclusion map* or *embedding* of S into A, denoted by $i\colon S \hookrightarrow A$ is the function such that, for every $x \in S$,

$$i(x) = x$$

The *restriction* of any function $f\colon A \to B$, denoted by $f|_S$ is the function from S into B such that, for any $x \in S$,

$$f|_S(x) = f(x)$$

Functions as Relations

There is another point of view from which functions may be considered. First of all, every function $f\colon A \to B$ gives rise to a relation from A to B called the *graph of* f and defined by

$$\text{Graph of } f = \{(a, b) \mid a \in A, b = f(a)\}$$

Two functions $f\colon A \to B$ and $g\colon A \to B$ are defined to be *equal*, written $f = g$, if $f(a) = g(a)$ for every $a \in A$; that is, if they have the same graph. Accordingly, we do not distinguish between a function and its graph. Now, such a graph relation has the property that each a in A belongs to a unique ordered pair (a, b) in the relation. On the other hand, any relation f from A to B that has this property gives rise to a function $f\colon A \to B$, where $f(a) = b$ for each (a, b) in f. Consequently, one may equivalently define a function as follows:

Definition: A function $f\colon A \to B$ is a relation from A to B (i.e., a subset of $A \times B$) such that each $a \in A$ belongs to a unique ordered pair (a, b) in f.

Although we do not distinguish between a function and its graph, we will still use the terminology "graph of f" when referring to f as a set of ordered pairs. Moreover, since the graph of f is a relation, we can draw its picture as was done for relations in general, and this pictorial representation is itself sometimes called the graph of f. Also, the defining condition of a function, that each $a \in A$ belongs to a unique pair (a, b) in f, is equivalent to the geometrical condition of each vertical line intersecting the graph in exactly one point.

EXAMPLE 3.2

(a) Let $f: A \to B$ be the function defined in Example 3.1 (b). Then the graph of f is as follows:

$$\{(a, s),\ (b, u),\ (c, r),\ (d, s)\}$$

(b) Consider the following three relations on the set $A = \{1, 2, 3\}$:

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

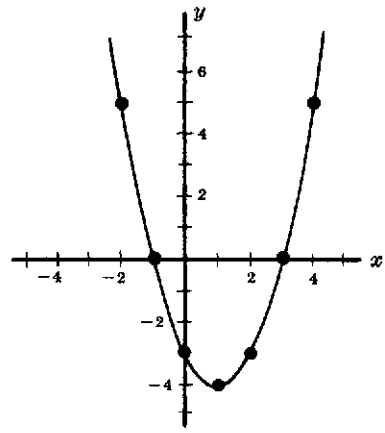
f is a function from A into A since each member of A appears as the first coordinate in exactly one ordered pair in f; here $f(1) = 3$, $f(2) = 3$, and $f(3) = 1$. g is not a function from A into A since $2 \in A$ is not the first coordinate of any pair in g and so g does not assign any image to 2. Also h is not a function from A into A since $1 \in A$ appears as the first coordinate of two distinct ordered pairs in h, (1, 3) and (1, 2). If h is to be a function it cannot assign both 3 and 2 to the element $1 \in A$.

(c) By a *real polynomial function*, we mean a function $f: \mathbf{R} \to \mathbf{R}$ of the form

$$f(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0$$

where the a_i are real numbers. Since $\mathbf{R}$ is an infinite set, it would be impossible to plot each point of the graph. However, the graph of such a function can be approximated by first plotting some of its points and then drawing a smooth curve through these points. The points are usually obtained from a table where various values are assigned to x and the corresponding values of $f(x)$ are computed. Figure 3-2 illustrates this technique using the function $f(x) = x^2 - 2x - 3$.

x	$f(x)$
−2	5
−1	0
0	−3
1	−4
2	−3
3	0
4	5



Graph of $f(x) = x^2 - 2x - 3$

Fig. 3-2

Composition Function

Consider functions $f: A \to B$ and $g: B \to C$; that is, where the codomain of f is the domain of g. Then we may define a new function from A to C, called the *composition* of f and g and written $g \circ f$, as follows:

$$(g \circ f)(a) \equiv g(f(a))$$

That is, we find the image of a under f and then find the image of $f(a)$ under g. This definition is not really new. If we view f and g as relations, then this function is the same as the composition of f and g as relations (see Section 2.6) except that here we use the functional notation $g \circ f$ for the composition of f and g instead of the notation $f \circ g$ which was used for relations.

Consider any function $f: A \to B$. Then

$$f \circ 1_A = f \quad \text{and} \quad 1_B \circ f = f$$

where 1_A and 1_B are the identity functions on A and B, respectively.

3.3 ONE-TO-ONE, ONTO, AND INVERTIBLE FUNCTIONS

A function $f: A \to B$ is said to be *one-to-one* (written 1-1) if different elements in the domain A have distinct images. Another way of saying the same thing is that f is *one-to-one* if $f(a) = f(a')$ implies $a = a'$.

A function $f: A \to B$ is said to be an *onto* function if each element of B is the image of some element of A. In other words, $f: A \to B$ is onto if the image of f is the entire codomain, i.e., if $f(A) = B$. In such a case we say that f is a function from A onto B or that f maps A onto B.

A function $f: A \to B$ is *invertible* if its inverse relation f^{-1} is a function from B to A. In general, the inverse relation f^{-1} may not be a function. The following theorem gives simple criteria which tells us when it is.

Theorem 3.1: A function $f: A \to B$ is invertible if and only if f is both one-to-one and onto.

If $f: A \to B$ is one-to-one and onto, then f is called a *one-to-one correspondence* between A and B. This terminology comes from the fact that each element of A will then correspond to a unique element of B and vice versa.

Some texts use the terms *injective* for a one-to-one function, *surjective* for an onto function, and *bijective* for a one-to-one correspondence.

EXAMPLE 3.3 Consider the functions $f_1: A \to B$, $f_2: B \to C$, $f_3: C \to D$ and $f_4: D \to E$ defined by the diagram of Fig. 3-3. Now f_1 is one-to-one since no element of B is the image of more than one element of A. Similarly, f_2 is one-to-one. However, neither f_3 nor f_4 is one-to-one since $f_3(r) = f_3(u)$ and $f_4(v) = f_4(w)$

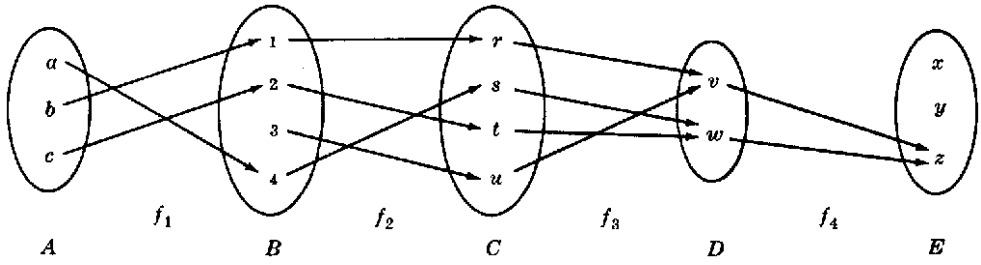


Fig. 3-3

As far as being onto is concerned, f_2 and f_3 are both onto functions since every element of C is the image under f_2 of some element of B and every element of D is the image under f_3 of some element of C, $f_2(B) = C$ and $f_3(C) = D$. On the other hand, f_1 is not onto since $3 \in B$ is not the image under f_4 of any element of A. and f_4 is not onto since $x \in E$ is not the image under f_4 of any element of D.

Thus f_1 is one-to-one but not onto, f_3 is onto but not one-to-one and f_4 is neither one-to-one nor onto. However, f_2 is both one-to-one and onto, i.e., is a one-to-one correspondence between A and B. Hence f_2 is invertible and f_2^{-1} is a function from C to B.

Geometrical Characterization of One-to-One and Onto Functions

Consider now functions of the form $f: \mathbf{R} \to \mathbf{R}$. Since the graphs of such functions may be plotted in the Cartesian plane $\mathbf{R}^2$ and since functions may be identified with their graphs, we might wonder

whether the concepts of being one-to-one and onto have some geometrical meaning. The answer is yes. Specifically:

(1) f:**R** $\rightarrow$ **R** is one-to-one if each horizontal line intersects the graph of f in at most one point.

(2) f:**R** $\rightarrow$ **R** is an onto function if each horizontal line intersects the graph of f at one or more points.

Accordingly, if f is both one-to-one and onto, i.e. invertible, then each horizontal line will intersect the graph of f at exactly one point.

EXAMPLE 3.4 Consider the following four functions from **R** into **R**:

$$f_1(x) = x^2, \quad f_2(x) = 2^x, \quad f_3(x) = x^3 - 2x^2 - 5x + 6, \quad f_4(x) = x^3$$

The graphs of these functions appear in Fig. 3-4. Observe that there are horizontal lines which intersect the graph of f_1 twice and there are horizontal lines which do not intersect the graph of f_1 at all; hence f_1 is neither one-to-one nor onto. Similarly, f_2 is one-to-one but not onto, f_3 is onto but not one-to-one and f_4 is both one-to-one and onto. The inverse of f_4 is the cube root function, i.e., $f_4^{-1}(x) = \sqrt[3]{x}$.

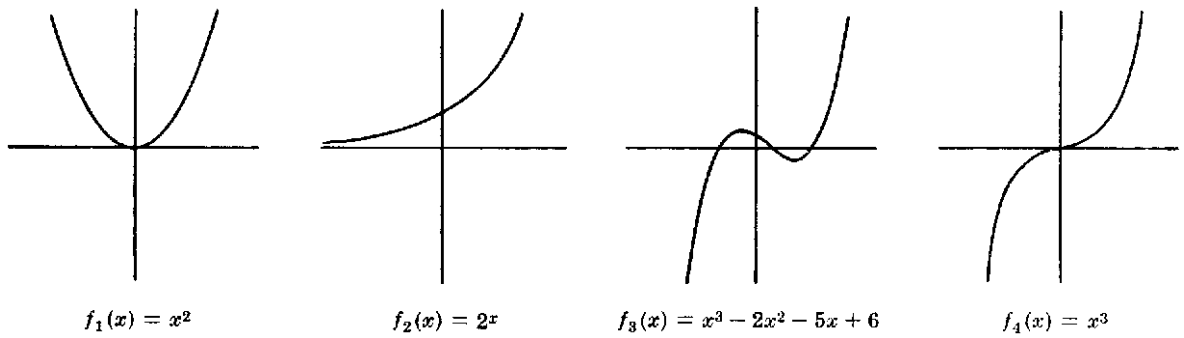


Fig. 3-4

Permutations

An invertible (bijective) function σ: $X \rightarrow X$ is called a *permutation* on X. The composition and inverses of permutations on X and the identity function on X are also permutations on X.

Suppose $X = \{1, 2, \ldots, n\}$. Then a permutation σ on X is frequently denoted by

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & \cdots & n \\ j_1 & j_2 & j_3 & \cdots & j_n \end{pmatrix}$$

where $j_1 = \sigma(i)$. The set of all such permutations is denoted by S_n, and there are $n! = n(n-1)\cdots 3 \cdot 2 \cdot 1$ of them. For example,

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 4 & 6 & 2 & 5 & 1 & 3 \end{pmatrix} \quad \text{and} \quad \tau = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 6 & 4 & 3 & 1 & 2 & 5 \end{pmatrix}$$

are permutations in S_6, and there are $6! = 720$ of them. Sometimes, we only write the second line of the permutation, that is, we denote the above permutations by writing $\sigma = 462513$ and $\tau = 643125$.

3.4 MATHEMATICAL FUNCTIONS, EXPONENTIAL AND LOGARITHMIC FUNCTIONS

This section presents various mathematical functions which appear often in the analysis of algorithms, and in computer science in general, together with their notation. We also discuss the exponential and logarithmic functions, and their relationship.

Floor and Ceiling Functions

Let x be any real number. Then x lies between two integers called the floor and the ceiling of x. Specifically,

$\lfloor x \rfloor$, called the *floor* of x, denotes the greatest integer that does not exceed x.

$\lceil x \rceil$, called the *ceiling* of x, denotes the least integer that is not less than x.

If x is itself an integer, then $\lfloor x \rfloor = \lceil x \rceil$; otherwise $\lfloor x \rfloor + 1 = \lceil x \rceil$. For example,

$$\lfloor 3.14 \rfloor = 3, \quad \left\lfloor \sqrt{5} \right\rfloor = 2, \quad \lfloor -8.5 \rfloor = -9, \quad \lfloor 7 \rfloor = 7, \quad \lfloor -4 \rfloor = -4,$$

$$\lceil 3.14 \rceil = 4, \quad \left\lceil \sqrt{5} \right\rceil = 3, \quad \lceil -8.5 \rceil = -8, \quad \lceil 7 \rceil = 7, \quad \lceil -4 \rceil = -4$$

Integer and Absolute Value Functions

Let x be any real number. The *integer value* of x, written $\text{INT}(x)$, converts x into an integer by deleting (truncating) the fractional part of the number. Thus

$$\text{INT}(3.14) = 3, \quad \text{INT}(\sqrt{5}) = 2, \quad \text{INT}(-8.5) = -8, \quad \text{INT}(7) = 7$$

Observe that $\text{INT}(x) = \lfloor x \rfloor$ or $\text{INT}(x) = \lceil x \rceil$ according to whether x is positive or negative.

The *absolute value* of the real number x, written $\text{ABS}(x)$ or $|x|$, is defined as the greater of x or $-x$. Hence $\text{ABS}(0) = 0$, and, for $x \neq 0$, $\text{ABS}(x) = x$ or $\text{ABS}(x) = -x$, depending on whether x is positive or negative. Thus

$$|-15| = 15, \quad |7| = 7, \quad |-3.33| = 3.33, \quad |4.44| = 4.44, \quad |-0.075| = 0.075$$

We note that $|x| = |-x|$ and, for $x \neq 0$, $|x|$ is positive.

Remainder Function and Modular Arithmetic

Let k be any integer and let M be a positive integer. Then

$$k \pmod{M}$$

(read: k modulo M) will denote the integer remainder when k is divided by M. More exactly, $k \pmod{M}$ is the unique integer r such that

$$k = Mq + r \quad \text{where} \quad 0 \leq r < M$$

When k is positive, simply divide k by M to obtain the remainder r. Thus

$$25 \pmod{7} = 4, \quad 25 \pmod{5} = 0, \quad 35 \pmod{11} = 2, \quad 3 \pmod{8} = 3$$

If k is negative, divide $|k|$ by M to obtain a remainder r'; then $k \pmod{M} = M - r'$ when $r' \neq 0$. Thus

$$-26 \pmod{7} = 7 - 5 = 2, \quad -371 \pmod{8} = 8 - 3 = 5, \quad -39 \pmod{3} = 0$$

The term "mod" is also used for the mathematical congruence relation, which is denoted and defined as follows:

$$a \equiv b \pmod{M} \quad \text{if any only if} \quad M \text{ divides } b - a$$

M is called the *modulus*, and $a \equiv b \pmod{M}$ is read "a is congruent to b modulo M". The following aspects of the congruence relation are frequently useful:

$$0 \equiv M \pmod{M} \quad \text{and} \quad a \pm M \equiv a \pmod{M}$$

Arithmetic modulo M refers to the arithmetic operations of addition, multiplication, and subtraction where the arithmetic value is replaced by its equivalent value in the set

$$\{0, 1, 2, \ldots, M-1\} \quad \text{or in the set} \quad \{1, 2, 3, \ldots, M\}$$

For example, in arithmetic modulo 12, sometimes called "clock" arithmetic,

$$6 + 9 \equiv 3, \quad 7 \times 5 \equiv 11, \quad 1 - 5 \equiv 8, \quad 2 + 10 \equiv 0 \equiv 12$$

(The use of 0 or M depends on the application.)

Exponential Functions

Recall the following definitions for integer exponents (where m is a positive integer):

$$a^m = a \cdot a \cdots a (m \text{ times}), \quad a^0 = 1, \quad a^{-m} = \frac{1}{a^m}$$

Exponents are extended to include all rational numbers by defining, for any rational number m/n,

$$a^{m/n} = \sqrt[n]{a^m} = (\sqrt[n]{a})^m$$

For example,

$$2^4 = 16, \quad 2^{-4} = \frac{1}{2^4} = \frac{1}{16}, \quad 125^{2/3} = 5^2 = 25$$

In fact, exponents are extended to include all real numbers by defining, for any real number x,

$$a^x = \lim_{r \to x} a^r, \quad \text{where } r \text{ is a rational number}$$

Accordingly, the exponential function $f(x) = a^x$ is defined for all real numbers.

Logarithmic Functions

Logarithms are related to exponents as follows. Let b be a positive number. The logarithm of any positive number x to be the base b, written

$$\log_b x$$

represents the exponent to which b must be raised to obtain x. That is,

$$y = \log_b x \quad \text{and} \quad b^y = x$$

are equivalent statements. Accordingly,

$$\log_2 8 = 3 \quad \text{since} \quad 2^3 = 8; \qquad \log_{10} 100 = 2 \quad \text{since} \quad 10^2 = 100$$
$$\log_2 64 = 6 \quad \text{since} \quad 2^6 = 64; \qquad \log_{10} 0.001 = -3 \quad \text{since} \quad 10^{-3} = 0.001$$

Furthermore, for any base b, we have $b^0 = 1$ and $b^1 = b$; hence

$$\log_b 1 = 0 \quad \text{and} \quad \log_b b = 1$$

The logarithm of a negative number and the logarithm of 0 are not defined.

Frequently, logarithms are expressed using approximate values. For example, using tables or calculators, one obtains

$$\log_{10} 300 = 2.4771 \quad \text{and} \quad \log_e 40 = 3.6889$$

as approximate answers. (Here $e = 2.718281....$)

Three classes of logarithms are of special importance: logarithms to base 10, called *common logarithms*; logarithms to base e, called *natural logarithms;* and logarithms to base 2, called *binary logarithms*. Some texts write

$$\ln x \text{ for } \log_e x \qquad \text{and} \qquad \lg x \text{ or } \log x \text{ for } \log_2 x$$

The term $\log x$, by itself, usually means $\log_{10} x$; but it is also used for $\log_e x$ in advanced mathematical texts and for $\log_2 x$ in computer science texts.

Frequently, we will require only the floor or the ceiling of a binary logarithm. This can be obtained by looking at the powers of 2. For example,

$$\lfloor \log_2 100 \rfloor = 6 \quad \text{since} \quad 2^6 = 64 \quad \text{and} \quad 2^7 = 128$$
$$\lceil \log_2 1000 \rceil = 9 \quad \text{since} \quad 2^8 = 512 \quad \text{and} \quad 2^9 = 1024$$

and so on.

Relationship between the Exponential and Logarithmic Functions

The basic relationship between the exponential and the logarithmic functions

$$f(x) = b^x \quad \text{and} \quad g(x) = \log_b x$$

is that they are inverses of each other; hence the graphs of these functions are related geometrically. This relationship is illustrated in Fig. 3-5 where the graphs of the exponential function $f(x) = 2^x$, the logarithmic function $g(x) = \log_2 x$, and the linear function $h(x) = x$ appear on the same coordinate axis. Since $f(x) = 2^x$ and $g(x) = \log_2 x$ are inverse functions, they are symmetric with respect to the linear function $h(x) = x$ or, in other words, the line $y = x$.

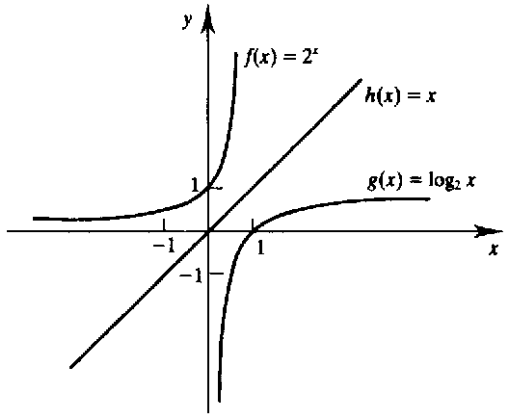


Fig. 3-5

Figure 3-5 also indicates another important property of the exponential and logarithmic functions. Specifically, for any positive c, we have

$$g(c) < h(c) < f(c), \quad \text{that is,} \quad g(c) < c < f(c)$$

In fact, as c increases in value, the vertical distances $h(c) - g(c)$ and $f(c) - g(c)$ increase in value. Moreover, the logarithmic function $g(x)$ grows very slowly compared with the linear function $h(x)$, and the exponential function $f(x)$ grows very quickly compared with $h(x)$.

3.5 SEQUENCES, INDEXED CLASSES OF SETS

Sequences and indexed classes of sets are special types of functions with their own notation. We discuss these objects in this section. We also discuss the summation notation here.

Sequences

A *sequence* is a function from the set $\mathbf{N} = \{1, 2, 3, \ldots\}$ of positive integers into a set A. The notation a_n is used to denote the image of the integer n. Thus a sequence is usually denoted by

$$a_1, a_2, a_3, \ldots \quad \text{or} \quad \{a_n : n \in \mathbf{N}\} \quad \text{or} \quad \text{simply} \quad \{a_n\}$$

Sometimes the domain of a sequence is the set $\{0, 1, 2, \ldots\}$ of nonnegative integers rather than $\mathbf{N}$. In such a ease we say n begins with 0 rather than 1.

A *finite sequence* over a set A is a function from $\{1, 2, \ldots, m\}$ into A, and it is usually denoted by

$$a_1, a_2, \ldots, a_m$$

Such a finite sequence is sometimes called a *list* or an *m-tuple*.

EXAMPLE 3.5

(a) The following are two familiar sequences:

(i) $1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \ldots$ which may be defined by $a_n = \frac{1}{n}$;

(ii) $1, \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \ldots$ which may be defined by $b_n = 2^{-n}$

Note that the first sequence begins with $n = 1$ and the second sequence begins with $n = 0$.

(b) The important sequence $1, -1, 1, -1, \ldots$ may be formally defined by

$$a_n = (-1)^{n+1} \quad \text{or, equivalently, by} \quad b_n = (-1)^n$$

where the first sequence begins with $n = 1$ and the second sequence begins with $n = 0$.

(c) ***Strings*** Suppose a set A is finite and A is viewed as a character set or an alphabet. Then a finite sequence over A is called a *string* or *word*, and it is usually written in the form $a_1a_2 \ldots a_m$, that is, without parentheses. The number m of characters in the string is called its *length*. One also views the set with zero characters as a string; it is called the *empty string* or *null string*. Strings over an alphabet A and certain operations on these strings will be discussed in detail in Chapter 13.

Summation Symbol, Sums

Here we introduce the summation symbol $\sum$ (the Greek letter sigma). Consider a sequence $a_1, a_2, a_3, \ldots$. Then we define the following:

$$\sum_{J=1}^{n} a_j = a_1 + a_2 + \cdots + a_n \quad \text{and} \quad \sum_{j=m}^{n} a_j = a_m + a_{m+1} + \cdots + a_n$$

The letter j in the above expressions is called a *dummy index* or *dummy variable*. Other letters frequently used as dummy variables are i, k, s, and t.

EXAMPLE 3.6

$$\sum_{i=1}^{n} a_i b_i = a_1 b_1 + a_2 b_2 + \cdots + a_n b_n$$

$$\sum_{j=2}^{5} j^2 = 2^2 + 3^2 + 4^2 + 5^2 = 4 + 9 + 16 + 25 = 54$$

$$\sum_{j=1}^{n} j = 1 + 2 + \cdots + n$$

The last sum appears very often. It has the value $n(n+1)/2$. That is:

$$1+2+3+\cdots+n=\frac{n(n+1)}{2}, \quad \text{for example,} \quad 1+2+\cdots+50=\frac{50(51)}{2}=1275$$

Indexed Classes of Sets

Let I be any nonempty set, and let S be a collection of sets. An *indexing function* from I to S is a function $f\colon I \to S$. For any $i \in I$, we denote the image $f(i)$ by A_i. Thus the indexing function f is usually denoted by

$$\{A_i \mid i \in I\} \quad \text{or} \quad \{A_i\}_{i\in I} \quad \text{or simply} \quad \{A_i\}$$

The set I is called the *indexing set*, and the elements of I are called *indices*. If f is one-to-one and onto, we say that S is indexed by I.

The concepts of union and intersection are defined for indexed classes of sets as follows:

$$\cup_{i\in I} A_i = \{x \mid x \in A_i \text{ for some } i \in I\} \quad \text{and} \quad \cap_{i\in I}\ A_i = \{x \mid x \in A_i \text{ for all } i \in I\}$$

In the case that I is a finite set, this is just the same as our previous definition of union and intersection. If I is $\mathbf{N}$, we may denote the union and intersection, respectively, as follows:

$$A_1 \cup A_2 \cup A_3 \cup \ldots \quad \text{and} \quad A_1 \cap A_2 \cap A_3 \cap \ldots$$

EXAMPLE 3.7 Let I be the set $\mathbf{Z}$ of integers. To each $n \in \mathbf{Z}$, we assign the following infinite interval in $\mathbf{R}$:

$$A_n = \{x \mid x \le n\} = (-\infty, n]$$

For any real number a, there exists integers n_1 and n_2 such that $n_1 < a < n_2$; so $a \in A_{n_2}$ but $a \notin A_{n_1}$. Hence

$$a \in \cup_n A_n \quad \text{but} \quad a \notin \cap_n A_n$$

Accordingly,

$$\cup_n A_n = \mathbf{R} \quad \text{but} \quad \cap_n\ A_n = \emptyset$$

3.6 RECURSIVELY DEFINED FUNCTIONS

A function is said to be *recursively defined* if the function definition refers to itself. In order for the definition not to be circular, the function definition must have the following two properties:

(1) There must be certain arguments, called *base values*, for which the function does not refer to itself.
(2) Each time the function does refer to itself, the argument of the function must be closer to a base value.

A recursive function with these two properties is said to be *well-defined*.

The following examples should help clarify these ideas.

Factorial Function

The product of the positive integers from 1 to n, inclusive, is called "n factorial" and is usually denoted by $n!$. That is,

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

It is also convenient to define $0! = 1$, so that the function is defined for all nonnegative integers. Thus:

$$0! = 1, \quad 1! = 1, \quad 2! = 2\cdot 1 = 2, \quad 3! = 3\cdot 2\cdot 1 = 6, \quad 4! = 4\cdot 3\cdot 2\cdot 1 = 24$$

$$5! = 5\cdot 4\cdot 3\cdot 2\cdot 1 = 120, \quad 6! = 6\cdot 5\cdot 4\cdot 3\cdot 2\cdot 1 = 720$$

And so on. Observe that

$$5! = 5 \cdot 4! = 5 \cdot 24 = 120 \quad \text{and} \quad 6! = 6 \cdot 5! = 6 \cdot 120 = 720$$

This is true for every positive integer n; that is,

$$n! = n \cdot (n - 1)!$$

Accordingly, the factorial function may also be defined as follows:

Definition 3.1 (Factorial Function):

(a) If $n = 0$, then $n! = 1$.
(b) If $n > 0$, then $n! = n \cdot (n - 1)!$

Observe that the above definition of $n!$ is recursive, since it refers to itself when it uses $(n - 1)!$. However:

(1) The value of $n!$ is explicitly given when $n = 0$ (thus 0 is a base value).
(2) The value of $n!$ for arbitrary n is defined in terms of a smaller value of n which is closer to the base value 0.

Accordingly, the definition is not circular, or, in other words, the function is well-defined.

EXAMPLE 3.8 Figure 3-6 shows the nine steps to calculate 4! using the recursive definition. Specifically:

Step 1. This defines 4! in terms of 3!, so we must postpone evaluating 4! until we evaluate 3. This postponement is indicated by indenting the next step.

Step 2. Here 3! is defined in terms of 2!, so we must postpone evaluating 3! until we evaluate 2!.

Step 3. This defines 2! in terms of 1!.

Step 4. This defines 1! in terms of 0!.

Step 5. This step can explicitly evaluate 0!, since 0 is the base value of the recursive definition.

Steps 6 to 9. We backtrack, using 0! to find 1!, using 1! to find 2!, using 2! to find 3!, and finally using 3! to find 4!. This backtracking is indicated by the "reverse" indention.

Observe that we backtrack in the reverse order of the original postponed evaluations.

(1) $4! = 4 \cdot 3!$
(2) $3! = 3 \cdot 2!$
(3) $2! = 2 \cdot 1!$
(4) $1! = 1 \cdot 0!$
(5) $0! = 1$
(6) $1! = 1 \cdot 1 = 1$
(7) $2! = 2 \cdot 1 = 2$
(8) $3! = 3 \cdot 2 = 6$
(9) $4! = 4 \cdot 6 = 24$

Fig. 3-6

Level Numbers

Let P be a procedure or recursive formula which is used to evaluate $f(X)$ where f is a recursive function and X is the input. We associate a *level number* with each execution of P as follows. The original execution of P is assigned level 1; and each time P is executed because of a recursive call, its level is one more than the level of the execution that made the recursive call. The *depth* of recursion in evaluating $f(X)$ refers to the maximum level number of P during its execution.

Consider, for example, the evaluation of 4! Example 3.8, which uses the recursive formula $n! = n(n-1)!$. Step 1 belongs to level 1 since it is the first execution of the formula. Thus:

Step 2 belongs to level 2; Step 3 to level 3, . . . ; Step 5 to level 5.

On the other hand, Step 6 belongs to level 4 since it is the result of a return from level 5. In other words, Step 6 and Step 4 belong to the same level of execution. Similarly,

Step 7 belongs to level 3; Step 8 to level 2; and Step 9 to level 1.

Accordingly, in evaluating 4!, the depth of the recursion is 5.

Fibonacci Sequence

The celebrated Fibonacci sequence (usually denoted by $F_0, F_1, F_2, \ldots$) is as follows:

$$0, \quad 1, \quad 1, \quad 2, \quad 3, \quad 5, \quad 8, \quad 13, \quad 21, \quad 34, \quad 55, \quad \ldots$$

That is, $F_0 = 0$ and $F_1 = 1$ and each succeeding term is the sum of the two preceding terms. For example, the next two terms of the sequence are

$$34 + 55 = 89 \quad \text{and} \quad 55 + 89 = 144$$

A formal definition of this function follows:

Definition 3.2 (Fibonacci Sequence):

(a) If $n = 0$, or $n = 1$, then $F_n = n$.
(b) If $n > 1$, then $F_n = F_{n-2} + F_{n-1}$.

This is another example of a recursive definition, since the definition refers to itself when it uses F_{n-2} and F_{n-1} However:

(1) The base values are 0 and 1.
(2) The value of F_n is defined in terms of smaller values of n which are closer to the base values.

Accordingly, this function is well-defined.

Ackermann Function

The Ackermann function is a function with two arguments, each of which can be assigned any nonnegative interger, that is, 0, 1, 2, ... This function is defined as:

Definition 3.3 (Ackermann function):

(a) If $m = 0$, then $A(m, n) = n + 1$.
(b) If $m \neq 0$ but $n = 0$, then $A(m, n) = A(m - 1, 1)$.
(c) If $m \neq 0$ and $n \neq 0$, then $A(m, n) = A(m - 1, A(m, n - 1))$.

Once more, we have a recursive definition, since the definition refers to itself in parts (b) and (c). Observe that $A(m, n)$ is explicitly given only when $m = 0$. The base criteria are the pairs

$$(0, 0), \quad (0, 1), \quad (0, 2), \quad (0, 3), \quad \ldots, (0, n), \quad \ldots$$

Although it is not obvious from the definition, the value of any $A(m, n)$ may eventually be expressed in terms of the value of the function on one or more of the base pairs.

The value of $A(1, 3)$ is calculated in Problem 3.21. Even this simple case requires 15 steps. Generally speaking, the Ackermann function is too complex to evaluate on any but a trivial example. Its importance comes from its use in mathematical logic. The function is stated here mainly to give another example of a classical recursive function and to show that the recursion part of a definition may be complicated.

3.7 CARDINALITY

Two sets A and B are said to be *equipotent*, or to have the *same number of elements* or the *same cardinality*, written $A \simeq B$, if there exists a one-to-one correspondence $f: A \to B$. A set A is *finite* if A is empty or if A has the same cardinality as the set $\{1, 2, \ldots, n\}$ for some positive integer n. A set is *infinite* if it is not finite. Familiar examples of infinite sets are the natural numbers **N**, the integers **Z**, the rational numbers **Q**, and the real numbers **R**.

We now introduce the idea of "cardinal numbers". We will consider cardinal numbers simply as symbols assigned to sets in such a way that two sets are assigned the same symbol if and only if they have the same cardinality. The cardinal number of a set A is commonly denoted by $|A|$, $n(A)$, or card (A). We will use $|A|$.

The obvious symbols are used for the cardinality of finite sets. That is, 0 is assigned to the empty set $\emptyset$, and n is assigned to the set $\{1, 2, \ldots, n\}$. Thus $|A| = n$ if and only if A has n elements. For example,

$$|\{x, y, z\}| = 3 \quad \text{and} \quad |\{1, 3, 5, 7, 9\}| = 5$$

The cardinal number of the infinite set **N** of positive integers is $\aleph_0$ ("aleph-naught"). This symbol was introduced by Cantor. Thus $|A| = \aleph_0$ if and only if A has the same cardinality as **N**.

EXAMPLE 3.9 Let $E = \{2, 4, 6, \ldots\}$, the set of even positive integers. The function $f: \mathbf{N} \to E$ defined by $f(n) = 2n$ is a one-to-one correspondence between the positive integers **N** and E. Thus E has the same cardinality as **N** and so we may write

$$|E| = \aleph_0$$

A set with cardinality $\aleph_0$ is said to be *denumerable* or *countably infinite*. A set which is finite or denumerable is said to be *countable*. One can show that the set **Q** of rational numbers is countable. In fact, we have the following theorem (proved in Problem 3.13) which we will use subsequently.

Theorem 3.2: A countable union of countable sets is countable.

That is, if $A_1, A_2, \ldots$ are each countable sets, then the following union is countable:

$$A_1 \cup A_2 \cup A_3 \cup \ldots$$

An important example of an infinite set which is uncountable, i.e., not countable, is given by the following theorem which is proved in Problem 3.14.

Theorem 3.3: The set **I** of all real numbers between 0 and 1 is uncountable.

Inequalities and Cardinal Numbers

One also wants to compare the size of two sets. This is done by means of an inequality relation which is defined for cardinal numbers as follows. For any sets A and B, we define $|A| \leq |B|$ if there exists a function $f: A \to B$ which is one-to-one. We also write

$$|A| < |B| \quad \text{if} \quad |A| \leq |B| \quad \text{but} \quad |A| \neq |B|$$

For example, $|\mathbf{N}| < |\mathbf{I}|$, where $\mathbf{I} = \{x: 0 \leq x \leq 1\}$, since the function $f: \mathbf{N} \to \mathbf{I}$ defined by $f(n) = 1/n$ is one-to-one, but $|\mathbf{N}| \neq |\mathbf{I}|$ by Theorem 3.3.

Cantor's Theorem, which follows and which we prove in Problem 3.25, tells us that the cardinal numbers are unbounded.

Theorem 3.4 (Cantor): For any set A, we have $|A| < |\text{Power}(A)|$ (where Power(A) is the power set of A, i.e., the collection of all subsets of A).

The next theorem tells us that the inequality relation for cardinal numbers is antisymmetric.

Theorem 3.5: (Schroeder-Bernstein): Suppose A and B are sets such that

$$|A| \leq |B| \quad \text{and} \quad |B| \leq |A|$$

Then $|A| = |B|$.

We prove an equivalent formulation of this theorem in Problem 3.26.

3.8 ALGORITHMS AND FUNCTIONS

An algorithm M is a finite step-by-step list of well-defined instructions for solving a particular problem, say, to find the output $f(X)$ for a given function f with input X. (Here X may be a list or set of values.) Frequently, there may be more than one way to obtain $f(X)$, as illustrated by the following examples. The particular choice of the algorithm M to obtain $f(X)$ may depend on the "efficiency" or "complexity" of the algorithm; this question of the complexity of an algorithm M is formally discussed in the next section.

EXAMPLE 3.10 (Polynomial Evaluation) Suppose, for a given polynomial $f(x)$ and value $x = a$, we want to find $f(a)$, say,

$$f(x) = 2x^3 - 7x^2 + 4x - 15 \quad \text{and} \quad a = 5$$

This can be done in the following two ways.

(a) (***Direct Method***)**:** Here we substitute $a = 5$ directly in the polynomial to obtain

$$f(5) = 2(125) - 7(25) + 4(5) - 7 = 250 - 175 + 20 - 15 = 80$$

Observe that there are $3 + 2 + 1 = 6$ multiplications and 3 additions. In general, evaluating a polynomial of degree n directly would require approximately

$$n + (n-1) + \cdots + 1 = \frac{n(n+1)}{2} \text{ multiplications and } n \text{ additions.}$$

(b) (***Horner's Method or Synthetic Division***)**:** Here we rewrite the polynomial by successively factoring out x (on the right) as follows:

$$f(x) = (2x^2 - 7x + 4)x - 15 = ((2x - 7)x + 4)x - 15$$

Then

$$f(5) = ((3)5 + 4)5 - 15 = (19)5 - 15 = 95 - 15 = 80$$

For those familiar with synthetic division, the above arithmetic is equivalent to the following synthetic division:

$$\begin{array}{r|cccccccc} 5 & 2 & - & 7 & + & 4 & - & 15 \\ & & & 10 & + & 15 & + & 95 \\ \hline & 2 & + & 3 & + & 19 & + & 80 \end{array}$$

Observe that here there are 3 multiplications and 3 additions. In general, evaluating a polynomial of degree n by Horner's method would require approximately

$$n \text{ multiplications and } n \text{ additions}$$

Clearly Horner's method (b) is more efficient than the direct method (a).

EXAMPLE 3.11 (Greatest Common Divisor) Let a and b be positive integers with, say, $b < a$; and suppose we want to find $d = \text{GCD}(a, b)$, the greatest common divisor of a and b. This can be done in the following two ways.

(a) (***Direct Method***)**:** Here we find all the divisors of a, say by testing all the numbers from 2 to $a/2$, and all the divisors of b. Then we pick the largest common divisor. For example, suppose $a = 258$ and $b = 60$. The divisors of a and b follow:

$$\begin{array}{lll} a = 258; & \text{divisors:} & 1,\ 2,\ 3,\ 6,\ 86,\ 129,\ 258 \\ b = 60; & \text{divisors:} & 1,\ 2,\ 3,\ 4,\ 5,\ 6,\ 10,\ 12,\ 15,\ 20,\ 30,\ 60 \end{array}$$

Accordingly, $d = \text{GCD}(258, 60) = 6$.

(b) (***Euclidean Algorithm***)**:** Here we divide a by b to obtain a remainder r_1. (Note $r_1 < b$.) Then we divide b by the remainder r_1 to obtain a second remainder r_2. (Note $r_2 < r_1$.) Next we divide r_1 by r_2 to obtain a third remainder r_3. (Note $r_3 < r_2$.) We continue dividing r_k by r_{k+1} to obtain a remainder r_{k+2}. Since

$$a > b > r_1 > r_2 > r_3 \ldots \qquad (*)$$

eventually we obtain a remainder $r_m = 0$. Then $r_{m-1} = \text{GCD}\ (a, b)$. For example, suppose $a = 258$ and $b = 60$. Then:

(1) Dividing $a = 258$ by $b = 60$ yields the remainder $r_1 = 18$.
(2) Dividing $b = 60$ by $r_1 = 18$ yields the remainder $r_2 = 6$.
(3) Dividing $r_1 = 18$ by $r_2 = 6$ yields the remainder $r_3 = 0$.

Thus $r_2 = 6 = \text{GCD}(258, 60)$.

The Euclidean algorithm is a very efficient way to find the greatest common divisor of two positive integers a and b. The fact that the algorithm ends follows from $(*)$. The fact that the algorithm yields $d = \text{GCD}(a, b)$ is not obvious; it is discussed in Section 11.6.

3.9 COMPLEXITY OF ALGORITHMS

The analysis of algorithms is a major task in computer science. In order to compare algorithms, we must have some criteria to measure the efficiency of our algorithms. This section discusses this important topic.

Suppose M is an algorithm, and suppose n is the size of the input data. The time and space used by the algorithm are the two main measures for the efficiency of M. The time is measured by counting the number of "key operations;" for example:

(a) In sorting and searching, one counts the number of comparisons.
(b) In arithmetic, one counts multiplications and neglects additions.

Key operations are so defined when the time for the other operations is much less than or at most proportional to the time for the key operations. The space is measured by counting the maximum of memory needed by the algorithm.

The *complexity* of an algorithm M is the function $f(n)$ which gives the running time and/or storage space requirement of the algorithm in terms of the size n of the input data. Frequently, the storage space required by an algorithm is simply a multiple of the data size. Accordingly, unless otherwise stated or implied, the term "complexity" shall refer to the running time of the algorithm.

The complexity function $f(n)$, which we assume gives the running time of an algorithm, usually depends not only on the size n of the input data but also on the particular data. For example, suppose we want to search through an English short story TEXT for the first occurrence of a given 3-letter word W. Clearly, if W is the 3-letter word "the," then W likely occurs near the beginning of TEXT, so $f(n)$ will be small. On the other hand, if W is the 3-letter word "zoo," then W may not appear in TEXT at all, so $f(n)$ will be large.

The above discussion leads us to the question of finding the complexity function $f(n)$ for certain cases. The two cases one usually investigates in complexity theory are as follows:

(1) *Worst case*: The maximum value of $f(n)$ for any possible input.
(2) *Average case*: The expected value of $f(n)$.

The analysis of the average case assumes a certain probabilistic distribution for the input data; one possible assumption might be that the possible permutations of a data set are equally likely. The average case also uses the following concept in probability theory. Suppose the numbers $n_1, n_2, \ldots, n_k$ occur with respective probabilities $p_1, p_2, \ldots, p_k$. Then the *expectation* or *average value* E is given by

$$E = n_1 p_1 + n_2 p_2 + \cdots + n_k p_k$$

These ideas are illustrated below.

Linear Search

Suppose a linear array DATA contains n elements, and suppose a specific ITEM of information is given. We want either to find the location LOC of ITEM in the array DATA, or to send some message, such as LOC $= 0$, to indicate that ITEM does not appear in DATA. The linear search algorithm solves this problem by comparing ITEM, one by one, with each element in DATA. That is, we compare ITEM with DATA[1], then DATA[2], and so on, until we find LOC such that ITEM $=$ DATA[LOC].

The complexity of the search algorithm is given by the number C of comparisons between ITEM and DATA[K]. We seek $C(n)$ for the worst case and the average case.

(1) ***Worst Case***: Clearly the worst case occurs when ITEM is the last element in the array DATA or is not there at all. In either situation, we have

$$C(n) = n$$

Accordingly, $C(n) = n$ is the worst-case complexity of the linear search algorithm.

(2) ***Average Case***: Here we assume that ITEM does appear in DATA, and that it is equally likely to occur at any position in the array. Accordingly, the number of comparisons can be any of the numbers $1, 2, 3, \ldots, n$, and each number occurs with probability $p = 1/n$. Then

$$\begin{aligned} C(n) &= 1 \cdot \frac{1}{n} + 2 \cdot \frac{1}{n} + \cdots + n \cdot \frac{1}{n} \\ &= (1 + 2 + \cdots + n) \cdot \frac{1}{n} \\ &= \frac{n(n+1)}{2} \cdot \frac{1}{n} = \frac{n+1}{2} \end{aligned}$$

This agrees with our intuitive feeling that the average number of comparisons needed to find the location of ITEM is approximately equal to half the number of elements in the DATA list.

Remark: The complexity of the average case of an algorithm is usually much more complicated to analyze than that of the worst case. Moreover, the probabilistic distribution that one assumes for the average case may not actually apply to real situations. Accordingly, unless otherwise stated or implied, the complexity of an algorithm shall mean the function which gives the running time of the worst case in terms of the input size. This is not too strong an assumption, since the complexity of the average case for many algorithms is proportional to the worst case.

Rate of Growth; Big O Notation

Suppose M is an algorithm, and suppose n is the size of the input data. Clearly the complexity $f(n)$ of M increases as n increases. It is usually the rate of increase of $f(n)$ that we want to examine. This is usually done by comparing $f(n)$ with some standard function, such as

$$\log n, \quad n, \quad n \log n, \quad n^2, \quad n^3, \quad 2^n$$

The rates of growth for these standard functions are indicated in Fig. 3-7, which gives their approximate values for certain values of n. Observe that the functions are listed in the order of their rates of growth: the logarithmic function $\log_2 n$ grows most slowly, the exponential function 2^n grows most rapidly, and the polynomial functions n^c grow according to the exponent c.

n \ $g(n)$	$\log n$	n	$n \log n$	n^2	n^3	2^n
5	3	5	15	25	125	32
10	4	10	40	100	10^3	10^3
100	7	100	700	10^4	10^6	10^{30}
1000	10	10^3	10^4	10^6	10^9	10^{300}

Fig. 3-7 Rate of growth of standard functions

The way we compare our complexity function $f(n)$ with one of the standard functions is to use the functional "big O" notation which we formally define below.

Definition 3.4: Let $f(x)$ and $g(x)$ be arbitrary functions defined on **R** or a subset of **R**. We say "$f(x)$ is of order $g(x)$," written

$$f(x) = O(g(x))$$

if there exists a real number k and a positive constant C such that, for all $x > k$, we have

$$|f(x)| \leq C|g(x)|$$

In other words, $f(x) = 0(g(x))$ if a constant multiple of $|g(x)|$ exceeds $|f(x)|$ for all x greater than some real number k.

We also write:

$$f(x) = h(x) + O(g(x)) \quad \text{when} \quad f(x) - h(x) = O(g(x))$$

(The above is called the "big O" notation since $f(x) = o(g(x))$ has an entirely different meaning.)

Consider now a polynomial $P(x)$ of degree m. We show in Problem 3.24 that $P(x) = O(x^m)$. Thus, for example,

$$7x^2 - 9x + 4 = O(x^2) \quad \text{and} \quad 8x^3 - 576x^2 + 832x - 248 = O(x^3)$$

Complexity of Well-known Algorithms

Assuming $f(n)$ and $g(n)$ are functions defined on the positive integers, then

$$f(n) = O(g(n))$$

means that $f(n)$ is bounded by a constant multiple of $g(n)$ for almost all n.

To indicate the convenience of this notation, we give the complexity of certain well-known searching and sorting algorithms in computer science:

(a) Linear search: $O(n)$
(b) Binary search: $O(\log n)$
(c) Bubble sort: $O(n^2)$
(d) Merge-sort: $O(n \log n)$

Solved Problems

FUNCTIONS

3.1. Let $X = \{1, 2, 3, 4\}$. Determine whether each relation on X is a function from X into X.

(*a*) $f = \{(2, 3), (1, 4), (2, 1), (3.2), (4, 4)\}$

(*b*) $g = \{(3, 1), (4, 2), (1, 1)\}$

(*c*) $h = \{(2, 1), (3, 4), (1, 4), (2, 1), (4, 4)\}$

Recall that a subset f of $X \times X$ is a function $f: X \to X$ if and only if each $a \in X$ appears as the first coordinate in exactly one ordered pair in f.

(*a*) No. Two different ordered pairs (2, 3) and (2, 1) in f have the same number 2 as their first coordinate.

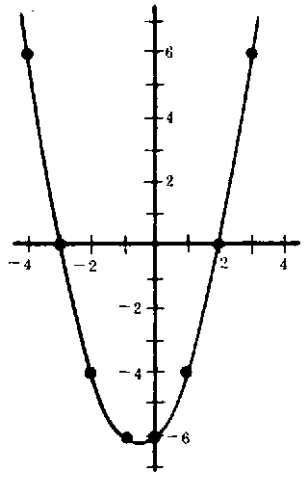
(*b*) No. The element $2 \in X$ does not appear as the first coordinate in any ordered pair in g.

(*c*) Yes. Although $2 \in X$ appears as the first coordinate in two ordered pairs in h, these two ordered pairs are equal.

3.2. Sketch the graph of: (*a*) $f(x) = x^2 + x - 6$; (*b*) $g(x) = x^3 - 3x^2 - x + 3$.

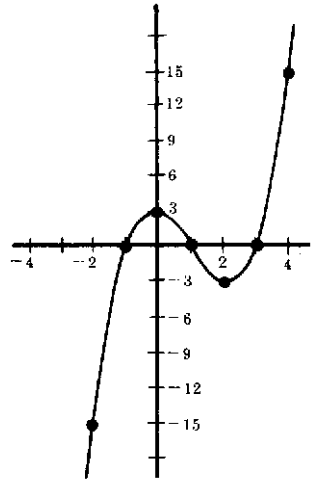
Set up a table of values for x and then find the corresponding values of the function. Since the functions are polynomials, plot the points in a coordinate diagram and then draw a smooth continuous curve through the points. See Fig. 3-8.

x	$f(x)$
−4	6
−3	0
−2	−4
−1	−6
0	−6
1	−4
2	0
3	6



Graph of $f = x^2 + x - 6$

x	$g(x)$
−2	−15
−1	0
0	3
1	0
2	−3
3	0
4	15



Graph of $g = x^3 - 3x^2 - x - 3$

Fig. 3-8

3.3. Let $A = \{a, b, c\}$, $B = \{x, y, z\}$, $C = \{r, s, t\}$. Let $f: A \to B$ and $g: B \to C$ be defined by:

$$f = \{(a, y)(b, x), (c, y)\} \quad \text{and} \quad g = \{(x, s), (y, t), (z, r)\}.$$

Find: (*a*) composition function $g \circ f: A \to C$; (*b*) $\mathrm{Im}(f), \mathrm{Im}(g), \mathrm{Im}(g \circ f)$.

(*a*) Use the definition of the composition function to compute:

$$(g \circ f)(a) = g(f(a)) = g(y) = t$$
$$(g \circ f)(b) = g(f(b)) = g(x) = s$$
$$(g \circ f)(c) = g(f(c)) = g(y) = t$$

That is $g \circ f = \{(a, t), (b, s), (c, t)\}$.

(*b*) Find the image points (or second coordinates):

$$\mathrm{Im}(f) = \{x, y\}, \quad \mathrm{Im}(g) = \{r, s, t\}, \quad \mathrm{Im}(g \circ f) = \{s, t\}$$

3.4. Let f: $\mathbf{R} \to \mathbf{R}$ and g: $\mathbf{R} \to \mathbf{R}$ be defined by $f(x) = 2x + 1$ and $g(x) = x^2 - 2$. Find the formula for the composition function $g \circ f$.

Compute $g \circ f$ as follows: $(g \circ f)(x) = g(f(x)) = g(2x + 1) = (2x + 1)^2 - 2 = 4x^2 + 4x - 1$.

Observe that the same answer can be found by writing

$$y = f(x) = 2x + 1 \quad \text{and} \quad z = g(y) = y^2 - 2$$

and then eliminating y from both equations:

$$z = y^2 - 2 = (2x + 1)^2 - 2 = 4x^2 + 4x - 1$$

ONE-TO-ONE, ONTO, AND INVERTIBLE FUNCTIONS

3.5. Let the functions f: $A \to B$, g: $B \to C$, h: $C \to D$ be defined by Fig. 3-9. Determine if each function is: (a) onto, (b) one-to-one, (c) invertible.

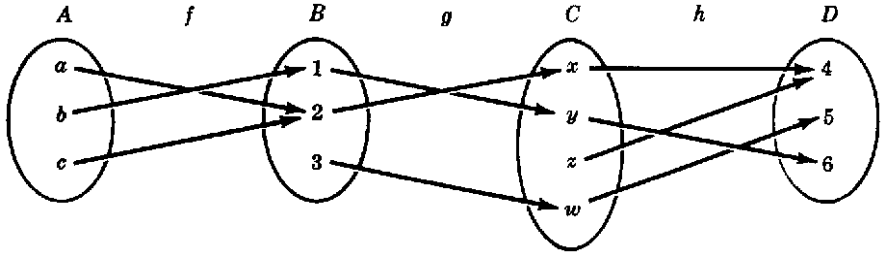


Fig. 3-9

(a) The function f: $A \to B$ is not onto since $3 \in B$ is not the image of any element in A.

The function g: $B \to C$ is not onto since $z \in C$ is not the image of any element in B.

The function h: $C \to D$ is onto since each element in D is the image of some element of C.

(b) The function f: $A \to B$ is not one-to-one since a and c have the same image 2.

The function g: $B \to C$ is one-to-one since 1, 2 and 3 have distinct images.

The function h: $C \to D$ is not one-to-one since x and z have the same image 4.

(c) No function is one-to-one and onto; hence no function is invertible.

3.6. Consider permutations $\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 3 & 6 & 4 & 5 & 1 & 2 \end{pmatrix}$ and $\tau = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 2 & 4 & 6 & 5 & 3 & 1 \end{pmatrix}$ in S_6.
Find: (a) composition $\tau \circ \sigma$; (b) σ^{-1}.

(a) Note that σ sends 1 into 3 and τ sends 3 into 6. So the composition $\tau \circ \sigma$ sends 1 into 6. I.e. $(\tau \circ \sigma)(1) = 6$. Moreover, $\tau \circ \sigma$ sends 2 into 6 into 1 that is, $(\tau \circ \sigma)(2) = 1$, Similarly,

$$(\tau \circ \sigma)(3) = 5, \quad (\tau \circ \sigma)(4) = 3, \quad (\tau \circ \sigma) = 2, \quad (\tau \circ \sigma)(6) = 4$$

Thus

$$\tau \circ \sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 6 & 1 & 5 & 3 & 2 & 4 \end{pmatrix}$$

(b) Look for 1 in the second row of σ. Note σ sends 5 into 1. Hence $\sigma^{-1}(1) = 5$. Look for 2 in the second row of σ. Note σ sends 6 into 2. Hence $\sigma^{-1}(2) = 6$. Similarly, $\sigma^{-1}(3) = 1$, $\sigma^{-1}(4) = 3$, $\sigma^{-1}(5) = 4$, $\sigma^{-1}(6) = 2$. Thus

$$\sigma^{-1} = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 5 & 6 & 1 & 3 & 4 & 2 \end{pmatrix}$$

3.7. Consider functions $f\colon A \to B$ and $g\colon B \to C$. Prove the following:

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

(b) If f and g are onto functions, then $g \circ f$ is an onto function.

(a) Suppose $(g \circ f)(x) = (g \circ f)(y)$; then $g(f(x)) = g(f(y))$. Hence $f(x) = f(y)$ because g is one-to-one. Furthermore, $x = y$ since f is one-to-one. Accordingly $g \circ f$ is one-to-one.

(b) Let c be any arbitrary element of C. Since g is onto, there exists a $b \in B$ such that $g(b) = c$. Since f is onto, there exists an $a \in A$ such that $f(a) = b$. But then

$$(g \circ f)(a) = g(f(a)) = g(b) = c$$

Hence each $c \in C$ is the image of some element $a \in A$. Accordingly, $g \circ f$ is an onto function.

3.8. Let $f\colon \mathbf{R} \to \mathbf{R}$ be defined by $f(x) = 2x - 3$. Now f is one-to-one and onto; hence f has an inverse function f^{-1}. Find a formula for f^{-1}.

Let y be the image of x under the function f:

$$y = f(x) = 2x - 3$$

Consequently, x will be the image of y under the inverse function f^{-1}. Solve for x in terms of y in the above equation:

$$x = (y + 3)/2$$

Then $f^{-1}(y) = (y + 3)/2$. Replace y by x to obtain

$$f^{-1}(x) = \frac{x + 3}{2}$$

which is the formula for f^{-1} using the usual independent variable x.

3.9. Prove the following generalization of DeMorgan's law: For any class of sets $\{A_i\}$ we have

$$(\cup_i\ A_i)^c = \cap_i A_i^c$$

We have:

$$x \in (\cup_i A_i)^c \quad \text{iff } x \notin \cup_i A_i, \quad \text{iff } \forall_i \in I, x \notin A_i, \quad \text{iff } \forall_i \in I, x \in A_i^c, \quad \text{iff } x \in \cap_i A_i^c$$

Therefore, $(\cup_i A_i)^c = \cap_i A_i^c$. (Here we have used the logical notations iff for "if and only" if and $\forall$ for "for all.")

CARDINALITY

3.10. Find the cardinal number of each set:

(a) $A = \{a, b, c, \ldots, y, z\}$, (c) $C = \{10, 20, 30, 40, \ldots\}$,

(b) $B = \{x \mid x \in \mathbf{N},\ x^2 = 5\}$, (d) $D = \{6, 7, 8, 9, \ldots\}$.

(a) $|A| = 26$ since there are 26 letters in the English alphabet.

(b) $|B| = 0$ since there is no positive integer whose square is 5, that is, B is empty.

(c) $|C| = \aleph_0$ because $f\colon \mathbf{N} \to C$, defined by $f(n) = 10_n$, is a one-to-one correspondence between $\mathbf{N}$ and C.

(d) $|D| = \aleph_0$ because $g\colon \mathbf{N} \to D$, defined by $g(n) = n + 5$ is a one-to-one correspondence between $\mathbf{N}$ and D.

3.11. Show that the set $\mathbf{Z}$ of integers has cardinality $\aleph_0$.

The following diagram shows a one-to-one correspondence between $\mathbf{N}$ and $\mathbf{Z}$:

$$\begin{array}{lcccccccccc} \mathbf{N} = & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & \ldots \\ & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \ldots \\ \mathbf{Z} = & 0 & 1 & -1 & 2 & -2 & 3 & -3 & 4 & \ldots \end{array}$$

That is, the following function $f\colon \mathbf{N} \to \mathbf{Z}$ is one-to-one and onto:

$$f(n) = \begin{cases} n/2 & \text{if } n \text{ is even} \\ (1 - n)/2 & \text{if } n \text{ is odd} \end{cases}^{n/2}$$

Accordingly, $|\mathbf{Z}| = |\mathbf{N}| = \aleph_0$.

3.12. Let $A_1, A_2, \ldots$ be a countable number of finite sets. Prove that the union $S = \cup_i A_i$ is countable.

Essentially, we list the elements of A_1, then we list the elements of A_2 which do not belong to A_1, then we list the elements of A_3 which do not belong to A_1 or A_2, i.e., which have not already been listed, and so on. Since the A_i are finite, we can always list the elements of each set. This process is done formally as follows.

First we define sets $B_1, B_2, \ldots$ where B_i contains the elements of A_i which do not belong to preceding sets, i.e., we define

$$B_1 = A_1 \quad \text{and} \quad B_k = A_k \backslash (A_1 \cup A_2 \cup \cdots \cup A_{k-1})$$

Then the B_i are disjoint and $S = \cup_i B_i$. Let $b_{i1}, b_{i2}, \ldots, b_{im}$, be the elements of B_i. Then $S = \{b_{ij}\}$. Let $f: S \to \mathbf{N}$ be defined as follows:

$$f(b_{ij}) = m_1 + m_2 + \cdots + m_{i-1} + j$$

If S is finite, then S is countable. If S is infinite then f is a one-to-one correspondence between S and $\mathbf{N}$. Thus S is countable.

3.13. Prove Theorem 3.2: A countable union of countable sets is countable.

Suppose $A_1, A_2, A_3, \ldots$ are a countable number of countable sets. In particular, suppose $a_{i1}, a_{i2}, a_{i3}, \ldots$ are the elements of A_i. Define sets $B_2, B_3, B_4, \ldots$ as follows:

$$B_k = \{a_{ij} \mid i + j = k\}$$

For example, $B_6 = \{a_{15},\ a_{24},\ a_{33},\ a_{12},\ a_{51}\}$. Observe that each B_k is finite and

$$S = \cup_i A_i = \cup_k B_k$$

By the preceding problem $\cup_k B_k$ is countable. Hence $S = \cup_i A_i$ is countable and the theorem is proved.

3.14. Prove Theorem 3.3: The set $\mathbf{I}$ of all real numbers between 0 and 1 inclusive is uncountable.

The set $\mathbf{I}$ is clearly infinite, since it contains $1, \frac{1}{2}, \frac{1}{3}, \ldots$. Suppose $\mathbf{I}$ is denumerable. Then there exists a one-to-one correspondence $f: N \to \mathbf{I}$. Let $f(1) = a_1$, $f(2) = a_2, \ldots$; that is, $\mathbf{I} = \{a_1, a_2, a_3, \ldots\}$. We list the elements $a_1, a_2, \ldots$. in a column and express each in its decimal expansion:

$$\begin{aligned} a_1 &= 0.x_{11}x_{12}x_{13}x_{14}\ldots \\ a_2 &= 0.x_{21}x_{22}x_{23}x_{24}\ldots \\ a_3 &= 0.x_{31}x_{32}x_{33}x_{34}\ldots \\ a_4 &= 0.x_{41}x_{42}x_{43}x_{44}\ldots \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots \end{aligned}$$

where $x_{ij} \in \{0,\ 1,\ 2,\ \ldots,\ 9\}$. (For those numbers which can be expressed in two different decimal expansions, e.g., $0.2000000\ldots = 0.1999999\ldots$, we choose the expansion which ends with nines.)

Let $b = 0.y_1y_2y_3y_4\ldots$ be the real number obtained as follows:

$$y_i = \begin{cases} 1 & \text{if } x_{ii} \neq 1 \\ 2 & \text{if } x_{ii} = 1 \end{cases}$$

Now $b \in \mathbf{I}$. But

$$\begin{aligned} &b \neq a_1 \text{ because } y_1 \neq x_{11} \\ &b \neq a_2 \text{ because } y_2 \neq x_{22} \\ &b \neq a_3 \text{ because } y_3 \neq x_{33} \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots \end{aligned}$$

Therefore b does not belong to $\mathbf{I} = \{a_1, a_2, \ldots\}$. This contradicts the fact that $b \in \mathbf{I}$. Hence the assumption that $\mathbf{I}$ is denumerable must be false, so $\mathbf{I}$ is uncountable.

SPECIAL MATHEMATICAL FUNCTIONS

3.15. Find: (*a*) $\lfloor 7.5 \rfloor$, $\lfloor -7.5 \rfloor$, $\lfloor -18 \rfloor$; (*b*) $\lceil 7.5 \rceil$, $\lceil -7.5 \rceil$, $\lceil -18 \rceil$.

(*a*) By definition, $\lfloor x \rfloor$ denotes the greatest integer that does not exceed x, hence $\lfloor 7.5 \rfloor = 7$, $\lfloor -7.5 \rfloor = -8$, $\lfloor -18 \rfloor = -18$.

(*b*) By definition, $\lceil x \rceil$ denotes the least integer that is not less than x, hence $\lceil 7.5 \rceil = 8$, $\lceil -7.5 \rceil = -7$, $\lceil -18 \rceil = -18$.

3.16. Find: (*a*) 25 (mod7); (*b*) 25 (mod5); (*c*) $-35 \pmod{11}$; (*d*) $-3 \pmod{8}$.

When k is positive, simply divide k by the modulus M to obtain the remainder r. Then $r = k(\text{mod } M)$. If k is negative, divide $|k|$ by M to obtain the remainder r'. Then $k(\text{mod } M) = M - r'$ (when $r' \neq 0$). Thus:

(*a*) $25 \pmod 7 = 4$ (*c*) $-35 \pmod{11} = 11 - 2 = 9$

(*b*) $25 \pmod 5 = 0$ (*d*) $-3 \pmod 8 = 8 - 3 = 5$

3.17. Evaluate modulo $M = 15$: (*a*) $9 + 13$; (*b*) $7 + 11$; (*c*) $4 - 9$; (*d*) $2 - 10$.

Use $a \pm M = a(\text{mod } M)$:

(*a*) $9 + 13 = 22 = 22 - 15 = 7$ (*c*) $4 - 9 = -5 = -5 + 15 = 10$

(*b*) $7 + 11 = 18 = 18 - 15 = 3$ (*d*) $2 - 10 = -8 = -8 + 15 = 7$

3.18. Simplify: (*a*) $\dfrac{n!}{(n-1)!}$; (*b*) $\dfrac{(n+2)!}{n!}$.

(*a*) $\dfrac{n!}{(n-1)!} = \dfrac{n(n-1)(n-2)\cdots 3\cdot 2\cdot 1}{(n-1)(n-2)\cdots 3\cdot 2\cdot 1} = n$ or, simply, $\dfrac{n!}{(n-1)!} = \dfrac{n(n-1)!}{(n-1)!} = n$

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

3.19. Evaluate: (*a*) $\log_2 8$; (*b*) $\log_2 64$; (*c*) $\log_{10} 100$; (*d*) $\log_{10} 0.001$.

(*a*) $\log_2 8 = 3$ since $2^3 = 8$ (*c*) $\log_{10} 100 = 2$ since $10^2 = 100$

(*b*) $\log_2 64 = 6$ since $2^6 = 64$ (*d*) $\log_{10} 0.001 = -3$ since $10^{-3} = 0.001$

RECURSIVE FUNCTIONS

3.20. Let a and b be positive integers, and suppose Q is defined recursively as follows:

$$Q(a, b) = \begin{cases} 0 & \text{if } a < b \\ Q(a-b, b) + 1 & \text{if } b \leq a \end{cases}$$

(*a*) Find: (i) $Q(2, 5)$; (ii) $Q(12, 5)$.

(*b*) What does this function Q do? Find $Q(5861, 7)$.

(*a*) (i) $Q(2, 5) = 0$ since $2 < 5$.

(ii) $Q(12, 5) = Q(7, 5) + 1$
$= [Q(2, 5) + 1] + 1 = Q(2, 5) + 2$
$= 0 + 2 = 2$

(*b*) Each time b is subtracted from a, the value of Q is increased by 1. Hence $Q(a, b)$ finds the quotient when a is divided by b. Thus $Q(5861, 7) = 837$.

3.21. Use the definition of the Ackermann function to find $A(1, 3)$.

Figure 3-10 shows the 15 steps that are used to evaluate $A(1, 3)$.

The forward indention indicates that we are postponing an evaluation and are recalling the definition, and the backward indention indicates that we are backtracking. Observe that (*a*) of the definition is used in Steps 5, 8, 11 and 14; (*b*) in Step 4; and (*c*) in Steps 1, 2, and 3. In the other steps we are backtracking with substitutions.

(1) $A(1, 3) = A(0, A(1, 2))$
(2) $A(1, 2) = A(0, A(1, 1))$
(3) $A(1, 1) = A(0, A(1, 0))$
(4) $A(1, 0) = A(0, 1)$
(5) $A(0, 1) = 1 + 1 = 2$
(6) $A(1, 0) = 2$
(7) $A(1, 1) = A(0, 2)$
(8) $A(0, 2) = 2 + 1 = 3$
(9) $A(1, 1) = 3$
(10) $A(1, 2) = A(0, 3)$
(11) $A(0, 3) = 3 + 1 = 4$
(12) $A(1, 2) = 4$
(13) $A(1, 3) = A(0, 4)$
(14) $A(0, 4) = 4 + 1 = 5$
(15) $A(1, 3) = 5$

Fig. 3-10

MISCELLANEOUS PROBLEMS

3.22. Find the domain D of each of the following real-valued functions of a real variable:

(a) $f(x) = \frac{1}{x-2}$ (c) $f(x) = \sqrt{25 - x^2}$
(b) $f(x) = x^2 - 3x - 4$ (d) x^2 where $0 \le x \le 2$

When a real-valued function of a real variable is given by a formula $f(x)$, then the domain D consists of the largest subset of $\mathbf{R}$ for which $f(x)$ has meaning and is real, unless otherwise specified.

(a) f is not defined for $x - 2 = 0$, i.e., for $x = 2$; hence $D = \mathbf{R}\backslash\{2\}$.

(b) f is defined for every real number; hence $D = \mathbf{R}$.

(c) f is not defined when $25 - x^2$ is negative; hence $D = [-5, 5] = \{x \mid -5 \le x \le 5\}$.

(d) Here, the domain of f is explicitly given as $D = \{x \mid 0 \le x \le 2\}$.

3.23. For any $n \in \mathbf{N}$, let $D_n = (0, 1/n)$, the open interval from 0 to $1/n$. Find:

(a) $D_3 \cup D_4$; (b) $D_3 \cap D_{20}$; (c) $D_s \cup D_t$; (d) $D_s \cap D_t$.

(a) Since $(0, 1/3)$ is a superset of $(0, 1/7)$, $D_3 \cup D_4 = D_3$.

(b) Since $(0, 1/20)$ is a subset of $(0, 1/3)$, $D_3 \cap D_{20} = D_{20}$.

(c) Let $m = \min(s, t)$, that is, the smaller of the two numbers s and t; then D_m is equal to D_s or D_t contains the other as a subset. Hence $D_s \cap D_t = D_m$.

(d) Let $M = \max(s, t)$, that is, the larger of the two numbers s and t; then $D_s \cap D_t = D_m$.

3.24. Suppose $P(n) = a_0 + a_1 n + a_2 n^2 + \cdots + a_m n^2$ has degree m. Prove $P(n) = O(n^m)$.

Let $b_0 = |a_0|$, $b_1 = |a_1|, \ldots, b_m = |a_m|$. Then for $n \ge 1$,

$$\begin{aligned} p(n) &\le b_0 + b_1 n + b_2 n^2 + \cdots + b_m n^m = \left(\tfrac{b_0}{n^m} + \tfrac{b_1}{n^{m-1}} + \cdots + b_m\right) n^m \\ &\le (b_0 + b_1 + \cdots + b_m) n^m = M n^m \end{aligned}$$

where $M = |a_0| + |a_1| + \cdots + |a_m|$. Hence $P(n) = O(n^m)$.

For example, $5x^3 + 3x = O(x^3)$ and $x^5 - 4000000x^2 = O(x^5)$.

3.25. Prove Theorem 3.4 (Cantor): $|A| < |\text{Power}(A)|$ (where Power (A) is the power set *of* A).

The function $g\colon A \to \text{Power}(A)$ defined by $g(a) = \{a\}$ is clearly one-to-one; hence $|A| \le |\text{Power}(A)|$.

If we show that $|A| \ne |\text{Power}(A)|$, then the theorem will follow. Suppose the contrary, that is, suppose $|A| = |\text{Power}(A)|$ and that $f\colon A \to \text{Power}(A)$ is a function which is both one-to-one and onto. Let $a \in A$ be called a "bad" element if $a \notin f(a)$, and let B be the set of bad elements. In other words,

$$B = \{x : x \in A,\ x \notin f(x)\}$$

Now B is a subset of A. Since $f\colon A \to \text{Power}(A)$ is onto, there exists $b \in A$ such that $f(b) = B$. Is b a "bad" element or a "good" element? If $b \in B$ then, by definition of B, $b \notin f(b) = B$, which is impossible. Likewise, if $b \notin B$ then $b \in f(b) = B$, which is also impossible. Thus the original assumption that $|A| = |\text{Power}(A)|$ has led to a contradiction. Hence the assumption is false, and so the theorem is true.

3.26. Prove the following equivalent formulation of the Schroeder–Bernstein Theorem 3.5:

Suppose $X \supseteq Y \supseteq X_1$ and $X \simeq X_1$. Then $X \simeq Y$.

Since $X \simeq X_1$ there exists a one-to-one correspondence (bijection) $f\colon X \to X_1$ Since $X \supseteq Y$, the restriction of f to Y, which we also denote by f, is also one-to-one. Let $f(Y) = Y_1$. Then Y and Y_1 are equipotent,

$$X \supseteq Y \supseteq X_1 \supseteq Y_1$$

and $f\colon Y \to Y_1$ is bijective. But now $Y \supseteq X_1 \supseteq Y_1$ and $Y \simeq Y_1$. For similar reasons, X_1 and $f(X_1) = X_2$ are equipotent,

$$X \supseteq Y \supseteq X_1 \supseteq Y_1 \supseteq X_2$$

and $f\colon X_1 \to X_2$ is bijective. Accordingly, there exist equipotent sets $X, X_1, X_2, \ldots$ and equipotent sets $Y, Y_1, Y_2, \ldots$ such that

$$X \supseteq Y \supseteq X_1 \supseteq Y_1 \supseteq X_2 \supseteq Y_2 \supseteq X_3 \supseteq Y_3 \supseteq \cdots$$

and $f\colon X_k \to X_{k+1}$ and $f : Y_k \to Y_{k+1}$ are bijective.

Let

$$B = X \cap Y \cap X_1 \cap Y_1 \cap X_2 \cap Y_2 \cap \cdots$$

Then

$$X = (X \backslash Y) \cup (Y \backslash X_1) \cup (X_1 \backslash Y_1) \cup \cdots \cup B$$
$$Y = (Y \backslash X_1) \cup (X_1 \backslash Y_1) \cup (Y_1 \backslash X_2) \cup \cdots \cup B$$

Furthermore, $X \backslash Y$, $X_1 \backslash Y_1$, $X_2 \backslash Y_2, \ldots$ are equipotent. In fact, the function

$$f : (X_k \backslash Y_k) \to (X_{k+1} \backslash Y_{k+1})$$

is one-to-one and onto.

Consider the function $g\colon X \to Y$ defined by the diagram in Fig. 3-11. That is,

$$g(x) = \begin{cases} f(x) & \text{if } x \in X_k \backslash Y_k \text{ or } x \in X \backslash Y \\ x & \text{if } x \in Y_k \backslash X_k \text{ or } x \in B \end{cases}$$

Then g is one-to-one and onto. Therefore $X \simeq Y$.

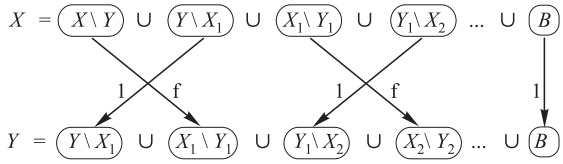


Fig. 3-11

Supplementary Problems

FUNCTIONS

3.27. Let $W = \{a, b, c, d\}$. Decide whether each set of ordered pairs is a function from W into W.

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

3.28. Let $V = \{1, 2, 3, 4\}$. For the following functions $f\colon V \to V$ and $g\colon V \to V$, find:

(a) $f \circ g$; (b), $g \circ f$; (c) $f \circ f$:

$$f = \{(1, 3), (2, 1), (3, 4), (4, 3)\} \quad \text{and} \quad g = \{(1, 2), (2, 3), (3, 1), (4, 1)\}$$

3.29. Find the composition function $h \circ g \circ f$ for the functions in Fig. 3-9.

ONE-TO-ONE, ONTO, AND INVERTIBLE FUNCTIONS

3.30. Determine if each function is one-to-one.

(a) To each person on the earth assign the number which corresponds to his age.

(b) To each country in the world assign the latitude and longitude of its capital.

(c) To each book written by only one author assign the author.

(d) To each country in the world which has a prime minister assign its prime minister.

3.31. Let functions f, g, h from $V = \{1, 2, 3, 4\}$ into V be defined by: $f(n) = 6 - n$, $g(n) = 3$, $h = \{(1, 2), (2, 3), (3, 4), (4, 1)\}$. Decide which functions are:

(a) one-to-one; (b) onto; (c) both; (d) neither.

3.32. Let functions f, g, h from $\mathbf{N}$ into $\mathbf{N}$ be defined by $f(n) = n + 2$, (b) $g(n) = 2^n$; $h(n) =$ number of positive divisors of n. Decide which functions are:

(a) one-to-one; (b) onto; (c) both; (d) neither; (e) Find $h'(2) = \{x | h(x) = 2\}$.

3.33. Decide which of the following functions are: (a) one-to-one; (b) onto; (c) both; (d) neither.

(1) f: $\mathbf{Z}^2 \to \mathbf{Z}$ where $f(n, m) = n - m$; (3) h: $\mathbf{Z} \times (\mathbf{Z} \backslash 0) \to \mathbf{Q}$ where $h(n, m) = n/m$;

(2) g: $\mathbf{Z}^2 \to \mathbf{Z}^2$ where $g(n, m) = (m, n)$; (4) k: $\mathbf{Z} \to \mathbf{Z}^2$ where $k(n) = (n, n)$.

3.34. Let f: $\mathbf{R} \to \mathbf{R}$ be defined by $f(x) = 3x - 7$. Find a formula for the inverse function f^{-1}: $\mathbf{R} \to \mathbf{R}$.

3.35. Consider permutations $\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 2 & 5 & 6 & 1 & 3 & 4 \end{pmatrix}$ and $\tau = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 6 & 4 & 3 & 1 & 2 & 5 \end{pmatrix}$ in S_6.

Find: (a) $\tau \circ \sigma$; (b) $\sigma \circ \tau$; (c) σ^2; (d) σ^{-1}; (e) τ^{-1}

PROPERTIES OF FUNCTIONS

3.36. Prove: Suppose f: $A \to B$ and g: $B \to A$ satisfy $g \circ f = 1_A$. Then f is one-to-one and g is onto.

3.37. Prove Theorem 3.1: A function f: $A \to B$ is invertible if and only if f is both one-to-one and onto.

3.38. Prove: Suppose f: $A \to B$ is invertible with inverse function f^{-1}: $B \to A$. Then $f^{-1} \circ f = 1_A$ and $f \circ f^{-1} = 1_B$.

3.39. Suppose f: $A \to B$ is one-to-one and $g : A \to B$ is onto. Let x be a subset of A.

(a) Show $f1_x$, the restriction of f to x, is one-to-one.

(b) Show $g1_x$, need not be onto.

3.40. For each $n \in \mathbf{N}$, consider the open interval $A_n = (0, 1/n) = \{x \mid 0 < x < 1/n\}$. Find:

(a) $A_2 \cup A_8$; (c) $\cup(A_i \mid i \in J)$; (e) $\cup(A_i \mid i \in K)$;

(b) $A_3 \cap A_7$; (d) $\cap(A_i \mid i \in J)$; (f) $\cap(A_i \mid i \in K)$.

where J is a finite subset of $\mathbf{N}$ and $\mathbf{K}$ is an infinite subset of $\mathbf{N}$.

3.41. For each $n \in N$, let $D_n = \{n, 2n, 3n, \ldots\} = \{\text{multiples of } n\}$.

(a) Find: (i) $D_2 \cap D_7$; (ii) $D_6 \cap D_8$; (iii) $D_3 \cap D_{12}$; (iv) $D_3 \cup D_{12}$.

(b) Prove that $\cap(D_i | i \in K) = \emptyset$ where K is an infinite subset of $\mathbf{N}$.

3.42. Consider an indexed class of sets $\{A_i \mid i \in I\}$, a set $\mathbf{B}$ and an index i_0 in I.
Prove: (a) $B \cap (\cup_i A_i) = \cup_i (B \cap A_i)$; (b) $\cap (A_i \mid i \in I) \subseteq A_{i_0} \subseteq \cup(A_i \mid i \in I)$.

CARDINAL NUMBERS

3.43. Find the cardinal number of each set: (a) $\{x \mid x$ is a letter in "BASEBALL"$\}$; (b) Power set of $A = \{a, b, c, d, e\}$; (c) $\{x \mid x^2 = 9, 2x = 8\}$.

3.44. Find the cardinal number of:

(a) all functions from $A = \{a, b, c, d\}$ into $B = \{1, 2, 3, 4, 5\}$;

(b) all functions from P into Q where $|P| = r$ and $|Q| = s$;

(c) all relations on $A = \{a, b, c, d\}$;

(d) all relations on P where $|P| = r$.

3.45. Prove:

(a) Every infinite set A contains a denumerable subset D.

(b) Each subset of a denumerable set is finite or denumerable.

(c) If A and B are denumerable, then $A \times B$ is denumerable.

(e) The set $\mathbf{Q}$ of rational numbers is denumerable.

3.46. Prove: (a) $|A \times B| = |B \times A|$; (b) If $A \subseteq B$ then $|A| \leq |B|$; (c) If $|A| = |B|$ then $P(A)| = |P(B)|$.

SPECIAL FUNCTIONS

3.47. Find: (a) $\lfloor 13.2 \rfloor$, $\lfloor -0.17 \rfloor$, $\lfloor 34 \rfloor$; (b) $\lceil 13.2 \rceil$, $\lceil -0.17 \rceil$, $\lceil 34 \rceil$.

3.48. Find:

(a) 29 (mod 6); (c) 5 (mod 12); (e) -555 (mod 11).

(b) 200 (mod 20); (d) -347 (mod 6);

3.49. Find: (a) $3! + 4!$; (b) $3!(3! + 2!)$; (c) $6!/5!$; (d) $30!/28!$.

3.50. Evaluate: (a) $\log_2 16$; (b) $\log_3 27$; (c) $\log_{10} 0.01$.

MISCELLANEOUS PROBLEMS

3.51. Let n be an integer. Find $L(25)$ and describe what the function L does where L is defined by:

$$L(n) = \begin{cases} 0 & \text{if } n = 1 \\ L(\lfloor n/2 \rfloor) + 1 & \text{if } n > 1 \end{cases}$$

3.52. Let a and b be integers. Find $Q(2, 7)$, $Q(5, 3)$, and $Q(15, 2)$, where $Q(a, b)$ is defined by:

$$Q(a, b) = \begin{cases} 5 & \text{if } a < b \\ Q(a - b,\ b + 2) + a & \text{if } a \geq b \end{cases}$$

3.53. Prove: The set P of all polynomials $p(x) = a_0 + a_1 x + \cdots + a_x^m$ with integral coefficients (that is, where $a_0, a_1, \ldots, a_m$ are integers) is denumerable.

Answers to Supplementary Problems

3.27. (a) Yes; (b) No; (c) Yes; (d) No.

3.28. (a) $\{(1, 1), (2, 4), (3, 3), (4, 3)\}$;
(b) $\{(1, 1), (2, 2), (3, 1), (4, 1)\}$;
(c) $\{(1, 4), (2, 3), (3, 3), (4, 4)\}$.

3.29. $\{(a, 4), (b, 6), (c, 4)\}$

3.30. (a) No, (b) yes, (c) no, (d) yes.

3.31. (a) f, h; (b) f, h; (c) f, h; (d) g.

3.32. (a) f, g; (b) h; (c) none; (d) none; (e) {all prime numbers}.

3.33. (a) g, k; (b) f, g, h; (c) g; (d) none.

3.34. $f^{-1}(x) = (x + 7)/3$

3.35. (a) 425631; (b) 416253; (c) 534261; (d) 415623; (e) 453261.

3.40. (a) A_2; (b) A_7; (c) A_r where r is the smallest integer in J; (d) A_s where s is the largest integer in J; (e) A_r where r is the smallest integer in K; (f) $\emptyset$.

3.41. (i) D_{14}; (ii) D_{24}; (iii) D_{12} (iv) D_3.

3.43. (a) 5; (b) $2^5 = 32$; (c) 0.

3.44. (a) $5^4 = 625$; (b) s^r; (c) $2^{16} = 65\,536$; (d) 2.

3.47. (a) 13, −1, 34; (b) 14, 0, 34.

3.48. (a) 5; (b) 0; (c) 2; (d) $6 - 5 = 1$; (e) $11 - 5 = 6$.

3.49. (a) 30; (b) 48; (c) 6; (d) 870.

3.50. (a) 4; (b) 3; (c) −2.

3.51. $L(25) = 4$. Each time n is divided by 2, the value of L is increased by 1. Hence L is the greatest integer such that $2^L < N$. Thus $L(n) = \lfloor \log_2 n \rfloor$.

3.52. $Q(2, 7) = 5$, $Q(5, 3) = 10$, $Q(15, 2) = 42$.

3.53. Hint: Let P_k denote the set of polynomials $p(x)$ such that $m \leq k$ and each $|a_i| \leq k$. P_k is finite and $P = \cup_k P_k$.

CHAPTER 14

Ordered Sets and Lattices

14.1 INTRODUCTION

Order and precedence relationships appear in many different places in mathematics and computer science. This chapter makes these notions precise. We also define a lattice, which is a special kind of an ordered set.

14.2 ORDERED SETS

Suppose R is a relation on a set S satisfying the following three properties:

$[\mathbf{O}_1]$ (Reflexive) For any $a \in S$, we have aRa.

$[\mathbf{O}_2]$ (Antisymmetric) If aRb and bRa, then $a = b$.

$[\mathbf{O}_3]$ (Transitive) If aRb and bRc, then aRc.

Then R is called a *partial order* or, simply an *order* relation, and R is said to define a *partial ordering* of S. The set S with the partial order is called a *partially ordered set* or, simply, an *ordered set* or *poset*. We write (S, R) when we want to specify the relation R.

The most familiar order relation, called the *usual order*, is the relation $\leq$ (read "less than or equal") on the positive integers $\mathbf{N}$ or, more generally, on any subset of the real numbers $\mathbf{R}$. For this reason, a partial order relation is usually denoted by $\preceq$; and

$$a \preceq b$$

is read "a precedes b." In this case we also write:

$a \prec b$ means $a \preceq b$ and $a \neq b$;	read "a strictly precedes b."
$b \succeq a$ means $a \preceq b$;	read "b succeeds a."
$b \succ a$ means $a \prec b$;	read "b strictly succeeds a."

$\not\preceq, \not\prec, \not\succeq$, and $\not\succ$ are self-explanatory.

When there is no ambiguity, the symbols $\leq$, $<$, $>$, and $\geq$ are frequently used instead of $\preceq$, $\prec$, $\succ$, and $\succeq$, respectively.

EXAMPLE 14.1

(a) Let S be any collection of sets. The relation $\subseteq$ of set inclusion is a partial ordering of S. Specifically, $A \subseteq A$ for any set A; if $A \subseteq B$ and $B \subseteq A$ then $A = B$; and if $A \subseteq B$ and $B \subseteq C$ then $A \subseteq C$.

(b) Consider the set $\mathbf{N}$ of positive integers. We say "a divides b," written $a \mid b$, if there exists an integer c such that $ac = b$. For example, $2 \mid 4$, $3 \mid 12$, $7 \mid 21$, and so on. This relation of divisibility is a partial ordering of $\mathbf{N}$.

(c) The relation "$\mid$" of divisibility is not an ordering of the set $\mathbf{Z}$ of integers. Specifically, the relation is not antisymmetric. For instance, $2 \mid -2$ and $-2 \mid 2$, but $2 \neq -2$.

(d) Consider the set $\mathbf{Z}$ of integers. Define aRb if there is a positive integer r such that $b = a^r$. For instance, $2\ R\ 8$ since $8 = 2^3$. Then R is a partial ordering of $\mathbf{Z}$.

Dual Order

Let $\preceq$ be any partial ordering of a set S. The relation $\succeq$, that is, a succeeds b, is also a partial ordering of S; it is called the *dual order*. Observe that $a \preceq b$ if and only if $b \succeq a$; hence the dual order $\succeq$ is the inverse of the relation $\preceq$, that is, $\succeq = \preceq^{-1}$.

Ordered Subsets

Let A be a subset of an ordered set S, and suppose $a, b \in A$. Define $a \preceq b$ as elements of A whenever $a \preceq b$ as elements of S. This defines a partial ordering of A called the *induced order* on A. The subset A with the induced order is called an *ordered subset* of S. Unless otherwise stated or implied, any subset of an ordered set S will be treated as an ordered subset of S.

Quasi-order

Suppose $\prec$ is a relation on a set S satisfying the following two properties:

$[\mathbf{Q}_1]$ (Irreflexive) For any $a \in A$, we have $a \not\prec a$.

$[\mathbf{Q}_2]$ (Transitive) If $a \prec b$, and $b \prec c$, then $a \prec c$.

Then $\prec$ is called a *quasi-order* on S.

There is a close relationship between partial orders and quasi-orders. Specifically, if $\preceq$ is a partial order on a set S and we define $a \prec b$ to mean $a \preceq b$ but $a \neq b$, then $\prec$ is a quasi-order on S. Conversely, if $\prec$ is a quasi-order on a set S and we define $a \preceq b$ to mean $a \prec b$ or $a = b$, then $\preceq$ is a partial order on S. This allows us to switch back and forth between a partial order and its corresponding quasi-orders using whichever is more convenient.

Comparability, Linearly Ordered Sets

Suppose a and b are elements in a partially ordered set S. We say a and b are *comparable* if

$$a \preceq b \quad \text{or} \quad b \preceq a$$

that is, if one of them precedes the other. Thus a and b are *noncomparable*, written

$$a \parallel b$$

if neither $a \preceq b$ nor $b \preceq a$.

The word "partial" is used in defining a partially ordered set S since some of the elements of S need not be comparable. Suppose, on the other hand, that every pair of elements of S are comparable. Then S is said to be *totally ordered* or *linearly ordered*, and S is called a *chain*. Although an ordered set S may not be linearly ordered, it is still possible for a subset A of S to be linearly ordered. Clearly, every subset of a linearly ordered set S must also be linearly ordered.

EXAMPLE 14.2

(a) Consider the set **N** of positive integers ordered by divisibility. Then 21 and 7 are comparable since $7 \mid 21$. On the other hand, 3 and 5 are noncomparable since neither $3 \mid 5$ nor $5 \mid 3$. Thus **N** is not linearly ordered by divisibility. Observe that $A = \{2, 6, 12, 36\}$ is a linearly ordered subset of **N** since $2 \mid 6$, $6 \mid 12$ and $12 \mid 36$.

(b) The set **N** of positive integers with the usual order $\leq$ (less than or equal) is linearly ordered and hence every ordered subset of **N** is also linearly ordered.

(c) The power set $P(A)$ of a set A with two or more elements is not linearly ordered by set inclusion. For instance, suppose a and b belong to A. Then $\{a\}$ and $\{b\}$ are noncomparable. Observe that the empty set $\varnothing$, $\{a\}$, and A do form a linearly ordered subset of $P(A)$ since $\varnothing \subseteq \{a\} \subseteq A$. Similarly, $\varnothing$, $\{b\}$, and A form a linearly ordered subset of $P(A)$.

Product Sets and Order

There are a number of ways to define an order relation on the Cartesian product of given ordered sets. Two of these ways follow:

(*a*) ***Product Order*:** Suppose S and T are ordered sets. Then the following is an order relation on the product set $S \times T$, called the *product order*:

$$(a, b) \precsim (a', b') \quad \text{if} \quad a \leq a' \text{ and } b \leq b'$$

(*b*) ***Lexicographical Order*:** Suppose S and T are linearly ordered sets. Then the following is an order relation on the product set $S \times T$, called the *lexicographical* or *dictionary order*:

$$(a, b) \prec (a', b') \quad \text{if} \quad a < b \quad \text{or if} \quad a = a' \text{ and } b < b'$$

This order can be extended to $S_1 \times S_2 \times \cdots \times S_n$ as follows:

$$(a_1, a_2, \ldots, a_n) \prec (a'_1, a'_2, \ldots, a'_n) \quad \text{if} \quad a_i = a'_i \text{ for } i = 1, 2, \ldots, k-1 \text{ and } a_k < a'_k$$

Note that the lexicographical order is also linear.

Kleene Closure and Order

Let A be a (nonempty) linearly ordered alphabet. Recall that A^*, called the Kleene closure of A, consists of all words w on A, and $|w|$ denotes the length of w. Then the following are two order relations on A^*.

(*a*) ***Alphabetical (Lexicographical) Order*:** The reader is no doubt familiar with the usual alphabetical ordering of A^*. That is:

(i) $\lambda \prec w$, where λ is the empty word and w is any nonempty word.

(ii) Suppose $u = au'$ and $v = bv'$ are distinct nonempty words where $a, b \in A$ and $u', v' \in A^*$. Then

$$u \prec v \quad \text{if } a < b \quad \text{or} \quad \text{if } a = b \text{ but } u' \prec v'$$

(*b*) ***Short-lex Order*:** Here A^* is ordered first by length, and then alphabetically. That is, for any distinct words u, v in A^*,

$$u \prec v \quad \text{if } |u| < |v| \quad \text{or} \quad \text{if } |u| = |v| \text{ but } u \text{ precedes } v \text{ alphabetically}$$

For example, "to" precedes "and" since $|\text{to}| = 2$ but $|\text{and}| = 3$. However, "an" precedes "to" since they have the same length, but "an" precedes "to" alphabetically. This order is also called the *free semigroup order*.

14.3 HASSE DIAGRAMS OF PARTIALLY ORDERED SETS

Let S be a partially ordered set, and suppose a, b belong to S. We say that a is an *immediate predecessor* of b, or that b is an *immediate successor* of a, or that b is a *cover* of a, written

$$a \ll b$$

if $a < b$ but no element in S lies between a and b, that is, there exists no element c in S such that $a < c < b$.

Suppose S is a finite partially ordered set. Then the order on S is completely known once we know all pairs a, b in S such that $a \ll b$, that is, once we know the relation $\ll$ on S. This follows from the fact that $x < y$ if and only if $x \ll y$ or there exist elements $a_1, a_2, \ldots, a_m$ in S such that

$$x \ll a_1 \ll a_2 \ll \cdots \ll a_m \ll y$$

The *Hasse diagram* of a finite partially ordered set S is the directed graph whose vertices are the elements of S and there is a directed edge from a to b whenever $a \ll b$ in S. (Instead of drawing an arrow from a to b, we sometimes place b higher than a and draw a line between them. It is then understood that movement upwards indicates succession.) In the diagram thus created, there is a directed edge from vertex x to vertex y if and only if $x \ll y$. Also, there can be no (directed) cycles in the diagram of S since the order relation is antisymmetric.

The Hasse diagram of a poset S is a picture of S; hence it is very useful in describing types of elements in S. Sometimes we define a partially ordered set by simply presenting its Hasse diagram. We note that the Hasse diagram of a poset S need not be connected.

Remark: The Hasse diagram of a finite poset S turns out to be a directed cycle-free graph (DAG) studied in Section 9.9. The investigation here is independent of the previous investigation. Here we mainly think of order in terms of "less than" or "greater than" rather than in terms of directed adjacency relations. Accordingly, there will be some overlap in the content.

EXAMPLE 14.3

(a) Let $A = \{1, 2, 3, 4, 6, 8, 9, 12, 18, 24\}$ be ordered by the relation "x divides y." The diagram of A is given in Fig. 14-1(a). (Unlike rooted trees, the direction of a line in the diagram of a poset is always upward.)

(b) Let $B = \{a, b, c, d, e\}$. The diagram in Fig. 14-1(b) defines a partial order on B in the natural way. That is, $d \leq b$, $d \leq a$, $e \leq c$ and so on.

(c) The diagram of a finite linearly ordered set, i.e., a finite chain, consists simply of one path. For example, Fig. 14-1(c) shows the diagram of a chain with five elements.

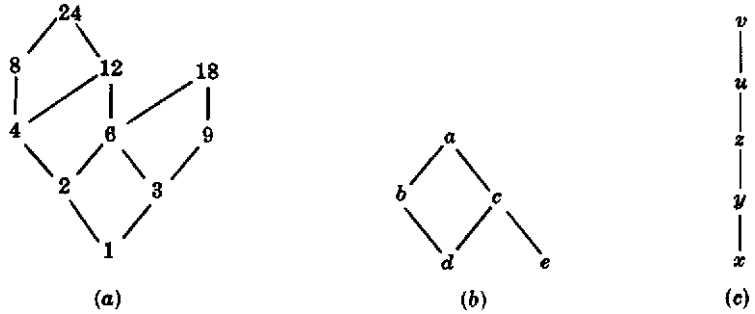


Fig. 14-1

EXAMPLE 14.4 A *partition* of a positive integer m is a set of positive integers whose sum is m. For instance, there are seven partitions of $m = 5$ as follows:

$$5, \quad 3-2, \quad 2-2-1, \quad 1-1-1-1-1, \quad 4-1, \quad 3-1-1, \quad 2-1-1-1$$

We order the partitions of an integer m as follows. A partition P_1 precedes a partition P_2 if the integers in P_1 can be added to obtain the integers in P_2 or, equivalently, if the integers in P_2 can be further subdivided to obtain the integers in P_1. For example,

$$2-2-1 \quad \text{precedes} \quad 3-2$$

since $2+1=3$. On the other hand, $3-1-1$ and $2-2-1$ are noncomparable.

Figure 14-2 gives the Hasse diagram of the partitions of $m = 5$.

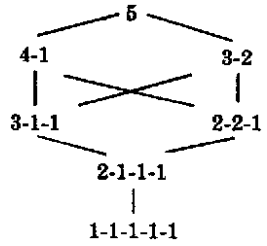


Fig. 14-2

Minimal and Maximal, and First and Last Elements

Let S be a partially ordered set. An element a in S is called a *minimal* element if no other element of S strictly precedes (is less than) a. Similarly, an element b in S is called a *maximal* element if no element of S strictly succeeds (is larger than) b. Geometrically speaking, a is a minimal element if no edge enters a (from below), and b is a maximal element if no edge leaves b (in the upward direction). We note that S can have more than one minimal and more than one maximal element.

If S is infinite, then S may have no minimal and no maximal element. For instance, the set $\mathbf{Z}$ of integers with the usual order $\leq$ has no minimal and no maximal element. On the other hand, if S is finite, then S must have at least one minimal element and at least one maximal element.

An element a is S is called a *first* element if for every element x in S,

$$a \precsim x$$

that is, if a precedes every other element in S. Similarly, an element b in S is called a *last* element if for every element y in S,

$$y \precsim b$$

that is, if b succeeds every other element in S. We note that S can have at most one first element, which must be a minimal element, and S can have at most one last element, which must be a maximal element. Generally speaking, S may have neither a first nor a last element, even when S is finite.

EXAMPLE 14.5

(a) Consider the three partially ordered sets in Example 14-3 whose Hasse diagrams appear in Fig. 14-1.

(i) A has two maximal elements 18 and 24 and neither is a last element. A has only one minimal element, 1, which is also a first element.

(ii) B has two minimal elements, d and e, and neither is a first element. B has only one maximal element, a, which is also a last element.

(iii) The chain has one minimal element, x, which is a first element, and one maximal element, v, which is a last element.

(b) Let A be any nonempty set and let $P(A)$ be the power set of A ordered by set inclusion. Then the empty set $\varnothing$ is a first element of $P(A)$ since, for any set X, we have $\varnothing \subseteq X$. Moreover, A is a last element of $P(A)$ since every element Y of $P(A)$ is, by definition, a subset of A, that is, $Y \subseteq A$.

14.4 CONSISTENT ENUMERATION

Suppose S is a finite partially ordered set. Frequently we want to assign positive integers to the elements of S in such a way that the order is preserved. That is, we seek a function $f: S \to \mathbf{N}$ so that if $a \prec b$ then $f(a) < f(b)$. Such a function is called a *consistent enumeration* of S. The fact that this can always be done is the content of the following theorem.

Theorem 14.1: There exists a consistent enumeration for any finite poset A.

We prove this theorem in Problem 14.8. In fact, we prove that if S has n elements then there exists a consistent enumeration $f: S \to \{1, 2, \ldots, n\}$.

We emphasize that such an enumeration need not be unique. For example, the following are two such enumerations for the poset in Fig. 14-1(b):

(i) $f(d) = 1$, $f(e) = 2$, $f(b) = 3$, $f(c) = 4$, $f(a) = 5$.

(ii) $g(e) = 1$, $g(d) = 2$, $g(c) = 3$, $g(b) = 4$, $g(a) = 5$.

However the chain in Fig. 14-1(c) admits only one consistent enumeration if we map the set into $\{1, 2, 3, 4, 5\}$. Specifically, we must assign:

$$h(x) = 1, \quad h(y) = 2, \quad h(z) = 3, \quad h(u) = 4, \quad h(v) = 5$$

14.5 SUPREMUM AND INFIMUM

Let A be a subset of a partially ordered set S. An element M in S is called an *upper bound* of A if M succeeds every element of A, i.e., if, for every x in A, we have

$$x \precsim M$$

If an upper bound of A precedes every other upper bound of A, then it is called the *supremum* of A and is denoted by

$$\sup(A)$$

We also write $\sup(a_1, \ldots, a_n)$ instead of $\sup(A)$ if A consists of the elements $a_1, \ldots, a_n$. We emphasize that there can be at most one $\sup(A)$; however, $\sup(A)$ may not exist.

Analogously, an element m in a poset S is called a *lower bound* of a subset A of S if m precedes every element of A, i.e., if, for every y in A, we have

$$m \precsim y$$

If a lower bound of A succeeds every other lower bound of A, then it is called the *infimum* of A and is denoted by

$$\inf(A), \quad \text{or} \quad \inf(a_1, \ldots, a_n)$$

if A consists of the elements $a_1, \ldots, a_n$. There can be at most one $\inf(A)$ although $\inf(A)$ may not exist.

Some texts use the term *least upper bound* instead of supremum and then write $\text{lub}(A)$ instead of $\sup(A)$, and use the term *greatest lower bound* instead of infimum and write $\text{glb}(A)$ instead of $\inf(A)$.

If A has an upper bound we say A is *bounded above*, and if A has a lower bound we say A is *bounded below*. In particular, A is *bounded* if A has an upper and lower bound.

EXAMPLE 14.6

(a) Let $S = \{a, b, c, d, e, f\}$ be ordered as pictured in Fig. 14-3(a), and let $A = \{b, c, d\}$. The upper bounds of A are e and f since only e and f succeed every element in A. The lower bounds of A are a and b since only a and b precede every element of A. Note that e and f are noncomparable; hence $\sup(A)$ does not exist. However, b also succeeds a, hence $\inf(A) = b$.

(b) Let $S = \{1, 2, 3, \ldots, 8\}$ be ordered as pictured in Fig. 14-3(b), and let $A = \{4, 5, 7\}$. The upper bounds of A are 1, 2, and 3, and the only lower bound is 8. Note that 7 is not a lower bound since 7 does not precede 4. Here $\sup(A) = 3$ since 3 precedes the other upper bounds 1 and 2. Note that $\inf(A) = 8$ since 8 is the only lower bound.

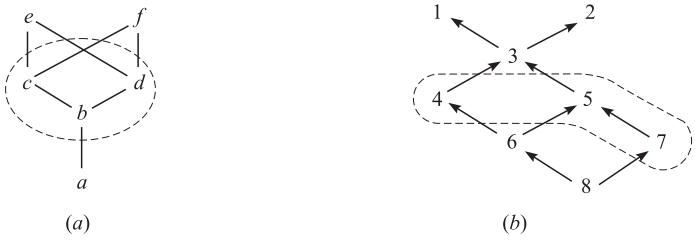


Fig. 14-3

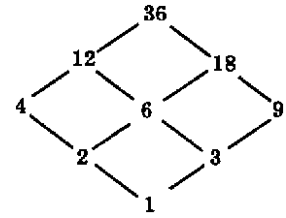


Fig. 14-4

Generally speaking, $\sup(a, b)$ and $\inf(a, b)$ need not exist for every pair of elements a and b in a poset S. We now give two examples of partially ordered sets where $\sup(a, b)$ and $\inf(a, b)$ do exist for every a, b in the set.

EXAMPLE 14.7

(a) Let the set **N** of positive integers be ordered by divisibility. The *greatest common divisor* of a and b in **N**, denoted by

$$\gcd(a, b)$$

is the largest integer which divides a and b. The *least common multiple* of a and b, denoted by

$$\text{lcm}(a, b)$$

is the smallest integer divisible by both a and b.

An important theorem in number theory says that every common divisor of a and b divides $\gcd(a, b)$. One can also prove that $\text{lcm}(a, b)$ divides every multiple of a and b. Thus

$$\gcd(a, b) = \inf(a, b) \quad \text{and} \quad \text{lcm}(a, b) = \sup(a, b)$$

In other words, $\inf(a, b)$ and $\sup(a, b)$ do exist for any pair of elements of **N** ordered by divisibility.

(b) For any positive integer m, we will let $\mathbf{D}_m$ denote the set of divisors of m ordered by divisibility. The Hasse diagram of

$$\mathbf{D}_{36} = \{1, 2, 3, 4, 6, 9, 12, 18, 36\}$$

appears in Fig. 14-4. Again, $\inf(a, b) = \gcd(a, b)$ and $\sup(a, b) = \text{lcm}(a, b)$ exist for any pair a, b in $\mathbf{D}_m$.

14.6 ISOMORPHIC (SIMILAR) ORDERED SETS

Suppose X and Y are partially ordered sets. A one-to-one (injective) function $f: X \to Y$ is called a *similarity mapping* from X into Y if f preserves the order relation, that is, if the following two conditions hold for any pair a and a' in X:

(1) If $a \precsim a'$ then $f(a) \precsim f(a')$.

(2) If $a \parallel a'$ (noncomparable), then $f(a) \parallel f(a')$.

Accordingly, if A and B are linearly ordered, then only (1) is needed for f to be a similarity mapping.

Two ordered sets X and Y are said to be *isomorphic* or *similar*, written

$$X \simeq Y$$

if there exists a one-to-one correspondence (bijective mapping) $f: X \to Y$ which preserves the order relations, i.e., which is a similarity mapping.

EXAMPLE 14.8 Suppose $X = \{1, 2, 6, 8, 12\}$ is ordered by divisibility and suppose $Y = \{a, b, c, d, e\}$ is isomorphic to X; say, the following function f is a similarity mapping from X onto Y:

$$f = \{(1, e), (2, d), (6, b), (8, c), (12, a)\}$$

Draw the Hasse diagram of Y.

The similarity mapping preserves the order of the initial set X and is one-to-one and onto. Thus the mapping can be viewed simply as a relabeling of the vertices in the Hasse diagram of the initial set X. The Hasse diagrams for both X and Y appear in Fig. 14-5.

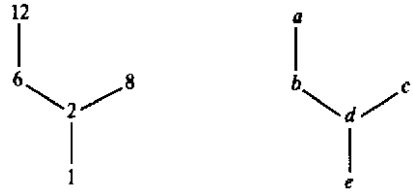


Fig. 14-5

14.7 WELL-ORDERED SETS

We begin with a definition.

Definition 14.1: An ordered set S is said to be *well-ordered* if every subset of S has a first element.

The classical example of a well-ordered set is the set $\mathbf{N}$ of positive integers with the usual order $\leq$. The following facts follow from the definition.

(1) A well-ordered set is linearly ordered. For if $a, b, \in S$, then $\{a, b\}$ has a first element; hence a and b are comparable.
(2) Every subset of a well-ordered set is well-ordered.
(3) If X is well-ordered and Y is isomorphic to X, then Y is well-ordered.
(4) All finite linearly ordered sets with the same number n of elements are well-ordered and are all isomorphic to each other. In fact, they are all isomorphic to $\{1, 2, \ldots, n\}$ with the usual order $\leq$.

(5) Every element $a \in S$, other than a last element, has an immediate successor. For, let $M(a)$ denote the set of elements which strictly succeed a. Then the first element of $M(a)$ is the immediate successor of a.

EXAMPLE 14.9

(a) The set **Z** of integers with the usual order $\leq$ is linearly ordered and every element has an immediate successor and an immediate predecessor, but **Z** is not well-ordered. For example, **Z** itself has no first element. However, any subset of **Z** which is bounded from below is well-ordered.

(b) The set **Q** of rational numbers with the usual order $\leq$ is linearly ordered, but no element in **Q** has an immediate successor or an immediate predecessor. For if $a, b \in Q$, say $a < b$, then $(a + b)/2 \in \mathbf{Q}$ and

$$a < \frac{a+b}{2} < b$$

(*c*) Consider the disjoint well-ordered sets

$$A = \{1, 3, 5, \ldots, \} \quad \text{and} \quad B = \{2, 4, 6, \ldots\}$$

Then the following ordered set

$$S = \{A; B\} = \{1, 3, 5, \ldots; 2, 4, 6, \ldots\}$$

is well-ordered. Note that, besides the first element 1, the element 2 does not have an immediate predecessor.

Notation: Here and subsequently, if $A, B, \ldots$ are disjoint ordered sets, then $\{A; B; \ldots\}$ means the set $A \cup B \cup \ldots$ ordered positionwise from left to right; that is, the elements in the same set keep their order, and any element in a set on the left precedes any element in a set on its right. Thus every element in A precedes every element in B, and so on.

Transfinite Induction

First we restate the principle of mathematical induction. (See Section 1.8 and 11.3.)

Principle of Mathematical Induction: Let A be a subset of the set **N** of positive integers with the following two properties:

(i) $1 \in A$.
(ii) If $k \in A$, then $k + 1 \in A$.

Then $A = \mathbf{N}$.

The above principle is one of Peano's axioms for the natural numbers (positive integers) **N**. There is another form which is sometimes more convenient to use. Namely:

Principle of Mathematical Induction (Second Form): Let A be a subset of **N** with the following two properties:

(i) $1 \in A$.
(ii) If j belongs to A for $1 \leq j < k$, then $k \in A$.

Then $A = \mathbf{N}$.

The second form of induction is equivalent to the fact that **N** is well-ordered (Theorem 11.6). In fact, there is a somewhat similar statement which is true for every well-ordered set.

Principle of Transfinite Induction: Let A be a subset of a well-ordered set S with the following two properties:

(i) $a_0 \in A$.
(ii) If $s(a) \subseteq A$, then $a \in A$.

Then $A = S$.

Here a_0 is the first element of S, and $s(a)$, called the *initial segment* of a, is defined to be the set of all elements of S which strictly precede a.

Axiom of Choice, Well-Ordering Theorem

Let $\{A_i \mid i \in I\}$ be a collection of nonempty disjoint sets. We assume every $A_i \subseteq X$. A function $f\colon \{A_i\} \to X$ is called a *choice function* if $f(A_i) = a_i \in A_i$. In other words, f "chooses" a point $a_i \in A_i$ for each set A_i.

The axiom of choice lies at the foundations of mathematics and, in particular, the theory of sets. This "innocent looking" axiom, which follows, has as a consequence some of the most powerful and important results in mathematics.

Axiom of Choice: There exists a choice function for any nonempty collection of nonempty disjoint sets.

One of the consequences of the axiom of choice is the following theorem, which is attributed to Zermelo.

Well-Ordering Theorem: Every set S can be well-ordered.

The proof of this theorem lies beyond the scope of this text. Moreover, since all of our structures are finite or countable, we will not need to use this theorem. Ordinary mathematical induction suffices.

14.8 LATTICES

There are two ways to define a lattice L. One way is to define L in terms of a partially ordered set. Specifically, a lattice L may be defined as a partially ordered set in which $\inf(a, b)$ and $\sup(a, b)$ exist for any pair of elements $a, b \in L$. Another way is to define a lattice L axiomatically. This we do below.

Axioms Defining a Lattice

Let L be a nonempty set closed under two binary operations called *meet* and *join*, denoted respectively by $\wedge$ and $\vee$. Then L is called *lattice* if the following axioms hold where a, b, c are elements in L:

$[\mathbf{L}_1]$ Commutative law:
$(1a)$ $a \wedge b = b \wedge a$ $\qquad$ $(1b)$ $a \vee b = b \vee a$

$[\mathbf{L}_2]$ Associative law:
$(2a)$ $(a \wedge b) \wedge c = a \wedge (b \wedge c)$ $\qquad$ $(2b)$ $(a \vee b) \vee c = a \vee (b \vee c)$

$[\mathbf{L}_3]$ Absorption law:
$(3a)$ $a \wedge (a \vee b) = a$ $\qquad$ $(3b)$ $a \vee (a \wedge b) = a$

We will sometimes denote the lattice by $(L, \wedge, \vee)$ when we want to show which operations are involved.

Duality and the Idempotent Law

The *dual* of any statement in a lattice $(L, \wedge, \vee)$ is defined to be the statement that is obtained by interchanging $\wedge$ and $\vee$. For example, the dual of

$$a \wedge (b \vee a) = a \vee a \quad \text{is} \quad a \vee (b \wedge a) = a \wedge a$$

Notice that the dual of each axiom of a lattice is also an axiom. Accordingly, the principle of duality holds; that is:

Theorem 14.2 (Principle of Duality): The dual of any theorem in a lattice is also a theorem.

This follows from the fact that the dual theorem can be proven by using the dual of each step of the proof of the original theorem.

An important property of lattices follows directly from the absorption laws.

Theorem 14.3 (Idempotent Law): (i) $a \wedge a = a$; (ii) $a \vee a = a$.

The proof of (i) requires only two lines:

$$\begin{aligned} a \wedge a &= a \wedge (a \vee (a \wedge b)) && \text{(using } (3b)\text{)} \\ &= a && \text{(using } (3a)\text{)} \end{aligned}$$

The proof of (ii) follows from the above principle of duality (or can be proved in a similar manner).

Lattices and Order

Given a lattice L, we can define a partial order on L as follows:

$$a \precsim b \quad \text{if} \quad a \wedge b = a$$

Analogously, we could define

$$a \precsim b \quad \text{if} \quad a \vee b = b$$

We state these results in a theorem.

Theorem 14.4: Let L be a lattice. Then:

(i) $a \wedge b = a$ if and only if $a \vee b = b$.

(ii) The relation $a \precsim b$ (defined by $a \wedge b = a$ or $a \vee b = b$) is a partial order on L.

Now that we have a partial order on any lattice L, we can picture L by a diagram as was done for partially ordered sets in general.

EXAMPLE 14.10 Let C be a collection of sets closed under intersection and union. Then $(C, \cap, \cup)$ is a lattice. In this lattice, the partial order relation is the same as the set inclusion relation. Figure 14-6 shows the diagram of the lattice L of all subsets of $\{a, b, c\}$.

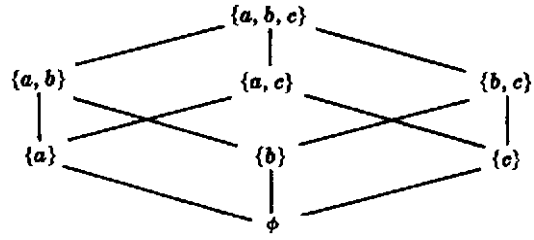


Fig. 14-6

We have shown how to define a partial order on a lattice L. The next theorem tells us when we can define a lattice on a partially ordered set P such that the lattice will give back the original order on P.

Theorem 14.5: Let P be a poset such that the $\inf(a, b)$ and $\sup(a, b)$ exist for any a, b in P. Letting

$$a \wedge b = \inf(a, b) \quad \text{and} \quad a \vee b = \sup(a, b)$$

we have that $(P, \wedge, \vee)$ is a lattice. Furthermore, the partial order on P induced by the lattice is the same as the original partial order on P.

The converse of the above theorem is also true. That is, let L be a lattice and let $\precsim$ be the induced partial order on L. Then $\inf(a, b)$ and $\sup(a, b)$ exist for any pair a, b in L and the lattice obtained from the poset $(L, \precsim)$ is the original lattice. Accordingly, we have the following:

Alternate Definition: A lattice is a partially ordered set in which

$$a \wedge b = \inf(a, b) \quad \text{and} \quad a \vee b = \sup(a, b)$$

exist for any pair of elements a and b.

We note first that any linearly ordered set is a lattice since $\inf(a, b) = a$ and $\sup(a, b) = b$ whenever $a \precsim b$. By Example 14.7, the positive integers $\mathbf{N}$ and the set $\mathbf{D}_m$ of divisors of m are lattices under the relation of divisibility.

Sublattices, Isomorphic Lattices

Suppose M is a nonempty subset of a lattice L. We say M is *a sublattice* of L if M itself is a lattice (with respect to the operations of L). We note that M is a sublattice of L if and only if M is closed under the operations of $\wedge$ and $\vee$ of L. For example, the set $\mathbf{D}_m$ of divisors of m is a sublattice of the positive integers $\mathbf{N}$ under divisibility.

Two lattices L and L' are said to be *isomorphic* if there is a one-to-one correspondence $f: L \to L'$ such that

$$f(a \wedge b) = f(a) \wedge f(b) \quad \text{and} \quad f(a \vee b) = f(a) \vee f(b)$$

for any elements a, b in L.

14.9 BOUNDED LATTICES

A lattice L is said to have a *lower bound* 0 if for any element x in L we have $0 \precsim x$. Analogously, L is said to have an *upper bound* I if for any x in L we have $x \precsim I$. We say L is *bounded* if L has both a lower bound 0 and an upper bound I. In such a lattice we have the identities

$$a \vee I = I, \quad a \wedge I = a, \quad a \vee 0 = a, \quad a \wedge 0 = 0$$

for any element a in L.

The nonnegative integers with the usual ordering,

$$0 < 1 < 2 < 3 < 4 < \cdots$$

have 0 as a lower bound but have no upper bound. On the other hand, the lattice $P(U)$ of all subsets of any universal set $\mathbf{U}$ is a bounded lattice with $\mathbf{U}$ as an upper bound and the empty set $\varnothing$ as a lower bound.

Suppose $L = \{a_1, a_2, \ldots, a_n\}$ is a finite lattice. Then

$$a_1 \vee a_2 \vee \cdots \vee a_n \quad \text{and} \quad a_1 \wedge a_2 \wedge \cdots \wedge a_n$$

are upper and lower bounds for L, respectively. Thus we have

Theorem 14.6: Every finite lattice L is bounded.

14.10 DISTRIBUTIVE LATTICES

A lattice L is said to be *distributive* if for any elements a, b, c in L we have the following:

[$\mathbf{L}_4$] Distributive law:

$$(4a)\ a \wedge (b \vee c) = (a \wedge b) \vee (a \wedge c) \qquad (4b)\ a \vee (b \wedge c) = (a \vee b) \wedge (a \vee c)$$

Otherwise, L is said to be *nondistributive*. We note that by the principle of duality the condition $(4a)$ holds if and only if $(4b)$ holds.

Figure 14-7(a) is a nondistributive lattice since

$$a \vee (b \wedge c) = a \vee 0 = a \quad \text{but} \quad (a \vee b) \wedge (a \vee c) = I \wedge c = c$$

Figure 14-7(b) is also a nondistributive lattice. In fact, we have the following characterization of such lattices.

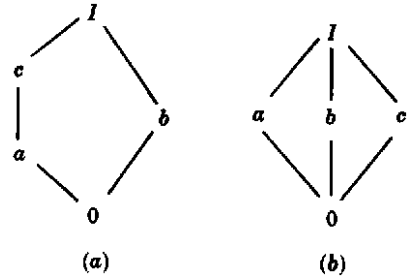


Fig. 14-7

Theorem 14.7: A lattice L is nondistributive if and only if it contains a sublattice isomorphic to Fig. 14-7(a) or to Fig. 14.7(b).

The proof of this theorem lies beyond the scope of this text.

Join Irreducible Elements, Atoms

Let L be a lattice with a lower bound 0. An element a in L is said to be *join irreducible* if $a = x \vee y$ implies $a = x$ or $a = y$. (Prime numbers under multiplication have this property, i.e., if $p = ab$ then $p = a$ or $p = b$ where p is prime.) Clearly 0 is join irreducible. If a has at least two immediate predecessors, say, b_1 and b_2 as in Fig. 14-8(a), then $a = b_1 \vee b_2$, and so a is not join irreducible. On the other hand, if a has a unique immediate predecessor c, then $a \neq \sup(b_1, b_2) = b_1 \vee b_2$ for any other elements b_1 and b_2 because c would lie between the b's and a as in Fig. 14-8(b). In other words, $a \neq 0$, is join irreducible if and only if a has a unique immediate predecessor. Those elements which immediately succeed 0, called *atoms*, are join irreducible. However, lattices can have other join irreducible elements. For example, the element c in Fig. 14-7(a) is not an atom but is join irreducible since a is its only immediate predecessor.

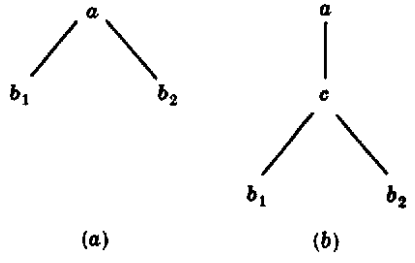


Fig. 14-8

If an element a in a finite lattice L is not join irreducible, then we can write $a = b_1 \vee b_2$. Then we can write b_1 and b_2 as the join of other elements if they are not join irreducible; and so on. Since L is finite we finally have

$$a = d_1 \vee d_2 \vee \cdots \vee d_n$$

where the d's are join irreducible. If d_i precedes d_j then $d_i \vee d_j = d_j$; so we can delete the d_i from the expression. In other words, we can assume that the d's are *irredundant*, i.e., no d precedes any other d. We emphasize that such an expression need not be unique, e.g., $I = a \vee b$ and $I = b \vee c$ in both lattices in Fig. 14-7. We now state the main theorem of this section (proved in Problem 14.28.)

Theorem 14.8: Let L be a finite distributive lattice. Then every a in L can be written uniquely (except for order) as the join of irredundant join irreducible elements.

Actually this theorem can be generalized to lattices with *finite length*, i.e., where all linearly ordered subsets are finite. (Problem 14.30 gives an infinite lattice with finite length.)

14.11 COMPLEMENTS, COMPLEMENTED LATTICES

Let L be a bounded lattice with lower bound 0 and upper bound I. Let a be an element of L. An element x in L is called a *complement* of a if

$$a \vee x = I \quad \text{and} \quad a \wedge x = 0$$

Complements need not exist and need not be unique. For example, the elements a and c are both complements of b in Fig. 14-7(a). Also, the elements y, z, and u in the chain in Fig. 14-1 have no complements. We have the following result.

Theorem 14.9: Let L be a bounded distributive lattice. Then complements are unique if they exist.

Proof: Suppose x and y are complements of any element a in L. Then

$$a \vee x = I, \quad a \vee y = I, \quad a \wedge x = 0, \quad a \wedge y = 0$$

Using distributivity,

$$x = x \vee 0 = x \vee (a \wedge y) = (x \vee a) \wedge (x \vee y) = I \wedge (x \vee y) = x \vee y$$

Similarly,

$$y = y \vee 0 = y \vee (a \wedge x) = (y \vee a) \wedge (y \vee x) = I \wedge (y \vee x) = y \vee x$$

Thus

$$x = x \vee y = y \vee x = y$$

and the theorem is proved.

Complemented Lattices

A lattice L is said to be *complemented* if L is bounded and every element in L has a complement. Figure 14-7(b) shows a complemented lattice where complements are not unique. On the other hand, the lattice $P(\mathbf{U})$ of all subsets of a universal set $\mathbf{U}$ is complemented, and each subset A of $\mathbf{U}$ has the unique complement $A^c = \mathbf{U} \backslash A$.

Theorem 14.10: Let L be a complemented lattice with unique complements. Then the join irreducible elements of L, other than 0, are its atoms.

Combining this theorem and Theorems 14.8 and 14.9, we get an important result.

Theorem 14.11: Let L be a finite complemented distributive lattice. Then every element a in L is the join of a unique set of atoms.

Remark: Some texts define a lattice L to be complemented if each a in L has a unique complement. Theorem 14.10 is then stated differently.

Solved Problems

ORDERED SETS AND SUBSETS

14.1. Let $\mathbf{N} = \{1, 2, 3, \ldots\}$ be ordered by divisibility. State whether each of the following subsets of $\mathbf{N}$ are linearly (totally) ordered.

(a) $\{24, 2, 6\}$; (c) $\mathbf{N} = \{1, 2, 3 \ldots\}$; ($e$) $\{7\}$;
(b) $\{3, 15, 5\}$; (d) $\{2, 8, 32, 4\}$; (f) $\{15, 5, 30\}$.

(a) Since 2 divides 6 which divides 24, the set is linearly ordered.

(b) Since 3 and 5 are not comparable, the set is not linearly ordered.

(c) Since 2 and 3 are not comparable, the set is not linearly ordered.

(d) This set is linearly ordered since $2 \prec 4 \prec 8 \prec 32$.

(e) Any set consisting of one element is linearly ordered.

(f) Since 5 divides 15 which divides 30, the set is linearly ordered.

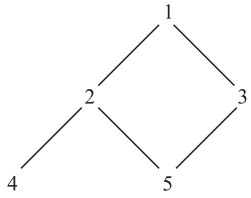
14.2. Let $A = \{1, 2, 3, 4, 5\}$ be ordered by the Hasse diagram in Fig. 14-9(a).

(a) Insert the correct symbol, $\prec$, $\succ$, or $\|$ (not comparable), between each pair of elements:
(i) 1 ___ 5; (ii) 2 ___ 3; (iii) 4 ___ 1; (iv) 3 ___ 4.

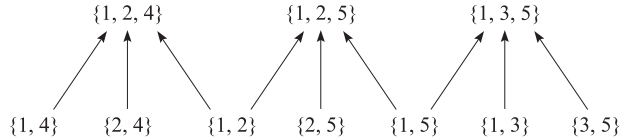
(b) Find all minimal and maximal elements of A.

(c) Does A have a first element or a last element?

(d) Let $L(A)$ denote the collection of all linearly ordered subsets of A with 2 or more elements, and let $L(A)$ be ordered by set inclusion. Draw the Hasse diagram of $L(A)$.



(a)



(b)

Fig. 14-9

(a) (i) Since there is a "path" (edges slanting upward) from 5 to 3 to 1, 5 precedes 1; hence $1 \succ 5$.
(ii) There is no path from 2 to 3, or vice versa; hence $2 \| 3$.
(iii) There is a path from 4 to 2 to 1; hence $4 \prec 1$.
(iv) Neither $3 \prec 4$ nor $4 \prec 3$; hence $3 \| 4$.

(b) No element strictly precedes 4 or 5, so 4 and 5 are minimal elements of A. No element strictly succeeds 1, so 1 is a maximal element of A.

(c) A has no first element. Although 4 and 5 are minimal elements of A, neither precedes the other. However, 1 is a last element of A since 1 succeeds every element of A.

(d) The elements of $L(A)$ are as follows:

$$\{1, 2, 4\}, \quad \{1, 2, 5\}, \quad \{1, 3, 5\}, \quad \{1, 2\}, \quad \{1, 4\}, \quad \{1, 3), \quad \{1, 5\}, \quad \{2, 4\}, \quad \{2, 5\}, \quad \{3, 5\}$$

(Note $\{2, 5\}$ and $\{3, 4\}$ are not linearly ordered.) The diagram of $L(A)$ appears in Fig. 14-9(b).

14.3. Prerequisites in college is a familiar partial ordering of available classes. We write $A \prec B$ if course A is a prerequisite for course B. Let C be the ordered set consisting of the mathematics courses and their prerequisites appearing in Fig. 14-10(a).

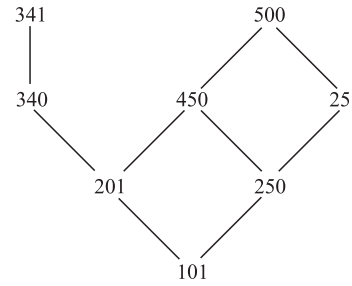
(a) Draw the Hasse diagram for the partial ordering C of these classes.

(b) Find all minimal and maximal elements of C.

(c) Does C have a first element or a last element?

Class	Prerequisites
Math 101	None
Math 201	Math 101
Math 250	Math 101
Math 251	Math 250
Math 340	Math 201
Math 341	Math 340
Math 450	Math 201, Math 250
Math 500	Math 450, Math 251

(a)



(b)

Fig. 14-10

(a) Math 101 must be on the bottom of the diagram since it is the only course with no prerequisites. Since Math 201 and Math 250 only require Math 101, we have Math 101 $\ll$ Math 201and Math 101 $\ll$ Math 250; hence draw a line slanting upward from Math 101 to Math 201 and one from Math 101 to Math 250. Continuing this process, we obtain the Hasse diagram in Fig. 14-10(b).

(b) No element strictly precedes Math 101 so Math 101 is a minimal element of C. No element strictly succeeds Math 341 or Math 500, so each is a maximal element of C.

(c) Math 101 is a first element of C since it precedes every other element of C. However, C has no last element. Although Math 341 and Math 500 are maximal elements, neither is a last element since neither precedes the other.

PRODUCT SETS AND ORDER

14.4. Suppose $\mathbf{N^2} = \mathbf{N} \times \mathbf{N}$ is given the product order (Section 14.2) where $\mathbf{N}$ has the usual order $\leq$.

Insert the correct symbol, $\prec$, $\succ$, or $\|$ (not comparable), between each of the following pairs of elements of $\mathbf{N} \times \mathbf{N}$:

(a) (5, 7) ___ (7, 1); (c) (5, 5) ___ (4, 8); (e) (7, 9) ___ (4, 1);
(b) (4, 6) ___ (4, 2); (d) (1, 3) ___ (1, 7); (f) (7, 9) ___ (8, 2).

Here $(a, b) \prec (a', b')$ if $a < a'$ and $b \leq b'$ or if $a \leq a'$ and $b < b'$. Thus:

(a) $\|$ since $5 < 7$ but $7 > 1$. (c) $\|$ since $5 > 4$ and $5 < 8$. (e) $\succ$ since $7 > 4$ and $9 > 1$.
(b) $\succ$ since $4 = 4$ and $6 > 2$. (d) $\prec$ since $1 = 1$ and $3 < 7$. (f) $\|$ since $7 < 8$ and $9 > 2$.

14.5. Repeat Problem 14.4 using the lexicographical ordering of $\mathbf{N^2} = \mathbf{N} \times \mathbf{N}$.

Here $(a, b) \prec (a', b')$ if $a < a'$ or if $a = a'$ but $b < b'$. Thus:

(a) $\prec$ since $5 < 7$. (c) $\succ$ since $5 > 4$. (e) $\succ$ since $7 > 4$.
(b) $\succ$ since $4 = 4$ and $6 > 2$. (d) $\prec$ since $1 = 1$ but $3 < 7$. (f) $\prec$ since $7 < 8$.

14.6. Consider the English alphabet $\mathbf{A} = \{a, b, c, \ldots, y, z\}$ with the usual (alphabetical) order. (Recall that $\mathbf{A}^*$ consisting of all words in $\mathbf{A}$.) Consider the following list of words in $\mathbf{A}^*$:

went, forget, to, medicine, me, toast, melt, for, we, arm

(a) Sort the list of words using the short-lex (free semigroup) order.

(*b*) Sort the list of words using the usual (alphabetical) order of $\mathbf{A}^*$.

(*a*) First order the elements by length and then order them lexicographically (alphabetically):

me, to, we, arm, for, melt, went, toast, forget, medicine

(*b*) The usual (alphabetical) ordering yields:

arm, for, forget, me, medicine, melt, to, toast, we, went

CONSISTENT ENUMERATIONS

14.7. Suppose a student wants to take all eight mathematics courses in Problem 14.3, but only one per semester.

(*a*) Which choice or choices does she have for her first and for her last (eighth) semester?

(*b*) Suppose she wants to take Math 250 in her first year (first or second semester) and Math 340 in her senior year (seventh or eighth semester). Find all the ways that she can take the eight courses.

(*a*) By Fig. 14-10, Math 101 is the only minimal element and hence must be taken in the first semester, and Math 341 and 500 are the maximal elements and hence one of them must be taken in the last semester.

(*b*) Math 250 is not a minimal element and hence must be taken in the second semester, and Math 340 is not a maximal element so it must be taken in the seventh semester and Math 341 in the eighth semester. Also Math 500 must be taken in the sixth semester. The following gives the three possible ways to take the eight courses:

101, 250, 251, 201, 450, 500, 340, 341, 101, 250, 201, 251, 450, 500, 340, 341,
101, 250, 201, 450, 251, 500, 340, 341

14.8. Prove Theorem 14.1: Suppose S is a finite poset with n elements. Then there exists a consistent enumeration $f: S \to \{1, 2, \ldots, n\}$.

The proof is by induction on the number n of elements in S. Suppose $n = 1$, say $S = \{s\}$. Then $f(s) = 1$ is a consistent enumeration of S. Now suppose $n > 1$ and the theorem holds for posets with fewer than n elements. Let $a \in S$ be a minimal element. (Such an element a exists since S is finite.) Let $T = S \backslash \{a\}$. Then T is a finite poset with $n - 1$ elements and hence, by induction, T admits a consistent enumeration; say $g: T \to \{1, 2, \ldots, n-1\}$. Define $f: S \to \{1, 2, \ldots, n\}$ by:

$$f(x) = \begin{cases} 1, & \text{if } x = a \\ g(x) + 1 & \text{if } x \neq a \end{cases}$$

Then f is the required consistent enumeration.

UPPER AND LOWER BOUNDS, SUPREMUM AND INFIMUM

14.9. Let $S = \{a, b, c, d, e, f, g\}$ be ordered as in Fig. 14-11(*a*), and let $X = \{c, d, e\}$.

(*a*) Find the upper and lower bounds of X.

(*b*) Identify $\sup(X)$, the supremum of X, and $\inf(X)$, the infimum of X, if either exists.

(*a*) The elements e, f, and g succeed every element of X; hence e, f, and g are the upper bounds of X. The element a precedes every element of X; hence a is the lower bound of X. Note that b is not a lower bound since b does not precede c; in fact, b and c are not comparable.

(*b*) Since e precedes both f and g, we have $e = \sup(X)$. Likewise, since a precedes (trivially) every lower bound of X, we have $a = \inf(X)$. Note that $\sup(X)$ belongs to X but $\inf(X)$ does not belong to X.

14.10. Let $S = \{1, 2, 3, \ldots, 8\}$ be ordered as in Fig. 14-11(*b*), and let $A = (2, 3, 6\}$.
(*a*) Find the upper and lower bounds of A. (*b*) Identify $\sup(A)$ and $\inf(A)$ if either exists.

(*a*) The upper bound is 2, and the lower bounds are 6 and 8.

(*b*) Here $\sup(A) = 2$ and $\inf(A) = 6$.

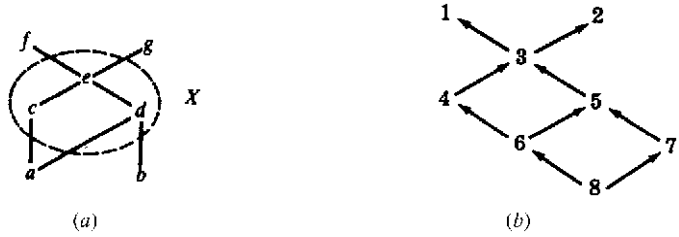


Fig. 14-11

14.11. Repeat Problem 14.10 for the subset $B = \{1, 2, 5\}$.

(*a*) There is no upper bound for B since no element succeeds both 1 and 2. The lower bounds are 6, 7, 8.

(*b*) Trivially, $\sup(A)$ does not exist since there are no upper bounds. Although A has three lower bounds, $\inf(A)$ does not exist since no lower bound succeeds both 6 and 7.

14.12. Consider the set $\mathbf{Q}$ of rational numbers with the usual order $\leq$. Consider the subset D of $\mathbf{Q}$ defined by

$$D = \{x \mid x \in \mathbf{Q} \text{ and } 8 < x^3 < 15\}$$

(*a*) Is D bounded above or below? (*b*) Does $\sup(D)$ or $\inf(D)$ exist?

(*a*) The subset D is bounded both above and below. For example, 1 is a lower bound and 100 an upper bound.

(*b*) We claim that $\sup(D)$ does not exist. Suppose, on the contrary, $\sup(D) = x$. Since $\sqrt[3]{15}$ is irrational, $x > \sqrt[3]{15}$. However, there exists a rational number y such that $\sqrt[3]{15} < y < x$. Thus y is also an upper bound for D. This contradicts the assumption that $x = \sup(D)$. On the other hand, $\inf(D)$ does exist. Specifically, $\inf(D) = 2$.

ISOMORPHIC (SIMILAR) SETS, SIMILARITY MAPPINGS

14.13. Suppose a poset A is isomorphic (similar) to a poset B and $f\colon A \to B$ is a similarity mapping. Are the following statements true or false?

(*a*) An element $a \in A$ is a first (last, minimal, or maximal) element of A if and only if $f(a)$ is a first (last, minimal, or maximal) element of B.

(*b*) An element $a \in A$ immediately precedes an element $a' \in A$, that is, $a \ll a'$, if and only if $f(a) \ll f(a')$.

(*c*) An element $a \in A$ has r immediate successors in A if and only if $f(a)$ has r immediate successors in B.

All the statements are true; the order structure of A is the same as the order structure of B.

14.14. Let $S = \{a, b, c, d, e\}$ be the ordered set in Fig. 14-12(*a*). Suppose $A = \{1, 2, 3, 4, 5\}$ is isomorphic to S. Draw the Hasse diagram of A if the following is a similarity mapping from S to A:

$$f = \{(a, 1), (b, 3), (c, 5), (d, 2), (e, 4)\}$$

The similarity mapping f preserves the order structure of S and hence f may be viewed simply as a relabeling of the vertices in the diagram of S. Thus Fig. 14-12(*b*) shows the Hasse diagram of A.

14.15. Let $A = \{1, 2, 3, 4, 5\}$ is ordered as in Fig. 14-12(*b*). Find the number n of similarity mappings $f\colon A \to A$.

Since 1 is the only minimal element of A and 4 is the only maximal element, we must have $f(1) = 1$ and $f(4) = 4$. Also, $f(3) = 3$ since 3 is the only immediate successor of 1. On the other hand, there are two possibilities for $f(2)$ and $f(5)$, that is, we can have $f(2) = 2$ and $f(5) = 5$, or $f(2) = 5$ and $f(5) = 2$. Accordingly, $n = 2$.

14.16. Give an example of a finite nonlinearly ordered set $X = (A, R)$ which is isomorphic to $Y = (A, R^{-1})$, the set A with the inverse order.

Let R be the partial ordering of $A = \{a, b, c, d, e\}$ pictured in Fig. 14-13(*a*).

Then Fig. 14-13(*b*) shows A with the inverse order R. (The diagram of R is simply turned upside down to obtain R^{-1}.) Notice that the two diagrams are identical except for the labeling. Thus X is isomorphic to Y.

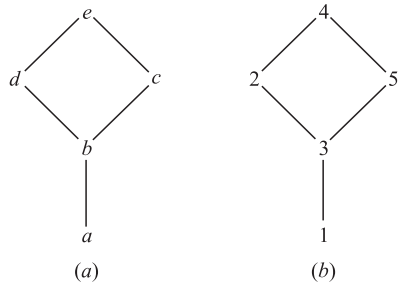


Fig. 14-12

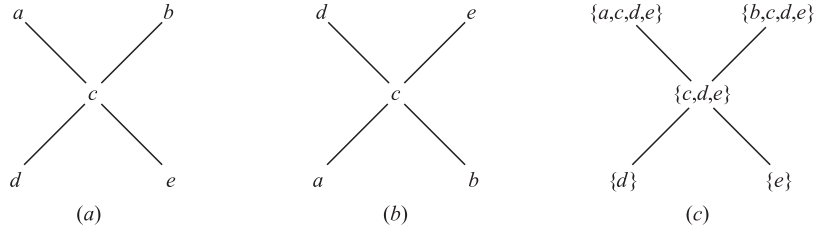


Fig. 14-13

14.17. Let A be an ordered set and, for each $a \in A$, let $p(a)$ denote the set of predecessors of a:

$$p(a) = \{x \mid x \precsim a\}$$

(called the *predecessor set* of a). Let $p(A)$ denote the collection of all predecessor sets of the elements in A ordered by set inclusion.

(a) Show that A and $p(A)$ are isomorphic by showing that the map $f\colon A \to p(A)$, defined by $f(a) = p(a)$, is a similarity mapping of A onto $p(A)$.

(b) Find the Hasse diagram of $p(A)$ for the set A in Fig. 14-13(a).

(a) First we show that f preserves the order relation of A. Suppose $a \precsim b$. Let $x \in p(a)$. Then $x \precsim a$, and hence $a \precsim b$; so $x \in p(b)$. Thus $p(a) \subseteq p(b)$. Suppose $a \parallel b$ (noncomparable). Then $a \in p(a)$ but $a \notin p(b)$; hence $p(a) \not\subseteq p(b)$. Similarly, $b \in p(b)$ but $b \notin p(a)$; hence $p(b) \not\subseteq p(a)$. Therefore, $p(a) \parallel p(b)$. Thus f preserves order.

We now need only show that f is one-to-one and onto. Suppose $y \in p(A)$. Then $y = p(a)$ for some $a \in A$. Thus $f(a) = p(a) = y$ so f is onto $p(A)$. Suppose $a \neq b$. Then $a \prec b$, $b \prec a$ or $a \parallel b$. In the first and third cases, $b \in p(b)$ but $b \notin p(a)$, and in the second case $a \in p(a)$ but $a \notin p(b)$. Accordingly, in all three cases, we have $p(a) \neq p(b)$. Therefore f is one-to-one.

Consequently, f is a similarity mapping of A onto $p(A)$ and so $A \simeq p(A)$.

(b) The elements of $p(A)$ follow:

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

Figure 14-13(c) gives the diagram of $p(A)$ ordered by set inclusion. Observe that the diagrams in Fig. 14-13(a) and (c) are identical except for the labeling of the vertices.

WELL-ORDERED SETS

14.18. Prove the Principle of Transfinite Induction: Let A be a subset of a well-ordered set S with the following two properties: (i) $a_0 \in A$. (ii) If $s(a) \subseteq A$ then $a \in A$. Then $A = S$.

(Here a_0 is the first element of A, and $s(a)$ is the initial segment of a, i.e., the set of all elements strictly preceding a.) Suppose $A \neq S$. Let $B = S \backslash A$. Then $B \neq \varnothing$. Since S is well-ordered, B has a first element b_0. Each element $x \in s(b_0)$

precedes b_0 and hence does not belong to B. Thus every $x \in s(b_0)$ belongs to A; hence $s(b_0) \subseteq A$. By (ii), $b_0 \in A$. This contradicts the assumption that $b_0 \in S \backslash A$. Thus the original assumption that $A \neq S$ is not true. Therefore $A = S$.

14.19. Let S be a well-ordered set with first element a_0. Define *a limit element* of S.

An element $b \in$ S is a limit element if $b \neq a_0$ and b has no immediate predecessor.

14.20. Consider the set $\mathbf{N} = \{1, 2, 3, \ldots\}$ of positive integers. Every number in $\mathbf{N}$ can be written uniquely as a product of a nonnegative power of 2 times an odd number. Suppose $a, a' \in \mathbf{N}$ and

$$a = 2^r(2s + 1) \quad \text{and} \quad a' = 2^{r'}(2s' + 1)$$

where r, r' and s, s' are nonnegative integers. Define:

$$a \prec a' \quad \text{if } r < r' \quad \text{or} \quad \text{if } r = r' \text{ but } s < s'.$$

(a) Insert the correct symbol, $\prec$ or $\succ$, between each pair of numbers:

(i) 5 ___ 14; (ii) 6 ___ 9; (iii) 3 ___ 20; (iv) 14 ___ 21

(b) Let $S = (\mathbf{N}, \prec)$. Show that S is well-ordered.

(c) Does S have any limit elements?

(a) The elements of $\mathbf{N}$ can be listed as in Fig. 14-14. The first row consists of the odd numbers, the second row of 2 times the odd numbers, the third row of $2^2 = 4$ times the odd numbers, and so on. Then $a \prec a'$ if a is in higher row then, a' or if a and a' are in the same row but a comes before a' in the row. Accordingly:

(i) $5 \prec 14$; (ii) $6 \succ 9$; (iii) $3 \succ 20$; (iv) $14 \succ 20$.

r \ s	0	1	2	3	4	5	6	7	
0	1	3	5	7	9	11	13	15	...
1	2	6	10	14	18	22	26	30	...
2	4	12	20	28	36	44	52	60	...
	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	

Fig. 14-14

(b) Let A be a subset of S. The rows are well-ordered. Let r_0 denote the minimum row of elements in A. In r_0 there may be many elements of A. The columns are well-ordered, so let s_0 denote the minimum column of the elements of A in row r_0. Then $x = (r_0, s_0)$ is the first element of A. Thus S is well-ordered.

(c) As indicated by Fig. 14-14, every power of 2, that is, 1, 2, 4, 8, ..., has no immediate predecessor. Thus each, other than 1, is a limit element of S.

14.21. Let S be a well-ordered set. Let $f\colon S \to S$ be a similarity mapping of S into S. Prove that, for every $a \in S$, we have $a \precsim f(a)$.

Let $D = \{x | f(x) \prec x\}$. If D is empty, then the statement is true. Suppose $D \neq \varnothing$. Since D is well-ordered, D has a first element, say d_0. Since $d_0 \in D$, we have $f(d_0) \prec d_0$. Since f is a similarity mapping:

$$f(d_0) \prec d_0 \quad \text{implies} \quad f(f(d_0)) \prec f(d_0)$$

Thus $f(d_0)$ also belongs to D. But $f(d_0) \prec d_0$ and $f(d_0) \in D$ contradicts the fact that d_0 is the first element of D. Hence the original assumption that $D \neq \varnothing$ leads to a contradiction. Therefore D is empty and the statement is true.

14.22. Let A be a well-ordered set. Let $s(A)$ denote the collection of all initial segments $s(a)$ of elements $a \in A$ ordered by set inclusion. Prove A is isomorphic to $s(A)$ by showing that the map $f\colon A \to s(A)$, defined by $f(a) = s(a)$, is a similarity mapping of A onto $s(A)$. (Compare with Problem 14.17.)

First we show that f is one-to-one and onto. Suppose $y \in s(A)$. Then $y = s(a)$ for some $a \in A$. Thus $f(a) = s(a) = y$, so f is onto $s(A)$. Suppose $x \neq y$. Then one precedes the other, say, $x \prec y$. Then $x \in s(y)$. But $x \notin s(x)$. Thus $s(x) \neq s(y)$. Therefore, f is also one-to-one.

We now need only show that f preserves order, that is,

$$x \precsim y \quad \text{if and only if} \quad s(x) \subseteq s(y)$$

Suppose $x \precsim y$. If $a \in s(x)$, then $a \prec x$ and hence $a \prec y$; thus $a \in s(y)$. Thus $s(x) \subseteq s(y)$. On the other hand, suppose $x \not\precsim y$, that is, $x \succ y$. Then $y \in s(x)$. But $y \notin s(y)$; hence $s(x) \not\subseteq s(y)$. In other words, $x \precsim y$ if and only if $s(x) \subseteq s(y)$. Accordingly, f is a similarity mapping of A onto $S(A)$, and so $A \cong s(A)$.

LATTICES

14.23. Write the dual of each statement:

(a) $(a \wedge b) \vee c = (b \vee c) \wedge (c \vee a)$; $\quad (b)$ $(a \wedge b) \vee a = a \wedge (b \vee a)$.

Replace $\vee$ by $\wedge$ and replace $\wedge$ by $\vee$ in each statement to obtain the dual statement:

(a) $(a \vee b) \wedge c = (b \wedge c) \vee (c \wedge a)$; $\quad (b)$ $(a \vee b) \wedge a = a \vee (b \wedge a)$

14.24. Prove Theorem 14.4: Let L be a lattice. Then:

(i) $a \wedge b = a$ if and only if $a \vee b = b$.

(ii) The relation $a \precsim b$ (defined by $a \wedge b = a$ or $a \vee b = b$) is a partial order on L.

(i) Suppose $a \wedge b = a$. Using the absorption law in the first step we have:

$$b = b \vee (b \wedge a) = b \vee (a \wedge b) = b \vee a = a \vee b$$

Now suppose $a \vee b = b$. Again using the absorption law in the first step we have:

$$a = a \wedge (a \vee b) = a \wedge b$$

Thus $a \wedge b = a$ if and only if $a \vee b = b$.

(ii) For any a in L, we have $a \wedge a = a$ by idempotency. Hence $a \precsim a$, and so $\precsim$ is reflexive.

Suppose $a \precsim b$ and $b \precsim a$. Then $a \wedge b = a$ and $b \wedge a = b$. Therefore, $a = a \wedge b = b \wedge a = b$, and so $\precsim$ is antisymmetric.

Lastly, suppose $a \precsim b$ and $b \precsim c$. Then $a \wedge b = a$ and $b \wedge c = b$. Thus

$$a \wedge c = (a \wedge b) \wedge c = a \wedge (b \wedge c) = a \wedge b = a$$

Therefore $a \precsim c$, and so $\precsim$ is transitive. Accordingly, $\precsim$ is a partial order on L.

14.25. Which of the partially ordered sets in Fig. 14-15 are lattices?

A partially ordered set is a lattice if and only if $\sup(x, y)$ and $\inf(x, y)$ exist for each pair x, y in the set. Only (c) is not a lattice since $\{a, b\}$ has three upper bounds, c, d and I, and no one of them precedes the other two, that is, $\sup(a, b)$ does not exist.

14.26. Consider the lattice L in Fig. 14-15(a).

(a) Which nonzero elements are join irreducible?

(b) Which elements are atoms?

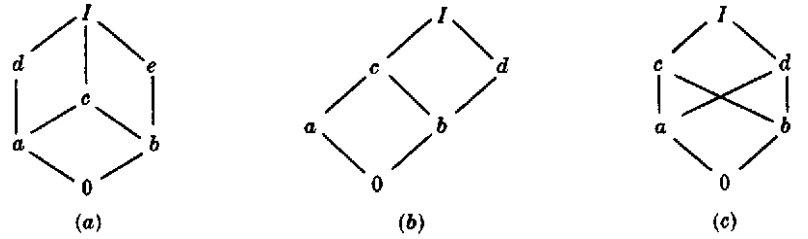


Fig. 14-15

(*c*) Which of the following are sublattices of L:

$$L_1 = \{0, a, b, I\}, \quad L_2 = \{0, a, e, I\}, \quad L_3 = \{a, c, d, I\}, \quad L_4 = \{0, c, d, I\}$$

(*d*) Is L distributive?

(*e*) Find complements, if they exist, for the elements, a, b and c.

(*f*) Is L a complemented lattice?

(*a*) Those nonzero elements with a unique immediate predecessor are join irreducible. Hence a, b, d, and e are join irreducible.

(*b*) Those elements which immediately succeed 0 are atoms, hence a and b are the atoms.

(*c*) A subset L' is a sublattice if it is closed under $\wedge$ and $\vee$. L_1 is not a sublattice since $a \vee b = c$, which does not belong to L_1. The set L_4 is not a sublattice since $c \wedge d = a$ does not belong to L_4. The other two sets, L_2 and L_3, are sublattices.

(*d*) L is not distributive since $M = \{0, a, d, e, I\}$ is a sublattice which is isomorphic to the nondistributive lattice in Fig. 14-7(*a*).

(*e*) We have $a \wedge e = 0$ and $a \vee e = I$, so a and e are complements. Similarly, b and d are complements. However, c has no complement.

(*f*) L is not a complemented lattice since c has no complement.

14.27. Consider the lattice M in Fig. 14-15(*b*).

(*a*) Find the nonzero join irreducible elements and atoms of M.

(*b*) Is M (i) distributive? (ii) complemented?

(*a*) The nonzero elements with a unique predecessor are a, b, and d, and of these three only a and b are atoms since their unique predecessor is 0.

(*b*) (i) M is distributive since M does not have a sublattice isomorphic to one of the lattices in Fig. 14-7. (ii) M is not complemented since b has no complement. Note a is the only solution to $b \wedge x = 0$ but $b \wedge a = c \neq I$.

14.28. Prove Theorem 14.8: Let L be a finite distributive lattice. Then every $a \in L$ can be written uniquely (except for order) as the join of irredundant join irreducible elements.

Since L is finite we can write a as the join of irredundant join irreducible elements as we discussed in Section 14.9. Thus we need to prove uniqueness. Suppose

$$a = b_1 \vee b_2 \vee \cdots \vee b_r = c_1 \vee c_2 \vee \cdots \vee c_s$$

where the b's are irredundant and join irreducible and the c's are irredundant and irreducible. For any given i we have

$$b_i \precsim (b_1 \vee b_2 \vee \cdots \vee b_r) = (c_1 \vee c_2 \vee \cdots \vee c_s)$$

Hence

$$b_i = b_i \wedge (c_1 \vee c_2 \vee \cdots \vee c_s) = (b_i \wedge c_1) \vee (b_i \wedge c_2) \vee \cdots \vee (b_i \wedge c_s)$$

Since b_i is join irreducible, there exists a j such that $b_i = b_i \wedge c_j$, and so $b_i \precsim c_j$. By a similar argument, for c_j there exists a b_k such that $c_j \precsim b_k$. Therefore

$$b_i \precsim c_j \precsim b_k$$

which gives $b_i = c_j = b_k$ since the b's are irredundant. Accordingly, the b's and c's may be paired off. Thus the representation for a is unique except for order.

14.29. Prove Theorem 14.10: Let L be a complemented lattice with unique complements. Then the join irreducible elements of L, other than 0, are its atoms.

Suppose a is join irreducible and a is not an atom. Then a has a unique immediate predecessor $b \neq 0$. Let b' be the complement of b. Since $b \neq 0$ we have $b' \neq I$. If a precedes b', then $b \precsim a \precsim b'$, and so $b \wedge b' = b'$, which is impossible since $b \wedge b' = I$. Thus a does not precede b', and so $a \wedge b'$ must strictly precede a. Since b is the unique immediate predecessor of a, we also have that $a \wedge b'$ precedes b as in Fig. 14-16(a). But $a \wedge b'$ precedes b'. Hence

$$a \wedge b' \precsim \inf(b, b') = b \wedge b' = 0$$

Thus $a \wedge b' = 0$. Since $a \vee b = a$, we also have that

$$a \vee b' = (a \vee b) \vee b' = a \vee (b \vee b') = a \vee I = I$$

Therefore b' is a complement of a. Since complements are unique, $a = b$. This contradicts the assumption that b is an immediate predecessor of a. Thus the only join irreducible elements of L are its atoms.

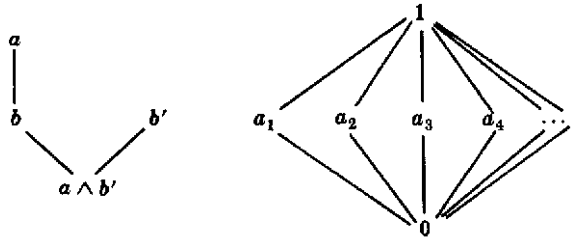


Fig. 14-16

14.30. Give an example of an infinite lattice L with finite length.

Let $L = \{0, 1, a_1, a_2, a_3, \ldots\}$ and let L be ordered as in Fig. 14-16(b). Accordingly, for each $n \in \mathbf{N}$, we have $0 < a_n < 1$. Then L has finite length since L has no infinite linearly ordered subset.

Supplementary Problems

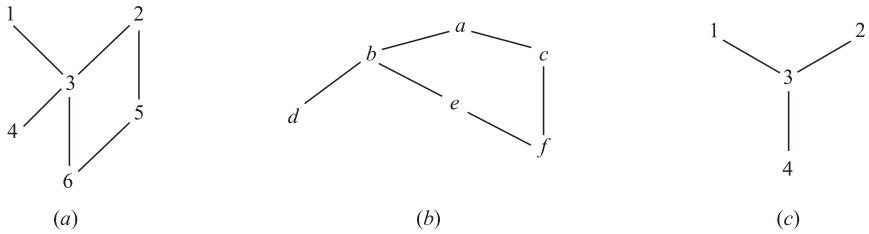
ORDERED SETS AND SUBSETS

14.31. Let $A = \{1, 2, 3, 4, 5, 6\}$ be ordered as in Fig. 14-17(a).

(a) Find all minimal and maximal elements of A.

(b) Does A have a first or last element?

(c) Find all linearly ordered subsets of A, each of which contains at least three elements.



(a) (b) (c)

Fig. 14-17

14.32. Let $B = \{a, b, c, d, e, f\}$ be ordered as in Fig. 14-17(b).

(a) Find all minimal and maximal elements of B.

(b) Does B have a first or last element?

(c) List two and find the number of consistent enumerations of B into the set $\{1, 2, 3, 4, 5, 6\}$.

14.33. Let $C = \{1, 2, 3, 4\}$ be ordered as in Fig. 14-17(c). Let $L(C)$ denote the collection of all nonempty linearly ordered subsets of C ordered by set inclusion. Draw a diagram of $L(C)$.

14.34. Draw the diagrams of the partitions of m (see Example 14.4) where: (a) $m = 4$; (b) $m = 6$.

14.35. Let $\mathbf{D}_m$ denote the positive divisors of m ordered by divisibility. Draw the Hasse diagrams of:
(a) $\mathbf{D}_{12}$; (b) $\mathbf{D}_{15}$; (c) $\mathbf{D}_{16}$; (d) $\mathbf{D}_{17}$.

14.36. Let $S = \{a, b, c, d, e, f\}$ be a poset. Suppose there are exactly six pairs of elements where the first immediately precedes the second as follows:

$$f \ll a, \quad f \ll d, \quad e \ll b, \quad c \ll f, \quad e \ll c, \quad b \ll f$$

(a) Find all minimal and maximal elements of S.

(b) Does S have any first or last element?

(c) Find all pairs of elements, if any, which are noncomparable.

14.37. State whether each of the following is true or false and, if it is false, give a counterexample.

(a) If a poset S has only one maximal element a, then a is a last element.

(b) If a finite poset S has only one maximal element a, then a is a last element.

(c) If a linearly ordered set S has only one maximal element a, then a is a last element.

14.38. Let $S = \{a, b, c, d, e\}$ be ordered as in Fig. 14-18(a).

(a) Find all minimal and maximal elements of S.

(b) Does S have any first or last element?

(c) Find all subsets of S in which c is a minimal element.

(d) Find all subsets of S in which c is a first element.

(e) List all linearly ordered subsets with three or more elements

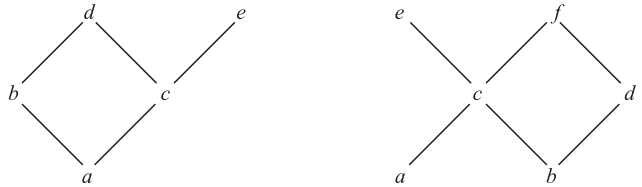


Fig. 14-18

14.39. Let $S = \{a, b, c, d, e, f\}$ be ordered as in Fig. 14-18(b).

(a) Find all minimal and maximal elements of S.

(b) Does S have any first or last element?

(c) List all linearly ordered subsets with three or more elements.

14.40. Let $S = \{a, b, c, d, e, f, g\}$ be ordered as in Fig. 14-11(a). Find the number n of linearly ordered subsets of S with: (a) four elements; (b) five elements.

14.41. Let $S = \{1, 2, \ldots, 7, 8\}$ be ordered as in Fig. 14-11(b). Find the number n of linearly ordered subsets of S with: (a) five elements; (b) six elements.

CONSISTENT ENUMERATIONS

14.42. Let $S = \{a, b, c, d, e\}$ be ordered as in Fig. 14-18(a). List all consistent enumerations of S into $\{1, 2, 3, 4, 5\}$.

14.43. Let $S = \{a, b, c, d, e, f\}$ be ordered as in Fig. 14-18(b). Find the number n of consistent enumerations of S into $\{1, 2, 3, 4, 5, 6\}$.

14.44. Suppose the following are three consistent enumerations of an ordered set $A = \{a, b, c, d\}$:

$$[(a, 1), (b, 2), (c, 3), (d, 4)], \quad [(a, 1), (b, 3), (c, 2), (d, 4)], \quad [(a, 1), (b, 4), (c, 2), (d, 3)]$$

Assuming the Hasse diagram D of A is connected, draw D.

ORDER AND PRODUCT SETS

14.45. Let $M = \{2, 3, 4, \ldots\}$ and let $M^2 = M \times M$ be ordered as follows:

$$(a, b) \prec (c, d) \quad \text{if} \quad a \mid c \text{ and } b < d$$

Find all minimal and maximal elements of $M \times M$.

14.46. Consider the English alphabet $A = \{a, b, c, \ldots, y, z\}$ with the usual (alphabetical) order. Recall A^* consists of all words in A. Let L consist of the following list of elements in A^*:

gone, or, arm, go, an, about, gate, one, at, occur

(a) Sort L according to the short-lex order, i.e., first by length and then alphabetically.

(b) Sort L alphabetically.

14.47. Consider the ordered sets A and B appearing in Fig. 14-17(a) and (b), respectively. Suppose $S = A \times B$ is given the product order. Insert the correct symbol, $\prec$, $\succ$ or $\|$, between each pair of elements of S:

(a) $(4, b)$___$(2, e)$; (b) $(3, a)$___$(6, f)$; (c) $(5, d)$___$(1, a)$; (d) $(6, e)$___$(2, b)$.

14.48. Suppose $\mathbf{N} = \{1, 2, 3, \ldots\}$ and $\mathbf{A} = \{a, b, c, \ldots, y, z\}$ are given the usual orders, and $S = \mathbf{N} \times \mathbf{A}$ is ordered lexicographically. Sort the following elements of S:

$$(2, z), \quad (1, c), \quad (2, c), \quad (1, y), \quad (4, b), \quad (4, z), \quad (3, b), \quad (2, a)$$

UPPER AND LOWER BOUNDS, SUPREMUM AND INFIMUM

14.49. Let $S = \{a, b, c, d, e, f, g\}$ be ordered as in Fig. 14-11(a). Let $A = \{a, c, d\}$.

(a) Find the set of upper bounds of A. (c) Does $\sup(A)$ exist?

(b) Find the set of lower bounds of A. (d) Does $\inf(A)$ exist?

14.50. Repeat Problem 14.49 for subset $B = \{b, c, \mathrm{e}\}$ of S.

14.51. Let $S = \{1, 2, \ldots, 7, 8\}$ be ordered as in Fig. 14-11(b). Consider the subset $A = \{3, 6, 7\}$ of S.

(a) Find the set of upper bounds of A. (c) Does $\sup(A)$ exist?

(b) Find the set of lower bounds of A. (d) Does $\inf(A)$ exist?

14.52. Repeat Problem 14.51 for the subset $B = \{1, 2, 4, 7\}$ of S.

14.53. Consider the rational numbers $\mathbf{Q}$ with the usual order $\leq$. Let $A = \{x \mid x \in \mathbf{Q} \text{ and } 5 < x^3 < 27\}$.

(a) Is A bounded above or below?

(b) Does $\sup(A)$ or $\inf(A)$ exist?

14.54. Consider the real numbers $\mathbf{R}$ with the usual order $\leq$. Let $A = \{x \mid x \in \mathbf{Q} \text{ and } 5 < x^3 < 27\}$.

(a) Is A bounded above or below? (b) Does $\sup(A)$ or $\inf(A)$ exist?

ISOMORPHIC (SIMILAR) SETS, SIMILARITY MAPPINGS

14.55. Find the number of non-isomorphic posets with three elements a, b, c, and draw their diagrams.

14.56. Find the number of connected non-isomorphic posets with four elements a, b, c, d, and draw their diagrams.

14.57. Find the number of similarity mapings $f\colon S \to S$ where S is the ordered set in:

(a) Fig. 14-17(a); (b) Fig. 14-17(b); (c) Fig. 14-17(c).

14.58. Show that the isomorphism relation $A \cong B$ for ordered sets is an equivalence relation, that is:

(a) $A \cong A$ for any ordered set A. (b) If $A \cong B$, then $B \cong A$. (c) If $A \cong B$ and $B \cong C$, then $A \cong C$.

WELL-ORDERED SETS

14.59. Let the union S of sets $A = \{a_1, a_2, a_3, \ldots\}$, $B = \{b_1, b_2, b_3, \ldots\}$, $C = \{c_1, c_2, c_3, \ldots\}$ be ordered by:

$$S = \{A; B; C\} = \{a_1, a_2, \ldots, b_1, b_2, \ldots, c_1, c_2, \ldots\}$$

(a) Show that S is well-ordered.

(b) Find all limit elements of S.

(c) Show that S is not isomorphic to $\mathbf{N} = \{1, 2, \ldots\}$ with the usual order $\leq$.

14.60. Let $A = \{a, b, c\}$ be linearly ordered by $a < b < c$, and let $\mathbf{N}$ have the usual order $\leq$.

(a) Show that $S = \{A; \mathbf{N}\}$ is isomorphic to $\mathbf{N}$.

(b) Show that $S' = \{\mathbf{N}; A\}$ is not isomorphic to $\mathbf{N}$.

14.61. Suppose A is a well-ordered set under the relation $\precsim$, and suppose A is also well-ordered under the inverse relation $\succsim$. Describe A.

14.62. Suppose A and B are well-ordered isomorphic sets. Show that there is only one similarity mapping $f\colon A \to B$.

14.63. Let S be a well-ordered set. For any $a \in S$, the set $s(a) = \{x \mid x \prec a\}$ is called an *initial segment* of a. Show that S cannot be isomorphic to one of its *initial segments*. (*Hint*: Use Problem 14.21.)

14.64. Suppose $s(a)$ and $s(b)$ are distinct initial segments of a well-ordered set S. Show that $s(a)$ and $s(b)$ cannot be isomorphic. (*Hint*: Use Problem 14.63.)

LATTICES

14.65. Consider the lattice L in Fig. 14-19(a).

(a) Find all sublattices with five elements.

(b) Find all join-irreducible elements and atoms.

(c) Find complements of a and b, if they exist.

(d) Is L distributive? Complemented?

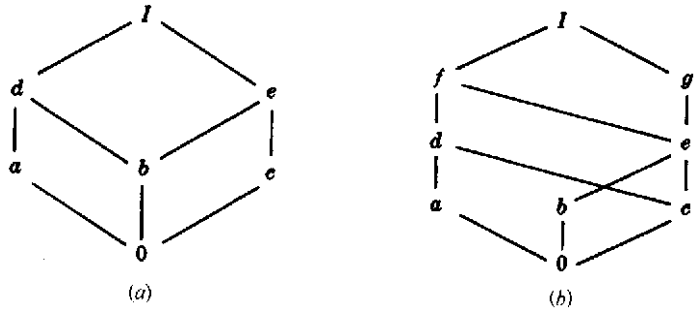


Fig. 14-19

14.66. Consider the lattice M in Fig. 14-19(b).

(a) Find all join-irreducible elements.

(b) Find the atoms.

(c) Find complements of a and b, if they exist.

(d) Express each x in M as the join of irredundant join-irreducible elements.

(e) Is M distributive? Complemented?

14.67. Consider the bounded lattice L in Fig. 14-20(a).

(a) Find the complements, if they exist, of e and f.

(b) Express I in an irredundant join-irreducible decomposition in as many ways as possible.

(c) Is L distributive?

(d) Describe the isomorphisms of L with itself.

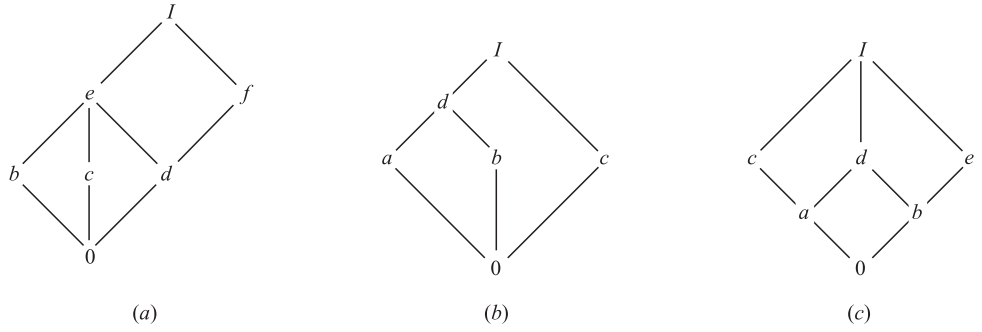


Fig. 14-20

14.68. Consider the bounded lattice L in Fig. 14-20(b).

(a) Find the complements, if they exist, of a and c.

(b) Express I in an irredundant join-irreducible decomposition in as many ways as possible.

(c) Is L distributive?

(d) Describe the isomorphisms of L with itself.

14.69. Consider the bounded lattice L in Fig. 14-20(c).

(a) Find the complements, if they exist, of a and c.

(b) Express I in an irredundant join-irreducible decomposition in as many ways as possible.

(c) Is L distributive?

(d) Describe the isomorphisms of L with itself.

14.70. Consider the lattice $\mathbf{D}_{60} = \{1, 2, 3, 4, 5, 6, 10, 12, 15, 20, 30, 60\}$, the divisors of 60 ordered by divisibility.

(a) Draw the diagram of $\mathbf{D}_{60}$.

(b) Which elements are join-irreducible and which are atoms?

(c) Find complements of 2 and 10, if they exist.

(d) Express each number x as the join of a minimum number of irredundant join irreducible elements.

14.71. Consider the lattice $\mathbf{N}$ of positive integers ordered by divisibility.

(a) Which elements are join-irreducible ?

(b) Which elements are atoms?

14.72. Show that the following "weak" distributive laws hold for any lattice L:

(a) $a \vee (b \wedge c) \leq (a \vee b) \wedge (a \vee c)$; (b) $a \wedge (b \vee c) \geq (a \wedge b) \vee (a \wedge c)$.

14.73. Let $S = \{1, 2, 3, 4\}$. We use the notation $[12, 3, 4] \equiv [\{1, 2\}, \{3\}, \{4\}]$. Three partitions of S follow:

$$P_1 = [12, 3, 4], \quad P_2 = [12, 34], \quad P_3 = [13, 2, 4]$$

(a) Find the other twelve partitions of S.

(b) Let L be the collection of the 12 partitions of S ordered by *refinement*, that is, $P_i \prec P_j$ if each cell of P_i is a subset of a cell of P_j. For example $P_1 \prec P_2$, but P_2 and P_3 are noncomparable. Show that L is a bounded lattice and draw its diagram.

14.74. An element a in a lattice L is said to be meet-irreducible if $a = x \wedge y$ implies $a = x$ or $a = y$. Find all meet-irreducible elements in: (a) Fig. 14-19(a); (b) Fig. 14-19(b); (c) $\mathbf{D}_{60}$ (see Problem 14.70.)

14.75. A lattice M is said to be *modular* if whenever $a \leq c$ we have the law

$$a \vee (b \wedge c) = (a \vee b) \wedge c$$

(a) Prove that every distributive lattice is modular.

(b) Verify that the non-distributive lattice in Fig. 14-7(b) is modular; hence the converse of (a) is not true.

(c) Show that the nondistributive lattice in Fig. 14-7(a) is non-modular. (In fact, one can prove that every non-modular lattice contains a sublattice isomorphic to Fig. 14-7(a).)

14.76. Let R be a ring. Let L be the collection of all ideals of R. Prove that L is a bounded lattice where, for any ideals J and K of R, we define: $J \vee K = J + K$ and $J \wedge K = J \cap K$.

Answers to Supplementary Problems

14.31. (a) Minimal, 4 and 6; maximal, 1 and 2. (b) First, none; last, none, (c) {1, 3, 4}, {1, 3, 6}, {2, 3, 4}, {2, 3, 6}, {2, 5, 6}.

14.32. (a) Minimal, d and f; maximal, a. (b) First, none; last, a. (c) There are eleven: *dfebca*, *dfecba*, *dfceba*, *fdebca*, *fdecba*, *fdceba*, *fedbca*, *fedcba*, *fcdeba*, *fecdba*, *fcedba*.

14.33. See Fig. 14-21.

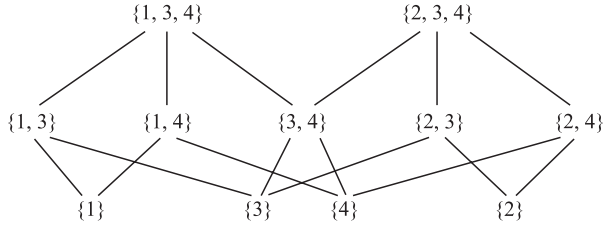


Fig. 14-21

14.34. See Fig. 14-22.

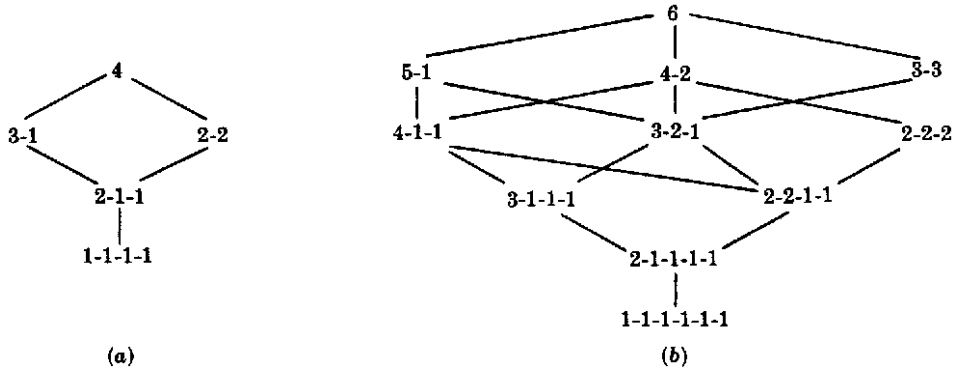


Fig. 14-22

14.35. See Fig. 14-23.

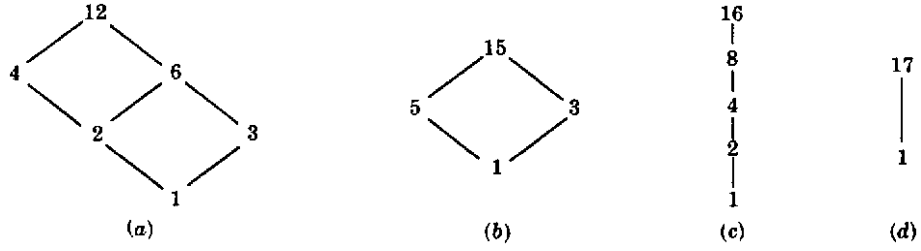


Fig. 14-23

14.36. *Hint*: Draw the diagram of S.

(a) Minimal, e; maximal, a, d.

(b) First, e; Last, none.

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

14.37. (a) False. Example: $\mathbf{N} \cup \{a\}$ where $1 \ll a$, and $\mathbf{N}$ ordered by $\leq$. (b) True. (c) True.

14.38. (a) Minimal, a; maximal, d and e. (b) First, a; last, none. (c) Any subset which contains c and omits a; that is: c, cb, cd, ce, cbd, cbe, cde, $cbde$. (d) c, cd, ce, cde. (e) abd, acd, ace.

14.39. (a) Minimal, a and b; maximal, e and f. (b) First, none; last, none. (c) ace, acf, bce, bcf, bdf.

14.40. (a) Four. (b) None.

14.41. (a) Six. (b) None.

14.42. $abcde$, $abced$, $acbde$, $acbed$, $acebd$.

14.43. Eleven.

14.44. $a \ll b,\ a \ll c,\ c \ll d$.

14.45. Minimal, $(p, 2)$ where p is a prime. Maximal, none.

14.46. (a) an, at, go, or, arm, one, gate, gone, about, occur. (b) an, about, arm, at, gate, go, gone, occur, one, or.

14.47. (a) $\|$; (b) $>$; (c) $\|$; (d) $<$.

14.48. $1c$, $1y$, $2a$, $2c$, $2z$, $3b$, $4b$, $4z$

14.49. (a) e, f, g; (b) a; (c) $\sup(A) = e$; (d) $\inf(A) = a$.

14.50. (a) e, f, g; (b) none; (c) $\sup(B) = e$; (d) none.

14.51. (a) 1, 2, 3; (b) 8; (c) $\sup(A) = 3$; (d) $\inf(A) = 8$.

14.52. (a) None; (b) 8; (c) none; (d) $\inf(B) = 8$.

14.53. (a) Both; (b) $\sup(A) = 3$; $\inf(A)$ does not exist.

14.54. (a) Both; (b) $\sup(A) = 3$; $\inf(A) = \sqrt[3]{5}$

14.55. Four: (1) a, b, c; (2) $a, b \ll c$; (3) $a \ll b, a \ll c$. (4) $a \ll b \ll c$.

14.56. Four: See Fig. 14-24.

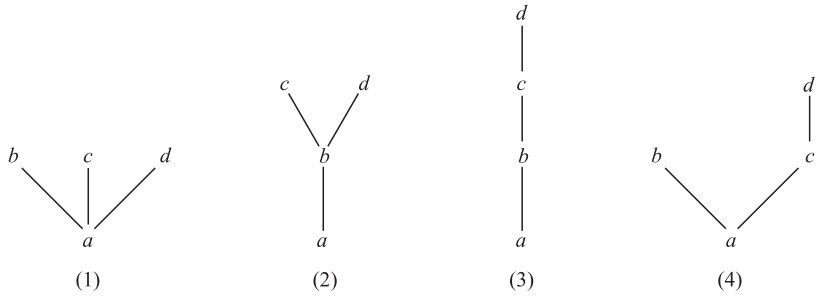


Fig. 14-24

14.57. (a) One: Identity mapping; (b) one; (c) two.

14.59. (b) b_1, c_1; (c) **N** has no limit points.

14.60. (a) Define $f: S \to \mathbf{N}$ by $f(a) = 1$, $f(b) = 2$, $f(3) = 3$, $f(n) = n + 3$.
(b) The element a is a limit point of S', but **N** has no limit points.

14.61. A is a finite linearly ordered set.

14.65. (a) Six: 0*abdI*, 0*acdI*, 0*adeI*, 0*bceI*, 0*aceI*, 0*cdeI*; (b) (i) a, b, c, 0; (ii) a, b, c. (c) c and e are complements of a. b has no complement. (d) No. No.

14.66. (a) a, b, c, g, 0. (b) a, b, c. (c) a has g; b has none. (d) $I = a \vee g$, $f = a \vee b$, $e = b \vee c$, $d = a \vee c$. Other elements are join-irreducible. (e) No. No.

14.67. (a) e has none; f has b and c. (b) $I = c \vee f = b \vee f = b \vee d \vee f$. (c) No, since decompositions are not unique. (d) Two: 0, d, e, f, I must be mapped into themselves. Then $F = 1_L$, identity map on L, or $F = \{(b, c), (c, b)\}$.

14.68. (a) a has c; c has a and b. (b) $I = a \vee c = b \vee c$. (c) No. (d) Two: 0, c, d, I must be mapped into themselves. Then $f = 1_L$ or $f = \{(a, b), (b, a)\}$.

14.69. (a) a has e, c has b and e. (b) $I = a \vee e = b \vee c = c \vee e$. (c) No. (d) Two: 0, d, I are mapped into themselves. Then $f = 1_L$ or $f = \{(a, b), (b, a), (c, d), (d, c)\}$.

14.70. (a) See Fig. 14-25. (b) 1, 2, 3, 4, 5. The atoms are 2, 3 and 5. (c) 2 has none, 10 has none. (d) $60 = 4 \vee 3 \vee 5$; $30 = 2 \vee 3 \vee 5$; $20 = 4 \vee 5$; $15 = 3 \vee 5$; $12 = 3 \vee 4$; $10 = 2 \vee 5$; $6 = 2 \vee 3$.

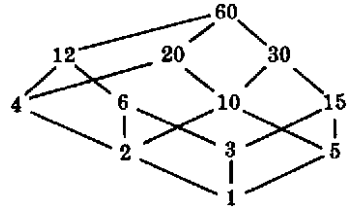


Fig. 14-25

14.73. (a) [1, 2, 3, 4], [14, 2, 3], [13, 24], [14, 23], [123, 4], [124, 3], [134, 2], [234, 1], [1234], [23, 1, 4] [24, 1, 3], [34, 1, 2]. (b) See Fig. 14-26.

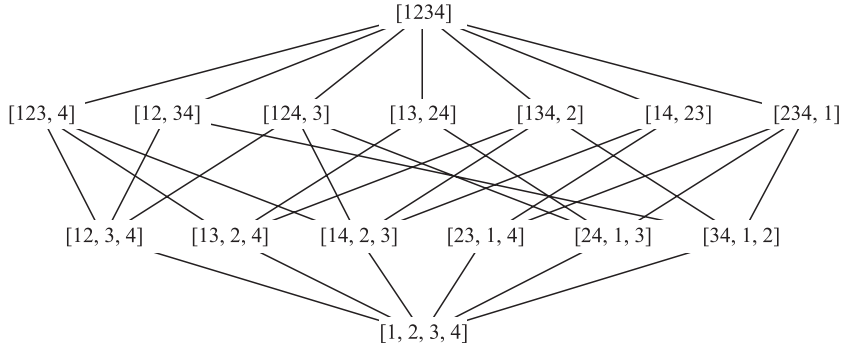


Fig. 14-26

14.74. Geometrically, an element $a \neq I$ is meet-irreducible if and only if a has only one immediate successor. (a) a, c, d, e, I; (b) a, b, d, f, g, I; (c) 4, 6, 12, 15, 60.

14.75. (a) If $a \leq c$ then $a \vee c = c$. Hence $a \vee (b \wedge c) = (a \vee b) \wedge (a \vee c) = (a \vee b) \wedge c$; (b) Here $a \leq c$. But $a \vee (b \wedge c) = a \vee 0 = a$ and $(a \vee b) \wedge c = I \wedge c = c$; hence $a \vee (b \wedge c) \neq (a \vee b) \wedge c$.

CHAPTER 5

Techniques of Counting

5.1 INTRODUCTION

This chapter develops some techniques for determining, without direct enumeration, the number of possible outcomes of a particular event or the number of elements in a set. Such sophisticated counting is sometimes called *combinatorial analysis*. It includes the study of permutations and combinations.

5.2 BASIC COUNTING PRINCIPLES

There are two basic counting principles used throughout this chapter. The first one involves addition and the second one multiplication.

Sum Rule Principle:

Suppose some event E can occur in m ways and a second event F can occur in n ways, and suppose both events cannot occur simultaneously. Then E or F can occur in $m + n$ ways.

Product Rule Principle:

Suppose there is an event E which can occur in m ways and, independent of this event, there is a second event F which can occur in n ways. Then combinations of E and F can occur in mn ways.

The above principles can be extended to three or more events. That is, suppose an event E_1 can occur in n_1 ways, a second event E_2 can occur in n_2 ways, and, following E_2; a third event E_3 can occur in n_3 ways, and so on. Then:

Sum Rule: If no two events can occur at the same time, then one of the events can occur in:

$$n_1 + n_2 + n_3 + \cdots \text{ ways.}$$

Product Rule: If the events occur one after the other, then all the events can occur in the order indicated in:

$$n_1 \cdot n_2 \cdot n_3 \cdot \ldots \text{ ways.}$$

EXAMPLE 5.1 Suppose a college has 3 different history courses, 4 different literature courses, and 2 different sociology courses.

(a) The number m of ways a student can choose one of each kind of courses is:

$$m = 3(4)(2) = 24$$

(b) The number n of ways a student can choose just one of the courses is:

$$n = 3 + 4 + 2 = 9$$

There is a set theoretical interpretation of the above two principles. Specifically, suppose $n(A)$ denotes the number of elements in a set A. Then:

(1) ***Sum Rule Principle:*** Suppose A and B are disjoint sets. Then

$$n(A \cup B) = n(A) + n(B)$$

(2) ***Product Rule Principle:*** Let $A \times B$ be the Cartesian product of sets A and B. Then

$$n(A \times B) = n(A) \cdot n(B)$$

5.3 MATHEMATICAL FUNCTIONS

We discuss two important mathematical functions frequently used in combinatorics.

Factorial Function

The product of the positive integers from 1 to n inclusive is denoted by $n!$, read "n factorial." Namely:

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

Accordingly, $1! = 1$ and $n! = n(n-l)!$. It is also convenient to define $0! = 1$.

EXAMPLE 5.2

(a) $3! = 3 \cdot 2 \cdot 1 = 6, \quad 4! = 4 \cdot 3 \cdot 2 \cdot 1 = 24, \quad 5 = 5 \cdot 4! = 5(24) = 120.$

(b) $\dfrac{12 \cdot 11 \cdot 10}{3 \cdot 2 \cdot 1} = \dfrac{12 \cdot 11 \cdot 10 \cdot 9!}{3 \cdot 2 \cdot 1 \cdot 9!} = \dfrac{12!}{3!\,9!}$ and, more generally,

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

(c) For large n, one uses Stirling's approximation (where $e = 2.7128...$):

$$n! = \sqrt{2\pi n}\, n^n e^{-n}$$

Binomial Coefficients

The symbol $\binom{n}{r}$, read "*nCr*" or "*n* Choose *r*," where *r* and *n* are positive integers with $r \leq n$, is defined as follows:

$$\binom{n}{r} = \frac{n(n-1)\cdots(n-r+1)}{r(r-1)\ldots 3\cdot 2\cdot 1} \qquad \text{or equivalently} \qquad \binom{n}{r} = \frac{n!}{r!(n-r)!}$$

Note that $n-(n-r)=r$. This yields the following important relation.

Lemma 5.1: $\binom{n}{n-r} = \binom{n}{r}$ or equivalently, $\binom{n}{a} = \binom{n}{b}$ where $a+b=n$.

Motivated by that fact that we defined $0! = 1$, we define:

$$\binom{n}{0} = \frac{n!}{0!n!} = 1 \quad \text{and} \quad \binom{0}{0} = \frac{0!}{0!\,0!} = 1$$

EXAMPLE 5.3

(a) $\binom{8}{2} = \frac{8\cdot 7}{2\cdot 1} = 28; \quad \binom{9}{4} = \frac{9\cdot 8\cdot 7\cdot 6}{4\cdot 3\cdot 2\cdot 1} = 126; \quad \binom{12}{5} = \frac{12\cdot 11\cdot 10\cdot 9\cdot 8}{5\cdot 4\cdot 3\cdot 2\cdot 1} = 792.$

Note that $\binom{n}{r}$ has exactly *r* factors in both the numerator and the denominator.

(b) Suppose we want to compute $\binom{10}{7}$. There will be 7 factors in both the numerator and the denominator. However, $10-7=3$. Thus, we use Lemma 5.1 to compute:

$$\binom{10}{7} = \binom{10}{3} = \frac{10\cdot 9\cdot 8}{3\cdot 2\cdot 1} = 120$$

Binomial Coefficients and Pascal's Triangle

The numbers $\binom{n}{r}$ are called *binomial coefficients*, since they appear as the coefficients in the expansion of $(a+b)^n$. Specifically:

Theorem (Binomial Theorem) 5.2: $(a+b)^n = \sum_{k=0}^{n} \binom{n}{r} a^{n-k}b^k$

The coefficients of the successive powers of $a+b$ can be arranged in a triangular array of numbers, called Pascal's triangle, as pictured in Fig. 5-1. The numbers in Pascal's triangle have the following interesting properties:

(i) The first and last number in each row is 1.

(ii) Every other number can be obtained by adding the two numbers appearing above it. For example:

$$10 = 4+6, \quad 15 = 5+10, \quad 20 = 10+10.$$

Since these numbers are binomial coefficients, we state the above property formally.

$$(a+b)^0 = 1$$
$$(a+b)^1 = a + b$$
$$(a+b)^2 = a^2 + 2ab + b^2$$
$$(a+b)^3 = a^3 + 3a^2b + 3ab^2 + b^3$$
$$(a+b)^4 = a^4 + 4a^3b + 6a^2b^2 + 4ab^3 + b^4$$
$$(a+b)^5 = a^5 + 5a^4b + 10a^3b^2 + 10a^2b^3 + 5ab^4 + b^5$$
$$(a+b)^6 = a^6 + 6a^5b + 15a^4b^2 + 20a^3b^3 + 15a^2b^4 + 6ab^5 + b^6$$
$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

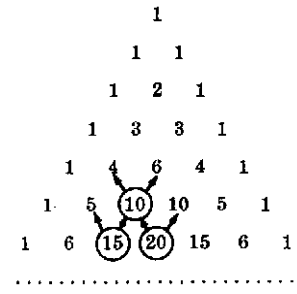


Fig. 5-1 Pascal's triangle

Theorem 5.3: $\binom{n+1}{r} = \binom{n}{r-1} + \binom{n}{r}$.

5.4 PERMUTATIONS

Any arrangement of a set of n objects in a given order is called a *permutation* of the object (taken all at a time). Any arrangement of any $r \leq n$ of these objects in a given order is called an "r-permutation" or "a permutation of the n objects taken r at a time." Consider, for example, the set of letters A, B, C, D. Then:

(i) $BDCA$, $DCBA$, and $ACDB$ are permutations of the four letters (taken all at a time).
(ii) BAD, ACB, DBC are permutations of the four letters taken three at a time.
(iii) AD, BC, CA are permutations of the four letters taken two at a time.

We usually are interested in the number of such permutations without listing them.
The number of permutations of n objects taken r at a time will be denoted by

$$P(n, r) \quad \text{(other texts may use } {}_nP_r,\ P_{n,r},\ \text{or } (n)_r).$$

The following theorem applies.

Theorem 5.4: $P(n, r) = n(n-1)(n-2)\cdots(n-r+1) = \dfrac{n!}{(n-r)!}$

We emphasize that there are r factors in $n(n-1)(n-2)\cdots(n-r+1)$.

EXAMPLE 5.4 Find the number m of permutations of six objects, say, A, B, C, D, E, F, taken three at a time. In other words, find the number of "three-letter words" using only the given six letters without repetition.

Let us represent the general three-letter word by the following three positions:

$$\underline{\quad}, \underline{\quad}, \underline{\quad}$$

The first letter can be chosen in 6 ways; following this the second letter can be chosen in 5 ways; and, finally, the third letter can be chosen in 4 ways. Write each number in its appropriate position as follows:

$$\underline{6}, \underline{5}, \underline{4}$$

By the Product Rule there are $m = 6 \cdot 5 \cdot 4 = 120$ possible three-letter words without repetition from the six letters. Namely, there are 120 permutations of 6 objects taken 3 at a time. This agrees with the formula in Theorem 5.4:

$$P(6, 3) = 6 \cdot 5 \cdot 4 = 120$$

In fact, Theorem 5.4 is proven in the same way as we did for this particular case.

Consider now the special case of $P(n,\ r)$ when $r = n$. We get the following result.

Corollary 5.5: There are $n!$ permutations of n objects (taken all at a time).

For example, there are $3! = 6$ permutations of the three letters A, B, C. These are:

$$ABC, \quad ACB, \quad BAC, \quad BCA, \quad CAB, \quad CBA.$$

Permutations with Repetitions

Frequently we want to know the number of permutations of a multiset, that is, a set of objects some of which are alike. We will let

$$P(n;\ n_1,\ n_2,\ \ldots,\ n_r)$$

denote the number of permutations of n objects of which n_1 are alike, n_2 are alike, ..., n_r are alike. The general formula follows:

Theorem 5.6: $P(n;\ n_1,\ n_2, \ldots,\ n_r) = \dfrac{n!}{n_1!\, n_2! \ldots n_r!}$

We indicate the proof of the above theorem by a particular example. Suppose we want to form all possible five-letter "words" using the letters from the word "*BABBY*." Now there are $5! = 120$ permutations of the objects B_1, A, B_2, B_3, Y, where the three B's are distinguished. Observe that the following six permutations

$$B_1B_2\,B_3AY, \quad B_2B_1\,B_3AY, \quad B_3B_1\,B_2AY, \quad B_1B_3\,B_2AY, \quad B_2B_3\,B_1AY, \quad B_3B_2\,B_1AY$$

produce the same word when the subscripts are removed. The 6 comes from the fact that there are $3! = 3{\cdot}2{\cdot}1 = 6$ different ways of placing the three B's in the first three positions in the permutation. This is true for each set of three positions in which the B's can appear. Accordingly, the number of different five-letter words that can be formed using the letters from the word "*BABBY*" is:

$$P(5;\ 3) = \frac{5!}{3!} = 20$$

EXAMPLE 5.5 Find the number m of seven-letter words that can be formed using the letters of the word "*BENZENE*."

We seek the number of permutations of 7 objects of which 3 are alike (the three E's), and 2 are alike (the two N's). By Theorem 5.6,

$$m = P(7;\ 3,\ 2) = \frac{7!}{3!2!} = \frac{7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}{3 \cdot 2 \cdot 1 \cdot 2 \cdot 1} = 420$$

Ordered Samples

Many problems are concerned with choosing an element from a set S, say, with n elements. When we choose one element after another, say, r times, we call the choice an *ordered sample* of size r. We consider two cases.

(1) Sampling with replacement

Here the element is replaced in the set S before the next element is chosen. Thus, each time there are n ways to choose an element (repetitions are allowed). The Product rule tells us that the number of such samples is:

$$n \cdot n \cdot n \cdots n \cdot n (r \text{ factors}) = n^r$$

(2) Sampling without replacement

Here the element is not replaced in the set S before the next element is chosen. Thus, there is no repetition in the ordered sample. Such a sample is simply an r-permutation. Thus the number of such samples is:

$$P(n, r) = n(n-1)(n-2)\cdots(n-r+1) = \frac{n!}{(n-r)!}$$

EXAMPLE 5.6 Three cards are chosen one after the other from a 52-card deck. Find the number m of ways this can be done: (a) with replacement; (b) without replacement.

(a) Each card can be chosen in 52 ways. Thus $m = 52(52)(52) = 140\,608$.

(b) Here there is no replacement. Thus the first card can be chosen in 52 ways, the second in 51 ways, and the third in 50 ways. Therefore:

$$m = P(52, 3) = 52(51)(50) = 132\,600$$

5.5 COMBINATIONS

Let S be a set with n elements. A *combination* of these n elements taken r at a time is any selection of r of the elements where order does not count. Such a selection is called an *r-combination*; it is simply a subset of S with r elements. The number of such combinations will be denoted by

$$C(n, r) \qquad \text{(other texts may use } {}_nC_r,\ C_{n,r},\ \text{or } C_r^n).$$

Before we give the general formula for $C(n,\ r)$, we consider a special case.

EXAMPLE 5.7 Find the number of combinations of 4 objects, A, B, C, D, taken 3 at a time.

Each combination of three objects determines $3! = 6$ permutations of the objects as follows:

$$\begin{array}{lllllll}
ABC: & ABC, & ACB, & BAC, & BCA, & CAB, & CBA \\
ABD: & ABD, & ADB, & BAD, & BDA, & DAB, & DBA \\
ACD: & ACD, & ADC, & CAD, & CDA, & DAC, & DCA \\
BCD: & BDC, & BDC, & CBD, & CDB, & DBC, & DCB
\end{array}$$

Thus the number of combinations multiplied by 3! gives us the number of permutations; that is,

$$C(4, 3) \cdot 3! = P(4, 3) \quad \text{or} \quad C(4, 3) = \frac{P(4, 3)}{3!}$$

But $P(4,\ 3) = 4 \cdot 3 \cdot 2 = 24$ and $3! = 6$; hence $C(4, 3) = 4$ as noted above.

As indicated above, any combination of n objects taken r at a time determines $r!$ permutations of the objects in the combination; that is,

$$P(n, r) = r!\, C(n, r)$$

Accordingly, we obtain the following formula for $C(n,\ r)$ which we formally state as a theorem.

Theorem 5.7: $C(n, r) = \dfrac{P(n,\ r)}{r!} = \dfrac{n!}{r!(n-r)!}$

Recall that the binomial coefficient $\binom{n}{r}$ was defined to be $\dfrac{n!}{r!(n-r)!}$; hence

$$\boxed{C(r, n) = \binom{n}{r}}$$

We shall use $C(n,\ r)$ and $\binom{n}{r}$ interchangeably.

EXAMPLE 5.8 A farmer buys 3 cows, 2 pigs, and 4 hens from a man who has 6 cows, 5 pigs, and 8 hens. Find the number m of choices that the farmer has.

The farmer can choose the cows in $C(6,\ 3)$ ways, the pigs in $C(5, 2)$ ways, and the hens in $C(8,\ 4)$ ways. Thus the number m of choices follows:

$$m = \binom{6}{3}\binom{5}{2}\binom{8}{4} = \frac{6\cdot 5\cdot 4}{3\cdot 2\cdot 1}\cdot\frac{5\cdot 4}{2\cdot 1}\cdot\frac{8\cdot 7\cdot 6\cdot 5}{4\cdot 3\cdot 2\cdot 1} = 20\cdot 10\cdot 70 = 14\ 000$$

5.6 THE PIGEONHOLE PRINCIPLE

Many results in combinational theory come from the following almost obvious statement.

Pigeonhole Principle: If n pigeonholes are occupied by $n + 1$ or more pigeons, then at least one pigeonhole is occupied by more than one pigeon.

This principle can be applied to many problems where we want to show that a given situation can occur.

EXAMPLE 5.9

(a) Suppose a department contains 13 professors, then two of the professors (pigeons) were born in the same month (pigeonholes).

(b) Find the minimum number of elements that one needs to take from the set $S = \{1, 2, 3, \ldots, 9\}$ to be sure that two of the numbers add up to 10.
Here the pigeonholes are the five sets $\{1, 9\}$, $\{2, 8\}$, $\{3, 7\}$, $\{4, 6\}$, $\{5\}$. Thus any choice of six elements (pigeons) of S will guarantee that two of the numbers add up to ten.

The Pigeonhole Principle is generalized as follows.

Generalized Pigeonhole Principle: If n pigeonholes are occupied by $kn + 1$ or more pigeons, where k is a positive integer, then at least one pigeonhole is occupied by $k + 1$ or more pigeons.

EXAMPLE 5.10 Find the minimum number of students in a class to be sure that three of them are born in the same month.

Here the $n = 12$ months are the pigeonholes, and $k + 1 = 3$ so $k = 2$. Hence among any $kn + 1 = 25$ students (pigeons), three of them are born in the same month.

5.7 THE INCLUSION–EXCLUSION PRINCIPLE

Let A and B be any finite sets. Recall Theorem 1.9 which tells us:

$$n(A \cup B) = n(A) + n(B) - n(A \cap B)$$

In other words, to find the number $n(A \cup B)$ of elements in the union of A and B, we add $n(A)$ and $n(B)$ and then we subtract $n(A \cap B)$; that is, we "include" $n(A)$ and $n(B)$, and we "exclude" $n(A \cap B)$. This follows from the fact that, when we add $n(A)$ and $n(B)$, we have counted the elements of $(A \cap B)$ twice.

The above principle holds for any number of sets. We first state it for three sets.

Theorem 5.8: For any finite sets A, B, C we have

$$n(A \cup B \cup C) = n(A) + n(B) + n(C) - n(A \cap B) - n(A \cap C) - n(B \cap C) + n(A \cap B \cap C)$$

That is, we "include" $n(A)$, $n(B)$, $n(C)$, we "exclude" $n(A \cap B)$, $n(A \cap C)$, $n(B \cap C)$, and finally "include" $n(A \cap B \cap C)$.

EXAMPLE 5.11 Find the number of mathematics students at a college taking at least one of the languages French, German, and Russian, given the following data:

65 study French, 20 study French and German,
45 study German, 25 study French and Russian, 8 study all three languages.
42 study Russian, 15 study German and Russian,

We want to find $n(F \cup G \cup R)$ where F, G, and R denote the sets of students studying French, German, and Russian, respectively.

By the Inclusion–Exclusion Principle,

$$\begin{aligned} n(F \cup G \cup R) &= n(F) + n(G) + n(R) - n(F \cap G) - n(F \cap R) - n(G \cap R) + n(F \cap G \cap R) \\ &= 65 + 45 + 42 - 20 - 25 - 15 + 8 = 100 \end{aligned}$$

Namely, 100 students study at least one of the three languages.

Now, suppose we have any finite number of finite sets, say, $A_1, A_2, \ldots, A_m$. Let s_k be the sum of the cardinalities

$$n(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_K})$$

of all possible k-tuple intersections of the given m sets. Then we have the following general Inclusion–Exclusion Principle.

Theorem 5.9: $n(A_1 \cup A_2 \cup \cdots \cup A_m) = s_1 - s_2 + s_3 - \cdots + (-1)^{m-1} s_m$.

5.8 TREE DIAGRAMS

A *tree diagram* is a device used to enumerate all the possible outcomes of a sequence of events where each event can occur in a finite number of ways. The construction of tree diagrams is illustrated in the following example.

EXAMPLE 5.12

(a) Find the product set $A \times B \times C$, where $A = \{1, 2\}$, $B = \{a, b, c\}$, $C = \{x, y\}$.

The tree diagram for $A \times B \times C$ appears in Fig. 5-2(a). Here the tree is constructed from left to right, and the number of branches at each point corresponds to the possible outcomes of the next event. Each endpoint (leaf) of the tree is labeled by the corresponding element of $A \times B \times C$. As noted previously, $A \times B \times C$ has $n = 2(3)(2) = 12$ elements.

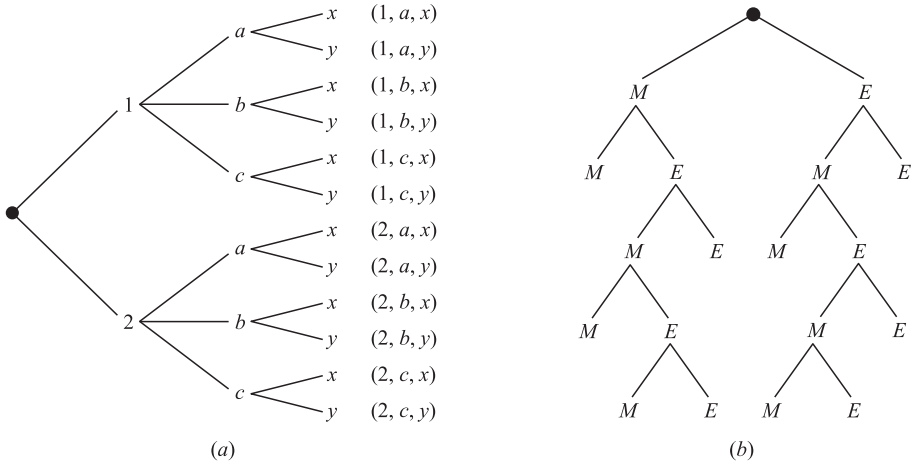


Fig. 5-2

(b) Mark and Erik are to play a tennis tournament. The first person to win two games in a row or who wins a total of three games wins the tournament. Find the number of ways the tournament can occur.

The tree diagram showing the possible outcomes of the tournament appears in Fig. 5-2(*b*). Here the tree is constructed from top-down rather than from left-right. (That is, the "root" is on the top of the tree.) Note that there are 10 endpoints, and the endpoints correspond to the following 10 ways the tournament can occur:

$$MM,\ MEMM,\ MEMEM,\ MEMEE,\ MEE,\ EMM,\ EMEMM,\quad EMEME,\ EMEE,\ EE$$

The path from the beginning (top) of the tree to the endpoint describes who won which game in the tournament.

Solved Problems

FACTORIAL NOTATION AND BINOMIAL COEFFICIENTS

5.1. Compute: (*a*) 4!, 5!; (*b*) 6!, 7!, 8!, 9!; (*c*) 50!.

(*a*) $4! = 4 \cdot 3 \cdot 2 \cdot 1 = 24$, $5! = 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 5(24) = 120$.
(*b*) Now use $(n+1)! = (n+1)n!$:

$$6! = 5(5!) = 6(120) = 720, \qquad 8! = 8(7!) = 8(5040) = 40\,320,$$
$$7! = 7(6!) = 7(720) = 5\,040, \quad 9! = 9(8!) = 9(40\,320) = 362\,880.$$

(*c*) Since n is very large, we use Sterling's approximation: $n! = \sqrt{2\pi n}\, n^{\pi} e^{-n}$ (where $e \approx 2.718$). Thus:

$$50! \approx N = \sqrt{100\pi}\, 50^{50} e^{-50}$$

Evaluating N using a calculator, we get $N = 3.04 \times 10^{64}$ (which has 65 digits).

5.2. Compute: (*a*) $\dfrac{13!}{11!}$; (*b*) $\dfrac{7!}{10!}$.

(*a*) $$\frac{13!}{11!} = \frac{13 \cdot 12 \cdot 11 \cdot 10 \cdot 9 \cdot 8 \cdot 7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}{11 \cdot 10 \cdot 9 \cdot 8 \cdot 7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1} = 13 \cdot 12 = 156.$$

Alternatively, this could be solved as follows:

$$\frac{13!}{11!} = \frac{13 \cdot 12 \cdot 11!}{11!} = 13 \cdot 12 = 156.$$

(*b*) $\dfrac{7!}{10!} = \dfrac{7!}{10 \cdot 9 \cdot 8 \cdot 7!} = \dfrac{1}{10 \cdot 9 \cdot 8} = \dfrac{1}{720}$.

5.3. Simplify: (*a*) $\dfrac{n!}{(n-1)!}$; (*b*) $\dfrac{(n+2)!}{n!}$.

(*a*) $\dfrac{n!}{(n-1)!} = \dfrac{n(n-1)(n-2)\cdots 3 \cdot 2 \cdot 1}{(n-1)(n-2)\cdots 3 \cdot 2 \cdot 1} = n$; alternatively, $\dfrac{n!}{(n-1)!} = \dfrac{n(n-1)!}{(n-1)!} = n$.

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

5.4. Compute: (*a*) $\binom{16}{3}$; (*b*) $\binom{12}{4}$; (*c*) $\binom{8}{5}$.

Recall that there are as many factors in the numerator as in the denominator.

(*a*) $\binom{16}{3} = \dfrac{16 \cdot 15 \cdot 14}{3 \cdot 2 \cdot 1} = 560$; (*b*) $\binom{12}{4} = \dfrac{12 \cdot 11 \cdot 10 \cdot 9}{4 \cdot 3 \cdot 2 \cdot 1} = 495$;

(*c*) Since $8 - 5 = 3$, we have $\binom{8}{5} = \binom{8}{3} = \dfrac{8 \cdot 7.6}{3 \cdot 2 \cdot 1} = 56$.

5.5. Prove: $\binom{17}{6} = \binom{16}{5} + \binom{16}{6}$.

Now $\binom{16}{5} + \binom{16}{6} = \dfrac{16!}{5!11!} + \dfrac{16!}{6!10!}$. Multiply the first fraction by $\frac{6}{6}$ and the second by $\frac{11}{11}$ to obtain the same denominator in both fractions; and then add:

$$\begin{aligned}\binom{16}{5} + \binom{16}{6} &= \frac{6 \cdot 16!}{6 \cdot 5! \cdot 11!} + \frac{11 \cdot 16!}{6! \cdot 11 \cdot 10!} = \frac{6 \cdot 16!}{6! \cdot 11!} + \frac{11 \cdot 16!}{6! \cdot 11!} \\ &= \frac{6 \cdot 16! + 11 \cdot 16!}{6! \cdot 11!} = \frac{(6+11) \cdot 16!}{6! \cdot 11!} = \frac{17 \cdot 16!}{6! \cdot 11!} = \frac{17!}{6! \cdot 11!} = \binom{17}{6}\end{aligned}$$

5.6. Prove Theorem 5.3: $\binom{n+1}{r} = \binom{n}{r-1} + \binom{n}{r}$.

(The technique in this proof is similar to that of the preceding problem.)

Now $\binom{n}{r-1} + \binom{n}{r} = \dfrac{n!}{(r-1)! \cdot (n-r+1)!} + \dfrac{n!}{r! \cdot (n-r)!}$.

To obtain the same denominator in both fractions, multiply the first fraction by $\frac{r}{r}$ and the second fraction by $\dfrac{n-r+1}{n-r+1}$. Hence

$$\begin{aligned}\binom{n}{r-1} + \binom{n}{r} &= \frac{r \cdot n!}{r \cdot (r-1)! \cdot (n-r+1)!} + \frac{(n-r+1) \cdot n!}{r! \cdot (n-r+1) \cdot (n-r)!} \\ &= \frac{r \cdot n!}{r!(n-r+1)!} + \frac{(n-r+1) \cdot n!}{r!(n-r+1)!} \\ &= \frac{r \cdot n! + (n-r+1) \cdot n!}{r!(n-r+1)!} = \frac{[r + (n-r+1)] \cdot n!}{r!(n-r+1)!} \\ &= \frac{(n+1)n!}{r!(n-r+1)!} = \frac{(n+1)!}{r!(n-r+1)!} = \binom{n+1}{r}\end{aligned}$$

COUNTING PRINCIPLES

5.7. Suppose a bookcase shelf has 5 History texts, 3 Sociology texts, 6 Anthropology texts, and 4 Psychology texts. Find the number n of ways a student can choose:

(*a*) one of the texts; (*b*) one of each type of text.

(*a*) Here the Sum Rule applies; hence, $n = 5 + 3 + 6 + 4 = 18$.

(*b*) Here the Product Rule applies; hence, $n = 5 \cdot 3 \cdot 6 \cdot 4 = 360$.

5.8. A history class contains 8 male students and 6 female students. Find the number n of ways that the class can elect: (a) 1 class representative; (b) 2 class representatives, 1 male and 1 female; (c) 1 president and 1 vice president.

(a) Here the Sum Rule is used; hence, $n = 8 + 6 = 14$.

(b) Here the Product Rule is used; hence, $n = 8 \cdot 6 = 48$.

(c) There are 14 ways to elect the president, and then 13 ways to elect the vice president. Thus $n = 14 \cdot 13 = 182$.

5.9. There are four bus lines between A and B, and three bus lines between B and C. Find the number m of ways that a man can travel by bus: (a) from A to C by way of B; (b) roundtrip from A to C by way of B; (c) roundtrip from A to C by way of B but without using a bus line more than once.

(a) There are 4 ways to go from A to B and 3 ways from B to C; hence $n = 4 \cdot 3 = 12$.

(b) There are 12 ways to go from A to C by way of B, and 12 ways to return. Thus $n = 12 \cdot 12 = 144$.

(c) The man will travel from A to B to C to B to A. Enter these letters with connecting arrows as follows:

$$A \to B \to C \to B \to A$$

The man can travel four ways from A to B and three ways from B to C, but he can only travel two ways from C to B and three ways from B to A since he does not want to use a bus line more than once. Enter these numbers above the corresponding arrows as follows:

$$A \overset{4}{\to} B \overset{3}{\to} C \overset{2}{\to} B \overset{3}{\to} A$$

Thus, by the Product Rule, $n = 4 \cdot 3 \cdot 2 \cdot 3 = 72$.

PERMUTATIONS

5.10. State the essential difference between permutations and combinations, with examples.

Order counts with permutations, such as words, sitting in a row, and electing a president, vice president, and treasurer. Order does not count with combinations, such as committees and teams (without counting positions). The product rule is usually used with permutations, since the choice for each of the ordered positions may be viewed as a sequence of events.

5.11. Find: (a) $P(7, 3)$; (b) $P(14, 2)$.

Recall $P(n, r)$ has r factors beginning with n.

(a) $P(7, 3) = 7 \cdot 6 \cdot 5 = 210$; (b) $P(14, 2) = 14 \cdot 13 = 182$.

5.12. Find the number m of ways that 7 people can arrange themselves:

(a) In a row of chairs; (b) Around a circular table.

(a) Here $m = P(7, 7) = 7!$ ways.

(b) One person can sit at any place at the table. The other 6 people can arrange themselves in 6! ways around the table; that is $m = 6!$.

This is an example of a *circular permutation*. In general, n objects can be arranged in a circle in $(n - 1)!$ ways.

5.13. Find the number n of distinct permutations that can be formed from all the letters of each word:

(a) *THOSE*; (b) *UNUSUAL*; (c) *SOCIOLOGICAL*.

This problem concerns permutations with repetitions.

(a) $n = 5! = 120$, since there are 5 letters and no repetitions.

(b) $n = \dfrac{7!}{3!} = 840$, since there are 7 letters of which 3 are U and no other letter is repeated.

(*c*) $n = \dfrac{12!}{3!2!2!2!}$, since there are 12 letters of which 3 are *O*, 2 are *C*, 2 are *I*, and 2 are *L*. (We leave the answer using factorials, since the number is very large.)

5.14. A class contains 8 students. Find the number n of samples of size 3:
(*a*) With replacement; (*b*) Without replacement.

(*a*) Each student in the ordered sample can be chosen in 8 ways; hence, there are

$$n = 8 \cdot 8 \cdot 8 = 8^3 = 512 \text{ samples of size 3 with replacement.}$$

(*b*) The first student in the sample can be chosen in 8 ways, the second in 7 ways, and the last in 6 ways. Thus, there are $n = 8 \cdot 7 \cdot 6 = 336$ samples of size 3 without replacement.

5.15. Find n if $P(n, 2) = 72$.

$$P(n,\ 2) = n(n-1) = n^2 - n. \quad \text{Thus, we get}$$

$$n^2 - n = 72 \qquad \text{or} \qquad n^2 - n - 72 = 0 \quad \text{or} \quad (n-9)(n+8) = 0$$

Since n must be positive, the only answer is $n = 9$.

COMBINATIONS

5.16. A class contains 10 students with 6 men and 4 women. Find the number n of ways to:

(*a*) Select a 4-member committee from the students.
(*b*) Select a 4-member committee with 2 men and 2 women.
(*c*) Elect a president, vice president, and treasurer.

(*a*) This concerns combinations, not permutations, since order does not count in a committee. There are "10 choose 4" such committees. That is:

$$n = C(10,\ 4) = \binom{10}{4} = \frac{10 \cdot 9 \cdot 8 \cdot 7}{4 \cdot 3 \cdot 2 \cdot 1} = 210$$

(*b*) The 2 men can be chosen from the 6 men in $C(6, 2)$ ways, and the 2 women can be chosen from the 4 women in $C(4, 2)$ ways. Thus, by the Product Rule:

$$n = \binom{6}{2}\binom{4}{2} = \frac{6 \cdot 5}{2 \cdot 1} \cdot \frac{4 \cdot 3}{2 \cdot 1} = 15(6) = 90$$

(*c*) This concerns permutations, not combinations, since order does count. Thus,

$$n = P(6, 3) = 6 \cdot 5 \cdot 4 = 120$$

5.17. A box contains 8 blue socks and 6 red socks. Find the number of ways two socks can be drawn from the box if:
(*a*) They can be any color. (*b*) They must be the same color.

(*a*) There are "14 choose 2" ways to select 2 of the 14 socks. Thus:

$$n = C(14,\ 2) = \binom{14}{2} = \frac{14 \cdot 13}{2 \cdot 1} = 91$$

(*b*) There are $C(8, 2) = 28$ ways to choose 2 of the 8 blue socks, and $C(6, 2) = 15$ ways to choose 2 of the 4 red socks. By the Sum Rule, $n = 28 + 15 = 43$.

5.18. Find the number m of committees of 5 with a given chairperson that can be selected from 12 people.

The chairperson can be chosen in 12 ways and, following this, the other 4 on the committee can be chosen from the 11 remaining in $C(11, 4)$ ways. Thus $m = 12 \cdot C(11, 4) = 12 \cdot 330 = 3960$.

PIGEONHOLE PRINCIPLE

5.19. Find the minimum number n of integers to be selected from $S = \{1, 2, \ldots, 9\}$ so that: (a) The sum of two of the n integers is even. (b) The difference of two of the n integers is 5.

(a) The sum of two even integers or of two odd integers is even. Consider the subsets $\{1, 3, 5, 7, 9\}$ and $\{2, 4, 6, 8\}$ of S as pigeonholes. Hence $n = 3$.

(b) Consider the five subsets $\{1, 6\}, \{2, 7\}, \{3, 8\}, \{4, 9\}, \{5\}$ of S as pigeonholes. Then $n = 6$ will guarantee that two integers will belong to one of the subsets and their difference will be 5.

5.20. Find the minimum number of students needed to guarantee that five of them belong to the same class (Freshman, Sophomore, Junior, Senior).

Here the $n = 4$ classes are the pigeonholes and $k + 1 = 5$ so $k = 4$. Thus among any $kn + 1 = 17$ students (pigeons), five of them belong to the same class.

5.21. Let L be a list (not necessarily in alphabetical order) of the 26 letters in the English alphabet (which consists of 5 vowels, A, E, I, O, U, and 21 consonants).

(a) Show that L has a sublist consisting of four or more consecutive consonants.

(b) Assuming L begins with a vowel, say A, show that L has a sublist consisting of five or more consecutive consonants.

(a) The five letters partition L into $n = 6$ sublists (pigeonholes) of consecutive consonants. Here $k + 1 = 4$ and so $k = 3$. Hence $nk + 1 = 6(3) + 1 = 19 < 21$. Hence some sublist has at least four consecutive consonants.

(b) Since L begins with a vowel, the remainder of the vowels partition L into $n = 5$ sublists. Here $k + 1 = 5$ and so $k = 4$. Hence $kn + 1 = 21$. Thus some sublist has at least five consecutive consonants.

INCLUSION–EXCLUSION PRINCIPLE

5.22. There are 22 female students and 18 male students in a classroom. Find the total number t of students.

The sets of male and female students are disjoint; hence $t = 22 + 18 = 40$.

5.23. Suppose among 32 people who save paper or bottles (or both) for recycling, there are 30 who save paper and 14 who save bottles. Find the number m of people who:

(a) save both; (b) save only paper; (c) save only bottles.

Let P and B denote the sets of people saving paper and bottles, respectively. Then:

$$(a) \quad m = n(P \cap B) = n(P) + n(B) - n(P \cup B) = 30 + 14 - 32 = 12$$
$$(b) \quad m = n(P \backslash B) = n(P) - n(P \cap B) = 30 - 12 = 18$$
$$(c) \quad m = n(B \backslash P) = n(B) - n(P \cap B) = 14 - 12 = 2$$

5.24. Let A, B, C, D denote, respectively, art, biology, chemistry, and drama courses. Find the number N of students in a dormitory given the data:

12 take A,	5 take A and B,	4 take B and D,	2 take B, C, D,
20 take B,	7 take A and C,	3 take C and D,	3 take A, C, D,
20 take C,	4 take A and D,	3 take A, B, C,	2 take all four,
8 take D,	16 take B and C,	2 take A, B, D,	71 take none.

Let T be the number of students who take at least one course. By the Inclusion–Exclusion Principle Theorem 5.9, $T = s_1 - s_2 + s_3 - s_4$ where:

$$s_1 = 12 + 20 + 20 + 8 = 60, \quad s_2 = 5 + 7 + 4 + 16 + 4 + 3 = 39,$$
$$s_3 = 3 + 2 + 2 + 3 = 10, \quad s_4 = 2.$$

Thus $T = 29$, and $N = 71 + T = 100$.

TREE DIAGRAMS

5.25. Teams A and B play in a tournament. The first team to win three games wins the tournament. Find the number n of possible ways the tournament can occur.

Construct the appropriate tree diagram in Fig. 5-3(a). The tournament can occur in 20 ways:

$$AAA,\ AABA,\ AABBA,\ AABBB,\ ABAA,\ ABABA,\ ABABB,\ ABBAA,\ ABBAB,\ ABBB,$$
$$BBB,\ BBAB,\ BBAAB,\ BBAAA,\ BABB,\ BABAB,\ BABAA,\ BAABB,\ BAABA,\ BAAA$$

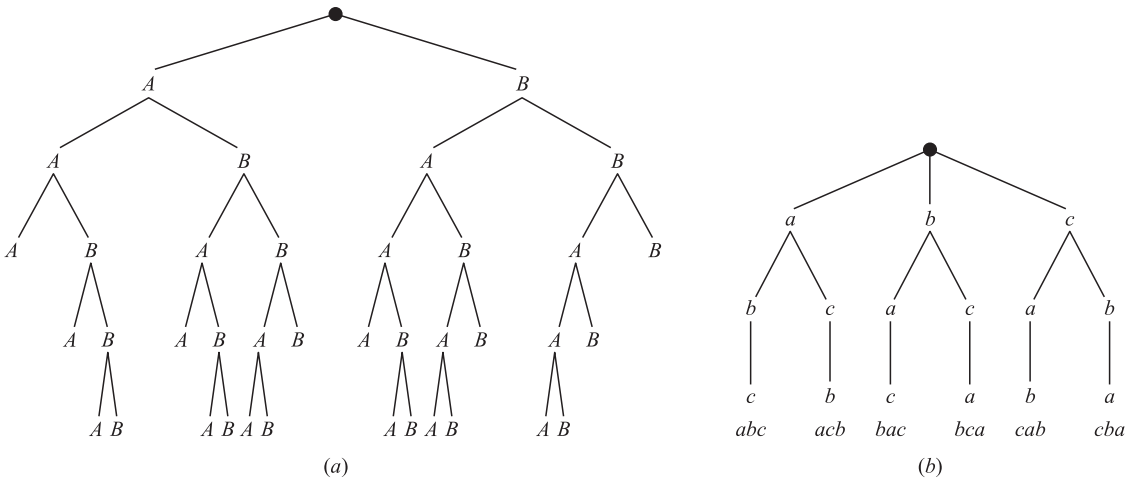


Fig. 5-3

5.26. Construct the tree diagram that gives the permutations of $\{a, b, c\}$.

The tree diagram appears in Fig. 5-3(b). There are six permutations, and they are listed on the bottom of the diagram.

MISCELLANEOUS PROBLEMS

5.27. There are 12 students in a class. Find the number n of ways that the 12 students can take 3 tests if 4 students are to take each test.

There are $C(12, 4) = 495$ ways to choose 4 of the 12 students to take the first test. Following this, there are $C(8, 4) = 70$ ways to choose 4 of the remaining 8 students to take the second test. The remaining students take the third test. Thus:

$$n = 70(495) = 34\,650$$

5.28. Prove Theorem (Binomial Theorem) 5.2: $(a+b)^n = \sum_{r=0}^{n} \binom{n}{r} a^{n-r} b^r$.

The theorem is true for $n = 1$, since

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

We assume the theorem is true for $(a+b)^n$ and prove it is true for $(a+b)^n + 1$.

$$(a+b)^{n+1} = (a+b)\,(a+b)^n$$

$$= (a+b)[a^n + \tbinom{n}{1} a^{n-1} b + \cdots + \tbinom{n}{r-1} a^{n-r+1} b^{r-1} + \tbinom{n}{r} a^{n-r} b^r + \cdots \tbinom{n}{1} a\, b^{n-1} + b^n]$$

Now the term in the product which contains b^r is obtained from

$$b[\tbinom{n}{r-1} a^{n-r+1} b^{r-1}] + a[\tbinom{n}{r} a^{n-r} b^r] = \tbinom{n}{r-1} a^{n-r+1} b^r + \tbinom{n}{r} a^{n-r+1} b^r$$
$$= [\tbinom{n}{r-1} + \tbinom{n}{r}] a^{n-r+1} b^r$$

But, by Theorem 5.3, $\binom{n}{r-1} = \binom{n}{r} = \binom{n+1}{r}$. Thus, the term containing b^r is:

$$\binom{n+1}{r} a^{n-r+1} b^r$$

Note that $(a+b)(a+b)^n$ is a polynomial of degree $n+1$ in b. Consequently:

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

which was to be proved.

5.29. Let n and $n_1, n_2, \ldots, n_r$ be nonnegative integers such that $n_1 + n_2 + \cdots + n_r = n$. The *multinomial coefficients* are denoted and defined by:

$$\binom{n}{n_1,\ n_2, \ldots, n_r} = \frac{n!}{n_1! n_2! \ldots n_r!}$$

Compute the following multinomial coefficients:

$$(a)\ \binom{6}{3,\ 2,\ 1}; \quad (b)\ \binom{8}{4,\ 2,\ 2,\ 0}; \quad (c)\ \binom{10}{5,\ 3,\ 2,\ 2}.$$

$$(a)\ \binom{6}{3,\ 2,\ 1} = \frac{6!}{3!\,2!\,1!} = \frac{6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}{3 \cdot 2 \cdot 1 \cdot 2 \cdot 1 \cdot 1} = 60$$

$$(b)\ \binom{8}{4,\ 2,\ 2,\ 0} = \frac{8!}{4!\,2!\,2!\,0!} = \frac{8 \cdot 7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}{4 \cdot 3 \cdot 2 \cdot 1 \cdot 2 \cdot 1 \cdot 2 \cdot 1 \cdot 1} = 420$$

(c) $\binom{10}{5,\ 3,\ 2,\ 2}$ has no meaning, since $5 + 3 + 2 + 2 \neq 10$.

5.30. A student must take five classes from three areas of study. Numerous classes are offered in each discipline, but the student cannot take more than two classes in any given area.

(a) Using the pigeonhole principle, show that the student will take at least two classes in one area.

(b) Using the Inclusion–Exclusion Principle, show that the student will have to take at least one class in each area.

(a) The three areas are the pigeonholes and the student must take five classes (pigeons). Hence, the student must take at least two classes in one area.

(b) Let each of the three areas of study represent three disjoint sets, A, B, and C. Since the sets are disjoint, $m(A \cup B \cup C) = 5 = n(A) + n(B) + n(C)$. Since the student can take at most two classes in any area of study, the sum of classes in any two sets, say A and B, must be less than or equal to four. Hence, $5 - [n(A) + n(B)] = n(C) \geq 1$. Thus, the student must take at least one class in any area.

Supplementary Problems

FACTORIAL NOTATION, BINOMIAL COEFFICIENTS

5.31. Find: (a) 10!, 11!, 12!; (b) 60!. (Hint: Use Sterling's approximation to $n!$.)

5.32. Evaluate: (a) 16!/14!; (b) 14!/11!; (c) 8!/10!; (d) 10!/13!.

5.33. Simplify: (a) $\dfrac{(n+1)!}{n!}$; (b) $\dfrac{n!}{(n-2)!}$; (c) $\dfrac{(n-1)!}{(n+2)!}$; (d) $\dfrac{(n-r+1)!}{(n-r-1)!}$.

5.34. Find: (a) $\binom{5}{2}$; (b) $\binom{7}{3}$; (c) $\binom{14}{2}$; (d) $\binom{6}{4}$; (e) $\binom{20}{17}$; (f) $\binom{18}{15}$.

5.35. Show that: (a) $\binom{n}{0} + \binom{n}{n} + \binom{n}{2} + \binom{n}{3} + \cdots + \binom{n}{n} = 2^n$

(b) $\binom{n}{0} - \binom{n}{1} + \binom{n}{2} - \binom{n}{3} + \cdots + \binom{n}{n} = 0$

5.36. Given the following eighth row of Pascal's triangle, find: (a) the ninth row; (b) the tenth row.

$$1 \quad 8 \quad 28 \quad 56 \quad 70 \quad 56 \quad 28 \quad 8 \quad 1$$

5.37. Evaluate the following multinomial coefficients (defined in Problem 5.29):

(a) $\begin{pmatrix} 6 \\ 2,\ 3,\ 1 \end{pmatrix}$; (b) $\begin{pmatrix} 7 \\ 3,\ 2,\ 2,\ 0 \end{pmatrix}$; (c) $\begin{pmatrix} 9 \\ 3,\ 5,\ 1 \end{pmatrix}$; (d) $\begin{pmatrix} 8 \\ 4,\ 3,\ 2 \end{pmatrix}$.

COUNTING PRINCIPLES

5.38. A store sells clothes for men. It has 3 kinds of jackets, 7 kinds of shirts, and 5 kinds of pants. Find the number of ways a person can buy: (a) one of the items; (b) one of each of the three kinds of clothes.

5.39. A class has 10 male students and 8 female students. Find the number of ways the class can elect: (a) a class representative; (b) 2 class representatives, one male and one female; (c) a class president and vicepresident.

5.40. Suppose a code consists of five characters, two letters followed by three digits. Find the number of: (a) codes; (b) codes with distinct letter; (c) codes with the same letters.

PERMUTATIONS

5.41. Find the number of automobile license plates where: (a) Each plate contains 2 different letters followed by 3 different digits. (b) The first digit cannot be 0.

5.42. Find the number m of ways a judge can award first, second, and third places in a contest with 18 contestants.

5.43. Find the number of ways 5 large books, 4 medium-size books, and 3 small books can be placed on a shelf where: (a) there are no restrictions; (b) all books of the same size are together.

5.44. A debating team consists of 3 boys and 3 girls. Find the number of ways they can sit in a row where:
(a) there are no restrictions; (b) the boys and girls are each to sit together; (c) just the girls are to sit together.

5.45. Find the number of ways 5 people can sit in a row where: (a) there are no restrictions; (b) two of the people insist on sitting next to each other.

5.46. Repeat Problem 5.45 if they sit around a circular table.

5.47. Consider all positive integers with three different digits. (Note that zero cannot be the first digit.) Find the number of them which are: (a) greater than 700; (b) odd; (c) divisible by 5.

5.48. Suppose repetitions are not permitted. (a) Find the number of three-digit numbers that can be formed from the six digits 2, 3, 5, 6, 7, and 9. (b) How many of them are less than 400? (c) How many of them are even?

5.49. Find the number m of ways in which 6 people can ride a toboggan if one of 3 of them must drive.

5.50. Find n if: (a) $P(n, 4) = 42P(n, 2)$; (b) $2P(n, 2) + 50 = P(2n, 2)$.

PERMUTATIONS WITH REPETITIONS, ORDERED SAMPLES

5.51. Find the number of permutations that can be formed from all the letters of each word: (a) QUEUE; (b) COMMITTEE; (c) PROPOSITION; (d) BASEBALL.

5.52. Suppose we are given 4 identical red flags, 2 identical blue flags, and 3 identical green flags. Find the number m of different signals that can be formed by hanging the 9 flags in a vertical line.

5.53. A box contains 12 lightbulbs. Find the number n of ordered samples of size 3:
(a) with replacement; (b) without replacement.

5.54. A class contains l0 students. Find the number n of ordered samples of size 4:
(a) with replacement; (b) without replacement.

COMBINATIONS

5.55. A restaurant has 6 different desserts. Find the number of ways a customer can choose:
(a) 1 dessert; (b) 2 of the desserts; (c) 3 of the desserts.

5.56. A class contains 9 men and 3 women. Find the number of ways a teacher can select a committee of 4 from the class where there is:
(a) no restrictions; (b) 2 men and 2 women; (c) exactly one woman; (d) at least one woman.

5.57. A woman has 11 close friends. Find the number of ways she can invite 5 of them to dinner where:
(a) There are no restrictions.
(b) Two of the friends are married to each other and will not attend separately.
(c) Two of the friends are not speaking with each other and will not attend together.

5.58. A class contains 8 men and 6 women and there is one married couple in the class. Find the number m of ways a teacher can select a committee of 4 from the class where the husband or wife but not both can be on the committee.

5.59. A box has 6 blue socks and 4 white socks. Find the number of ways two socks can be drawn from the box where:
(a) There are no restrictions. (b) They are different colors. (c) They are the same color.

5.60. A women student is to answer 10 out of 13 questions. Find the number of her choices where she must answer:
(a) the first two questions;
(b) the first or second question but not both;
(c) exactly 3 out of the first 5 questions;
(d) at least 3 of the first 5 questions.

INCLUSION–EXCLUSION PRINCIPLE

5.61. Suppose 32 students are in an art class A and 24 students are in a biology class B, and suppose 10 students are in both classes. Find the number of students who are:
(a) in class A or in class B; (b) only in class A; (c) only in class B.

5.62. A survey of 80 car owners shows that 24 own a foreign-made car and 60 own a domestic-made car. Find the number of them who own:
(a) both a foreign made car and a domestic made car;
(b) only a foreign made car;
(c) only a domestic made car.

5.63. Consider all integers from 1 up to and including 100. Find the number of them that are:
(a) odd or the square of an integer; (b) even or the cube of an integer.

5.64. In a class of 30 students, 10 got A on the first test, 9 got A on a second test, and 15 did not get an A on either test. Find: the number of students who got:
(a) an A on both tests;
(b) an A on the first test but not the second;
(c) an A on the second test but not the first.

5.65. Consider all integers from 1 up to and including 300. Find the number of them that are divisible by:
(a) at least one of 3, 5, 7;
(b) 3 and 5 but not by 7;
(c) by 5, but by neither 3 nor 7;
(d) by none of the numbers 3, 5, 7.

5.66. In a certain school, French (F), Spanish (S), and German (G) are the only foreign languages taught. Among 80 students:

(i) 20 study F, 25 study S, 15 study G.

(ii) 8 study F and S, 6 study S and G, 5 study F and G.

(iii) 2 study all three languages.

Find the number of the 80 students who are studying:

(a) none of the languages; (c) only one language; (e) exactly two of the languages.

(b) only French; (d) only Spanish and German;

5.67. Find the number m of elements in the union of sets A, B, C, D where:

(i) A, B, C, D have 50, 60, 70, 80 elements, respectively.

(ii) Each pair of sets has 20 elements in common.

(iii) Each three of the sets has 10 elements in common.

(iv) All four of the sets have 5 elements in common.

PIGEONHOLE PRINCIPLE

5.68. Find the minimum number of students needed to guarantee that 4 of them were born: (a) on the same day of the week; (b) in the same month.

5.69. Find the minimum number of students needed to guarantee that 3 of them:

(a) have last names which begin with the same first letter;

(b) were born on the same day of a month (with 31 days).

5.70. Consider a tournament with n players where each player plays against every other player. Suppose each player wins at least once. Show that at least 2 of the players have the same number of wins.

5.71. Suppose 5 points are chosen at random in the interior of an equilateral triangle T where each side has length two inches. Show that the distance between two of the points must be less than one inch.

5.72. Consider any set $X = \{x_1, x_2, \ldots, x_7\}$ of seven distinct integers. Show that there exist $x, y \in X$ such that $x + y$ or $x - y$ is divisible by 10.

MISCELLANEOUS PROBLEMS

5.73. Find the number m of ways 10 students can be divided into three teams where one team has 4 students and the other teams have 3 students.

5.74. Assuming a cell can be empty, find the number n of ways that a set with 3 elements can be partitioned into: (a) 3 ordered cells; (b) 3 unordered cells.

5.75. Assuming a cell can be empty, find the number n of ways that a set with 4 elements can be partitioned into: (a) 3 ordered cells; (b) 3 unordered cells.

5.76. The English alphabet has 26 letters of which 5 are vowels. Consider only 5-letter "words" consisting of 3 different consonants and 2 different vowels. Find the number of such words which:

(a) have no restrictions; (c) contain the letters B and C;

(b) contain the letter B; (d) begin with B and contain the letter C.

5.77. Teams A and B play in the World Series of baseball, where the team that first wins four games wins the series. Suppose A wins the first game, and that the team that wins the second game also wins the fourth game.

(a) Find and list the number n of ways the series can occur.

(b) Find the number of ways that B wins the series.

(c) Find the number of ways the series lasts seven games.

5.78. Find the number of ways a coin can be tossed:

(a) 6 times so that there is exactly 3 heads and no two heads occur in a row.

(b) $2n$ times so that there is exactly n heads and no two heads occur in a row.

5.79. Find the number of ways 3 elements a, b, c, can be assigned to 3 cells, so exactly 1 cell is empty.

5.80. Find the number of ways n distinct elements can be assigned to n cells so exactly 1 cell is empty.

Answers to Supplementary Problems

5.31. (a) 3 628 800; 39 916 800; 479 001 600;
(b) $\log(60!) = 81.92$, so $60! = 6.59 \times 10^{81}$.

5.32. (a) 240; (b) 2 184; (c) 1/90; (d) 1/1716.

5.33. (a) $n + 1$; (b) $n(n-1)$; (c) $1/[n(n+1)(n+2)$;
(d) $(n-r)(n-r+1)$.

5.34. (a) 10; (b) 35; (c) 91; (d) 15; (e) 1140; (f) 816.

5.35. Hints: (a) Expand $(1+1)^n$; (b) Expand $(1-1)^n$.

5.36. (a) 1, 9, 36, 84, 126, 126, 84, 36, 9, 1;
(b) 1, 10, 45, 120, 210, 252, 210, 120, 45, 10, 1.

5.37. (a) 60; (b) 210; (c) 504; (d) not defined.

5.38. (a) 15; (b) 105.

5.39. (a)18; (b) 80; (c) 306.

5.40. (a) $26^2 \cdot 10^3$; (b) $26 \cdot 25 \cdot 10^3$; (b); (c) $26 \cdot 10^3$.

5.41. (a) $26 \cdot 25 \cdot 10 \cdot 9 \cdot 8 = 468\,000$; (b) $26 \cdot 25 \cdot 9 \cdot 9 \cdot 8 = 421\,200$.

5.42. $m = 18 \cdot 17 \cdot 16 = 4896$.

5.43. (a) 12!; (b) 3!5!4!3! = 103 680.

5.44. (a) 6! = 720; (b) $2 \cdot 3! \cdot 3! \cdot = 72$; (c) $4 \cdot 3! \cdot 3! = 144$.

5.45. (a) 120; (b) 48.

5.46. (a) 24; (b) 12.

5.47. (a) $3 \cdot 9 \cdot 8$; (b) $9 \cdot 8 \cdot 5$; (c) $9 \cdot 8 \cdot 7/2$; (d) $9 \cdot 8 \cdot 7/5$.

5.48. (a) $P(6, 3) = 120$; (b) $2 \cdot 5 \cdot 4 = 40$; (c) $2 \cdot 5 \cdot 4 = 40$.

5.49. $m = 360$.

5.50. (a) 9; (b) 5.

5.51. (a) 30; (b) 9!/[2!2!2!] = 45 360; (c) 11!/[2!3!2!] = 1 663 200; (d) 8!/[2!2!2!] = 5040.

5.52. $m = 9!/[4!2!3!] = 1260$.

5.53. (a) $12^3 = 1\,728$; (b) $P(12, 3) = 1320$.

5.54. (a) $10^4 = 10\,000$; (b) $P(10, 4) = 5040$.

5.55. (a) 6; (b) 15; (c) 20.

5.56. (a) $C(12, 4)$; (b) $C(9, 2) \cdot C(3, 2) = 108$;
(c) $C(9, 3) \cdot 3 = 252$; (d) $9 + 108 + 252 = 369$ or $C(12, 4) - C(9, 4) = 369$.

5.57. (a) $C(11, 5) = 462$; (b) $126 + 84 = 210$;
(c) $C(9, 5) + 2C(9, 4) = 378$.

5.58. $m = C(12, 4) + 2C(12, 3) = 935$.

5.59. (a) $C(10, 2) = 45$; (b) $6 \cdot 4 = 24$; (c) $C(6, 2) + C(4, 2) = 21$ or $45 - 24 = 21$.

5.60. (a) 165; (b) 110; (c) 80; (d) 276.

5.61. (a) 46; (b) 22; (c) 14.

5.62. (a) 4; (b) 20; (c) 56.

5.63. (a) 55; (b) 52.

5.64. (a) 4; (b) 6; (c) 5.

5.65. (a) $100 + 60 + 42 - 20 - 14 - 8 + 2 = 162$;
(b) $20 - 2 = 18$; (c) $60 - 20 - 8 + 2 = 34$;
(d) $300 - 162 = 138$.

5.66. (a) 37; (b) 9; (c) 28; (d) 4; (e) 13.

5.67. $m = 175$

5.68. (a) 22; (b) 37.

5.69. (a) 53; (b) 63.

5.70. Each player will win anywhere from 1 up to $n - 1$ games (pigeonholes). There are n players (pigeons).

5.71. Draw three lines between the midpoints of the sides of T. This partitions T into 4 equilateral triangles (pigeonholes) where each side has length 1. Two of the 5 points (pigeons) must lie in one of the triangles.

5.72. Let r_i be the remainder when x_i is divisible by 10. Consider the six pigeonholes: $H_1 = \{x_i \mid r_i = 0\}$, $H_2 = \{x_i \mid r_i = 5\}$, $H_3 = \{x_i \mid r_i = 1 \text{ or } 9\}$, $H_4 = \{x_i \mid r_i = 2 \text{ or } 8\}$, $H_5 = \{x_i \mid r_i = 3 \text{ or } 7\}$, $H_6 = \{x_i \mid r_i = 4 \text{ or } 6\}$. Then some x and y belong to some H_k.

5.73. $m = C(10, 4) \cdot C(6, 3) = 420$

5.74. (a) $n = 3^3 = 27$ (Each element can be placed in any of the three cells.) (b) The number of elements in three cells can be distributed as follows: [3, 0, 0], [2,1,0], or [1,1,1]. Thus $n = 1 + 3 + 1 = 5$.

5.75. (a) $n = 3^4 = 81$ (Each element can be placed in any of the three cells.) (b) The number of elements in three cells can be distributed as follows: [4, 0, 0], [3, 1, 0], [2, 2, 0], or [2, 1, 1]. Thus $n = 1 + 4 + 3 + 6 = 14$.

5.76. (a) $C(21, 3) \cdot C(5, 2) \cdot 5!$; (b) $C(20, 2) \cdot C(5, 2) \cdot 5!$; (c) $19 \cdot C(5, 2) \cdot 5!$; (d) $19 \cdot C(5, 2) \cdot 4!$.

5.77. Draw tree diagram T as in Fig. 5-4. Note T begins at A, the winner of the first game, and there is only one choice in the fourth game, the winner of the second game.
(a) $n = 15$ as listed below; (b) 6; (c) 8:
AAAA, *AABAA*, *AABABA*, *AABABBA*, *AABABBB*, *ABABAA*, *ABABABA*, *ABABABB*, *ABABBAA*, *ABABBAB*, *ABABBB*, *ABBBAAA*, *ABBBAAB*, *ABBBAB*, *ABBBB*.

5.78. (a) 4, *HTHTHT*, *HTTHTH*, *HTHTTH*, *THTHTH*;
(b) $n + 1$.

5.79. 18.

5.80. $n!C(n, 2)$.

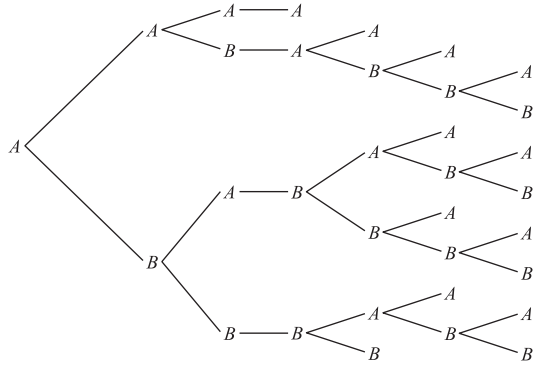


Fig. 5-4

CHAPTER 6

Advanced Counting Techniques, Recursion

6.1 INTRODUCTION

Here we consider more sophisticated counting techniques and problems. This includes problems involving combinations with repetition, ordered and unordered partitions, and the Inclusion–Exclusion Principle and the Pigeonhole Principle.

We also discuss recursion in this chapter.

6.2 COMBINATIONS WITH REPETITIONS

Consider the following problem. A bakery makes only $M = 4$ kinds of cookies: apple (a), banana (b), carrot (c), dates (d). Find the number of ways a person can buy $r = 8$ of the cookies.

Observe that order does not count. This is an example of combinations with repetitions. In particular, each combination can be listed with the a's first, then the b's, then the c's, and finally the d's. Four such combinations follow:

$$r_1 = aa,\ bb,\ cc,\ dd; \quad r_2 = aaa,\ c,\ ddd; \quad r_3 = bbbb,\ c,\ ddd; \quad r_4 = aaaaa,\ ddd.$$

Counting the number m of such combinations may not be easy.

Suppose we want to code the above combinations using only two symbols, say 0 and 1. This can be done by letting 0 denote a cookie, and letting 1 denote a change from one kind of cookie to another. Then each combination will require $r = 8$ zeros, one for each cookie, and $M - 1 = 3$ ones, where the first one denotes the change from a to b, the second one from b to c, and the third one from c to d. Thus the above four combinations will be coded as follows:

$$r_1 = 00100100100, \quad r_2 = 00001101000, \quad r_3 = 10000101000, \quad r_4 = 00000111000.$$

Counting the number m of these "codewords" is easy. Each codeword contains $R + M - 1 = 11$ digits where $r = 8$ are 0's and hence $M - 1 = 3$ are 1's. Accordingly,

$$M = C(11, 8) = C(11, 3) = \frac{11 \cdot 10 \cdot 9}{3 \cdot 2 \cdot 1} = 165$$

A similar argument gives us the following theorem.

Theorem 6.1: Suppose there are M kinds of objects. Then the number of combinations of r such objects is $C(r + M - 1, r) = C(r + M - 1, M - 1)$.

EXAMPLE 6.1 Find the number m of nonnegative integer solutions of $x + y + z = 18$.

We can view each solution, say $x = 3$, $y = 7$, $z = 8$, as a combination of $r = 18$ objects consisting of 3 a's, 7 b's, and 8 c's, where there are $M = 3$ kinds of objects, a's, b's, and c's. By Theorem 6.1,

$$m = C(r + M - 1, M - 1) = C(20, 2) = 190.$$

6.3 ORDERED AND UNORDERED PARTITIONS

Suppose a set has 7 elements. We want to find the number m of ordered partitions of S into three cells, say $[A_1, A_2, A_3]$, so they contain 2, 3, and 2 elements, respectively.

Since S has 7 elements, there are $C(7, 2)$ ways of choosing the first two elements for A_1. Following this, there are $C(5, 3)$ ways of choosing the 3 elements for A_2. Lastly, there are $C(2, 2)$ ways of choosing the 2 elements for A_3 (or, the last 2 elements form the cell A_3). Thus:

$$m = C(7, 2)C(5, 3)C(2, 2) = \binom{7}{2}\binom{5}{3}\binom{2}{2} = \frac{7 \cdot 6}{2 \cdot 1} \cdot \frac{5 \cdot 4 \cdot 3}{3 \cdot 2 \cdot 1} \cdot \frac{2 \cdot 1}{2 \cdot 1} = 210$$

Observe that

$$m = \binom{7}{2}\binom{5}{3}\binom{2}{2} = \frac{7!}{2!5!} \cdot \frac{5!}{3!2!} \cdot \frac{2!}{2!0!} = \frac{7!}{2!3!2!}$$

since each numerator after the first is cancelled by a term in the denominator of the previous factor.

The above discussion can be shown to be true in general. Namely:

Theorem 6.2: The number m of ordered partitions of a set S with n elements into r cells $[A_1, A_2, \ldots, A_r]$ where, for each i, $n(A_i) = n_i$, follows:

$$m = \frac{n!}{n_1!n_2! \ldots n_r!}$$

Unordered Partitions

Frequently, we want to partition a set S into cells $[A_1, A_2, \ldots, A_r]$ where the cells are now unordered. The number m of such unordered partitions is obtained from the number m' of ordered partitions by dividing m' by each $k!$ where k of the cells have the same number of elements.

EXAMPLE 6.2 Find the number m of ways to partition 10 students into four teams $[A_1, A_2, A_3, A_4]$ so that two teams contain 3 students and two teams contain 2 students.

By Theorem 6.2, there are $m' = 10!/(3!3!2!2!) = 25\,200$ such ordered partitions.

Since the teams form an unordered partition, we divide m' by 2! because of the two cells with 3 elements each and 2! because of the two cells with 2 elements each.

Thus $m = 25\,200/(2!2!) = 6300$.

6.4 INCLUSION–EXCLUSION PRINCIPLE REVISITED

Let $A_1, A_2, \ldots, A_r$ be subsets of a universal set U. Suppose we let s_k denote the sum of the cardinalities of all possible k-tuple intersections of the sets, that is, the sum of all of the cardinalities

$$n(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k})$$

For example,

$$s_1 \sum_i n(A_i), \quad s_2 = \sum_{i<j} n(A_i \cap A_j), \quad s_3 = \sum_{i_1<i_2<i_3} n(A_{i_1} \cap A_{i_2} \cap A_{i_3})$$

The Inclusion–Exclusion Principle, which appears in Section 5.7, gave a formula for the number of elements in the union of the sets. Specifically, (Theorem 5.9) we have

$$n(A_1 \cup A_2 \cup \cdots \cup A_r) = s_1 - s_2 + s_3 - \cdots + (-1)^{r-1} s_r$$

On the other hand, using DeMorgan's law,

$$n(A_1^{\mathrm{C}} \cap A_2^{\mathrm{C}} \cap \ldots \cap A_r^{\mathrm{C}}) = n([A_1 \cup A_2 \cup \ldots \cup A_r]^{\mathrm{C}}) = |\mathbf{U}| - n(A_1 \cup A_2 \cup \ldots \cup A_r)$$

Accordingly, we obtain an alternate form for Theorem 5.9:

Theorem (Inclusion–Exclusion Principle) 6.3: Let $A_1, A_2, \ldots, A_r$ be subsets of a universal set $\mathbf{U}$. Then the number m of elements which do not appear in any of the subsets $A_1, A_2, \ldots, A_r$ of $\mathbf{U}$ is:

$$m = n(A_1^{\mathrm{C}} \cap A_2^{\mathrm{C}} \cap \ldots \cap A_r^{\mathrm{C}}) = |\mathbf{U}| - s_1 + s_2 - s_3 + \cdots + (-1)^r s_r$$

EXAMPLE 6.3 Let $\mathbf{U}$ be the set of positive integers not exceeding 1000. Then $|U| = 1000$. Find $|S|$ where S is the set of such integers which are not divisible by 3, 5, or 7.

Let A be the subset of integers which are divisible by 3, B which are divisible by 5, and C which are divisible by 7. Then $S = A^C \cap B^C \cap C^C$ since each element of S is not divisible by 3, 5 or 7. By integer division,

$$|A| = 1000/3 = 33, \quad |B| = 1000/5 = 200, \quad |C| = 1000/7 = 142,$$
$$|A \cap B| = 1000/15 = 66, \quad |A \cap C| = 1000/21 = 47, \quad |B \cap C| = 1000/35 = 28,$$
$$|A \cap B \cap C| = 1000/105 = 9$$

Thus, by the Inclusion–Exclusion Principle Theorem 6.3,

$$|S| = 1000 - (333 + 200 + 142) + (66 + 47 + 28) - 9 = 1000 - 675 + 141 - 9 = 457$$

Number of Onto Functions

Let A and B be sets such that $|A| = 6$ and $|B| = 4$. We want to find the number of surjective (onto) functions from A onto B.

Let b_1, b_2, b_3, b_4 be the four elements in B. Let $\mathbf{U}$ be the set of all functions from A into B. Furthermore, let F_1 be the set of functions which do not send any element of A into b_1 that is, b_1 is not in the range of any function in F_1. Similarly, let F_2, F_3, and F_4 be the sets of functions which do not send any element of A into b_2, b_3, and b_4, respectively.

We are looking for the number of functions in $S = F_1^C \cap F_2^C \cap F_3^C \cap F_4^C$, that is, those functions which do send at least one element of A into b_1, at least one element of A into b_2, and so on. We will use the Inclusion–Exclusion Principle Theorem 6.3 as follows.

(i) For each function in $\mathbf{U}$, there are 4 choices for each of the 6 elements in A; hence $|U| = 4^6 = 4096$

(ii) There are $C(4, 1) = 4$ functions F_i. In each case, there are 3 choices for each of the 6 elements in A, hence $|F_i| = 3^6 = 729$.

(iii) There are $C(4, 2) = 6$ pairs $F_i \cap F_j$. In each case, there are 2 choices for each of the 6 elements in A, hence $|F_i \cap F_j| = 2^6 = 64$.

(iv) There are $C(4, 3) = 4$ triplets $F_i \cap F_j \cap F_k$. In each case, there is only one choice for each of the 6 elements in A. Hence $|F_i \cap F_j \cap F_k| = 1^6 = 1$.

(v) $F_1 \cap F_2 \cap F_3 \cap F_4$ has no element, that is, is empty. Hence $|F_1 \cap F_2 \cap F_3 \cap F_4| = 0$. By the Inclusion–Exclusion Principle Theorem 6.3,

$$|S| = |F_1^C \cap F_2^C \cap F_3^C \cap F_4^C| = 4^6 - C(4,1)3^6 + C(4,2)2^6 - C(4,3)1^7$$
$$= 4096 - 2916 + 384 - 1 = 795$$

The above result is true in general. Namely:

Theorem 6.4: Suppose $|A| = m$ and $|B| = n$ where $m \geq n$. Then the number N of surjective (onto) functions from A onto B is:

$$N = n^m - C(n,1)(n-1)^m + C(n,2)(n-2)^m - \cdots + (-1)^{n-1}C(n,n-1)1^m$$

Derangements

A *derangement* is a permutation of objects where each object is not in its original position. For example, 453162 is not a derangement of 123456 since 3 is in its correct position, but 264531 is a derangement of 123456. (Alternately, a permutation $\sigma\colon X \to X$ is a derangement if $\sigma(i) \neq i$ for every $i \in X = \{1, 2, \ldots, n\}$.)

Let D_n denote the number of derangements of n objects. For example, 231 and 312 are the only derangements of 123. Hence $D_3 = 2$. The following theorem, proved in Problem 6.6, applies.

Theorem 6.5: $D_n = n![1 - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} + \cdots + (-1)^n \frac{1}{n!}]$

The probability (Chapter 7) that a derangement of n objects occurs equals D_n divided by $n!$, the number of permutations of the n objects. Thus Theorem 6.5 yields:

Corollary 6.6: Let p be the probability of a derangement of n objects. Then

$$p = 1 - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} + \cdots + (-1)^n \frac{1}{n!}$$

EXAMPLE 6.4 (Hat Check Problem) Suppose $n = 5$ people check in their hats at a restaurant and they are given back their hats at random. Find the probability p that no person receives his/her own hat.

This is an example of a derangement with $n = 5$. By Corollary 6.6,

$$p = 1 - 1 + 1/2 - 1/6 + 1/24 - 1/120 = 44/120 = 11/30 \approx 0.367$$

Note that the signs alternate and the terms get very, very small in Corollary 6.6. Figure 6-1 gives the values of p for the first few values of n. Note that, for $n > 4$, p is very close to the following value (where $e = 2.718$):

$$e^{-1} = 1 - \tfrac{1}{1!} + \tfrac{1}{2!} - \tfrac{1}{3!} + \cdots + (-1)^n \tfrac{1}{n!} + \cdots \approx 0.368$$

n	1	2	3	4	5	6	7
$p = D_n/n!$	0.0000	0.5000	0.3333	0.3750	0.3667	0.3681	0.3679

Fig. 6-1

6.5 PIGEONHOLE PRINCIPLE REVISITED

The Pigeonhole Principle (with its generalization) is stated with simple examples in Section 5.6. Here we give examples of more sophisticated applications of this principle.

EXAMPLE 6.5 Consider six people, where any two of them are either friends or strangers. Show that there are three of them which are either mutual friends or mutual strangers.

Let A be one of the people. Let X consist of those which are friends of A, and Y consist of those which are strangers of A. By the Pigeonhole Principle, either X or Y has at least three people. Suppose X has three people. If two of them are friends, then the two with A are three mutual friends. If not, then X has three mutual strangers. Alternately, suppose Y has three people. If two of them are strangers, then the two with A are three mutual strangers. If not, then X has three mutual friends.

EXAMPLE 6.6 Consider five *lattice* points $(x_1, y_1), \ldots, (x_5, y_5)$ in the plane, that is, points with integer coordinates. Show that the midpoint of one pair of the points is also a lattice point.

The midpoint of points $P(a, b)$ and $Q(c, d)$ is $([a + c]/2, [b + d]/2)$. Note that $(r + s)/2$ is an integer if r and s are integers with the same *parity*, that is, both are odd or both are even. There are four pairs of parities: (odd, odd), (odd, even), (even, odd), and (even, even). There are five points. By the Pigeonhole Principle, two of the points have the same pair of parities. The midpoint of these two points has integer coordinates.

An important application of the Pigeonhole Principle follows.

Theorem 6.7: Every sequence of distinct $n^2 + 1$ real numbers contains a subsequence of length $n + 1$ which is strictly increasing or strictly decreasing.

For example, consider the following sequence of $10 = 3^2 + 1$ numbers (where $n = 3$): 2, 1, 8, 6, 7, 5, 9, 4, 12, 3. There are many subsequences of length $n + 1 = 4$ which are strictly increasing or strictly decreasing; for example,

$$2, 6, 9, 12; \quad 1, 5, 9, 12; \quad 8, 6, 5, 4; \quad 7, 5, 4, 3.$$

On the other hand, the following sequence of $9 = 3^2$ numbers has no subsequence of length $n + 1 = 4$ which is strictly increasing or strictly decreasing:

$$3, \quad 2, \quad 1, \quad 6, \quad 5, \quad 4, \quad 9, \quad 8, \quad 7.$$

The proof of Theorem 6.7 appears in Problem 6.10.

6.6 RECURRENCE RELATIONS

Previously, we discussed recursively defined functions such as

(a) Factorial function, (b) Fibonacci sequence, (c) Ackermann function.

Here we discuss certain kinds of recursively defined sequences $\{a_n\}$ and their solution. We note that a *sequence* is simply a function whose domain is

$$\mathbf{N} = \{1, 2, 3, \ldots\} \quad \text{or} \quad \mathbf{N}_0 = \mathbf{N} \cup \{0\} = \{0, 1, 2, 3, \ldots\}$$

We begin with some examples.

EXAMPLE 6.7 Consider the following sequence which begins with the number 3 and for which each of the following terms is found by multiplying the previous term by 2:

$$3, \quad 6, \quad 12, \quad 24, \quad 48, \quad \ldots$$

It can be defined recursively by:

$$a_0 = 3, \quad a_k = 2a_{k-1} \text{ for } k \geq 1 \quad \text{or} \quad a_0 = 3, \quad a_{k+1} = 2a_k \quad \text{for } k \geq 0$$

The second definition may be obtained from the first by setting $k = k + 1$. Clearly, the formula $a_n = 3(2^n)$ gives us the nth term of the sequence without calculating any previous term.

The following remarks about the above example are in order.

(1) The equation $a_k = 2a_{k-1}$ or, equivalently, $a_{k+1} = 2a_k$, where one term of the sequence is defined in terms of previous terms of the sequence, is called a *recurrence relation*.
(2) The equation $a_0 = 3$, which gives a specific value to one of the terms, is called an *initial condition*.
(3) The function $a_n = 3(2^n)$, which gives a formula for a_n as a function of n, not of previous terms, is called a *solution* of the recurrence relation.
(4) There may be many sequences which satisfy a given recurrence relation. For example, each of the following is a solution of the recurrence relation $a_k = 2a_{k-1}$.

$$1, 2, 4, 8, 16, \ldots \quad \text{and} \quad 7, 14, 28, 56, 112, \ldots$$

All such solutions form the so-called *general solution* of the recurrence relation.
(5) On the other hand, there may be only a unique solution to a recurrence relation which also satisfies given initial conditions. For example, the initial condition $a_0 = 3$ uniquely yields the solution $3, 6, 12, 24, \ldots$ of the recurrence relation $a_k = 2a_{k-1}$.

This chapter shows how to solve certain recurrence relations. First we give two important sequences the reader may have previously studied.

EXAMPLE 6.8

(a) *Arithmetic Progression*

An arithmetic progression is a sequence of the form

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

That is, the sequence begins with the number a and each successive term is obtained from the previous term by adding d (the common difference between any two terms). For example:

(i) $a = 5, d = 3$: $5, 8, 9, 11, \ldots$
(ii) $a = 2, d = 5$: $2, 7, 12, 17, \ldots$
(iii) $a = 1, d = 0$: $1, 1, 1, 1, 1, \ldots$

We note that the general arithmetic progression may be defined recursively by:

$$a_1 = a \quad \text{and} \quad a_{k+1} = a_k + d \quad \text{for } k \geq 1$$

where the solution is $a_n = a + (n - 1)d$.

(b) *Geometric Progression*

A geometric progression is a sequence of the form

$$a, ar, ar^2, ar^3, \ldots$$

That is, the sequence begins with the number a and each successive term is obtained from the previous term by multiplying by r (the common ratio between any two terms) for example:

(i) $a = 1, r = 3$: $1, 3, 9, 27, 81, \ldots$
(ii) $a = 5, r = 2$: $5, 10, 20, 40, \ldots$
(iii) $a = 1, r = \frac{1}{2}$: $1, \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \ldots$

We note that the general geometric progression may be defined recursively by:

$$a_1 = a \quad \text{and} \quad a_{k+1} = ra_k \quad \text{for } k \geq 1$$

where the solution is $a_{n+1} = ar^n$.

6.7 LINEAR RECURRENCE RELATIONS WITH CONSTANT COEFFICIENTS

A *recurrence relation of order k* is a function of the form

$$a_n = \Phi(a_{n-1}, a_{n-2}, \ldots, a_{n-k}, n)$$

that is, where the nth term a_n of a sequence is a function of the preceding k terms $a_{n-1}, a_{n-2}, \ldots, a_{n-k}$ (and possibly n). In particular, a *linear kth-order recurrence relation with constant coefficients* is a recurrence relation of the form

$$a_n = C_1 a_{n-1} + C_2 a_{n-2} + \cdots + C_k a_{n-k} + f(n)$$

where $C_1, C_2, \ldots, C_k$ are constants with $C_k \neq 0$, and $f(n)$ is a function of n. The meanings of the names linear and constant coefficients follow:

Linear: There are no powers or products of the a_j's.
Constant coefficients: The $C_1 C_2, \ldots, C_k$ are constants (do not depend on n).

If $f(n) = 0$, then the relation is also said to be *homogeneous*.

Clearly, we can uniquely solve for a_n if we know the values of $a_{n-1}, a_{n-2}, \ldots, a_{n-k}$. Accordingly, by mathematical induction, there is a unique sequence satisfying the recurrence relation if we are given *initial values* for the first k elements of the sequence.

EXAMPLE 6.9 Consider each of the following recurrence relations.

(a) $a_n = 5a_{n-1} - 4a_{n-2} + n^2$

This is a second-order recurrence relation with constant coefficients. It is nonhomogeneous because of the n^2. Suppose we are given the initial conditions $a_1 = 1$, $a_2 = 2$. Then we can find sequentially the next few elements of the sequence:

$$a_3 = 5(2) - 4(1) + 3^2 = 15, \quad a_4 = 5(15) - 4(2) + 4^2 = 83$$

(b) $a_n = 2a_{n-1}a_{n-2} + n^2$

The product $a_{n-1}a_{n-2}$ means the recurrence relation is not linear. Given initial conditions $a_1 = 1$, $a_2 = 2$, we can still find the next few elements of the sequence:

$$a_3 = 2(2)(1) + 3^2 = 13, \quad a_4 = 2(13)(2) + 4^2 = 68$$

(c) $a_n = na_{n-1} + 3a_{n-2}$

This is a homogeneous linear second-order recurrence relation but it does not have constant coefficients because the coefficient of a_{n-1} is n, not a constant. Given initial conditions $a_1 = 1$, $a_2 = 2$, the next few elements of the sequence follow:

$$a_3 = 3(2) + 3(1) = 9, \quad a_4 = 4(9) + 3(2) = 42$$

(d) $a_n = 2a_{n-1} + 5a_{n-2} - 6a_{n-3}$

This is a homogeneous linear third-order recurrence relation with constant coefficients. Thus we need three, not two, initial conditions to yield a unique solution of the recurrence relation. Suppose we are given the initial conditions $a_1 = 1$, $a_2 = 2$, $a_3 = 1$. Then, the next few elements of the sequence follow:

$$a_4 = 2(1) + 5(2) - 6(1) = 6, \quad a_5 = 2(2) + 5(1) - 6(6) = -37$$
$$a_6 = 2(1) + 5(6) - 6(-37) = 254$$

This chapter will investigate the solutions of homogeneous linear recurrence relations with constant coefficients. The theory of nonhomogeneous recurrence relations and recurrence relations without constant coefficients lies beyond the scope of this text.

For computational convenience, most of our sequences will begin with a rather than a. The theory is not affected at all.

6.8 SOLVING SECOND-ORDER HOMOGENEOUS LINEAR RECURRENCE RELATIONS

Consider a homogeneous second-order recurrence relation with constant coefficients which has the form

$$a_n = sa_{n-1} + ta_{n-2} \quad \text{or} \quad a_n - sa_{n-1} - ta_{n-2} = 0$$

where s and t are constants with $t \neq 0$. We associate the following quadratic polynomial with the above recurrence relation:

$$\Delta(x) = x^2 - sx - t$$

This polynomial $\Delta(x)$ is called the *characteristic polynomial* of the recurrence relation, and the roots of $\Delta(x)$ are called its *characteristic roots*.

Theorem 6.8: Suppose the characteristic polynomial $\Delta(x) = x^2 - sx - t$ of the recurrence relation

$$a_n = sa_{n-1} + ta_{n-2}$$

has distinct roots r_1 and r_2. Then the general solution of the recurrence relation follows, where c_1 and c_2 are arbitrary constants:

$$a_n = c_1 r_1^n + c_2 r_2^n$$

We emphasize that the constants c_1 and c_2 may be uniquely computed using initial conditions. We note that the theorem is true even when the roots are not real. Such cases lie beyond the scope of this text.

EXAMPLE 6.10 Consider the following homogeneous recurrence relation:

$$a_n = 2a_{n-1} + 3a_{n-2}$$

The general solution is obtained by first finding its characteristic polynomial $\Delta(x)$ and its roots r_1 and r_2:

$$\Delta(x) = x^2 - 2x - 3 = (x-3)(x+1); \quad \text{roots } r_1 = 3, r_2 = -1$$

Since the roots are distinct, we can use Theorem 6.8 to obtain the general solution:

$$a_n = c_1 3^n + c_2(-1)^n$$

Thus any values for c_1 and c_2 will give a solution to the recurrence relation.

Suppose we are also given the initial conditions $a_0 = 1$, $a_1 = 2$. Using the recurrence relation we can compute the next few terms of the sequence:

$$1, \quad 2, \quad 8, \quad 28, \quad 100, \quad 356, \quad 1268, \quad 3516, \quad \ldots$$

The unique solution is obtained by finding c_1 and c_2 using the initial conditions. Specifically:

$$\text{For } n = 0 \text{ and } a_0 = 1, \text{ we get:} \quad c_1 3^0 + c_2(-1)^0 = 1 \quad \text{or} \quad c_1 + c_2 = 1$$
$$\text{For } n = 1 \text{ and } a_1 = 2, \text{ we get:} \quad c_1 3^1 + c_2(-1)^1 = 2 \quad \text{or} \quad 3c_1 - c_2 = 2$$

Solving the system of the two equations in the unknowns c_1 and c_2 yields:

$$c_1 = \frac{3}{4} \quad \text{and} \quad c_2 = \frac{1}{4}$$

Thus the following is the unique solution of the given recurrence relation with the given initial conditions $a_0 = 1$, $a = 2$:

$$a_n = \frac{3}{4}3^n + \frac{1}{4}(-1)^n = \frac{3^{n+1} + (-1)^n}{4}$$

EXAMPLE 6.11 Consider the celebrated Fibonacci sequence:

$$a_n = a_{n-1} + a_{n-2}, \quad \text{with} \quad a_0 = 0, a_1 = 1$$

The first 10 terms of the sequence follow:

$$0, 1, 1, 2, 3, 5, 8, 13, 21, 34, \ldots$$

Sometimes the Fibonacci sequence is defined using the initial conditions $a_0 = 1, a_1 = 1$ or the initial conditions $a_1 = 1, a_2 = 2$. We use $a_0 = 0, a_1 = 1$ for computational convenience. (All three initial conditions yield the same sequence after the pair of terms 1, 2.)

Observe that the Fibonacci sequence is a homogeneous linear second-order recurrence relation, so it can be solved using Theorem 6.8. Its characteristic polynomial follows:

$$\Delta(x) = x^2 - x - 1$$

Using the quadratic formula, we obtain the roots:

$$r_1 = \frac{1+\sqrt{5}}{2}, \quad r_2 = \frac{1-\sqrt{5}}{2}$$

By Theorem 6.8, we obtain the general solution:

$$a_n = c_1 \left(\frac{1+\sqrt{5}}{2}\right)^n + c_2 \left(\frac{1-\sqrt{5}}{2}\right)^n$$

The initial conditions yield the following system of two linear equations in c_1 and c_2

$$\text{For } n = 0 \text{ and } a_0 = 0, \text{ we get:} \qquad 0 = c_1 + c_2$$

$$\text{For } n = 1 \text{ and } a_1 = 1, \text{ we get:} \qquad 1 = c_1 \left(\frac{1+\sqrt{5}}{2}\right) + c_2 \left(\frac{1-\sqrt{5}}{2}\right)$$

The solution of the system follows:

$$c_1 = \frac{1}{\sqrt{5}}, c_2 = -\frac{1}{\sqrt{5}}$$

Accordingly, the following is the solution of the Fibonacci recurrence relation:

$$a_n = \frac{1}{\sqrt{5}} \left(\frac{1+\sqrt{5}}{2}\right)^n - \frac{1}{\sqrt{5}} \left(\frac{1-\sqrt{5}}{2}\right)^n$$

One can show that the absolute value of the above second term for a_n is always less than 1/2. Thus a_n is also the closest integer to the number

$$\frac{1}{\sqrt{5}} \left(\frac{1+\sqrt{5}}{2}\right)^n \approx (0.4472)(1.6180)^n$$

Solution when Roots of the Characteristic Polynomial are Equal

Suppose the roots of the characteristic polynomial are not distinct. Then we have the following result.

Theorem 6.9: Suppose the characteristic polynomial $\Delta(x) = x^2 - sx - t$ of the recurrence relation

$$a_n = sa_{n-1} + ta_{n-2}$$

has only one root r_0. Then the general solution of the recurrence relation follows, where c_1 and c_2 are arbitrary constants:

$$a_n = c_1 r_0^n + c_2 n r_0^n$$

The constants c_1 and c_2 may be uniquely computed using initial conditions.

EXAMPLE 6.12 Consider the following homogeneous recurrence relation:

$$a_n = 6a_{n-1} - 9a_{n-2}$$

The characteristic polynomial $\Delta(x)$ follows:

$$\Delta(x) = x^2 - 6x + 9 = (x-3)^2$$

Thus $\Delta(x)$ has only the one root $r_0 = 3$. Now we use Theorem 6.9 to obtain the following general solution of the recurrence relation:

$$a_n = c_1 3^n + c_2 n 3^n$$

Thus any values for c_1 and c_2 will give a solution to the recurrence relation.

Suppose we are also given the initial conditions $a_1 = 3$, $a_2 = 27$. Using the recurrence relation we can compute the next few terms of the sequence:

$$3, \quad 27, \quad 135, \quad 567, \quad 2187, \quad 8109, \quad \ldots$$

The unique solution is obtained by finding c_1 and c_2 using the initial conditions. Specifically:

$$\begin{array}{llll} \text{For } n = 1 \text{ and } a_1 = 3, & \text{we get:} & c_1 3^1 + c_2(1)(3)^1 = 3 & \text{or} \quad 3c_1 + 3c_2 = 3 \\ \text{For } n = 2 \text{ and } a_2 = 27, & \text{we get:} & c_1 3^2 + c_2(2)(3)^2 = 27 & \text{or} \quad 9c_1 + 18c_2 = 27 \end{array}$$

Solving the system of the two equations in the unknowns c_1 and c_2 yields:

$$c_1 = -1 \quad \text{and} \quad c_2 = 2$$

Thus the following is the unique solution of the recurrence relation with the given initial conditions:

$$a_n = -3^n + 2n3^n = 3^n(2n-1)$$

6.9 SOLVING GENERAL HOMOGENEOUS LINEAR RECURRENCE RELATIONS

Consider now a general linear homogeneous kth-order recurrence relation with constant coefficients which has the form

$$a_n = C_1 a_{n-1} + C_2 a_{n-2} + C_3 a_{n-3} + \cdots + C_k a_{n-k} = \sum_{i=1}^{k} C_i a_{n-1} \tag{6.1}$$

where $C_1, C_2, \ldots, C_k$ are constants with $C_k \neq 0$. The *characteristic polynomial* $\Delta(x)$ of the recurrence relation (*6.1*) follows:

$$\Delta(x) = x^k - C_1 x^{k-1} - C_2 x^{k-2} - C_3 x^{k-3} - \cdots - C_k = x^k - \sum_{i=1}^{k} C_1 x^{k-1}$$

The roots of $\Delta(x)$ are called the *characteristic roots* of the recurrence relation.

The following remarks are in order.

Remark 1: If $p(n)$ and $q(n)$ are solutions of (*6.1*), then any linear combination

$$c_1 p(n) + c_2 q(n)$$

of $p(n)$ and $q(n)$ is also a solution. (This is not true if the recurrence relation is nonhomogeneous.)

Remark 2: If r is a root of multiplicity m of the characteristic polynomial $\Delta(x)$ of (*6.1*), then each of the following

$$r^n, nr^n, n^2r^n, \ldots, n^{n-1}r^n$$

is a solution of (*6.1*). Thus any linear combination

$$c_1r^n + c_2nr^n + c_3n^2r^n + \cdots + c_mn^{m-1}r^n = (c_1 + c_2n + c_3n^2 + \cdots + c_mn^{m-1})r^n$$

is also a solution.

EXAMPLE 6.13 Consider the following third-order homogeneous recurrence relation:

$$a_n = 11a_{n-1} - 39a_{n-2} + 45a_{n-3}$$

The characteristic polynomial $\Delta(x)$ of the recurrence relation follows:

$$\Delta(x) = x^3 - 11x^2 + 39x - 45 = (x - 3)^2(x - 5)$$

Thus $\Delta(x)$ has two roots, $r_1 = 3$ of multiplicity 2 and $r_2 = 5$ of multiplicity 1. Thus, by the above remarks, the following is the general solution of the recurrence relation:

$$a_n = c_1(3^n) + c_2n(3^n) + c_3(5^n) = (c_1 + c_2n)(3^n) + c_3(5^n)$$

Thus any values for c_1, c_2, c_3 will give a solution to the recurrence relation.

Suppose we are also given the initial conditions $a_0 = 5$, $a_1 = 11$, $a_3 = 25$. Using the recurrence relation we can compute the next few terms of the sequence:

$$5, \quad 11, \quad 25, \quad 71, \quad 301, \quad 1667, \quad \ldots$$

The unique solution is obtained by finding c_1, c_2, c_3 using the initial conditions. Specifically:

$$\begin{array}{lll} \text{For } n = 0 \text{ and } a_0 = 5, & \text{we get:} & c_1 + c_3 = 5 \\ \text{For } n = 1 \text{ and } a_1 = 11, & \text{we get:} & 3c_1 + 3c_2 + 5c_3 = 11 \\ \text{For } n = 2 \text{ and } a_2 = 25, & \text{we get:} & 9c_1 + 18c_2 + 25c_3 = 25 \end{array}$$

Solving the system of the three equations in the unknowns c_1, c_2, c_3 yields:

$$c_1 = 4, \quad c_2 = -2, \quad c_3 = 1$$

Thus the following is the unique solution of the recurrence relation with the given initial conditions:

$$a_n = (4 - 2n)(3^n) + 5^n$$

Remark: Finding the roots of the characteristic polynomial $\Delta(x)$ is an important step in solving recurrence relations. Generally speaking, this may be difficult when the degree of $\Delta(x)$ is greater than 2. (Example B.16 indicates one way to find the roots of some polynomials of degree 3 or more.)

Solved Problems

ADVANCED COUNTING TECHNIQUES, INCLUSION–EXCLUSION

6.1. A bagel shop sells $M = 5$ kinds of bagels. Find the number m of ways a customer can buy: (a) 8 bagels; (b) a dozen bagels.

Use $m = C(r+M-1, r) = C(r+M-1, M-1)$, that is, Theorem 6.1, since this problem concerns combinations with repetitions.

(a) Here $r = 8$, so $m = C(8+4, 4) = C(12, 4) = 494$.

(b) Here $r = 12$, so $m = C(12+4, 4) = C(16, 4) = 1820$.

6.2. Find the number m of nonnegative solutions to $x + y + z = 18$ with the conditions that $x \geq 3$, $y \geq 2$, $z \geq 1$.

Let $x' = x-3$, $y' = y-2$ and $z' = z-1$. Then m is also the number of nonnegative solutions to $x' + y' + z' = 12$. As in Example 6.1, this second problem concerns combinations with repetitions with $M = 3$ and $r = 12$. Thus

$$m = C(12+2, 2) = C(14, 2) = 91.$$

6.3. Let E be the equation $x + y + z = 18$. Find the number m of nonnegative solutions to E with the conditions that $x < 7$, $y < 8$, $z < 9$.

Let S be the set of all nonnegative solutions of E. Let A be the set of solutions for which $x \geq 7$, let B be the set of solutions for which $y \geq 8$, and let C be the set of solutions for which $z \geq 9$. Then

$$m = |A^C \cap B^C \cap C^C|$$

As in Problem 6.2, we obtain:

$$\begin{array}{ll} |A| = C(11+2, 2) = 78, & |A \cap B| = C(3+2, 2) = 10 \\ |B| = C(10+2, 2) = 66, & |A \cap C| = C(2+2, 2) = 6 \\ |C| = C(9+2, 2) = 55, & |B \cap C| = C(1+2, 2) = 3 \end{array}$$

Also, $|S| = C(18+2, 2) = 190$ and $|A \cap B \cap C| = 0$. By the Inclusion–Exclusion Principle,

$$m = 190 - (78 + 66 + 55) + (10 + 6 + 3) - 0 = 10$$

6.4. There are 9 students in a class. Find the number m of ways: (a) the 9 students can take 3 different tests if 3 students are to take each test; (b) the 9 students can be partitioned into 3 teams A, B, C so that each team contains 3 students,

(a) Method 1: We seek the number m of partitions of the 9 students into cells containing 3 students. By Theorem 6.2, $m = 9!/(3!3!3!) = 5040$.

Method 2: There are $C(9, 3)$ to choose three students to take the first test; then there are $C(6, 3)$ ways to choose 3 students to take the second test; and the remaining students take the third test. Thus $m = C(9, 3)C(6, 3) = 5040$.

(b) Each partition $\{A, B, C\}$ of the students can be arranged in $3! = 6$ ways as an ordered partition. By (a), there are 5040 such ordered partitions. Hence $m = 5040/6 = 840$.

6.5. Find the number N of ways a company can assign 7 projects to 4 people so that each person gets at least one project.

We want to find the number N of onto functions from a set with $m = 7$ elements onto a set with $n = 4$ elements. We use Theorem 6.4:

$$\begin{aligned} N &= 4^2 - C(4, 1)(3^7) + C(4, 2)(2^7) - C(4, 3)(1^7) \\ &= 4^7 - 4(3^7) + 6(2^7) - 4(1^7) = 16\,384 - 8748 + 768 - 4 = 8400 \end{aligned}$$

6.6 Prove Theorem 6.5: $D_n = n![1 - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} + \cdots + (-1)^n \frac{1}{n!}]$

Recall (Section 3.3) that S_n denotes the set of permutations on $X = \{1, 2, \ldots, n\}$ and $|S_n| = n!$. For $i = 1, \ldots, n$, let F_i denote all permutations in S_n which "fix i," that is, $F_i = \{\sigma \in S_n \mid \sigma(i) = i\}$. Then, for distinct subscripts,

$$|F_r| = (n-1)!, \quad |F_i \cap F_i| = (n-2)!, \ldots |F_{i_1} \cap F_{i_2} \cap \ldots \cap F_{i_r}| = (n-r)!$$

Let Y denote the set of all derangements in S_n. Then

$$D_n = |Y| = |F_1^C \cap F_2^C \cap \cdots \cap F_n^C|$$

By the Inclusion–Exclusion principle,

$$D_n = |S_n| - s_1 + s_2 - s_3 + \cdots + (-1)^n s_n$$

where

$$s_r = \sum_{i_1 < i_2 < \ldots < i_n} |F_{i_1} \cap F_{i_2} \cap \ldots \cap F_{i_r}| = C(n, r)\,(n-r)! = \frac{n!}{r!}$$

Setting $|S_n| = n!$ and $s_r = n!/r!$ in the formula for D_n gives us our theorem.

PIGEONHOLE PRINCIPLE

6.7. Suppose five points are chosen from the interior of a square S where each side has length two inches. Show that the distance between two of the points must be less than $\sqrt{2}$ inches.

Draw two lines between the opposite sides of S which partitions S into four subsquares each whose sides have length one inch. By the Pigeonhole Principle, two of the points lie in one of the subsquares. The diagonal of each subsquare is $\sqrt{2}$ inches, so the distance between the two points is less than $\sqrt{2}$ inches.

6.8. Let p and q be positive integers. A number r is said to satisfy the (p, q)-Ramsey property if a set of r people must have a subset of p mutual friends or a subset of q mutual strangers. The Ramsey number $R(p, q)$ is the smallest such integer r. Show that R(3, 3) = 6.

By Example 6.5, $R(3, 3) \geq 6$. We show that $R(3, 3) > 5$. Consider five people who are sitting around a circular table, and suppose each person is only friends with the people sitting next to him/her. No three people can be strangers since two of the three people must be sitting next to each other. Also no three people can be mutual friends since they cannot be sitting next to each other. Thus $R(3, 3) > 5$. Accordingly $R(3, 3) = 6$.

6.9. Suppose a team X plays 18 games in a two-week 14-day period, and plays at least one game a day. Show that there is a period of days in which exactly 9 games were played.

Let $S = \{s_1, s_2, \ldots, s_{14}\}$ where s_i is the number of games X played from the first day to the ith day. Then $s_{14} = 18$, and all the s_i are distinct. Let $T = \{t_1, t_2, \ldots, t_{14}\}$ where $t_i = s_i + 9$. Then $t_{14} = 18 + 9 = 27$, and the t_i are distinct. Together S and T have $14 + 14 = 28$ numbers, which lie between 1 and 27. By the Pigeonhole Principle, two of the numbers must be equal. However the entries in S and the entries in T are distinct. Thus there is $s_j \in S$ and $t_n \in T$ such that $s_j = t_k = s_k + 9$. Therefore,

$$9 = s_j - s_n = \text{number of games played in days } k+1, k+2, \ldots, j-1, j$$

6.10. Prove Theorem 6.7: Every sequence of distinct $n^2 + 1$ real numbers contains a subsequence of length $n + 1$ which is strictly increasing or strictly decreasing.

Let $a_1, a_2, \ldots, a_{n^2+i}$ be a sequence of $n^2 + 1$ distinct real numbers. To each a_t we associate the pair (i_t, d_t) where: (1) i_t is the longest increasing subsequence beginning at a_t and (2) d_t is the longest decreasing subsequence beginning at a_t Thus there are $n^2 + 1$ such ordered pairs, one for each number in the sequence.

Now suppose that no subsequence is longer than n. Then i_t and d_t cannot exceed n. Thus there are at most n^2 distinct pairs (i_t, d_t). By the Pigeonhole Principle, two of the $n^2 + 1$ pairs are equal, that is, there are two distinct points a_r and a_s such that $(i_r, d_r) = (i_s, d_s)$. WLOG, we can assume $r < s$. Then a_r occurs before a_s in the sequence (See Fig. 6-2(a)). Then a_r followed by the increasing subsequence of i_s numbers beginning at a_s gives a subsequence of length $i_s + 1 = i_r + 1$ beginning a_r (See Fig. 6-2(b)). This contradicts the definition of i_r. Similarly, suppose $a_r > a_s$. Then a_n followed by the decreasing subsequence of d_s numbers beginning at a_s gives a subsequence of length $d_r + 1 = d_s + 1$ beginning at a_r which contradicts the definition of d_r (See Fig. 6-2(c)). In each case we get a contradiction. Thus the assumption that no subsequence exceeds n is not true, and the theorem is proved.

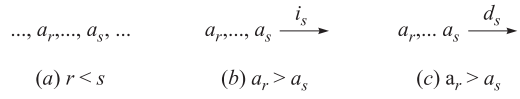


Fig. 6-2

RECURSION

6.11. Consider the second-order homogeneous recurrence relation $a_n = a_{n-1} + 2a_{n-2}$ with initial conditions $a_0 = 2, a_1 = 7$,

(*a*) Find the next three terms of the sequence.

(*b*) Find the general solution.

(*c*) Find the unique solution with the given initial conditions.

(*a*) Each term is the sum of the preceding term plus twice its second preceding term. Thus:

$$a_2 = 7 + 2(2) = 11, \quad a_3 = 11 + 2(7) = 25, \quad a_4 = 25 + 2(11) = 46$$

(*b*) First we find the characteristic polynomial $\Delta(t)$ and its roots:

$$\Delta(x) = x^2 - x - 2 = (x - 2)(x + 1); \quad \text{roots } r_1 = 2, r_2 = -1$$

Since the roots are distinct, we use Theorem 6.8 to obtain the general solution:

$$a_n = c_1(2^n) + c_2(-1)^n$$

(*c*) The unique solution is obtained by finding c_1 and c_2 using the initial conditions:

$$\begin{array}{ll} \text{For } n = 0, a_0 = 2, \text{ we get:} & c_1(2^0) + c_2(-1)^0 = 2 \quad \text{or} \quad c_1 + c_2 = 2 \\ \text{For } n = 1, a_1 = 7, \text{ we get:} & c_1(2^1) + c_2(-1)^1 = 7 \quad \text{or} \quad 2c_1 - c_2 = 7 \end{array}$$

Solving the two equations for c_1 and c_2 yields $c_1 = 3$ and $c_2 = 1$. The unique solution follows:

$$a_n = 3(2^n) - (-1)^n$$

6.12. Consider the third-order homogeneous recurrence relation $a_n = 6a_{n-1} - 12a_{n-2} + 8a_{n-3}$

(*a*) Find the general solution.

(*b*) Find the solution with initial conditions $a_0 = 3, a_1 = 4, a_2 = 12$.

(*a*) First we find the characteristic polynomial

$$\Delta(x) = x^3 - 6x^2 + 12x - 8 = (x - 2)^3$$

Then $\Delta(x)$ has only one root $r_0 = 2$ which has multiplicity 3. Thus the general solution of the recurrence relation follows:

$$a_n = c_1(2^n) + c_2n(2^n) + c_3n^2(2^n) = (c_1 + c_2n + c_3n^2)(2^n)$$

(*b*) We find the values for c_1, c_2, and c_3 as follows:

$$\begin{array}{lll} \text{For } n = 0, a_0 = 3 & \text{we get:} & c_1 = 3 \\ \text{For } n = 1, a_1 = 4 & \text{we get:} & 2c_1 + 2c_2 + 2c_3 = 4 \\ \text{For } n = 2, a_2 = 12 & \text{we get:} & 4c_1 + 8c_2 + 16c_3 = 12 \end{array}$$

Solving the system of three equations in c_1, c_2, c_3 yields the solution

$$c_1 = 3, \quad c_2 = -2, \quad c_3 = 1$$

Thus the unique solution of the recurrence relation follows:

$$a_n = (3 - 2n + n^2)(2^n)$$

Supplementary Problems

ADVANCED COUNTING TECHNIQUES, INCLUSION–EXCLUSION

6.13. A store sells $M = 4$ kinds of cookies. Find the number of ways a customer can buy:
(a) 10 cookies; (b) 15 cookies.

6.14. Find the number m of nonnegative solutions to $x + y + z = 20$ with the conditions that $x \geq 5$, $y \geq 3$, and $z \geq 1$.

6.15. Let E be the equation $x + y + z = 20$. Find the number m of nonnegative solutions to E with the conditions that $x < 8$, $y < 9$, $z < 10$.

6.16. Find the number m of positive integers not exceeding 1000 which are not divisible by 3, 7, or 11.

6.17. Find the number of ways that 14 people can be partitioned into 6 committees, such that 2 committees contain 3 people and the other committees contain 2 people.

6.18. Assume that a cell can be empty. Find the number m of ways that a set:
(a) With 3 people can be partitioned into: (i) three ordered cells; (ii) three unordered cells.
(b) With 4 people can be partitioned into: (i) three ordered cells; (ii) three unordered cells.

6.19. Find the number N of surjective (onto) functions from a set A to a set B where:

(a) $|A| = 8, |B| = 3$; (b) $|A| = 6, |B| = 4$; (c) $|A| = 5, |B| = 5$; (d) $|A| = 5, |B| = 7$.

6.20. Find the number of derangements of $X = \{1, 2, 3, \ldots, 2m\}$ such that the first m elements of each derangement are: (a) the first m elements of X; (b) the last m elements of X.

PIGEONHOLE PRINCIPLE

6.21. Find the minimum number of students that can be admitted to a college so that there are at least 15 students from one of the 50 states.

6.22. Consider nine lattice points in space. Show that the midpoint of two of the points is also a lattice point.

6.23. Find an increasing subsequence of maximum length and a decreasing subsequence of maximum length in the sequence: 14, 2, 8, 3, 25, 15, 10, 20, 9, 4.

6.24. Consider a line of 50 people with distinct heights. Show there is a subline of 8 people which is either increasing or decreasing.

6.25. Give an example of a sequence of 25 distinct integers which does not have a subsequence of 6 integers which is either increasing or decreasing.

6.26. Suppose a team X plays 19 games in a two-week period of 14 days, and plays at least one game per day. Show there is a period of consecutive days that X played exactly 8 games.

6.27. Suppose 10 points are chosen at random in the interior of an equilateral triangle T where each side has length three inches. Show that the distance between two of the points must be less than one inch.

6.28. Let $X = \{x_i\}$ be a set of n positive integers. Show that the sum of the integers of a subset of X is divisible by n.

6.29. Consider a group of 10 people (where each pair are either friends or strangers). Show that there is either a subgroup of 4 mutual friends or a subgroup of 3 mutual strangers.

6.30. For the Ramsey numbers $R(p, q)$ show that: (a) $R(p, q) = R(q, p)$; (b) $R(p, 1) = 1$; (c) $R(p, 2) = p$.

RECURSION

6.31. For each recurrence relation and initial conditions, find: (i) general solution; (ii) unique solution with the given initial conditions:

(a) $a_n = 3a_{n-1} + 10a_{n-2}; a_0 = 5, a_1 = 11$
(b) $a_n = 4a_{n-1} + 21a_{n-2}; a_0 = 9, a_1 = 13$
(c) $a_n = 3a_{n-1} - 2a_{n-2}; a_0 = 5, a_1 = 8$
(d) $a_n = 5a_{n-1} - 6a_{n-2}; a_0 = 2, a_1 = 8$
(e) $a_n = 3a_{n-1} - a_{n-2}; a_0 = 0, a_1 = 1$
(f) $a_n = 5a_{n-1} - 3a_{n-2}; a_0 = 0, a_1 = 1$

6.32. Repeat Problem 6.31 for the following recurrence relations and initial conditions:

(a) $a_n = 6a_{n-1}; a_0 = 5$
(b) $a_n = 7a_{n-1}; a_0 = 5$
(c) $a_n = 4a_{n-1} - 4a_{n-2}; a_0 = 1, a_1 = 8$
(d) $a_n = 10a_{n-1} - 25a_{n-2}; a_0 = 2, a_1 = 15$

6.33. Find the unique solution to each recurrence relation with the given initial conditions:

(a) $a_n = 10a_{n-1} - 32a_{n-2} + 32a_{n-3}$ with $a_0 = 5, a_1 = 18, a_2 = 76$
(b) $a_n = 9a_{n-1} - 27a_{n-2} + 27a_{n-3}$ with $a_0 = 5, a_1 = 24, a_2 = 117$

6.34. Consider the following second-order recurrence relation and its characteristic polynomial $\Delta(x)$:

$$a_n = sa_{n-1} + ta_{n-2} \quad \text{and} \quad \Delta(x) = x^2 - sx - t \tag{$*$}$$

(a) Suppose $p(n)$ and $q(n)$ are solutions of $(*)$. Show that, for any constants c_1 and c_2, $c_1 p(n) + c_2 q(n)$ is also a solution of $(*)$.

(b) Suppose r is a root of $\Delta(x)$. Show that $a_n = r^n$ is a solution to $(*)$.

(c) Suppose r is a double root of $\Delta(x)$. Show that: (i) $s = 2r$ and $t = -r^2$; (ii) $a_n = nr^n$ is also a root of $(*)$.

6.35. Repeat Problem 6.34(a) and (b) for any linear kth-order homogeneous recurrence relation with constant coefficients and its characteristic polynomial $\Delta(x)$ which have the form:

$$a_n = C_1 a_{n-1} + C_2 a_{n-2} + \cdots + C_k a_{n-k} = \sum_{i=1}^{k} C_1 a_{n-1} \quad \text{and} \quad \Delta(x) = x^k - \sum_{i=1}^{k} C_1 x^{k-i}$$

Answers to Supplementary Problems

6.13. (a) 286; (b) 646.

6.14. 78.

6.15. 15.

6.16. 520.

6.17. $(14!)/[(3!3!2!2!2!2!)(2!4!)] = 3\,153\,150$.

6.18. (a) (i) $3^3 = 27$; (ii) They may be distributed as: [3, 0, 0], [2, 1, 0], or [1, 1, 1]. Hence $m = 1 + 3 + 1 = 5$.
(b) (i) $3^4 = 81$; (ii) They may be distributed as: [4, 0, 0], [3, 1, 0], [2, 2, 0] or [2, 1, 1]. Hence $m = 1+4+3+6 = 14$.

6.19. (a) 5796; (b) 1560; (c) $5! = 120$; (d) 0.

6.20. (a) $(D_m)^2$; (b) $(m!)^2$.

6.21. 701.

6.22. There are eight triplets of parities: (odd, odd, odd), (odd, odd, even), …. Thus 2 of the 9 points have the same triplet of parities.

6.23. 2, 3, 10, 20; 25, 15, 10, 8, 4.

6.24. Use Theorem 6.7 with $n = 9$.

6.25. 5, 4, 3, 2, 1, 10, 9, 8, 7, 6, …, 25, 24, 23, 22, 21.

6.26. (Hint: See Problem 6.9.)

6.27. (Hint: Partition T into 9 equilateral triangles where each side has length one inch.)

6.28. Let $s_i = x_1 + \cdots + x_i$. The result is true if n divides some s_i. Otherwise, let r' be the remainder when s_i is divided by n. Two of the rs must be equal. Say $r_p = r_q$ where $p < q$. Then n divides $s_q - s_p = x_{p+1} + \cdots + x_q$.

6.31. (a) $a_n = c_1(5^n) + c_2(-2)^n$; $c_1 = 3, c_2 = 2$
(b) $a_n = c_1(7^n) + c_2(-3)^n$; $c_1 = 4, c_2 = 5$
(c) $a_n = c_1 + c_2(2^n)$; $c_1 = 2, c_2 = 3$
(d) $a_n = c_1(2^n) + c_2(3^n)$; $c_1 = -2, c_2 = 4$
(e) $a_n = c_1[(3+t)/2]^n + c_2[(3-t)/2]^n$; $c_1 = 1/t$, $c_2 = -1/t$ where $t = \sqrt{5}$
(f) $a_n = c_1[(5+s)/2]^n + c_2[(5-s)/2]^n$; $c_1 = 1/s$, $c_2 = -1/s$ where $= \sqrt{13}$

6.32. (a) $a_n = c_1(6^n)$, $c_1 = 5$
(b) $a_n = c_1(7^n)$, $c_1 = 5$
(c) $a_n = c_1(2^n) + c_2 n(2^n)$, $c_1 = 1, c_2 = 3$
(d) $a_n = c_1(5^n) + c_2 n(5^n)$, $c_1 = 2, c_2 = 1$

6.33. (a) $a_n = 2(4^n) + n(4^n) + 3(2^n)$; (b) $a_n = 5(3^n) + 2n(3^n) + n^2(3^n) = (5 + 2n + n^2)3^n$.

6.34. (b) r is a root of $\Delta(x)$ so $r^2 - sr - t = 0$ or $r^2 = sr + t$. Let $a_n = r^n$. Then $sa_{n-1} + ta_{n-2} = sr^{n-1} + tr^{n-2} = (sr + t)r^{n-2} = r^2(r^{n-2}) = r^n = a_n$

(c) (i) r is a double root of $\Delta(x)$; hence $\Delta(x) = (x - r)^2 = x^2 - 2rx + r^2 = x^2 - sx - t$. Thus $s = 2r$ and $t = -r^2$. (ii) Let a_n nr_n. Then $sa_{n-1} + ta_{n-2} = nr^n = an$.

APPENDIX B

Algebraic Systems

B.1 INTRODUCTION

This Appendix investigates some of the major algebraic systems in mathematics: semigroups, groups, rings, and fields. We also define the notion of a homomorphism and the notion of a quotient structure. We begin with the formal definition of an operation, and discuss various types of operations.

B.2 OPERATIONS

The reader is familiar with the operations of addition and multiplication of numbers, union and intersection of sets, and the composition of functions. These operations are denoted as follows:

$$a + b = c, \quad a \cdot b = c, \quad A \cup B = C, \quad A \cap B = C, \quad g \circ f = h.$$

In each situation, an element (c, C, or h) is assigned to an original pair of elements. We make this notion precise.

Definition B.1: Let S be a nonempty set. An *operation* on S is a function $*$ from $S \times S$ into S. In such a case, instead of $*(a, b)$, we usually write

$$a * b \quad \text{or sometimes} \quad ab$$

The set S and an operation $*$ on S is denoted by $(S, *)$ or simply S when the operation is understood.

Remark: An operation $*$ from $S \times S$ into S is sometimes called a *binary operation*. A *unary* operation is a function from S into S. For example, the absolute value $|n|$ of an integer n is a unary operation on $\mathbf{Z}$, and the complement A^C of a set A is a unary operation on the power set $P(X)$ of a set X. A *ternary* (3-ary) operation is a function from $S \times S \times S$ into S. More generally, an n-ary operation is a function from $S \times S \times \cdots \times S$ (n factors) into S. Unless otherwise stated, the word operation shall mean binary operation. We will also assume that our underlying set S is nonempty.

Suppose S is a finite set. Then an operation $*$ on S can be presented by its operation (multiplication) table where the entry in the row labeled a and the column labeled b is $a * b$.

Suppose S is a set with an operation $*$, and suppose A is a subset of S. Then A is said to be *closed under* $*$ if $a * b$ belongs to A for any elements a and b in A.

EXAMPLE B.1 Consider the set **N** of positive integers.

(a) Addition ($+$) and multiplication ($\times$) are operations on **N**. However, subtraction ($-$) and division ($/$) are not operations on **N** since the difference and the quotient of positive integers need not be positive integers. For example, $2 - 9$, and $7/3$ are not positive integers.

(b) Let A and B denote, respectively, the set of even and odd positive integers. Then A is closed under addition and multiplication since the sum and product of any even numbers are even. On the other hand, B is closed under multiplication but not addition since, for example, $3 + 5 = 8$ is even.

EXAMPLE B.2 Let $S = \{a, b, c, d\}$. The tables in Fig. B-1 define operations $*$ and $\cdot$ on S. Note that $*$ can be defined by the following operation where x and y are any elements of S:

$$x * y = x$$

$*$	a	b	c	d
a	a	a	a	a
b	b	b	b	b
c	c	c	c	c
d	d	d	d	d

(a)

$\cdot$	a	b	c	d
a	a	b	c	d
b	b	a	a	b
c	c	b	a	a
d	d	a	a	a

(b)

Fig. B-1

Next we list a number of important properties of our operations.

Associative Law:

An operation $*$ on a set S is said to be *associative* or to satisfy the *Associative Law* if, for any elements a, b, c in S, we have

$$(a * b) * c = a * (b * c)$$

Generally speaking, if an operation is not associative, then there may be many ways to form a product. For example, the following shows five ways to form the product *abcd*:

$$((ab)c)d, \quad (ab)(cd), \quad (a(bc))d, \quad a((bc)d), \quad a(b(cd))$$

If the operation is associative, then the following theorem (proved in Problem B.4) applies.

Theorem B.1: Suppose $*$ is an associative operation on a set S. Then any product $a_1 * a_2 * \cdots * a_n$ requires no parentheses, that is, all possible products are equal.

Commutative Law:

An operation $*$ on a set S is said to be *commutative* or satisfy the *Commutative Law* if, for any elements a, b in S,

$$a * b = b * a$$

EXAMPLE B.3

(a) Consider the set **Z** of integers. Addition and multiplication of integers are associative and commutative. On the other hand, subtraction is nonassociative. For example,

$$(8 - 4) - 3 = 1 \quad \text{but} \quad 8 - (4 - 3) = 7$$

Moreover, subtraction is not commutative since, for example, $3 - 7 \neq 7 - 3$.

(b) Consider the operation of matrix multiplication on the set M of n-square matrices. One can prove that matrix multiplication is associative. On the other hand, matrix multiplication is not commutative. For example,

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}\begin{bmatrix} 5 & 6 \\ 0 & -2 \end{bmatrix} = \begin{bmatrix} 5 & 2 \\ 15 & 10 \end{bmatrix} \quad \text{but} \quad \begin{bmatrix} 5 & 6 \\ 0 & -2 \end{bmatrix}\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} = \begin{bmatrix} 23 & 34 \\ -6 & -8 \end{bmatrix}$$

Identity Element:

Consider an operation $*$ on a set S. An element e in S is called an *identity* element for $*$ if, for any element a in S,

$$a * e = e * a = a$$

More generally, an element e is called a *left identity* or a *right identity* according as $e * a = a$ or $a * e = a$ where a is any element in S. The following theorem applies.

Theorem B.2: Suppose e is a left identity and f is a right identity for an operation on a set S. Then $e = f$

The proof is very simple. Since e is a left identity, $ef = f$; but since f is a right identity, $ef = e$. Thus $e = f$. This theorem tells us, in particular, that an identity element is unique, and that if an operation has more than one left identity then it has no right identity, and vice versa.

Inverses:

Suppose an operation $*$ on a set S does have an identity element e. The *inverse* of an element a in S is an element b such that

$$a * b = b * a = e$$

If the operation is associative, then the inverse of a, if it exists, is unique (Problem B.2). Observe that if b is the inverse of a, then a is the inverse of b. Thus the inverse is a symmetric relation, and we can say that the elements a and b are inverses.

Notation: If the operation on S is denoted by $a * b$, $a \times b$, $a \cdot b$, or ab, then S is said to be written *multiplicatively* and the inverse of an element $a \in S$ is usually denoted by a^{-1}. Sometimes, when S is commutative, the operation is denoted by $+$ and then S is said to be written *additively*. In such a case, the identity element is usually denoted by 0 and it is called the *zero* element; and the inverse is denoted by $-a$ and it is called the *negative* of a.

EXAMPLE B.4 Consider the rational numbers $\mathbf{Q}$. Under addition, 0 is the identity element, and -3 and 3 are (additive) inverses since

$$(-3) + 3 = 3 + (-3) = 0$$

On the other hand, under multiplication, 1 is the identity element, and -3 and $-1/3$ are (multiplicative) inverses since

$$(-3)(-1/3) = (-1/3)(-3) = 1$$

Note 0 has no multiplicative inverse.

Cancellation Laws:

An operation $*$ on a set S is said to satisfy the *left cancellation law* or the *right cancellation law* according as:

$$a * b = a * c \text{ implies } b = c \quad \text{or} \quad b * a = c * a \text{ implies } b = c$$

Addition and subtraction of integers in $\mathbf{Z}$ and multiplication of nonzero integers in $\mathbf{Z}$ do satisfy both the left and right cancellation laws. On the other hand, matrix multiplication does not satisfy the cancellation laws. For example, suppose

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

Then $AB = AC = D$, but $B \neq C$.

B.3 SEMIGROUPS

Let S be a nonempty set with an operation. Then S is called a *semigroup* if the operation is associative. If the operation also has an identity element, then S is called a *monoid*.

EXAMPLE B.5

(a) Consider the positive integers $\mathbf{N}$. Then $(\mathbf{N}, +)$ and $(\mathbf{N}, \times)$ are semigroups since addition and multiplication on $\mathbf{N}$ are associative. In particular, $(\mathbf{N}, \times)$ is a monoid since it has the identity element 1. However, $(\mathbf{N}, +)$ is not a monoid since addition in $\mathbf{N}$ has no zero element.

(b) Let S be a finite set, and let $F(S)$ be the collection of all functions $f\colon S \to S$ under the operation of composition of functions. Since the composition of functions is associative, $F(S)$ is a semigroup. In fact, $F(S)$ is a monoid since the identity function is an identity element for $F(S)$.

(c) Let $S = \{a, b, c, d\}$. The multiplication tables in Fig. B-1 define operations $*$ and $\cdot$ on S. Note that $*$ can be defined by the formula $x * y = x$ for any x and y in S. Hence

$$(x * y) * z = x * z = x \quad \text{and} \quad x * (y * z) = x * y = x$$

Therefore, $*$ is associative and hence $(S, *)$ is a semigroup. On the other hand, $\cdot$ is not associative since, for example,

$$(b \cdot c) \cdot c = a \cdot c = c \quad \text{but} \quad b \cdot (c \cdot c) = b \cdot a = b$$

Thus $(S, \cdot)$ is not a semigroup.

Free Semigroup, Free Monoid

Let A be a nonempty set. A *word* w on A is a finite sequence of its elements. For example, the following are words on $A = \{a, b, c\}$:

$$u = ababbbb = abab^4 \quad \text{and} \quad v = baccaaaa = bac^2a^4$$

(We write a^2 for aa, a^3 for aaa, and so on.) The *length* of a word w, denoted by $l(w)$, is the number of elements in w. Thus $l(u) = 7$ and $l(v) = 8$.

The concatenation of words u and v on a set A, written $u * v$ or uv, is the word obtained by writing down the elements of u followed by the elements of v. For example,

$$uv = (abab^4)(bac^2a^4) = abab^5c^2a^4$$

Now let $F = F(A)$ denote the collection of all words on A under the operation of concatenation. Clearly, for any words u, v, w, the words $(uv)w$ and $u(vw)$ are identical; they simply consist of the elements of u, v, w written down one after the other. Thus F is a semigroup; it is called the *free semigroup* on A, and the elements of A are called the *generators* of F.

The empty sequence, denoted by λ, is also considered as a word on A. However, we do not assume that λ belongs to the free semigroup $F = F(A)$. The set of all words on A including λ is frequently denoted by A^*. Thus A^* is a monoid under concatenation; it is called the *free monoid* on A.

Subsemigroups

Let A be a nonempty subset of a semigroup S. Then A is called a subsemigroup of S if A itself is a semigroup with respect to the operation on S. Since the elements of A are also elements of S, the Associative Law automatically holds for the elements of A. Therefore, A is a subsemigroup of S if and only if A is closed under the operation on S.

EXAMPLE B.6

(a) Let A and B denote, respectively, the set of even and odd positive integers. Then $(A, \times)$ and $(B, \times)$ are subsemigroups of $(\mathbf{N}, \times)$ since A and B are closed under multiplication. On the other hand, $(A, +)$ is a subsemigroup of $(\mathbf{N}, +)$ since A is closed under addition, but $(B, +)$ is not a subsemigroup of $(\mathbf{N}, +)$ since B is not closed under addition.

(b) Let F be the free semigroup on the set $A = \{a, b\}$. Let H consist of all even words, that is, words with even length. The concatenation of two such words is also even. Thus H is a subsemigroup of F.

Congruence Relations and Quotient Structures

Let S be a semigroup and let $\sim$ be an equivalence relation on S. Recall that the equivalence relation $\sim$ induces a partition of S into equivalence classes. Also, $[a]$ denotes the equivalence class containing the element $a \in S$, and that the collection of equivalence classes is denoted by $S/\sim$.

Suppose that the equivalence relation $\sim$ on S has the following property:

$$\boxed{\text{If } a \sim a' \text{ and } b \sim b', \text{ then } ab \sim a'b'.}$$

Then $\sim$ is called a *congruence relation* on S. Furthermore, we can now define an operation on the equivalence classes by

$$[a] * [b] = [a * b] \quad \text{or, simply,} \quad [a]\,[b] = [ab]$$

Furthermore, this operation on $S/\sim$ is associative; hence $S/\sim$ is a semigroup. We state this result formally.

Theorem B.3: Let $\sim$ be a congruence relation on a semigroup S. Then $S/\sim$, the equivalence classes under $\sim$, form a semigroup under the operation $[a]\,[b] = [ab]$.

This semigroup $S/\sim$ is called the quotient of S by $\sim$.

EXAMPLE B.7

(a) Let F be the free semigroup on a set A. Define $u \sim u'$ if u and u' have the same length. Then $\sim$ is an equivalence relation on F. Furthermore, suppose $u \sim u'$ and $v \sim v'$, say,

$$l(u) = l(u') = m \quad \text{and} \quad l(v) = l(v') = n$$

Then $l(uv) = l(u'v') = m + n$, and so $uv \sim u'v'$. Thus $\sim$ is a congruence relation on F.

(b) Consider the integers $\mathbf{Z}$ and a positive integer $m > 1$. Recall (Section 11.8) that we say that a is congruent to b modulo m, written

$$a \equiv b \pmod{m}$$

if m divides the difference $a - b$. Theorem 11.21 states that this relation is an equivalence relation on $\mathbf{Z}$. Furthermore, Theorem 11.22 tells us that if $a \equiv c \pmod{m}$ and $b \equiv d \pmod{m}$ then:

$$a + b \equiv c + d \pmod{m} \quad \text{and} \quad ab \equiv cd \pmod{m}$$

In other words, this relation is a congruence relation on $\mathbf{Z}$.

Homomorphism of Semigroups

Consider two semigroups $(S, *)$ and $(S', *')$. A function $f\colon S \to S'$ is called a *semigroup homomorphism* or, simply, a *homomorphism* if

$$f(a * b) = f(a) *' f(b) \quad \text{or, simply} \quad f(ab) = f(a)f(b)$$

Suppose f is also one-to-one and onto. Then f is called an *isomorphism* between S and S', and S and S′ are said to be *isomorphic* semigroups, written $S \cong S$.

EXAMPLE B.8

(a) Let M be the set of all 2×2 matrices with integer entries. The determinant of any matrix $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is denoted and defined by $\det(A) = |A| = ad - bc$. One proves in Linear Algebra that the determinant is a *multiplicative function*, that is, for any matrices A and B,

$$\det(AB) = \det(A) \cdot \det(B)$$

Thus the determinant function is a semigroup homomorphism on $(M, \times)$, the matrices under matrix multiplication. On the other hand, the determinant function is not additive, that is, for some matrices,

$$\det(A + B) \neq \det(A) + \det(B)$$

Thus the determinant function is not a semigroup homomorphism on $(M, +)$.

(b) Figure B-2(a) gives the addition table for $\mathbf{Z_4}$, the integers modulo 4 under addition; and Fig. B-2(b) gives the multiplication table for $S = \{1, 3, 7, 9\}$ in $\mathbf{Z_{10}}$. (We note that S is a reduced residue system for the integers $\mathbf{Z}$ modulo 10.) Let $f\colon \mathbf{Z_4} \to S$ be defined by

$$f(0) = 1, \quad f(1) = 3, \quad f(2) = 9, \quad f(3) = 7$$

+	0	1	2	3
0	0	1	2	3
1	1	2	3	0
2	2	3	0	1
3	3	0	1	2

(a)

×	1	3	7	9
1	1	3	7	9
3	3	9	1	7
7	7	1	9	3
9	9	7	3	1

(b)

Fig. B-2

One can show that f is a homomorphism. Since f is also one-to-one and onto, f is an isomorphism. Thus $\mathbf{Z_4}$ and S are isomorphic semigroups.

(c) Let $\sim$ be a congruence relation on a semigroup S. Let $\phi\colon S \to S/\sim$ be the *natural mapping* from S into the factor semigroup $S/\sim$ defined by

$$\phi(a) = [a]$$

That is, each element a in S is assigned its equivalence class $[a]$. Then ϕ is a homomorphism since

$$\phi(ab) = [ab] = [a]\,[b] = \phi(a)\phi(b)$$

Fundamental Theorem of Semigroup Homomorphisms

Recall that the image of a function $f\colon S \to S'$, written $f(S)$ of Im f, consists of the images of the elements of S under f. Namely:

$$\text{Im } f = \{b \in S' \mid \text{there exists } a \in S \text{ for which } f(a) = b\}$$

The following theorem (proved in Problem B.5) is fundamental to semigroup theory.

Theorem B.4: Let $f : S \to S'$ be a semigroup homomorphism. Let $a \sim b$ if $f(a) = f(b)$. Then: (i) $\sim$ is a congruence relation on S. (ii) $S/\sim$ is isomorphic to $f(S)$.

EXAMPLE B.9

(a) Let F be the free semigroup on $A = \{a, b\}$. The function $f\colon F \to \mathbf{Z}$ defined by

$$f(u) = l(u)$$

is a homomorphism. Note $f(F) = \mathbf{N}$. Thus $F/\sim$ is isomorphic to $\mathbf{N}$.

(b) Let M be the set of 2×2 matrices with integer entries. Consider the determinant function det: $M \to \mathbf{Z}$. We note that the image of det is $\mathbf{Z}$. By Theorem B.4, $M/\sim$ is isomorphic to $\mathbf{Z}$.

Semigroup Products

Let $(S_1, *_1)$ and $(S_2, *_2)$ be semigroups. We form a new semigroup $S = S_1 \otimes S_2$, called the direct product of S_1 and S_2, as follows.

(1) The elements of S come from $S_1 \times S_2$, that is, are ordered pairs (a, b) where $a \in S_1$ and $b \in S_2$

(2) The operation $*$ in S is defined componentwise, that is,

$$(a, b) * (a', b') = (a *_1 a',\ b *_2 b') \quad \text{or simply} \quad (a, b)(a', b') = (aa', bb')$$

One can easily show (Problem B.3) that the above operation is associative.

B.4 GROUPS

Let G be a nonempty set with a binary operation (denoted by juxtaposition). Then G is called a *group* if the following axioms hold:

[$\mathbf{G_1}$] Associative Law: For any a, b, c in G, we have $(ab)c = a(bc)$.
[$\mathbf{G_2}$] Identity element: There exists an element e in G such that $ae = ea = a$ for every a in G.
[$\mathbf{G_3}$] Inverses: For each a in G, there exists an element a^{-1} in G (the *inverse* of a) such that

$$aa^{-1} = a^{-1}a = e$$

A group G is said to be *abelian* (or *commutative*) if $ab = ba$ for every $a, b \in G$, that is, if G satisfies the Commutative Law.

When the binary operation is denoted by juxtaposition as above, the group G is said to be written *multiplicatively*. Sometimes, when G is abelian, the binary operation is denoted by $+$ and G is said to be written *additively*. In such a case the identity element is denoted by 0 and it is called the *zero* element; and the inverse is denoted by $-a$ and it is called the *negative* of a.

The number of elements in a group G, denoted by $|G|$, is called the *order* of G. In particular, G is called a *finite group* if its order is finite.

Suppose A and B are subsets of a group G. Then we write:

$$AB = \{ab \mid a \in A,\ b \in B\} \quad \text{or} \quad A + B = \{a + b \mid a \in A,\ \ b \in B\}$$

EXAMPLE B.10

(a) The nonzero rational numbers $\mathbf{Q}\backslash\{0\}$ form an abelian group under multiplication. The number 1 is the identity element and q/p is the multiplicative inverse of the rational number p/q.

(b) Let S be the set of 2×2 matrices with rational entries under the operation of matrix multiplication. Then S is not a group since inverses do not always exist. However, let G be the subset of 2×2 matrices with a nonzero determinant. Then G is a group under matrix multiplication. The identity element is

$$I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \text{ and the inverse of } A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \text{ is } A^{-1} = \begin{bmatrix} d/|A| & -b/|A| \\ -c/|A| & a/|A| \end{bmatrix}$$

This is an example of a nonabelian group since matrix multiplication is noncommutative.

(c) Recall that $\mathbf{Z}_m$ denotes the integers modulo m. $\mathbf{Z}_m$ is a group under addition, but it is not a group under multiplication. However, let $\mathbf{U}_m$ denote a reduced residue system modulo m which consists of those integers relatively prime to m. Then $\mathbf{U}_m$ is a group under multiplication (modulo m). Figure B-3 gives the multiplication table for $\mathbf{U}_{12} = \{1, 5, 7, 11\}$.

$\times$	1	5	7	11
1	1	5	7	11
5	5	1	11	7
7	7	11	1	5
11	11	7	5	1

Fig. B-3

	ε	σ_1	σ_2	σ_3	ϕ_1	ϕ_2
ε	ε	ϕ_1	σ_3	σ_3	ϕ_1	ϕ_2
σ_1	σ_1	ε	ϕ_1	ϕ_2	σ_2	σ_3
σ_2	σ_2	ϕ_2	ε	ϕ_1	σ_3	σ_1
σ_3	σ_3	ϕ_1	ϕ_2	ε	σ_1	σ_2
ϕ_1	ϕ_1	σ_3	σ_1	σ_2	ϕ_2	ε
ϕ_2	ϕ_2	σ_2	σ_3	σ_1	ε	ϕ_1

Fig. B-4

Symmetric Group S_n

A one-to-one mapping σ of the set $\{1, 2, \ldots, n\}$ onto itself is called a *permutation*. Such a permutation may be denoted as follows where $j_i = \sigma(i)$:

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & \cdots & n \\ j_1 & j_2 & j_3 & \cdots & j_n \end{pmatrix}$$

The set of all such permutations is denoted by S_n, and there are $n! = n(n-1) \cdot \ldots \cdot 2 \cdot 1$ of them. The composition and inverses of permutations in S_n belong to S_n, and the identity function ε belongs to S_n. Thus S_n forms a group under composition of functions called the *symmetric group of degree n*.

The symmetric group S_3 has $3! = 6$ elements as follows:

$$\varepsilon = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix}, \quad \phi_1 = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}$$

$$\sigma_1 = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix}, \quad \sigma_3 = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix}, \quad \phi_2 = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix}$$

The multiplication table of S_3 appears in Fig. B-4.

MAP(A), PERM(A), and AUT(A)

Let A be a nonempty set. The collection MAP(A) of all functions (mappings) $f: A \to A$ is a semigroup under composition of functions; it is not a group since some functions may have no inverses. However, the subsemigroup PERM(A) of all one-to-one correspondences of A with itself (called *permutations* of A) is a group under composition of functions.

Furthermore, suppose A contains some type of geometric or algebraic structure; for example, A may be the set of vertices of a graph, or A may be an ordered set or a semigroup. Then the set AUT(A) of all isomorphisms of A with itself (called *automorphisms* of A) is also a group under compositions of functions.

B.5 SUBGROUPS, NORMAL SUBGROUPS, AND HOMOMORPHISMS

Let H be a subset of a group G. Then H is called a *subgroup* of G if H itself is a group under the operation of G. Simple criteria to determine subgroups follow.

Proposition B.5: A subset H of a group G is a subgroup of G if:

(i) The identity element $e \in H$.
(ii) H is closed under the operation of G, i.e. if $a, b \in H$, then $ab \in H$.
(iii) H is closed under inverses, that is, if $a \in H$, then $a^{-1} \in H$.

Every group G has the subgroups $\{e\}$ and G itself. Any other subgroup of G is called a *nontrivial subgroup*.

Cosets

Suppose H is a subgroup of G and $a \in G$. Then the set

$$Ha = \{ha \mid h \in H\}$$

is called a *right coset* of H. (Analogously, aH is called a *left coset* of H.) We have the following important results (proved in Problems B.13 and B.15).

Theorem B.6: Let H be a subgroup of a group G. Then the right cosets Ha form a partition of G.

Theorem B.7 (Lagrange): Let H be a subgroup of a finite group G. Then the order of H divides the order of G.

The number of right cosets of H in G, called the index of H in G, is equal to the number of left cosets of H in G; and both numbers are equal to $|G|$ divided by $|H|$.

Normal Subgroups

The following definition applies.

Definition B.2: A subgroup H of G is a *normal* subgroup if $a^{-1}Ha \subseteq H$, for every $a \in G$, or, equivalently, if $aH = Ha$, i.e., if the right and left cosets coincide.

Note that every subgroup of an abelian group is normal.

The importance of normal subgroups comes from the following result (proved in Problem B.17).

Theorem B.8: Let H be a normal subgroup of a group G. Then the cosets of H form a group under coset multiplication:

$$(aH)(bH) = abH$$

This group is called the *quotient group* and is denoted by G/H.

Suppose the operation in G is addition or, in other words, G is written additively. Then the cosets of a subgroup H of G are of the form $a + H$. Moreover, if H is a normal subgroup of G, then the cosets form a group under coset addition, that is,

$$(a + H) + (b + H) = (a + b) + H$$

EXAMPLE B.11

(a) Consider the permutation group S_3 of degree 3 which is investigated above. The set $H = \{\varepsilon, \sigma_1\}$ is a subgroup of S_3. Its right and left cosets follow:

Right Cosets	**Left Cosets**
$H = \{\varepsilon, \sigma_1\}$	$H = \{\varepsilon, \sigma_1\}$
$H\phi_1 = \{\phi_1, \sigma_2\}$	$\phi_1 H = \{\phi_1, \sigma_3\}$
$H\phi_2 = \{\phi_2, \sigma_3\}$	$\phi_2 H = \{\phi_2, \sigma_2\}$

Observe that the right cosets and the left cosets are distinct; hence H is not a normal subgroup of S_3.

(b) Consider the group G of 2×2 matrices with rational entries and nonzero determinants. (See Example A.10.) Let H be the subset of G consisting of matrices whose upper-right entry is zero; that is, matrices of the form

$$\begin{bmatrix} a & 0 \\ c & d \end{bmatrix}$$

Then H is a subgroup of G since H is closed under multiplication and inverses and $I \in H$. However, H is not a normal subgroup since, for example, the following product does not belong to H:

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

On the other hand, let K be the subset of G consisting of matrices with determinant 1. One can show that K is also a subgroup of G. Moreover, for any matrix X in G and any matrix A in K, we have

$$\det(X^{-1}AX) = 1$$

Hence $X^{-1}AX$ belongs to K, so K is a normal subgroup of G.

Integers Modulo *m*

Consider the group $\mathbf{Z}$ of integers under addition. Let H denote the multiples of 5, that is,

$$H = \{\ldots, -10, -5, 0, 5, 10, \ldots\}$$

Then H is a subgroup (necessarily normal) of $\mathbf{Z}$. The cosets of H in $\mathbf{Z}$ appear in Fig. B-5(a). By the above Theorem B.8, $\mathbf{Z}/H = \{0, 1, 2, 3, 4\}$ is a group under coset addition; its addition table appears in Fig. B-5(b).

This quotient group $\mathbf{Z}/H$ is referred to as the integers modulo 5 and it is frequently denoted by $\mathbf{Z_5}$. Analogously, for any positive integer n, there exists the quotient group $\mathbf{Z_n}$ called the *integers modulo n*.

$\overline{0} = 0 + H = H = \{..., -10, -5, 0, 5, 10, ...\}$
$\overline{1} = 1 + H = \{..., -9, -4, 1, 6, 11, ...\}$
$\overline{2} = 2 + H = \{..., -8, -3, 2, 7, 12, ...\}$
$\overline{3} = 3 + H = \{..., -7, -2, 3, 8, 13, ...\}$
$\overline{4} = 4 + H = \{..., -6, -1, 4, 9, 14, ...\}$

(a)

$+$	$\overline{0}$	$\overline{1}$	$\overline{2}$	$\overline{3}$	$\overline{4}$
$\overline{0}$	$\overline{0}$	$\overline{1}$	$\overline{2}$	$\overline{3}$	$\overline{4}$
$\overline{1}$	$\overline{1}$	$\overline{2}$	$\overline{3}$	$\overline{4}$	$\overline{0}$
$\overline{2}$	$\overline{2}$	$\overline{3}$	$\overline{4}$	$\overline{0}$	$\overline{1}$
$\overline{3}$	$\overline{3}$	$\overline{4}$	$\overline{0}$	$\overline{1}$	$\overline{2}$
$\overline{4}$	$\overline{4}$	$\overline{0}$	$\overline{1}$	$\overline{2}$	$\overline{3}$

(b)

Fig. B-5

Cyclic Subgroups

Let G be any group and let a be any element of G. As usual, we define $a^0 = e$ and $a^{n+1} = a^n \cdot a$. Clearly, $a^m a^n = a^{m+n}$ and $(a^m)^n = a^{mn}$, for any integers m and n. Let S denote the set of all the powers of a; that is

$$S = \{\cdots, a^{-3},\ a^{-2},\ a^{-1},\ e,\ a,\ a^2,\ a^3,\ \cdots\}$$

Then S is a subgroup of G called the cyclic group generated by a. We denote this group by $gp(a)$.

Furthermore, suppose that the powers of a are not distinct, say $a^r = a^s$ with, say, $r > s$. Then $a^{r-s} = e$ where $r, s > 0$. The smallest positive integer m such that $a^m = e$ is called the *order* of a and it will be denoted by $|a|$. If $|a| = m$, then the cyclic subgroup $gp(a)$ has m elements as follows:

$$gp(a) = \{e, a, a^2, a^3, \ldots, a^{m-1}\}$$

Consider, for example, the element ϕ_1 in the symmetric group S_3 discussed above. Then:

$$\phi_1{}^1 = \phi_1, \quad \phi_1{}^2 = \phi_2, \quad \phi_1{}^3 = \phi_2 \cdot \phi_1 = e$$

Hence $|\phi_1| = 3$ and $gp(\phi_1) = \{e,\ \phi_1,\ \phi_2\}$. Observe that $|\phi_1|$ divides the order of S_3. This is true in general; that is, for any element a in a group G, $|a|$ equals the order of $gp(a)$ and hence $|a|$ divides $|G|$ by Lagrange's Theorem B.7. We also remark that a group G is said to be *cyclic* if it has an element a such that $G = gp(a)$.

Generating Sets, Generators

Consider any subset A of a group G. Let $gp(A)$ denote the set of all elements x in G such that x is equal to a product of elements where each element comes from the set $A \cup A^{-1}$ (where A^{-1} denotes the set of inverses of elements of A). That is,

$$gp(A) = \{x \in G \mid x = b_1 b_2 \ldots b_m \text{ where each } b_i \in A \cup A^{-1}\}$$

Then $gp(A)$ is a subgroup of G with *generating set* A. In particular, A is said to generate the group G if $G = gp(A)$, that is, if every g in G is a product of elements from $A \cup A^{-1}$. We say A is a *minimal set of generators* of G if A generates G and if no set with fewer elements than A generates G. For example, the permutations $a = \sigma_1$ and $b = \phi_1$ form a minimal set of generators of the symmetric group S_3 (Fig. B-4). Specifically,

$$e = a^2, \quad \sigma_1 = a, \quad \sigma_2 = ab, \quad \sigma_3 = ab^2, \quad \phi_1 = b, \quad \phi_2 = b^2$$

and S_3 is not cyclic so it cannot be generated by one element.

Homomorphisms

A mapping f from a group G into a group G' is called a homomorphism if, for every $a, b \in G$,

$$f(ab) = f(a) f(b)$$

In addition, if f is one-to-one and onto, then f is called an *isomorphism*; and G and G' are said to be *isomorphic*, written $G \cong G'$.

If $f : G \to G'$ is a homomorphism, then the kernel of f, written Ker f, is the set of elements whose image is the identity element e' of G'; that is,

$$\text{Ker } f = \{a \in G \mid f(a) = e'\}$$

Recall that the image of f, written $f(G)$ or Im f, consists of the images of the elements under f; that is,

$$\text{Im } f = \{b \in G' \mid \text{there exists } a \in G \text{ for which } f(a) = b\}\ .$$

The following theorem (proved in Problem B.19) is fundamental to group theory.

Theorem B.9: Suppose $f\colon G \to G'$ is a homomorphism with kernel K. Then K is a normal subgroup of G, and the quotient group G/K is isomorphic to $f(G)$.

EXAMPLE B.12

(a) Let G be the group of real numbers under addition, and let G' be the group of positive real numbers under multiplication. The mapping $f\colon G \to G'$ defined by $f(a) = 2^a$ is a homomorphism because

$$f(a+b) = 2^{a+b} = 2^a 2^b = f(a)f(b)$$

In fact, f is also one-to-one and onto; hence G and G' are isomorphic.

(b) Let a be any element in a group G. The function $f\colon \mathbf{Z} \to G$ defined by $f(n) = a^n$ is a homomorphism since

$$f(m+n) = a^{m+n} = a^m \cdot a^n = f(m) \cdot f(n)$$

The image of f is $gp(a)$, the cyclic subgroup generated by a. By Theorem B.9,

$$gp(a) \cong \mathbf{Z}/\mathbf{K}$$

where K is the kernel of f. If $K = \{0\}$, then $gp(a) = \mathbf{Z}$. On the other hand, if m is the order of a, then $K = \{\text{multiples of } m\}$, and so $gp(a) \cong \mathbf{Z}_m$. In other words, any cyclic group is isomorphic to either the integers $\mathbf{Z}$ under addition, or to $\mathbf{Z}_m$, the integers under addition modulo m.

B.6 RINGS, INTEGRAL DOMAINS, AND FIELDS

Let R be a nonempty set with two binary operations, an operation of addition (denoted by $+$) and an operation of multiplication (denoted by juxtaposition). Then R is called a *ring* if the following axioms are satisfied:

$[\mathbf{R_1}]$ For any $a, b, c \in R$, we have $(a+b)+c = a+(b+c)$.

$[\mathbf{R_2}]$ There exists an element $0 \in R$, called the *zero* element, such that, for every $a \in R$,

$$a + 0 = 0 + a = a.$$

$[\mathbf{R_3}]$ For each $a \in R$ there exists an element $-a \in R$, called the *negative* of a, such that

$$a + (-a) = (-a) + a = 0.$$

$[\mathbf{R_4}]$ For any $a, b \in R$, we have $a + b = b + a$.

$[\mathbf{R_5}]$ For any $a, b, c \in R$, we have $(ab)c = a(bc)$.

$[\mathbf{R_6}]$ For any $a, b, c \in R$, we have: (i) $a(b+c) = ab + ac$, and (ii) $(b+c)a = ba + ca$.

Observe that the axioms $[\mathbf{R_1}]$ through $[\mathbf{R_4}]$ may be summarized by saying that R is an abelian group under addition.

Subtraction is defined in R by $a - b = a + (-b)$.

One can prove (Problem B.21) that $a \cdot 0 = 0 \cdot a = 0$ for every $a \in R$.

A subset S of R is a *subring* of R if S itself is a ring under the operations in R. We note that S is a subring of R if: (i) $0 \in S$, and (ii) for any $a, b \in S$, we have $a - b \in S$ and $ab \in S$.

Special Kinds of Rings: Integral Domains and Fields

This subsection defines a number of different kinds of rings, including integral domains and fields.

R is called a *commutative ring* if $ab = ba$ for every $a, b \in R$.

R is called a *ring with an identity element 1* if the element 1 has the property that $a \cdot 1 = 1 \cdot a = a$ for every element $a \in R$. In such a case, an element $a \in R$ is called a *unit* if a has a multiplicative inverse, that is, an element a^{-1} in R such that $a \cdot a^{-1} = a^{-1} \cdot a = 1$.

R is called a *ring with zero divisors* if there exist nonzero elements $a, b \in R$ such that $ab = 0$. In such a case, a and b are called *zero divisors*.

Definition B.3: A commutative ring R is an *integral domain* if R has no zero divisors, that is, if $ab = 0$ implies $a = 0$ or $b = 0$.

Definition B.4: A commutative ring R with an identity element 1 (not equal to 0) is a field if every nonzero $a \in R$ is a unit, that is, has a multiplicative inverse.

A field is necessarily an integral domain; for if $ab = 0$ and $a \neq 0$, then

$$b = 1 \cdot b = a^{-1}ab = a^{-1} \cdot 0 = 0$$

We remark that a field may also be viewed as a commutative ring in which the nonzero elements form a group under multiplication.

EXAMPLE B.13

(a) The set $\mathbf{Z}$ of integers with the usual operations of addition and multiplication is the classical example of an integral domain (with an identity element). The units in $\mathbf{Z}$ are only 1 and -1, that is, no other element in $\mathbf{Z}$ has a multiplicative inverse.

(b) The set $\mathbf{Z}_m = \{0, 1, 2, \ldots, m-1\}$ under the operation of addition and multiplication modulo m is a ring; it is called the *ring of integers modulo m*. If m is a prime, then $\mathbf{Z}_m$ is a field. On the other hand, if m is not a prime then $\mathbf{Z}_m$ has zero divisors. For instance, in the ring $\mathbf{Z}_6$,

$$2 \cdot 3 = 0 \quad \text{but} \quad 2 \not\equiv 0 \pmod 6 \quad \text{and} \quad 3 \not\equiv 0 \pmod 6$$

(c) The rational numbers $\mathbf{Q}$ and the real numbers $\mathbf{R}$ each form a field with respect to the usual operations of addition and multiplication.

(d) Let M denote the set of 2×2 matrices with integer or real entries. Then M is a noncommutative ring with zero divisors under the operations of matrix addition and matrix multiplication. M does have an identity element, the identity matrix.

(e) Let R be any ring. Then the set $R[x]$ of all polynomials over R is a ring with respect to the usual operations of addition and multiplication of polynomials. Moreover, if R is an integral domain then $R[x]$ is also an integral domain.

Ideals

A subset J of a ring R is called an *ideal* in R if the following three properties hold:

(i) $0 \in J$.

(ii) For any $a, b \in J$, we have $a - b \in J$.

(iii) For any $r \in R$ and $a \in J$, we have $ra, ar \in J$.

Note first that J is a subring of R. Also, J is a subgroup (necessarily normal) of the additive group of R. Thus we can form the following collection of cosets which form a partition of R:

$$\{a + J \mid a \in R\}$$

The importance of ideals comes from the following theorem which is analogous to Theorem B.7 for normal subgroups.

Theorem B.10: Let J be an ideal in a ring R. Then the cosets $\{a + J \mid a \in R\}$ form a ring under the coset operations

$$(a + J) + (b + J) = a + b + J \quad \text{and} \quad (a + J)(b + J) = ab + J$$

This ring is denoted by R/J and is called the *quotient ring*.

Now let R be a commutative ring with an identity element 1. For any $a \in R$, the following set is an ideal:

$$(a) = \{ra \mid r \in R\} = aR$$

It is called the *principal ideal generated* by a. If every ideal in R is a principal ideal, then R is called a *principal ideal ring*. In particular, if R is also an integral domain, then R is called a *principal ideal domain* (PID).

EXAMPLE B.14

(a) Consider the ring $\mathbf{Z}$ of integers. Then every ideal J in $\mathbf{Z}$ is a principal ideal, that is, $J = (m) = m\mathbf{Z}$, for some integer m. Thus $\mathbf{Z}$ is a principal ideal domain (PID). The quotient ring $\mathbf{Z_m} = \mathbf{Z}/(m)$ is simply the ring of integers modulo m. Although $\mathbf{Z}$ is an integral domain (no zero divisors), the quotient ring $\mathbf{Z_m}$ may have zero divisors, e.g., 2 and 3 are zero divisors in $\mathbf{Z_6}$.

(b) Let R be any ring. Then $\{0\}$ and R are ideals. In particular, if R is a field, then $\{0\}$ and R are the only ideals.

(c) Let K be a field. Then the ring $K[x]$ of polynomials over K is a PID (principal ideal domain). On the other hand, the ring $K[x, y]$ of polynomials in two variables is not a PID.

Ring Homomorphisms

A mapping f from a ring R into a ring R' is called a *ring homomorphism* or, simply, *homomorphism* if, for every $a, b \in R$,

$$f(a + b) = f(a) + f(b), \quad f(ab) = f(a)f(b)$$

In addition, if f is one-to-one and onto, then f is called an *isomorphism*; and R and R' are said to be *isomorphic*, written $R \cong R'$.

Suppose $f: R \to R'$ is a homomorphism. Then the kernel of f, written Ker f, is the set of elements whose image is the zero element 0 of R'; that is,

$$\text{Ker } f = \{r \in R \mid f(r) = 0\}$$

The following theorem (analogous to Theorem B.9 for groups) is fundamental to ring theory.

Theorem B.11: Let $f: R \to R'$ be a ring homomorphism with kernel K. Then K is an ideal in R, and the quotient ring R/K is isomorphic to $f(R)$.

Divisibility in Integral Domains

Now let D be an integral domain. We say that b divides a in D if $a = bc$ for some $c \in D$. An element $u \in D$ is called a *unit* if u divides 1, i.e., if u has a multiplicative inverse. An element $b \in D$ is called an *associate* of $a \in D$ if $b = ua$ for some unit $u \in D$. A nonunit $p \in D$ is said to be *irreducible* if $p = ab$ implies a or b is a unit.

An integral domain D is called a *unique factorization domain* (UFD), if every nonunit $a \in D$ can be written uniquely (up to associates and order) as a product of irreducible elements.

EXAMPLE B.15

(a) The ring $\mathbf{Z}$ of integers is the classical example of a unique factorization domain. The units of $\mathbf{Z}$ are 1 and -1. The only associates of $n \in Z$ are n and $-n$. The irreducible elements of $\mathbf{Z}$ are the prime numbers.

(b) The set $D = \{a + b\sqrt{13} \mid a, b \text{ integers}\}$ is an integral domain. The units of D follow:

$$\pm 1, \quad 18 \pm 5\sqrt{13}, \quad -18 \pm 5\sqrt{13}$$

The elements 2, $3 - \sqrt{13}$ and $-3 - \sqrt{13}$ are irreducible in D. Observe that

$$4 = 2 \cdot 2 = (3 - \sqrt{13})(-3 - \sqrt{13})$$

Thus D is not a unique factorization domain. (See Problem B.97.)

B.7 POLYNOMIALS OVER A FIELD

This section investigates polynomials whose coefficients come from some integral domain or field K. In particular, we show that polynomials over a field K have many of the same properties as the integers.

Basic Definitions

Let K be an integral domain or a field. Formally, a polynomial f over K is an infinite sequence of elements from K in which all except a finite number of them are 0; that is,

$$f = (\ldots, 0, a_n, \ldots, a_1, a_0) \quad \text{or, equivalently,} \quad f(t) = a_n t^n + \cdots + a_1 t + a_0$$

where the symbol t is used as an indeterminate. The entry a_k is called the kth coefficient of f. If n is the largest integer for which $a \neq 0$, then we say that the degree of f is n, written $\deg(f) = n$. We also call a_n the leading coefficient of f. If $a_n = 1$, we call f a *monic* polynomial. On the other hand, if every coefficient of f is 0 then f is called the *zero* polynomial, written $f \equiv 0$. The degree of the zero polynomial is not defined.

Let $K[t]$ be the collection of all polynomials $f(t)$ over K. Consider the polynomials

$$f(t) = a_n t^n + \cdots + a_1 t + a_0 \quad \text{and} \quad g(t) = b_m t^m + \cdots + b_1 t + b_0$$

Then the sum $f + g$ is the polynomial obtained by adding corresponding coefficients; that is, if $m \leq n$, then

$$f(t) + g(t) = a_n t^n + \cdots + (a_m + b_m)t^m + \cdots + (a_1 + b_1)t + (a_0 + b_0)$$

Furthermore, the product of f and g is the polynomial

$$f(t)g(t) = (a_n b_m)t^{n+m} + \cdots + (a_1 b_0 + a_0 b_1)t + (a_0 b_0)$$

That is,

$$f(t)g(t) = c_{n+m}t^{n+m} + \cdots + c_1 t + c_0 \quad \text{where} \quad c_k = \sum_{i=0}^{k} a_i b_{k-i} = a_0 b_k + a_1 b_{k-1} + \cdots + a_k b_0$$

The set K of scalars is viewed as a subset of $K[t]$. Specifically, we identify the scalar $a_0 \in K$ with the polynomial

$$f(t) = a_0 \quad \text{or} \quad a_0 = (\cdots, 0, 0, a_0)$$

Then the operators of addition and scalar multiplication are preserved by this identification. Thus, the mapping $\psi\colon K \to K[t]$ defined by $\psi(a_0) = a_0$ is an isomorphism which embeds K into $K[t]$.

Theorem B.12: Let K be an integral domain. Then $K[t]$ under the operations of addition and multiplication of polynomials is a commutative ring with an identity element 1.

The following simple result has important consequences.

Lemma B.13: Suppose f and g are polynomials over an integral domain K. Then

$$\deg(fg) = \deg(f) + \deg(g).$$

The proof follows directly from the definition of the product of polynomials. Namely, suppose

$$f(t) = a_n t^n + \cdots + a_1 t + a_0 \quad \text{and} \quad g(t) = b_m t^m + \cdots + b_1 t + b_0$$

where $a_n \neq 0$ and $b_m \neq 0$. Thus $\deg(f) = n$ and $\deg(g) = m$. Then

$$f(t)g(t) = a_n b_m t^{n+m} + \text{ terms of lower degree}$$

Also, since K is an integral domain with no zero divisors, $a_n b_m \neq 0$. Thus

$$\deg(fg) = m + n = \deg(f) + \deg(g)$$

and the lemma is proved.

The following proposition lists many properties of our polynomials. (Recall that a polynomial g is said to *divide* a polynomial f if there exists a polynomial h such that $f(t) = g(t)h(t)$.)

Proposition B.14: Let K be an integral domain and let f and g be polynomials over K.

(i) $K[t]$ is an integral domain.

(ii) The units of $K[t]$ are the units in K.

(iii) If g divides f, then $\deg(g) \leq \deg(f)$ or $f \equiv 0$.

(iv) If g divides f and f divides g, then $f(t) = kg(t)$ where k is a unit in K.

(v) If d and d' are monic polynomials such that d divides d' and d' divides d, then $d = d'$.

Euclidean Algorithm, Roots of Polynomials

This subsection discusses the roots of a polynomial $f(t)$, where we now assume the coefficients of $f(t)$ come from a field K. Recall that a scalar $a \in K$ is a *root* of a polynomial $f(t)$ if $f(a) = 0$. First we begin with an important theorem which is very similar to a corresponding theorem for the integers $\mathbf{Z}$.

Theorem B.15 (Euclidean Division Algorithm): Let $f(t)$ and $g(t)$ be polynomials over a field K with $g(t) \neq 0$. Then there exist polynomials $q(t)$ and $r(t)$ such that

$$f(t) = q(t)g(t) + r(t)$$

where either $r(t) \equiv 0$ or $\deg(r) < \deg(g)$.

The above theorem (proved in Problem B.30) formalizes the process known as "long division." The polynomial $q(t)$ is called the *quotient* and the polynomial $r(t)$ is called the *remainder* when $f(t)$ is divided by $g(t)$.

Corollary B.16 (Remainder Theorem): Suppose $f(t)$ is divided by $g(t) = t - a$. Then $f(a)$ is the remainder.

The proof follows from the Euclidean Algorithm. That is, dividing $f(t)$ by $t - a$ we get

$$f(t) = q(t)(t - a) + r(t)$$

where $\deg(r) < \deg(t - a) = 1$. Hence $r(t) = r$ is a scalar. Substituting $t = a$ in the equation for $f(t)$ yields

$$f(a) = q(a)(a - a) + r = q(t) \cdot 0 + r = r$$

Thus $f(a)$ is the remainder, as claimed.

Corollary B.16 also tells us that $f(a) = 0$ if and only if the remainder $r = r(t) \equiv 0$. Accordingly:

Corollary B.17 (Factor Theorem): The scalar $a \in K$ is a root of $f(t)$ if and only if $t - a$ is a factor of $f(t)$.

The next theorem (proved in Problem B.31) tells us the number of possible roots of a polynomial.

Theorem B.18: Suppose $f(t)$ is a polynomial over a field K, and $\deg(f) = n$. Then $f(t)$ has at most n roots.

The following theorem (proved in Problem B.32) is the main tool for finding rational roots of a polynomial with integer coefficients.

Theorem B.19: Suppose a rational number p/q (reduced to lowest terms) is a root of the polynomial

$$f(t) = a_n t^n + \cdots + a_1 t + a_0$$

where all the coefficients $a_n, \ldots, a_1, a_0$ are integers. Then p divides the constant term a_0 and q divides the leading coefficient a_n. In particular, if $c = p/q$ is an integer, then c divides the constant term a_0.

EXAMPLE B.16

(a) Suppose $f(t) = t^3 + t^2 - 8t + 4$. Assuming $f(t)$ has a rational root, find all the roots of $f(t)$.

Since the leading coefficient is 1, the rational roots of $f(t)$ must be integers from among ± 1, ± 2, ± 4. Note $f(1) \neq 0$ and $f(-1) \neq 0$. By synthetic division, or dividing by $t - 2$, we get

$$\begin{array}{r|rrrrrrr} 2 & 1 & + & 1 & - & 8 & + & 4 \\ & & & 2 & + & 6 & - & 4 \\ \hline & 1 & + & 3 & - & 2 & + & 0 \end{array}$$

Therefore $t = 2$ is a root and $f(t) = (t-2)(t^2 + 3t - 2)$. Using the quadratic formula for $t^2 + 3t - 2 = 0$, we obtain the following three roots of $f(t)$:

$$t = 2, \quad t = (-3 + \sqrt{17})/2, \quad t = (-3 - \sqrt{17})/2$$

(b) Suppose $h(t) = t^4 - 2t^3 + 11t - 10$. Find all the real roots of $h(t)$ assuming there are two integer roots.

The integer roots must be among ± 1, ± 2, ± 5, ± 10. By synthetic division (or dividing by $t - 1$ and then $t + 2$) we get

$$\begin{array}{r|rrrrrrrrr} 1 & 1 & - & 2 & + & 0 & + & 11 & - & 10 \\ & & & 1 & - & 1 & - & 1 & + & 10 \\ \hline -2 & 1 & - & 1 & - & 1 & + & 10 & + & 0 \\ & & - & 2 & + & 6 & - & 10 & & \\ \hline & 1 & - & 3 & + & 5 & + & 0 & & \end{array}$$

Thus $t = 1$ and $t = -2$ are roots and $h(t) = (t-1)(t+2)(t^2 - 3t + 5)$. The quadratic formula with $t^2 - 3t + 5$ tells us that there are no other real roots. That is, $t = 1$ and $t = -2$ are the only real roots of $h(t)$.

$K[t]$ as a PID and UFD

The following theorems (proved in Problems B.33 and B.34) apply.

Theorem B.20: The ring $K[t]$ of polynomials over a field K is a principal ideal domain (PID). That is, if J is an ideal in $K[t]$, then there exists a unique monic polynomial d which generates J, that is, every polynomial f in J is a multiple of d.

Theorem B.21: Let f and g be polynomials in $K[t]$, not both zero. Then there exists a unique monic polynomial d such that:

(i) d divides both f and g. (ii) If d' divides f and g, then d′ divides d.

The polynomial d in the above Theorem B.21 is called the *greatest common divisor* of f and g, written $d = gcd(f, g)$. If $d = 1$, then f and g are said to be *relatively prime*.

Corollary B.22: Let d be the greatest common divisor of f and g. Then there exist polynomials m and n such that $d = mf + ng$. In particular, if f and g are relatively prime, then there exist polynomials m and n such that $mf + ng = 1$.

A polynomial $p \in K[t]$ is said to be *irreducible* if p is not a scalar and if $p = fg$ implies f or g is a scalar. In other words, p is irreducible if its only divisors are its associates (scalar multiples). The following lemma (proved in Problem B.36) applies.

Lemma B.23: Suppose $p \in K[t]$ is irreducible. If p divides the product fg of polynomials f and g in $K[t]$, then p divides f or p divides g. More generally, if p divides the product $f_1 f_2 \cdots f_n$ of n polynomials, then p divides one of them.

The next theorem (proved in Problem B.37) states that the polynomials over a field form a *unique factorization domain* (UFD).

Theorem B.24 (Unique Factorization Theorem): Let f be a nonzero polynomial in $K[t]$. Then f can be written uniquely (except for order) as a product

$$f = kp_1 p_2 \cdots p_n$$

where $k \in K$ and the p's are monic irreducible polynomials in $K[t]$.

Fundamental Theorem of Algebra

The proof of the following theorem lies beyond the scope of this text.

Fundamental Theorem of Algebra: Any nonzero polynomial $f(t)$ over the complex field **C** has a root in **C**.

Thus $f(t)$ can be written uniquely (except for order) as a product

$$f(t) = k(t - r_1)(t - r_2) \cdots (t - r_n)$$

where k and the r_i are complex numbers and $\deg(f) = n$.

The above theorem is certainly not true for the real field **R**. For example, $f(t) = t^2 + 1$ is a polynomial over R, but $f(t)$ has no real root.

The following theorem (proved in Problem B.38) does apply.

Theorem B.25: Suppose $f(t)$ is a polynomial over the real field **R**, and suppose the complex number $z = a + bi$, $b \neq 0$, is a root of $f(t)$. Then the complex conjugate $\bar{z} = a - bi$ is also a root of $f(t)$. Hence the following is a factor of $f(t)$:

$$c(t) = (t - z)(t - \bar{z}) = t^2 - 2at + a^2 + b^2$$

The following theorem follows from Theorem B.25 and the Fundamental Theorem of Algebra.

Theorem B.26: Let $f(t)$ be a nonzero polynomial over the real field **R**. Then $f(t)$ can be written uniquely (except for order) as a product

$$f(t) = kp_1(t) p_2(t) \cdots p_n(t)$$

where $k \in \mathbf{R}$ and the $p_i(t)$ are real monic polynomials of degree 1 or 2.

EXAMPLE B.17 Let $f(t) = t^4 - 3t^3 + 6t^2 + 25t - 39$. Find all the roots of $f(t)$ given that $t = 2 + 3i$ is a root.

Since $2 + 3i$ is a root, then $2 - 3i$ is a root and $c(t) = t^2 - 4t + 13$ is a factor of $f(t)$. Dividing $f(t)$ by $c(t)$ we get

$$f(t) = (t^2 - 4t + 13)(t^2 + t - 3)$$

The quadratic formula with $t^2 + t - 3$ gives us the other roots of $f(t)$. That is, the four roots of $f(t)$ are as follows:

$$t = 2 + 3i, \quad t = 2 - 3i, \quad t = (-1 + \sqrt{13})/2, \quad t = (-1 - \sqrt{13})/2$$

Solved Problems

OPERATIONS AND SEMIGROUPS

B.1. Consider the set $\mathbf{Q}$ of rational numbers, and let $*$ be the operation on $\mathbf{Q}$ defined by

$$a * b = a + b - ab$$

(*a*) Find: (i) $3 * 4$; (ii) $2 * (-5)$; (iii) $7 * (1/2)$.

(*b*) Is $(\mathbf{Q}, *)$ a semigroup? Is it commutative?

(*c*) Find the identity element for $*$.

(*d*) Do any of the elements in $\mathbf{Q}$ have an inverse? What is it?

(*a*) (i) $3 * 4 = 3 + 4 - 3(4) = 3 + 4 - 12 = -5$
(ii) $2 * (-5) = 2 + (-5) + 2(-5) = 2 - 5 + 10 = 7$
(iii) $7 * (1/2) = 7 + (1/2) - 7(1/2) = 4$

(*b*) We have:

$$\begin{aligned}(a * b) * c &= (a + b - ab) * c = (a + b - ab) + c - (a + b - ab)c \\ &= a + b - ab + c - ac - bc + abc = a + b + c - ab - ac - bc + abc \\ a * (b * c) &= a * (b + c - bc) = a + (b + c - bc) - a(b + c - bc) \\ &= a + b + c - bc - ab - ac + abc\end{aligned}$$

Hence $*$ is associative and $(\mathbf{Q}, *)$ is a semigroup. Also

$$a * b = a + b - ab = b + a - ba = b * a$$

Hence $(\mathbf{Q}, *)$ is a commutative semigroup.

(*c*) An element e is an identity element if $a * e = a$ for every $a \in \mathbf{Q}$. Compute as follows:

$$a * e = a, \quad a + e - ae = a, \quad e - ea = 0, \quad e(1 - a) = 0, \quad e = 0$$

Accordingly, 0 is the identity element.

(*d*) In order for a to have an inverse x, we must have $a * x = 0$ since 0 is the identity element by Part (*c*). Compute as follows:

$$a * x = 0, \quad a + x - ax = 0, \quad a = ax - x, \quad a = x(a - l), \quad x = a/(a - l)$$

Thus if $a \neq 1$, then a has an inverse and it is $a/(a - 1)$.

B.2. Let S be a semigroup with identity e, and let b and b' be inverses of a. Show that $b = b'$, that is, that inverses are unique if they exist.

We have:

$$b * (a * b') = b * e = b \quad \text{and} \quad (b * a) * b' = e * b' = b'$$

Since S is associative, $(b * a) * b' = b * (a * b')$; hence $b = b'$.

B.3. Let $S = \mathbf{N} \times \mathbf{N}$. Let $*$ be the operation on S defined by $(a, b) * (a', b') = (aa', bb')$.

(*a*) Show that $*$ is associative. (Hence S is a semigroup.)

(*b*) Define $f: (S, *) \to (\mathbf{Q}, \times)$ by $f(a, b) = a/b$. Show that f is a homomorphism.

(*c*) Find the congruence relation $\sim$ in S determined by the homomorphism f, that is, where $x \sim y$ if $f(x) = f(y)$. (See Theorem B.4.)

(*d*) Describe $S/\sim$. Does $S/\sim$ have an identity element? Does it have inverses?

Suppose $x = (a, b)$, $y = (c, d)$, $z = (e, f)$.

(*a*) We have

$$(xy)z = (ac,\ bd) * (e,\ f) = [(ac)e, (bd)f]$$
$$x(yz) = (a,\ b) * (ce,\ df) = [a(ce), b(df)]$$

Since a, b, c, d, e, f, are positive integers, $(ac)e = a(ce)$ and $(bd)f = b(df)$. Thus $(xy)z = x(yz)$ and hence $*$ is associative. That is, $(S, *)$ is a semigroup.

(*b*) f is a homomorphism since

$$f(x * y) = f(ac,\ bd) = (ac)/(bd) = (a/b)(c/d) = f(x)f(y)$$

(*c*) Suppose $f(x) = f(y)$. Then $a/b = c/d$ and hence $ad = bc$. Thus f determines the congruence relation $\sim$ on S defined by $(a, b) \sim (c, d)$ if $ad = bc$.

(*d*) The image of f is $\mathbf{Q}^+$, the set of positive rational numbers. By Theorem B.3, $S/\sim$ is isomorphic to $\mathbf{Q}^+$. Thus $S/\sim$ does have an identity element, and every element has an inverse.

B.4. Prove Theorem B.1. Suppose $*$ is an associative operation on a set S. Then any product $a_1 * a_2 * \ldots * a_n$ requires no parenthesis, that is, all possible products are equal.

The proof is by induction on n. Since n is associative, the theorem holds for $n = 1, 2$, and 3. Suppose $n \geq 4$. We use the notation:

$$(a_1a_2, \cdots a_n) = (\cdots((a_1a_2)a_3)\cdots)a_n \quad \text{and} \quad [a_1a_2\cdots a_n] = \text{any product}$$

We show $[a_1a_2\cdots a_n] = (a_1a_2\cdots a_n)$ and so all such products will be equal. Since $[a_1a_2\cdots a_n]$ denotes some product, there exists an $r < n$ such that $[a_1a_2\cdots a_n] = [a_1a_2\cdots a_r]\,[a_{r+1}\cdots a_n]$. Therefore, by induction,

$$\begin{aligned}[a_1a_2\cdots a_n] &= [a_1a_2\cdots a_r]\,[a_{r+1}\cdots a_n] = [a_1a_2\cdots a_r](a_{r+1}\cdots a_n)\\ &= [a_1\cdots a_r]\,((a_{r+1}\cdots a_{n-1})a_n) = ([a_1\cdots a_r]\,(a_{r-1}\cdots a_{n-1}))a_n\\ &= [a_1\cdots a_{n-1}]a_n = (a_1\cdots a_{n-1})a_n = (a_1a_2\cdots a_n)\end{aligned}$$

Thus the theorem is proved.

B.5. Prove Theorem B.4: Let $f : S \to S'$ be a semigroup homomorphism. Let $a \sim b$ if $f(a) = f(b)$. Then: (i) $\sim$ is a congruence relation; (ii) $S/\sim$ is isomorphic to $f(S)$.

(i) First we show that $\sim$ is an equivalence relation. Since $f(a) = f(a)$, we have $a \sim a$.
If $a \sim b$, then $f(a) = f(b)$ or $f(b) = f(a)$; hence $b \sim a$. Lastly, if $a \sim b$ and $b \sim c$, then $f(a) = f(b)$ and $f(b) = f(c)$; hence $f(a) = f(c)$. Thus $a \sim c$. That is, $\sim$ is an equivalence relation. Suppose now $a \sim a'$ and $b \sim b'$. Then $f(a) = f(a')$ and $f(b) = f(b')$.

Since f is a homomorphism,

$$f(ab) = f(a)f(b) = f(a')\,f(b') = f(a'b')$$

Therefore $ab \sim a'b'$. That is, $\sim$ is a congruence relation.

(ii) Define $\Psi: S/\sim \to f(S)$ by $\Psi([a]) = f(a)$. We need to prove: (1) Ψ is well-defined, that is, $\Psi([a]) \in f(S)$, and if $[a] = [b]$ then $f([a]) = f([b])$. (2) Ψ is an isomorphism, that is, Ψ is a homomorphism, one-to-one and onto.

(1) *Proof that* Ψ *is well-defined*: We have $\Psi([a]) = f(a)$. Since $a \in S$, we have $f(a) \in f(S)$. Hence $\Psi([a]) \in f(S)$, as required. Now suppose $[a] = [b]$. Then $a \sim b$ and hence $f(a) = f(b)$. Thus

$$\Psi([a]) = f(a) = f(b) = \Psi([b])$$

That is, Ψ is well-defined.

(2) *Proof that* Ψ *is an isomorphism*: Since f is a homomorphism,

$$\Psi([a][b]) = \Psi[ab] = f(ab) = f(a)f(b) = \Psi([a])\Psi([b])$$

Hence Ψ is a homomorphism. Suppose $\Psi([a]) = \Psi([b])$. Then $f(a) = f(b)$, and so $a \sim b$. Thus $[a] = [b]$ and Ψ is one-to-one. Lastly, let $y \in f(S)$. Then, $f(a) = y$ for some $a \in S$. Hence $\Psi([a]) = f(a) = y$. Thus Ψ is onto $f(S)$. Accordingly, Ψ is an isomorphism.

GROUPS

B.6. Consider the group $G = \{1, 2, 3, 4, 5, 6\}$ under multiplication modulo 7.

(*a*) Find the multiplication table of G.
(*b*) Find $2^{-1}, 3^{-1}, 6^{-1}$.
(*c*) Find the orders and subgroups generated by 2 and 3.
(*d*) Is G cyclic?

(*a*) To find $a * b$ in G, find the remainder when the product ab is divided by 7.
For example, $5 \cdot 6 = 30$ which yields a remainder of 2 when divided by 7; hence $5 * 6 = 2$ in G. The multiplication table of G appears in Fig. B-6(*a*).

(*b*) Note first that 1 is the identity element of G. Recall that a^{-1} is that element of G such that $aa^{-1} = 1$. Hence $2^{-1} = 4$, $3^{-1} = 5$ and $6^{-1} = 6$.

(*c*) We have $2^1 = 2$, $2^2 = 4$, but $2^3 = 1$. Hence $|2| = 3$ and $gp(2) = \{1, 2, 4\}$. We have $3^1 = 3$, $3^2 = 2$, $3^3 = 6$, $3^4 = 4$, $3^5 = 5$, $3^6 = 1$. Hence $|3| = 6$ and $gp(3) = G$.

(*d*) G is cyclic since $G = gp(3)$.

*	1	2	3	4	5	6
1	1	2	3	4	5	6
2	2	4	6	1	3	5
3	3	6	2	5	1	4
4	4	1	5	2	6	3
5	5	3	1	6	4	2
6	6	5	4	3	2	1

(*a*)

*	1	2	4	7	8	11	13	14
1	1	2	4	7	8	11	13	14
2	2	4	8	14	1	7	11	13
4	4	8	1	13	2	14	7	11
7	7	14	13	4	11	2	1	8
8	8	1	2	11	4	13	14	7
11	11	7	14	2	13	1	8	4
13	13	11	7	1	14	8	4	2
14	14	13	11	8	7	4	2	1

(*b*)

Fig. B-6

B.7. Let G be a reduced residue system modulo 15, say, $G = \{1, 2, 4, 7, 8, 11, 13, 14\}$ (the set of integers between 1 and 15 which are coprime to 15). Then G is a group under multiplication modulo 15.

(*a*) Find the multiplication table of G.
(*b*) Find $2^{-1}, 7^{-1}, 11^{-1}$.
(*c*) Find the orders and subgroups generated by 2, 7, and 11.
(*d*) Is G cyclic?

(*a*) To find $a * b$ in G, find the remainder when the product ab is divided by 15. The multiplication table appears in Fig. B-6(*b*).

(*b*) The integers r and s are inverses if $r * s = 1$. Hence: $2^{-1} = 8$, $7^{-1} = 13$, $11^{-1} = 11$.

(*c*) We have $2^2 = 4$, $2^3 = 8$, $2^4 = 1$. Hence $|2| = 4$ and $\text{gp}(2) = \{1, 2, 4, 8\}$. Also, $7^2 = 4$, $7^3 = 4 * 7 = 13$, $7^4 = 13 * 7 = 1$. Hence $|7| = 4$ and $\text{gp}(7) = \{1, 4, 7, 13\}$. Lastly, $11^2 = 1$. Hence $|11| = 2$ and $\text{gp}(11) = \{1, 11\}$.

(*d*) No, since no element generates G.

B.8. Consider the symmetric group S_3 whose multiplication table is given in Fig. B-4.

(*a*) Find the order and the group generated by each element of S_3.

(*b*) Find the number and all subgroups of S_3.

(*c*) Let $A = \{\sigma_1, \sigma_2\}$ and $B = \{\phi_1, \phi_2\}$. Find AB, $\sigma_3 A$, and $A\sigma_3$.

(*d*) Let $H = gp(\sigma_1)$ and $K = gp(\sigma_2)$. Show that HK is not a subgroup of S_3.

(*e*) Is S_3 cyclic?

(*a*) There are six elements: (1) ε, (2) σ_1, (3) σ_2, (4) σ_3, (5) ϕ_1, (6) ϕ_2. Find the powers of each element x until $x^n = \varepsilon$. Then $|x| = n$, and $gp(x) = \{\varepsilon, x^1, x^2, \ldots, x^{n-1}\}$. Note $x^1 = x$, so we need only begin with $n = 2$ when $x \neq \varepsilon$.

(1) $\varepsilon^1 = \varepsilon$; so $|\varepsilon| = 1$ and $g(\varepsilon) = \{\varepsilon\}$.

(2) $\sigma_1^2 = \varepsilon$; hence $|\sigma_1| = 2$ and $gp(\sigma_1) = \{\varepsilon, \sigma_1\}$.

(3) $\sigma_2^2 = \varepsilon$; hence $|\sigma_2| = 2$ and $gp(\sigma_2) = \{\varepsilon, \sigma_2\}$.

(4) $\sigma_3^2 = \varepsilon$; hence $|\sigma_3| = 2$ and $gp(\sigma_3) = \{\varepsilon, \sigma_3\}$.

(5) $\phi_1^2 = \phi_2$, $\phi_1^3 = \phi_2\phi_1 = \varepsilon$; hence $|\phi_1| = 3$ and $gp(\phi_1) = \{\varepsilon, \phi_1, \phi_2\}$.

(6) $\phi_2^2 = \phi_1$, $\phi_2^3 = \phi_1\phi_2 = \varepsilon$; hence $|\phi_2| = 3$ and $gp(\phi_1) = \{\varepsilon, \phi_2, \phi_1\}$.

(*b*) First of all, $H_1 = \{\varepsilon\}$ and $H_2 = S_3$ are subgroups of S_3. Any other subgroup of S_3 must have order 2 or 3 since its order must divide $|S_3| = 6$. Since 2 and 3 are prime numbers, these subgroups must be cyclic (Problem B.61) and hence must appear in part (a). Thus the other subgroups of S_3 follow:

$$H_3 = \{\varepsilon, \sigma_1\}, \quad H_4 = \{\varepsilon, \sigma_2\}, \quad H_5 = \{\varepsilon, \sigma_3\}, \quad H_6 = \{\varepsilon, \phi_1, \phi_2\}$$

Accordingly, S_3 has six subgroups.

(*c*) Multiply each element of A by each element of B:

$$\sigma_1\phi_1 = \sigma_2, \quad \sigma_1\phi_2 = \sigma_3, \quad \sigma_3\phi_1 = \sigma_3, \quad \sigma_2\phi_2 = \sigma_1$$

Hence $\text{AB} = \{\sigma_1, \sigma_2, \sigma_3\}$.

Multiply σ_3 by each element of A:

$$\sigma_3\sigma_1 = \phi_1, \quad \sigma_3\sigma_2 = \phi_2, \quad \text{hence} \quad \sigma_3 A = \{\phi_1, \phi_2\}$$

Multiply each element of A by σ_3:

$$\sigma_1\sigma_3 = \phi_2, \quad \sigma_2\sigma_3 = \phi_1, \quad \text{hence} \quad A\sigma_3 = \{\phi_1, \phi_2\}$$

(*d*) $H = \{e, \sigma_1\}$, $K = \{e, \sigma_2\}$ and then $HK = \{e, \sigma_1, \sigma_2, \phi_1\}$, which is not a subgroup of S_3 since HK has four elements.

(*e*) S_3 is not cyclic since S_3 is not generated by any of its elements.

B.9. Let σ and τ be the following elements of the symmetric group S_6:

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 3 & 1 & 5 & 4 & 6 & 2 \end{pmatrix} \quad \text{and} \quad \tau = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 5 & 3 & 1 & 6 & 2 & 4 \end{pmatrix}$$

Find: $\tau\sigma$, $\sigma\tau$, σ^2, and σ^{-1}. (Since σ and τ are functions, $\tau\sigma$ means apply σ and then τ.)

Figure B-7 shows the effect on 1, 2, …, 6 of the composition of the permutations:

(*a*) σ and then τ; (*b*) τ and then σ; (*c*) σ and then σ, i.e. σ^2.

Thus:

$$\tau\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 1 & 5 & 2 & 6 & 4 & 3 \end{pmatrix}, \quad \sigma\tau = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 6 & 5 & 3 & 2 & 1 & 4 \end{pmatrix}, \quad \sigma^2 = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 5 & 3 & 6 & 4 & 2 & 1 \end{pmatrix}$$

We obtain σ^{-1} by interchanging the top and bottom rows of σ and then rearranging:

$$\sigma^{-1} = \begin{pmatrix} 3 & 1 & 5 & 4 & 6 & 2 \\ 1 & 2 & 3 & 4 & 5 & 6 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 2 & 6 & 1 & 4 & 3 & 5 \end{pmatrix}$$

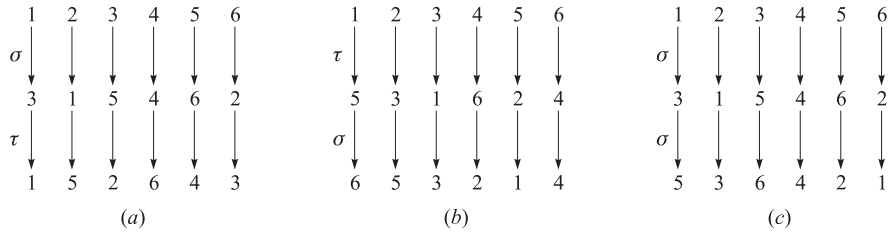


Fig. B-7

B.10. Let H and K be groups.

(*a*) Define the direct product $G = H \times K$ of H and K.

(*b*) What is the identity element and the order of $G = H \times K$?

(*c*) Describe and find the multiplication table of the group $G = \mathbf{Z_2} \times \mathbf{Z_2}$.

(*a*) Let $G = H \times K$, the Cartesian product of H and K, with the operation $*$ defined componentwise by

$$(h,\ k) * h',\ k' = (hh', kk')$$

Then G is a group (Problem B.68), called the *direct product* of H and K.

(*b*) The element $e = (e_H, e_K)$ is the identity element of G, and $|G| = |H| \cdot |K|$.

(*c*) Since $\mathbf{Z_2}$ has two elements, G has four elements. Let

$$e = (0, 0), \quad a = (1, 0), \quad b = (0, 1), \quad c = (1, 1)$$

The multiplication table of G appears in Fig. B-8(*a*). Note that G is abelian since the table is symmetric. Also, $a^2 = e$, $b^2 = e$, $c^2 = e$. Thus G is not cyclic, and hence $G \ncong \mathbf{Z_4}$.

B.11. Let S be the square in the plane $\mathbf{R^2}$ pictured in Fig. B-8(*b*), with its center at the origin 0. Note that the vertices of S are numbered counterclockwise from 1 to 4.

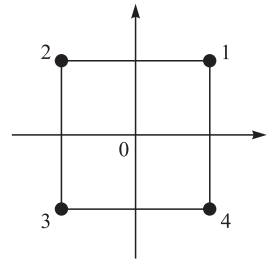
(*a*) Define the group G of symmetries of S.

(*b*) List the elements of G.

(*c*) Find a minimum set of generators of G.

$*$	e	a	b	c
e	e	a	b	c
a	a	e	c	b
b	b	c	e	a
c	c	b	a	e

(a)



(b)

Fig. B-8

(a) A symmetry σ of S is a rigid one-to-one correspondence between S and itself. (Here rigid means that distances between points do not change.) The group G of symmetries of S is the set of all symmetries of S under composition of mappings.

(b) There are eight symmetries as follows. For $\alpha = 0°, 90°, 180°, 270°$, let $\sigma(\alpha)$ be the symmetry obtained by rotating S about its center α degrees, and let $\tau(\alpha)$ be the symmetry obtained by reflecting S about the y-axis and then rotating S about its center α degrees. Note that any symmetry σ of S is completely determined by its effect on the vertices of S and hence σ can be represented as a permutation in S_4. Thus:

$$\sigma(0°) = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 2 & 3 & 4 \end{pmatrix}, \quad \sigma(90°) = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 1 \end{pmatrix},$$

$$\sigma(180°) = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 4 & 1 & 4 \end{pmatrix}, \quad \sigma(270°) = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 1 & 2 & 3 \end{pmatrix}$$

$$\tau(0°) = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 1 & 4 & 3 \end{pmatrix}, \quad \tau(90°) = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 2 & 1 & 4 \end{pmatrix},$$

$$\tau(180°) = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 3 & 2 & 1 \end{pmatrix}, \quad \tau(270°) = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 4 & 3 & 2 \end{pmatrix}$$

(c) Let $a = \sigma(90°)$ and $b = \tau(0°)$. Then a and b form a maximum set of generators of G. Specifically,

$$\sigma(0°) = a^4, \quad \sigma(90°) = a, \quad \sigma(180°) = a^2, \quad \sigma(270°) = a^3$$
$$\tau(0°) = b, \quad \tau(90°) = ba, \quad \tau(180°) = ba^2, \quad \tau(270°) = ba^3$$

and G is not cyclic so it is not generated by one element. (One can show that the relations $a^4 = e$, $b^2 = e$, and $bab = a^{-1}$ completely describe G.)

B.12. Let G be a group and let A be a nonempty set.

(a) Define the meaning of the statement "G acts on A."

(b) Define the stabilizer H_a of an element $a \in A$.

(c) Show that H_a is a subgroup of G.

(a) Let PERM(A) denote the group of all permutations of A. Let $\psi : G \to \text{PERM}(A)$ be any homomorphism. Then G is said to act on A where each element g in G defines a permutation $g : A \to A$ by

$$g(a) = (\psi(g))(a)$$

(Frequently, the permutation $g : A \to A$ is given directly and hence the homomorphism is implicitly defined.)

(b) The stabilizer H_a of $a \in A$ consists of all elements of G which "fix a," that is,

$$H_a = \{g \in G \mid g(a) = a\}$$

(*c*) Since $e(a) = a$, we have $e \in H_a$. Suppose $g, g' \in H_a$. Then $(gg')(a) = g(g'(a)) = g(a) = a$; hence $gg' \in H_a$. Also, $g^{-1}(a) = a$ since $g(a) = a$; hence $g^{-1} \in H_a$. Thus H_a is a subgroup of G.

B.13. Prove Theorem B.6: Let H be a subgroup of a group G. Then the right cosets Ha form a partition of G.

Since $e \in H$, we have $a = ea \in Ha$; hence every element belongs to a coset. Now suppose Ha and Hb are not disjoint. Say $c \in Ha \cap Hb$. The proof is complete if we show that $Ha = Hb$.

Since c belongs to both Ha and Hb, we have $c = h_1 a$ and $c = h_2 b$, where $h_1, h_2 \in H$. Then $h_1 a = h_2 b$, and so $a = h_1^{-1} h_2 b$. Let $x \in Ha$. Then

$$x = h_3 a = h_3 h_1^{-1} h_2 b$$

where $h_3 \in H$. Since H is a subgroup, $h_3 h_1^{-1} h_2 \in H$; hence $x \in Hb$. Since x was any element of Ha, we have $Ha \subseteq Hb$. Similarly, $Hb \subseteq Ha$. Both inclusions imply $Ha = Hb$, and the theorem is proved.

B.14. Let H be a finite subgroup of G. Show that H and any coset Ha have the same number of elements.

Let $H = \{h_1, h_2, \ldots, h_k\}$, where H has k elements. Then $Ha = \{h_1 a, h_2 a, \ldots, h_k a\}$.

However, $h_i a = h_j a$ implies $h_i = h_j$; hence the k elements listed in Ha are distinct. Thus H and Ha have the same number of elements.

B.15. Prove Theorem B.7 (Lagrange): Let H be a subgroup of a finite group G. Then the order of H divides the order of G.

Suppose H has r elements and there are s right cosets; say

$$Ha_1, Ha_2, \ldots, Ha_s$$

By Theorem B.6, the cosets partition G and by Problem B.14, each coset has r elements. Therefore G has rs elements, and so the order of H divides the order of G.

B.16. Prove: Every subgroup of a cyclic group G is cyclic.

Since G is cyclic, there is an element $a \in G$ such that $G = gp(a)$. Let H be a subgroup of G. If $H = \{e\}$, then $H = gp(e)$ and H is cyclic. Otherwise, H contains a nonzero power of a. Since H is a subgroup, it must be closed under inverses and so H contains positive powers of a. Let m be the smallest positive power of a such that a^m belongs to H. We claim that $b = a^m$ generates H. Let x be any other element of H; since x belongs to G we have $x = a^n$ for some integer n. Dividing n by m we get a quotient q and a remainder r, that is,

$$n = mq + r$$

where $0 \leq r < m$. Then

$$a^n = a^{mq+r} = a^{mq} \cdot a^r = b^q \cdot a^r \quad \text{so} \quad a^r = b^{-q} a^n$$

But $a^n, b \in H$. Since H is a subgroup, $b^{-q} a^n \in H$, which means $a^r \in H$. However, m is the smallest positive power of a belonging to H. Therefore, $r = 0$. Hence $x = a^n = b^q$. Thus b generates H, and H is cyclic.

B.17. Prove Theorem B.8: Let H be a normal subgroup of a group G. Then the cosets of H in G form a group under coset multiplication defined by $(aH)\,(bH) = abH$.

Coset multiplication is well-defined, since

$$(aH)(bH) = a(Hb)H = a(bH)H = ab(HH) = abH$$

(Here we have used the fact that H is normal, so $Hb = bH$, and, from Problem B.57, that $HH = H$.) Associativity of coset multiplication follows from the fact that associativity holds in G. H is the identity element of G/H, since

$$(aH)H = a(HH) = aH \quad \text{and} \quad H(aH) = (Ha)H = (aH)H = aH$$

Lastly, $a^{-1}H$ is the inverse of aH since

$$(a^{-1}H)(aH) = a^{-1}aHH = eH = H \quad \text{and} \quad (aH)(a^{-1}H) = aa^{-1}HH = eH = H$$

Thus G/H is a group under coset multiplication.

B.18. Suppose $F : G \to G'$ is a group homomorphism. Prove: (*a*) $f(e) = e'$; (*b*) $(fa^{-6}) = f(a)^{-1}$.

(*a*) Since $e = ee$ and f is a homomorphism, we have

$$f(e) = f(ee) = f(e)f(e)$$

Multiplying both sides by $f(e)^{-1}$ gives us our result.

(*b*) Using part (a) and that $aa^{-1} = a^{-1}a = e$, we have

$$e' = f(e) = f(aa^{-1}) = f(a)f(a^{-1}) \quad \text{and} \quad e' = f(e) = f(a^{-1}a) = f(a^{-1})f(a)$$

Hence $f(a^{-1})$ is the inverse of $f(a)$; that is, $f(a^{-1}) = f(a)^{-1}$.

B.19. Prove Theorem B.9: Let $f : G \to G'$ be a homomorphism with kernel K. Then K is a normal subgroup of G, and G/K is isomorphic to the image of f. (Compare with Problem B.5, the analogous theorem for semigroups.)

Proof that K is normal: By Problem B.18, $f(e) = e'$, so $e \in K$. Now suppose $a, b \in K$ and $g \in G$. Then $f(a) = e'$ and $f(b) = e'$. Hence

$$\begin{aligned} f(ab) &= f(a)f(b) = e'e' = e' \\ f(a^{-1}) &= f(a)^{-1} = e'^{-1} = e' \\ f(gag^{-1}) &= f(g)f(a)f(g^{-1}) = f(g)e'f(g)^{-1} = e' \end{aligned}$$

Hence ab, a^{-1}, and gag^{-1} belong to K, so K is a normal subgroup.

Proof that $G/K \cong H$, where H is the image of f: Let φ: $G/K \to H$ be defined by

$$\varphi(Ka) = f(a)$$

We show that φ is well-defined, i.e., if $Ka = Kb$ then $\varphi(Ka) = \varphi(Kb)$. Suppose $Ka = Kb$. Then $ab^{-1} \in K$ (Problem B.57). Then $f(ab^{-1}) = e'$, and so

$$f(a)f(b)^{-1} = f(a)f(b^{-1}) = f(ab^{-1}) = e'$$

Hence $f(a) = f(b)$, and so $\varphi(Ka) = \varphi(Kb)$. Thus φ is well-defined.

We next show that φ is a homomorphism:

$$\varphi(KaKb) = \varphi(Kab) = f(ab) = f(a)f(b) = \varphi(Ka)\varphi(Kb)$$

Thus φ is a homomorphism. We next show that φ is one-to-one. Suppose $\varphi(Ka) = \varphi(Kb)$. Then

$$f(a) = f(b) \quad \text{or} \quad f(a)f(b)^{-1} = e' \quad \text{or} \quad f(a)f(b^{-1}) = e' \quad \text{or} \quad f(ab^{-1}) = e'$$

Thus $ab^{-1} \in K$, and by Problem B.57 we have $Ka = Kb$. Thus φ is one-to-one. We next show that φ is onto. Let $h \in H$. Since H is the image of f, there exists $a \in G$ such that $f(a) = h$. Thus $\varphi(Ka) = f(a) = h$, and so φ is onto. Consequently $G/K \cong H$ and the theorem is proved.

RINGS, INTEGRAL DOMAINS, FIELDS

B.20. Consider the ring $\mathbf{Z_{10}} = \{0, 1, 2, \ldots, 9\}$ of integers modulo 10. (*a*) Find the units of $\mathbf{Z_{10}}$. (*b*) Find -3, -8, and 3^{-1}. (*c*) Let $f(x) = 2x^2 + 4x + 4$. Find the roots of $f(x)$ over $\mathbf{Z_{10}}$.

(*a*) By Problem B.78 those integers relatively prime to the modulus $m = 10$ are the units in $\mathbf{Z}_{10}$. Hence the units are 1, 3, 7, and 9.

(*b*) Recall that $-a$ in a ring R is the element such that $a+(-a) = (-a)+a = 0$. Hence $-3 = 7$ since $3+7 = 7+3 = 0$ in $\mathbf{Z_{10}}$. Similarly $-8 = 2$. Recall that a^{-1} in a ring R is the element such that $a \cdot a^{-1} = a^{-1} \cdot a = 1$. Hence $3^{-1} = 7$ since $3 \cdot 7 = 7 \cdot 3 = 1$ in $\mathbf{Z_{10}}$.

(*c*) Substitute each of the ten elements of $\mathbf{Z_{10}}$ into $f(x)$ to see which elements yield 0. We have:

$$f(0) = 4, \quad f(2) = 0, \quad f(4) = 2, \quad f(6) = 0, \quad f(8) = 4$$
$$f(1) = 0, \quad f(3) = 4, \quad f(5) = 4, \quad f(7) = 0, \quad f(9) = 2$$

Thus the roots are 1, 2, 6, and 7. (This example shows that a polynomial of degree n can have more than n roots over an arbitrary ring. This cannot happen if the ring is a field.)

B.21. Prove that in a ring R: (i) $a \cdot 0 = 0 \cdot a = 0$; (ii) $a(-b) = (-a)b = -ab$; (iii) $(-1)a = -a$ (when R has an identity element 1).

(i) Since $0 = 0 + 0$, we have

$$a \cdot 0 = a(0 + 0) = a \cdot 0 + a \cdot 0$$

Adding $-(a \cdot 0)$ to both sides yields $0 = a \cdot 0$. Similarly $0 \cdot a = 0$.

(ii) Using $b + (-b) = (-b) + b = 0$, we have

$$ab + a(-b) = a(b + (-b)) = a \cdot 0 = 0$$
$$a(-b) + ab = a((-b) + b) = a \cdot 0 = 0$$

Hence $a(-b)$ is the negative of ab; that is, $a(-b) = -ab$. Similarly, $(-a)b = -ab$.

(iii) We have

$$a + (-1)a = 1 \cdot a + (-1)a = (1 + (-1))a = 0 \cdot a = 0$$
$$(-1)a + a = (-1)a + 1 \cdot a = ((-1) + 1)a = 0 \cdot a = 0$$

Hence $(-1)a$ is the negative of a; that is, $(-1)a = -a$.

B.22. Let D be an integral domain. Show that if $ab = ac$ with $a \neq 0$ then $b = c$.

Since $ab = ac$, we have

$$ab - ac = 0 \quad \text{and so} \quad a(b - c) = 0$$

Since $a \neq 0$, we must have $b - c = 0$, since D has no zero divisors. Hence $b = c$.

B.23. Suppose J and K are ideals in a ring R. Prove that $J \cap K$ is an ideal in R.

Since J and K are ideals, $0 \in J$ and $0 \in K$. Hence $0 \in J \cap K$. Now let $a, b \in J \cap K$ and let $r \in R$. Then $a, b \in J$ and $a, b \in K$. Since J and K are ideals,

$$a - b, ra, ar \in J \quad \text{and} \quad a - b, ra, ar \in K$$

Hence $a - b, ra, ar \in J \cap K$. Therefore $J \cap K$ is an ideal.

B.24. Let J be an ideal in a ring R with an identity element 1. Prove: (*a*) If $1 \in J$ then $J = R$; (*b*) If any unit $u \in J$ then $J = R$.

(*a*) If $1 \in J$ then for any $r \in R$ we have $r \cdot 1 \in R$ or $r \in J$. Hence $J = R$.

(*b*) If $u \in J$ then $u^{-1} \cdot u \in J$ or $1 \in J$. Hence $J = R$ by part (a).

B.25. Prove: (*a*) A finite integral domain D is a field. (*b*) $\mathbf{Z_p}$ is a field where p is a prime number. (*c*) (Fermat) If p is prime, then $a^p \equiv a \pmod{p}$ for any integer a.

(*a*) Suppose D has n elements, say $D = \{a_1, a_2, \ldots, a_n\}$. Let a be any nonzero element of D. Consider the n elements

$$aa_1, aa_2, \ldots, a_n$$

Since $a \neq 0$, we have $aa_i = aa_k$ implies $a_i = a_k$ (Problem B.22). Thus the above n elements are distinct, and so they must be a rearrangement of the elements of D. One of them, say aa_k, must equal the identity element 1 of D; that is, $aa_k = 1$. Thus a_k is the inverse of a. Since a was any nonzero element of D, we have that D is a field.

(*b*) Recall $\mathbf{Z_p} = \{0, 1, 2, \ldots, p - 1\}$. We show that $\mathbf{Z_p}$ has no zero divisors. zero divisors. Suppose a $*b = 0$ in $\mathbf{Z_p}$; that is, 0 (mod p). Then p divides ab. Since p is prime, p divides a or p divides b. Thus $a \equiv 0 \pmod{p}$ or $b \equiv 0 \pmod{p}$; that is, $a = 0$ or $b = 0$ in $\mathbf{Z_p}$. Accordingly, $\mathbf{Z_p}$ has no zero divisors and hence $\mathbf{Z_p}$ is an integral domain. By part (a), $\mathbf{Z_p}$ is a field.

(*c*) If p divides a, then $a \equiv 0 \pmod{p}$ and so $a^p \equiv a \equiv 0 \pmod{p}$. Suppose p does not divide a, then a may be viewed as a nonzero element of $\mathbf{Z_p}$ is a field, its nonzero elements form a group G under multiplication of order $p - 1$. By Problem B.45, $a^{p-1} = 1$ in $\mathbf{Z_p}$.

In other words, $a^{p-1} \equiv 1 \pmod{p}$. Multiplying by a gives $a^p \equiv a \pmod{p}$, and the theorem is proved.

POLYNOMIALS OVER A FIELD

B.26. Suppose $f(t) = 2t^3 - 3t^2 - 6t - 2$. Find all the roots of $f(t)$ knowing that $f(t)$ has a rational root.

The rational roots of $f(t)$ must be among $\pm 1, \pm 2, \pm 1/2$. Testing each possible root, we get, by synthetic division (or dividing by $2t + 1$),

$$\begin{array}{r|l} -\frac{1}{2} & 2 - 3 - 6 - 2 \\ & \quad -1 + 2 + 2 \\ \hline & 2 - 4 - 4 + 0 \end{array}$$

Therefore $t = -1/2$ is a root and

$$f(t) = (t + 1/2)(2t^2 - 4t - 4) = (2t + 1)(t^2 - 2t - 2)$$

We can now use the quadratic formula on $t^2 - 2t - 2$ to obtain the following three roots of $f(t)$:

$$t = -1/2, \quad t = 1 + \sqrt{3}, \quad t = 1 - \sqrt{3}$$

B.27. Let $f(t) = t^4 - 3t^3 + 3t^2 + 3t - 20$. Find all the roots of $f(t)$ given that $t = 1 + 2i$ is a root.

Since $1 + 2i$ is a root, then $1 - 2i$ is a root and $c(t) = t^2 - 2t + 5$ is a factor of $f(t)$. Dividing $f(t)$ by $c(t)$ we get

$$f(t) = (t^2 - 2t + 5)(t^2 - t - 4)$$

The quadratic formula with $t^2 - t - 4$ gives us the other roots of $f(t)$. That is, the four roots of $f(t)$ follow:

$$t = 1 + 2i, \quad t = 1 - 2i, \quad t = (1 + \sqrt{17})/2, \quad t = (1 - \sqrt{17})/2$$

B.28. Let $K = \mathbf{Z_8}$. Find all roots of $f(t) = t^2 + 6t$.

Here $\mathbf{Z_8} = \{0, 1, 2, \ldots, 7\}$. Substitute each element of $\mathbf{Z_8}$ into $f(t)$ to obtain:

$$f(0) = 0, \quad f(2) = 0, \quad f(4) = 0, \quad f(6) = 0$$

Then $f(t)$ has four roots, $t = 0, 2, 4, 6$. (Theorem B.21 does not hold here since K is not a field.)

B.29. Suppose $f(t)$ is a real polynomial with odd degree n. Show that $f(t)$ has a real root.

The complex (nonreal) roots come in pairs. Since $f(t)$ has an odd number n of roots (counting multiplicity), $f(t)$ must have at least one real root.

B.30. Prove Theorem B.15 (Euclidean Division Algorithm): Let $f(t)$ and $g(t)$ be polynomials over a field K with $g(t) \neq 0$. Then there exist polynomials $q(t)$ and $r(t)$ such that

$$f(t) = q(t)g(t) + r(t)$$

where either $r(t) \equiv 0$ or $\deg(r) < \deg(g)$.

If $f(t) = 0$ or if $\deg(f) < \deg(g)$, then we have the required representation $f(t) = 0g(t) + f(t)$. Now suppose $\deg(f) \geq \deg(g)$, say

$$f(t) = a_n t^n + \cdots + a_1 t + a_0 \quad \text{and} \quad g(t) = b_m t^m + \cdots + b_1 t + b_0$$

where $a_n, b_m \neq 0$ and $n > m$. We form the polynomial

$$f_1(t) = f(t) - \frac{a_n}{b_m} t^{n-m} g(t) \tag{1}$$

(This is the first subtraction step in "long division.") Then $\deg(f_1) < \deg(f)$. By induction, there exist polynomials $q_1(t)$ and $r(t)$ such that $f_1(t) = q_1(t)g(t) + r(t)$ where either $r(t) \equiv 0$ or $\deg(r) < \deg(g)$. Substituting this into (1) and solving for $f(t)$, we get

$$f(t) = \left[q_1(t) + \frac{a_n}{b_m} t^{n-m}\right] g(t) + r(t)$$

which is the desired representation.

B.31. Prove Theorem B.18: Suppose $f(t)$ is a polynomial over a field K, and $\deg(f) = n$. Then $f(t)$ has at most n roots.

The proof is by induction on n. If $n = 1$, then $f(t) = at + b$ and $f(t)$ has the unique root $t = -b/a$. Suppose $n > 1$. If $f(t)$ has no roots, then the theorem is true. Suppose $a \in K$ is a root of $f(t)$. Then

$$f(t) = (t - a)g(t) \tag{1}$$

where $\deg(g) = n - 1$. We claim that any other root of $f(t)$ must also be a root of g(t).

Suppose $b \neq a$ is another root of $f(t)$. Substituting $t = b$ in (1) yields $0 = f(b) = (b - a)g(b)$.

Since K has no zero divisors and $b - a \neq 0$, we must have $g(b) = 0$. By induction, $g(t)$ has at most $n - 1$ roots. Thus $f(t)$ has at most $n - 1$ roots other than a. Thus $f(t)$ has at most n roots.

B.32. Prove Theorem B.19: Suppose a rational number p/q (reduced to lowest terms) is a root of the polynomial

$$f(t) = a_n t^n + \cdots + a_1 t + a_0$$

where all the coefficients $a_n, \ldots, a_1, a_0$ are integers. Then p divides the constant term a_0 and q divides the leading coefficients a_n. In particular, if $c = p/q$ is an integer, then c divides the constant term a_0.

Substitute $t = p/q$ into $f(t) = 0$ to obtain $a_n(p/q)^n + \cdots + a_1(p/q) + a_0 = 0$. Multiply both sides of the equation by q^n to obtain

$$a_n p^n + a_{n-1} p^{n-1} q + a_{n-2} p^{n-2} q^2 + \cdots + a_1 p q^{n-1} + a_0 q^n = 0 \tag{1}$$

Since p divides all of the first n terms of (1), p must divide the last term $a_0 q^n$. Assuming p and q are relatively prime, p divides a_0. Similarly, q divides the last n terms of (1), hence q divides the first term $a_n p^n$. Since p and q are relatively prime, q divides a_n.

B.33. Prove Theorem B.20: The ring $K[t]$ of polynomials over a field K is a principal ideal domain (PID). If J is an ideal in $K[t]$, then there exists a unique monic polynomial d which generates J, that is, every polynomial f in J is a multiple of d.

Let d be a polynomial of lowest degree in J. Since we can multiply d by a nonzero scalar and still remain in J, we can assume without loss in generality that d is a monic polynomial (leading coefficient equal 1). Now suppose $f \in J$. By the division algorithm there exist polynomials q and r such that $f = qd + r$ where either $r \equiv 0$ or $\deg(r) < \deg(d)$. Now $f, d \in J$ implies $qd \in J$ and hence $r = f - qd \in J$. But d is a polynomial of lowest degree in J. Accordingly, $r \equiv 0$ and $f = qd$, that is, d divides f. It remains to show that d is unique. If d' is another monic polynomial which generates J, then d divides d' and d' divides d. This implies that $d = d'$, because d and d' are monic. Thus the theorem is proved.

B.34. Prove Theorem B.21: Let f and g be polynomials in $K[t]$, not both the zero polynomial. Then there exists a unique monic polynomial d such that: (i) d divides both f and g. (ii) If d' divides f and g, then d' divides d.

The set $I = \{mf + ng \mid m, n \in K[t]\}$ is an ideal. Let d be the monic polynomial which generates I. Note $f, g \in I$; hence d divides f and g. Now suppose d' divides f and g. Let J be the ideal generated by d'. Then $f, g \in J$ and hence $I \subseteq J$. Accordingly, $d \in J$ and so d' divides d as claimed. It remains to show that d is unique. If d_1 is another (monic) greatest common divisor of f and g, then d divides d_1 and d_1 divides d. This implies that $d = d_1$ because d and d_1 are monic. Thus the theorem is proved.

B.35. Prove Corollary B.22: Let d be the greatest common divisor of f and g. Then there exist polynomials m and n such that $d = mf + ng$. In particular, if f and g are relatively prime, then there exist polynomials m and n such that $mf + ng = 1$.

From the proof of Theorem B.21 in Problem B.34, the greatest common divisor d generates the ideal $I = \{mf + ng \mid m, n \in K[t]\}$. Thus there exist polynomials m and n such that $d = mf + ng$.

B.36. Prove Lemma B.23: Suppose $p \in K[t]$ is irreducible. If p divides the product fg of polynomials $f, g \in K[t]$, then p divides f or p divides g. More generally, if p divides the product $f_1 f_2 \cdots f_n$ of n polynomials, then p divides one of them.

Suppose p divides fg but not f. Since p is irreducible, the polynomials f and p must then be relatively prime. Thus there exist polynomials $m, n \in K[t]$ such that $mf + np = 1$. Multiplying this equation by g, we obtain $mfg + npg = g$. But p divides fg and so p divides mfg. Also, p divides npg. Therefore, p divides the sum $g = mfg + npg$.

Now suppose p divides $f_1 f_2 \cdots f_n$. If p divides f_1, then we are through. If not, then by the above result p divides the product $f_2 \cdots f_n$. By induction on n, p divides one of the polynomials in the product $f_2 \cdots f_n$. Thus the lemma is proved.

B.37. Prove Theorem B.24 (Unique Factorization Theorem): Let f be a nonzero polynomial in $K[t]$. Then f can be written uniquely (except for order) as a product $f = kp_1 p_2 \cdots p_n$ where $k \in K$ and the p's are monic irreducible polynomials in $K[t]$.

We prove the existence of such a product first. If f is irreducible or if $f \in K$, then such a product clearly exists. On the other hand, suppose $f = gh$ where g and h are nonscalars. Then g and h have degrees less than that of f. By induction, we can assume $g = k_1 g_1 g_2 \cdots g_r$ and $h = k_2 h_1 h_2 \cdots h_s$ where $k_1, k_2 \in K$ and the g_i and h_j are monic irreducible polynomials. Accordingly, our desired representation follows:

$$f = (k_1 k_2) g_1 g_2 \cdots g_r h_1 h_2 \cdots h_s$$

We next prove uniqueness (except for order) of such a product for f. Suppose

$$f = kp_1 p_2 \ldots p_n = k' q_1 q_2 \ldots q_m \quad \text{where} \quad k, k' \in K$$

and the $p_1, \ldots, p_n, q_1, \ldots, q_m$ are monic irreducible polynomials. Now p_1 divides $k'q_1 \ldots q_m$. Since p_1 is irreducible it must divide one of the q's by Lemma B.23. Say p_1 divides q_1. Since p_1 and q_1 are both irreducible and monic, $p_1 = q_1$. Accordingly, $kp_2 \ldots p_n = k' q_2 \ldots q_m$. By induction, we have that $n = m$ and $p_2 = q_2, \ldots, p_n = q_m$ for some rearrangement of the q's. We also have that $k = k'$. Thus the theorem is proved.

B.38. Prove Theorem B.25: Suppose $f(t)$ is a polynomial over the real field R, and suppose the complex number $z = a + bi$, $b \neq 0$, is a root of $f(t)$. Then the complex conjugate $\bar{z} = a - bi$ is also a root of $f(t)$. Hence the following is a factor of $f(t)$:

$$c(t) = (t - z)(t - \bar{z}) = t^2 - 2at + a^2 + b^2$$

Dividing $f(t)$ by $c(t)$ where $\deg(c) = 2$, there exist $q(t)$ and real numbers M and N such that

$$f(t) = c(t)q(t) + Mt + N \tag{1}$$

Since $z = a + bi$ is a root of $f(t)$ and $c(t)$, we have, by substituting $t = a + bi$ in (1),

$$f(z) = c(z)q(z) + M(z) + n \quad \text{or} \quad 0 = 0q(z) + M(z) + N \quad \text{or} \quad M(a + bi) + N = 0$$

Thus $Ma + N = 0$ and $Mb = 0$. Since $b \neq 0$, we must have $M = 0$. Then $0 + N = 0$ or $N = 0$. Accordingly, $f(t) = c(t)q(t)$ and $\bar{z} = a - bi$ is a root of $f(t)$.

Supplementary Problems

OPERATIONS AND SEMIGROUPS

B.39. Consider the set **N** of positive integers, and let $*$ denote least common multiple (lcm) operation on N.

(a) Find $4 * 6$, $3 * 5$, $9 * 18$, $1 * 6$.

(b) Is $(\mathbf{N}, *)$ a semigroup? Is it commutative?

(c) Find the identity element of $*$.

(d) Which elements in N, if any, have inverses and what are they?

B.40. Let $*$ be the operation on the set **R** of real numbers defined by $a * b = a + b + 2ab$.

(a) Find $2 * 3$, $3 * (-5)$, and $7 * (1/2)$. (b) Is $(\mathbf{R}, *)$ a semigroup? Is it commutative?

(c) Find the identity element of $*$. (d) Which elements have inverses and what are they?

B.41. Let A be a nonempty set with the operation $*$ defined by $a * b = a$, and assume A has more than one element.

(a) Is A a semigroup? (c) Does A have an identity element?

(b) Is A commutative? (d) Which elements, if any, have inverses and what are they?

B.42. Let $A = \{a, b\}$. (a) Find the number of operations on A. (b) Exhibit one which is neither associative nor commutative.

B.43. For each of the following sets, state which are closed under: (a) multiplication; (b) addition.

$$A = \{0, 1\}, \quad B = \{1, 2\}, \quad C = \{x \mid x \text{ is prime}\}, \quad D = \{2, 4, 8, \ldots\} = \{x \mid x = 2^n\}.$$

B.44. Let $A = \{\ldots, -9, -6, -3, 0, 3, 6, 9, \ldots\}$, the multiples of 3. Is A closed under:

(a) addition; (b) multiplication; (c) subtraction; (d) division (except by 0)?

B.45. Find a set A of three integers which is closed under: (a) multiplication; (b) addition.

B.46. Let S be an infinite set. Let A be the collection of finite subsets of S and let B be the collection of infinite subsets of S

(a) Is A closed under: (i) union; (ii) intersection; (iii) complements?

(b) Is B closed under: (i) union; (ii) intersection; (iii) complements?

B.47. Let $S = \mathbf{Q} \times \mathbf{Q}$, the set of ordered pairs of rational numbers, with the operation $*$ defined by

$$(a, b) * (x, y) = (ax, ay + b)$$

(a) Find $(3, 4) * (1, 2)$ and $(-1, 3) * (5, 2)$. (c) Find the identity element of S.

(b) Is S a semigroup? Is it commutative? (d) Which elements, if any, have inverses and what are they?

B.48. Let $S = \mathbf{N} \times N$, the set of ordered pairs of positive integers, with the operation $*$ defined by

$$(a, b) * (c, d) = (ad + bc, bd)$$

(a) Find $(3, 4) * (1, 5)$ and $(2, 1) * (4, 7)$.

(b) Show that $*$ is associative. (Hence that S is a semigroup.)

(c) Define $f : (S, *) \to (\mathbf{Q}, +)$ by $f(a, b) = a/b$. Show that f is a homomorphism.

(d) Find the congruence relation $\sim$ in S determined by the homomorphism f, that is, $x \sim y$ if $f(x) = f(y)$.

(e) Describe $S/\sim$. Does $S/\sim$ have an identity element? Does it have inverses?

B.49. Let $S = \mathbf{N} \times \mathbf{N}$. Let $*$ be the operation on S defined by

$$(a, b) * (a', b') = (a + a', b + b')$$

(a) Find $(3, 4) * (1, 5)$ and $(2, 1) * (4, 7)$.

(b) Show that $*$ is associative. (Hence that S is a semigroup.)

(c) Define $f: (S, *) \to (\mathbf{Z}, +)$ by $f(a, b) = a - b$. Show that f is a homomorphism.

(d) Find the congruence relation $\sim$ in S determined by the homomorphism f.

(e) Describe $S/\sim$. Does $S/\sim$ have an identity element? Does it have inverses?

GROUPS

B.50. Consider $\mathbf{Z_{20}} = \{0, 1, 2, \ldots, 19\}$ under addition modulo 20. Let H be the subgroup generated by 5. (a) Find the elements and order of H. (b) Find the cosets of H in $\mathbf{Z_{20}}$.

B.51. Consider $G = \{1, 5, 7, 11\}$ under multiplication modulo 12. (a) Find the order of each element. (b) Is G cyclic? (c) Find all subgroups of G.

B.52. Consider $G = \{1, 5, 7, 11, 13, 17\}$ under multiplication modulo 18. (a) Construct the multiplication table of G. (b) Find 5^{-1}, 7^{-1}, and 17^{-1}. (c) Find the order and group generated by: (i) 5; (ii) 13; (d) Is G cyclic?

B.53. Consider the symmetric group S_4. Let $\alpha = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 4 & 2 & 1 \end{pmatrix}$ and $\beta = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 3 & 1 \end{pmatrix}$.

(a) Find $\alpha\beta$, $\beta\alpha$, α^2, α^{-1}. (b) Find the orders of α, β, and $\alpha\beta$.

B.54. Prove the following results for a group G.

(a) The identity element e is unique.

(b) Each a in G has a unique inverse a^{-1}.

(c) $(a^{-1})^{-1} = a$, $(ab)^{-1} = b^{-1}a^{-1}$, and, more generally, $(a_r a_2 \ldots a_n) = a_n^{-1} \ldots a_2^{-1} a_1^{-1}$.

(d) $ab = ac$ implies $b = c$, and $ba = ca$ implies $b = c$.

(e) For any integers r and s, we have $a^r a^s = a^{r+s}$, $(a^r)^s = a^{rs}$.

(f) G is abelian if and only if $(ab)^2 = a^2b^2$ for all $a, b \in G$.

B.55. Let H be a subgroup of G. Prove: (a) $H = Ha$ if and only if $a \in H$. (b) $Ha = Hb$ if and only if $ab^{-1} \in H$, (c) $HH = H$.

B.56. Prove Proposition B.5: *A* subset H of a group G is a subgroup of G if: (i) $e \in H$, (ii) for all $a, b \in H$, we have $ab, a^{-1} \in H$.

B.57. Let G be a group. Prove:

(a) The intersection of any number of subgroups of G is a subgroup of G.

(b) For any $A \subseteq G$, $gp(A)$ is equal to the intersection of all subgroups of G containing A.

(c) The intersection of any number of normal subgroups of G is a normal subgroup of G.

B.58. Suppose G is an abelian group. Show that any factor group G/H is also abelian.

B.59. Suppose $|G| = p$, where p is a prime. Prove: (a) G has no subgroups except G and $\{e\}$. (b) G is cyclic and every element $a \neq e$ generates G.

B.60. Show that $G = \{1, -1, i, -i\}$ is a group under multiplication, and show that $G \cong \mathbf{Z_4}$ by giving an explicit isomorphism $f: G \to \mathbf{Z_4}$.

B.61. Let H be a subgroup of G with only two right cosets. Show that H is normal.

B.62. Let $S = \mathbf{R^2}$, the Cartesian plane. Find the stabilizer H_a of $a = (1, 0)$ in S where G is the following group acting on S:

(a) $G = \mathbf{Z} \times \mathbf{Z}$ and G acts on S by $g(x, y) = (x + m, y + n)$ where $g = (m, n)$. That is, each element g in G is a translation of S.

(b) $G = (\mathbf{R}, +)$ and G acts on S by $g(x, y) = (x\cos g - y\sin g, x\sin g + y\cos g)$. That is, each element in G rotates S about the origin by an angle g.

B.63. Let S be the regular polygon with n sides, and let G be the group of symmetries of S.

(a) Find the order of G.

(b) Show that G is generated by two elements a and b such that $a^n = e$, $b^2 = e$, and $b^{-1}ab = a^{-1}$. (G is called the *dihedral group*.)

B.64. Suppose a group G acts on a set S, say by the homomorphism $\psi: \to \mathrm{PERM}(S)$.

(a) Prove that, for any $s \in S$: (i) $e(s) = s$, and (ii) $(gg')(s) = g(g'(s))$ where $g, g' \in G$.

(b) The orbit G_s of any $s \in S$ is defined by $G_s = \{g(s) \mid g \in G\}$. Show that the orbits form a partition of S.

(c) Show that $|G_s| =$ the number of cosets of the stabilizer H_s of s in G. (Recall $H_s = \{g \in G \mid g(s) \in s\}$.)

B.65. Let G be an abelian group and let n be a fixed positive integer. Show that the function $f: G \to G$ defined by $f(a) = a^n$ is a homomorphism.

B.66. Let G be the multiplicative group of complex numbers z such that $|z| = 1$, and let $\mathbf{R}$ be the additive group of real numbers. Prove $G \cong \mathbf{R}/\mathbf{Z}$.

B.67. Suppose H and N are subgroups of G with N normal. Show that: (a) HN is a subgroup of G. (b) $H \cap N$ is a normal subgroup of H. (c) $H/(H \cap N) \cong HN/N$.

B.68. Let H and K be groups. Let G be the product set $H \times K$ with the operation

$$(h, k) * (h', k') = (hh', kk').$$

(a) Show that G is a group (called the *direct product* of H and K).

(b) Let $H' = H \times \{e\}$. Show that: (i) $H' \cong H$; (ii) H' is a normal subgroup of G; (iii) $G/H' \cong K$.

RINGS

B.69. Consider the ring $\mathbf{Z}_{12} = \{0, 1, \ldots, 11\}$ of integers modulo 12. (a) Find the units of $\mathbf{Z}_{12}$. (b) Find the roots of $f(x) = x^2 + 4x + 4$ over $\mathbf{Z}_{12}$. (c) Find the associates of 2.

B.70. Consider the ring $\mathbf{Z}_{30} = \{0, 1, \ldots, 29\}$ of integers modulo 30.

(a) Find -2, -7, and -11. (b) Find: 7^{-1}, 11^{-1}, and 26^{-1}.

B.71. Show that in a ring R: (a) $(-a)(-b) = ab$; (b) $(-1)(-1) = 1$, if R has an identity element 1.

B.72. Suppose $a^2 = a$ for every $a \in R$. (Such a ring is called a *Boolean* ring). Prove that R is commutative.

B.73. Let R be a ring with an identity element 1. We make R into another ring R' by defining:

$$a \oplus b = a + b + 1 \quad \text{and} \quad a * b = ab + a + b$$

(a) Verify that R' is a ring. (b) Determine the 0-element and the 1-element of R'.

B.74. Let G be any (additive) abelian group. Define a multiplication in G by $a * b = 0$ for every $a, b \in G$. Show that this makes G into a ring.

B.75. Let J and K be ideals in a ring R. Prove that $J + K$ nd $J \cap K$ are also ideals.

B.76. Let R be a ring with unity 1. Show that $(a) = \{ra \mid r \in R\}$ is the smallest ideal containing a.

B.77. Show that R and $\{0\}$ are ideals of any ring R.

B.78. Prove: (a) The units of a ring R form a group under multiplication. (b) The units in $\mathbf{Z}_m$ are those integers which are relatively prime to m.

B.79. For any positive integer m, verify that $m\mathbf{Z} = \{\text{rm} \mid r \in \mathbf{Z}\}$ is a ring. Show that $2\mathbf{Z}$ and $3\mathbf{Z}$ are not isomorphic.

B.80. Prove Theorem B.10: Let J be an ideal in a ring R. Then the cosets $\{a + J \mid a \in R\}$ form a ring under the coset operations $(a + J) + (b + J) = a + b + J$ and $(a + J)(b + J) = ab + J$.

B.81. Prove Theorem B.11: Let $f: R \to R'$ be a ring homomorphism with kernel K. Then K is an ideal in R, and the quotient ring R/K is isomorphic to $f(R)$.

B.82. Let J be an ideal in a ring R. Consider the (canonical) mapping $f: R \to R/J$ defined by $f(a) = a + J$. Show that: (a) f is a ring homomorphism; (b) f is an onto mapping.

B.83. Suppose J is an ideal in a ring R. Show that: (a) If R is commutative, then R/J is commutative. (b) If R has a unity element 1 and $1 \notin J$, then $1 + R$ is a unity element for R/J.

INTEGRAL DOMAINS AND FIELDS

B.84. Prove that if $x^2 = 1$ in an integral domain D, then $x = -1$ or $x = 1$.

B.85. Let $R \neq \{0\}$ be a finite commutative ring with no zero divisors. Show that R is an integral domain, that is, that R has an identity element 1.

B.86. Prove that $F = \{a + b\sqrt{2} \mid a, b \text{ rational}\}$ is a field.

B.87. Prove that $F = \{a + b\sqrt{2} \mid a, b \text{ integers}\}$ is an integral domain but not a field.

B.88. A complex number $a + bi$ where a, b are integers is called a *Gaussian integer*. Show that the set G of Gaussian integers is an integral domain. Also show that the units are ± 1, $\pm i$.

B.89. Let R be an integral domain and let J be an ideal in R. Prove that the factor ring R/J is an integral domain if and only if J is a prime ideal. (An ideal J is *prime* if $J \neq R$ and if $ab \in J$ implies $a \in J$ or $b \in J$.)

B.90. Let R be a commutative ring with unity element 1, and let J be an ideal in R. Prove that the factor ring R/J is a field if and only if J is a maximal ideal. (An ideal J is maximal if $J \neq R$ and no ideal K lies strictly between J and R, that is, if $J \subseteq K \subseteq R$ then $J = K$ or $K = R$.)

B.91. Let D be the ring of real 2×2 matrices of the form $\begin{bmatrix} a & -b \\ b & a \end{bmatrix}$. Show that D is isomorphic to the complex field C, when D is a field.

B.92. Show that the only ideal in a field K is $\{0\}$ or K itself.

B.93. Suppose $f: K \to K'$ is a homomorphism from a field K to a field K'. Show that f is an *embedding*; that is, f is one-to-one. (We assume $f(1) \neq 0$.)

B.94. Consider the integral domain $D = \{a + b\sqrt{13} \mid a, b \text{ integers}\}$. (See Example B.15(*b*).) If $\alpha = a + \sqrt{13}$, we define $N(\alpha) = a^2 - 13b^2$. Prove:

(i) $N(\alpha\beta) = N(\alpha)N(\beta)$.

(ii) α is a unit if and only if $N(\alpha) = +1$.

(iii) Among the units of D are ± 1, $18 \pm 5\sqrt{13}$; and $-18 \pm 5\sqrt{13}$.

(iv) The numbers 2, $3 - \sqrt{13}$ and $-3 - \sqrt{13}$ are irreducible.

POLYNOMIALS OVER A FIELD

B.95. Find the roots of $f(t)$ assuming $f(t)$ has an integer root: (a) $f(t) = t^3 - 2t^2 - 6t - 3$; (b) $f(t) = t^3 - t^2 - 11t - 10$; (c) $f(t) = t^3 + 2t^2 - 13t - 6$.

B.96. Find the roots of $f(t)$ assuming $f(t)$ has a rational root: (a) $f(t) = 2t^3 - 3t^2 - 16t - 7$; (b) $f(t) = 2t^3 - t^2 - 9t + 9$.

B.97. Find the roots of $f(t) = t^4 - 5t^3 + 16t^2 - 9t - 13$, given that $t = 2 + 3i$ is a root.

B.98. Find the roots of $f(t) = t^4 - t^3 - 5t^2 + 12t - 10$, given that $t = 1 - i$ is a root.

B.99. For any scalar $a \in K$, define the *evaluation map* $\psi_a\colon K[t] \to K$ by $\psi_a(f(t)) = f(a)$. Show that ψ_a is a ring homomorphism.

B.100. Prove: (a) Proposition B.14. (b) Theorem B.26.

Answers to Supplementary Problems

B.39. (a) 12, 15, 18, 6; (b) Yes, yes; (c) 1; (d) Only 1 and it is its own inverse.

B.40. (a) 17, -32, 29/2; (b) Yes, yes; (c) Zero; (d) If $a \neq 1/2$, then a has an inverse which is $-a/(1 + 2a)$.

B.41. (a) Yes; (b) No; (c) No; (d) It is meaningless to talk about inverses when no identity element exists.

B.42. (a) Sixteen, since there are two choices, a or b, for each of the four products aa, ab, ba, and bb. (b) Let $aa = b$, $ab = a$, $ba = b$, $bb = a$. Then $ab \neq ba$. Also, $(aa)b = bb = a$, but $a(ab)\ as = b$.

B.43. (a) A, D; (b) none.

B.44. (a) Yes; (b) yes; (c) yes; (d) no.

B.45. (a) $\{1, -1, 0\}$; (b) There is no set.

B.46. (a) Yes, yes, no; (b) Yes, no, no.

B.47. (a) (3, 10), $(-5, 1)$; (b) yes, no; (c) (1, 0); (d) The element (a, b) has an inverse if $a \neq 0$, and its inverse is $(1/a, -b/a)$.

B.48. (a) (19, 20), (18, 7). (d) $(a, b) \sim (c, d)$ if $ad = bc$. (e) $S/\sim$ is isomorphic to the positive rational numbers under addition. Thus $S/\sim$ has no identity element and no inverses.

B.49. (a) (4, 9), (6, 8); (d) $(a, b) \sim (c, d)$ if $a + d = b + c$. (e) $S/\sim$ is isomorphic $\mathbf{Z}$ since every integer is the difference of two positive integers. Thus $S/\sim$ has an identity element, and every element has an inverse.

B.50. (a) $H = l\{0, 5, 10, 15\}$ and $|H| = 4$. (b) H, $1 + H = \{1, 6, 11, 16\}$, $2 + H = \{2, 7, 12, 17\}$, $3 + H = \{3, 8, 13, 18\}$, $4 + H = \{4, 9, 14, 19\}$.

B.51. (a) $x^2 = 1$ if $x \neq 1$. (b) No. (c) $\{1\}$, $\{1, 5\}$, $\{1, 7\}$, $\{1, 11\}$, G.

B.52. (a) See Fig. B-9(*a*). (b) 11, 13, 17; (c) (i) $|\%| = 6$, $gp(5) = G$; (ii) $|13| = 3$, $gp(13) = \{1, 7, 13\}$; (d) Yes, since $G = gp(5)$.

B.53. (a) See Fig. B-9(*b*). (b) 4, 3, 4.

×	1	5	7	11	13	17
1	1	5	7	11	13	17
5	5	7	17	1	11	13
7	7	17	13	5	1	11
11	11	1	5	13	17	7
13	13	11	1	17	7	5
17	17	13	11	7	5	1

(*a*)

$$\alpha\beta = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 1 & 4 & 2 \end{pmatrix}, \quad \beta\alpha = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 1 & 2 & 3 \end{pmatrix}$$

$$\alpha^2 = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 1 & 4 & 3 \end{pmatrix}, \quad \alpha^{-1} = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 3 & 1 & 2 \end{pmatrix}$$

(*b*)

Fig. B-9

B.60. $f(1) = 0$, $f(i) = 1$, $f(-1) = 2$, $f(-i) = 3$

B.62. (a) $\{(0, 0)\}$, (b) $\{2\pi r \mid r \in Z\}$.

B.69. (a) 1, 5, 7, 11; (b) 4, 10; (c) $\{2, 10\}$.

B.70. (a) 28, 23, 19; (b) 13, 11, 26^{-1} does not exist since 26 is not a unit.

B.72. Show $-a = a$ using $a + a = (a + a)^2$. Then show $ab = -ba$ by $(a + b) = (a + b)^2$.

B.73. (b) $-1 =$ 0−element, $0 =$ 1−element.

B.91. Show f is an isomorphism where $f\left(\begin{bmatrix} a & -b \\ b & a \end{bmatrix}\right) = a + bi$.

B.93. *Hint*: Use Problem B.92.

B.95. (a) -1, $(3 \pm \sqrt{21})/2$; (b) -2, $(3 \pm \sqrt{29})/2$; (c) 3, $(-5 \pm \sqrt{17})/2$

B.96. (a) $-1/2$, $1 \pm 2\sqrt{2}$; (b) 3/2, $(-1 \pm \sqrt{13})/2$

B.97. 2 ± 3i, $(1 \pm \sqrt{5})/2$

B.98. $1 \pm$ i, $(-1 \pm \sqrt{21})/2$