Subject:DISCRETE MATHEMATICS

Branch: B.Tech CSE 4th sem Code: CSE 208-C

Sets ,Relations and Functions

1.1 SETS AND ELEMENTS, SUBSETS

A *set* may be viewed as any well-defined collection of objects, called the *elements* or *members* of the set. One usually uses capital letters, $A, B, X, Y, \ldots,$ to denote sets, and lowercase letters, $a, b, x, y, \ldots,$ to denote elements of sets. Synonyms for "set" are "class," "collection," and "family."

Membership in a set is denoted as follows:

$a \in S$ denotes that a belongs to a set S

$a, b \in S$ denotes that a and b belong to a set S

Here $\in$ is the symbol meaning "is an element of." We use $/\in$ to mean "is not an element of."

Specifying Sets

There are essentially two ways to specify a particular set. One way, if possible, is to list its members separated by commas and contained in braces { }. A second way is to state those properties which characterized the elements in the set. Examples illustrating these two ways are:

$$A = \{1, 3, 5, 7, 9\} \text{ and } B = \{x \mid x \text{ is an even integer, } x > 0\}$$

That is, A consists of the numbers 1, 3, 5, 7, 9. The second set, which reads:

B is the set of x such that x is an even integer and x is greater than 0,

denotes the set B whose elements are the positive integers. Note that a letter, usually x, is used to denote a typical member of the set; and the vertical line | is read as "such that" and the comma as "and."

EXAMPLE 1.1

(a) The set A above can also be written as $A = \{x \mid x \text{ is an odd positive integer, } x < 10\}$.

(b) We cannot list all the elements of the above set B although frequently we specify the set by

$$B = \{2, 4, 6, \ldots\}$$

where we assume that everyone knows what we mean. Observe that $8 \in B$, but $3 \notin B$.

(c) Let $E = \{x \mid x^2 - 3x + 2 = 0\}$, $F = \{2, 1\}$ and $G = \{1, 2, 2, 1\}$. Then $E = F = G$.

We emphasize that a set does not depend on the way in which its elements are displayed. A set remains the same if its elements are repeated or rearranged.

Even if we can list the elements of a set, it may not be practical to do so. That is, we describe a set by listing its elements only if the set contains a few elements; otherwise we describe a set by the property which characterizes its elements.

Subsets

Suppose every element in a set A is also an element of a set B, that is, suppose $a \in A$ implies a B. Then A is called a *subset* of B. We also say that A is *contained* in B or that B *contains* A. This relationship is written

$$A \subseteq B \quad \text{or} \quad B \supseteq A$$

Two sets are equal if they both have the same elements or, equivalently, if each is contained in the other. That is:

$$\boxed{A = B \text{ if and only if } A \subseteq B \text{ and } B \subseteq A}$$

If A is not a subset of B, that is, if at least one element of A does not belong to B, we write A $/\subseteq$ B.

EXAMPLE 1.2 Consider the sets:

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

Then $C \subseteq A$ and $C \subseteq B$ since 1 and 3, the elements of C, are also members of A and B. But $B \,/\!\subseteq A$ since some of the elements of B, e.g., 2 and 5, do not belong to A. Similarly, $A \,/\!\subseteq B$.

Property 1: It is common practice in mathematics to put a vertical line “ |’ or slanted line “/” through a symbol to indicate the opposite or negative meaning of a symbol.

Property 2: The statement $A \subseteq B$ does not exclude the possibility that $A = B$. In fact, for every set A we have $A \subseteq A$ since, trivially, every element in A belongs to A. However, if $A \subseteq B$ and $A \,/\!= B$, then we say A is a proper subset of B (sometimes written $A \subset B$).

Property 3: Suppose every element of a set A belongs to a set B and every element of B belongs to a set C. Then clearly every element of A also belongs to C. In other words, if $A \subseteq B$ and $B \subseteq C$, then $A \subseteq C$.

The above remarks yield the following theorem.

Theorem 1.1: Let A, B, C be any sets. Then:

(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 $A \subseteq C$

Special symbols

Some sets will occur very often in the text, and so we use special symbols for them. Some such symbols are:

$\mathbf{N}$ = the set of *natural numbers* or positive integers: 1, 2, 3,...
$\mathbf{Z}$ = the set of all integers: ..., −2, −1, 0, 1, 2,...
$\mathbf{Q}$ = the set of rational numbers
$\mathbf{R}$ = the set of real numbers
$\mathbf{C}$ = the set of complex numbers

Observe that $\mathbf{N} \subseteq \mathbf{Z} \subseteq \mathbf{Q} \subseteq \mathbf{R} \subseteq \mathbf{C}$.

Universal Set, Empty Set

All sets under investigation in any application of set theory are assumed to belong to some fixed large set called the *universal set* which we denote by

$$\mathbf{U}$$

unless otherwise stated or implied.

Given a universal set **U** and a property P, there may not be any elements of **U** which have property P. For example, the following set has no elements:

$$S = \{x \mid x \text{ is a positive integer, } x^2 = 3\}$$

Such a set with no elements is called the *empty set* or *null set* and is denoted by

$$\emptyset$$

There is only one empty set. That is, if S and T are both empty, then $S=T$, since they have exactly the same elements, namely, none.

The empty set $\emptyset$is also regarded as a subset of every other set. Thus we have the following simple result which we state formally.

Theorem 1.2: For any set A, we have $\emptyset \subseteq A \subseteq \mathbf{U}$.

Disjoint Sets

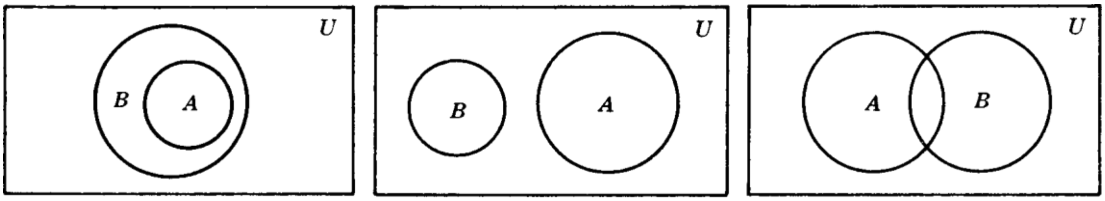
Two sets A and B are said to be *disjoint* if they have no elements in common. For example, suppose

$$A = \{1, 2\}, \quad B = \{4, 5, 6\}, \quad \text{and} \quad C = \{5, 6, 7, 8\}$$

Then A and B are disjoint, and A and C are disjoint. But B and C are not disjoint since B and C have elements in common, e.g., 5 and 6. We note that if A and B are disjoint, then neither is a subset of the other (unless one is the empty set).

1.2 VENN DIAGRAMS

A Venn diagram is a pictorial representation of sets in which sets are represented by enclosed areas in the plane. The universal set **U** is represented by the interior of a rectangle, and the other sets are represented by disks lying within the rectangle. If $A \subseteq B$, then the disk representing A will be entirely within the disk representing B as in Fig. 1-1*(a)*. If A and B are disjoint, then the disk representing A will be separated from the disk representing B as in Fig. 1-1*(b)*.



(*a*) $A \subseteq B$ (*b*) A **and** B **are disjoint** (*c*)

Fig. 1-1

1.3 SET OPERATIONS

This section introduces a number of set operations, including the basic operations of union, intersection, and complement.

Union and Intersection

The *union* of two sets A and B, denoted by $A \cup B$, is the set of all elements which belong to A or to B; that is,

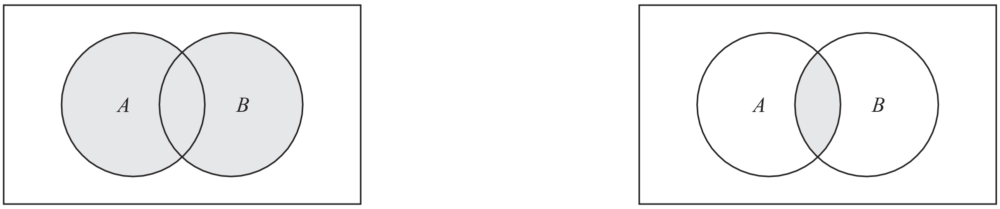
$$A \cup B = \{x \mid x \in A \text{ or } x \in B\}$$

Here "or" is used in the sense of and/or. Figure 1-3*(a)* is a Venn diagram in which $A \cup B$ is shaded.

The *intersection* of two sets A and B, denoted by $A \cap B$, is the set of elements which belong to both A and B; that is,

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

Figure 1-3*(b)* is a Venn diagram in which $A \cap B$ is shaded.



(a) $A \cup B$ is shaded

(b) $A \cap B$ is shaded

Fig. 1-3

Recall that sets A and B are said to be *disjoint* or *nonintersecting* if they have no elements in common or, using the definition of intersection, if $A \cap B = \varnothing$, the empty set. Suppose

$$S = A \cup B \quad \text{and} \quad A \cap B = \varnothing$$

Then S is called the *disjoint union* of A and B.

EXAMPLE 1.4

(a) Let $A = \{1, 2, 3, 4\}$, $B = \{3, 4, 5, 6, 7\}$, $C = \{2, 3, 8, 9\}$. Then

$A \cup B = \{1, 2, 3, 4, 5, 6, 7\}$, $A \cup C = \{1, 2, 3, 4, 8, 9\}$, $B \cup C = \{2, 3, 4, 5, 6, 7, 8, 9\}$,
$A \cap B = \{3, 4\}$, $A \cap C = \{2, 3\}$, $B \cap C = \{3\}$.

(b) Let $\mathbf{U}$ be the set of students at a university, and let M denote the set of male students and let F denote the set of female students. The $\mathbf{U}$ is the disjoint union of M of F ; that is,

$$\mathbf{U} = M \cup F \quad \text{and} \quad M \cap F = \varnothing$$

This comes from the fact that every student in $\mathbf{U}$ is either in M or in F , and clearly no student belongs to both M and F , that is, M and F are disjoint.

The following properties of union and intersection should be noted.

Property 1: Every element x in $A \cap B$ belongs to both A and B; hence x belongs to A and x belongs to B. Thus $A \cap B$ is a subset of A and of B; namely

$$A \cap B \subseteq A \quad \text{and} \quad A \cap B \subseteq B$$

Property 2: An element x belongs to the union $A \cup B$ if x belongs to A or x belongs to B; hence every element in A belongs to $A \cup B$, and every element in B belongs to $A \cup B$. That is,

$$A \subseteq A \cup B \quad \text{and} \quad B \subseteq A \cup B$$

We state the above results formally:

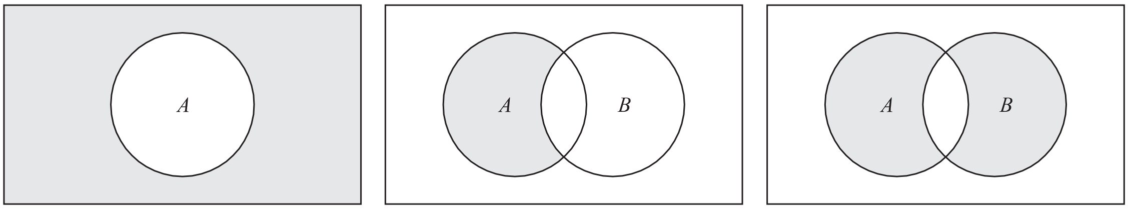
Theorem 1.3: For any sets A and B, we have:

(i) $A \cap B \subseteq A \subseteq A \cup B$ and (ii) $A \cap B \subseteq B \subseteq A \cup B$.

The operation of set inclusion is closely related to the operations of union and intersection, as shown by the following theorem.

Theorem 1.4: The following are equivalent: $A \subseteq B$, $A \cap B = A$, $A \cup B = B$.

This theorem is proved in Problem 1.8. Other equivalent conditions to are given in Problem 1.31.



(*a*) A^C is shaded

(*b*) $A\backslash B$ is shaded

(*c*) $A \oplus B$ is shaded

Fig. 1-4

Complements, Differences, Symmetric Differences

Recall that all sets under consideration at a particular time are subsets of a fixed universal set **U**. The *absolute complement* or, simply, *complement* of a set *A*, denoted by A^C, is the set of elements which belong to **U** but which do not belong to *A*. That is,

$$A^C = \{x \mid x \in \mathbf{U}, x \notin A\}$$

Some texts denote the complement of *A* by A' or $\bar{A}$. Fig. 1-4*(a)* is a Venn diagram in which A^C is shaded.

The *relative complement* of a set *B* with respect to a set *A* or, simply, the *difference* of *A* and *B*, denoted by $A\backslash B$, is the set of elements which belong to *A* but which do not belong to *B*; that is

$$A\backslash B = \{x \mid x \in A, x \notin B\}$$

The set $A\backslash B$ is read "*A* minus *B*." Many texts denote $A\backslash B$ by $A - B$ or $A \sim B$. Fig. 1-4*(b)* is a Venn diagram in which $A\backslash B$ is shaded.

The *symmetric difference* of sets *A* and *B*, denoted by $A \oplus B$, consists of those elements which belong to *A* or *B* but not to both. That is,

$$A \oplus B = (A \cup B)\backslash(A \cap B) \quad \text{or} \quad A \oplus B = (A\backslash B) \cup (B\backslash A)$$

Figure 1-4*(c)* is a Venn diagram in which $A \oplus B$ is shaded.

EXAMPLE 1.5 Suppose $\mathbf{U} = \mathbf{N} = \{1, 2, 3, \ldots\}$ is the universal set. Let

$$A = \{1, 2, 3, 4\},\ B = \{3, 4, 5, 6, 7\},\ C = \{2, 3, 8, 9\},\ E = \{2, 4, 6, \ldots\}$$

(Here *E* is the set of even integers.) Then:

$$A^C = \{5, 6, 7, \ldots\},\ B^C = \{1, 2, 8, 9, 10, \ldots\},\ E^C = \{1, 3, 5, 7, \ldots\}$$

That is, E^C is the set of odd positive integers. Also:

$$\begin{aligned}
&A\backslash B = \{1, 2\}, \quad && A\backslash C = \{1, 4\}, \quad && B\backslash C = \{4, 5, 6, 7\}, \quad && A\backslash E = \{1, 3\}, \\
&B\backslash A = \{5, 6, 7\}, \quad && C\backslash A = \{8, 9\}, \quad && C\backslash B = \{2, 8, 9\}, \quad && E\backslash A = \{6, 8, 10, 12, \ldots\}.
\end{aligned}$$

Furthermore:

$$\begin{aligned}
&A \oplus B = (A\backslash B) \cup (B\backslash A) = \{1, 2, 5, 6, 7\}, \qquad && B \oplus C = \{2, 4, 5, 6, 7, 8, 9\}, \\
&A \oplus C = (A\backslash C) \cup (B\backslash C) = \{1, 4, 8, 9\}, \qquad && A \oplus E = \{1, 3, 6, 8, 10, \ldots\}.
\end{aligned}$$

Fundamental Products

Consider *n* distinct sets $A_1, A_2, \ldots, A_n$. A *fundamental product* of the sets is a set of the form

$$A_1^* \cap A_2^* \cap \ldots \cap A_n^* \quad \text{where} \quad A_i^* = A \quad \text{or} \quad A_i^* = A^C$$

We note that:

$=$

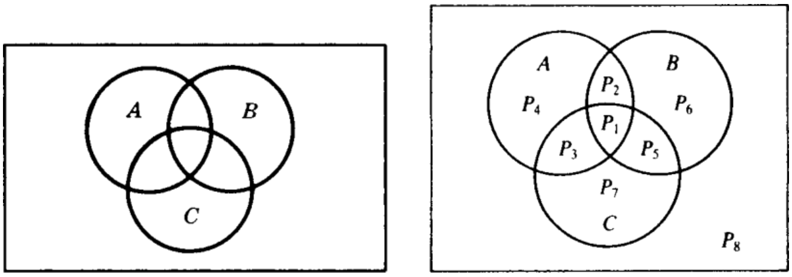
(i) There are m 2^n such fundamental products.
(ii) Any two such fundamental products are disjoint.
(iii) The universal set **U** is the union of all fundamental products.

Thus **U** is the disjoint union of the fundamental products (Problem 1.60). There is a geometrical description of these sets which is illustrated below.

EXAMPLE 1.6 Figure 1-5*(a)* is the Venn diagram of three sets *A*, *B*, *C*. The following lists the $m = 2^3 = 8$ fundamental products of the sets *A*, *B*, C:

$P_1 = A \cap B \cap C,$ $P_3 = A \cap B^C \cap C,$ $P_5 = A^C \cap B \cap C,$ $P_7 = A^C \cap B^C \cap C,$
$P_2 = A \cap B \cap C^C,$ $P_4 = A \cap B^C \cap C^C,$ $P_6 = A^C \cap B \cap C^{C,}$ $P_8 = A^C \cap B^C \cap C^C.$

The eight products correspond precisely to the eight disjoint regions in the Venn diagram of sets *A*, *B*, *C* as indicated by the labeling of the regions in Fig. 1-5*(b).*



(*a*) (*b*)

Fig. 1-5

1.4 ALGEBRA OF SETS, DUALITY

Sets under the operations of union, intersection, and complement satisfy various laws (identities) which are listed in Table 1-1. In fact, we formally state this as:

Theorem 1.5: Sets satisfy the laws in Table 1-1.

Table 1-1 Laws of the algebra of sets

Idempotent laws:	(1a) $A \cup A = A$	(1b) $A \cap A = A$
Associative laws:	(2a) $(A \cup B) \cup C = A \cup (B \cup C)$	(2b) $(A \cap B) \cap C = A \cap (B \cap C)$
Commutative laws:	(3a) $A \cup B = B \cup A$	(3b) $A \cap B = B \cap A$
Distributive laws:	(4a) $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$	(4b) $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
Identity laws:	(5a) $A \cup \varnothing = A$	(5b) $A \cap \mathbf{U} = A$
	(6a) $A \cup \mathbf{U} = \mathbf{U}$	(6b) $A \cap \varnothing = \varnothing$
Involution laws:	(7) $(A^C)^C = A$	
Complement laws:	(8a) $A \cup A^C = \mathbf{U}$	(8b) $A \cap A^C = \varnothing$
	(9a) $\mathbf{U}^C = \varnothing$	(9b) $\varnothing^C = \mathbf{U}$
DeMorgan’s laws:	(10a) $(A \cup B)^C = A^C \cap B^C$	(10b) $(A \cap B)^C = A^C \cup B^C$

Remark: Each law in Table 1-1 follows from an equivalent logical law. Consider, for example, the proof of DeMorgan's Law 10(a):

$(A \cup B)^C = \{x \mid x \not\in (A \text{ or } B)\} = \{x \mid x \not\in A \text{ and } x \not\in B\} = A^C \cap B^C$ Here we use the

equivalent (DeMorgan's) logical law:

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

where $\neg$ means "not," $\vee$ means "or," and means "and." (Sometimes Venn diagrams are used to illustrate the laws in Table 1-1 as in Problem 1.17.)

Duality

The identities in Table 1-1 are arranged in pairs, as, for example, (2a) and (2b). We now consider the principle behind this arrangement. Suppose E is an equation of set algebra. The dual E^* of E is the equation obtained by replacing each occurrence of $\cup$, $\cap$, **U** and $\varnothing$ in E by $\cap$, $\cup$, $\varnothing$, and **U**, respectively. For example, the dual of

$$(\mathbf{U} \cap A) \cup (B \cap A) = A \text{ is } (\varnothing \cup A) \cap (B \cup A) = A$$

Observe that the pairs of laws in Table 1-1 are duals of each other. It is a fact of set algebra, called the *principle of duality*, that if any equation E is an identity then its dual E^* is also an identity.

1.5 FINITE SETS, COUNTING PRINCIPLE

Sets can be finite or infinite. A set S is said to be *finite* if S is empty or if S contains exactly m elements where m is a positive integer; otherwise S is *infinite*.

EXAMPLE 1.7

(a) The set A of the letters of the English alphabet and the set D of the days of the week are finite sets. Specifically, A has 26 elements and D has 7 elements.

(b) Let E be the set of even positive integers, and let **I** be the *unit interval*, that is,

$$E = \{2, 4, 6, \ldots\} \quad \text{and} \quad \mathbf{I} = [0, 1] = \{x \mid 0 \le x \le 1\}$$

Then both E and **I** are infinite.

A set ***S*** is *countable* if ***S*** is finite or if the elements of ***S*** can be arranged as a sequence, in which case ***S*** is said to be *countably infinite*; otherwise ***S*** is said to be *uncountable*. The above set E of even integers is countably infinite, whereas one can prove that the unit interval $\mathbf{I} = [0, 1]$ is uncountable.

Counting Elements in Finite Sets

The notation $n(S)$ or S will denote the number of elements in a set S. (Some texts use #(S) or card(S) instead of $n(S)$.) Thus $n(A)$ 26, where A is the letters in the English alphabet, and $n(D)$ 7, where D is the days of the week. Also $n(\ )$ 0 since the empty set has no elements.
The following lemma applies.

Lemma 1.6: Suppose A and B are finite disjoint sets. Then $A \cup B$ is finite and

$$n(A \cup B) = n(A) + n(B)$$

This lemma may be restated as follows:

Lemma 1.6: Suppose S is the disjoint union of finite sets A and B. Then S is finite and

$$n(S) = n(A) + n(B)$$

Proof. In counting the elements of $A \cup B$, first count those that are in A. There are $n(A)$ of these. The only other elements of $A \cup B$ are those that are in B but not in A. But since A and B are disjoint, no element of B is in A, so there are $n(B)$ elements that are in B but not in A. Therefore, $n(A\ B) \cup n(A) = n(B)$. $+$

For any sets A and B, the set A is the disjoint union of $A\ B$ and $A\ B$.\Thus Lemma 1.6 gives us the following useful result.

Corollary 1.7: Let A and B be finite sets. Then

$$n(A \backslash B) = n(A) - n(A \cap B)$$

For example, suppose an art class A has 25 students and 10 of them are taking a biology class B. Then the number of students in class A which are not in class B is:

$$n(A \backslash B) = n(A) - n(A \cap B) = 25 - 10 = 15$$

Given any set A, recall that the universal set **U** is the disjoint union of A and A^{C}. Accordingly, Lemma 1.6 also gives the following result.

Corollary 1.8: Let A be a subset of a finite universal set **U**. Then

$$n(A^{C}) = n(\mathbf{U}) - n(A)$$

For example, suppose a class **U** with 30 students has 18 full-time students. Then there are $-30\ 18 = 12$ part-time students in the class **U**.

Inclusion–Exclusion Principle

There is a formula for $n(A \cup B)$ even when they are not disjoint, called the Inclusion–Exclusion Principle. Namely:

Theorem (Inclusion–Exclusion Principle) 1.9: Suppose A and B are finite sets. Then $A \cup B$ and $A\ B$ are finite and

$$n(A \cup B) = n(A) + n(B) - n(A \cap B)$$

That is, we find the number of elements in A or B (or both) by first adding $n(A)$ and $n(B)$ (inclusion) and then subtracting $n(A \cap B)$ (exclusion) since its elements were counted twice.

We can apply this result to obtain a similar formula for three sets:

Corollary 1.10: Suppose A, B, C are finite sets. Then $A \cup B \cup C$ is finite and

$$n(A \cup B \cup C) = n(A) + n(B) + n(C) - n(A \cap B) - n(A \cap C) - n(B \cap C) + n(A \cap B \cap C)$$

Mathematical induction (Section 1.8) may be used to further generalize this result to any number of finite sets.

EXAMPLE 1.8 Suppose a list A contains the 30 students in a mathematics class, and a list B contains the 35 students in an English class, and suppose there are 20 names on both lists. Find the number of students: (a) only on list A, (b) only on list B, (c) on list A or B (or both), (d) on exactly one list.

(a) List A has 30 names and 20 are on list B; hence $30 - 20 = 10$ names are only on list A.

(b) Similarly, $35 - 20 = 15$ are only on list B.

(c) We seek $n(A \cup B)$. By inclusion–exclusion,

$$n(A \cup B) = n(A) + n(B) - n(A \cap B) = 30 + 35 - 20 = 45.$$

In other words, we combine the two lists and then cross out the 20 names which appear twice.

(d) By (a) and (b), $10 + 15 = 25$ names are only on one list; that is, $n(A \oplus B) = 25$.

However, if A and B are two arbitrary sets, it is possible that some objects are in A but not in B, some are in B but not in A, some are in both A and B, and some are in neither A nor B; hence in general we represent A and B as in Fig. 1-1*(c)*.

Power Sets

For a given set S, we may speak of the class of all subsets of S. This class is called the *power set* of S, and will be denoted by $P(S)$. If S is finite, then so is $P(S)$. In fact, the number of elements in $P(S)$ is 2 raised to the power $n(S)$. That is,

$$n(P(S)) = 2^{n(S)}$$

(For this reason, the power set of S is sometimes denoted by 2^S.)

EXAMPLE 1.10 Suppose $S = \{1, 2, 3\}$. Then

$$P(S) = [\emptyset, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, S]$$

Note that the empty set $\emptyset$ belongs to $P(S)$ since $\emptyset$ is a subset of S. Similarly, S belongs to $P(S)$. As expected from the above remark, $P(S)$ has $2^3 = 8$ elements.

1.6 MATHEMATICAL INDUCTION

An essential property of the set $\mathbf{N} = \{1, 2, 3, \ldots\}$ of positive integers follows:

Principle of Mathematical Induction I: Let P be a proposition defined on the positive integers $\mathbf{N}$; that is, $P(n)$ is either true or false for each $n \in \mathbf{N}$. Suppose P has the following two properties:

(i) $P(1)$ is true.

(ii) $P(k + 1)$ is true whenever $P(k)$ is true.

Then P is true for every positive integer $n \in \mathbf{N}$.

We shall not prove this principle. In fact, this principle is usually given as one of the axioms when $\mathbf{N}$ is developed axiomatically.

EXAMPLE 1.13 Let P be the proposition that the sum of the first n odd numbers is n^2; that is,

$$P(n) : 1 + 3 + 5 + \cdots + (2n - 1) = n^2$$

(The kth odd number is $2k - 1$, and the next odd number is $2k + 1$.) Observe that $P(n)$ is true for $n = 1$; namely,

$$P(1) = 1^2$$

Assuming $P(k)$ is true, we add $2k + 1$ to both sides of $P(k)$, obtaining

$$1 + 3 + 5 + \cdots + (2k - 1) + (2k + 1) - k^2 + (2k + 1) = (k + 1)^2$$

which is $P(k + 1)$. In other words, $P(k + 1)$ is true whenever $P(k)$ is true. By the principle of mathematical induction, P is true for all n.

There is a form of the principle of mathematical induction which is sometimes more convenient to use. Although it appears different, it is really equivalent to the above principle of induction.

Principle of Mathematical Induction II: Let P be a proposition defined on the positive integers $\mathbf{N}$ such that:

(i) $P(1)$ is true.

(ii) $P(k)$ is true whenever $P(j)$ is true for all $1 \leq j < k$.

Then P is true for every positive integer $n \in \mathbf{N}$.

Remark: Sometimes one wants to prove that a proposition P is true for the set of integers

$$\{a, a + 1, a + 2, a + 3, \ldots\}$$

where a is any integer, possibly zero. This can be done by simply replacing 1 by a in either of the above Principles of Mathematical Induction.

Relations

2.1 PRODUCT SETS

Consider two arbitrary sets A and B. The set of all ordered pairs (a, b) where a A a∈nd b B is called the *product,* or *Cartesian product,* of A and B. A short designation of this product is A 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 **R** denotes the set of real numbers and so $\mathbf{R}^2$=**R** **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, $A \times A = \{(1, 1), (1, 2), (2, 1), (2, 2)\}$

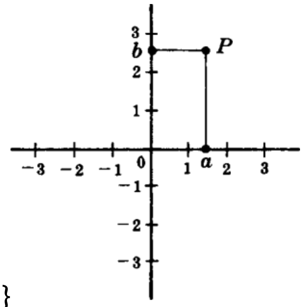


Fig. 2-1

There are two things worth noting in the above examples. First of all A B B×A. T/he C×artesian 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 A_1, a_2 A∈$_2, \ldots,$ a_n ∈A_n is called the *product* of the sets $A_1, \ldots, A_n$ and is denoted by

$$A_1 \times A_2 \times \cdots \times A_n \qquad \text{or } \overset{n}{\underset{i=1}{A_1}}$$

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 spac

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 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) $(a, b) \not\in R$; 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.

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,

$1Ry, 1Rz, 3Ry,$ but $1\not Rx, 2\not Rx, 2\not Ry, 2\not Rz, 3\not Rx, 3\not Rz$

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 $A\ /\subseteq B$.

(c) A familiar relation on the set $\mathbf{Z}$ of integers is "m divides n." A common notation for this relation is to write $m|n$ when m divides n. Thus 6 | 30 but 7 | 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 "$\|$," is a relation on L since either $a \parallel b$ or $a\ /\parallel b$.

(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 O_A or simply O.

(f) Let A be any set. Then A A×and are ∅subsets of A 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

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

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.

2.2 PICTORIAL REPRESENTATIVES OF RELATIONS

There are various ways of picturing relations.

Relations on R

Let *S* be a relation on the set **R** of real numbers; that is, *S* is a subset of $\mathbf{R}^2$ **R** =**R**. F×requently, *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 circ=le having its center at the origin and radius 5. See Fig. 2-2(*a*).

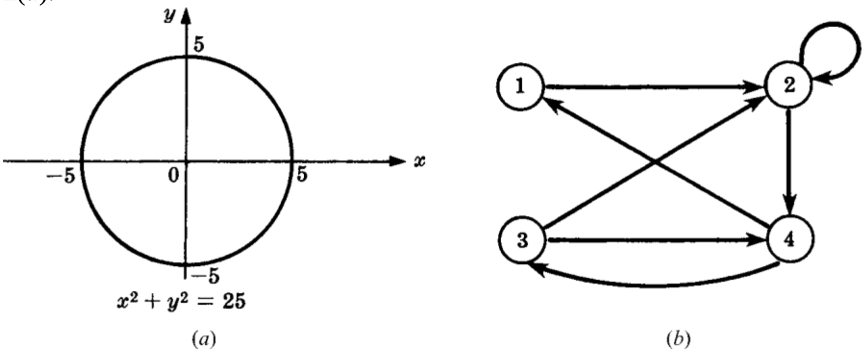


Fig. 2-2

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

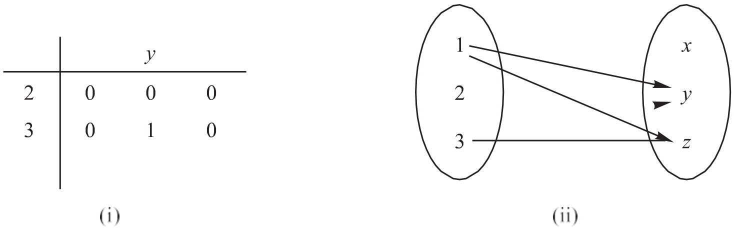
Suppose *A* and *B* are finite sets. There are two ways of picturing a relation *R* from *A* to *B*.

(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.

$$\begin{array}{cccc} & x & & z \\ 1 & 0 & 1 & 1 \end{array}$$



	y		
2	0	0	0
3	0	1	0

(i) (ii)

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

Fig. 2-3

2.3 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 B and S i×s a subset of B C. Then R×and S give rise to a relation from A to C denoted by R S and defined by:

That is ,$a(R\circ S)c$ if for some $b \in B$ we have aRb and bSc.

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

The relation R 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 R, t∘he composition of R with itself, is always defined. Also, R R is s∘ometimes denoted by R^2. Similarly, R^3 R^2 R =R R∘R, =and∘so on. Thus R^n is defined for all positive n.

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

$3(R\circ S)x$ and $3(R\circ S)z$

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

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

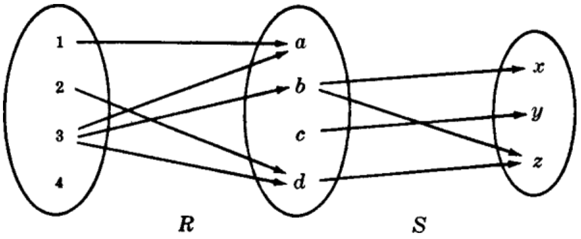


Fig. 2-4

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

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

Multiplying M_R and M_S we obtain the matrix

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

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.

2.4 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 $A = \{1, 2, 3, 4\}$:

$R_1 = \{(1, 1), (1, 2), (2, 3), (1, 3), (4, 4)\}$
$R_2 = \{(1, 1)(1, 2), (2, 1), (2, 2), (3, 3), (4, 4)\}$
$R_3 = \{(1, 3), (2, 1)\}$
$R_4 = \emptyset$, the empty relation
$R_5 = A \times A$, 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 **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 **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, $x \leq x$ for every $x \in \mathbf{Z}$, $A \subseteq A$ for any set $A \in C$, and $n|n$ 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 $(b, a) \not\in R$.

EXAMPLE 2.7

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

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

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

The relation is sym⊥metric since if line a is perpendicular to line b then b is perpendicular to a. Also, is symme∥tric since if line a is parallel to line b then b is parallel to line a. The other relations are not symmetric. For example:

$$3 \leq 4 \text{ but } 4 /\!\leq 3; \quad \{1, 2\} \subseteq \{1, 2, 3\} \text{ but } \{1, 2, 3\} /\!\subseteq \{1, 2\}; \quad \text{and} \quad 2 \mid 6 \text{ but } 6 / \mid 2.$$

A relation R on a set A is *antisymmetric* if whenever aRb and bRa then $a \quad = b$, that is, if $a \quad b$ 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 2. Simil/arly, 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 $3 /= -3$.) 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 $R^{J} = \{(1, 1), (2, 2)\}$ 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 $(a, c) \not\in R$.

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 $(2, 3) \not\in R_3$. 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 $a \nparallel a$. 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 $R^n \subseteq R$.

2.5 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

reflexive(R), symmetric(R), and 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 $O_A = \{(a, a) \mid a \in A\}$ is the diagonal or equality relation on A.

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

(i) $R \cup O_A$ 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 $A = \{1, 2, 3, 4\}$. The

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

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

$$R^{*} = \overset{\infty}{\underset{i=1}{{}_{i}}} \quad R$$

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:

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

Accordingly,

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

2.6 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

$$a \equiv b \ (\text{mod } m)$$

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

$$11 \equiv 3 \ (\text{mod } 4) \text{ and } 22 \equiv 6 \ (\text{mod } 4)$$

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 /= A_j then $A_i \cap A_j = \varnothing$.

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 $[a]$ /= $[b]$, 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 **Z** of integers denoted by

$$x \equiv y \ (\text{mod } 5)$$

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

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

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, **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.

2.7 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 **R** of real numbers is reflexive, antisymmetric, and transitive. Thus is a partial ordering on **R**.

(c) The relation "*a* divides *b*," written $a \mid b$, is a partial ordering on the set **N** of positive integers. However, "*a* divides *b*" is not a partial ordering on the set **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 /= -3.

2.8 *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 **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 $O = (0, 0, 0)$.

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 in$\in$to B.” If a 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 A B is denoted by Ran(f), 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 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

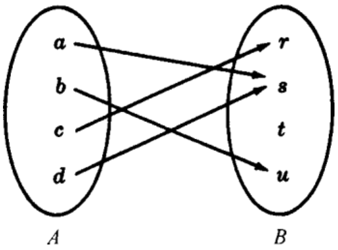


Fig. 3-1

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

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

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 $i\colon S \rightarrowtail A$ 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 A 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 A B and g A B are defined to be *equal*, written f g, if $f(a)$ $g(a)$ for every a ∈ 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 A B is a relation from A to B (i.e., a subset of A B) such that each a 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 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 $f\colon A \to B$ 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 $A = \{1, 2, 3\}$:

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

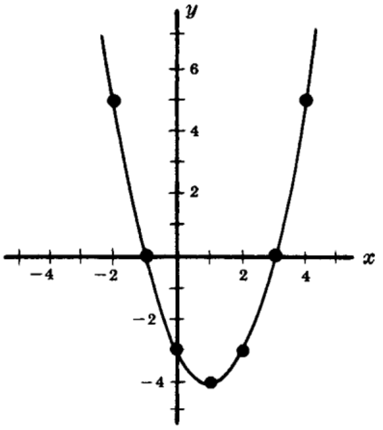
f is a function from A into A since each member of A appears as the first coordinate in exactly one ordered pair in f; here $f(1)$ 3, $f(2)$ 3, and $f(3)$ 1. g is not a function from A into A since 2 A is not the first coordinate of any pair in g and so g does not assign any image to 2. Also h is not 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 $f\colon \mathbf{R} \to \mathbf{R}$ 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 **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.

3.1 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 $f(a) = f(a^{\rfloor})$ implies $a = a^{\rfloor}$.
A functiọn f→A B is said to be an *onto* function if each element of B is the image of some element of A. In other wọrds, f A 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 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.

Ịf f→A 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 : B→f_2 B : C→f_3 C D and→f_4 D 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 $f_4(v) = f_4(w)$

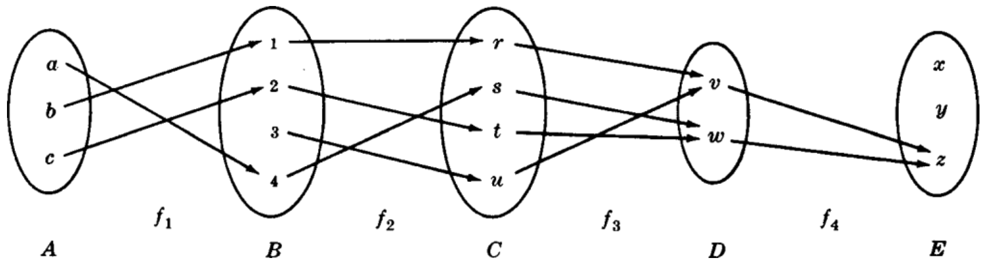


Fig. 3-3

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 $\mathbf{R} \rightarrow$ $\mathbf{R}$. 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** → **R** is one-to-one if each horizontal line intersects the graph of f in at most one point.

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

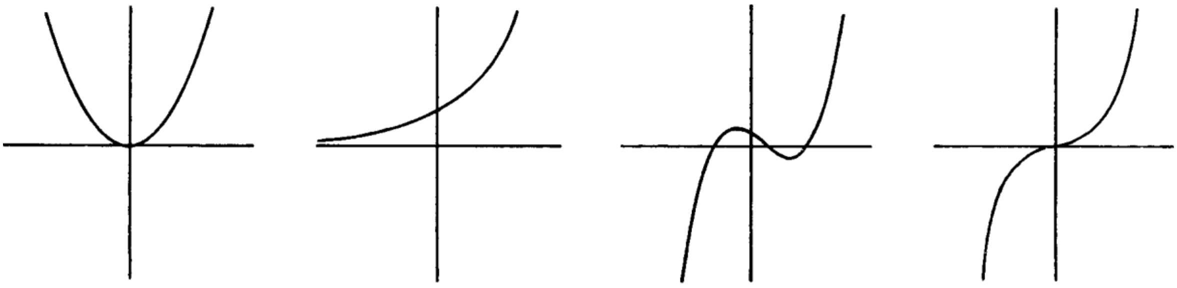
Accordingly, if f is both one-to-one and onto, i.e. invertible, then each horizontal line will intersect the graph of

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., $f_4^{-1}(x) = \sqrt[3]{x}$.



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

Fig. 3-4

Permutations

An invertible (bijective) function $\sigma: X \rightarrow X$ 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 $.n(n\ .\ 1)$ 3 2 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 $\tau = 643125$.

3.2 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,

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

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

Integer and Absolute Value Functions

Let x be any real number. The *integer value* of x, written 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 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. Thu

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

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

DISCRETE MATHEMATICS

Branch: CSE 4th SEM Code : CSE 208C

Unit II

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 Suppose a college has 3 different history courses, 4 different literature courses, and 2 different sociology courses.

(a) The number m of waysa 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

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

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 .2

(a) $3! = 3 \cdot 2 \cdot 1 = 6$, $4! = 4 \cdot 3 \cdot 2 \cdot 1 = 24$, $5 = 5 \cdot 4! = 5(24) = 120$.

(c) $\frac{12 \cdot 11 \cdot 10}{\cdot 2 \cdot 1} = \frac{12 \cdot 11 \cdot 10 \cdot 9!}{3 \cdot 2 \cdot 1 \cdot 9!} = \frac{12!}{3!\,9!}$

(d)

and, more generally,

$$\frac{n(n-1)\cdots(n-r+1)}{r(r-1)\cdots 3 \cdot 2 \cdot 1} = \frac{n(n-1)\cdots(n-r+1)(n-r)!}{r(r-1)\cdots 3 \cdot 2 \cdot 1 \cdot (n-r)!} = \frac{n!}{r!(n-r)!}$$

(e) For large n, one uses Stirling's approximation (where $e = 2.7128\ldots$):

$$n! = \sqrt{2\pi n}\, n^n e^{-n}$$

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:

$$\binom{n}{r} = \frac{n(n-1)\cdots(n-r+1)}{r(r-1)\ldots 3 \cdot 2 \cdot 1} \quad \text{or equivalently} \quad \binom{n}{r} = \frac{n!}{r!(n-r)!}$$

Note that $n - (n - r) = r$. This yields the following important relation.

Lemma .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:

$$\binom{n}{} = \frac{n!}{} = 1 \quad \text{and} \quad \binom{0}{} = \frac{0!}{} = 1$$

0 0!n! 0 0! 0!

EXAMPLE 3

(a) $\binom{8}{2} = \frac{8 \cdot 7}{2 \cdot 1} = 28$; $\binom{9}{4} = \frac{9 \cdot 8 \cdot 7 \cdot 6}{4 \cdot 3 \cdot 2 \cdot 1} = 126$; $\binom{12}{5} = \frac{12 \cdot 11 \cdot 10 \cdot 9 \cdot 8}{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1} = 792$.

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) 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.

$$
\begin{aligned}
(a+b)^0 &= 1\\
(a+b)^1 &= a + b\\
(a+b)^2 &= a^2 + 2ab + b^2\\
(a+b)^3 &= a^3 + 3a^2b + 3ab^2 + b^3\\
(a+b)^4 &= a^4 + 4a^3b + 6a^2b^2 + 4ab^3 + b^4\\
(a+b)^5 &= a^5 + 5a^4b + 10a^3b^2 + 10a^2b^3 + 5ab^4 + b^5\\
(a+b)^6 &= a^6 + 6a^5b + 15a^4b^2 + 20a^3b^3 + 15a^2b^4 + 6ab^5 + b^6\\
&\cdots\cdots\cdots\cdots\cdots\cdots
\end{aligned}
$$

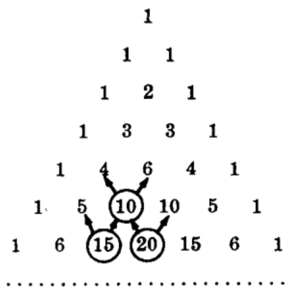


Fig. -1 Pascal's triangle

Theorem 3: $\binom{n+1}{r} = \binom{n}{r-1} + \binom{n}{r}$.

5.1 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

$P(n, r)$ (other texts may use ${}_nP_r$, $P_{n,r}$, or $(n)_r$).

The following theorem applies.

Theorem 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 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:

$$\text{——}, \text{——}, \text{——}$$

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 5 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 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: 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, ACB, BAC, BCA, CAB, 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 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

$$B_1B_2B_3AY,\ B_2B_1B_3AY,\ B_3B_1B_2AY,\ B_1B_3B_2AY,\ B_2B_3B_1AY,\ B_3B_2B_1AY$$

produce the same word when the subscripts are removed. The 6 comes from the fact that there are 3 3·2·1 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 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 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:

$$m = P(52, 3) = 52(51)(50) = 132\ 600$$

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 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:

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

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 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 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:

$$m = \binom{6}{3}\binom{5}{2}\binom{8}{4} = \frac{6 \cdot 5 \cdot 4}{3 \cdot 2 \cdot 1} \cdot \frac{5 \cdot 4}{2 \cdot 1} \cdot \frac{8 \cdot 7 \cdot 6 \cdot 5}{4 \cdot 3 \cdot 2 \cdot 1} = 20 \cdot 10 \cdot 70 = 14\,000$$

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 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 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.

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 $n(A\ \ B)$ of elements in the union of A and B, we add $n(A)$ and $n(B)$ and then we subtract $n(A\ \ B)$; that is, we “includ∪e” $n(A)$ and $n(B)$, and we “exclude” $n(A\ \ B)$. This follows from the fact that, when we ad∩d $n(A)$ and $n(B)$, we have counted the elements of $(A\ B)$ twice. ∩

The above principle holds for any number of sets. We first state it for th∩ree sets.

Theorem 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 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,

$$n(F \cup G \cup R) = n(F) + n(G) + n(R) - n(F \cap G) - n(F \cap R) - n(G \cap R) + n(F \cap G \cap R)$$

$$= 65 + 45 + 42 - 20 - 25 - 15 + 8 = 100$$

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 .9: $n(A_1 \cup A_2 \cup \cdots \cup A_m) = s_1 - s_2 + s_3 - \cdots + (-1)^{m-1}s_m$.

.

Solved Questions

1.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, $n = 5 \cdot 3 \cdot 6 \cdot 4 = 360$.

2.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.

(b) Here the Sum Rule is used; hence, $n = 8 + 6 = 14$.

(c) Here the Product Rule is used; hence, $n = 8 \cdot 6 = 48$.

(d) There are 14 ways to elect the president, and then 13 ways to elect the vice president. Thus $n = 14 \cdot 13 = 182$.

3.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.

(e) There are 4 ways to go from *A* to *B* and 3 ways from *B* to *C*; hence $n = 4 \cdot 3 = 12$.

(f) There are 12 ways to go from *A* to *C* by way of *B*, and 12 ways to return. Thus $n = 12 \cdot 12 = 144$.

(g) The man will travel from *A* to *B* to *C* to *B* to *A*. Enter these letters with connecting arrows as follows:

$$A \rightarrow B \rightarrow C \rightarrow B \rightarrow A$$

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:

$$A \xrightarrow{4} B \xrightarrow{3} C \xrightarrow{2} B \xrightarrow{3} A$$

Thus, by the Product Rule, $n = 4 \cdot 3 \cdot 2 \cdot 3 = 72$.

3.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.

4. 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.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.

6.Find the number *n* of distinct permutations that can be formed from all the letters of each word:

(b) THOSE; (*b*) *UNUSUAL*; (*c*) *SOCIOLOGICAL.*

This problem concerns permutations with repetitions.

(a) $n = 5! = 120$, since there are 5 letters and no repetitions.

!

7.A class contains 10 students with 6 men and 4 women. Find the number *n* of ways to:

(c) Select a 4-member committee from the students.

(d) Select a 4-member committee with 2 men and 2 women.

(e) 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,

$$n = P(6, 3) = 6 \cdot 5 \cdot 4 = 120$$

8.A box contains 8 blue socks and 6 red socks. Find the number of ways two socks can be drawn from the box if:

(f) 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.

9.Find the number *m* of committees of 5 with a given chairperson that can be selected from 12people.

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 $m = 12 \cdot C(11, 4) = 12 \cdot 330 = 3960$.

10.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.

(g) 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.

(h) 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.

11.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.

12.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).

(i) Show that *L* has a sublist consisting of four or more consecutive consonants.

(j) 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.

13.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 $t = 22 + 18 = 40$.

14.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:

(k) 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:

$$(a)\ m = n(P \cap B) = n(P) + n(B) - n(P \cup B) = 30 + 14 - 32 = 12$$

$$(b)\ m = n(P \setminus B) = n(P) - n(P \cap B) = 30 - 12 = 18$$

$$(c)\ m = n(B \setminus P) = n(B) - n(P \cap B) = 14 - 12 = 2$$

RECURRENCE RELATIONS

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.

(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,... of the recurrence relation $a_k = 2a_{k-1}$.

EXAMPLE

(a) Arithmetic Progression

An arithmetic progression is a sequence of the form

$$a, a + d, a + 2d, a + 3d, \dots$$

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,...

(ii) $a = 2, d = 5$: 2, 7, 12, 17,...

(iii) $a = 1, d = 0$: 1, 1, 1, 1, 1,...

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

$$a, ar, ar^2, ar^3, \dots$$

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,...

(ii) $a = 5, r = 2$: 5, 10, 20, 40,...

(iii) $a = 1, r = \frac{1}{2}$: $1, \frac{1}{2}, \frac{1}{4}, \frac{1}{8}$.

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 $a_{n+1} = ar^n$.

LINEAR RECURRENCE RELATIONS WITH CONSTANT COEFFICIENTS

A *recurrence relation of order k* is a function of the form

$$a_n = \alpha(a_{n-1}, a_{n-2}, \dots, a_{n-k}, n)$$

that is, where the *n*th term a_n of a sequence is a function of the preceding *k* terms $a_{n-1}, a_{n-2}, \dots, a_{n-k}$ (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 $C_1, C_2, \dots, C_k$ 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 $C_1 C_2, \dots, C_k$ 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 $a_{n-1}, a_{n-2}, \dots, a_{n-k}$. 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 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.

6.7. 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 $O(t)$ and its roots:

$$O(x) = x^2 - x - 2 = (x - 2)(x + 1); \quad \text{roots } r_1 = 2, r_2 = -1$$

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:

For $n = 0$, $a_0 = 2$, we get: $c_1(2^0) + c_2(-1)^0 = 2$ or $c_1 + c_2 = 2$
For $n = 1$, $a_1 = 7$, we get: $c_1(2^1) + c_2(-1)^1 = 7$ or $2c_1 - c_2 = 7$

Solving the two equations for c_1 and c_2 yields $c_1 = 3$ and $c_2 = 1$. The unique solution follows:

$$a_n = 3(2^n) - (-1)^n$$

6.8. 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

$$O(x) = x^3 - 6x^2 + 12x - 8 = (x - 2)^3$$

Then $O(x)$ 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:

For $n = 0, a_0 = 3$ we get: $c_1 = 3$
For $n = 1, a_1 = 4$ we get: $2c_1 + 2c_2 + 2c_3 = 4$
For $n = 2, a_2 = 12$ we get: $4c_1 + 8c_2 + 16c_3 = 12$

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:

$$a_n = (3 - 2n + n^2)(2^n)$$

Propositional logic

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.

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, $p \wedge q$

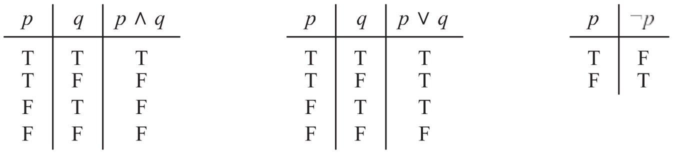
Any two propositions can be combined by the word “and” to form a compound proposition called the *conjunction* of the original propositions. Symbolically,

$$p \wedge q$$

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 q may∨be defined equivalently by the table in Fig. 4-1(*b*). Observe that p 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)$2 + 2 /= 5$

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:

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

4.2 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 , , and (a∧nd∨others ¬discussed subsequently). Such an expression $P(p,q,\ldots)$ will be called a *proposition*.

The main property of a proposition *P (p,q,.. .)* 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 *(p ¬ q)*∧ Figure 4-2(*a*) indicates how the truth table of *(p¬ ∧q)¬* is constructed. Observe that the first columns of the table are for the variables *p,q,*... 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 , , . Finally∧ ∨ ¬ we obtain the truth value of the proposition, which appears in the last column.

The actual truth table of the proposition *(p ¬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,

¬ has precedence over ∧ which has precedence over ∨

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 (p	¬		∧	¬	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	(p ¬		∧	¬	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	(p ¬		∧	¬	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

4.3 TAUTOLOGIES AND CONTRADICTIONS

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

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

4.4 LOGICAL EQUIVALENCE

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

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

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

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

4.6 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 ∨and ¬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

4.7 ARGUMENTS

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

$$P_1, P_2, ..., P_n \text{ € } Q$$

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

Definition 4.4: An argument $P_1, P_2, \ldots, P_n$ € 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:

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

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:

$$p \to q, q \text{ € } p$$

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.

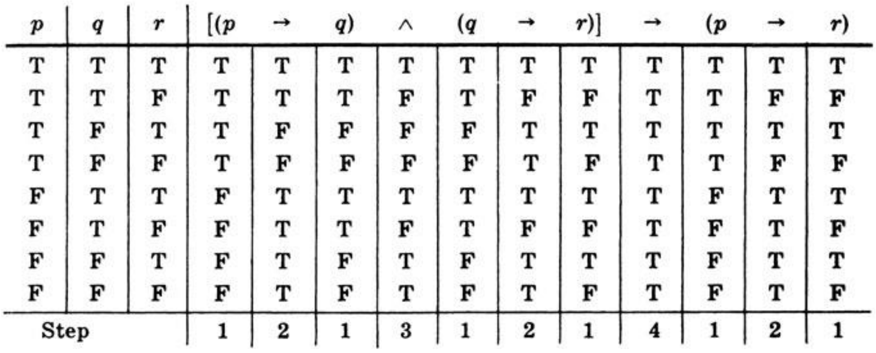
Now the propositions $P_1, P_2,\ldots, P_n$ are true simultaneously if and only if the proposition $P_1 \wedge P_2 \wedge \ldots P_n$ is true. Thus the argument $P_1, P_2,\ldots, P_n$ € Q is valid if and only if Q is true whenever $P_1 \wedge P_2 \wedge \ldots \wedge P_n$ is true or, equivalently, if 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 $P_1, P_2,\ldots, P_n$ € 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:

$p \rightarrow q, q \rightarrow r$ € $p \rightarrow r$ ***(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 € S is valid. For the argument is of the form

$$p \rightarrow q, q \rightarrow r \text{ € } p \rightarrow r$$

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.

4.8 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 A. That is, $p(x)$ becomes a statement (with a truth value) whenever any element a 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,

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

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 **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 **Ø**. That is, $p(x)$ is not true for any integer in **N**.

(c) Let $p(x)$ be "$x + 5 > 1$." Its truth set is **N**. That is, $p(x)$ is true for every element in **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 A, for some x A, or for no x 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, $x\, p(x)$, that is $p(x)$ preceded by the quantifier , 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 $\{n \mid n + 2 > 8\} = \{7, 8, \ldots\} /= \mathbf{N}$.

(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

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

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

Q_2: If $\{x \mid p(x)\} \, / = \varnothing$ 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 $\{n \mid n + 4 < 7\} = \{1, 2\} \, / = \varnothing$.

(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:

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

4.9 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

$\neg(\forall x \in M)(x$ is male$) \equiv (\exists \; x \in M)$ (x 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:

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

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:

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

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 is cha∀nged to and each∃ is change∃d to as the nega∀tion symbol passes through¬ the statement from left to right. For example,

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

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,… 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,... 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

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

(a) It is not cold. (*c*) It is cold or it is raining.

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

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

4.2. 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.

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

Construct the truth tables for $(p\ q)$ and $p\ q$ 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 $(p\ q)$ and $p\ q$ 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

4.4. 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

4.5. 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, $p \rightarrow q \equiv \neg p \vee q$. Hence,

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

(b) Productivity does not increase or wages rise.

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

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

				Conditional	Converse	Inverse	Contrapositive
p	q	$\neg p$	$\neg q$	$p \rightarrow q$	$q \rightarrow p$	$\neg p \rightarrow \neg q$	$\neg q \rightarrow \neg p$
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

4.7. 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 $p \rightarrow q$ is $\neg q \rightarrow \neg p$. 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.

4.8. 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 $\neg(p \rightarrow q) \equiv p \wedge \neg q$; 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 $\neg(p \to \neg q) \equiv p \wedge \neg\neg q \equiv p \wedge q$. Hence the negation of the statement is:

It snows and they drive the car.

ARGUMENTS

4.9. Show that the following argument is a fallacy: $p \to q, \neg p$ € $\neg q$.

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

p	q	$p \to q$	$\neg p$	$(p \to q) \wedge \neg p$	$\neg q$	$[(p \to q) \wedge \neg p] \to \neg p$
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

4.10. Determine the validity of the following argument: $p \to q, \neg p$ € $\neg p$.

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

p	q	$[(p$	$\to$	$q)$	$\wedge$	$\neg$	$q]$	$\to$	$\neg$	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

4.11. Prove the following argument is valid: $p \to \neg q, r \to q, r$ € $\neg p$.

Construct the truth table of the premises and conclusions as in Fig. 4-14(a). Now, $p \to \neg q$, $r \to q$, 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	$p \to \neg q$	$r \to q$	$\neg q$
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	$\neg q$	$p \to \neg q$	$\neg p$
T	T	F	F	F
T	F	T	T	F
F	T	F	T	T
F	F	T	T	T

(b)

Fig. 4-14

4.12. 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

$p \rightarrow \neg q,\ \neg q$ € 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 $p \rightarrow \neg q$ 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.

4.13. 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 $p \rightarrow q, \neg p$€ $\neg\ q$, 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

4.14. 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.

4.15. 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.

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

4.17. 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,. . .; (*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.

4.18. 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.)

SUBJECT:DISCRETE MATHEMATICS Code:CSE 208-C

B.Tech CSE 4th SEM

UNIT : Algebraic Structure & Morphism

Definition:

1.Algebraic Structure: A system consists of a non empty set and one or more operstion on that set is called Algebraic system or Algebraic Structure.

2. Binary Operations: A non empty set A such that f: AXA→A is called a binary operation on A.

If a,b ϵA & * is a binary operation on A , then it is denoted by a*b .

For Example ☹i)The operation of addition on the set of Natural Numbers.

(ii)The operation of substraction on the set of integers is a binary operation but not with the set of Natural numbers. Because

Table of Operation: Let A ={a1,a2,a3……….an} be a non empty set. '*' be a binary operation on A .Then

*	a_1	a_2		a_n
a_1	a_1*a_1			a_1*a_n
a_2		a_2*a_2		
a_n	a_n*a_1			a_n*a_n

Example:

Consider A={1,2,3} and a binary operation * on A is defined as a*b= 2a + 2b . Represents its table

*	1	2	3
1	4	6	8
2	6	8	10
3	8	10	12

Viz: 1*1=2.1+2.1=4; 1*2=2.1+2.2=6; etc…

Properties of Binary Operations

Let A be a non empty set with a binary operation * . Then

(1) Closure property: If a,b ϵA , then a*bϵA
Viz:The operation of addition on the set of integers is a closed operation.

(2) **Associative Property:** If for every a,b,cϵA , WE HAVE

(a*b)*c=a*(b*c)

Example: Let * be a binary operation on set of rational number Q defined as

a*b=a+b-ab

Then * is Associative because

If ∀a,b,c ϵQ

(a*b)*c = (a+b-ab)*c = (a+b-ab)+c –(a+b-ab)c =a+b-ab+c-ac-ab+abc=a+b+c-ab-bc-ca+abc

a*(b*c)=a*(b+c-bc)=a+b+c-bc-a(b+c-bc)=a+b+c-ab-bc-ca+abc

Therefore,(a*b)*c=a*(b*c)

(3) **Commutative Property:** If for all a,bϵA , a*b=b*a

Example:

* on set Q ,set of Rational numbers defined as a*b=$a^2 + b^2$,∀a,b ϵ Q

Is commutative, because

∀a,b ϵ Q , a*b= $a^2 + b^2 = b^2 + a^2$ = b* a

(4) **Identity Property :** If there exists an element e ϵ A such that

a * e =e *a ∀a ϵ A

(Right Identity) = (Left Identity)

Theorem : Prove that $e_1 = e_2$, where e1 amd e2 are the R.H.I and L.H.I of binary operation *.

Proof: Since e1 is the right identity

Therefore, e1* e2 = e1

Again since, e2 is the left identity

Therefore , e1=e2, Hence proved.

(5) **Inverse Property :** The binary operation * has the inverse property if for each a ϵ A
∋ an element b ϵ A such that a * b =b * a = e…… e is the identity in *

Then b is said to be invers of a.

(6) **Idempotent :** The operation * has the Idempotent Type equation here.Property if for every aϵ A , we have

a*a=a , ∀a ϵ A

(7) **Distributivity :** Let two binaryboperations * and + on A , then * distributes over + ,if for every a, b ,c ϵ A

We have

a *(b + c) = (a *b) +(a *c) [Left Distribution]

(b + c) * a = (b* c) +(c*a)   [Right Distribution]

(8) **Cancellation :** The operation * has the cancellation Property , if for every a, b, c ϵ A we have

a * b =a * c implies b=c (Left cancellation)

b * a = c * a implies b=c (Right cancellation)

(9) Semi group : An Algebraic Structure (A,*) is said to be a semi group if it satisfies the following properties

(i) The operation * is closed on set A.

(ii) The operation * is an associative operation.

Example: If A= { 1,3,5,7,9………} set of all odd positive integers and '*' is a ordinary multiplication

Then (A,*) is a semi group.

(10) Monoid : An algebraic structure (A ,o) , where 'o' is a binary operation on A is said to be Monoid if it satisfies the following properties

(i) The operation 'o' is a closed operation on set A.

(ii) The operation 'o' is an associative operation.

(iii) There exists an identity element w.r.t the operation 'o'.

Example: An algebraic structure (N,+) , where N is the set of Natural Numbers with addition as a binary operation is a Monoid.

(11) Group : An Algebraic Structure (G,*) ,where '*' is a binary operation on G. Then the system (G,*) is said to be a group if it satisfies the following properties ,

(i) The operation '*' is a closed operation.

(ii) the operation '*' is an associative operation.

(iii) There exists an identity element w.r.t the operation '*'.

(iv) For every aϵ G there exits an element a^{-1} ϵ G such that

$a^{-1} * a= a* a^{-1}$ =e ,e is the identity element in G.

Example : The algebraic structure (I,+) is a group ,where I is the set of integers and "+" is the operation addition. Here 0 is the identity element and for every a ϵ I , -a ϵ I is the inverse .

(12) Sub group: Let (G,*) be a group and let S ⊆ G ,Then (S,*) is called a subgroup if it satisfies the following conditions :

(ii) The operation * is an associative operation.

(iii) If e ϵ G , identity element in G , then e ϵ S .

(iv) For every a ϵ S , a^{-1} ϵ S .

Example : (I,+) is a group ,where I is the set of integers ,the the algebraic structure (H,+) , where H is the set of even integers is the subgroup of I.

13) Abelian Group : A Group (G, *) is said to be Abelian Group if

a * b = b*a , ∀a,b ϵ G.

xample: Consider an algebraic system (G,*) where G is the set of all non zero real numbers and * is a binary operation defined by

$a * b =\frac{ab}{4}$ Show that (G,*) is an abelian group.

f: (i) Closure Property : The set G is closed under the operation * ,Since a*b=(ab/4) is a real number ,hence belongs to G.

(ii) Associative Property : Let a,b,c ∈ G then

(a*b)*c (ab/4)*c= (ab)c/16==abc/16

a *(b*c)= a*(bc/4)=a(bc)/16=abc/16

Hence (a*b)*c= a*(b*c)

iii) Identity: let e be a positive real Number then e a=a i.e ea/4=a i.e e=4

arly. a * e=4 I,ae e=4

the identity element in G.

nverse : Let a ∈ G , If a^{-1}∈ G, then a * a^{-1} =4

Hence $aa^{-1}/4 = 4$ or $a^{-1} = 16/a$

arly,,

$a^{-1} * a = 4$ or $a^{-1} = 16/a$

, the inverse of the element a in G

5/a

he * in G is commutative

e a*b=ab/4=ba/4= b*a

efore, the system (G,*) is an abelian Group.

Hence Proved

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

(a) Let *A* and *B* denote, respectively, the set of even and odd positive integers. Then $(A, \times)$ and $(B, \times)$ are subsemigroups of $(\mathbf{N}, +)$ since *A* and *B* are closed under multiplication. On the other hand, $(A, +)$ is a subsemigroup of (**N**,) since *A* is closed under addition, but (*B*,) is not a subsemigroup of (**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 : 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 $\sim$.

Homomorphism of Semigroups

Consider two semigroups $(S, *)$ and $(S^J, *^J)$. A function $f : S \to S^J$ is called a *semigroup homomorphism*

or, simply, a *homomorphism* if

$$f(a * b) = f(a) *^J f(b) \quad \text{or, simply} \quad f(ab) = f(a)f(b)$$

Suppose f is also one-to-one and onto. Then f is called an *isomorphism* between S and S^J, and S and S^J are said to be *isomorphic* semigroups, written $S \cong S$.

EXAMPLE

(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 $(M, \quad)$, the matrices under matrix multi- plication. On the other hand, the determinant function is not additive, that is, for some matrices,

$$\det(A + B) \ /= \det(A) + \det(B)$$

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 $f : \mathbf{Z_4} \to S$ be defined by

$$f(0) = 1,\ f(1) = 3,\ f(2) = 9,\ f(3) = 7$$

	0	1	2	3
0	0	1	2	3
1	1	2	3	0

	1	3	7	9
1	1	3	7	9
3	3	9	1	7

+				
2	2	(a) 3	0	1
3	3	0	1	2

×				
7	7	(b) 1	9	3
9	9	7	3	1

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 φ: $S \to S/\sim$ be the *natural mapping* from S into the factor semigroup $S/\sim$ defined by

$$\varphi(a) = [a]$$

That is, each element a in S is assigned its equivalence class $[a]$. Then φ is a homomorphism since

$$\varphi(ab) = [ab] = [a]\,[b] = \varphi(a)\varphi(b)$$

Fundamental Theorem of Semigroup Homomorphisms

Recall that the image of a function f: S S^{J}, written $f(S)$ of Im f, consists of the images of the elements of S under f. Namely:

$$\text{Im}\, f = \{b \in S^{J} \mid \text{there exists } a \in S \text{ for which } f(a) = b\}$$

The following theorem (proved in Problem B.5) is fundamental to semigroup theory.

Theorem : Let f : $S \to S^{J}$ 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

(a) Let F be the free semigroup on $A = \{a, b\}$. The function f : $F \to \mathbf{Z}$ 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.

5 5 1 11 7

7 7 11 1 5

11 11 7 5 1

(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. **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:

$$\varepsilon = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix}, \quad \varphi_1 = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}$$

$$\sigma_1 = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \\ & & 1 \end{pmatrix}, \quad \sigma_3 = \begin{pmatrix} & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix}, \quad \varphi_2 = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix}$$

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 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.1 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**
$H = \{\varepsilon, \sigma_1\}$	$H = \{\varepsilon, \sigma_1\}$

$H\varphi_1 = \{\varphi_1, \sigma_2\}$ $\varphi_1 H = \{\varphi_1,$
$\sigma_3\}$ $H\varphi_2 = \{\varphi_2, \sigma_3\}$ $\varphi_2 H$
$= \{\varphi_2, \sigma_2\}$

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*.

$\bar{0} = 0 + H = H = \{\ldots, -10, -5, 0, 5, 10, \ldots\}$

+	$\bar{0}$	$\bar{1}$	$\bar{2}$	$\bar{3}$	$\bar{4}$
$\bar{0}$	$\bar{0}$	$\bar{1}$	$\bar{2}$	$\bar{3}$	$\bar{4}$

$$\overline{2} = 2 + H = \{\ldots, -8, -3, 2, 7, 12, \ldots\}$$

$$\overline{3} = 3 + H = \{\ldots, -7, -2, 3, 8, 13, \ldots\}$$

$$\overline{4} = 4 + H = \{\ldots, -6, -1, 4, 9, 14, \ldots\}$$

(*a*)

(*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:

$$\varphi_1{}^1 = \varphi_1, \varphi_1{}^2 = \varphi_2, \varphi_1{}^3 = \varphi_2 \cdot \varphi_1 = e$$

Hence $\varphi_1 = 3$ and $gp(\varphi_1) = \{e, \varphi_1, \varphi_2\}$. Observe that φ_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 $A\,A^{-1}$ (where A^{-1} denotes the set of inverses of elements of A). That is,

$$gp(A) = \{x \in G \cdot x = b_1 b_2 \ldots b_m \text{ where each } b_i \in A \cup A^{-1}\}$$

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 $b = \varphi_1$ form a minimal set of generators of the symmetric group S_3 (Fig. B-4). Specifically,

$$e = a^2, \sigma_1 = a, \sigma_2 = ab, \sigma_3 = ab^2, \varphi_1 = b, \varphi_2 = b^2$$

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^J 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^J are said to be *isomorphic*,

written $G \cong G^J$.
J

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^J of G^J; that is,

$$\text{Ker}\, f = \{a \in G \mid f(a) = e^J\}$$

Recall that the image of f, written $f(G)$ or Im f, consists of the images of the elements under f; that is,

$$\text{Im}\, f = \{b \in G^J \mid \text{there exists } a \in G \text{ for which } f(a) = b\}\ .$$

The following theorem (proved in Problem B.19) is fundamental to group theory.

Theorem B.9: Suppose f: $G \rightarrow G'$ 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: $G \rightarrow G'$ 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: $\mathbf{Z} \rightarrow G$ 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 {multiples of m}, and so $gp(a)$ $\mathbf{Z}_m$. 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.2 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:

[**R1**] For any $a, b, c \in R$, we have $(a + b) + c = a + (b + c)$.

[**R2**] There exists an element $0 \in R$, called the *zero* element, such that, for every $a \in R$,

$$a + 0 = 0 + a = a.$$

[**R3**] For each $a \in R$ there exists an element $-a \in R$, called the *negative* of a, such that

$$a + (-a) = (-a) + a = 0.$$

[**R4**] For any $a, b \in R$, we have $a + b = b +$

a. [**R5**] For any $a, b, c \in R$, we have $(ab)c =$

$a(bc)$.

[**R₆**] 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 [**R₁**] through [**R₄**] may be summarized by saying that R is an abelian group under addition.

Subtraction is defined in R by $-ab = a + (-b)$.

One can prove (Problem B.21) that $a = 0 \cdot = 0 \quad a \quad \in 0$ for every $a \quad R$.

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 *a*=*b ba* for every *a*, *b*∈*R*.

R is called a *ring with an identity element 1* if the element 1 has the property= that *a*· 1 1 *a a* for every ele∈ment *a R*. In such a case, an ele∈ment *a R* is called a *unit* if *a* has a multiplicative inverse, that is, an element a^{-1} in *R* such that *a* $\bar{\bar{a}}^{-1}$ a^{-1} *a* 1.

R is called a *ring with zero divisors* if there exist nonzero elements∈*a*, *b 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 $a \;/\!= 0$, 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 **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 **Z** are only 1 and 1, that is, no other element in **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 $\mathbf{Z}_6$,

$$2 \cdot 3 = 0 \text{ but } 2 \;/\!\equiv 0 \;(\text{mod } 6) \text{ and } 3 \;/\!\equiv 0 \;(\text{mod } 6)$$

(c) The rational numbers **Q** and the real numbers **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 inte[]gral 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 co{ets a |+ $J \in a$ R form a ring under the coset operations

$$(a + J) + (b + J) = a + b + J \text{ and } (a + J)(b + J) = ab + J$$

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 **Z** of integers. Then every ideal J in **Z** is a principal ideal, that is, $J = (m) = m\mathbf{Z}$, for some integer m. Thus **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 **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^J is called a *ring homomorphism* or, simply, *homomorphism* if, for every $a, b \in R$,

$$f(a + b) = f(a) + f(b), f(ab) = f(a)f(b)$$

In addition, if f is one-to-one and onto, then f is called an *isomorphism*; and R and R^J are said to be *isomorphic*,

written $R \cong R^J$.

Suppose $f : R \to R^J$ 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^J; 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 **Z** of integers is the classical example of a unique factorization domain. The units of **Z** are 1 and -1. The only associates of $n \in Z$ are n and $-n$. The irreducible elements of **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.3 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,

$$f = (\ldots, 0, a_n, \ldots, a_1, a_0) \text{ or, equivalently, } f(t) = a_n t^n + \cdots + a_1 t + a_0$$

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

$$f(t) = a_n t^n + \cdots + a_1 t + a_0 \text{ and } g(t) = b_m t^m + \cdots + b_1 t + b_0$$

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

$$f(t) = a_0 \text{ or } a_0 = (\cdots, 0, 0, a_0)$$

Then the operators of addition and scalar multiplication are preserved by this identification. Thus, the mapping

$\psi: K \to K[t]$ 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

$$f(t) = a_n t^n + \cdots + a_1 t + a_0 \qquad \text{and } g(t) = b_m t^m + \cdots + b_1 t + b_0$$

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 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 **Z**.

Theorem B.15 (Euclidean Division Algorithm): Let $f(t)$ and $g(t)$ be polynomials over a field K with $g(t)$ 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 poly- nomial $q(t)$ 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 $f(1) \,/\!= 0$ and $f(-1) \,/\!= 0$. By synthetic division, or dividing by $t - 2$, we get

$$\begin{array}{cccccccc} 2 & 1 & + & 1 & - & 8 & + & 4 \\ \cdot & & & 2 & + & 6 & - & 4 \\ \hline & 1 & + & 3 & - & 2 & + & 0 \end{array}$$

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 $f(t)$:

$$t = 2, \quad t = (-3 + \sqrt{17})/2, \; t = (-3 - \sqrt{17})/2$$

(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

$$\begin{array}{cccccccccc} 1 & 1 & - & 2 & + & 0 & + & 11 & - & 10 \\ \cdot & & & 1 & - & 1 & - & 1 & + & 10 \\ \hline \cdot & & & & & & & & & \\ \hline -2 & 1 & - & 1 & - & 1 & + & 10 & + & 0 \\ & - & 2 & + & 6 & - & 10 & & & \\ & 1 & - & 3 & + & 5 & + & 0 & & \end{array}$$

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 [d]eal 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 $f_1 f_2 \cdot f_n$ 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

$$f = kp_1 p_2 \ldots p_n$$

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 $k \in$ **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 $t =$ root.
$2 + 3i$ is a

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

Q1 consider the set **Q** of rational numbers, and let $*$ be the operation on **Q** defined by

$$a * b = a + b - ab$$

(a) Find: (i) $3 * 4$; (ii) $2 * (-5)$; (iii) $7 * (1/2)$.

(b) Is (**Q**, $*$) a semigroup? Is it commutative?

(c) Find the identity element for $*$.

(d) Do any of the elements in **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:

$(a * b) * c = (a + b - ab) * c = (a + b - ab) + c - (a + b - ab)c$

$= a + b - ab + c - ac - bc + abc = a + b + c - ab -$
$ac - bc + abc$ $a * (b * c) = a * (b + c - bc) = a + (b + c - bc) -$
$a(b + c - bc)$

$= a + b + c - bc - ab - ac + abc$

Hence $*$ is associative and (**Q**, $*$) is a semigroup. Also

$$a * b = a + b - ab = b + a - ba = b * a$$

Hence (**Q**, $*$) is a commutative semigroup.

(c) An element e is an identity element if $a * e = a$ for every $a \in$ **Q**. Compute as follows:

$a * e = a, a + e - ae = a, \qquad e - ea = 0, \quad e(1$

$- a) = 0, \qquad e = 0$Accordingly, 0 is the identity element.
$* \quad =$

(d) In order for a to have an inverse x, we must have $a \qquad x \qquad 0$ since 0 is the identity element by Part (c). Compute as follows:

$a * x = 0, a + x - ax = 0, \qquad a = ax - x, \quad a = x(a - l), \quad x = a/(a -$
$l)$

Thus if $a \,/\!= 1$, then a has an inverse and it is $a/(a - 1)$.

Q .Let S be a semigroup with identity e, and let b and b^J be inverses of a. Show that bb^J, that is, that inverses are unique if they exist.

We have:

$$b * (a * b^J) = b * e = b \quad \text{and} \quad (b * a) * b^J = e * b^J = b^J$$

Since S is associative, $(b * a) * b^J = b* (a * b^J)$; hence $b = b^J$.

Q. Let $S = \mathbf{N} \times \mathbf{N}$. Let $*$ be the operation on S defined by $(a, b) * (a^J, b^J) = (aa^J, bb^J)$.

(a) Show that $*$ is associative. (Hence S is a semigroup.)

(b) Define f: $(S, *) \rightarrow (\mathbf{Q}, \times)$ 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.2. 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 $n \geq 1$, 2, and 3. Suppose n 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,

$$[a_1a_2\cdots a_n] = [a_1a_2\cdots a_r]\,[a_{r+1}\cdots a_n] = [a_1a_2\cdots a_r](a_{r+1}\cdots a_n)$$

$$= [a_1 \cdots a_r]\,((a_{r+1} \cdots a_{n-1})a_n) = ([a_1 \cdots a_r]\,(a_{r-1} \cdots a_{n-1}))a_n$$

$$= [a_1 \cdots a_{n-1}]a_n = (a_1 \cdots a_{n-1})a_n = (a_1a_2 \cdots a_n)$$

Thus the theorem is proved.

B.3. 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 $W: S/\sim\ \to f(S)$ by $W([a]) = f(a)$. We need to prove: (1) W is well-defined, that is, $W([a]) \in f(S)$, and if $[a] = [b]$ then $f([a]) = f([b])$. (2) W is an isomorphism, that is, W is a homomorphism, one-to-one and onto.

(1) *Proof that W is well-defined*: We have $W([a]) = f(a)$. Since $a \in S$, we have $f(a) \in f(S)$. Hence $W([a]) \in f(S)$, as required. Now suppose $[a] = [b]$. Then $a \sim b$ and hence $f(a) = f(b)$. Thus

$$W([a]) = f(a) = f(b) = W([b])$$

That is, W is well-defined.

(2) *Proof that W is an isomorphism*: Since f is a homomorphism,

$$W([a][b]) = W[ab] = f(ab) = f(a)f(b) = W([a])W([b])$$

Hence W is a homomorphism. Suppose $W([a]) = W([b])$. Then $f(a) = f(b)$, and so $a \sim b$. Thus $[a] = [b]$ and W is one-to-one. Lastly, let $y \in f(S)$. Then, $f(a) = y$ for some $a \in S$. Hence $W([a]) = f(a) = y$. Thus W is onto $f(S)$. Accordingly, W is an isomorphism.

B.4. 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?

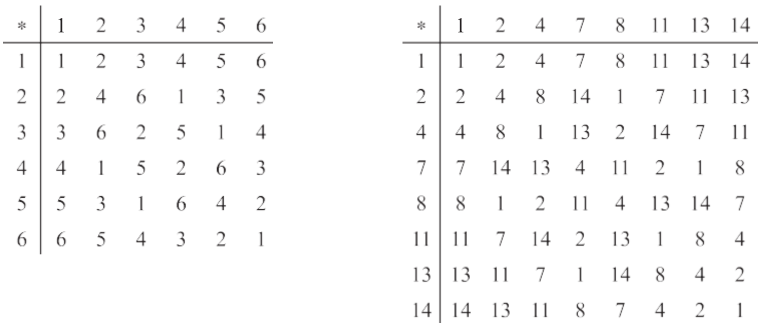
(a) To find $a * b$ in G, find the remainder when the product ab is divided by 7. For example; 5=6 30 which yields a remainder of 2 when divided by 7; hence 5 6 2 in G. The multiplication table of G appears in Fig. B-6(*a*).

(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.5. 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 gp(2) $= \{1, 2, 4, 8\}$. Also,

$7^2 = 4, 7^3 = 4 * 7 = 13, 7^4 = 13 * 7 = 1$. Hence $|7| = 4$ and gp(7) $= \{1, 4, 7, 13\}$. Lastly, $11^2 = 1$. Hence $|11| = 2$

and gp(11) $= \{1, 11\}$.

(*d*) No, since no element generates *G*.

B.6. 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 $B = \{\varphi_1, \varphi_2\}$. 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 $gp(x) = \{\varepsilon, x^1, x^2, ..., x^{n-1}\}$. Note $x^1 = x$, so we need only begin with $n = 2$ when $x /= \varepsilon$.

(1) $\varepsilon^1 = \varepsilon$; so $|\varepsilon| = 1$ and $g(\varepsilon) = \{\varepsilon\}$.

(2) $\sigma_1^2 = \varepsilon$; hence $\cdot\sigma_1\cdot = 2$ and $gp(\sigma_1) = \{\varepsilon, \sigma_1\}$.

(3) $\sigma_2^2 = \varepsilon$; hence $\cdot\sigma_2\cdot = 2$ and $gp(\sigma_2) = \{\varepsilon, \sigma_2\}$.

(4) $\sigma_3^2 = \varepsilon$; hence $\cdot\sigma_3\cdot = 2$ and $gp(\sigma_3) = \{\varepsilon, \sigma_3\}$.

(5) $\varphi_1^2 = \varphi_2, \varphi_1^3 = \varphi_2\varphi_1 = \varepsilon$; hence $\cdot\varphi_1\cdot = 3$ and $gp(\varphi_2) = \{\varepsilon, \varphi_1, \varphi_2\}$.

(6) $\varphi_2^2 = \varphi_1, \varphi_2^3 = \varphi_1\varphi_2 = \varepsilon$; hence $\cdot\varphi_2\cdot = 3$ and $gp(\varphi_1) = \{\varepsilon, \varphi_1, \varphi_2\}$.

(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:

$$H_3 = \{\varepsilon, \sigma_1\}, \qquad H_4 = \{\varepsilon, \sigma_2\}, \quad H_5 = \{\varepsilon, \sigma_3\}, \; H_6 = \{\varepsilon, \varphi_1, \varphi_2\}$$

Accordingly, S_3 has six subgroups.

(c) Multiply each element of *A* by each element of *B*:

$$\sigma_1\varphi_1 = \sigma_2, \qquad \sigma_1\varphi_2 = \sigma_3, \; \sigma_3\varphi_1 = \sigma_3, \quad \sigma_2\varphi_2 = \sigma_1$$

Hence AB $= \{\sigma_1, \sigma_2, \sigma_3\}$.

Multiply σ_3 by each element of A:

$$\sigma_3\sigma_1 = \varphi_1, \qquad \sigma_3\sigma_2 = \varphi_2, \text{ hence } \sigma_3 A = \{\varphi_1, \varphi_2\}$$

Multiply each element of A by σ_3:

$$\sigma_1\sigma_3 = \varphi_2, \qquad \sigma_2\sigma_3 = \varphi_1, \text{ hence } A\sigma_3 = \{\varphi_1, \varphi_2\}$$

(d) $H = \{e, \sigma_1\}$, $K = \{e, \sigma_2\}$ and then $HK = \{e, \sigma_1, \sigma_2, \varphi_1\}$, 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.7. Let σ and τ be the following elements of the symmetric group S_6:

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

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:

$$\sigma^{-1} = \begin{pmatrix} 3 & 1 & 5 & 4 & 6 & 2 \\ 1 & 2 & 3 & 4 & 5 & 6 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 2 & 6 & 1 & 4 & 3 & 5 \end{pmatrix}$$

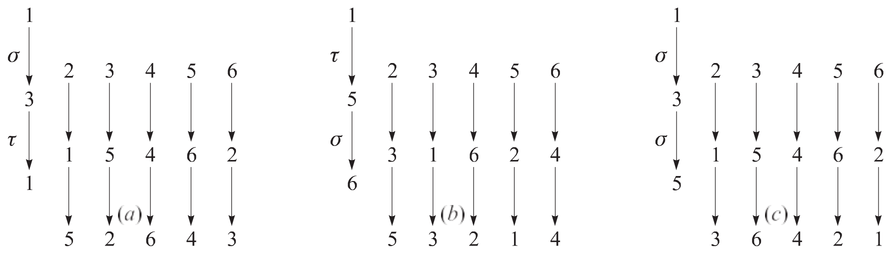


Fig. B-7

B.8. 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

$$(h,\ k) * h^{J},\ k^{J} = (hh^{J}, kk^{J})$$

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

$$e = (0, 0),\ a = (1, 0),\ b = (0, 1),\ c = (1, 1)$$

The multiplication table of G appears in Fig. B-8(a). Note that G is abelian since the table is symmetric. Also,

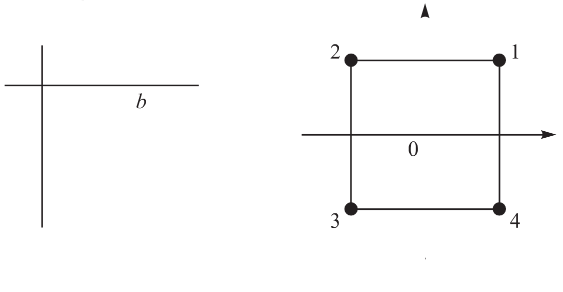
$a^2 = e, b^2 = e, c^2 = e$. Thus G is not cyclic, and hence $G \cong \mathbf{Z_4}$.

B.9. 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.



(a) (b)

a	*a*	*e*	*c*	*b*
b	*b*	*c*	*e*	*a*
c	*c*	*b*	*a*	*e*

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 α 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:

$$\sigma(0^\circ) = \begin{pmatrix}1&2&3&4\\1&2&3&4\end{pmatrix}, \qquad \sigma(90^\circ) = \begin{pmatrix}1&2&3&4\\2&3&4&1\end{pmatrix},$$

$$\sigma(180^\circ) = \begin{pmatrix}1&2&3&4\\3&4&1&4\end{pmatrix}, \qquad \sigma(270^\circ) = \begin{pmatrix}1&2&3&4\\4&1&2&3\end{pmatrix}$$

$$\tau(0^\circ) = \begin{pmatrix}1&2&3&4\\2&1&4&3\end{pmatrix}, \qquad \tau(90^\circ) = \begin{pmatrix}1&2&3&4\\3&2&1&4\end{pmatrix},$$

$$\tau(180^\circ) = \begin{pmatrix}1&2&3&4\\4&3&2&1\end{pmatrix}, \qquad \tau(270^\circ) = \begin{pmatrix}1&2&3&4\\1&4&3&2\end{pmatrix}$$

(c) Let $a = \sigma(90^\circ)$ and $b = \tau(0^\circ)$. Then *a* and *b* form a maximum set of generators of *G*. Specifically,

$$\sigma(0^\circ) = a^4, \qquad \sigma(90^\circ) = a, \quad \sigma(180^\circ) = a^2, \qquad \sigma(270^\circ) = a^3$$

$$\tau(0^\circ) = b, \quad \tau(90^\circ) = ba, \qquad \tau(180^\circ) = ba^2, \qquad \tau(270^\circ) = ba^3$$

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.10. 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 $\psi : G \to$ 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.11. 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 $c = h_1a$ and $c = h_2b$, where $h_1, h_2 \in H.$ Then $h_1a = h_2b$, and so $a = h_1{}^1h_2b$. Let $x \in Ha$. Then

$$x = h_3a = h_3h_1{}^1h_2b$$

where $h_3 \in H$. Since H is a subgroup, $h_3h_1{}^1h_2 \in H$; 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.12. 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 $Ha = \{h_1a, h_2a, \ldots, h_ka\}$.

However, $h_ia = h_ja$ 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.13. 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.14. Prove: Every subgroup of a cyclic group G is cyclic.

Since G is cyclic, there is an element a 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 $$a^n = a^{mq+r} = a^{mq} \cdot a^r = b^q \cdot a^r \text{ so } a^r = b^{-q}\, a^n$$

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.15. 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

$$(aH)H = a(HH) = aH \qquad \text{and } H\,(aH) = (Ha)H = (aH)H = aH$$

Lastly, $a^{-1}H$ is the inverse of aH since

$$(a^{-1}H)(aH) = a^{-1}aHH = eH = H \text{ and } (aH)(a^{-1}H) = aa^{-1}HH = eH = H$$

Thus G/H is a group under coset multiplication.

B.16. Suppose $F : G \to G^J$ is a group homomorphism. Prove: (a) $f(e) = e^J$; (b) $(f\,a^{-6}) = f(a)^{-1}$.

(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

$$e^J = f(e) = f(aa^{-1}) = f(a)f(a^{-1}) \quad \text{and} \quad e^J = f(e) = f(a^{-1}a) = f(a^{-1})f(a)$$

Hence $f\,(a^{-1})$ is the inverse of $f\,(a)$; that is, $f\,(a^{-1}) = f\,(a)^{-1}$.

B.17. Prove Theorem B.9: Let f $G \to G^J$ 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, $f(e) = e^J$, so $e \in K$. Now suppose $a, b \in K$ and $g \in G$. Then

$f(a) = e^J$ and $f(b) = e^J$. Hence

$$f(ab) = f(a)f(b) = \\ e^Je^J = e^J\, f(a^{-1}) = f \\ (a)^{-1} = e^{J-1} = e^J$$

$$f(gag^{-1}) = f(g)f(a)f(g^{-1}) = f(g)e^{J}f(g)^{-1} = e^{J}$$

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

$$\phi(Ka) = f(a)$$

We show that ϕ is well-defined, i.e., if $Ka = Kb$ then $\phi(Ka) = \phi(Kb)$. Suppose $Ka = Kb$. Then $ab^{-1} \in K$

(Problem B.57). Then $f(ab^{-1}) = e^{J}$, and so

$$f(a)f(b)^{-1} = f(a)f(b^{-1}) = f(ab^{-1}) = e^{J}$$

Hence $f(a) = f(b)$, and so $\phi(Ka) = \phi(Kb)$. Thus ϕ is well-defined.

We next show that ϕ is a homomorphism:

$$\phi(KaKb) = \phi(Kab) = f(ab) = f(a)f(b) = \phi(Ka)\phi(Kb)$$

Thus ϕ is a homomorphism. We next show that ϕ is one-to-one. Suppose $\phi(Ka) = \phi(Kb)$. Then

$$f(a) = f(b) \quad \text{or} \quad f(a)f(b)^{-1} = e^{J} \quad \text{or} \quad f(a)f(b^{-1}) = e^{J} \quad \text{or} \quad f(ab^{-1}) = e^{J}$$

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 $\phi(Ka) = f(a) = h$, and so ϕ is onto. Consequently $G/K = H$ and the theorem is proved.

B.18. 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:

$$f(0) = 4, f(2) = 0, f(4) = 2, f(6) = 0, f(8) = 4$$

$$f(1) = 0, f(3) = 4, f(5) = 4, f(7) = 0, f(9) = 2$$

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.19. Prove that in a ring *R*: (i) $a \; 0 = 0 \cdot a = 0$; (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.20. Let *D* be an integral domain. Show that if $ab = ac$ with $a /= 0$ then $b = c$.

Since $ab = ac$, we have

$$ab - ac = 0 \quad \text{and so} \quad a(b - c) = 0$$

Since $a /= 0$, we must have $b - c = 0$, since *D* has no zero divisors.

Hence $b = c$.

B.21. 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,

$$a - b, ra, ar \in J \qquad \text{and } a - b, ra, ar \in K$$

Hence $a - b, ra, ar \in J \cap K$. Therefore $J \cap K$ is an ideal.

B.22. 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.23. 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 $a /= 0$, we have $aa_i = aa_k$ implies $a_i = a_k$ (Problem B.22). Thus the above n elements are distinct, and sothey 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 prove

B.24. 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 ± 1, ± 2, $\pm 1/2$. Testing each possible root, we get, by synthetic division

(or dividing by $2t + 1$),

$$-\ 1\ .\ 2 - 3 - 6 - 2$$

$$2$$

$$.\quad -1 + 2 + 2$$

$$2 - 4 - 4 + 0$$

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 $f(t)$:

$$t = -1/2,\ t = 1 + \sqrt{3},\ t = 1 - \sqrt{3}$$

B.25. 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:

$$t = 1 + 2i,\ t = 1 - 2i,\ t = (1 + \sqrt{17})/2, \quad t = (1 - \sqrt{17})/2$$

B.26. 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:

$$f(0) = 0, f(2) = 0, f(4) = 0, f(6) = 0$$

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.27. 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.28. 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

$$f(t) = a_n t^n + \cdots + a_1 t + a_0 \text{ and } g(t) = b_m t^m + \cdots + b_1 t + b_0$$

where $a_n, b_m \neq 0$ and $n > m$. We form the polynomial

$$f_1(t) = f(t) - \frac{a_n}{b_m} t^{n-m} g(t) \qquad (1)$$

(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 $r(t) = 0$ 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.29. 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

$$f(t) = (t - a)g(t) \qquad (1)$$

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.30. 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

$$a_n p^n + a_{n-1} p^{n-1} q + a_{n-2} p^{n-2} q^2 + \cdots + a_1 p q^{n-1} + a_0 q^n = 0 \qquad (1)$$

Since p divides all of the first n terms of (1), p must divide the last term a_0q^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_np^n. Since p and q are relatively prime, q divides a_n.

B.31. 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^J is another monic polynomial which generates J, then d divides d^J and d^J divides d. This implies that $d = d^J$, because d and d^J are monic. Thus the theorem is proved.

B.32. 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^J divides f and g, then d^J 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^J divides f and g. Let J be the ideal generated by d^J. Then $f, g \in J$ and hence $I \subseteq J$. Accordingly, $d \in J$ and so d^J 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.33. 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.34. 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 $f_1f_2 \cdots f_n$ 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 $f_1f_2 \cdots f_n$. 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.

Boolean Algebra

BASIC DEFINITIONS

Let B be a nonempty set with two binary operations $+$ and $*$, a unary operation $'$, and two distinct elements 0 and 1. Then B is called a *Boolean algebra* if the following axioms hold where a, b, c are any elements in B:

[$\mathbf{B_1}$] Commutative laws:
(1a) $a + b = b + a$ (1b) $a * b = b * a$

[$\mathbf{B_2}$] Distributive laws:
(2a) $a + (b * c) = (a + b) * (a + c)$ (2b) $a * (b + c) = (a * b) + (a * c)$

[$\mathbf{B_3}$] Identity laws:
(3a) $a + 0 = a$ (3b) $a * 1 = a$

[$\mathbf{B_4}$] Complement laws:
(4a) $a + a' = 1$ (4b) $a * a' = 0$

We will sometimes designate a Boolean algebra by $\langle B, +, *, ', 0, 1\rangle$ when we want to emphasize its six parts. We say 0 is the *zero* element, l, is the *unit* element, and a' is the *complement* of a. We will usually drop the symbol $*$ and use juxtaposition instead. Then (2b) is written $a(b + c) = ab + ac$ which is the familiar algebraic identity of rings and fields. However, (2a) becomes $a + bc = (a + b)(a + c)$, which is certainly not a usual identity in algebra.

The operations $+, *$, and $'$ are called sum, product, and complement, respectively. We adopt the usual convention that, unless we are guided by parentheses, $'$ has precedence over $*$, and $*$ has precedence over $+$. For example,

$a + b * c$ means $a + (b * c)$ and not $(a + b) * c$; $a * b'$ means $a * (b')$ and not $(a * b)'$

Of course when $a + b * c$ is written $a + bc$ then the meaning is clear.

EXAMPLE 1

(a) Let $\mathbf{B} = \{0, 1\}$, the set of *bits* (binary digits), with the binary operations of $+$ and $*$ and the unary operation $'$ defined by Fig. 15-1. Then $\mathbf{B}$ is a Boolean algebra. (Note $'$ simply changes the bit, i.e., $1' = 0$ and $0' = 1$.)

1	1	1

1	1	0

	0	1

+	1	0
0	1	0

$*$	1	0
0	0	0

$'$	1	0

Fig. 15-1

(b) Let $\mathbf{B}^n = \mathbf{B} \times \mathbf{B} \times \cdots \times \mathbf{B}$ (n factors) where the operations of $+, *$, and ^{J} are defined componentwise using Fig. 15-1. For notational convenience, we write the elements of $\mathbf{B}^n$ as n-bit sequences without commas, e.g., $x = 110011$ and $y = 111000$ belong to $\mathbf{B}^n$. Hence

$$x + y = 111011, \quad x * y = 110000, \quad x^J = 001100$$

Then $\mathbf{B}^n$ is a Boolean algebra. Here 0 000 0 is the zero element, and 1 111 1 is the unit element. We note that $\mathbf{B}^n$ has 2^n elements. $=$ $\cdots$ $=$ $\cdots$

(c) Let $\mathbf{D}_{70} = \{1, 2, 5, 7, 10, 14, 35, 70\}$, the divisors of 70. Define $+, *$, and ^{J} on $\mathbf{D}_{70}$ by

$$a + b = \mathrm{lcm}(a, b), \quad a * b = \gcd(a, b), \quad a^J = \frac{70}{a}$$

Then $\mathbf{D}_{70}$ is a Boolean algebra with 1 the zero element and 70 the unit

element.

(d) Let C be a collection of sets closed under the set operations of union, intersection, and complement. Then C is a Boolean algebra with the empty set $\varnothing$ as the zero element and the universal set $\mathbf{U}$ as the unit element.

Subalgebras, Isomorphic Boolean Algebras

Suppose C is a nonempty subset of a Boolean algebra B. We say C is a *subalgebra* of B if C itself is a Boolean algebra (with respect to the operations of B). We note that C is a subalgebra of B if and only if C is closed under the three operations of B, i.e., $+, *$, and ^{J}. For example, $\{1, 2, 35, 70\}$ is a subalgebra of $\mathbf{D}_{70}$ in Example 15.1(c).

Two Boolean algebras B and B^J are said to be *isomorphic* if there is a one-to-one correspondence $f\colon B \to B^J$

which preserves the three operations, i.e., such that, for any elements, a, b in B,

$$f(a + b) = f(a) + f(b), \quad f(a * b) = f(a) * f(b) \quad \text{and} \quad f(a^J) = f(a)^J$$

15.1 DUALITY

The *dual* of any statement in a Boolean algebra B is the statement obtained by interchanging the operations $+$ and $*$, and interchanging their identity elements 0 and 1 in the original statement. For example, the dual of

$$(1 + a) * (b + 0) = b \quad \text{is} \quad (0 * a) + (b * 1) = b$$

Observe the symmetry in the axioms of a Boolean algebra B. That is, the dual of the set of axioms of B is the same as the original set of axioms. Accordingly, the important principle of duality holds in B. Namely,

Theorem 15.1 (Principle of Duality): The dual of any theorem in a Boolean algebra is also a theorem.

In other words, if any statement is a consequence of the axioms of a Boolean algebra, then the dual is also a consequence of those axioms since the dual statement can be proven by using the dual of each step of the proof of the original statement.

15.2 BASIC THEOREMS

Using the axioms [$\mathbf{B_1}$] through [$\mathbf{B_4}$], we prove (Problem 15.5) the following theorem.

Theorem 15.2: Let a, b, c be any elements in a Boolean algebra B.

(i) Idempotent laws:
(5*a*) $a + a = a$ (5*b*) $a * a = a$

(ii) Boundedness laws:
(6*a*) $a + 1 = 1$ (6*b*) $a * 0 = 0$

(iii) Absorption laws:
(7*a*) $a + (a * b) = a$ (7*b*) $a * (a + b) = a$

(iv) Associative laws:
(8*a*) $(a + b) + c = a + (b + c)$ (8*b*) $(a * b) * c = a * (b * c)$

Theorem 15.2 and our axioms still do not contain all the properties of sets listed in Table 1-1. The next two theorems give us the remaining properties.

Theorem 15.3: Let a be any element of a Boolean algebra B.

(i) (Uniqueness of Complement) If $a + x = 1$ and $a * x = 0$, then $x = a'$.

(ii) (Involution law) $(a')' = a$. (iii)

(9*a*) $0' = 1$. (9*b*) $1' = 0$.

Theorem 15.4 (DeMorgan's laws): (10*a*) $(a + b)' = a' * b'$. (10*b*) $(a * b)' = a' + b'$.

We prove these theorems in Problems 15.6 and 15.7.

15.3 BOOLEAN ALGEBRAS AS LATTICES

By Theorem 15.2 and axiom [$\mathbf{B_1}$], every Boolean algebra B satisfies the associative, commutative, and absorption laws and hence is a lattice where $+$ and $*$ are the join and meet operations, respectively. With respect to this lattice, $a + 1 = 1$ implies $a \leq 1$ and $a * 0 = 0$ implies $0 \leq a$, for any element $a \in B$. Thus B is a bounded lattice. Furthermore, axioms $\mathbf{B_2}$ and [$\mathbf{B_4}$] show that B is also distributive and complemented. Conversely, every bounded, distributive, and complemented lattice L satisfies the axioms $\mathbf{B_1}$ through [$\mathbf{B_4}$]. Accordingly, we have the following

Alternate Definition: A Boolean algebra B is a bounded, distributive and complemented lattice.

Since a Boolean algebra B is a lattice, it has a natural partial ordering (and so its diagram can be drawn). Recall (Chapter 14) that we define $a \leq b$ when the equivalent conditions $a + b = b$ and $a * b = a$ hold. Since we are in a Boolean algebra, we can actually say much more.

Theorem 15.5: The following are equivalent in a Boolean algebra:

(1) $a + b = b$, (2) $a * b = a$, (3) $a' + b = 1$, (4) $a * b' = 0$

Thus in a Boolean alegbra we can write $a \leq b$ whenever any of the above four conditions is known to be true.

EXAMPLE 2

(a) Consider a Boolean algebra of sets. Then set A precedes set B if A is a subset of B. Theorem 15.4 states that if $A \subseteq B$ then the following conditions hold:

(1) $A \cup B = B$ (2) $A \cap B = A$ (3) $A^c \cup B = \mathbf{U}$ (4) $A \cap B^c = \boldsymbol{\emptyset}$

(b) Consider the Boolean algebra $\mathbf{D}_{70}$. Then *a* precedes *b* if *a* divides *b*. In such a case, lcm*(a, b)* = *b* and gcd*(a, b)* = *a*. For example, let *a* = 2 and *b* = 14. Then the following conditions hold:

(1) lcm(2, 14) = 14. (3) lcm(2^{J}, 14) = lcm(35, 14) = 70.

(2) gcd(2, 14) = 2. (4) gcd(2, 14^{J}) = gcd(2, 5) = 1.

15.4 REPRESENTATION THEOREM

Let *B* be a finite Boolean algebra. Recall (Section 14.10) that an element *a* in *B* is an atom if *a* immediately succeeds 0, that is if 0 *a*. Let *A* be the set of atoms of *B* and let *P (A)* be the Boolean algebra of all subsets of the set *A* of atoms. By Theorem 14.8, each *x* 0 in *B* can be expressed uniquely (except for order) as the sum (join) of atoms, i.e., elements of *A*. Say,

$$x = a_1 + a_2 + \cdots + a_r$$

is such a representation. Consider the function *f*: *B* → *P (A)* defined by

$$f(x) = \{a_1, a_2, \ldots, a_r\}$$

The mapping is well defined since the representation is unique.

Theorem 15.6: The above mapping *f*: *B* → *P (A)* is an isomorphism.

Thus we see the intimate relationship between set theory and abstract Boolean algebras in the sense that every finite Boolean algebra is structurally the same as a Boolean algebra of sets.

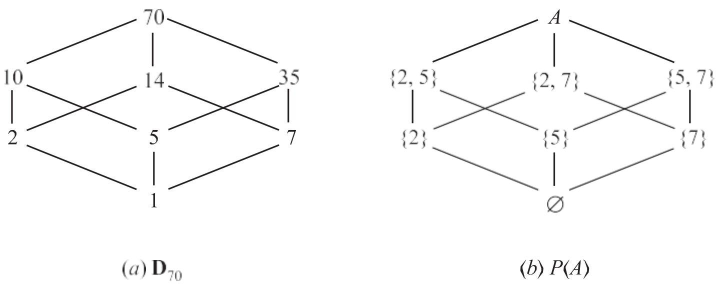
If a set *A* has *n* elements, then its power set *P (A)* has 2^n elements. Thus the above theorem gives us our next result.

Corollary 15.7: A finite Boolean algebra has 2^n elements for some positive integer *n*.

EXAMPLE 15.3 Consider the Boolean algebra $\mathbf{D}_{70}$ 1, 2{ 5, . . . , 70 who}se diagram is given in Fig. 15-2*(a)*. Note that *A* 2, 5, 7 {is the s}et of atoms of $\mathbf{D}_{70}$. The following is the unique representation of each nonatom by atoms:

$$10 = 2 \vee 5, \quad 14 = 2 \vee 7, \quad 35 = 5 \vee 7, \quad 70 = 2 \vee 5 \vee 7$$

Figure 15-2*(b)* gives the diagram of the Boolean algebra of the power set *P (A)* of the set *A* of atoms. Observe that the two diagrams are structurally the same.



(*a*) $\mathbf{D}_{70}$ (*b*) *P*(*A*)

Fig. 15-2

15.5 SUM-OF-PRODUCTS FORM FOR SETS

This section motivates the concept of the sum-of-products form in Boolean algebra by an example of set theory. Consider the Venn diagram in Fig. 15-3 of three sets *A*, *B*, and *C*. Observe that these sets partition the

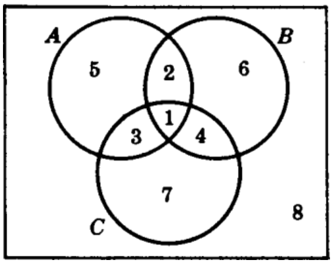


Fig. 15-3

rectangle (universal set) into eight numbered sets which can be represented as follows:

$$\begin{array}{llll}
(1)\ A \cap B \cap C & (3)\ A \cap B^c \cap C & (5)\ A \cap B^c \cap C^c & (7)\ A^c \cap B^c \cap C \\
(2)\ A \cap B \cap C^c & (4)\ A^c \cap B \cap C & (6)\ A^c \cap B \cap C^c & (8)\ A^c \cap B^c \cap C^c
\end{array}$$

Each of these eight sets is of the form $A^* \cap B^* \cap C^*$ where:

$$A^* = A \text{ or } A^c, \qquad B^* = B \text{ or } B^c, \qquad C^* = C \text{ or } C^c$$

Consider any nonempty set expression E involving the sets A, B, and C, say,

$$E = [(A \cap B^c)^c \cup (A^c \cap C^c)] \cap [(B^c \cup C)^c \cap (A \cup C^c)]$$

Then E will represent some area in Fig. 15-3 and hence will uniquely equal the union of one or more of the eight sets.

Suppose we now interpret a union as a sum and an intersection as a product. Then the above eight sets are products, and the unique representation of E will be a sum (union) of products. This unique representation of E is the same as the complete sum-of-products expansion in Boolean algebras which we discuss below.

15.6 SUM-OF-PRODUCTS FORM FOR BOOLEAN ALGEBRAS

Consider a set of variables (or letters or symbols), say $x_1, x_2, \ldots, x_n$. A *Boolean expression E* in these variables, sometimes written $E(x_1, \ldots, x_n)$, is any variable or any expression built up from the variables using the Boolean operations $+$, $*$, and ^{J}. (Naturally, the expression E must be *well-formed*, that is, where $+$ and $*$ are used as binary operations, and ^{J} is used as a unary operation.) For example,

$$E_1 = (x + y^Jz)^J + (xyz^J + x^Jy)^J \quad \text{and} \quad E_2 = ((xy^Jz^J + y)^J + x^Jz)^J$$

are Boolean expressions in x, y, and z.

A *literal* is a variable or complemented variable, such as x, x^J, y, y^J, and so on. A *fundamental product* is a literal or a product of two or more literals in which no two literals involve the same variable. Thus

$$xz^J, \quad xy^Jz, \quad x, \quad y^J, \quad x^Jyz$$

are fundamental products, but xyx^Jz and $xyzy$ are not. Note that any product of literals can be reduced to either 0 or a fundamental product, e.g., $xyx^Jz = 0$ since $xx^J = 0$ (complement law), and $xyzy = xyz$ since $yy = y$ (idempotent law).

A fundamental product P_1 is said to be *contained in* (or *included in*) another fundamental product P_2 if the literals of P_1 are also literals of P_2. For example, x^Jz is contained in x^Jyz, but x^Jz is not contained in xy^Jz since x^J is not a literal of xy^Jz. Observe that if P_1 is contained in P_2, say $P_2 = P_1 * Q$, then, by the absorption law,

$$P_1 + P_2 = P_1 + P_1 * Q = P_1$$

Thus, for instance, $x^Jz + x^Jyz = x^Jz$.

Definition Boolean expression E is called a *sum-of-products* expression if E is a fundamental product or the sum of two or more fundamental products none of which is contained in another.

Definition 15.2: Let E be any Boolean expression. A *sum-of-products form* of E is an equivalent Boolean sum-of-products expression.

EXAMPLE .4 Consider the expressions

$$E_1 = xz' + y'z + xyz' \quad \text{and} \quad E_2 = xz' + x'yz' + xy'z$$

Although the first expression E_1 is a sum of products, it is not a sum-of-products expression. Specifically, the product xz' is contained in the product xyz'. However, by the absorption law, E_1 can be expressed as

$$E_1 = xz' + y'z + xyz' = xz' + xyz' + y'z = xz' + y'z$$

This yields a sum-of-products form for E_1. The second expression E_2 is already a sum-of-products expression.

EXAMPLE 5 Suppose Algorithm 15.1 is applied to the following Boolean expression:

$$E = ((xy)'z)'((x' + z)(y' + z'))'$$

Step 1. Using DeMorgan's laws and involution, we obtain

$$E = (xy'' + z')((x' + z)' + (y' + z')') = (xy + z')(xz' + yz) \ E$$

now consists only of sums and products of literals.

Step 2. Using the distributive laws, we obtain

$$E = xyxz' + xyyz + xz'z' + yzz'$$

E now is a sum of products.

Step 3. Using the commutative, idempotent, and complement laws, we obtain

$$E = xyz' + xyz + xz' + 0$$

Each term in E is a fundamental product or 0.

Step 4. The product xz' is contained in xyz'; hence, by the absorption law,

$$xz' + (xz'y) = xz'$$

Thus we may delete xyz' from the sum. Also, by the identity law for 0, we may delete 0 from the sum. Accordingly,

$$E = xyz + xz'$$

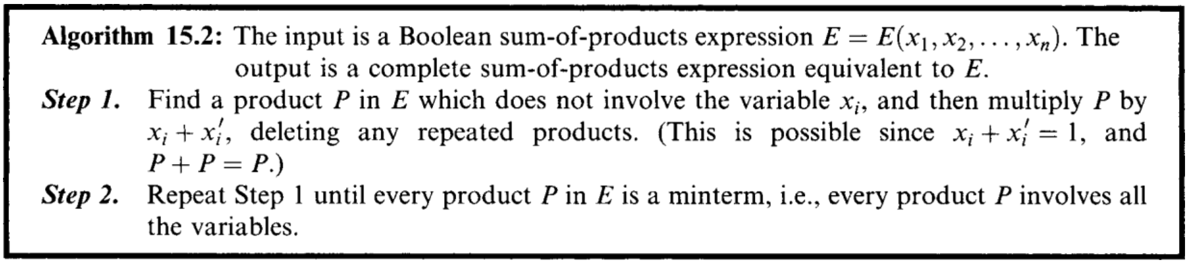
E is now represented by a sum-of-products expression.

Complete Sum-of-Products Forms

A Boolean expression $E = E(x_1, x_2, \ldots, x_n)$ is said to be a *complete sum-of-products* expression if E is a sum-of-products expression where each product P involves all the n variables. Such a fundamental product P which involves all the variables is called a *minterm*, and there is a maximum of 2^n such products for n variables. The following theorem applies.

Theorem : Every nonzero Boolean expression $E = E(x_1, x_2, \ldots, x_n)$ is equivalent to a complete sum-of-products expression and such a representation is unique.

The above unique representation of E is called the *complete sum-of-products form* of E. Algorithm 15-1 in Fig. 15-4 tells us how to transform E into a sum-of-products form. Figure 15-5 contains an algorithm which transforms a sum-of-products form into a complete sum-of-products form.



Algorithm 15.2: The input is a Boolean sum-of-products expression $E = E(x_1, x_2, \ldots, x_n)$. The output is a complete sum-of-products expression equivalent to E.

Step 1. Find a product P in E which does not involve the variable x_i, and then multiply P by $x_i + x_i'$, deleting any repeated products. (This is possible since $x_i + x_i' = 1$, and $P + P = P$.)

Step 2. Repeat Step 1 until every product P in E is a minterm, i.e., every product P involves all the variables.

Fig. 15-5

EXAMPLE .6 Express $E(x, y, z) = x(y'z)'$ into its complete sum-of-products form.

(a) Apply Algorithm 15.1 to E so E is represented by a sum-of-products expression:

$$E = x(y'z)' = x(y + z') = xy + xz'$$

(b) Now apply Algorithm 15.2 to obtain:

$$E = xy(z + z') + xz'(y + y') = xyz + xyz' + xyz' + xy'z'$$

$$= xyz + xyz' + xy'z'$$

Now E is reprsented by its complete sum-of-products form.

15.7 MINIMAL BOOLEAN EXPRESSIONS, PRIME IMPLICANTS

There are many ways of representing the same Boolean expression E. Here we define and investigate a minimal sum-of-products form for E. We must also define and investigate prime implicants of E since the minimal sum-of-products involves such prime implicants. Other minimal forms exist, but their investigation lies beyond the scope of this text.

Minimal Sum-of-Products

Consider a Boolean sum-of-products expression E. Let E_L denote the number of literals in E (counted according to multiplicity), and let E_S denote the number of summands in E. For instance, suppose

$$E = xyz' + x'y't + xy'z't + x'yzt$$

Then $E_L = 3 + 3 + 4 + 4 = 14$ and $E_S = 4$.

Suppose E and F are equivalent Boolean sum-of-products expressions. We say E is *simpler* than F if:

(i) $E_L < F_L$ and $E_S \leq F_L$, or *(ii)* $E_L \leq F_L$ and $E_S < F_L$

We say E is *minimal* if there is no equivalent sum-of-products expression which is simpler than E. We note that there can be more than one equivalent minimal sum-of-products expressions.

Prime Implicants

A fundamental product P is called a *prime implicant* of a Boolean expression E if

$$P + E = E$$

but no other fundamental product contained in P has this property. For instance, suppose

$$E = xy^{J} + xyz^{J} + x^{J}yz^{J}$$

One can show (Problem 15.15) that:

$$xz^{J} + E = E \quad \text{but} \quad x + E /= E \quad \text{and} \quad z^{J} + E /= E$$

Thus xz^{J} is a prime implicant of E.

The following theorem applies.

Theorem 15.9: A minimal sum-of-products form for a Boolean expression E is a sum of prime implicants of E.

The following subsections give a method for finding the prime implicants of E based on the notion of the consensus of fundamental products. This method can then be used to find a minimal sum-of-products form for E. Section 15.12 gives a geometric method for finding such prime implicants.

Consensus of Fundamental Products

Let P_1 and P_2 be fundamental products such that exactly one variable, say x_k, appears uncomplemented in one of P_1 and P_2 and complemented in the other. Then the *consensus* of P_1 and P_2 is the product (without repetitions) of the literals of P_1 and the literals of P_2 after x_k and x^{J}_k are deleted. (We do not define the consensus of $P_1 = x$ and $P_2 = x^{J}$.)

The following lemma (proved in Problem 15.19) applies.

Lemma 15.10: Suppose Q is the consensus of P_1 and P_2. Then $P_1 + P_2 + Q = P_1 + P_2$.

EXAMPLE .7 Find the consensus Q of P_1 and P_2 where:

(a) $P_1 = xyz^{J}s$ and $P_2 = xy^{J}t$.

Delete y and y^{J} and then multiply the literals of P_1 and P_2 (without repetition) to obtain $Q = xz^{J}st$.

(b) $P_1 = xy^{J}$ and $P_2 = y$.

Deleting y and y^{J} yields $Q = x$.

(c) $P_1 = x^{J}yz$ and $P_2 = x^{J}yt$.

No variable appears uncomplemented in one of the products and complemented in the other. Hence P_1 and P_2 have no consensus.

(d) $P_1 = x^{J}yz$ and $P_2 = xyz^{J}$.

Each of x and z appear complemented in one of the products and uncomplemented in the other. Hence P_1 and P_2 have no consensus.

Consensus Method for Finding Prime Implicants

Figure 15-6 contains an algorithm, called the *consensus method*, which is used to find the prime implicants of a Boolean expression E. The following theorem gives the basic property of this algorithm.

Theorem: The consensus method will eventually stop, and then E will be the sum of its prime implicants.

EXAMPLE .8 Let $E = xyz + x'z' + xyz' + x'y'z + x'yz'$. Then:

$$\begin{aligned}
E &= xyz + x'z' + xyz' + x'y'z && (x'yz' \text{ includes } x'z')\\
&= xyz + x'y' + xyz' + x'y'z + xy && (\text{consensus of } xyz \text{ and } xyz')\\
&= x'z' + x'y'z + xy && (xyz \text{ and } xyz' \text{ include } xy)\\
&= x'z' + x'y'z + xy + x'y' && (\text{consensus of } x'z' \text{ and } x'y'z)\\
&= x'z' + xy + x'y' && (x'y'z \text{ includes } x'y')\\
&= x'z' + xy + x'y' + yz' && (\text{consensus of } x'z' \text{ and } xy)
\end{aligned}$$

Now neither step in the consensus method will change E. Thus E is the sum of its prime implicants, which appear in the last line, that is, $x'z'$, xy, $x'y'$, and yz'.

Boolean Functions

Let E be a Boolean expression with n variables $x_1, x_2, \ldots, x_n$. The entire discussion above can also be applied to E where now the special sequences are assigned to the variables $x_1, x_2, \ldots, x_n$ instead of the input devices $A_1, A_2, \ldots, A_n$. The truth table $T = T(E)$ of E is defined in the same way as the truth table $T = T(L)$ for a logic circuit L. For example, the Boolean expression

$$E = xyz + xy'z + x'y$$

which is analogous to the logic circuit L in Example 15.12, yields the truth table

$$T(00001111, 00110011, 01010101) = 00110101$$

or simply $T(E) = 00110101$, where we assume the input consists of the special sequences.

Remark: The truth table for a Boolean expression $E = E(x_1, x_2, \ldots, x_n)$ with n variables may also be viewed as a "Boolean" function from $\mathbf{B}^n$ into $\mathbf{B}$. (The Boolean algebras $\mathbf{B}^n$ and $\mathbf{B} = \{0, 1\}$ are defined in Example 15.1.) That is, each element in $\mathbf{B}^n$ is a list of n bits which when assigned to the list of variables in E produces an element in $\mathbf{B}$. The truth table $T(E)$ of E is simply the graph of the function.

EXAMPLE

(a) Consider Boolean expressions $E = E(x, y, z)$ with three variables. The eight minterms (fundamental prod- ucts involving all three variables) are as follows:

$$xyz, \quad xyz', \quad xy'z, \quad x'yz, \quad xy'z', \quad x'yz', \quad x'y'z'$$

The truth tables for these minterms (using the special sequences for x, y, z) follow:

$$xyz = 00000001, \quad xyz' = 00000010, \quad xy'z = 00000100, \quad x'yz = 00001000$$

$$xy'z' = 00010000, \quad x'yz' = 00100000, \quad x'y'z = 01000000, \quad x'y'z' = 10000000$$

Observe that each minterm assumes the value 1 in only one of the eight positions.

(b) Consider the Boolean expression $E = xyz' + x'yz + x'y'z$. Note that E is a complete sum-of-products expression containing three minterms. Accordingly, the truth table $T\ T(E)$ for E, using the special sequences for x, y, z, can be easily obtained from the sequences in part (a). Specifically, the truth table $T(E)$ will contain exactly three 1's in the same positions as the 1's in the three minterms in E. Thus

$$T(00001111, 00110011, 01010101) = 01001010$$

or simply $T(E) = 01001010$.

Solved Problems

BOOLEAN ALGEBRAS

15.1. Write the dual of each Boolean equation: (a) $(a * 1) * (0 + a') = 0$; (b) $a + a'b = a + b$.

(a) To obtain the dual equation, interchange + and ∗, and interchange 0 and 1. Thus

$$(a + 0) + (1 * a') = 1$$

(b) First write the equation using ∗ to obtain $a + (a' * b) = a + b$. Then the dual is $a * (a' + b) = a * b$, which can be written as

$$a(a' + b) = ab$$

15.2. Recall (Chapter 14) that the set $\mathbf{D}_m$ of divisors of m is a bounded, distributive lattice with

$$a + b = a \vee b = \text{lcm}(a, b) \quad \text{and} \quad a * b = a \wedge b = \text{gcd}(a, b).$$

(a) Show that $\mathbf{D}_m$ is a Boolean algebra if m is square free, i.e., if m is a product of distinct primes.

(b) Find the atoms of $\mathbf{D}_m$.

(a) We need only show that $\mathbf{D}_m$ is complemented. Let x be in $\mathbf{D}_m$ and let $x' = m/x$. Since m is a product of distinct primes, x and x' have different prime divisors. Hence $x * x' = \text{gcd}(x, x') = 1$ and $x + x' = 1\text{cm}(x, x') = m$. Recall that 1 is the zero element (lower bound) of $\mathbf{D}_m$ and that m is the identity element (upper bound) of $\mathbf{D}_m$. Thus x' is a complement of x, and so $\mathbf{D}_m$ is a Boolean algebra.

(b) The atoms of $\mathbf{D}_m$ are the prime divisors of m.

15.3. Consider the Boolean algebra $\mathbf{D}_{210}$.

(a) List its elements and draw its diagram.

(b) Find the set A of atoms.

(c) Find two subalgebras with eight elements.

(d) Is $X = \{1, 2, 6, 210\}$ a sublattice of $\mathbf{D}_{210}$? A subalgebra?

(e) Is $Y = \{1, 2, 3, 6\}$ a sublattice of $\mathbf{D}_{210}$? A subalgebra?

(*a*) The divisors of 210 are 1, 2, 3, 5, 6, 7, 10, 14, 15, 21, 30, 35, 42, 70, 105, and 210. The diagram of $\mathbf{D}_{210}$ appears

in Fig. 15-25.

(*b*) $A = \{2, 3, 5, 7\}$, the set of prime divisors of 210.

(*c*) $B = \{1, 2, 3, 35, 6, 70, 105, 210\}$ and $C = \{1, 5, 6, 7, 30, 35, 42, 210\}$ are subalgebras of $\mathbf{D}_{210}$.

(d) X is a sublattice since it is linearly ordered. However, X is not a subalgebra since 35 is the complement of 2 in $\mathbf{D}_{210}$ but 35 does not belong to X. (In fact, no Boolean algebra with more than two elements is linearly ordered.)

(e) Y is a sublattice of $\mathbf{D}_{210}$ since it is closed under $+$ and $*$. However, Y is not a subalgebra of $\mathbf{D}_{210}$ since it is not closed under complements in $\mathbf{D}_{210}$, e.g., $35 = 2^J$ does not belong to Y. (We note that Y itself is a Boolean algebra; in fact, $Y = \mathbf{D}_6$.)

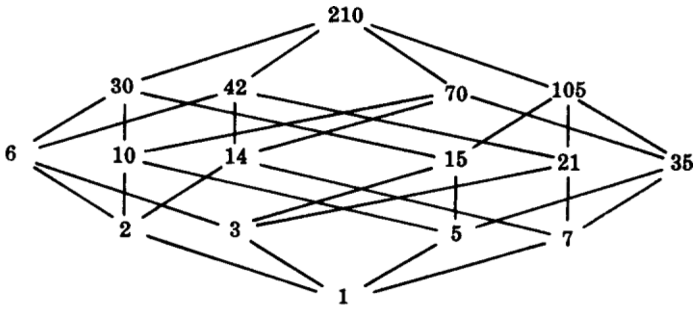


Fig. 15-25

15.4. Find the number of subalgebras of $\mathbf{D}_{210}$.

A subalgebra of $\mathbf{D}_{210}$ must contain two, four, eight or sixteen elements.

(i) There can be only one two-element subalgebra which consists of the upper bound 210 and lower bound 1, i.e.,
{1, 210}.

(ii) Since $\mathbf{D}_{210}$ contains sixteen elements, the only sixteen-element subalgebra is $\mathbf{D}_{210}$ itself.

(iii) Any four-element subalgebra is of the form $\{1, x, x^J, 210\}$, i.e., consists of the upper and lower bounds and a nonbound element and its complement. There are fourteen nonbound elements in $\mathbf{D}_{210}$ and so there are 14/2 = 7 pairs $\{x, x^J\}$. Thus $\mathbf{D}_{210}$ has seven four-element subalgebras.

(iv) Any eight-element subalgebra S will itself contain three atoms s_1, s_2, s_3. We can choose s_1 and s_2 to be any two of the four atoms of $\mathbf{D}_{210}$ and then s_3 must be the product of the other two atoms, e.g., we can let $s_1 = 2$, $s_2 = 3$, $s_3 = 5 \cdot 7 = 35$ (which determines the subalgebra B above), or we can let $s_1 = 5$, $s_2 = 7$, $s_3 = 2 \cdot 3 = 6$(which determines the subalgebra C above). There $\binom{4}{2} = 6$ ways to choose s_1 and s_2 from the four atoms of $\mathbf{D}_{210}$ are and so $\mathbf{D}_{210}$ has six eight-element

Accordingly, $\mathbf{D}_{210}$ has $1 + 1 + 7 + 6 = 15$ subalgebras.

15.5. Prove Theorem 15.2: Let a, b, c be any element in a Boolean algebra B.

(i) Idempotent laws:
(5a) $a + a = a$ (5b) $a * a = a$

(ii) Boundedness laws:
(6a) $a + 1 = 1$ (6b) $a * 0 = 0$

(iii) Absorption Laws:
(7a) $a + (a * b) = a$ (7b) $a * (a + b) = a$

(iv) Associative Laws:
(8a) $(a + b) + c = a + (b + c)$ (8b) $(a * b) * c = a * (b * c)$

(5b) $a = a * 1 = a * (a + a^J) = (a * a) + (a * a^J) = (a * a) + 0 = a * a$ (5a)
Follows from (5b) and duality.

(6b) $a * 0 = (a * 0) + 0 = (a * 0) + (a * a^J) = a * (0 + a^J) = a * (a^J + 0) = a * a^J = 0$

(6a) Follows from (6b) and duality.

(7b) $a * (a + b) = (a + 0) * (a + b) = a + (0 * b) = a + (b * 0) = a + 0 = a$ (7a)
Follows from (7b) and duality.

(8b) Let $L = (a \ast b) \ast c$ and $R = a \ast (b \ast c)$. We need to prove that $L = R$. We first prove that $a + L = a + R$. Using the absorption laws in the last two steps,

$$a + L = a + ((a \ast b) \ast c) = (a + (a \ast b)) \ast (a + c) = a \ast (a + c) = a$$

Also, using the absorption law in the last step,

$$a + R = a + (a * (b * c)) = (a + a) * (a + (b * c)) = a * (a + (b * c)) = a$$

Thus $a + L = a + R$. Next we show that $a' + L = a' + R$. We have,

$$a' + L = a' + ((a * b) * c) = (a' + (a * b)) * (a' + c)$$

$$= ((a' + a) * (a' + b)) * (a' + c) = (1 * (a' + b)) * (a' + c)$$

$$= (a' + b) * (a' + c) = a' + (b * c)$$

Also,

$$a' + R = a' + (a * (b * c)) = (a' + a) * (a' + (b * c))$$

$$= 1 * (a' + (b * c)) = a' + (b * c)$$

Thus $a' + L = a' + R$. Consequently,

$$L = 0 + L = (a * a') + L = (a + L) * (a' + L) = (a + R) * (a' + R)$$

$$= (a * a') + R = 0 + R = R$$

$(8a)$ Follows from $(8b)$ and duality.

15.6. Prove Theorem 15.3: Let a be any element of a Boolean algebra B.

(i) (Uniqueness of Complement) If $a + x = 1$ and $a * x = 0$, then $x = a'$.

(ii) (Involution Law) $(a')' = a$

(iii) $(9a)$ $0' = 1$; $(9b)$ $1' = 0$.

(i) We have:

$$a' = a' + 0 = a' + (a * x) = (a' + a) * (a' + x) = 1 * (a' + x) = a' + x$$

Also,

$$x = x + 0 = x + (a * a') = (x + a) * (x + a') = 1 * (x + a') = x + a'$$

Hence $x = x + a' = a' + x = a'$.

(ii) By definition of complement, $a + a' = 1$ and $a * a' = 0$. By commutativity, $a' + a = 1$ and $a' * a = 0$. By uniqueness of complement, a is the complement of a', that is, $a = (a')'$.

(iii) By boundedness law $(6a)$, $0 + 1 = 1$, and by identity axiom $(3b)$, $0 * 1 = 0$. By uniqueness of complement, 1 is the complement of 0, that is, $1 = 0'$. By duality, $0 = 1'$.

15.7. Prove Theorem 15.4: (DeMorgan's laws): $(10a)$ $(a + b)' = a' * b'$. $(10b)$ $(a * b)' = a' + b'$.

$(10a)$ We need to show that $(a + b) + (a' * b') = 1$ and $(a + b) * (a' * b') = 0$; then by uniqueness of complement,

$a' * b' = (a + b')$. We have:

$$(a + b) + (a' * b') = b + a + (a' * b') = b + (a + a') * (a + b')$$

$$= b + 1 \ast (a + b') = b + a + b' = b + b' + a = 1 + a = 1$$

Also,

$$(a + b) \ast (a' \ast b') = ((a + b) \ast a') \ast b'$$

$$= ((a \ast a') + (b \ast a')) \ast b' = (0 + (b \ast a')) \ast b'$$

$$= (b \ast a') \ast b' = (b \ast b') \ast a' = 0 \ast a' = 0$$

Thus $a' \ast b' = (a + b)'$.

*(*10*b)* Principle of duality (Theorem 15.1).

15.8. Prove Theorem 15.5: The following are equivalent in a Boolean algebra:

(1) $a + b = b$; *(2)* $a * b = a$; *(3)* $a^{J} + b = 1$; *(4)* $a * b^{J} = 0$.

By Theorem 14.4, (*1*) and (*2*) are equivalent. We show that (*1*) and (*3*) are equivalent. Suppose (*1*) holds. Then

$$a^{J} + b = a^{J} + (a + b) = (a^{J} + a) + b = 1 + b = 1$$

Now suppose *(3)* holds. Then

$$a + b = 1 * (a + b) = (a^{J} + b) * (a + b) = (a^{J} * a) + b = 0 + b = b$$

Thus *(1)* and *(3)* are equivalent.

We next show that *(3)* and *(4)* are equivalent. Suppose *(3)* holds. By DeMorgan's law and involution,

$$0 = 1^{J} = (a^{J} + b^{J})^{J} = a^{JJ} * b^{J} = a * b^{J}$$

Conversely, if *(4)* holds then

$$1 = 0^{J} = (a * b^{J})^{J} = a^{J} + b^{JJ} = a^{J} + b$$

Thus *(3)* and *(4)* are equivalent. Accordingly, all four are equivalent.

15.9. Prove Theorem 15.6: The mapping *f*: $B \rightarrow P(A)$ is an isomorphism where *B* is a Boolean algebra, *P(A)* is the power set of the set *A* of atoms, and

$$f(x) = \{a_1, a_2, \ldots, a_n\}$$

where $x = a_1 + \cdots + a_n$ is the unique representation of a as a sum of atoms.

Recall (Chapter 14) that if the *a*'s are atoms then $a^2 = a$ but $a\ a = 0$ for $a\ /= a$. Suppose *x*, *y* are in *B* and suppose

i *i* *i j* *i* *j*

$$x = a_1 + \cdots + a_r + b_1 + \cdots + b_s$$
$$y = b_1 + \cdots + b_s + c_1 + \cdots + c_t$$

where

$$A = \{a_1, \ldots, a_r, b_1, \ldots, b_s, c_1, \ldots, c_t, d_1, \ldots, d_k\}$$

is the set of atoms of *B*. Then

$$x + y = a_1 + \cdots + a_r + b_1 + \cdots + b_s + c_1 + \cdots + c_t$$
$$xy = b_1 + \cdots b_s$$

Hence

Let Thus

$$f(x+y) = \{a_1, \ldots, a_r, b_1, \ldots, b_s, c_1, \ldots, c_t\}$$

$$= \{a_1, \ldots, a_r, b_1$$

$$, \ldots, b_s\} \cup \{b_1, \ldots, b_s, c_1, \ldots, c_t\}$$

$$= f(x) \cup f(y)$$

$$f(xy) = \{b_1, \ldots, b_s\}$$

$$= \{a_1, \ldots, a_r, b_1, \ldots, b_s\} \cap \{b_1, \ldots, b_s, c_1, \ldots, c_t\}$$

$$= f(x) \cap f(y)$$

$y = c_1 + \cdots + c_t + d_1 + \cdots + d_k$. Then $x + y = 1$ and $xy = 0$, and so $y = x^J$

$$f(x^J) = \{c_1, \ldots, c_t, d_1, \ldots, d_k\} = \{a_1, \ldots, a_r, b_1, \ldots, b_s\}^c = (f(x))^c$$

Since the representation is unique, f is one-to-one and onto. Hence f is a Boolean algebra isomorphism.

Solved Problem

Q 1 Write the dual of each Boolean equation: (a) $(a * 1) * (0 + a') = 0$; (b) $a + a'b = a + b$.

(a) To obtain the dual equation, interchange + and ∗, and interchange 0 and 1. Thus

$$(a + 0) + (1 * a') = 1$$

(b) First write the equation using to obtain a $(a'$ $b)$ a b. Then the dual is a $(a'$ $b)$ a b, which can
be written as ∗ + ∗ = + ∗ + = ∗

$$a(a' + b) = ab$$

2 .Recall (Chapter 14) that the set $\mathbf{D}_m$ of divisors of m is a bounded, distributive lattice with

$$a + b = a \vee b = \text{lcm}(a, b) \quad \text{and} \quad a * b = a \wedge b = \text{gcd}(a, b).$$

i. Show that $\mathbf{D}_m$ is a Boolean algebra if m is square free, i.e., if m is a product of distinct primes.

ii. Find the atoms of $\mathbf{D}_m$.

(c) We need only show that $\mathbf{D}_m$ is complemented. Let x be in $\mathbf{D}_m$ and let $x' = m/x$. Since m is a product of distinct primes, x and x' have different prime divisors. Hence $x * x' = \text{gcd}(x, x') = 1$ and $x + x' = 1\text{cm}(x, x') = m$. Recall that 1 is the zero element (lower bound) of $\mathbf{D}_m$ and that m is the identity element (upper bound) of $\mathbf{D}_m$. Thus x' is a complement of x, and so $\mathbf{D}_m$ is a Boolean algebra.

(d) The atoms of $\mathbf{D}_m$ are the prime divisors of m.

b. Consider the Boolean algebra $\mathbf{D}_{210}$.

i. List its elements and draw its diagram.

ii. Find the set A of atoms.

iii. Find two subalgebras with eight elements.

iv. Is $X = \{1, 2, 6, 210\}$ a sublattice of $\mathbf{D}_{210}$? A subalgebra?

v. Is $Y = \{1, 2, 3, 6\}$ a sublattice of $\mathbf{D}_{210}$? A subalgebra?

(*a*) The divisors of 210 are 1, 2, 3, 5, 6, 7, 10, 14, 15, 21, 30, 35, 42, 70, 105, and 210. The diagram of $\mathbf{D}_{210}$ appears

(*b*) $A = \{2, 3, 5, 7\}$, the set of prime divisors of 210.

(*c*) $B = \{1, 2, 3, 35, 6, 70, 105, 210\}$ and $C = \{1, 5, 6, 7, 30, 35, 42, 210\}$ are subalgebras of $\mathbf{D}_{210}$.

(f) X is a sublattice since it is linearly ordered. However, X is not a subalgebra since 35 is the complement of 2 in $\mathbf{D}_{210}$ but 35 does not belong to X. (In fact, no Boolean algebra with more than two elements is linearly ordered.)

(g) Y is a sublattice of $\mathbf{D}_{210}$ since it is closed under + and ∗. However, Y is not a subalgebra of $\mathbf{D}_{210}$ since it is not closed under complements in $\mathbf{D}_{210}$, e.g., $35 = 2'$ does not belong to Y. (We note that Y itself is a Boolean algebra; in fact, $Y = \mathbf{D}_6$.)

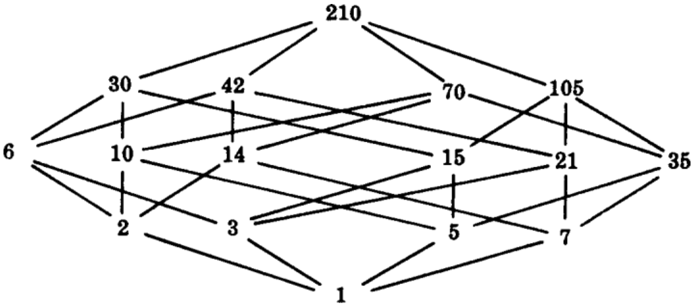


Fig. 15-25

c. Find the number of subalgebras of $\mathbf{D}_{210}$.

A subalgebra of $\mathbf{D}_{210}$ must contain two, four, eight or sixteen elements.

(v) There can be only one two-element subalgebra which consists of the upper bound 210 and lower bound 1, i.e.,
$\{1, 210\}$.

(vi) Since $\mathbf{D}_{210}$ contains sixteen elements, the only sixteen-element subalgebra is $\mathbf{D}_{210}$ itself.

(vii) Any four-element subalgebra is of the form $\{1, x, x^{J}, 210\}$, i.e., consists of the upper and lower bounds and a nonbound element and its complement. There are fourteen nonbound elements in $\mathbf{D}_{210}$ and so there are $14/2 = 7$ pairs $\{x, x^{J}\}$. Thus $\mathbf{D}_{210}$ has seven four-element subalgebras.

(viii) Any eight-element subalgebra S will itself contain three atoms s_1, s_2, s_3. We can choose s_1 and s_2 to be any two of the four atoms of $\mathbf{D}_{210}$ and then s_3 must be the product of the other two atoms, e.g., we can let $s_1 = 2$, $s_2 = 3$, $s_3 = 5 \cdot 7 = 35$ (which determines the subalgebra B 2 above), or we can let $s_1 = 5$, $s_2 = 7$, $s_3 = 2 \cdot 3 = 6$(which determines the subalgebra C above). There $4 \cdot$ $= 6$ ways to choose s_1 and s_2 from the four atoms of $\mathbf{D}_{210}$ are and so $\mathbf{D}_{210}$ has six eight-element

Accordingly, $\mathbf{D}_{210}$ has $1 + 1 + 7 + 6 = 15$ subalgebras.

d. Prove Theorem 15.2: Let a, b, c be any element in a Boolean algebra B.

(v) Idempotent laws:

(5a) $a + a = a$ *(5b)* $a * a = a$

(vi) Boundedness laws:

(6a) $a + 1 = 1$ *(6b)* $a * 0 = 0$

(vii) Absorption Laws:

(7a) $a + (a * b) = a$ *(7b)* $a * (a + b) = a$

(viii) Associative Laws:

(8a) $(a + b) + c = a + (b + c)$ *(8b)* $(a * b) * c = a * (b * c)$

(5b) $a = a * 1 = a * (a + a^{J}) = (a * a) + (a * a^{J}) = (a * a) + 0 = a * a$ *(5a)*
Follows from *(5b)* and duality.

(6b) $a * 0 = (a * 0) + 0 = (a * 0) + (a * a^{J}) = a * (0 + a^{J}) = a * (a^{J} + 0) = a * a^{J} = 0$

(6a) Follows from *(6b)* and duality.

(7b) $a * (a + b) = (a + 0) * (a + b) = a + (0 * b) = a + (b * 0) = a + 0 = a$ *(7a)*
Follows from *(7b)* and duality.

(8b) Let $L = (a * b) * c$ and $R = a * (b * c)$. We need to prove that $L = R$. We first prove that $a + L = a + R$. Using the absorption laws in the last two steps,

$$a + L = a + ((a * b) * c) = (a + (a * b)) * (a + c) = a * (a + c) = a$$

Also, using the absorption law in the last step,

$$a + R = a + (a * (b * c)) = (a + a) * (a + (b * c)) = a * (a + (b * c)) = a$$

Thus $a + L = a + R$. Next we show that $a' + L = a' + R$. We have,

$$a' + L = a' + ((a * b) * c) = (a' + (a * b)) * (a' + c)$$

$$= ((a' + a) * (a' + b)) * (a' + c) = (1 * (a' + b)) * (a' + c)$$

$$= (a' + b) * (a' + c) = a' + (b * c)$$

Also,

$$a' + R = a' + (a * (b * c)) = (a' + a) * (a' + (b * c))$$

$$= 1 * (a' + (b * c)) = a' + (b * c)$$

Thus $a' + L = a' + R$. Consequently,

$$L = 0 + L = (a * a') + L = (a + L) * (a' + L) = (a + R) * (a' + R)$$

$$= (a * a') + R = 0 + R = R$$

(8a) Follows from *(8b)* and duality.

e. Prove Theorem 15.3: Let a be any element of a Boolean algebra B.

(iii) (Uniqueness of Complement) If $a + x = 1$ and $a * x = 0$, then $x = a'$.

(iv) (Involution Law) $(a')' = a$

(iii) *(9a)* $0' = 1$; *(9b)* $1' = 0$.

(iv) We have:

$$a' = a' + 0 = a' + (a * x) = (a' + a) * (a' + x) = 1 * (a' + x) = a' + x$$

Also,

$$x = x + 0 = x + (a * a') = (x + a) * (x + a') = 1 * (x + a') = x + a'$$

Hence $x = x + a' = a' + x = a'$.

(v) By definition of complement, $a + a' = 1$ and $a * a' = 0$. By commutativity, $a' + a = 1$ and $a' * a = 0$. By uniqueness of complement, a is the complement of a', that is, $a = (a')'$.

(vi) By boundedness law *(6a)*, $0 + 1 = 1$, and by identity axiom *(3b)*, $0 * 1 = 0$. By uniqueness of complement, 1 is the complement of 0, that is, $1 = 0'$. By duality, $0 = 1'$.

f. Prove Theorem 15.4: (DeMorgan's laws): *(10a)* $(a + b)' = a' * b'$. *(10b)* $(a * b)' = a' + b'$.

(10a) We need to show that $(a + b) + (a' * b') = 1$ and $(a + b) * (a' * b') = 0$; then by uniqueness of complement,

$a' \ast b' = (a + b')$. We have:

$$(a + b) + (a' \ast b') = b + a + (a' \ast b') = b + (a + a') \ast (a + b')$$

$$= b + 1 \ast (a + b') = b + a + b' = b + b' + a = 1 + a = 1$$

Also,

$$(a + b) \ast (a' \ast b') = ((a + b) \ast a') \ast b'$$

$$= ((a \ast a') + (b \ast a')) \ast b' = (0 + (b \ast a')) \ast b'$$

$$= (b \ast a') \ast b' = (b \ast b') \ast a' = 0 \ast a' = 0$$

Thus $a' \ast b' = (a + b)'$.

Prove Theorem : The following are equivalent in a Boolean algebra:

$(1)\ a + b = b$; $(2)\ a \ast b = a$; $(3)\ a' + b = 1$; $(4)\ a \ast b' = 0$.

By Theorem 14.4, (*1*) and (*2*) are equivalent. We show that (*1*) and (*3*) are equivalent. Suppose (*1*) holds. Then

$$a' + b = a' + (a + b) = (a' + a) + b = 1 + b = 1$$

Now suppose *(3)* holds. Then

$$a + b = 1 \ast (a + b) = (a' + b) \ast (a + b) = (a' \ast a) + b = 0 + b = b$$

Thus *(1)* and *(3)* are equivalent.

We next show that *(3)* and *(4)* are equivalent. Suppose *(3)* holds. By DeMorgan's law and involution,

$$0 = 1' = (a' + b')' = a'' \ast b' = a \ast b'$$

Conversely, if *(4)* holds then

$$1 = 0' = (a \ast b')' = a' + b'' = a' + b$$

Thus *(3)* and *(4)* are equivalent. Accordingly, all four are equivalent.

g. Prove Theorem 15.6: The mapping f: $B \to P(A)$ is an isomorphism where B is a Boolean algebra, $P(A)$ is the power set of the set A of atoms, and

$$f(x) = \{a_1, a_2, \ldots, a_n\}$$

where $x = a_1 + \cdots + a_n$ is the unique representation of a as a sum of atoms.

Recall (Chapter 14) that if the a's are atoms then $a_i^2 = a_i$ but $a_i a_j = 0$ for $a_i \neq a_j$. Suppose x, y are in B and suppose

$$x = a_1 + \cdots + a_r + b_1 + \cdots + b_s$$
$$y = b_1 + \cdots + b_s + c_1 + \cdots + c_t$$

where

$$A = \{a_1,\ldots, a_r, b_1,\ldots, b_s, c_1,\ldots, c_t, d_1,\ldots, d_k\}$$

is the set of atoms of B. Then

$$x + y = a_1 + \cdots + a_r + b_1 + \cdots + b_s + c_1 + \cdots + c_t$$
$$xy = b_1 + \cdots b_s$$

Hence

$$f\,(x + y) = \{a_1,\ldots, a_r, b_1,\ldots, b_s, c_1,\ldots, c_t\}$$

$$= \{a_1,\ldots, a_r, b_1,\ldots, b_s\} \cup \{b_1,\ldots, b_s, c_1,\ldots, c_t\}$$

$$= f(x) \cup f(y)$$
$$f(xy) = \{b_1,\ldots, b_s\}$$

$$= \{a_1,\ldots, a_r, b_1,\ldots, b_s\} \cap \{b_1,\ldots, b_s, c_1,\ldots, c_t\}$$

Let

$$= f\,(x) \cap f\,(y)$$

Thus

$y = c_1 + \cdots + c_t + d_1 + \cdots + d_k$. Then $x + y = 1$ and $xy = 0$, and so $y = x^{J}$

$$f(x^{J}) = \{c_1, \ldots, c_t, d_1, \ldots, d_k\} = \{a_1, \ldots, a_r, b_1, \ldots, b_s\}^c = (f(x))^c$$

Since the representation is unique, f is one-to-one and onto. Hence f is a Boolean algebra isomorphism.

SUBJECT:DISCRETE MATHEMATICS

Code:CSE 208-C **B.Tech CSE 4th SEM**

Unit: Graphs and Trees

Definition: The is a figure consist of points or nodes called vertices which are connected to each other by way of lines called edges. These lines may be directed or undirected. A graph G consists of two things:

(i) A set V = V (G) whose elements are called vertices, points, or nodes of G.

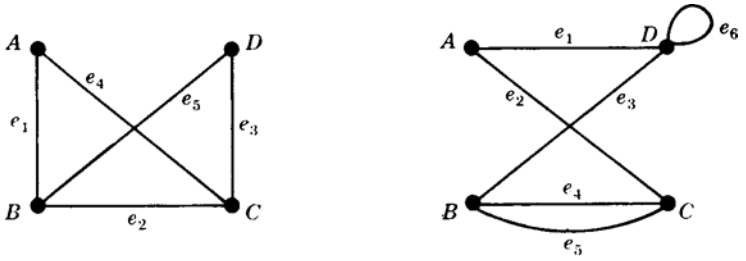
(ii) A set E = E(G) of unordered pairs of distinct vertices called edges of G.

We denote such a graph by G(V, E) when we want to emphasize the two parts of G. Vertices u and v are said to be adjacent or neighbors if there is an edge e u, v . In such a case, u and v are called the endpoints of e, and e is said to connect u and v. Also, the edge e is said to be incident on each of its endpoints u and v. Graphs are pictured by diagrams in the plane in a natural way. Specifically, each vertex v in V is represented by a dot (or small circle), and each edge e = {v1, v2} is represented by a curve which connects its endpoints v1 and v2 For example, Fig. 8-5(a) represents the graph G(V, E) where:

(i) V consists of vertices A, B, C, D.

(ii) E consists of edges e1 = {A, B}, e2 = {B, C}, e3 = {C, D}, e4 = {A, C}, e5 = {B, D}.

In fact, we will usually denote a graph by drawing its diagram rather than explicitly listing its vertices and edges.



(*a*) Graph (*b*) Multigraph

Fig. 1

Multigraphs

Consider the diagram in Fig. 1(b). The edges e4 and e5 are called multiple edges since they connect the same endpoints, and the edge e6 is called a loop since its endpoints are the same vertex. Such a diagram is called a multigraph; the formal definition of a graph permits neither multiple edges nor loops. Thus a graph may be defined to be a multigraph without multiple edges or loops.

Remark: Some texts use the term graph to include multigraphs and use the term simple graph to mean a graph without multiple edges and loops.

Degree of a Vertex

The degree of a vertex v in a graph G, written deg (v), is equal to the number of edges in G which contain v, that is, which are incident on v. Since each edge is counted twice in counting the degrees of the vertices of G, we have the following simple but important result.

Theorem 1: The sum of the degrees of the vertices of a graph G is equal to twice the number of edges in G.

Consider, for example, the graph in Fig. 1(a). We have

deg(A) = 2, deg(B) = 3, deg(C) = 3, deg(D) = 2.

The sum of the degrees equals 10 which, as expected, is twice the number of edges. A vertex is said to be even or odd according as its degree is an even or an odd number. Thus A and D are even vertices whereas B and C are odd vertices. Theorem 8.1: The sum of the degrees of the vertices of a graph G is equal to twice the number of edges in G.

Theorem 1 also holds for multigraphs where a loop is counted twice toward the degree of its endpoint. For example, in Fig. 1(b) we have deg(D) = 4 since the edge e6 is counted twice; hence D is an even vertex.

A vertex of degree zero is called an isolated vertex.

Finite Graphs, Trivial Graph

A multigraph is said to be finite if it has a finite number of vertices and a finite number of edges. Observe that a graph with a finite number of vertices must automatically have a finite number of edges and so must be finite. The finite graph with one vertex and no edges, i.e., a single point, is called the trivial graph. Unless otherwise specified, the multigraphs in this book shall be finite.

SUBGRAPHS, ISOMORPHIC AND HOMEOMORPHIC GRAPHS

Subgraphs

Consider a graph G = G(V, E). A graph H = H(V J, EJ) is called a subgraph of G if the vertices and edges of H are contained in the vertices and edges of G, that is, if V J ⊆ V and EJ ⊆ E. In particular:

(i) A subgraph H(V J, EJ) of G(V, E) is called the subgraph induced by its vertices V J if its edge set EJ

contains all edges in G whose endpoints belong to vertices in H .

(ii) If v is a vertex in G, then G v is the subgraph of G obtained by deleting v from G and deleting all edges in G which contain v.

(iii) If e is an edge in G, then G − e is the subgraph of G obtained by simply deleting the edge e from G.

Isomorphic Graphs

Graphs G(V, E) and G(V ∗, E∗) are said to be isomorphic if there exists a one-to-one correspondence f V V ∗ such that u, v is an edge of G if and only if f (u), f(v) is an edge of G∗. Normally, we do not distinguish between isomorphic graphs (even though their diagrams may “look different”). Figure 2 gives ten graphs pictured as letters. We note that A and R are isomorphic graphs. Also, F and T are isomorphic graphs, K and X are isomorphic graphs and M, S, V , and Z are isomorphic graphs

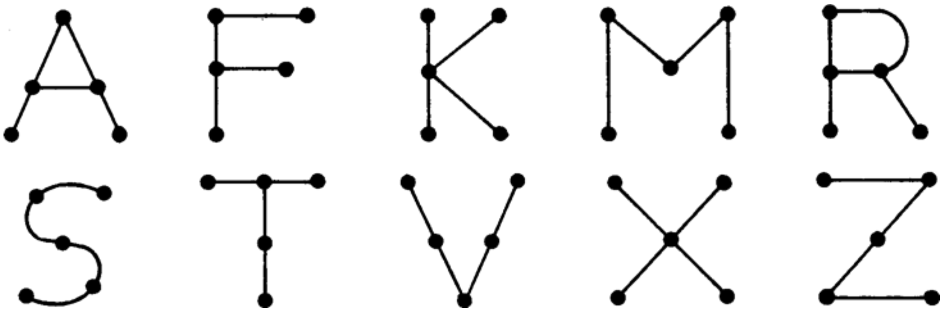


Fig.2

Homeomorphic Graphs

Given any graph G, we can obtain a new graph by dividing an edge of G with additional vertices. Two graphs G and G∗ are said to homeomorphic if they can be obtained from the same graph or isomorphic graphs by this method. The graphs (a) and (b) in Fig. 3 are not isomorphic, but they are homeomorphic since they can be obtained from the graph (c) by adding appropriate vertices.

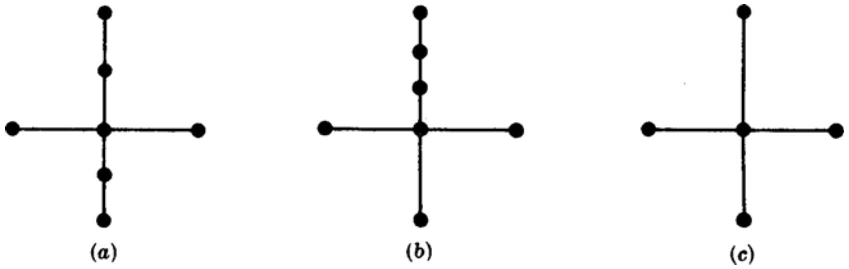


Fig.3

8.4 PATHS, CONNECTIVITY

A path in a multigraph G consists of an alternating sequence of vertices and edges of the form

v0, e1, v1, e2, v2, ..., en−1, vn−1, en, vn

where each edge ei contains the vertices vi 1 and vi (which appear on the sides of ei in the sequence). The number n of edges is called the length of the path. When there is no ambiguity, we denote a path by its sequence of vertices (v0, v1,..., vn). The path is said to be closed if v0 = vn. Otherwise, we say the path is from v0, to vn or between v0 and vn, or connects v0 to vn.

A simple path is a path in which all vertices are distinct. (A path in which all edges are distinct will be called

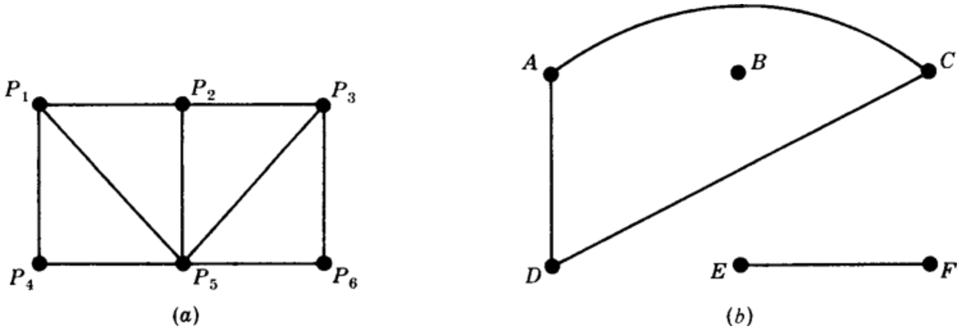
a trail.) A cycle is a closed path of length 3 or more in which all vertices are distinct except v0 vn. A cycle of length k is called a k-cycle.

EXAMPLE Consider the graph G in Fig. 4(a). Consider the following sequences:

α = (P4, P1, P2, P5, P1, P2, P3, P6), β = (P4, P1, P5, P2, P6),

γ = (P4, P1, P5, P2, P3, P5, P6), δ = (P4, P1, P5, P3, P6).

The sequence α is a path from P4 to P6; but it is not a trail since the edge {P1, P2} is used twice. The sequence β is not a path since there is no edge {P2, P6}. The sequence γ is a trail since no edge is used twice; but it is not a simple path since the vertex P5 is used twice. The sequence δ is a simple path from P4 to P6; but it is not the shortest path (with respect to length) from P4 to P6. The shortest path from P4 to P6 is the simple path (P4, P5, P6) which has length 2.



(a)

(b)

Fig.4

By eliminating unnecessary edges, it is not difficult to see that any path from a vertex u to a vertex v can be replaced by a simple path from u to v. We state this result formally.

Theorem 2: There is a path from a vertex u to a vertex v if and only if there exists a simple path from u to v.

Connectivity, Connected Components

A graph G is connected if there is a path between any two of its vertices. The graph in Fig. 8-8(a) is connected, but the graph in Fig. 4(b) is not connected since, for example, there is no path between vertices D and E.

Suppose G is a graph. A connected subgraph H of G is called a connected component of G if H is not contained in any larger connected subgraph of G. It is intuitively clear that any graph G can be partitioned into its connected components. For example, the graph G in Fig. 4(b) has three connected components, the subgraphs induced by the vertex sets A, C, D , E, F , and B .

The vertex B in Fig. 4(b) is called an isolated vertex since B does not belong to any edge or, in other words, deg(B) = 0. Therefore, as noted, B itself forms a connected component of the graph.

Remark: Formally speaking, assuming any vertex u is connected to itself, the relation “u is connected to v” is an equivalence relation on the vertex set of a graph G and the equivalence classes of the relation form the connected components of G.

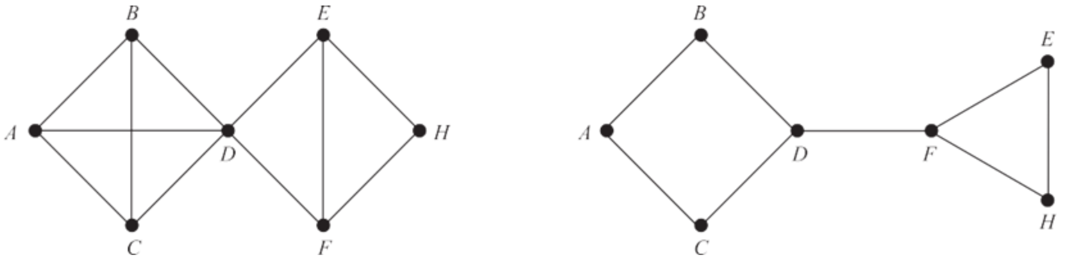
Distance and Diameter

Consider a connected graph G. The distance between vertices u and v in G, written d(u, v), is the length of the shortest path between u and v. The diameter of G, written diam(G), is the maximum distance between any two points in G. For example, in Fig. 5(a), d(A, F) = 2 and diam(G) = 3, whereas in Fig. 5(b),

d(A, F) = 3 and diam(G) = 4.

Cutpoints and Bridges

Let G be a connected graph. A vertex v in G is called a cutpoint if G v is disconnected. (Recall that G v is the graph obtained from G by deleting v and all edges containing v.) An edge e of G is called a bridge if G e is disconnected. (Recall that G e is the graph obtained from G by simply deleting the edge e). In Fig. 5(a), the vertex D is a cutpoint and there are no bridges. In Fig. 5(b), the edge D, F is a bridge. (Its endpoints D and F are necessarily cutpoints.)

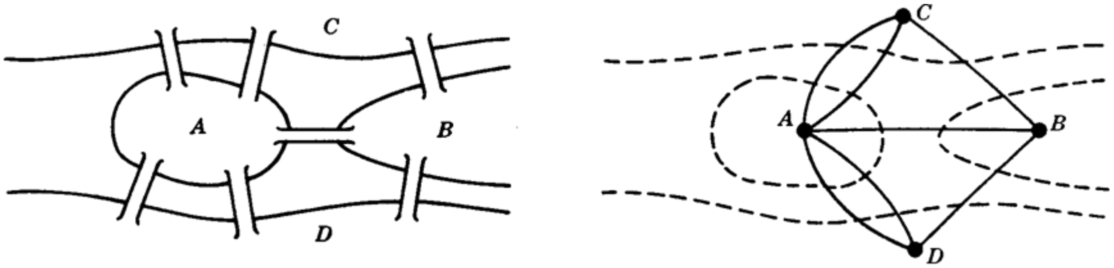


(a) Fig.5 (b)

TRAVERSABLE AND EULERIAN GRAPHS, BRIDGES OF KÖNIGSBERG

The eighteenth-century East Prussian town of Königsberg included two islands and seven bridges as shown in Fig.6(a). Question: Beginning anywhere and ending anywhere, can a person walk through town crossing all seven bridges but not crossing any bridge twice? The people of Königsberg wrote to the celebrated Swiss mathematician L. Euler about this question. Euler proved in 1736 that such a walk is impossible. He replaced the islands and the two sides of the river by points and the bridges by curves, obtaining Fig. 6(b).

Observe that Fig. 6(b) is a multigraph. A multigraph is said to be traversable if it “can be drawn without any breaks in the curve and without repeating any edges,” that is, if there is a path which includes all vertices and uses each edge exactly once. Such a path must be a trail (since no edge is used twice) and will be called a traversable trail. Clearly a traversable multigraph must be finite and connected.



(a) Königsberg in 1736

(b) Euler’s graphical representation

Fig.6

We now show how Euler proved that the multigraph in Fig. 8-10(b) is not traversable and hence that the walk in Königsberg is impossible. Recall first that a vertex is even or odd according as its degree is an even or an odd number. Suppose a multigraph is traversable and that a traversable trail does not begin or end at a vertex P . We claim that P is an even vertex. For whenever the traversable trail enters P by an edge, there must always be an edge not previously used by which the trail can leave P . Thus the edges in the trail incident with P must appear in pairs, and so P is an even vertex. Therefore if a vertex Q is odd, the traversable trail must begin or end at Q. Consequently, a multigraph with more than two odd vertices cannot be traversable. Observe that the multigraph corresponding to the Königsberg bridge problem has four odd vertices. Thus one cannot walk through Königsberg so that each bridge is crossed exactly once.

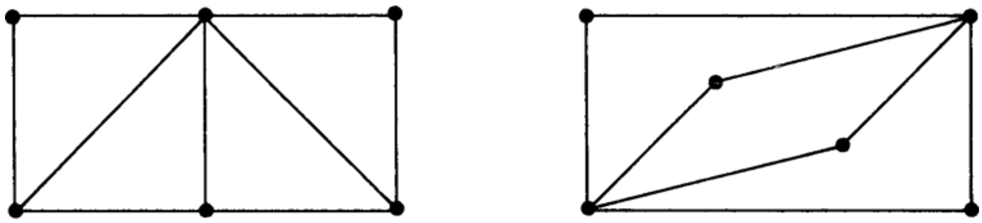
Euler actually proved the converse of the above statement, which is contained in the following theorem and corollary. A graph G is called an Eulerian graph if there exists a closed traversable trail, called an Eulerian trail.

Theorem .3 (Euler): A finite connected graph is Eulerian if and only if each vertex has even degree.

Corollary.4: Any finite connected graph with two odd vertices is traversable. A traversable trail may begin at either odd vertex and will end at the other odd vertex.

Hamiltonian Graphs

The above discussion of Eulerian graphs emphasized traveling edges; here we concentrate on visiting vertices. A Hamiltonian circuit in a graph G, named after the nineteenth-century Irish mathematician William Hamilton (1803–1865), is a closed path that visits every vertex in G exactly once. (Such a closed path must be a cycle.) If G does admit a Hamiltonian circuit, then G is called a Hamiltonian graph. Note that an Eulerian circuit traverses every edge exactly once, but may repeat vertices, while a Hamiltonian circuit visits each vertex exactly once but may repeat edges. Figure 7 gives an example of a graph which is Hamiltonian but not Eulerian, and vice versa.



(*a*) Hamiltonian and non-Eulerian

(*b*) Eulerian and non-Hamiltonian

Fig 7

Although it is clear that only connected graphs can be Hamiltonian, there is no simple criterion to tell us whether or not a graph is Hamiltonian as there is for Eulerian graphs. We do have the following sufficient condition which is due to G. A. Dirac.

Theorem 5: Let G be a connected graph with n vertices. Then G is Hamiltonian if n 3 and n deg(v) for each vertex v in G.

LABELED AND WEIGHTED GRAPHS

A graph G is called a labeled graph if its edges and/or vertices are assigned data of one kind or another. In particular, G is called a weighted graph if each edge e of G is assigned a nonnegative number w(e) called the weight or length of v. Figure 8 shows a weighted graph where the weight of each edge is given in the obvious way. The weight (or length) of a path in such a weighted graph G is defined to be the sum of the weights of the edges in the path. One important problem in graph theory is to find a shortest path, that is, a path of minimum weight (length), between any two given vertices. The length of a shortest path between P and Q in Fig. 8 is 14; one such path is

(P, A1, A2, A5, A3, A6, Q)

The reader can try to find another shortest path.

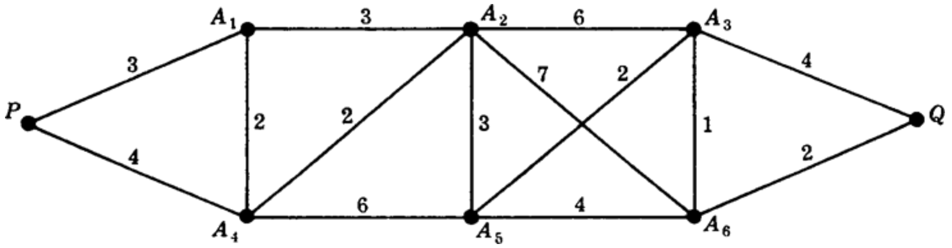


Fig 8

COMPLETE, REGULAR, AND BIPARTITE GRAPHS

There are many different types of graphs. This section considers three of them: complete, regular, and bipartite graphs.

Complete Graphs

A graph G is said to be complete if every vertex in G is connected to every other vertex in G. Thus a complete graph G must be connected. The complete graph with n vertices is denoted by Kn. Figure 9 shows the graphs K1 through K6.

Regular Graphs

A graph G is regular of degree k or k-regular if every vertex has degree k. In other words, a graph is regular if every vertex has the same degree.

The connected regular graphs of degrees 0, 1, or 2 are easily described. The connected 0-regular graph is the trivial graph with one vertex and no edges. The connected 1-regular graph is the graph with two vertices and one edge connecting them. The connected 2-regular graph with n vertices is the graph which consists of a single n-cycle. See Fig. 10.

The 3-regular graphs must have an even number of vertices since the sum of the degrees of the vertices is an even number (Theorem 1). Figure 11 shows two connected 3-regular graphs with six vertices. In general, regular graphs can be quite complicated. For example, there are nineteen 3-regular graphs with ten vertices. We note that the complete graph with n vertices Kn is regular of degree n − 1.

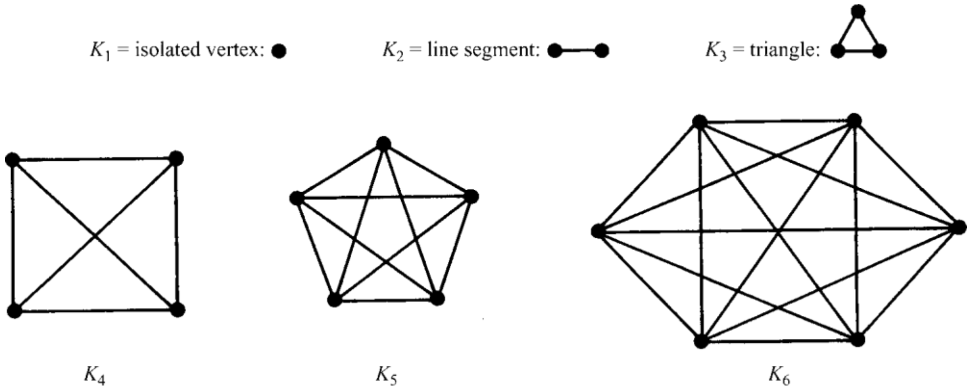


Fig.9

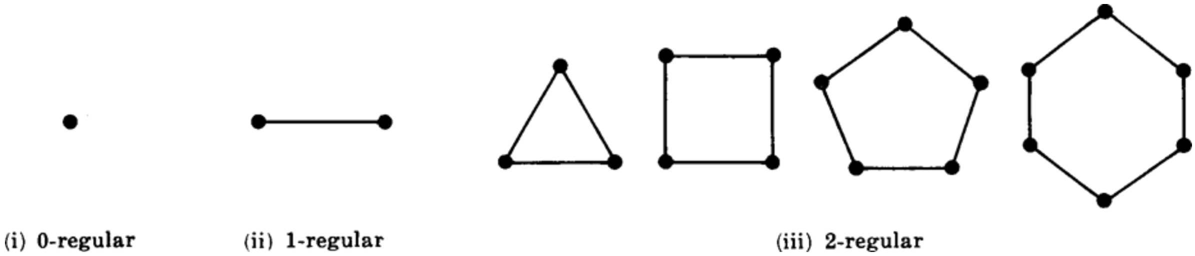


Fig 10.

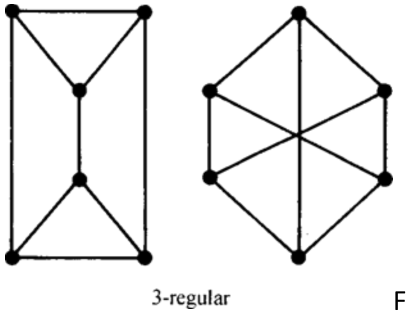


Fig 11

Bipartite Graphs

A graph G is said to be bipartite if its vertices V can be partitioned into two subsets M and N such that each edge of G connects a vertex of M to a vertex of N . By a complete bipartite graph, we mean that each vertex of M is connected to each vertex of N ; this graph is denoted by Km,n where m is the number of vertices in M and

n is the number of vertices in N , and, for standardization, we will assume m ≤ n. Figure 12 shows the graphs

K2,3, K3,3, and K2,4, Clearly the graph Km,n has mn edges.

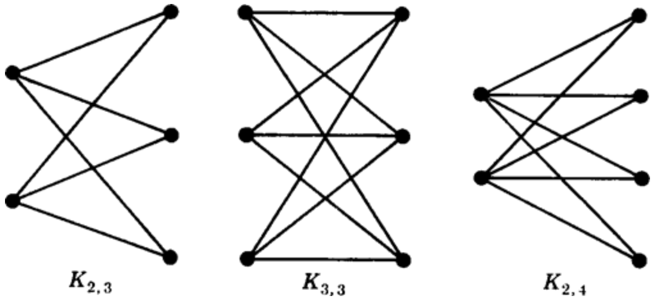


Fig.12

TREE GRAPHS

A graph T is called a tree if T is connected and T has no cycles. Examples of trees are shown in Fig. 8-17. A forest G is a graph with no cycles; hence the connected components of a forest G are trees. A graph without cycles is said to be cycle-free. The tree consisting of a single vertex with no edges is called the degenerate tree.

Consider a tree T . Clearly, there is only one simple path between two vertices of T ; otherwise, the two paths would form a cycle. Also:

(a) Suppose there is no edge u, v in T and we add the edge e u, v to T . Then the simple path from u to v

in T and e will form a cycle; hence T is no longer a tree.

(b) On the other hand, suppose there is an edge e u, v in T , and we delete e from T . Then T is no longer connected (since there cannot be a path from u to v); hence T is no longer a tree.

The following theorem applies when our graphs are finite.

Theorem 6: Let G be a graph with n > 1 vertices. Then the following are equivalent:

(i) G is a tree.

(ii) G is a cycle-free and has n − 1 edges.

(iii) G is connected and has n − 1 edges.

This theorem also tells us that a finite tree T with n vertices must have n 1 edges. For example, the tree in Fig.1 3(a) has 9 vertices and 8 edges, and the tree in Fig. 13(b) has 13 vertices and 12 edges.

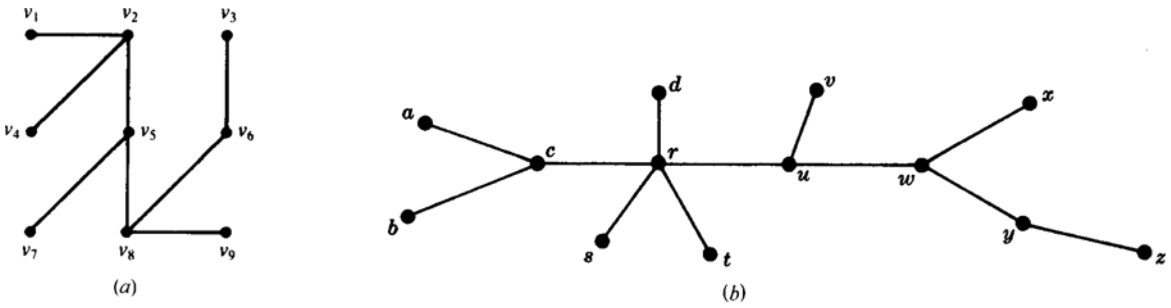


Fig 13

Spanning Trees

A subgraph T of a connected graph G is called a spanning tree of G if T is a tree and T includes all the vertices of G. Figure 14 shows a connected graph G and spanning trees T1, T2, and T3 of G.

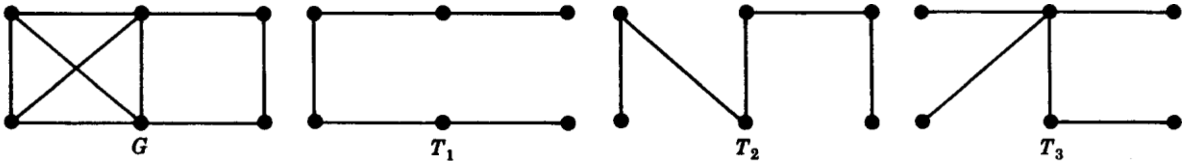


Fig 14.

8.1 PLANAR GRAPHS

A graph or multigraph which can be drawn in the plane so that its edges do not cross is said to be *planar*. Although the complete graph with four vertices K_4 is usually pictured with crossing edges as in Fig.15*(a)*, it can also be drawn with noncrossing edges as in Fig. 15*(b)*; hence K_4 is planar. Tree graphs form an important class of planar graphs. This section introduces our reader to these important graphs.

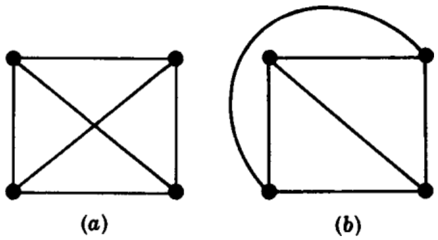


Fig 15

Maps, Regions

A particular planar representation of a finite planar multigraph is called a *map*. We say that the map is *connected* if the underlying multigraph is connected. A given map divides the plane into various regions. For example, the map in Fig. 16 with six vertices and nine edges divides the plane into five regions. Observe that four of the regions are bounded, but the fifth region, outside the diagram, is unbounded. Thus there is no loss in generality in counting the number of regions if we assume that our map is contained in some large rectangle rather than in the entire plane.

Observe that the border of each region of a map consists of edges. Sometimes the edges will form a cycle, but sometimes not. For example, in Fig. 16 the borders of all the regions are cycles except for r_3. However, if we do move counterclockwise around r_3 starting, say, at the vertex C, then we obtain the closed path

$$(C, D, E, F, E, C)$$

where the edge $\{E, F\}$ occurs twice. By the *degree* of a region r, written deg(r), we mean the length of the cycle or closed walk which borders r. We note that each edge either borders two regions or is contained in a region and will occur twice in any walk along the border of the region. Thus we have a theorem for regions which is analogous to Theorem 1 for vertices.

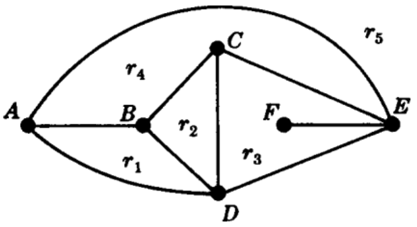


Fig. 16

Theorem .7: The sum of the degrees of the regions of a map is equal to twice the number of edges.

The degrees of the regions of Fig. 16 are:

$$\deg(r_1) = 3, \quad \deg(r_2) = 3, \quad \deg(r_3) = 5, \quad \deg(r_4) = 4, \quad \deg(r_5) = 3$$

The sum of the degrees is 18, which, as expected, is twice the number of edges.

For notational convenience we shall picture the vertices of a map with dots or small circles, or we shall assume that any intersections of lines or curves in the plane are vertices.

Euler's Formula

Euler gave a formula which connects the number V of vertices, the number E of edges, and the number R of regions of any connected map. Specifically:

Theorem 8 (Euler): $V - E + R = 2$.

(The proof of Theorem 8.8 appears in Problem 8.18.)

Observe that, in Fig. 8-22, $V = 6$, $E = 9$, and $R = 5$; and, as expected by Euler's formula.

$$V - E + R = 6 - 9 + 5 = 2$$

We emphasize that the underlying graph of a map must be connected in order for Euler's formula to hold.

Let G be a connected planar multigraph with three or more vertices, so G is neither K_1 nor K_2. Let M be a planar representation of G. It is not difficult to see that (1) a region of M can have degree 1 only if its border is a loop, and (2) a region of M can have degree 2 only if its border consists of two multiple edges. Accordingly, if G is a graph, not a multigraph, then every region of M must have degree 3 or more. This comment together with Euler's formula is used to prove the following result on planar graphs.

Theorem 9: Let G be a connected planar graph with p vertices and q edges, where $p \geq 3$. Then $q \geq 3p - 6$.

Note that the theorem is not true for K_1 where $p = 1$ and $q = 0$, and is not true for K_2 where $p = 2$ and $q - 1$.

Proof: Let r be the number of regions in a planar representation of G. By Euler's formula, $p - q + r = 2$. Now the sum of the degrees of the regions equals $2q$ by Theorem 8.7. But each region has degree 3 or more; hence $2q \geq 3r$. Thus $r \geq 2q/3$. Substituting this in Euler's formula gives

$$2 = p - q + r \leq p - q + \frac{2q}{3} \quad \text{or} \quad 2 \leq p - \frac{q}{3}$$

Multiplying the inequality by 3 gives $6 \leq 3p - q$ which gives us our result. □

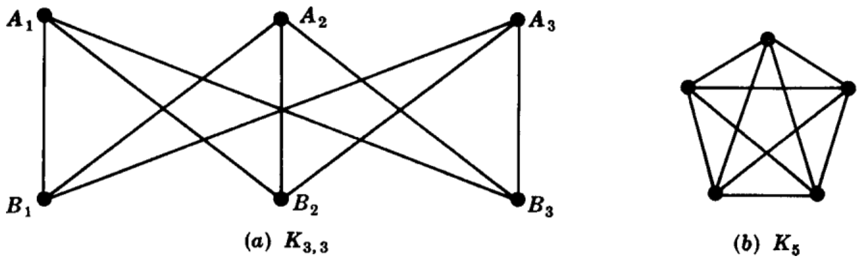
Nonplanar Graphs, Kuratowski's Theorem

We give two examples of nonplanar graphs. Consider first the *utility graph*; that is, three houses A_1, A_2, A_3 are to be connected to outlets for water, gas and electricity, B_1, B_2, B_3, as in Fig. 17*(a)*. Observe that this is the graph $K_{3,3}$ and it has $p = 6$ vertices and $q = 9$ edges. Suppose the graph is planar. By Euler's formula a planar representation has $r = 5$ regions. Observe that no three vertices are connected to each other; hence the degree of each region must be 4 or more and so the sum of the degrees of the regions must be 20 or more. By Theorem 8.7 the graph must have 10 or more edges. This contradicts the fact that the graph has $q = 9$ edges. Thus the utility graph $K_{3,3}$ is nonplanar.

Consider next the *star graph* in Fig. 17*(b)*. This is the complete graph K_5 on $p = 5$ vertices and has $q = 10$ edges. If the graph is planar, then by Theorem 9.

$$10 = q \leq 3p - 6 = 15 - 6 = 9$$

which is impossible. Thus K_5 is nonplanar.



(a) $K_{3,3}$ (b) K_5

Fig. 17

Theorem 10: (Kuratowski) A graph is nonplanar if and only if it contains a subgraph homeomorphic to $K_{3,3}$ or K_5.

8.2 GRAPH COLORINGS

Consider a graph *G*. *A vertex coloring*, or simply a *coloring* of *G* is an assignment of colors to the vertices of *G* such that adjacent vertices have different colors. We say that *G* is *n*-colorable if there exists a coloring of *G*which uses *n* colors. for example, “paint” *G* rather than “color” *G* when we are assigning colors to the vertices of *G*.) The minimum number of colors needed to paint *G* is called the *chromatic number* of *G* and is denoted by *χ(G)*.

Fig. 18 gives an algorithm by Welch and Powell for a coloring of a graph G. We emphasize that this algorithm does not always yield a minimal coloring of *G*.

Algorithm 8.4 (Welch-Powell): The input is a graph *G*.

Step 1. Order the vertices of *G* according to decreasing degrees.

Step 2. Assign the first color C_1 to the first vertex and then, in sequential order, assign C_1 to each vertex which is not adjacent to a previous vertex which was assigned C_1.

Step 3. Repeat Step 2 with a second color C_2 and the subsequence of noncolored vertices.

Step 4. Repeat Step 3 with a third color C_3, then a fourth color C_4, and so on until all vertices are colored.

Step 5. Exit.

Fig 18

EXAMPLE

(a) Consider the graph *G* in Fig. 19. We use the Welch-Powell Algorithm 8.4 to obtain a coloring of *G*. Ordering the vertices according to decreasing degrees yields the following sequence:

$$A_5,\quad A_3,\quad A_7,\quad A_1,\quad A_2,\quad A_4,\quad A_6,\quad A_8$$

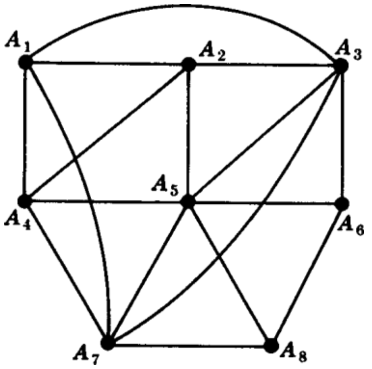


Fig. 19

The first color is assigned to vertices A_5 and A_1. The second color is assigned to vertices A_3, A_4, and A_8. The third color is assigned to vertices A_7, A_2, and A_6. All the vertices have been assigned a color, and so G is 3-colorable. Observe that G is not 2-colorable since vertices A_1, A_2, and A_3, which are connected to each other, must be assigned different colors. Accordingly, $\chi(G) = 3$.

(b) Consider the complete graph K_n with n vertices. Since every vertex is adjacent to every other vertex, K_n requires n colors in any coloring. Thus $\chi(K_n) = n$.

There is no simple way to actually determine whether an arbitrary graph is n-colorable. However, the following theorem gives a simple characterization of 2-colorable graphs.

Theorem 11: The following are equivalent for a graph G:

(i) G is 2-colorable.
(ii) G is bipartite.
(iii) Every cycle of G has even length.

There is no limit on the number of colors that may be required for a coloring of an arbitrary graph since, for example, the complete graph K_n requires n colors. However, if we restrict ourselves to planar graphs, regardless of the number of vertices, five colors suffice. Specifically, in Problem 8.20 we prove:

Theorem 12: Any planar graph is 5-colorable.

Solved Problems

GRAPH TERMINOLOGY

8.1. Consider the graph G in Fig. 8-36*(a)*.

(a) Describe G formally, that is, find the set $V(G)$ of vertices of G and the set $E(G)$ of edges of G.

(b) Find the degree of each vertex and verify Theorem 8.1 for this graph.

(a) There are five vertices so $V(G) = \{A, B, C, D, E$. T}here are seven pairs $x,\{y$ of }vertices where the vertex x is connected with the vertex y, hence

$$E(G) = [\{A, B\}, \{A, C\}, \{A, D\}, \{B, C\}, \{B, E\}, \{C, D\}, \{C, E\}]$$

(b) The degree of a vertex is equal to the number of edges to which it belongs; e.g., deg$(A) = 3$ since A belongs to the three edges $\{A, B\}$, $\{A, C\}$, $\{A, D\}$. Similarly,

$$\deg(B) = 3, \deg(C) = 4, \deg(D) = 2, \deg(E) = 2$$

The sum of the degrees is $3 + 3 + 4 + 2 + 2 = 14$ which does equal twice the number of edges.

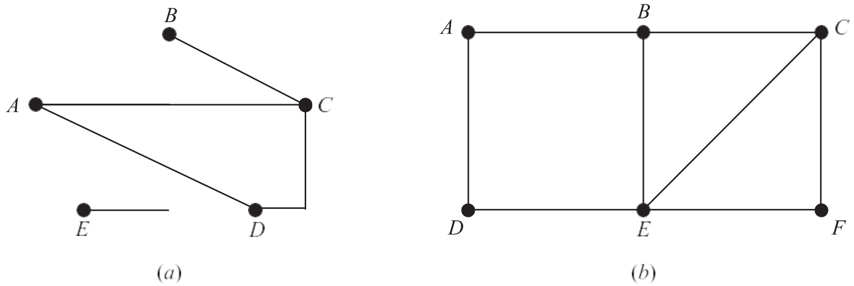


Fig. 8-36

8.2. Consider the graph G in Fig. 8-36*(b)*. Find:

(a) all simple paths from A to F; (*d*) diam(G), the diameter of G;
(b) all trails from A to F; (*e*) all cycles which include vertex A;
(c) $d(A, F)$, the distance from A to F; (*f*) all cycles in G.

(a) A simple path from A to F is a path such that no vertex, and hence no edge, is repeated. There are seven such paths, four beginning with the edges $\{A, B\}$ and three beginning with the ege $\{A, D\}$:

$$(A,B, C,F), \quad (A,B, C,E,F), \quad (A,B, E,F), \quad (A,B, E, C,F),$$
$$(A, D, E,F), \quad (A, D, E,B, C,F), \quad (A, D, E, C,F).$$

(b) A trail from A to F is a path such that no edge is repeated. There are nine such trails, the seven simple paths from
(a) together with

$$(A, D, E,B, C, E,F) \text{ and } (A, D, E, C,B, E,F).$$

(c) There is a path, e.g., (A,B, C,F), from A to F of length 3 and no shorter path from A to F; hence $d(A, F) = 3$.

(d) The distance between any two vertices is not greater than 3, and the distance from A to F is 3; hence diam$(G) = 3$.

(e) A cycle is a closed path in which no vertex is repeated (except the first and last). There are three cycles which include vertex A:

$$(A,B, E, D, A), \quad (A,B, C, E, D, A), \quad (A,B, C,F, E, D, A).$$

(f) There are six cycles in G; the three in *(e)* and

$$(B, C, E,B), \quad (C,F, E,C), \quad (B, C,F, E,B).$$

8.3. Consider the multigraphs in Fig. 8-37.

(a) Which of them are connected? If a graph is not connected, find its connected components.

(b) Which are cycle-free (without cycles)?

(c) Which are loop-free (without loops)?

(d) Which are (simple) graphs?

(a) Only (1) and (3) are connected, (2) is disconnected; its connected components are $\{A, D, E\}$ and $\{B, C\}$. (4) is disconnected; its connected components are $\{A, B, E\}$ and $\{C, D\}$.

(b) Only (1) and (4) are cycle-free. (2) has the cycle (A, D, E, A), and (3) has the cycle (A, B, E, A).

(c) Only (4) has a loop which is $\{B, B\}$.

(d) Only (1) and (2) are graphs. Multigraph (3) has multiple edges $\{A, E\}$ and $\{A, E\}$; and (4) has both multiple edges $\{C, D\}$ and $\{C, D\}$ and a loop $\{B, B\}$.

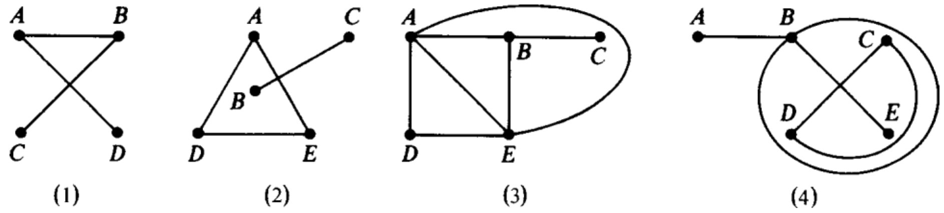


Fig. 8-37

8.4. Let G be the graph in Fig. 8-38*(a)*. Find:

(a) all simple paths from A to C; (*d*) $G - Y$;

(b) all cycles; (*e*) all cutpoints;

(c) subgraph H generated by $V^{J} = \{B, C, X, Y\}$; (*f*) all bridges.

(a) There are two simple paths from A to C: (A, X, Y, C) and (A, X, B, Y, C).

(b) There is only one cycle: (B, X, Y, B).

(c) As pictured in Fig. 8-38*(b)*, H consists of the vertices V^{J} and the set E^{J} of all edges whose endpoints belong to V^{J}, that is, $E^{J} = [\{B, X\}, \{X, Y\}, \{B, Y\}, \{C, Y\}]$.

(d) Delete vertex Y from G and all edges which contain Y to obtain the graph $G - Y$ in Fig. 8-38*(c)*. (Note Y is a cutpoint since $G - Y$ is disconnected.)

(e) Vertices A, X, and Y are cut points.

(f) An edge e is a bridge if $G - e$ is disconnected. Thus there are three bridges: $\{A, Z\}$, $\{A, X\}$, and $\{C, Y\}$.

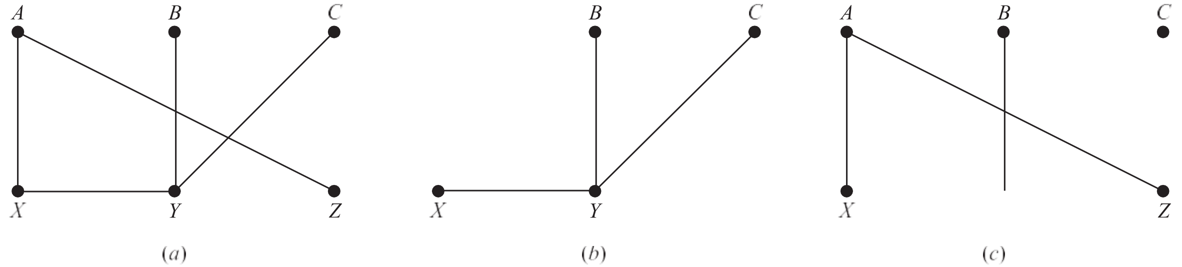


Fig. 8-38

8.5. Consider the graph *G* in Fig. 8-36*(b)*. Find the subgraphs obtained when each vertex is deleted. Does *G* have any cut points?

When we delete a vertex from *G*, we also have to delete all edges which contain the vertex. The six graphs obtained by deleting each of the vertices of *G* are shown in Fig. 8-39. All six graphs are connected; hence no vertex is a cut point.

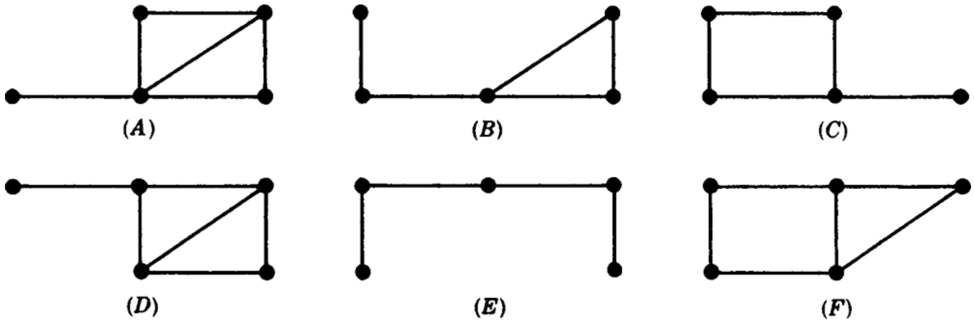


Fig. 8-39

8.6. Show that the six graphs obtained in Problem 8.5 are distinct, that is, no two of them are isomorphic. Also show that *(B)* and *(C)* are homeomorphic.

The degrees of the five vertices of any graph cannot be paired off with the degrees of any other graph, except for *(B)* and *(C)*. Hence none of the graphs is isomorphic except possibly *(B)* and *(C)*.

However if we delete the vertex of degree 3 in *(B)* and *(C)*, we obtain distinct subgraphs. Thus *(B)* and *(C)* are also nonisomorphic; hence all six graphs are distinct. However, *(B)* and *(C)* are homeomorphic since they can be obtained from isomorphic graphs by adding appropriate vertices.

TRAVERSABLE GRAPHS, EULER AND HAMILTONIAN CIRCUITS

8.7. Consider each graph in Fig. 8-40. Which of them are traversable, that is, have Euler paths? Which are Eulerian, that is, have an Euler circuit? For those that do not, explain why.

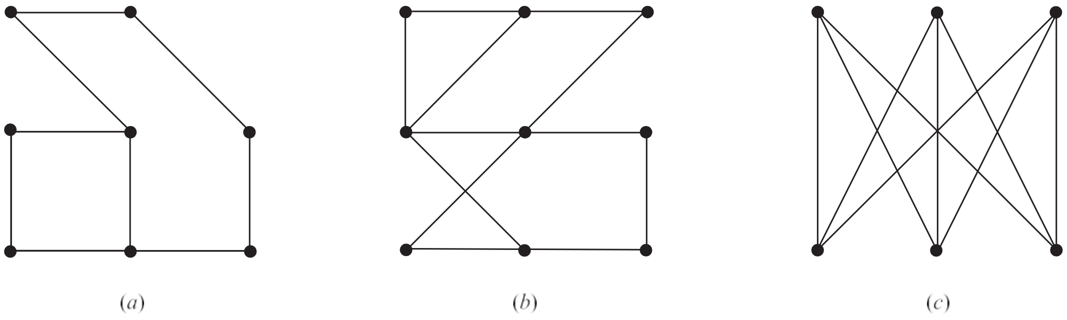


Fig. 8-40

G is traversable (has an Euler path) if only 0 or 2 vertices have odd degree, and *G* is Eulerian (has an Euler circuit) if all vertices are of even degree (Theorem 8.3).

(a) Traversable, since there are two odd vertices. The traversable path must begin at one of the odd vertices and will end at the other.

(b) Traversable, since all vertices are even. Thus G has an Euler circuit.

(c) Since six vertices have odd degrees, G is not traversable.

8.8. Which of the graphs in Fig. 8-40 have a Hamiltonian circuit? If not, why not?

Graphs *(a)* and *(c)* have Hamiltonian circuits. (The reader should be able to easily find one of them.) However, graph

(b) has no Hamiltonian circuit. For if α is a Hamiltonian circuit, then α must connect the middle vertex with the lower right vertex, then proceed along the bottom row to the lower right vertex, then vertically to the middle right, but then is forced back to the central vertex before visiting the remaining vertices.

8.9. Prove Theorem 8.3 (Euler): A finite connected graph G is Eulerian if and only if each vertex has even degree.

Suppose G is Eulerian and T is a closed Eulerian trail. For any vertex v of G, the trail T enters and leaves v the same number of times without repeating any edge. Hence v has even degree.

Suppose conversely that each vertex of G has even degree. We construct an Eulerian trail. We begin a trail T_1 at any edge e. We extend T_1 by adding one edge after the other. If T_1 is not closed at any step, say, T_1 begins at u but ends at v / = u, then only an odd number of the edges incident on v appear in T_1; hence we can extend T_1 by another edge incident on v. Thus we can continue to extend T_1 until T_1 returns to its initial vertex u, i.e., until T_1 is closed. If T_1 includes all the edges of G, then T_1 is our Eulerian trail.

Suppose T_1 does not include all edges of G. Consider the graph H obtained by deleting all edges of T_1 from G. H may not be connected, but each vertex of H has even degree since T_1 contains an even number of the edges incident on any vertex. Since G is connected, there is an edge e^J of H which has an endpoint u^J in T_1. We construct a trail T_2 in H beginning at u^J and using e^J. Since all vertices in H have even degree, we can continue to extend T_2 in H until T_2 returns to u^J as pictured in Fig. 8-41. We can clearly put T_1 and T_2 together to form a larger closed trail in G. We continue this process until all the edges of G are used. We finally obtain an Eulerian trail, and so G is Eulerian.

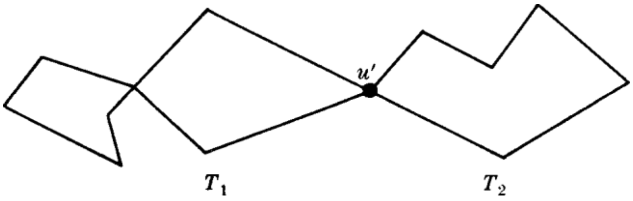


Fig. 8-41

TREES, SPANNING TREES

8.10. Draw all trees with exactly six vertices.

There are six such trees which are exhibited in Fig. 8-42. The first tree has diameter 5, the next two diameter 4, the next two diameter 3, and the last one diameter 2. Any other tree with 6 nodes is isormorphic to one of these trees.

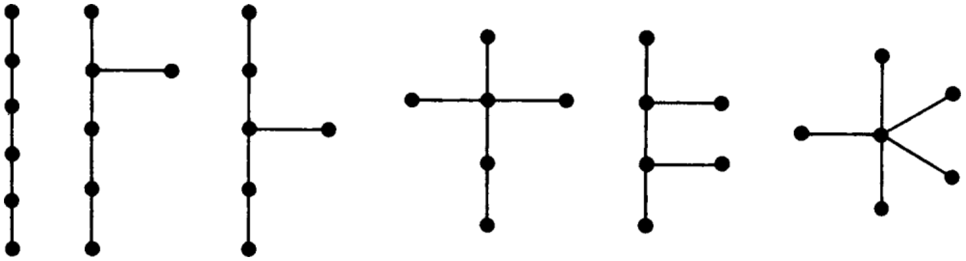
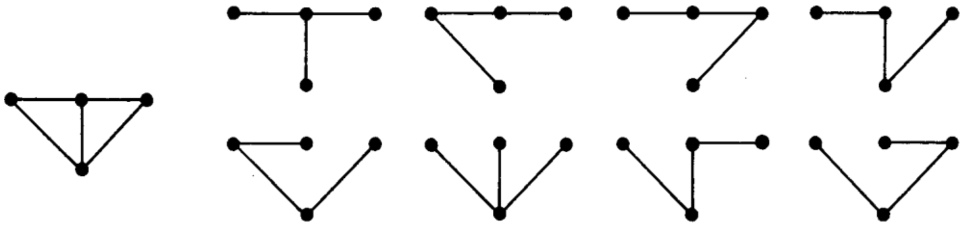


Fig. 8-42

8.11. Find all spanning trees of the graph G shown in Fig. 8-43*(a)*.

There are eight such spanning trees as shown in Fig. 8-43*(b)*. Each spanning tree must have $4 - 1 = 3$ edges since G has four vertices. Thus each tree can be obtained by deleting two of the five edges of G. This can be done in 10 ways, except that two of the ways lead to disconnected graphs. Hence the above eight spanning trees are all the spanning trees of G.

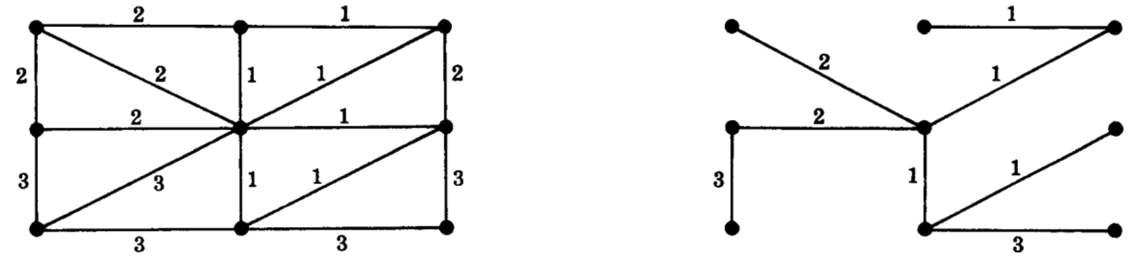


(*a*) Graph *G*

(*b*) Spanning trees

Fig. 8-43

8.12. Find a minimal spanning tree T for the weighted graph G in Fig. 8-44*(a)*.



(*a*)

(*b*)

Fig. 8-44

Since G has $n = 9$ vertices, T must have $n - 1 = 8$ edges. Apply Algorithm 8.2, that is, keep deleting edges with maximum length and without disconnecting the graph until only $n-1 = 8$ edges remain. Alternatively, apply Algorithm 8.3, that is, beginning with the nine vertices, keep adding edges with minimum length and without forming any circle until $n - 1 = 8$ edges are added. Both methods give a minimum spanning tree such as that shown in Fig. 8-44*(b)*.

8.13. Let G be a graph with more than one vertex. Prove the following are equivalent.

(i) G is a tree.

(ii) Each pair of vertices is connected by exactly one simple path.

(iii) G is connected; but $G - e$ is disconnected for any edge e of G.

(iv) G is cycle-free, but if any edge is added to G then the resulting graph has exactly one cycle.

(i) *implies* (ii) Let u and v be two vertices in G. Since G is a tree, G is connected so there is at least one path between u and v. By Problem 8.37 there can only be one simple path between u and v, otherwise G will contain a cycle.

(ii) *implies* (iii) Suppose we delete an edge $e = \{u, v\}$ from G. Note e is a path from u to v. Suppose the resulting graph $G - e$ has a path P from u to v. Then P and e are two distinct paths from u to v, which contradicts the hypothesis. Thus there is no path between u and v in $G - e$, so $G - e$ is disconnected.

(iii) *implies* (iv) Suppose G contains a cycle C which contains an edge e =u,{v . }By hypothesis, G is connected but G^J G e=is disconnected, with u and v belonging to different components of G^J (Problem 8.41) This contradicts the fact that u and v are connected by the path P C= e which lies in G^J. Hence G is cycle-free. Now let x and y be vertices of G and let H be the graph obtained by adjoining the edge e x, y =to {G. Si}nce G is connected, there is a path P from x to y in G; hence C Pe forms a=cycle in H . Suppose H contains another cycle C^J. Since G is cycle-free, C^J must contain the edge e, say C^J $P^J e$. Then P and=P^J are two simple paths in G from x to y. (See Fig. 8-45.) By Problem 8.37, G contains a cycle, which contradicts the fact that G is cycle-free. Hence H contains only one cycle.

(iv) *implies* (i) Since adding any edge $e = \{x, y\}$to G produces a cycle, the vertices x and y must already be connected in G. Hence G is connected and by hypothesis G is cycle-free; that is, G is a tree.

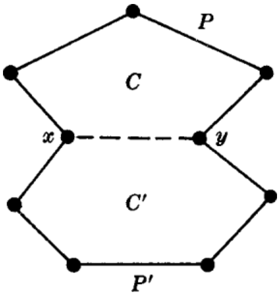


Fig. 8-45

8.14. Prove Theorem 8.6: Let G be a finite graph with $n \geq 1$ vertices. Then the following are equivalent. (i) G is a tree, (ii) G is a cycle-free and has $n - 1$ edges, (iii) G is connected and has $n - 1$ edges.

The proof is by induction on n. The theorem is certainly true for the graph with only one vertex and hence no edges. That is, the theorem holds for $n = 1$. We now assume that $n > 1$ and that the theorem holds for graphs with less than n vertices.

(i) *implies* (ii) Suppose G is a tree. Then G is cycle-free, so we only need to show that G has $n - 1$ edges. By Problem 8.38, G has a vertex of degree 1. Deleting this vertex and its edge, we obtain a tree T which has $n - 1$ vertices. The theorem holds for T , so T has $n - 2$ edges. Hence G has $n - 1$ edges.

(ii) *implies* (iii) Suppose G is cycle-free and has $n - 1$ edges. We only need show that G is connected. Suppose G is disconnected and has k components, $T_1, \dots, T_k$, which are trees since each is connected and cycle-free. Say T_i has n_i vertices. Note $n_i < n$. Hence the theorem holds for T_i, so T_i has $n_i - 1$ edges. Thus

$$n = n_1 + n_2 + \cdots + n_k$$

and

$$n - 1 = (n_1 - 1) + (n_2 - 1) + \cdots + (n_k - 1) = n_1 + n_2 + \cdots + n_k - k = n - k$$

Hence $k = 1$. But this contradicts the assumption that G is disconnected and has $k > 1$ components. Hence G is connected.

(iii) *implies* (i) Suppose G is connected and has $n - 1$ edges. We only need to show that G is cycle-free. Suppose G has a cycle containing an edge e. Deleting e we obtain the graph $H = G - e$ which is also connected. But H has n vertices and $n - 2$ edges, and this contradicts Problem 8.39. Thus G is cycle-free and hence is a tree.

PLANAR GRAPHS

8.15. Draw a planar representation, if possible, of the graphs (*a*), (*b*), and (*c*) in Fig. 8-46.

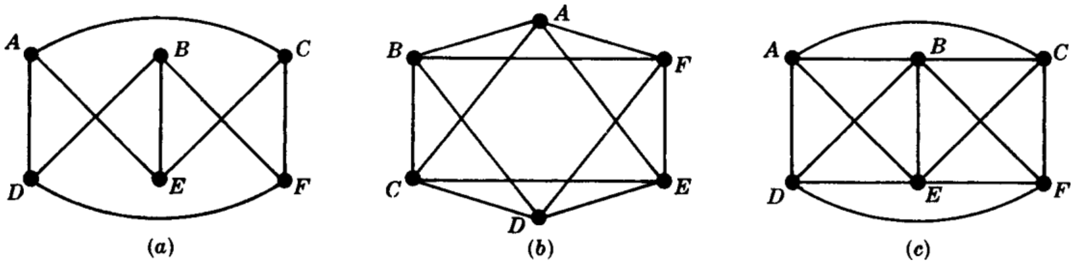


Fig. 8-46

(a) Redrawing the positions of B and E, we get a planar representation of the graph as in Fig. 8-47*(a)*.

(b) This is not the star graph K_5. This has a planar representation as in Fig. 8-47*(b)*.

(c) This graph is non-planar. The utility graph $K_{3,3}$ is a subgraph as shown in Fig. 8-47*(c)* where we have redrawn the positions of C and F.

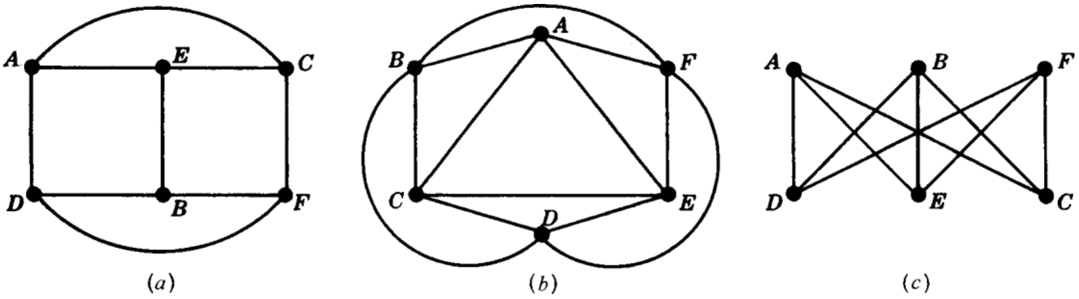


Fig. 8-47

8.16. Count the number V of vertices, the number E of edges, and the number R of regions of each map in Fig. 8-48; and verify Euler's formula. Also find the degree d of the outside region.

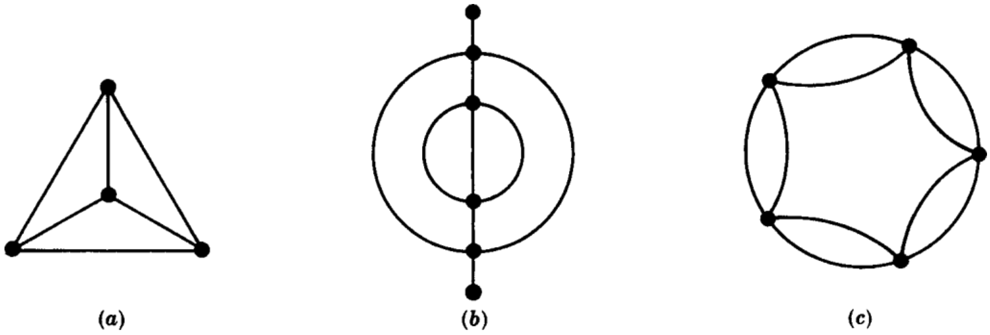


Fig. 8-48

(*a*) $V = 4, E = 6, R = 4$. Hence $V - E + R = 4 - 6 + 4 = 2$. Also $d = 3$.

(*b*) $V = 6, E = 9, R = 5$; so $V - E + R = 6 - 9 + 5 = 2$. Here $d = 6$ since two edges are counted twice. (*c*) $V = 5, E = 10, R = 7$. Hence $V - E + R = 5 - 10 + 7 = 2$. Here $d = 5$.

8.17. Find the minimum number n of colors required to paint each map in Fig. 8-48.

(*a*) $n = 4$; (*b*) $n = 3$; (*c*) $n = 2$.

8.18. Prove Theorem 8.8 (Euler): $V - E + R = 2$.

Suppose the connected map M consists of a single vertex P as in Fig. 8-49*(a)*. Then $V = 1$, $E = 0$, and $R = 1$.

Hence $V - E + R = 2$. Otherwise M can be built up from a single vertex by the following two constructions:

(1) Add a new vertex Q_2 and connect it to an existing vertex Q_1 by an edge which does not cross any existing edge as in Fig. 8-49*(b)*.

(2) Connect two existing vertices Q_1 and Q_2 by an edge e which does not cross any existing edge as in Fig. 8-49*(c)*.

Neither operation changes the value of $V - E + R$. Hence M has the same value of $V - E + R$ as the map consisting of a single vertex, that is, $V - E + R = 2$. Thus the theorem is proved.

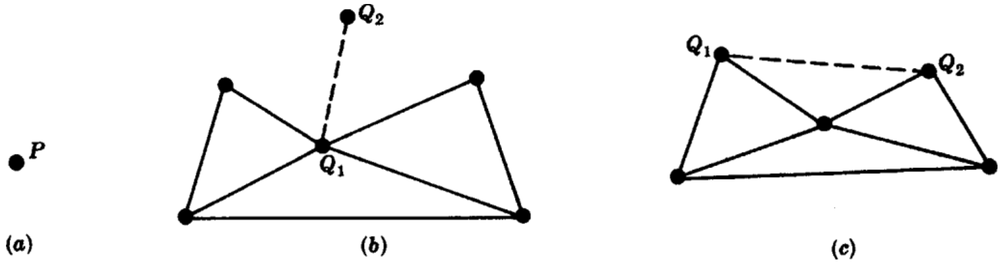


Fig. 8-49

8.19. Prove Theorem 8.11: The following are equivalent for a graph G: (i) G is 2-colorable. (ii) G is bipartite. (iii) Every cycle of G has even length.

(i) *implies* (ii). Suppose G is 2-colorable. Let M be the set of vertices painted the first color, and let N be the set of vertices painted the second color. Then M and N form a bipartite partition of the vertices of G since neither the vertices of M nor the vertices of N can be adjacent to each other since they are of the same color.

(ii) *implies* (iii). Suppose G is bipartite and M and N form a bipartite partition of the vertices of G. If a cycle begins at a vertex u of, say, M, then it will go to a vertex of N, and then to a vertex of M, and then to N and so on. Hence when the cycle returns to u it must be of even length. That is, every cycle of G will have even length.

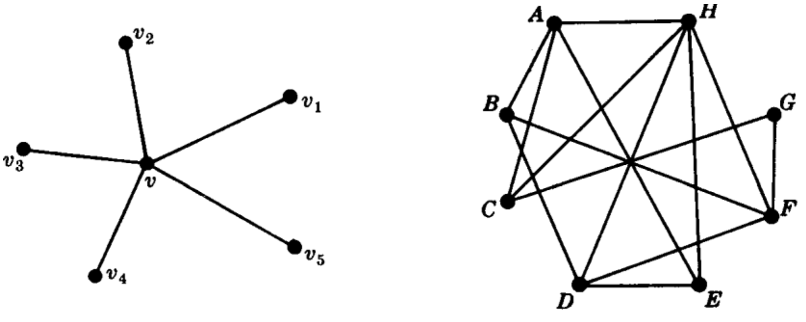
(iii) *implies* (i). Lastly, suppose every cycle of G has even length. We pick a vertex in each connected component and paint it the first color, say red. We then successively paint all the vertices as follows: If a vertex is painted red, then any vertex adjacent to it will be painted the second color, say blue. If a vertex is painted blue, then any vertex adjacent to it will be painted red. Since every cycle has even length, no adjacent vertices will be painted the same color. Hence G is 2-colorable, and the theorem is proved.

8.20. Prove Theorem 8.12: A planar graph G is5-colorable.

The proof is by induction on the number p of vertices of G. If $p \leq 5$, then the theorem obviously holds. Suppose $p > 5$, and the theorem holds for graphs with less than p vertices. By the preceding problem, G has a vertex v such that

$\leq$ $-$

deg*(v)* 5. By induction, the subgraph G v is 5-colorable. Assume one such coloring. If the vertices adjacent to v use less than the five colors, than we simply paint v with one of the remaining colors and obtain a 5-coloring of G. We are still left with the case that v is adjacent to five vertices which are painted different colors. Say the vertices, moving counterclockwise about v, are $v_1,\dots,$ v_5 and are painted respectively by the colors $c_1,\dots,$ c_5. (See Fig. 8-50*(a)*.)



(*a*) (*b*)

Fig. 8-50

Consider now the subgraph H of G generated by the vertices painted c_1 and c_3. Note H includes v_1 and v_3. If v_1 and v_3 belong to different components of H, then we can interchange the colors c_1 and c_3 in the component containing v_1 without destroying the coloring of $G - v$. Then v_1 and v_3 are painted by c_3, c_1 can be chosen to paint v, and we have a 5-coloring of G. On the other hand, suppose v_1 and v_3 are in the same component of H . Then there is a path P from v_1 to v_3 whose vertices are painted either c_1 or c_3. The path P together with the edges $\{v, v_1\}$ and $\{v, v_3\}$form a cycle C which encloses either v_2 or v_4. Consider now the subgraph K generated by the vertices painted c_3 or c_4. Since C encloses v_2 or v_4, but not both, the vertices v_2 and v_4 belong to different components of K. Thus we can interchange the colors c_2 and c_4 in the component containing v_2 without destroying the coloring of $G - v$. Then v_2 and v_4 are painted by c_4, and we can choose c_2 to paint v and obtain a 5-coloring of G. Thus G is 5-colorable and the theorem is proved.

8.21. Use the Welch-Powell Algorithm 8.4 (Fig. 8-24) to paint the graph in Fig. 8-50*(b)*.

First order the vertices according to decreasing degrees to obtain the sequence

$$H, \quad A, \quad D, \quad F, \quad B, \quad C, \quad E, \quad G$$

Proceeding sequentially, we use the first color to paint the vertices H , B, and then G. (We cannot paint A, D, or F the first color since each is connected to H , and we cannot paint C or E the first color since each is connected to either H or B.) Proceeding sequentially with the unpainted vertices, we use the second color to paint the vertices A and D. The remaining vertices F , C, and E can be painted with the third color. Thus the chromatic number n cannot be greater than 3. However, in any coloring, H, D, and E must be painted different colors since they are connected to each other. Hence $n = 3$.

8.22. Let G be a finite connected planar graph with at least three vertices. Show that G has at least one vertex of degree 5 or less.

Let p be the number of vertices and q the number of edges of G, and suppose deg*(u)* ≥ 6 for each vertex u of G.

But $2q$ equals the sum of the degrees of the vertices of G (Theorem 8.1); so $2q \geq 6p$. Therefore

$$q \geq 3p > 3p - 6$$

This contradicts Theorem 8.9. Thus some vertex of *G* has degree 5 or less.

9.1 ROOTED TREES

Recall that a tree graph is a connected cycle-free graph, that is, a connected graph without any cycles. A *rooted tree T* is a tree graph with a designated vertex *r* called the *root* of the tree. Since there is a unique simple path from the root *r* to any other vertex *v* in *T* , this determines a direction to the edges of *T* . Thus *T* may be viewed as a directed graph. We note that any tree may be made into a rooted tree by simply selecting one of the vertices as the root.

Consider a rooted tree *T* with root *r*. The length of the path from the root *r* to any vertex *v* is called the *level* (or *depth*) of *v*, and the maximum vertex level is called the *depth* of the tree. Those vertices with degree 1, other than the root *r*, are called the *leaves* of *T*, and a directed path from a vertex to a leaf is called a *branch*.

One usually draws a picture of a rooted tree *T* with the root at the top of the tree. Figure 9-2(*a*) shows a rooted tree *T* with root *r* and 10 other vertices. The tree has five leaves, *d,f, h, i*, and *j* . Observe that: *level(a)*=1, *level(f)* 2, *level(j)* 3. Furthermore, the depth of the tree is 3.

The fact that a rooted tree *T* gives a direction to the edges means that we can give a precedence relationship between the vertices. Specifically, we will say that a vertex *u precedes* a vertex *v* or that *v follows u* if there is a (directed) path from *v* to *u*. In particular, we say that *v immediately follows u* if *(u, v)* is an edge, that is, if *v* follows *u* and *v* is adjacent to *u*. We note that every vertex *v*, other than the root, immediately follows a unique vertex, but that *v* can be immediately followed by more than one vertex. For example, in Fig. 9-2(*a*), the vertex *j* follows c but immediately follows *g*. Also, both *i* and *j* immediately follow *g*.

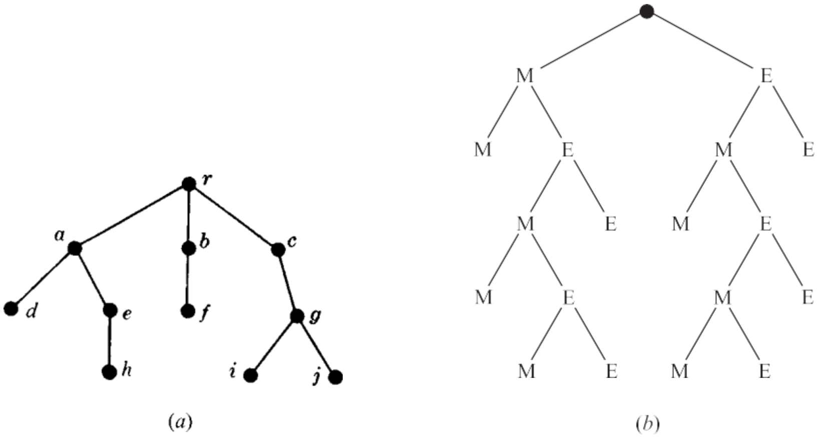


Fig. 9-2

A rooted tree *T* is also a useful device to enumerate all the logical possibilities of a sequence of events where each event can occur in a finite number of ways. This is illustrated in the following example.

EXAMPLE 9.5 Suppose Marc and Erik are playing a tennis tournament such that 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 proceed.

The rooted tree in Fig. 9-2(*b*) shows the various ways that the tournament could proceed. There are 10 leaves which correspond to the 10 ways that the tournament can occur:

MM, MEMM, MEMEM, MEMEE, MEE, EMM, EMEMM, EMEME, EMEE, EE

Specifically, the path from the root to the leaf describes who won which games in the particular tournament.

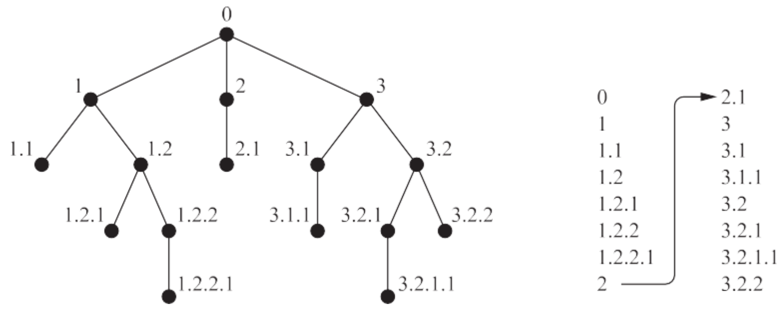
Ordered Rooted Trees

Consider a rooted tree T in which the edges leaving each vertex are ordered. Then we have the concept of an *ordered rooted tree*. One can systematically label (or *address*) the vertices of such a tree as follows: We first assign 0 to the root r. We next assign $1, 2, 3, \ldots$ to the vertices immediately following r according as the edges were ordered. We then label the remaining vertices in the following way. If a is the label of a vertex v, then $a.1, a.2, \ldots$ are assigned to the vertices immediately following v according as the edges were ordered. We illustrate this address system in Fig. 9-3(*a*), where edges are pictured from left to right according to their order. Observe that the number of decimal points in any label is one less than the level of the vertex. We will refer to this labeling system as the *universal address system* for an ordered rooted tree.

The universal address system gives us an important way of linearly describing (or storing) an ordered rooted tree. Specifically, given addresses a and $\neq b$, we let $a < b$ if b $a.c$, (that is, a is an *initial segment* of b), or if there exist positive integers m and n with $m < n$ such that

$$a = \text{r.m.s} \quad \text{and} \quad b = \text{r.n.t}$$

This order is called the *lexicographic order* since it is similar to the way words are arranged in a dictionary. For example, the addresses in Fig. 9-3(*a*) are linearly ordered as pictured in Fig. 9-3(*b*). This lexicographic order is



(*a*) (*b*)

Fig. 9-3

identical to the order obtained by moving down the leftmost branch of the tree, then the next branch to the right, then the second branch to the right, and so on.