

16

Groups, Coding Theory, and Polya's Method of Enumeration

In our study of algebraic structures we examine properties shared by particular mathematical systems. Then we generalize our findings in order to study the underlying structure common to these particular examples.

In Chapter 14 we did this with the ring structure, which depended on two closed binary operations. Now we turn to a structure involving one closed binary operation. This structure is called a *group*.

Our study of groups will examine many ideas comparable to those for rings. However, here we shall dwell primarily on those aspects of the structure that are needed for applications in cryptology, coding theory, and a counting method developed by George Polya.

16.1

Definition, Examples, and Elementary Properties

Definition 16.1

If G is a nonempty set and \circ is a binary operation on G , then (G, \circ) is called a *group* if the following conditions are satisfied.

- 1) For all $a, b \in G$, $a \circ b \in G$. (Closure of G under \circ)
- 2) For all $a, b, c \in G$, $a \circ (b \circ c) = (a \circ b) \circ c$. (The Associative Property)
- 3) There exists $e \in G$ with $a \circ e = e \circ a = a$, for all $a \in G$. (The Existence of an Identity)
- 4) For each $a \in G$ there is an element $b \in G$ such that $a \circ b = b \circ a = e$. (Existence of Inverses)

Furthermore, if $a \circ b = b \circ a$ for all $a, b \in G$, then G is called a *commutative*, or *abelian*, group. The adjective *abelian* honors the Norwegian mathematician Niels Henrik Abel (1802–1829).

We realize that the first condition in Definition 16.1 could have been omitted if we simply required the binary operation for G to be a *closed* binary operation.

Following Definition 14.1 (for a ring) we mentioned how the associative laws for the closed binary operations of $+$ (ring addition) and \cdot (ring multiplication) could be extended by mathematical induction. The same type of situation arises for groups. If (G, \circ) is any group, and $r, n \in \mathbf{Z}^+$ with $n \geq 3$ and $1 \leq r < n$, then

$$(a_1 \circ a_2 \circ \cdots \circ a_r) \circ (a_{r+1} \circ \cdots \circ a_n) = a_1 \circ a_2 \circ \cdots \circ a_r \circ a_{r+1} \circ \cdots \circ a_n,$$

where $a_1, a_2, \dots, a_r, a_{r+1}, \dots, a_n$ are all elements from G .

EXAMPLE 16.1

Under ordinary addition, each of $\mathbf{Z}, \mathbf{Q}, \mathbf{R}, \mathbf{C}$ is an abelian group. None of these is a group under multiplication because 0 has no multiplicative inverse. However, $\mathbf{Q}^*, \mathbf{R}^*$, and \mathbf{C}^* (the nonzero elements of \mathbf{Q}, \mathbf{R} , and \mathbf{C} , respectively) are abelian groups under ordinary multiplication.

If $(R, +, \cdot)$ is a ring, then $(R, +)$ is an abelian group; the nonzero elements of a *field* $(F, +, \cdot)$ form the abelian group (F^*, \cdot) .

EXAMPLE 16.2

For $n \in \mathbf{Z}^+, n > 1$, we find that $(\mathbf{Z}_n, +)$ is an abelian group. When p is a prime, (\mathbf{Z}_p^*, \cdot) is an abelian group. Tables 16.1 and 16.2 demonstrate this for $n = 6$ and $p = 7$, respectively. (Recall that in \mathbf{Z}_n we often write a for $[a] = \{a + kn \mid k \in \mathbf{Z}\}$. The same notation is used in \mathbf{Z}_p^* .)

Table 16.1

+	0	1	2	3	4	5
0	0	1	2	3	4	5
1	1	2	3	4	5	0
2	2	3	4	5	0	1
3	3	4	5	0	1	2
4	4	5	0	1	2	3
5	5	0	1	2	3	4

Table 16.2

\cdot	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

Definition 16.2

For every group G the number of elements in G is called the *order* of G and this is denoted by $|G|$. When the number of elements in a group is not finite we say that G has infinite order.

EXAMPLE 16.3

For all $n \in \mathbf{Z}^+, |(\mathbf{Z}_n, +)| = n$, while $|(\mathbf{Z}_p^*, \cdot)| = p - 1$ for each prime p .

EXAMPLE 16.4

Let us start with the ring $(\mathbf{Z}_9, +, \cdot)$ and consider the subset $U_9 = \{a \in \mathbf{Z}_9 \mid a \text{ is a unit in } \mathbf{Z}_9\} = \{a \in \mathbf{Z}_9 \mid a^{-1} \text{ exists}\} = \{1, 2, 4, 5, 7, 8\} = \{a \in \mathbf{Z}^+ \mid 1 \leq a \leq 8 \text{ and } \gcd(a, 9) = 1\}$. The results in Table 16.3 show us that U_9 is closed under the multiplication for the ring $(\mathbf{Z}_9, +, \cdot)$ —namely, multiplication modulo 9. Furthermore, we also see that 1 is the identity element and that each element has an inverse (in U_9). For instance, 5 is the inverse for 2, and 7 is the inverse for 4. Finally, since every ring is associative under the operation

of (ring) multiplication, it follows that $a \cdot (b \cdot c) = (a \cdot b) \cdot c$ for all $a, b, c \in U_9$. Consequently, (U_9, \cdot) is a group of order 6—in fact, it is an abelian group of order 6.

Table 16.3

\cdot	1	2	4	5	7	8
1	1	2	4	5	7	8
2	2	4	8	1	5	7
4	4	8	7	2	1	5
5	5	1	2	7	8	4
7	7	5	1	8	4	2
8	8	7	5	4	2	1

In general, for each $n \in \mathbb{Z}^+$, where $n > 1$, if $U_n = \{a \in (\mathbb{Z}_n, +, \cdot) \mid a \text{ is a unit}\} = \{a \in \mathbb{Z}^+ \mid 1 \leq a \leq n-1 \text{ and } \gcd(a, n) = 1\}$, then (U_n, \cdot) is an abelian group under the (closed) binary operation of multiplication modulo n . The group (U_n, \cdot) is called the *group of units* for the ring $(\mathbb{Z}_n, +, \cdot)$ and it has order $\phi(n)$, where ϕ denotes the Euler phi function of Section 8.1.

From here on the group operation will be written multiplicatively, unless it is given otherwise. So $a \circ b$ now becomes ab .

The following theorem provides several properties shared by all groups.

THEOREM 16.1

For every group G ,

- a) the identity of G is unique.
- b) the inverse of each element of G is unique.
- c) if $a, b, c \in G$ and $ab = ac$, then $b = c$. (Left-cancellation property)
- d) if $a, b, c \in G$ and $ba = ca$, then $b = c$. (Right-cancellation property)

Proof:

- a) If e_1, e_2 are both identities in G , then $e_1 = e_1 e_2 = e_2$. (Justify each equality.)
- b) Let $a \in G$ and suppose that b, c are both inverses of a . Then $b = be = b(ac) = (ba)c = ec = c$. (Justify each equality.)

The proofs of properties (c) and (d) are left for the reader. (It is because of these properties that we find each group element appearing exactly once in each row and each column of the table for a finite group.)

On the basis of the result in Theorem 16.1(b) the unique inverse of a will be designated by a^{-1} . When the group is written additively, $-a$ is used to denote the (additive) inverse of a .

As in the case of multiplication in a ring, we have powers of elements in a group. We define $a^0 = e$, $a^1 = a$, $a^2 = a \cdot a$, and in general $a^{n+1} = a^n \cdot a$, for all $n \in \mathbb{N}$. Since each group element has an inverse, for $n \in \mathbb{Z}^+$, we define $a^{-n} = (a^{-1})^n$. Then a^n is defined for all $n \in \mathbb{Z}$, and it can be shown that for all $m, n \in \mathbb{Z}$, $a^m \cdot a^n = a^{m+n}$ and $(a^m)^n = a^{mn}$.

If the group operation is addition, then multiples replace powers and for all $m, n \in \mathbf{Z}$, and all $a \in G$, we find that

$$ma + na = (m + n)a \quad m(na) = (mn)a.$$

In this case the identity is written as 0, rather than e . And here, for all $a \in G$, we have $0a = 0$, where the "0" in front of a is the integer 0 (in \mathbf{Z}) while the "0" on the right side of the equation is the identity 0 (in G). [So these two "0"'s are different.]

For an abelian group G we also find that for all $n \in \mathbf{Z}$ and all $a, b \in G$, (1) $(ab)^n = a^n b^n$, when G is written multiplicatively; and (2) $n(a + b) = na + nb$, when the additive notation is used for G .

We now take a look at a special subset of a group.

EXAMPLE 16.5

Let $G = (\mathbf{Z}_6, +)$. If $H = \{0, 2, 4\}$, then H is a nonempty subset of G . Table 16.4 shows that $(H, +)$ is also a group under the binary operation of G .

Table 16.4

+	0	2	4
0	0	2	4
2	2	4	0
4	4	0	2

This situation motivates the following definition.

Definition 16.3

Let G be a group and $\emptyset \neq H \subseteq G$. If H is a group under the binary operation of G , then we call H a *subgroup* of G .

EXAMPLE 16.6

- Every group G has $\{e\}$ and G as subgroups. These are the *trivial* subgroups of G . All others are termed *nontrivial*, or *proper*.
- In addition to $H = \{0, 2, 4\}$, the subset $K = \{0, 3\}$ is also a (proper) subgroup of $G = (\mathbf{Z}_6, +)$.
- Each of the nonempty subsets $\{1, 8\}$ and $\{1, 4, 7\}$ is a subgroup of (U_9, \cdot) .
- The group $(\mathbf{Z}, +)$ is a subgroup of $(\mathbf{Q}, +)$, which is a subgroup of $(\mathbf{R}, +)$. Yet \mathbf{Z}^* under multiplication is not a subgroup of (\mathbf{Q}^*, \cdot) . (Why not?)

For a group G and $\emptyset \neq H \subseteq G$, the following tells us when H is a subgroup of G .

THEOREM 16.2

If H is a nonempty subset of a group G , then H is a subgroup of G if and only if (a) for all $a, b \in H$, $ab \in H$, and (b) for all $a \in H$, $a^{-1} \in H$.

Proof: If H is a subgroup of G , then by Definition 16.3 H is a group under the same binary operation. Hence it satisfies all the group conditions, including the two mentioned here. Conversely, let $\emptyset \neq H \subseteq G$ with H satisfying conditions (a) and (b). For all $a, b, c \in H$, $(ab)c = a(bc)$ in G , so $(ab)c = a(bc)$ in H . (We say that H "inherits" the associative

property from G .) Finally, as $H \neq \emptyset$, let $a \in H$. By condition (b), $a^{-1} \in H$ and by condition (a), $aa^{-1} = e \in H$, so H contains the identity element and is a group.

A finiteness condition modifies the situation.

THEOREM 16.3

If G is a group and $\emptyset \neq H \subseteq G$, with H finite, then H is a subgroup of G if and only if H is closed under the binary operation of G .

Proof: As in the proof of Theorem 16.2, if H is a subgroup of G , then H is closed under the binary operation of G . Conversely, let H be a finite nonempty subset of G that is so closed. If $a \in H$, then $aH = \{ah | h \in H\} \subseteq H$ because of the closure condition. By left-cancellation in G , $ah_1 = ah_2 \Rightarrow h_1 = h_2$, so $|aH| = |H|$. With $aH \subseteq H$ and $|aH| = |H|$, it follows from H being finite that $aH = H$. As $a \in H$, there exists $b \in H$ with $ab = a$. But (in G) $ab = a = ae$, so $b = e$ and H contains the identity. Since $e \in H = aH$, there is an element $c \in H$ such that $ac = e$. Then $(ca)^2 = (ca)(ca) = (c(ac))a = (ce)a = ca = (ca)e$, so $ca = e$, and $c = a^{-1} \in H$. Consequently, by Theorem 16.2, H is a subgroup of G .

The finiteness condition in Theorem 16.3 is crucial. Both \mathbf{Z}^+ and \mathbf{N} are nonempty closed subsets of the group $(\mathbf{Z}, +)$, yet neither has the additive inverses needed for the group structure.

The next example provides a nonabelian group.

EXAMPLE 16.7

Consider the first equilateral triangle shown in Fig. 16.1(a). When we rotate this triangle counterclockwise (within its plane) through 120° about an axis perpendicular to its plane and passing through its center C , we obtain the second triangle shown in Fig. 16.1(a). As a result, the vertex originally labeled 1 in Fig. 16.1(a) is now in the position that was originally labeled 3. Likewise, 2 is now in the position originally occupied by 1, and 3 has moved to where 2 was. This can be described by the function $\pi_1: \{1, 2, 3\} \rightarrow \{1, 2, 3\}$, where $\pi_1(1) = 3$, $\pi_1(2) = 1$, $\pi_1(3) = 2$. A more compact notation, $\begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix}$, where we write $\pi_1(i)$ below i for each $1 \leq i \leq 3$, emphasizes that π_1 is a permutation of $\{1, 2, 3\}$. If π_2 denotes the counterclockwise rotation through 240° , then $\pi_2 = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}$. For the identity π_0 —that is, the rotation through $n(360^\circ)$ for $n \in \mathbf{Z}$ —we write $\pi_0 = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{pmatrix}$. These rotations are called *rigid motions* of the triangle. They are two-dimensional motions that keep the center C fixed and preserve the shape of the triangle. Hence the triangle looks the same as when we started, except for a possible rearrangement of the labels on some of its vertices.

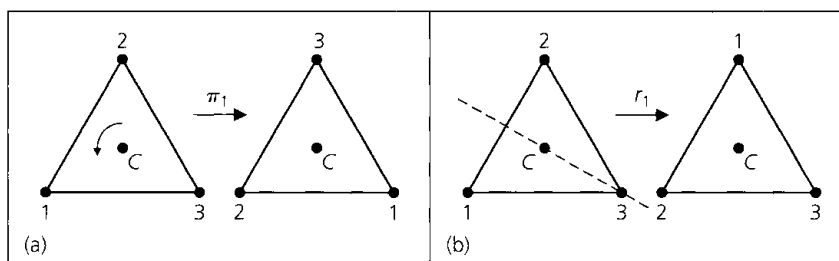


Figure 16.1

In addition to these rotations, the triangle can be reflected along an axis passing through a vertex and the midpoint of the opposite side. For the diagonal axis that bisects the base angle on the right, the reflection gives the result in Fig. 16.1(b). This we represent by $r_1 = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix}$. A similar reflection about the axis bisecting the left base angle yields the permutation $r_2 = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix}$. When the triangle is reflected about its vertical axis, we have $r_3 = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix}$. Each r_i , for $1 \leq i \leq 3$, is a three-dimensional rigid motion.

Let $G = \{\pi_0, \pi_1, \pi_2, r_1, r_2, r_3\}$, the set of rigid motions (in space) of the equilateral triangle. We make G into a group by defining the rigid motion $\alpha\beta$, for $\alpha, \beta \in G$, as that motion obtained by applying first α and then following up with β . Hence, for example, $\pi_1 r_1 = r_3$. We can see this geometrically, but it will be handy to consider the permutations as follows: $\pi_1 r_1 = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix}$, where, for example, $\pi_1(1) = 3$ and $r_1(3) = 3$ and we write $1 \xrightarrow{\pi_1} 3 \xrightarrow{r_1} 3$. So $1 \xrightarrow{\pi_1 r_1} 3$ in the product $\pi_1 r_1$. (Note that the order in which we write the product $\pi_1 r_1$ here is the opposite of the order for their composite function as defined in Section 5.6. The notation of Section 5.6 occurs in analysis, whereas in algebra there is a tendency to employ this opposite order.) Also, since $2 \xrightarrow{\pi_1} 1 \xrightarrow{r_1} 2$ and $3 \xrightarrow{\pi_1} 2 \xrightarrow{r_1} 1$, it follows that $\pi_1 r_1 = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix} = r_3$.

Table 16.5 verifies that under this binary operation G is closed, with identity π_0 . Also $\pi_1^{-1} = \pi_2$, $\pi_2^{-1} = \pi_1$, and every other element is its own inverse. Since the elements of G are actually functions, the associative property follows from Theorem 5.6 (although in reverse order).

Table 16.5

\cdot	π_0	π_1	π_2	r_1	r_2	r_3
π_0	π_0	π_1	π_2	r_1	r_2	r_3
π_1	π_1	π_2	π_0	r_3	r_1	r_2
π_2	π_2	π_0	π_1	r_2	r_3	r_1
r_1	r_1	r_2	r_3	π_0	π_1	π_2
r_2	r_2	r_3	r_1	π_2	π_0	π_1
r_3	r_3	r_1	r_2	π_1	π_2	π_0

We computed $\pi_1 r_1$ as r_3 , but from Table 16.5 we see that $r_1 \pi_1 = r_2$. With $\pi_1 r_1 = r_3 \neq r_2 = r_1 \pi_1$, it follows that G is nonabelian.

This group can also be obtained as the group of all permutations of the set $\{1, 2, 3\}$ under the binary operation of function composition. It is denoted by S_3 (the *symmetric* group on three symbols).

EXAMPLE 16.8

The symmetric group S_4 consists of the 24 permutations of $\{1, 2, 3, 4\}$. Here $\pi_0 = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 2 & 3 & 4 \end{pmatrix}$ is the identity. If $\alpha = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 1 & 4 & 3 \end{pmatrix}$, $\beta = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 1 & 2 & 4 \end{pmatrix}$, then $\alpha\beta = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 3 & 4 & 2 \end{pmatrix}$ but $\beta\alpha = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 2 & 1 & 3 \end{pmatrix}$, so S_4 is nonabelian. Also, $\beta^{-1} = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 1 & 4 \end{pmatrix}$ and $\alpha^2 = \pi_0 = \beta^3$. Within S_4 there is a subgroup of order 8 that represents the group of rigid motions for a square.

We turn now to a construction for making larger groups out of smaller ones.

THEOREM 16.4

Let (G, \circ) and $(H, *)$ be groups. Define the binary operation \cdot on $G \times H$ by $(g_1, h_1) \cdot (g_2, h_2) = (g_1 \circ g_2, h_1 * h_2)$. Then $(G \times H, \cdot)$ is a group and is called the *direct product* of G and H .

Proof: The verification of the group properties for $(G \times H, \cdot)$ is left to the reader.

EXAMPLE 16.9

Consider the groups $(\mathbf{Z}_2, +)$, $(\mathbf{Z}_3, +)$. On $G = \mathbf{Z}_2 \times \mathbf{Z}_3$, define $(a_1, b_1) \cdot (a_2, b_2) = (a_1 + a_2, b_1 + b_2)$. Then G is a group of order 6 where the identity is $(0, 0)$, and the inverse, for example, of the element $(1, 2)$ is $(1, 1)$.

EXERCISES 16.1

1. For each of the following sets, determine whether or not the set is a group under the stated binary operation. If so, determine its identity and the inverse of each of its elements. If it is not a group, state the condition(s) of the definition that it violates.

- $\{-1, 1\}$ under multiplication
- $\{-1, 1\}$ under addition
- $\{-1, 0, 1\}$ under addition
- $\{10n | n \in \mathbf{Z}\}$ under addition
- The set of all one-to-one functions $g: A \rightarrow A$, where $A = \{1, 2, 3, 4\}$, under function composition
- $\{a/2^n | a, n \in \mathbf{Z}, n \geq 0\}$ under addition

2. Prove parts (c) and (d) of Theorem 16.1.

3. Why is the set \mathbf{Z} not a group under subtraction?

4. Let $G = \{q \in \mathbf{Q} | q \neq -1\}$. Define the binary operation \circ on G by $x \circ y = x + y + xy$. Prove that (G, \circ) is an abelian group.

5. Define the binary operation \circ on \mathbf{Z} by $x \circ y = x + y + 1$. Verify that (\mathbf{Z}, \circ) is an abelian group.

6. Let $S = \mathbf{R}^* \times \mathbf{R}$. Define the binary operation \circ on S by $(u, v) \circ (x, y) = (ux, vx + y)$. Prove that (S, \circ) is a non-abelian group.

7. Find the elements in the groups U_{20} and U_{24} — the groups of units for the rings $(\mathbf{Z}_{20}, +, \cdot)$ and $(\mathbf{Z}_{24}, +, \cdot)$, respectively.

8. For any group G prove that G is abelian if and only if $(ab)^2 = a^2b^2$ for all $a, b \in G$.

9. If G is a group, prove that for all $a, b \in G$,

$$\text{a) } (a^{-1})^{-1} = a \qquad \text{b) } (ab)^{-1} = b^{-1}a^{-1}$$

10. Prove that a group G is abelian if and only if for all $a, b \in G$, $(ab)^{-1} = a^{-1}b^{-1}$.

11. Find all subgroups in each of the following groups.

$$\text{a) } (\mathbf{Z}_{12}, +) \qquad \text{b) } (\mathbf{Z}_{11}^*, \cdot) \qquad \text{c) } S_3$$

12. a) How many rigid motions (in two or three dimensions) are there for a square?

b) Make a group table for these rigid motions like the one in Table 16.5 for the equilateral triangle. What is the identity for this group? Describe the inverse of each element geometrically.

13. a) How many rigid motions (in two or three dimensions) are there for a regular pentagon? Describe them geometrically.

b) Answer part (a) for a regular n -gon, $n \geq 3$.

14. In the group S_5 , let

$$\alpha = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 3 & 1 & 4 & 5 \end{pmatrix} \quad \text{and} \quad \beta = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 1 & 5 & 3 & 4 \end{pmatrix}.$$

Determine $\alpha\beta$, $\beta\alpha$, α^3 , β^4 , α^{-1} , β^{-1} , $(\alpha\beta)^{-1}$, $(\beta\alpha)^{-1}$, and $\beta^{-1}\alpha^{-1}$.

15. If G is a group, let $H = \{a \in G | ag = ga \text{ for all } g \in G\}$. Prove that H is a subgroup of G . (The subgroup H is called the *center* of G .)

16. Let ω be the complex number $(1/\sqrt{2})(1 + i)$.

a) Show that $\omega^8 = 1$ but $\omega^n \neq 1$ for $n \in \mathbf{Z}^+$, $1 \leq n \leq 7$.

b) Verify that $\{\omega^n | n \in \mathbf{Z}^+, 1 \leq n \leq 8\}$ is an abelian group under multiplication.

17. a) Prove Theorem 16.4.

b) Extending the idea developed in Theorem 16.4 and Example 16.9 to the group $\mathbf{Z}_6 \times \mathbf{Z}_6 \times \mathbf{Z}_6 = \mathbf{Z}_6^3$, answer the following.

i) What is the order of this group?

ii) Find a subgroup of \mathbf{Z}_6^3 of order 6, one of order 12, and one of order 36.

iii) Determine the inverse of each of the elements $(2, 3, 4)$, $(4, 0, 2)$, $(5, 1, 2)$.

18. a) If H, K are subgroups of a group G , prove that $H \cap K$ is also a subgroup of G .

b) Give an example of a group G with subgroups H, K such that $H \cup K$ is not a subgroup of G .

19. a) Find all x in (\mathbf{Z}_5^*, \cdot) such that $x = x^{-1}$.

b) Find all x in $(\mathbf{Z}_{11}^*, \cdot)$ such that $x = x^{-1}$.

c) Let p be a prime. Find all x in (\mathbf{Z}_p^*, \cdot) such that $x = x^{-1}$.

d) Prove that $(p-1)! \equiv -1 \pmod{p}$, for p a prime. [This result is known as Wilson's Theorem, although it was only conjectured by John Wilson (1741–1793). The first proof was given in 1770 by Joseph Louis Lagrange (1736–1813).]

20. a) Find x in (U_8, \cdot) where $x \neq 1$, $x \neq 7$ but $x = x^{-1}$.
 b) Find x in (U_{16}, \cdot) where $x \neq 1$, $x \neq 15$ but $x = x^{-1}$.
 c) Let $k \in \mathbf{Z}^+$, $k \geq 3$. Find x in (U_{2^k}, \cdot) where $x \neq 1$, $x \neq 2^k - 1$ but $x = x^{-1}$.

16.2

Homomorphisms, Isomorphisms, and Cyclic Groups

We turn our attention once again to functions that preserve structure.

EXAMPLE 16.10

Let $G = (\mathbf{Z}, +)$ and $H = (\mathbf{Z}_4, +)$. Define $f: G \rightarrow H$ by

$$f(x) = [x] = \{x + 4k \mid k \in \mathbf{Z}\}.$$

For all $x, y \in G$,

$$\begin{array}{ccc} f(x+y) = [x+y] = [x] + [y] = f(x) + f(y), \\ \uparrow & & \uparrow \\ \text{The operation in } G & & \text{The operation in } H \end{array}$$

where the second equality follows from the way the addition of equivalence classes was developed in Section 14.3. Consequently, here f preserves the group operations and is an example of a special type of function that we shall now define.

Definition 16.4

If (G, \circ) and $(H, *)$ are groups and $f: G \rightarrow H$, then f is called a *group homomorphism* if for all $a, b \in G$, $f(a \circ b) = f(a) * f(b)$.

When we know that the given structures are groups, the function f is simply called a homomorphism.

Some properties of homomorphisms are given in the following theorem.

THEOREM 16.5

Let (G, \circ) , $(H, *)$ be groups with respective identities e_G, e_H . If $f: G \rightarrow H$ is a homomorphism, then

- a) $f(e_G) = e_H$.
 b) $f(a^{-1}) = [f(a)]^{-1}$ for all $a \in G$.
 c) $f(a^n) = [f(a)]^n$ for all $a \in G$ and all $n \in \mathbf{Z}$.
 d) $f(S)$ is a subgroup of H for each subgroup S of G .

Proof:

a) $e_H * f(e_G) = f(e_G) = f(e_G \circ e_G) = f(e_G) * f(e_G)$, so by right-cancellation [Theorem 16.1(d)], it follows that $f(e_G) = e_H$.

b) & c) The proofs of these parts are left for the reader.

- d) If S is a subgroup of G , then $S \neq \emptyset$, so $f(S) \neq \emptyset$. Let $x, y \in f(S)$. Then $x = f(a)$, $y = f(b)$, for some $a, b \in S$. Since S is a subgroup of G , it follows that $a \circ b \in S$, so $x * y = f(a) * f(b) = f(a \circ b) \in f(S)$. Finally, $x^{-1} = [f(a)]^{-1} = f(a^{-1}) \in f(S)$ because $a^{-1} \in S$ when $a \in S$. Consequently, by Theorem 16.2, $f(S)$ is a subgroup of H .

Definition 16.5

If $f: (G, \circ) \rightarrow (H, *)$ is a homomorphism, we call f an *isomorphism* if it is one-to-one and onto. In this case G, H are said to be *isomorphic groups*.

EXAMPLE 16.11

Let $f: (\mathbf{R}^+, \cdot) \rightarrow (\mathbf{R}, +)$ where $f(x) = \log_{10}(x)$. This function is both one-to-one and onto. (Verify these properties.) For all $a, b \in \mathbf{R}^+$, $f(ab) = \log_{10}(ab) = \log_{10} a + \log_{10} b = f(a) + f(b)$. Therefore, f is an isomorphism and the group of positive real numbers under multiplication is abstractly the same as the group of all real numbers under addition. Here the function f translates a problem in the multiplication of real numbers (a somewhat difficult problem without a calculator) into a problem dealing with the addition of real numbers (an easier arithmetic consideration). This was a major reason behind the use of logarithms before the advent of calculators.

EXAMPLE 16.12

Let G be the group of complex numbers $\{1, -1, i, -i\}$ under multiplication. Table 16.6 shows the multiplication table for this group. With $H = (\mathbf{Z}_4, +)$, consider $f: G \rightarrow H$ defined by

$$f(1) = [0] \quad f(-1) = [2] \quad f(i) = [1] \quad f(-i) = [3].$$

Then $f((i)(-i)) = f(1) = [0] = [1] + [3] = f(i) + f(-i)$, and $f((-1)(-i)) = f(i) = [1] = [2] + [3] = f(-1) + f(-i)$.

Although we have not checked all possible cases, the function is an isomorphism. Note that the image under f of the subgroup $\{1, -1\}$ of G is $\{[0], [2]\}$, a subgroup of H .

Table 16.6

\cdot	1	-1	i	$-i$
1	1	-1	i	$-i$
-1	-1	1	$-i$	i
i	i	$-i$	-1	1
$-i$	$-i$	i	1	-1

Let us take a closer look at this group G . Here $i^1 = i$, $i^2 = -1$, $i^3 = -i$, and $i^4 = 1$, so every element of G is a power of i , and we say that i *generates* G . This is denoted by $G = \langle i \rangle$. (It is also true that $G = \langle -i \rangle$. Verify this.)

The last part of the preceding example leads us to the following definition.

Definition 16.6

A group G is called *cyclic* if there is an element $x \in G$ such that for each $a \in G$, $a = x^n$ for some $n \in \mathbf{Z}$.

EXAMPLE 16.13

- a) The group $H = (\mathbf{Z}_4, +)$ is cyclic. Here the operation is addition, so we have multiples instead of powers. We find that both $[1]$ and $[3]$ generate H . For the case of $[3]$, we have $1 \cdot [3] = [3]$, $2 \cdot [3] (= [3] + [3]) = [2]$, $3 \cdot [3] = [1]$, and $4 \cdot [3] = [0]$. Hence $H = \langle [3] \rangle = \langle [1] \rangle$.
- b) Consider the multiplicative group $U_9 = \{1, 2, 4, 5, 7, 8\}$ that we examined in Example 16.4. Here we find that $2^1 = 2$, $2^2 = 4$, $2^3 = 8$, $2^4 = 7$, $2^5 = 5$, $2^6 = 1$, so U_9 is a cyclic group of order 6 and $U_9 = \langle 2 \rangle$. It is also true that $U_9 = \langle 5 \rangle$ because $5^1 = 5$, $5^2 = 7$, $5^3 = 8$, $5^4 = 4$, $5^5 = 2$, $5^6 = 1$.

The concept of a cyclic group leads to a related idea. Given a group G , if $a \in G$ consider the set $S = \{a^k | k \in \mathbf{Z}\}$. From Theorem 16.2 it follows that S is a subgroup of G . This subgroup is called the *subgroup generated by a* and is designated by $\langle a \rangle$. In Example 16.12 $\langle i \rangle = \langle -i \rangle = G$; also, $\langle -1 \rangle = \{-1, 1\}$ and $\langle 1 \rangle = \{1\}$. For part (a) of Example 16.13 we consider multiples instead of powers and find that $H = \langle [1] \rangle = \langle [3] \rangle$, $\langle [2] \rangle = \{[0], [2]\}$, and $\langle [0] \rangle = \{[0]\}$. When we examine the group U_9 in part (b) of that example we see that $U_9 = \langle 2 \rangle$ (or $\langle [2] \rangle$) $= \langle 5 \rangle$, $\langle 4 \rangle = \{1, 4, 7\} = \langle 7 \rangle$, $\langle 8 \rangle = \{1, 8\}$, and $\langle 1 \rangle = \{1\}$.

Definition 16.7

If G is a group and $a \in G$, the *order of a* , denoted $\text{ord}(a)$, is $|\langle a \rangle|$. (If $|\langle a \rangle|$ is infinite, we say that a has infinite order.)

In Example 16.12, $\text{ord}(1) = 1$, $\text{ord}(-1) = 2$, whereas both i and $-i$ have order 4.

Let us take a second look at the idea of order for the case where $|\langle a \rangle|$ is finite. When $|\langle a \rangle| = 1$ then $a = e$ because $a = a^1 \in \langle a \rangle$ and $e = a^0 \in \langle a \rangle$. If $|\langle a \rangle|$ is finite but $a \neq e$, then $\langle a \rangle = \{a^m | m \in \mathbf{Z}\}$ is finite, so $\{a, a^2, a^3, \dots\} = \{a^m | m \in \mathbf{Z}^+\}$ is also finite. Consequently, there exist $s, t \in \mathbf{Z}^+$, where $1 \leq s < t$ and $a^s = a^t$ —from which it follows that $a^{t-s} = e$, with $t-s \in \mathbf{Z}^+$. Since $e \in \{a^m | m \in \mathbf{Z}^+\}$, let n be the smallest positive integer such that $a^n = e$. We claim that $\langle a \rangle = \{a, a^2, a^3, \dots, a^{n-1}, a^n (= e)\}$.

First we observe that $\{a, a^2, a^3, \dots, a^{n-1}, a^n (= e)\} = \langle a \rangle$. Otherwise, we have $a^u = a^v$ for positive integers u, v where $1 \leq u < v \leq n$, and then $a^{v-u} = e$ with $0 < v-u < n$. This, however, contradicts the minimality of n . So now we know that $|\langle a \rangle| \geq n$. But for each $k \in \mathbf{Z}$, it follows from the division algorithm that $k = qn + r$, where $0 \leq r < n$, and so $a^k = a^{qn+r} = (a^n)^q (a^r) = (e^q)(a^r) = a^r \in \{a, a^2, a^3, \dots, a^{n-1}, a^n (= e = a^0)\}$. Therefore, $\langle a \rangle = \{a, a^2, a^3, \dots, a^{n-1}, a^n (= e)\}$ and we can also define $\text{ord}(a)$ as the *smallest positive integer n for which $a^n = e$* . This alternative definition for the order of a group element (of finite order) proves to be of value in the following theorem.

THEOREM 16.6

Let $a \in G$ with $\text{ord}(a) = n$. If $k \in \mathbf{Z}$ and $a^k = e$, then $n|k$.

Proof: By the division algorithm (again), we have $k = qn + r$, for $0 \leq r < n$, and so it follows that $e = a^k = a^{qn+r} = (a^n)^q (a^r) = (e^q)(a^r) = a^r$. If $0 < r < n$, we contradict the definition of n as $\text{ord}(a)$. Hence $r = 0$ and $k = qn$.

We now examine some further results on cyclic groups. The next example helps us to motivate part (b) of Theorem 16.7.

EXAMPLE 16.14

It is known from part (b) of Example 16.13 that $U_9 = \{1, 2, 4, 5, 7, 8\} = \langle 2 \rangle$. We use this fact to define the function $f: U_9 \rightarrow (\mathbf{Z}_6, +)$ as follows:

$$\begin{aligned} f(1) &= [0] & f(2) &= [1] & f(4) &= [2] \\ f(5) &= f(2^5) = [5] & f(7) &= f(2^4) = [4] & f(8) &= f(2^3) = [3]. \end{aligned}$$

So, in general, for each $a \in U_9$ we write $a = 2^k$, for some $0 \leq k \leq 5$, and have $f(a) = f(2^k) = [k]$. This function f is one-to-one and onto and we find, for example, that $f(2 \cdot 5) = f(1) = [0] = [1] + [5] = f(2) + f(5)$, and $f(7 \cdot 8) = f(2) = [1] = [4] + [3] = f(7) + f(8)$.

In general, for a, b in U_9 we may write $a = 2^m$ and $b = 2^n$, where $0 \leq m \leq 5$ and $0 \leq n \leq 5$. It then follows that

$$f(a \cdot b) = f(2^m \cdot 2^n) = f(2^{m+n}) = [m+n] = [m] + [n] = f(a) + f(b).$$

Consequently, the function f is an isomorphism and the groups U_9 and $(\mathbf{Z}_6, +)$ are isomorphic.

[Note how the function f links the generators of the two cyclic groups. Also note that the function $g: U_9 \rightarrow (\mathbf{Z}_6, +)$ where

$$\begin{aligned} g(1) &= [0] & g(5) &= [1] & g(7) &= g(5^2) = [2] \\ g(8) &= g(5^3) = [3] & g(4) &= g(5^4) = [4] & g(2) &= g(5^5) = [5] \end{aligned}$$

is another isomorphism between these two cyclic groups.]

THEOREM 16.7

Let G be a cyclic group.

- a) If $|G|$ is infinite, then G is isomorphic to $(\mathbf{Z}, +)$.
- b) If $|G| = n$, where $n > 1$, then G is isomorphic to $(\mathbf{Z}_n, +)$.

Proof:

- a) For $G = \langle a \rangle = \{a^k | k \in \mathbf{Z}\}$, let $f: G \rightarrow \mathbf{Z}$ be defined by $f(a^k) = k$. (Could we have $a^k = a^t$ with $k \neq t$? If so, f would not be a function.) For $a^m, a^n \in G$, $f(a^m \cdot a^n) = f(a^{m+n}) = m+n = f(a^m) + f(a^n)$, so f is a homomorphism. We leave to the reader the verification that f is one-to-one and onto.
- b) If $G = \langle a \rangle = \{a, a^2, \dots, a^{n-1}, a^n = e\}$, then the function $f: G \rightarrow \mathbf{Z}_n$ defined by $f(a^k) = [k]$ is an isomorphism. (Verify this.)

EXAMPLE 16.15

If $G = \langle g \rangle$, G is abelian because $g^m \cdot g^n = g^{m+n} = g^{n+m} = g^n \cdot g^m$ for all $m, n \in \mathbf{Z}$. The converse, however, is false. The group H of Table 16.7 is abelian, and $\mathfrak{e}(e) = 1$, $\mathfrak{e}(a) = \mathfrak{e}(b) = \mathfrak{e}(c) = 2$. Since no element of H has order 4, H cannot be cyclic. (The group H is the smallest noncyclic group and is known as the *Klein Four* group.)

Table 16.7

\cdot	e	a	b	c
e	e	a	b	c
a	a	e	c	b
b	b	c	e	a
c	c	b	a	e

Our last result concerns the structure of subgroups in a cyclic group.

THEOREM 16.8

Every subgroup of a cyclic group is cyclic.

Proof: Let $G = \langle a \rangle$. If H is a subgroup of G , each element of H has the form a^k , for some $k \in \mathbb{Z}$. For $H \neq \{e\}$, let t be the smallest positive integer such that $a^t \in H$. (How do we know such an integer t exists?) We claim that $H = \langle a^t \rangle$. Since $a^t \in H$, by the closure property for the subgroup H , $\langle a^t \rangle \subseteq H$. For the opposite inclusion, let $b \in H$, with $b = a^s$, for some $s \in \mathbb{Z}$. By the division algorithm, $s = qt + r$, where $q, r \in \mathbb{Z}$ and $0 \leq r < t$. Consequently, $a^s = a^{qt+r}$ and so $a^r = a^{-qt}a^s = (a^t)^{-q}b$. H is a subgroup of G , so $a^t \in H \Rightarrow (a^t)^{-q} \in H$. Then with $(a^t)^{-q}, b \in H$, it follows that $a^r = (a^t)^{-q}b \in H$. But if $a^r \in H$ with $r > 0$, then we contradict the minimality of t . Hence $r = 0$ and $b = a^{qt} = (a^t)^q \in \langle a^t \rangle$, so $H = \langle a^t \rangle$, a cyclic group.

EXERCISES 16.2

- Prove parts (b) and (c) of Theorem 16.5.
- Let $A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$.
 - Determine A^2, A^3 , and A^4 .
 - Verify that $\{A, A^2, A^3, A^4\}$ is an abelian group under ordinary matrix multiplication.
 - Prove that the group in part (b) is isomorphic to the group shown in Table 16.6.
- If $G = (\mathbb{Z}_6, +)$, $H = (\mathbb{Z}_3, +)$, and $K = (\mathbb{Z}_2, +)$, find an isomorphism for the groups $H \times K$ and G .
- Let $f: G \rightarrow H$ be a group homomorphism onto H . If G is abelian, prove that H is abelian.
- Let $(\mathbb{Z} \times \mathbb{Z}, \oplus)$ be the abelian group where $(a, b) \oplus (c, d) = (a + c, b + d)$ —here $a + c$ and $b + d$ are computed using ordinary addition in \mathbb{Z} —and let $(G, +)$ be an additive group. If $f: \mathbb{Z} \times \mathbb{Z} \rightarrow G$ is a group homomorphism where $f(1, 3) = g_1$ and $f(3, 7) = g_2$, express $f(4, 6)$ in terms of g_1 and g_2 .
- Let $f: (\mathbb{Z} \times \mathbb{Z}, \oplus) \rightarrow (\mathbb{Z}, +)$ be the function defined by $f(x, y) = x - y$. [Here $(\mathbb{Z} \times \mathbb{Z}, \oplus)$ is the same group as in Exercise 5, and $(\mathbb{Z}, +)$ is the group of integers under ordinary addition.]
 - Prove that f is a homomorphism onto \mathbb{Z} .
 - Determine all $(a, b) \in \mathbb{Z} \times \mathbb{Z}$ with $f(a, b) = 0$.
 - Find $f^{-1}(7)$.
 - If $E = \{2n | n \in \mathbb{Z}\}$, what is $f^{-1}(E)$?
- Find the order of each element in the group of rigid motions of (a) the equilateral triangle; and (b) the square.
- In S_5 find an element of order n , for all $2 \leq n \leq 5$. Also determine the (cyclic) subgroup of S_5 that each of these elements generates.
- Find all the elements of order 10 in $(\mathbb{Z}_{40}, +)$.
 - Let $G = \langle a \rangle$ be a cyclic group of order 40. Which elements of G have order 10?
- Determine U_{14} , the group of units for the ring $(\mathbb{Z}_{14}, +, \cdot)$.
 - Show that U_{14} is cyclic and find all of its generators.
- Verify that (\mathbb{Z}_p^*, \cdot) is cyclic for the primes 5, 7, and 11.
- For a group G , prove that the function $f: G \rightarrow G$ defined by $f(a) = a^{-1}$ is an isomorphism if and only if G is abelian.
- If $f: G \rightarrow H, g: H \rightarrow K$ are homomorphisms, prove that the composite function $g \circ f: G \rightarrow K$, where $(g \circ f)(x) = g(f(x))$, is a homomorphism.
- For $\omega = (1/\sqrt{2})(1 + i)$, let G be the multiplicative group $\{\omega^n | n \in \mathbb{Z}^+, 1 \leq n \leq 8\}$.
 - Show that G is cyclic and find each element $x \in G$ such that $\langle x \rangle = G$.
 - Prove that G is isomorphic to the group $(\mathbb{Z}_8, +)$.
- Find all generators of the cyclic groups $(\mathbb{Z}_{12}, +)$, $(\mathbb{Z}_{16}, +)$, and $(\mathbb{Z}_{24}, +)$.
 - Let $G = \langle a \rangle$ with $c(a) = n$. Prove that $a^k, k \in \mathbb{Z}^+$, generates G if and only if k and n are relatively prime.
 - If G is a cyclic group of order n , how many distinct generators does it have?
- Let $f: G \rightarrow H$ be a group homomorphism. If $a \in G$ with $c(a) = n$, and $c(f(a)) = k$ (in H), prove that $k | n$.

16.3

Cosets and Lagrange's Theorem

In the last two sections, for all finite groups G and subgroups H of G , we had $|H|$ dividing $|G|$. In this section we'll see that this was not mere chance but is true in general. To prove this we need one new idea.

Definition 16.8

If H is a subgroup of G , then for each $a \in G$, the set $aH = \{ah | h \in H\}$ is called a *left coset* of H in G . The set $Ha = \{ha | h \in H\}$ is a *right coset* of H in G .

If the operation in G is addition, we write $a + H$ in place of aH , where $a + H = \{a + h | h \in H\}$.

When the term *coset* is used in this chapter, it will refer to a left coset. For abelian groups there is no need to distinguish between left and right cosets. However, at the end of the next example we'll see that this is not the case for nonabelian groups.

EXAMPLE 16.16

If G is the group of Example 16.7 and $H = \{\pi_0, \pi_1, \pi_2\}$, the coset $r_1H = \{r_1\pi_0, r_1\pi_1, r_1\pi_2\} = \{r_1, r_2, r_3\}$. Likewise we have $r_2H = r_3H = \{r_1, r_2, r_3\}$, whereas $\pi_0H = \pi_1H = \pi_2H = H$.

We see that $|\alpha H| = |H|$ for each $\alpha \in G$ and that $G = H \cup r_1H$ is a partition of G .

For the subgroup $K = \{\pi_0, r_1\}$, we find $r_2K = \{r_2, \pi_2\}$ and $r_3K = \{r_3, \pi_1\}$. Again a partition of G arises: $G = K \cup r_2K \cup r_3K$. (Note: $Kr_2 = \{\pi_0r_2, r_1r_2\} = \{r_2, \pi_1\} \neq r_2K$.)

EXAMPLE 16.17

For $G = (\mathbb{Z}_{12}, +)$ and $H = \{[0], [4], [8]\}$, we find that

$$[0] + H = \{[0], [4], [8]\} = [4] + H = [8] + H = H$$

$$[1] + H = \{[1], [5], [9]\} = [5] + H = [9] + H$$

$$[2] + H = \{[2], [6], [10]\} = [6] + H = [10] + H$$

$$[3] + H = \{[3], [7], [11]\} = [7] + H = [11] + H,$$

and $H \cup ([1] + H) \cup ([2] + H) \cup ([3] + H)$ is a partition of G .

These examples now prepare us for the following results.

LEMMA 16.1

If H is a subgroup of the finite group G , then for all $a, b \in G$, (a) $|aH| = |H|$; and (b) either $aH = bH$ or $aH \cap bH = \emptyset$.

Proof:

- a) Since $aH = \{ah | h \in H\}$, it follows that $|aH| \leq |H|$. If $|aH| < |H|$, we have $ah_i = ah_j$ with h_i, h_j distinct elements of H . By left-cancellation in G we then get the contradiction $h_i = h_j$, so $|aH| = |H|$.
- b) If $aH \cap bH \neq \emptyset$, let $c = ah_1 = bh_2$, for some $h_1, h_2 \in H$. If $x \in aH$, then $x = ah$ for some $h \in H$, and so $x = (bh_2h_1^{-1})h = b(h_2h_1^{-1}h) \in bH$, and $aH \subseteq bH$. Similarly, $y \in bH \Rightarrow y = bh_3$, for some $h_3 \in H \Rightarrow y = (ah_1h_2^{-1})h_3 = a(h_1h_2^{-1}h_3) \in aH$, so $bH \subseteq aH$. Therefore aH and bH are either disjoint or identical.

We observe that if $g \in G$, then $g \in gH$ because $e \in H$. Also, by part (b) of Lemma 16.1, G can be partitioned into mutually disjoint cosets.

At this point we are ready to prove the main result of this section.

THEOREM 16.9

Lagrange's Theorem. If G is a finite group of order n with H a subgroup of order m , then m divides n .

Proof: If $H = G$ the result follows. Otherwise $m < n$ and there exists an element $a \in G - H$. Since $a \notin H$, it follows that $aH \neq H$, so $aH \cap H = \emptyset$. If $G = aH \cup H$, then $|G| = |aH| + |H| = 2|H|$ and the theorem follows. If not, there is an element $b \in G - (H \cup aH)$, with $bH \cap H = \emptyset = bH \cap aH$ and $|bH| = |H|$. If $G = bH \cup aH \cup H$, we have $|G| = 3|H|$. Otherwise we're back to an element $c \in G$ with $c \notin bH \cup aH \cup H$. The group G is finite, so this process terminates and we find that $G = a_1H \cup a_2H \cup \cdots \cup a_kH$. Therefore, $|G| = k|H|$ and m divides n .

An alternative method for proving this theorem is given in Exercise 12 for this section.

We close with the statements of two corollaries. Their proofs are requested in the Section Exercises.

COROLLARY 16.1

If G is a finite group and $a \in G$, then $\phi(a)$ divides $|G|$.

COROLLARY 16.2

Every group of prime order is cyclic.

EXERCISES 16.3

1. Let $G = S_4$. (a) For $\alpha = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 1 \end{pmatrix}$, find the subgroup $H = \langle \alpha \rangle$. (b) Determine the left cosets of H in G .

2. Answer Exercise 1 for the case where α is replaced by $\beta = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 1 & 4 \end{pmatrix}$.

3. If $\gamma = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 1 & 4 & 3 \end{pmatrix} \in S_4$, how many cosets does $\langle \gamma \rangle$ determine?

4. For $G = (\mathbb{Z}_{24}, +)$, find the cosets determined by the subgroup $H = \langle [3] \rangle$. Do likewise for the subgroup $K = \langle [4] \rangle$.

5. Let G be a group with subgroups H and K . If $|G| = 660$, $|K| = 66$, and $K \subset H \subset G$, what are the possible values for $|H|$?

6. Let R be a ring with unity u . Prove that the units of R form a group under the multiplication of the ring.

7. Let $G = S_4$, the symmetric group on four symbols, and let H be the subset of G where

$$H = \left\{ \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 2 & 3 & 4 \end{pmatrix}, \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 1 & 4 & 3 \end{pmatrix}, \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 4 & 1 & 2 \end{pmatrix}, \begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 3 & 2 & 1 \end{pmatrix} \right\}.$$

a) Construct a table to show that H is an abelian subgroup of G .

b) How many left cosets of H are there in G ?

c) Consider the group $(\mathbb{Z}_2 \times \mathbb{Z}_2, \oplus)$ where $(a, b) \oplus (c, d) = (a + c, b + d)$ — and the sums $a + c, b + d$ are computed using addition modulo 2. Prove that H is isomorphic to this group.

8. If G is a group of order n and $a \in G$, prove that $a^n = e$.

9. Let p be a prime. (a) If G has order $2p$, prove that every proper subgroup of G is cyclic. (b) If G has order p^2 , prove that G has a subgroup of order p .

10. Prove Corollaries 16.1 and 16.2.

11. Let H and K be subgroups of a group G , where e is the identity of G .

a) Prove that if $|H| = 10$ and $|K| = 21$, then $H \cap K = \{e\}$.

b) If $|H| = m$ and $|K| = n$, with $\gcd(m, n) = 1$, prove that $H \cap K = \{e\}$.

12. The following provides an alternative way to establish Lagrange's Theorem. Let G be a group of order n , and let H be a subgroup of G of order m .

a) Define the relation \mathcal{R} on G as follows: If $a, b \in G$, then $a \mathcal{R} b$ if $a^{-1}b \in H$. Prove that \mathcal{R} is an equivalence relation on G .

b) For $a, b \in G$, prove that $a \mathcal{R} b$ if and only if $aH = bH$.

- c) If $a \in G$, prove that $[a]$, the equivalence class of a under \mathcal{R} , satisfies $[a] = aH$.
- d) For each $a \in G$, prove that $|aH| = |H|$.
- e) Now establish the conclusion of Lagrange's Theorem, namely that $|H|$ divides $|G|$.
13. a) *Fermat's Theorem.* If p is a prime, prove that $a^p \equiv a \pmod{p}$ for each $a \in \mathbf{Z}$. [How is this related to Exercise 22(a) of Section 14.3?]
- b) *Euler's Theorem.* For each $n \in \mathbf{Z}^+$, $n > 1$, and each $a \in \mathbf{Z}$, prove that if $\gcd(a, n) = 1$, then $a^{\phi(n)} \equiv 1 \pmod{n}$.
- c) How are the theorems in parts (a) and (b) related?
- d) Is there any connection between these two theorems and the results in Exercises 6 and 8?

16.4

The RSA Cryptosystem (Optional)

This section provides us with an opportunity to use some of the theoretical ideas we encountered in Sections 14.3 and 16.3 in a more contemporary application.

In Example 14.15 of Section 14.3 we introduced two private-key cryptosystems: the cipher shift and the affine cipher. For an alphabet of m characters, the encryption function $E: \mathbf{Z}_m \rightarrow \mathbf{Z}_m$, for the cipher-shift cryptosystem, is given by $E(\theta) = (\theta + \kappa) \bmod m$, where $\theta, \kappa \in \mathbf{Z}_m$, for $\kappa (\neq 0)$ fixed. (Using $\kappa = 0$ would not alter any of the characters in a message.) Consequently, there are $m - 1$ possibilities to examine in an attempt to discover the value of the key κ . Further, once we know the value of κ , we also know the decryption function $D: \mathbf{Z}_m \rightarrow \mathbf{Z}_m$, for $D(\theta) = (\theta - \kappa) \bmod m$. In the case of the affine-cipher cryptosystem (also with an alphabet of m characters) the encryption function $E: \mathbf{Z}_m \rightarrow \mathbf{Z}_m$ is now given by $E(\theta) = (\alpha\theta + \kappa) \bmod m$, where $\theta, \alpha, \kappa \in \mathbf{Z}_m$, for fixed α, κ , with α invertible in \mathbf{Z}_m [or, equivalently, with $\gcd(\alpha, m) = 1$]. Here the decryption function $D: \mathbf{Z}_m \rightarrow \mathbf{Z}_m$ is given by $D(\theta) = [\alpha^{-1}(\theta - \kappa)] \bmod m$. Without prior knowledge of the key (α, κ) , now one would have to check $m\phi(m)$ possibilities to discover the appropriate values of α and κ for this private-key cryptosystem.

The security of either of the above cryptosystems depends on having the key [be it κ or (α, κ)] known only to the sender and the recipient of the messages.

The RSA cryptosystem is an example of a *public-key* cryptosystem. This cryptosystem was developed in the 1970s (and patented in 1983) by Ronald Rivest (1948–), Adi Shamir (1952–), and Leonard Adleman (1945–). (Taking the first letter from the surname of each of these three men provides the adjective RSA.)

We shall describe how this cryptosystem works and provide an example for encryption and decryption. In so doing, we shall find ourselves using some of the results from Sections 14.3 and 16.3.

EXAMPLE 16.18

As with the two private-key cryptosystems, once again we have an alphabet of m characters. We start with two distinct primes p, q . In practice, these should be large primes — each with 100 or more digits. (However, for our example we shall use much smaller primes.) After selecting the primes p, q , we then consider the integers $n = pq$ and $r = (p - 1) \cdot (q - 1) = \phi(p)\phi(q) = \phi(pq) = \phi(n)$, and, at this point, we choose an invertible element e in $\mathbf{Z}_r = (\mathbf{Z}_{\phi(n)})$.

[Here, if the element e is chosen at random, then the only time we fail to obtain an invertible element is when the element chosen is a multiple of p (there are q possibilities) or a multiple of q (there are p possibilities). In this count of $p + q$ elements we have accounted for pq twice, so there are only $p + q - 1$ possibilities for failure. Hence, the probability for

failure is $(p + q - 1)/(pq) = (1/q) + (1/p) - (1/(pq))$, a very small number if p and q each have 100 or more digits.]

For instance, consider $p = 61$, $q = 127$, with $n = (61)(127) = 7747$ and $r = \phi(61) \cdot \phi(127) = (60)(126) = 7560$. Now suppose we select e as 17.

Consider the following message that we wish to encrypt.

INVEST IN BONDS

Using the same plaintext assignments as in part (b) of Example 14.15, here we would replace the letter "I" by 08 (not merely 8). Then we replace "N" by 13. This provides us with the first block of four digits—namely, 0813—for the first two letters "IN". The assignment for the complete message is as follows [where we have appended the letter "X" to the right end, in order for the final block to have two letters (or, four digits)]:

I	N	V	E	S	T	I	N	B	O	N	D	S	X
08	13	21	04	18	19	08	13	01	14	13	03	18	23

We now encrypt each block B of four digits by the encryption function E , where $E(B) = B^e \bmod n$. (This modular exponentiation can be carried out efficiently by using the procedure in Example 14.16.) So here the domain of E is the concatenation of \mathbf{Z}_{26} with itself, and we find that

$$\begin{array}{lll}
 0813^{17} \bmod 7747 = 2169 & 2104^{17} \bmod 7747 = 0628 & 1819^{17} \bmod 7747 = 5540 \\
 0813^{17} \bmod 7747 = 2169 & 0114^{17} \bmod 7747 = 6560 & 1303^{17} \bmod 7747 = 6401 \\
 1823^{17} \bmod 7747 = 4829.
 \end{array}$$

Consequently, the recipient of the encrypted assignment (for the given plaintext message) receives the ciphertext

2169 0628 5540 2169 6560 6401 4829.

Now the question is: "How does the recipient decrypt the ciphertext received?"

Since e is a unit in $\mathbf{Z}_r (= \mathbf{Z}_{\phi(n)})$, we can use the Euclidean algorithm (as in Example 14.13) to compute $e^{-1} = d$. Then we define the decryption function D , where $D(C) = C^d \bmod n$, for a block C of four digits. Since $e^{-1} = d$, it follows that $ed \equiv 1 \bmod \phi(n)$ —that is, $ed \bmod \phi(n) = 1$. Therefore, $ed = k\phi(n) + 1$, for some $k \in \mathbf{Z}$. Now recall the argument given earlier for the probability that a randomly selected element e from \mathbf{Z}_n is invertible (or a unit in \mathbf{Z}_n). For any block B of four digits, we consider B as an element of \mathbf{Z}_n —in fact, we consider B as a unit in \mathbf{Z}_n . Since the units in the ring $(\mathbf{Z}_n, +, \cdot)$ form a group of order $\phi(n)$ under multiplication, it follows from the result in Exercise 8 of Section 16.3 that $B^{ed} = B^{k\phi(n)+1} = (B^{\phi(n)})^k B^1 \equiv B \pmod{n}$, or $B^{ed} \bmod n = B$. [This is also a consequence of Euler's Theorem, as stated in part (b) of Exercise 13 in Section 16.3.]

Applying the result from the previous paragraph in our example we have $p = 61$, $q = 127$, $n = pq = 7747$, $r = \phi(n) = (p - 1)(q - 1) = (60)(126) = 7560$, and $e = 17$. From the Euclidean algorithm we calculate $d = e^{-1} = 3113$. Now we find, for instance, that $2169^{3113} \bmod 7747 = 0813$ and that $0628^{3113} \bmod 7747 = 2104$. Continuing, the recipient determines the numeric assignment for the original plaintext and then the plaintext.

Now what makes the RSA cryptosystem more secure than the private-key cryptosystems we studied? First, we should relate that the RSA cryptosystem is *not* a private-key cryptosystem. This system is an example of a *public-key* cryptosystem, where the key (n, e) is made public. So it seems that all one needs to do to decrypt the encrypted assignment is

to determine $d = e^{-1}$ in \mathbf{Z}_r ($= \mathbf{Z}_{\phi(n)}$). Now it is time to realize that by knowing n we do not immediately know r . For to be able to determine $r = (p-1)(q-1)$, we need to know p, q , the prime factors of n . And this is what makes this system so much more secure than the other cryptosystems we mentioned. Determining the primes p, q , when they are 100 or more digits long, is not a feasible problem. However, as computer power continues to improve, to keep the RSA cryptosystem secure, one may need to redefine the key using primes with more and more digits.

In closing, we show how the problem of factoring the modulus n as pq is related to the problem of determining $r = (p-1)(q-1)$. We start by observing that

$$p + q = pq - (p-1)(q-1) + 1 = n - \phi(n) + 1 = n - r + 1,$$

while

$$\begin{aligned} p - q &= \sqrt{(p+q)^2 - 4pq} = \sqrt{(n-r+1)^2 - 4pq} \\ &= \sqrt{(p+q)^2 - 4n} = \sqrt{(n-r+1)^2 - 4n}. \end{aligned}$$

Then, from these two equations, we learn that

$$p = (1/2)[(p+q) + (p-q)] = (1/2)[(n-r+1) + \sqrt{(n-r+1)^2 - 4n}]$$

and

$$q = (1/2)[(p+q) - (p-q)] = (1/2)[(n-r+1) - \sqrt{(n-r+1)^2 - 4n}].$$

Consequently, when we know n and r , then we can readily determine the primes p, q such that $n = pq$.

EXERCISES 16.4

The use of a computer algebra system is strongly recommended for the first four exercises.

1. Determine the ciphertext for the plaintext INVEST IN STOCKS, when using RSA encryption with $e = 7$ and $n = 2573$.
2. Determine the ciphertext for the plaintext ORDER A PIZZA, when using RSA encryption with $e = 5$ and $n = 1459$.

3. Determine the plaintext for the RSA ciphertext 1418 1436 2370 1102 1805 0250, if $e = 11$ and $n = 2501$.

4. Determine the plaintext for the RSA ciphertext 0986 3029 1134 1105 1232 2281 2967 0272 1818 2398 1153, if $e = 17$ and $n = 3053$.

5. Find the primes p, q if $n = pq = 121,361$ and $\phi(n) = 120,432$.

6. Find the primes p, q if $n = pq = 5,446,367$ and $\phi(n) = 5,441,640$.

16.5 Elements of Coding Theory

In this and the next four sections we introduce an area of applied mathematics called *algebraic coding theory*. This theory was inspired by the fundamental paper of Claude Shannon (1948) along with results by Marcel Golay (1949) and Richard Hamming (1950). Since that time it has become an area of great interest where algebraic structures, probability, and combinatorics all play a role.

Our coverage will be held to an introductory level as we seek to model the transmission of information represented by strings of the signals 0 and 1.

In digital communications, when information is transmitted in the form of strings of 0's and 1's, certain problems arise. As a result of "noise" in the channel, when a certain signal is transmitted a different signal may be received, thus causing the receiver to make a wrong

decision. Hence we want to develop techniques to help us detect, and perhaps even correct, transmission errors. However, we can only improve the chances of correct transmission; there are no guarantees.

Our model uses a *binary symmetric channel*, as shown in Fig. 16.2. The adjective *binary* appears because an individual signal is represented by one of the bits 0 or 1. When a transmitter sends the signal 0 or 1 in such a channel, associated with either signal is a (constant) probability p for incorrect transmission. When that probability p is the same for both signals, the channel is called *symmetric*. Here, for example, we have probability p of sending 0 and having 1 received. The probability of sending signal 0 and having it received correctly is then $1 - p$. All possibilities are illustrated in Fig. 16.2.

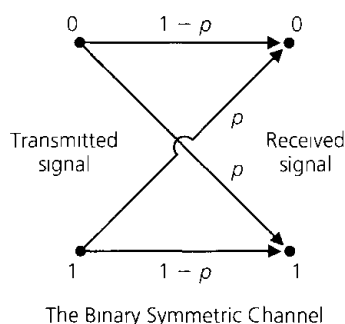


Figure 16.2

EXAMPLE 16.19

Consider the string $c = 10110$. We regard c as an element of the group \mathbf{Z}_2^5 , formed from the direct product of five copies of $(\mathbf{Z}_2, +)$. To shorten notation we write 10110 instead of $(1, 0, 1, 1, 0)$. When sending each bit (individual signal) of c through the binary symmetric channel, we assume that the probability of incorrect transmission is $p = 0.05$, so that the probability of transmitting c with no errors is $(0.95)^5 \doteq 0.77$.

Here, and throughout our discussion of coding theory, we assume that the transmission of each signal does not depend in any way on the transmissions of prior signals. Consequently, the probability of the occurrence of all of these *independent* events (in their prescribed order) is given by the product of their individual probabilities.

What is the probability that the party receiving the five-bit message receives the string $r = 00110$ —that is, the original message with an error in the first position? The probability of incorrect transmission for the first bit is 0.05, so with the assumption of independent events, $(0.05)(0.95)^4 \doteq 0.041$ is the probability of sending $c = 10110$ and receiving $r = 00110$. With $e = 10000$, we can write $c + e = r$ and interpret r as the result of the sum of the original message c and the particular *error pattern* $e = 10000$. Since $c, r, e \in \mathbf{Z}_2^5$ and $-1 = 1$ in \mathbf{Z}_2 , we also have $c + r = e$ and $r + e = c$.

In transmitting $c = 10110$, the probability of receiving $r = 00100$ is

$$(0.05)(0.95)^2(0.05)(0.95) \doteq 0.002,$$

so this multiple error is not very likely to occur.

Finally if we transmit $c = 10110$, what is the probability that r differs from c in exactly two places? To answer this we sum the probabilities for each error pattern consisting of two 1's and three 0's. Each such pattern has probability 0.002. There are $\binom{5}{2}$ such patterns, so

the probability of two errors in transmission is given by

$$\binom{5}{2}(0.05)^2(0.95)^3 \doteq 0.021.$$

These results lead us to the following theorem.

THEOREM 16.10

Let $c \in \mathbf{Z}_2^n$. For the transmission of c through a binary symmetric channel with probability p of incorrect transmission,

- a) the probability of receiving $r = c + e$, where e is a *particular* error pattern consisting of k 1's and $(n - k)$ 0's, is $p^k(1 - p)^{n-k}$.
- b) the probability that (exactly) k errors are made in the transmission is

$$\binom{n}{k}p^k(1 - p)^{n-k}.\dagger$$

In Example 16.19, the probability of making at most one error in the transmission of $c = 10110$ is $(0.95)^5 + \binom{5}{1}(0.05)(0.95)^4 \doteq 0.977$. Thus the chance for multiple errors in transmission will be considered negligible throughout the discussion in this chapter. Such an assumption is valid when p is small. In actuality, a binary symmetric channel is considered “good” when $p < 10^{-5}$. However, no matter what else we stipulate, we always want $p < 1/2$.

To improve the accuracy of transmission in a binary symmetric channel, certain types of coding schemes can be used where extra bits are provided.

For $m, n \in \mathbf{Z}^+$, let $n > m$. Consider $\emptyset \neq W \subseteq \mathbf{Z}_2^m$. The set W consists of the *messages* to be transmitted. To each $w \in W$ are appended $n - m$ extra bits to form the *code word* c , where $c \in \mathbf{Z}_2^n$. This process is called *encoding* and is represented by the function $E: W \rightarrow \mathbf{Z}_2^n$. Then $E(w) = c$ and $E(W) = C \subseteq \mathbf{Z}_2^n$. Since the function E simply appends extra bits to the (distinct) messages, the encoding process is one-to-one. Upon transmission, c is received as $T(c)$, where $T(c) \in \mathbf{Z}_2^n$. Unfortunately, T is not a function because $T(c)$ may be different at different transmission times (for the noise in the channel changes with time). (See Fig. 16.3.)

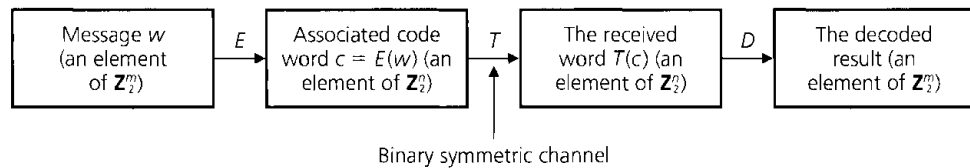


Figure 16.3

Upon receiving $T(c)$, we want to apply a decoding function $D: \mathbf{Z}_2^n \rightarrow \mathbf{Z}_2^m$ to remove the extra bits and, we hope, obtain the original message w . Ideally $D \circ T \circ E$ should be the identity function on W , with $D: C \rightarrow W$. Since this cannot be expected, we seek functions E and D such that there is a high probability of correctly decoding the received word $T(c)$ and recapturing the original message w . In addition, we want the ratio m/n to be as large as possible so that an excessive number of bits are not appended to w in getting the code

[†]This is the binomial probability distribution that was developed in (optional) Sections 3.5 and 3.7.

word $c = E(w)$. This ratio m/n measures the *efficiency* of our scheme and is called the *rate* of the code. Finally, the functions E and D should be more than theoretical results; they must be practical in the sense that they can be implemented electronically.

In such a scheme, the functions E and D are called the *encoding* and *decoding* functions, respectively, of an (n, m) *block code*.

We illustrate these ideas in the following two examples.

EXAMPLE 16.20

Consider the $(m + 1, m)$ block code for $m = 8$. Let $W = \mathbf{Z}_2^8$. For each $w = w_1 w_2 \cdots w_8 \in W$, define $E: \mathbf{Z}_2^8 \rightarrow \mathbf{Z}_2^9$ by $E(w) = w_1 w_2 \cdots w_8 w_9$, where $w_9 = \sum_{i=1}^8 w_i$, with the addition performed modulo 2. For example, $E(11001101) = 110011011$, and $E(00110011) = 001100110$.

For all $w \in \mathbf{Z}_2^8$, $E(w)$ contains an even number of 1's. So for $w = 11010110$ and $E(w) = 110101101$, if we receive $T(c) = T(E(w))$ as 100101101, from the odd number of 1's in $T(c)$ we know that a mistake has occurred in transmission. Hence we are able to *detect* single errors in transmission. But we seem to have no way to correct such errors.

The probability of sending the code word 110101101 and making at most one error in transmission is

$$\underbrace{(1-p)^9}_{\text{All nine bits are correctly transmitted.}} + \underbrace{\binom{9}{1}p(1-p)^8}_{\text{One bit is changed in transmission and an error is detected.}}$$

For $p = 0.001$ this gives $(0.999)^9 + \binom{9}{1}(0.001)(0.999)^8 \doteq 0.99996417$.

If we detect an error and we are able to relay a signal back to the transmitter to repeat the transmission of the code word, and continue this process until the received word has an even number of 1's, then the probability of sending the code word 110101101 and receiving the correct transmission is approximately 0.99996393.[†]

Should an even positive number of errors occur in transmission, $T(c)$ is unfortunately accepted as the correct code word and we interpret its first eight components as the original message. This scheme is called the $(m + 1, m)$ *parity-check code* and is appropriate only when multiple errors are not likely to occur.

If we send the message 11010110 through the channel, we have probability $(0.999)^8 = 0.99202794$ of correct transmission. By using this parity-check code, we increase our chances of getting the correct message to (approximately) 0.99996393. However, an extra signal is sent (and perhaps additional transmissions are needed) and the rate of the code has decreased from 1 to 8/9.

But suppose that instead of sending eight bits we sent 160 bits, in successive strings of length 8. The chances of receiving the correct message without any coding scheme would be

[†]For $p = 0.001$ the probability that an odd number of errors occurs in the transmission of the code word 110101101 is

$$\begin{aligned} p_{\text{odd}} &= \binom{9}{1}(0.999)^8(0.001) + \binom{9}{3}(0.999)^6(0.001)^3 + \binom{9}{5}(0.999)^4(0.001)^5 + \binom{9}{7}(0.999)^2(0.001)^7 + \binom{9}{9}(0.001)^9 \\ &\doteq 0.008928251 + 0.000000083 + 0.000000000 + 0.000000000 + 0.000000000 = 0.008928334. \end{aligned}$$

With $q =$ the probability of the correct transmission of 110101101 $= (0.999)^9$, the probability that this code word is transmitted and correctly received under these conditions (of retransmission) is then given by

$$q + p_{\text{odd}} \cdot q + (p_{\text{odd}})^2 q + (p_{\text{odd}})^3 q + \cdots = q/(1 - p_{\text{odd}}) \doteq 0.99996393 \text{ (to eight decimal places).}$$

$(0.999)^{160} \doteq 0.85207557$. With the parity-check method we send 180 bits, but the chances for correct transmission now increase to $(0.999964)^{20} \doteq 0.99928025$.

EXAMPLE 16.21

The $(3m, m)$ *triple repetition code* is one where we can both *detect* and *correct* single errors in transmission. With $m = 8$ and $W = \mathbf{Z}_2^8$, we define $E: \mathbf{Z}_2^8 \rightarrow \mathbf{Z}_2^{24}$ by $E(w_1 w_2 \cdots w_7 w_8) = w_1 w_2 \cdots w_8 w_1 w_2 \cdots w_8 w_1 w_2 \cdots w_8$.

Hence if $w = 10110111$, then $c = E(w) = 101101111011011110110111$.

The decoding function $D: \mathbf{Z}_2^{24} \rightarrow \mathbf{Z}_2^8$ is carried out by the majority rule. For example, if $T(c) = 101001110011011110110110$, then we have three errors occurring in positions 4, 9, and 24. We decode $T(c)$, by examining the first, ninth, and seventeenth positions to see which signal appears more times. Here it is 1 (which occurs twice), so we decode the first entry in the decoded message as 1. Continuing with the entries in the second, tenth, and eighteenth positions, the result for the second entry of the decoded message is 0 (which occurs all three times). As we proceed, we recapture the correct message, 10110111.

Although we have more than one transmission error here, all is well unless two (or more) errors occur with the second error eight or sixteen spaces after the first — that is, if two (or more) incorrect transmissions occur for the same bit of the original message.

Now how does this scheme compare with the other methods we have? With $p = 0.001$, the probability of correctly decoding a single bit is $(0.999)^3 + \binom{3}{1}(0.001)(0.999)^2 \doteq 0.99999700$. So the probability of receiving and correctly decoding the eight-bit message is $(0.99999700)^8 = 0.99997600$, just slightly better than the result from the parity-check method (where we may have to retransmit, thus increasing the overall transmission time). Here we transmit 24 signals for this message, so our rate is now $1/3$. For this increased accuracy and the ability to detect and now *correct* single errors (which we could not do in any previous schemes), we may pay with an increase in transmission time. But we do not waste time with retransmissions.

EXERCISES 16.5

1. Let C be a set of code words, where $C \subseteq \mathbf{Z}_2^7$. In each of the following, two of e (error pattern), r (received word) and c (code word) are given, with $r = c + e$. Determine the third term.

- a) $c = 1010110$, $r = 1011111$
- b) $c = 1010110$, $e = 0101101$
- c) $e = 0101111$, $r = 0000111$

2. A binary symmetric channel has probability $p = 0.05$ of incorrect transmission. If the code word $c = 011011101$ is transmitted, what is the probability that (a) we receive $r = 011111101$? (b) we receive $r = 111011100$? (c) a single error occurs? (d) a double error occurs? (e) a triple error occurs? (f) three errors occur, no two of them consecutive?

3. Let $E: \mathbf{Z}_2^3 \rightarrow \mathbf{Z}_2^9$ be the encoding function for the $(9, 3)$ triple repetition code.

- a) If $D: \mathbf{Z}_2^9 \rightarrow \mathbf{Z}_2^3$ is the corresponding decoding function, apply D to decode the received words (i) 111101100;

(ii) 000100011; (iii) 010011111.

- b) Find three different received words r for which $D(r) = 000$.

c) For each $w \in \mathbf{Z}_2^3$, what is $|D^{-1}(w)|$?

4. The $(5m, m)$ five-times repetition code has encoding function $E: \mathbf{Z}_2^m \rightarrow \mathbf{Z}_2^{5m}$, where $E(w) = wwwww$. Decoding with $D: \mathbf{Z}_2^{5m} \rightarrow \mathbf{Z}_2^m$ is accomplished by the majority rule. (Here we are able to correct single and double errors made in transmission.)

- a) With $p = 0.05$, what is the probability for the transmission and correct decoding of the signal 0?

b) Answer part (a) for the message 110 in place of the signal 0.

- c) For $m = 2$, decode the received word

$$r = 0111001001.$$

- d) If $m = 2$, find three received words r where $D(r) = 00$.

e) For $m = 2$ and $D: \mathbf{Z}_2^{10} \rightarrow \mathbf{Z}_2^2$, what is $|D^{-1}(w)|$ for each $w \in \mathbf{Z}_2^2$?

16.6

The Hamming Metric

In this section we develop the general principles for discussing the error-detecting and error-correcting capabilities of a coding scheme. These ideas were developed by Richard Wesley Hamming (1915–1998).

We start by considering a code $C \subseteq \mathbf{Z}_2^4$, where $c_1 = 0111$, $c_2 = 1111 \in C$. Now both the transmitter and the receiver know the elements of C . So if the transmitter sends c_1 but the person receiving the code word receives $T(c_1)$ as 1111, then he or she feels that c_2 was transmitted and makes whatever decision (a wrong one) c_2 implies. Consequently, although only one transmission error was made, the results could be unpleasant. Why is this? Unfortunately we have two code words that are almost the same. They are rather *close* to each other, for they differ in only one component.

We describe this notion of closeness more precisely as follows.

Definition 16.9

For each element $x = x_1x_2 \cdots x_n \in \mathbf{Z}_2^n$, where $n \in \mathbf{Z}^+$, the *weight* of x , denoted $\text{wt}(x)$, is the number of components x_i of x , for $1 \leq i \leq n$, where $x_i = 1$. If $y \in \mathbf{Z}_2^n$, the *distance* between x and y , denoted $d(x, y)$, is the number of components where $x_i \neq y_i$, for $1 \leq i \leq n$.

EXAMPLE 16.22

For $n = 5$, let $x = 01001$ and $y = 11101$. Then $\text{wt}(x) = 2$, $\text{wt}(y) = 4$, and $d(x, y) = 2$. In addition, $x + y = 10100$, so $\text{wt}(x + y) = 2$. Is it just by chance that $d(x, y) = \text{wt}(x + y)$? For each $1 \leq i \leq 5$, $x_i + y_i$ contributes a count of 1 to $\text{wt}(x + y) \iff x_i \neq y_i \iff x_i, y_i$ contribute a count of 1 to $d(x, y)$. [This is actually true for all $n \in \mathbf{Z}^+$, so $\text{wt}(x + y) = d(x, y)$ for all $x, y \in \mathbf{Z}_2^n$.]

When $x, y \in \mathbf{Z}_2^n$, we write $d(x, y) = \sum_{i=1}^n d(x_i, y_i)$ where,

$$\text{for each } 1 \leq i \leq n, \quad d(x_i, y_i) = \begin{cases} 0 & \text{if } x_i = y_i \\ 1 & \text{if } x_i \neq y_i. \end{cases}$$

LEMMA 16.2

For all $x, y \in \mathbf{Z}_2^n$, $\text{wt}(x + y) \leq \text{wt}(x) + \text{wt}(y)$.

Proof: We prove this lemma by examining, for each $1 \leq i \leq n$, the components $x_i, y_i, x_i + y_i$, of $x, y, x + y$, respectively. Only one situation would cause this inequality to be false: if $x_i + y_i = 1$ while $x_i = 0$ and $y_i = 0$, for some $1 \leq i \leq n$. But this never occurs because $x_i + y_i = 1$ implies that exactly one of x_i and y_i is 1.

In Example 16.22 we found that

$$\text{wt}(x + y) = \text{wt}(10100) = 2 \leq 2 + 4 = \text{wt}(01001) + \text{wt}(11101) = \text{wt}(x) + \text{wt}(y).$$

THEOREM 16.11

The distance function d defined on $\mathbf{Z}_2^n \times \mathbf{Z}_2^n$ satisfies the following for all $x, y, z \in \mathbf{Z}_2^n$.

- | | |
|--|---|
| <p>a) $d(x, y) \geq 0$</p> <p>c) $d(x, y) = d(y, x)$</p> | <p>b) $d(x, y) = 0 \iff x = y$</p> <p>d) $d(x, z) \leq d(x, y) + d(y, z)$</p> |
|--|---|

Proof: We leave the first three parts for the reader and prove part (d).

In \mathbf{Z}_2^n , $y + y = 0$, so $d(x, z) = \text{wt}(x + z) = \text{wt}(x + (y + y) + z) = \text{wt}((x + y) + (y + z)) \leq \text{wt}(x + y) + \text{wt}(y + z)$, by Lemma 16.2. With $\text{wt}(x + y) = d(x, y)$ and $\text{wt}(y + z) = d(y, z)$, the result follows. (This property is generally called the *Triangle Inequality*.)

When a function satisfies the four properties listed in Theorem 16.11, it is called a *distance function* or *metric*, and we call (\mathbf{Z}_2^n, d) a *metric space*. Hence d (as given above) is often referred to as the *Hamming metric*. This metric is used in the following.

Definition 16.10

For $n, k \in \mathbf{Z}^+$ and $x \in \mathbf{Z}_2^n$, the *sphere* of radius k centered at x is defined as $S(x, k) = \{y \in \mathbf{Z}_2^n \mid d(x, y) \leq k\}$.

EXAMPLE 16.23

For $n = 3$ and $x = 110 \in \mathbf{Z}_2^3$, $S(x, 1) = \{110, 010, 100, 111\}$ and $S(x, 2) = \{110, 010, 100, 111, 000, 101, 011\}$.

With these preliminaries in hand we turn now to the two major results of this section.

THEOREM 16.12

Let $E: W \rightarrow C$ be an encoding function with the set of messages $W \subseteq \mathbf{Z}_2^m$ and the set of code words $E(W) = C \subseteq \mathbf{Z}_2^n$, where $m < n$. If our objective is error detection, then for $k \in \mathbf{Z}^+$, we can detect all transmission errors of weight $\leq k$ if and only if the minimum distance between code words is at least $k + 1$.

Proof: The set C is known to both the transmitter and the receiver, so if $w \in W$ is the message and $c = E(w)$ is transmitted, let $c \neq T(c) = r$. If the minimum distance between code words is at least $k + 1$, then the transmission of c can result in as many as k errors and r will not be listed in C . Hence we can detect all errors e where $\text{wt}(e) \leq k$. Conversely, let c_1, c_2 be code words with $d(c_1, c_2) < k + 1$. Then $c_2 = c_1 + e$ where $\text{wt}(e) \leq k$. If we send c_1 and $T(c_1) = c_2$, then we would feel that c_2 had been sent, thus failing to detect an error of weight $\leq k$.

What can we say about error-correcting capability?

THEOREM 16.13

Let E, W , and C be as in Theorem 16.12. If our objective is error correction, then for $k \in \mathbf{Z}^+$, we can construct a decoding function $D: \mathbf{Z}_2^n \rightarrow W$ that corrects all transmission errors of weight $\leq k$ if and only if the minimum distance between code words is at least $2k + 1$.

Proof: For $c \in C$, consider $S(c, k) = \{x \in \mathbf{Z}_2^n \mid d(c, x) \leq k\}$. Define $D: \mathbf{Z}_2^n \rightarrow W$ as follows. If $r \in \mathbf{Z}_2^n$ and $r \in S(c, k)$ for some code word c , then $D(r) = w$ where $E(w) = c$. [Here c is the (unique) code word *nearest* to r .] If $r \notin S(c, k)$ for any $c \in C$, then we define $D(r) = w_0$, where w_0 is some arbitrary message that remains fixed once it is chosen. The only problem we could face here is that D might not be a function. This will happen if there is an element r in \mathbf{Z}_2^n with r in both $S(c_1, k)$ and $S(c_2, k)$ for distinct code words c_1, c_2 . But $r \in S(c_1, k) \Rightarrow d(c_1, r) \leq k$, and $r \in S(c_2, k) \Rightarrow d(c_2, r) \leq k$, so $d(c_1, c_2) \leq d(c_1, r) + d(r, c_2) \leq k + k < 2k + 1$. Consequently, if the minimum distance between code words is at least $2k + 1$, then D is a function, and it will decode all possible

received words, correcting any transmission error of weight $\leq k$. Conversely, if $c_1, c_2 \in C$ and $d(c_1, c_2) \leq 2k$, then c_2 can be obtained from c_1 by making at most $2k$ changes. Starting at code word c_1 we make approximately half (exactly, $\lfloor d(c_1, c_2)/2 \rfloor$) of these changes. This brings us to $r = c_1 + e_1$ with $\text{wt}(e_1) \leq k$. Continuing from r , we make the remaining changes to get to c_2 and find $r + e_2 = c_2$ with $\text{wt}(e_2) \leq k$. But then $r = c_2 + e_2$. Now with $c_1 + e_1 = r = c_2 + e_2$ and $\text{wt}(e_1), \text{wt}(e_2) \leq k$, how can one decide on the code word from which r arises? This ambiguity results in a possible error of weight $\leq k$ that cannot be corrected.

EXAMPLE 16.24

With $W = \mathbf{Z}_2^2$ let $E: W \rightarrow \mathbf{Z}_2^6$ be given by

$$E(00) = 000000 \quad E(10) = 101010 \quad E(01) = 010101 \quad E(11) = 111111.$$

Then the minimum distance between code words is 3, so we can correct all single errors. With

$$\begin{aligned} S(000000, 1) &= \{x \in \mathbf{Z}_2^6 \mid d(000000, x) \leq 1\} \\ &= \{000000, 100000, 010000, 001000, 000100, 000010, 000001\}, \end{aligned}$$

the decoding function $D: \mathbf{Z}_2^6 \rightarrow W$ gives $D(x) = 00$ for all $x \in S(000000, 1)$.

Similarly,

$$\begin{aligned} S(010101, 1) &= \{x \in \mathbf{Z}_2^6 \mid d(010101, x) \leq 1\} \\ &= \{010101, 110101, 000101, 011101, 010001, 010111, 010100\}, \end{aligned}$$

and here $D(x) = 01$ for each $x \in S(010101, 1)$. At this point our definition of D accounts for 14 of the elements in \mathbf{Z}_2^6 . Continuing to define D for the 14 elements in $S(101010, 1)$ and $S(111111, 1)$ there remain 36 other elements to account for. We define $D(x) = 00$ (or any other message) for these 36 other elements and have a decoding function that will correct single errors.

Beware! There is a subtle point that needs to be made about Theorems 16.12 and 16.13. For example, if the minimum distance between code words is $2k + 1$ one may feel that we can detect all errors of weight $\leq 2k$ and correct all errors of weight $\leq k$. This is not necessarily true. That is, error detection and error correction need not take place at the same time and at the maximum levels. To see this, reconsider the $(6, 2)$ -triple repetition code of Example 16.24. Here the encoding function $E: W (= \mathbf{Z}_2^2) \rightarrow \mathbf{Z}_2^6$ is given by $E(w_1 w_2) = w_1 w_2 w_1 w_2 w_1 w_2$ and the code comprises the four elements of \mathbf{Z}_2^6 in the range of E . Since the minimum distance between any two elements of \mathbf{Z}_2^2 is 1, it follows that the minimum distance between code words is 3 (as observed earlier in Example 16.24).

Now suppose that our major objective is error correction and that $r = 100000$ [$\notin E(W)$] is received. We see that $d(000000, r) = 1$, $d(101010, r) = 2$, $d(010101, r) = 4$, and $d(111111, r) = 5$. Consequently, we should choose to decode r as 000000, the unique code word nearest to r . Unfortunately, suppose that the actual message were 10 (with corresponding code word 101010), but we received $r = 100000$. Upon correcting r as 000000, we should then decode 000000 to get the incorrect message 00. And, in so doing, we have failed to detect an error of weight 2.

In this type of situation one can develop a scheme where a mixed strategy is used. Here both error correction and error detection may be carried out at some levels.

For $t \in \mathbb{N}$, if the received word is r and there is a unique code word c_1 such that $d(c_1, r) \leq t$, then we decode r as c_1 . (Note: The case where $r = c_1$ is covered when $t = 0$.) If there exists a second code word c_2 such that $d(c_2, r) = d(c_1, r)$, or if $d(c, r) > t$ for all code words c , then an error is declared (and retransmission is generally requested). Using this scheme, if the minimum distance between code words is at least $2t + s + 1$, for $s \in \mathbb{N}$, then we can correct all errors of weight $\leq t$ and detect all errors with weights between $t + 1$ and $t + s$, inclusive.

When using this scheme for the (6, 2)-triple repetition code, our options include:

- 1) $t = 0$; $s = 2$: Here we can detect all errors of weight ≤ 2 but we have no error-correction capability.
- 2) $t = 1$; $s = 0$: Single errors are corrected here but there is no error-detecting capability.

If we use the (10, 2)-five-times repetition code, then the minimum distance is 5. Applying the above scheme in this case, our options now include:

- 1) $t = 0$; $s = 4$: Here we can detect all errors of weight ≤ 4 but we have no error-correction capability.
- 2) $t = 1$; $s = 2$: Now single errors are corrected and we can also detect all errors e , where $2 \leq \text{wt}(e) \leq 3$.
- 3) $t = 2$; $s = 0$: All errors of weight ≤ 2 are corrected but there is no error-detecting capability.

[For more on this, the interested reader should examine Chapter 4 of the text by S. Roman [24].]

16.7

The Parity-Check and Generator Matrices

In this section we introduce an example where the encoding and decoding functions are given by matrices over \mathbb{Z}_2 . One of these matrices will help us to locate the *nearest* code word for a given received word. This will be especially helpful as the set C of code words grows larger.

EXAMPLE 16.25

Let

$$G = \begin{bmatrix} 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 1 \end{bmatrix}$$

be a 3×6 matrix over \mathbb{Z}_2 . The first three columns of G form the 3×3 identity matrix I_3 . Letting A denote the matrix formed from the last three columns of G , we write $G = [I_3 | A]$ to denote its structure. The (partitioned) matrix G is called a *generator matrix*.

We use G to define an encoding function $E: \mathbb{Z}_2^3 \rightarrow \mathbb{Z}_2^6$ as follows. For $w \in \mathbb{Z}_2^3$, $E(w) = wG$ is the element in \mathbb{Z}_2^6 obtained by multiplying w , considered as a three-dimensional row vector, by the matrix G on its right. Unlike the results on matrix multiplication in Chapter 7, in the calculations here we have $1 + 1 = 0$, not $1 + 1 = 1$.

(Even if the set W of messages is not all of \mathbb{Z}_2^3 , we'll assume that all of \mathbb{Z}_2^3 is encoded and that the transmitter and receiver will both know the real messages of importance and their corresponding code words.)

We find here, for example, that

$$E(110) = (110)G = [110] \begin{bmatrix} 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 1 \end{bmatrix} = [110101],$$

and

$$E(010) = (010)G = [010] \begin{bmatrix} 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 1 \end{bmatrix} = [010011].$$

Note that $E(110)$ can be obtained by adding the first two rows of G , whereas $E(010)$ is simply the second row of G .

The set of code words obtained by this method is

$$C = \{000000, 100110, 010011, 001101, 110101, 101011, 011110, 111000\} \subseteq \mathbf{Z}_2^6,$$

and one can recapture the corresponding message by simply dropping the last three components of the code word. In addition, the minimum distance between code words is 3, so we can detect errors of weight ≤ 2 or correct single errors. (We shall assume that multiple errors are rare and concentrate on error correction.)

For all $w = w_1 w_2 w_3 \in \mathbf{Z}_2^3$, $E(w) = w_1 w_2 w_3 w_4 w_5 w_6 \in \mathbf{Z}_2^6$. Since

$$\begin{aligned} E(w) &= [w_1 w_2 w_3] \begin{bmatrix} 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 1 \end{bmatrix} \\ &= [w_1 w_2 w_3 (w_1 + w_3)(w_1 + w_2)(w_2 + w_3)], \end{aligned}$$

we have $w_4 = w_1 + w_3$, $w_5 = w_1 + w_2$, $w_6 = w_2 + w_3$, and these equations are called the *parity-check equations*. Since $w_i \in \mathbf{Z}_2$ for each $1 \leq i \leq 6$, it follows that $w_i = -w_i$ and so the equations can be rewritten as

$$\begin{aligned} w_1 + w_3 + w_4 &= 0 \\ w_1 + w_2 + w_5 &= 0 \\ w_2 + w_3 + w_6 &= 0. \end{aligned}$$

Thus we find that

$$\begin{bmatrix} 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ w_4 \\ w_5 \\ w_6 \end{bmatrix} = H \cdot (E(w))^{\text{tr}} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix},$$

where $(E(w))^{\text{tr}}$ denotes the transpose of $E(w)$. Consequently, if $r = r_1 r_2 \cdots r_6 \in \mathbf{Z}_2^6$, we can identify r as a code word if and only if

$$H \cdot r^{\text{tr}} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

Writing $H = [B|I_3]$, we notice that if the rows and columns of B are interchanged, then we get A . Hence $B = A^{\text{tr}}$.

From the theory developed earlier on error correction, because the minimum distance between the code words of this example is 3, we should be able to develop a decoding function that corrects single errors.

Suppose we receive $r = 110110$. We want to find the code word c that is the *nearest neighbor* of r . If there is a long list of code words against which to check r , we would be better off to first examine $H \cdot r^{\text{tr}}$, which is called the *syndrome* of r . Here

$$H \cdot r^{\text{tr}} = \begin{bmatrix} 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 0 \\ 1 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix},$$

so r is not a code word. Hence we at least detect an error. Looking back at the list of code words, we see that $d(100110, r) = 1$. For all other $c \in C$, $d(r, c) \geq 2$. Writing $r = c + e = 100110 + 010000$, we find that the transmission error (of weight 1) occurs in the second component of r . Is it just a coincidence that the syndrome $H \cdot r^{\text{tr}}$ produced the second column of H ? If not, then we can use this result in order to realize that if a single transmission error occurred, it took place at the second component. Changing the second component of r , we get c ; the message w comprises the first three components of c .

Let $r = c + e$, where c is a code word and e is an error pattern of weight 1. Suppose that 1 is in the i th component of e , where $1 \leq i \leq 6$. Then

$$H \cdot r^{\text{tr}} = H \cdot (c + e)^{\text{tr}} = H \cdot (c^{\text{tr}} + e^{\text{tr}}) = H \cdot c^{\text{tr}} + H \cdot e^{\text{tr}}.$$

With c a code word, it follows that $H \cdot c^{\text{tr}} = \mathbf{0}$, so $H \cdot r^{\text{tr}} = H \cdot e^{\text{tr}} = i$ th column of matrix H . Thus c and r differ only in the i th component, and we can determine c by simply changing the i th component of r .

Since we are primarily concerned with transmissions where multiple errors are rare, this technique is of definite value. If we ask for more, however, we find ourselves expecting too much.

Suppose that we receive $r = 000111$. Computing the syndrome

$$H \cdot r^{\text{tr}} = \begin{bmatrix} 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix},$$

we obtain a result that is not one of the columns of H . Yet $H \cdot r^{\text{tr}}$ can be obtained as the sum of two columns from H . If $H \cdot r^{\text{tr}}$ came from the first and sixth columns of H , correcting these components in r results in the code word 100110. If we sum the third and fifth columns of H to get this syndrome, upon changing the third and fifth components of r we get a second code word, 001101. So we cannot expect H to correct multiple errors. This is no surprise since the minimum distance between code words is 3.

We summarize the results of Example 16.25 for the general situation. For $m, n \in \mathbf{Z}^+$ with $m < n$, the encoding function $E: \mathbf{Z}_2^m \rightarrow \mathbf{Z}_2^n$ is given by an $m \times n$ matrix G over \mathbf{Z}_2 . This matrix G is called the generator matrix for the code and has the form $[I_m | A]$, where

A is an $m \times (n - m)$ matrix. Here $E(w) = wG$ for each message $w \in \mathbf{Z}_2^m$, and the code $C = E(\mathbf{Z}_2^m) \subset \mathbf{Z}_2^n$.

The associated *parity-check matrix* H is an $(n - m) \times n$ matrix of the form $[A^T | I_{n-m}]$. This matrix can also be used to define the encoding function E , because if $w = w_1 w_2 \cdots w_m \in \mathbf{Z}_2^m$, then $E(w) = w_1 w_2 \cdots w_m w_{m+1} \cdots w_n$, where w_{m+1}, \dots, w_n can be determined from the set of $n - m$ (parity-check) equations that arise from $H \cdot (E(w))^T = \mathbf{0}$, the column vector of $n - m$ 0's.

This unique parity-check matrix H also provides a decoding scheme that corrects single errors in transmission if:

- a) H does not contain a column of 0's. (If the i th column of H had all 0's and $H \cdot r^T = \mathbf{0}$ for a received word r , we couldn't decide whether r was a code word or a received word whose i th component was incorrectly transmitted. We do not want to compare r with all code words when C is large.)
- b) No two columns of H are the same. (If the i th and j th columns of H are the same and $H \cdot r^T$ equals this repeated column, how would we decide which component of r to change?)

When H satisfies these two conditions, we get the following decoding algorithm. For each $r \in \mathbf{Z}_2^n$, if $T(c) = r$, then:

- 1) With $H \cdot r^T = \mathbf{0}$, we feel that the transmission was correct and that r is the code word that was transmitted. The decoded message then consists of the first m components of r .
- 2) With $H \cdot r^T$ equal to the i th column of H , we feel that there has been a single error in transmission and change the i th component of r in order to get the code word c . Here the first m components of c yield the original message.
- 3) If neither case 1 nor case 2 occurs, we feel that there has been more than one transmission error and we cannot provide a reliable way to decode in this situation.

We close with one final comment on the matrix H . If we start with a parity-check matrix $H = [B | I_{n-m}]$ and use it, as described above, to define the function E , then we obtain the same set of code words that is generated by the unique associated generator matrix $G = [I_m | B^T]$.

EXERCISES 16.6 and 16.7

1. For Example 16.24, list the elements in $S(101010, 1)$ and $S(111111, 1)$.

2. Decode each of the following received words for Example 16.24.

- | | |
|-----------|-----------|
| a) 110101 | b) 101011 |
| c) 001111 | d) 110000 |

3. a) If $x \in \mathbf{Z}_2^{10}$, determine $|S(x, 1)|$, $|S(x, 2)|$, $|S(x, 3)|$.

b) For $n, k \in \mathbf{Z}^+$ with $1 \leq k \leq n$, if $x \in \mathbf{Z}_2^n$, what is $|S(x, k)|$?

4. Let $E: \mathbf{Z}_2^5 \rightarrow \mathbf{Z}_2^{25}$ be an encoding function where the minimum distance between code words is 9. What is the largest value of k such that we can detect errors of weight $\leq k$? If we wish to correct errors of weight $\leq n$, what is the maximum value for n ?

5. For each of the following encoding functions, find the minimum distance between the code words. Discuss the error-detecting and error-correcting capabilities of each code.

- a) $E: \mathbf{Z}_2^2 \rightarrow \mathbf{Z}_2^5$
- | | |
|------------------------|------------------------|
| $00 \rightarrow 00001$ | $01 \rightarrow 01010$ |
| $10 \rightarrow 10100$ | $11 \rightarrow 11111$ |

- b) $E: \mathbf{Z}_2^2 \rightarrow \mathbf{Z}_2^{10}$
- | | |
|-----------------------------|-----------------------------|
| $00 \rightarrow 0000000000$ | $01 \rightarrow 0000011111$ |
| $10 \rightarrow 1111100000$ | $11 \rightarrow 1111111111$ |

c) $E: \mathbf{Z}_2^3 \rightarrow \mathbf{Z}_2^6$

000 \rightarrow 000111	001 \rightarrow 001001
010 \rightarrow 010010	011 \rightarrow 011100
100 \rightarrow 100100	101 \rightarrow 101010
110 \rightarrow 110001	111 \rightarrow 111000

d) $E: \mathbf{Z}_2^3 \rightarrow \mathbf{Z}_2^8$

000 \rightarrow 00011111	001 \rightarrow 00111010
010 \rightarrow 01010101	011 \rightarrow 01110000
100 \rightarrow 10001101	101 \rightarrow 10101000
110 \rightarrow 11000100	111 \rightarrow 11100011

6. a) Use the parity-check matrix H of Example 16.25 to decode the following received words.

i) 111101	ii) 110101
iii) 001111	iv) 100100
v) 110001	vi) 111111
vii) 111100	viii) 010100

b) Are all the results in part (a) uniquely determined?

7. The encoding function $E: \mathbf{Z}_2^2 \rightarrow \mathbf{Z}_2^5$ is given by the generator matrix

$$G = \begin{bmatrix} 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 1 \end{bmatrix}.$$

a) Determine all code words. What can we say about the error-detection capability of this code? What about its error-correction capability?

b) Find the associated parity-check matrix H .c) Use H to decode each of the following received words.

i) 11011	ii) 10101	iii) 11010
iv) 00111	v) 11101	vi) 00110

8. Define the encoding function $E: \mathbf{Z}_2^3 \rightarrow \mathbf{Z}_2^6$ by means of the parity-check matrix

$$H = \begin{bmatrix} 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 \end{bmatrix}.$$

a) Determine all code words.

b) Does this code correct all single errors in transmission?

9. Find the generator and parity-check matrices for the (9, 8) single parity-check coding scheme of Example 16.20.

10. a) Show that the 1×9 matrix $G = [1 \ 1 \ 1 \ \dots \ 1]$ is the generator matrix for the (9, 1) nine-times repetition code.b) What is the associated parity-check matrix H in this case?11. For an (n, m) code C with generator matrix $G = [I_m | A]$ and parity-check matrix $H = [A^T | I_{n-m}]$, the $(n, n-m)$ code C^d with generator matrix $[I_{n-m} | A^T]$ and parity-check matrix $[A | I_m]$ is called the *dual code* of C . Show that the codes in each of Exercises 9 and 10 constitute a pair of dual codes.12. Given $n \in \mathbf{Z}^+$, let the set $M(n, k) \subseteq \mathbf{Z}_2^n$ contain the maximum number of code words of length n , where the minimum distance between code words is $2k + 1$. Prove that

$$\frac{2^n}{\sum_{i=0}^{2k} \binom{n}{i}} \leq |M(n, k)| \leq \frac{2^n}{\sum_{i=0}^k \binom{n}{i}}.$$

(The upper bound on $|M(n, k)|$ is called the *Hamming bound*; the lower bound is referred to as the *Gilbert bound*.)

16.8

Group Codes: Decoding with Coset Leaders

Now that we've examined some introductory material on coding theory, it is time to see how the group structure enters the picture.

Definition 16.11

Let $E: \mathbf{Z}_2^m \rightarrow \mathbf{Z}_2^n$, for $n > m$, be an encoding function. The code $C = E(\mathbf{Z}_2^m)$ is called a *group code* if C is a subgroup of \mathbf{Z}_2^n .

Recall the encoding function $E: \mathbf{Z}_2^2 \rightarrow \mathbf{Z}_2^6$ (of Example 16.24) where

$$E(00) = 000000 \quad E(10) = 101010 \quad E(01) = 010101 \quad E(11) = 111111.$$

Here \mathbf{Z}_2^2 and \mathbf{Z}_2^6 are groups under componentwise addition modulo 2; the subset $C = E(\mathbf{Z}_2^2) = \{000000, 101010, 010101, 111111\}$ is a subgroup of \mathbf{Z}_2^6 , and an example of a group code. (Note that C contains 000000, the zero element of \mathbf{Z}_2^6 .)

In general when the code words form a group, we find that it is easier to compute the minimum distance between code words.

THEOREM 16.14

In a group code, the minimum distance between distinct code words is the minimum of the weights of the nonzero elements of the code.

Proof: Let $a, b, c \in C$ where $a \neq b$, $d(a, b)$ is minimum, and c is nonzero with minimum weight. By closure in the group C , $a + b$ is a code word. Since $d(a, b) = \text{wt}(a + b)$, by the choice of c we have $d(a, b) \geq \text{wt}(c)$. Also, $\text{wt}(c) = d(c, \mathbf{0})$, where $\mathbf{0}$ is a code word because C is a group. Then $d(c, \mathbf{0}) \geq d(a, b)$ by the choice of a, b , so $\text{wt}(c) \geq d(a, b)$. Consequently, $d(a, b) = \text{wt}(c)$.

If C is a set of code words and $|C| = 1024$, we have to compute $\binom{1024}{2} = 523,776$ distances to find the minimum distance between code words. But if we can recognize that C possesses a group structure, we need only compute the weights of the 1023 nonzero elements of C .

Is there some way to guarantee that the code words form a group? By Theorem 16.5(d), the homomorphic image of a subgroup is a subgroup, so if $E: \mathbf{Z}_2^m \rightarrow \mathbf{Z}_2^n$ is a group homomorphism, then $C = E(\mathbf{Z}_2^m)$ will be a subgroup of \mathbf{Z}_2^n . Our next result will use this fact to show that the codes we obtain when using a generator matrix G or a parity-check matrix H are group codes. Furthermore, the proof of this result reconfirms the observation we made (at the end of the previous section) about the code that arises from a generator matrix G or its associated parity-check matrix H .

THEOREM 16.15

Let $E: \mathbf{Z}_2^m \rightarrow \mathbf{Z}_2^n$ be an encoding function given by a generator matrix G or the associated parity-check matrix H . Then $C = E(\mathbf{Z}_2^m)$ is a group code.

Proof: We establish these results by proving that the function E arising from G or H is a group homomorphism.

If $x, y \in \mathbf{Z}_2^m$, then $E(x + y) = (x + y)G = xG + yG = E(x) + E(y)$. Hence E is a homomorphism and $C = E(\mathbf{Z}_2^m)$ is a group code [by virtue of part (d) of Theorem 16.5].

For the case of H , if x is a message, then $E(x) = x_1x_2 \cdots x_mx_{m+1} \cdots x_n$, where $x = x_1x_2 \cdots x_m \in \mathbf{Z}_2^m$ and $H \cdot (E(x))^{\text{tr}} = \mathbf{0}$. In particular, $E(x)$ is uniquely determined by these two properties. If y is also a message, then $x + y$ is likewise, and $E(x + y)$ has $(x_1 + y_1), (x_2 + y_2), \dots, (x_m + y_m)$ as its first m components, as does $E(x) + E(y)$. Further, $H \cdot (E(x) + E(y))^{\text{tr}} = H \cdot (E(x)^{\text{tr}} + E(y)^{\text{tr}}) = H \cdot E(x)^{\text{tr}} + H \cdot E(y)^{\text{tr}} = \mathbf{0} + \mathbf{0} = \mathbf{0}$. Since $E(x + y)$ is the unique element of \mathbf{Z}_2^n with $(x_1 + y_1), (x_2 + y_2), \dots, (x_m + y_m)$ as its first m components and with $H \cdot (E(x + y))^{\text{tr}} = \mathbf{0}$, it follows that $E(x + y) = E(x) + E(y)$. So E is a group homomorphism and, consequently, $C = \{c \in \mathbf{Z}_2^n \mid H \cdot c^{\text{tr}} = \mathbf{0}\}$ is a group code.

Now we use the group structure of C , together with its cosets in \mathbf{Z}_2^n , to develop a scheme for decoding. Our example uses the code developed in Example 16.25, but the procedure applies for every group code.

EXAMPLE 16.26

We develop a table for decoding as follows.

- 1) First list in a row the elements of the group code C , starting with the identity.

000000 100110 010011 001101 110101 101011 011110 111000.

- 2) Next select an element x of \mathbf{Z}_2^6 (\mathbf{Z}_2^n , in general) where x does not appear anywhere in the table developed so far and has minimum weight. Then list the elements of the

coset $x + C$, with $x + c$ directly below c for each $c \in C$. For $x = 100000$ we have

000000 100110 010011 001101 110101 101011 011110 111000
100000 000110 110011 101101 010101 001011 111110 011000.

- 3) Repeat step (2) until the cosets provide a partition of \mathbf{Z}_2^6 (\mathbf{Z}_2^n , in general). This results in the *decoding table* shown in Table 16.8.
- 4) Once the decoding table is constructed, for each received word r we find the column containing r and use the first three components of the code word c at the top of the column to decode r .

Table 16.8 Decoding Table for the Code of Example 16.25

000000	100110	010011	001101	110101	101011	011110	111000
100000	000110	110011	101101	010101	001011	111110	011000
010000	110110	000011	011101	100101	111011	001110	101000
001000	101110	011011	000101	111101	100011	010110	110000
000100	100010	010111	001001	110001	101111	011010	111100
000010	100100	010001	001111	110111	101001	011100	111010
000001	100111	010010	001100	110100	101010	011111	111001
010100	110010	000111	011001	100001	111111	001010	101100

From the table we find that the code words for the received words

$$r_1 = 101001 \quad r_2 = 111010 \quad r_3 = 001001 \quad r_4 = 111011$$

are

$$c_1 = 101011 \quad c_2 = 111000 \quad c_3 = 001101 \quad c_4 = 101011,$$

respectively. From these results the respective messages are

$$w_1 = 101 \quad w_2 = 111 \quad w_3 = 001 \quad w_4 = 101.$$

The entries in the first column of Table 16.8 are called the *coset leaders*. For the first seven rows, the coset leaders are the same in all tables, with some permutations of rows possible. However, for the last row, either 100001 or 001010 could have been used in place of 010100 because they also have minimum weight 2. So the table need not be unique. [As a result, not all double errors can be corrected because there may not be a unique code word at a minimum distance for each r in the last coset (the one with coset leader 010100). For example, $r = 001010$ has three closest code words (at distance 2) — namely, 000000, 101011, and 011110.]

How do the coset leaders really help us? It seems that the code words in the first row are what we used to decode r_1, r_2, r_3 , and r_4 above.

Consider the received words $r_1 = 101001$ and $r_2 = 111010$ in the sixth row, where the coset leader is $x = 000010$. Computing syndromes, we find that

$$H \cdot (r_1)^{\text{tr}} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = H \cdot (r_2)^{\text{tr}} = H \cdot x^{\text{tr}}.$$

This is not just a coincidence.

THEOREM 16.16

Let $C \subseteq \mathbb{Z}_2^n$ be a group code for a parity-check matrix H , and let $r_1, r_2 \in \mathbb{Z}_2^n$. For the table of cosets of C in \mathbb{Z}_2^n , r_1 and r_2 are in the same coset of C if and only if $H \cdot (r_1)^{\text{tr}} = H \cdot (r_2)^{\text{tr}}$.

Proof: If r_1 and r_2 are in the same coset, then $r_1 = x + c_1$ and $r_2 = x + c_2$, where x is the coset leader, and c_1 and c_2 are the code words at the tops of the respective columns for r_1 and r_2 . Then $H \cdot (r_1)^{\text{tr}} = H \cdot (x + c_1)^{\text{tr}} = H \cdot x^{\text{tr}} + H \cdot c_1^{\text{tr}} = H \cdot x^{\text{tr}} + \mathbf{0} = H \cdot x^{\text{tr}}$ because c_1 is a code word. Likewise, $H \cdot (r_2)^{\text{tr}} = H \cdot x^{\text{tr}}$, so r_1, r_2 have the same syndrome. Conversely, $H \cdot (r_1)^{\text{tr}} = H \cdot (r_2)^{\text{tr}} \Rightarrow H \cdot (r_1 + r_2)^{\text{tr}} = \mathbf{0} \Rightarrow r_1 + r_2$ is a code word c . Hence $r_1 + r_2 = c$, so $r_1 = r_2 + c$ and $r_1 \in r_2 + C$. Since $r_2 \in r_2 + C$, we have r_1, r_2 in the same coset.

In decoding received words, when Table 16.8 is used we must search through 64 elements to find a given received word. For $C \subseteq \mathbb{Z}_2^{12}$ there are 4096 strings, each with 12 bits. Such a searching process is tedious, so perhaps we should be thinking about having a computer do the searching. Presently it appears that this means storing the entire table: $6 \times 64 = 384$ bits of storage for Table 16.8; $12 \times 4096 = 49,152$ bits for $C \subseteq \mathbb{Z}_2^{12}$. We should like to improve this situation. Before things get better, however, they'll look worse as we enlarge Table 16.8, as shown in Table 16.9. This new table includes to the left of the coset leaders (the transposes of) the syndromes for each row.

Table 16.9 Decoding Table 16.8 with Syndromes

000	000000	100110	010011	001101	110101	101011	011110	111000
110	100000	000110	110011	101101	010101	001011	111110	011000
011	010000	110110	000011	011101	100101	111011	001110	101000
101	001000	101110	011011	000101	111101	100011	010110	110000
100	000100	100010	010111	001001	110001	101111	011010	111100
010	000010	100100	010001	001111	110111	101001	011100	111010
001	000001	100111	010010	001100	110100	101010	011111	111001
111	010100	110010	000111	011001	100001	111111	001010	101100

Now we can decode a received word r by the following procedure.

- 1) Compute the syndrome $H \cdot r^{\text{tr}}$.
- 2) Find the coset leader x to the right of $H \cdot r^{\text{tr}}$.
- 3) Add x to r to get c . (The code word c that we are seeking at the top of the column containing r satisfies $c + x = r$, or $c = x + r$.)

Consequently, all that is needed from Table 16.9 are the first two columns, which will require $(3)(8) + (6)(8) = 72$ storage bits. With 18 more storage bits for H we can store what we need for this decoding process, called *decoding by coset leaders*, in 90 storage bits, as opposed to the original estimate of 384 bits.

Applying this procedure to $r = 110110$, we find the syndrome

$$H \cdot r^{\text{tr}} = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}.$$

Since 011 is to the left of the coset leader $x = 010000$, the code word $c = x + r = 010000 + 110110 = 100110$, from which we recapture the original message, 100.

The code here is a group code where the minimum weight of the nonzero code words is 3, so we expected to be able to find a decoding scheme that corrected single errors. Here this is accomplished because the error patterns of weight 1 are all coset leaders. We cannot correct all double errors; only one error pattern of weight 2 is a coset leader. All error patterns of weight 1 or 2 would have to be coset leaders before our decoding scheme could correct both single and double errors in transmission.

Unlike the situation in Example 16.25, where syndromes were also used for decoding, things here are a bit different. Once we have a complete table listing all of the cosets of C in \mathbf{Z}_2^6 , the process of decoding by coset leaders will give us an answer for *all* received words, not just for those that are code words or have syndromes that appear among the columns of the parity-check matrix H . However, we do realize that there is still a problem here because the last row of our table is not unique. Nonetheless, as our last result will affirm, this method provides a decoding scheme that is as good as any other.

THEOREM 16.17

When we are decoding by coset leaders, if $r \in \mathbf{Z}_2^n$ is a received word and r is decoded as the code word c^* (which we then decode to recapture the message), then $d(c^*, r) \leq d(c, r)$ for all code words c .

Proof: Let x be the coset leader for the coset containing r . Then $r = c^* + x$, or $r + c^* = x$, so $d(c^*, r) = \text{wt}(r + c^*) = \text{wt}(x)$. If c is any code word, then $d(c, r) = \text{wt}(c + r)$, and we have $c + r = c + (c^* + x) = (c + c^*) + x$. Since C is a group code, it follows that $c + c^* \in C$ and so $c + r$ is in the coset $x + C$. Among the elements in the coset $x + C$, the coset leader x is chosen to have minimum weight, so $\text{wt}(c + r) \geq \text{wt}(x)$. Consequently, $d(c^*, r) = \text{wt}(x) \leq \text{wt}(c + r) = d(c, r)$.

16.9 Hamming Matrices

We found the parity-check matrix H helpful in correcting single errors in transmission when (a) H had no column of 0's and (b) no two columns of H were the same. For the matrix

$$H = \begin{bmatrix} 1 & 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 & 1 \end{bmatrix}$$

we find that H satisfies these two conditions and that for the number of rows ($r = 3$) in H we have the maximum number of columns possible. If an additional column is added, H will no longer be useful for correcting single errors.

The generator matrix G associated with H is

$$G = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{bmatrix}.$$

Consequently we have a $(7, 4)$ group code. The encoding function $E: \mathbf{Z}_2^4 \rightarrow \mathbf{Z}_2^7$ encodes four-bit messages into seven-bit code words. We realize that because H is determined by three parity-check equations, we have now maximized the number of bits we can have in

the messages (under our present coding scheme). In addition, the columns of H , read from top to bottom, are the binary equivalents of the integers from 1 to 7.

In general, if we start with r parity-check equations, then the parity-check matrix H can have as many as $2^r - 1$ columns and still be used to correct single errors. Under these circumstances $H = [B | I_r]$, where B is an $r \times (2^r - 1 - r)$ matrix, and $G = [I_m | B^T]$ with $m = 2^r - 1 - r$. The parity-check matrix H associated with a $(2^r - 1, 2^r - 1 - r)$ group code in this way is called a *Hamming matrix*, and the code is referred to as a *Hamming code*.

EXAMPLE 16.27

If $r = 4$, then $2^r - 1 = 15$ and $2^r - 1 - r = 11$. The one (up to a permutation of the columns) possible Hamming matrix H for $r = 4$ is

$$\begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & 1 \end{bmatrix}.$$

Once again, the columns of H contain the binary equivalents of the integers from 1 to 15 ($= 2^4 - 1$).

This matrix H is the parity-check matrix of a Hamming (15, 11) code whose rate is 11/15.

With regard to the rate of these Hamming codes, for all $r \geq 2$, the rate m/n of such a code is given by $m/n = (2^r - 1 - r)/(2^r - 1) = 1 - [r/(2^r - 1)]$. As r increases, $r/(2^r - 1)$ goes to 0 and the rate approaches 1.

We close our discussion on coding theory with one final observation. In Section 16.7 we presented G (and H) in what is called the *systematic form*. Other arrangements of the rows and columns of these matrices are also possible, and these yield *equivalent codes*. (More on this can be found in the text by L. L. Dornhoff and F. E. Hohn [4].) We mention this here because it is often common practice to list the columns in a Hamming matrix of r rows so that the binary representations of the integers from 1 to $2^r - 1$ appear as the columns of H are read from left to right. For the Hamming (7, 4) code, the matrix H mentioned at the start of this section would take the (equivalent) form

$$H_1 = \begin{bmatrix} 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{bmatrix}.$$

Here the identity appears in the first, second, and fourth columns instead of in the last three. Consequently, we would use these components for the parity checks and find that if we send the message $w = w_1 w_2 w_3 w_4$, then the corresponding code word $E(w)$ is $c_1 c_2 w_1 c_3 w_2 w_3 w_4$, where

$$\begin{aligned} c_1 &= w_1 + w_2 && + w_4 \\ c_2 &= w_1 && + w_3 + w_4 \\ c_3 &= && w_2 + w_3 + w_4, \end{aligned}$$

so that $H_1 \cdot (E(w))^T = \mathbf{0}$.

In particular, if we send the message $w = w_1 w_2 w_3 w_4 = 1010$, the corresponding code word would be $E(w) = c = c_1 c_2 w_1 c_3 w_2 w_3 w_4 = c_1 c_2 1 c_3 0 1 0$, where $c_1 = w_1 + w_2 +$

$w_4 = 1 + 0 + 0 = 1$, $c_2 = w_1 + w_3 + w_4 = 1 + 1 + 0 = 0$, and $c_3 = w_2 + w_3 + w_4 = 0 + 1 + 0 = 1$. Then $c = 1011010$ and $H_1 \cdot (E(w))^t = H_1 \cdot (E(1010))^t = H_1 \cdot (1011010)^t = \mathbf{0}$. (Verify this!) So if $c = 1011010$ is sent but $r = 1001010$ is received, we have $H_1 \cdot r^t = H_1 \cdot (1001010)^t = (011)^t$. (Verify this as well!) Since 011 is the binary representation for 3 we know that the error is in position 3 — and this time we did *not* have to examine the columns of H_1 . So using a parity-check matrix of the form H_1 simplifies syndrome decoding. In general, for $c = c_1c_2w_1c_3w_2w_3w_4$, let $r = c + e$, where e is an error pattern of weight 1. And suppose that the 1 in e is in position i , where $1 \leq i \leq 7$. Then the syndrome $H_1 \cdot r^t$ provides the binary representation for i and we can determine c without examining the columns of H_1 . From the third, fifth, sixth, and seventh components of c we can then recapture the original message w .

EXERCISES 16.8 and 16.9

1. Let $E: \mathbb{Z}_2^8 \rightarrow \mathbb{Z}_2^{12}$ be the encoding function for a code C . How many calculations are needed to find the minimum distance between code words? How many calculations are needed if E is a group homomorphism?

2. a) Use Table 16.9 to decode the following received words.

000011	100011	111110	100001
001100	011110	001111	111100

b) Do any of the results in part (a) change if a different set of coset leaders is used?

3. a) Construct a decoding table (with syndromes) for the group code given by the generator matrix

$$G = \begin{bmatrix} 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 1 \end{bmatrix}.$$

b) Use the table from part (a) to decode the following received words.

11110	11101	11011	10100
10011	10101	11111	01100

c) Does this code correct single errors in transmission?

4. Let

$$H = \begin{bmatrix} 1 & 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 & 1 \end{bmatrix}$$

be the parity-check matrix for a Hamming (7, 4) code.

a) Encode the following messages:

1000 1100 1011 1110 1001 1111.

b) Decode the following received words:

1100001 1110111 0010001 0011100.

c) Construct a decoding table consisting of the syndromes and coset leaders for this code.

d) Use the result in part (c) to decode the received words given in part (b).

5. a) What are the dimensions of the generator matrix for the Hamming (63, 57) code? What are the dimensions for the associated parity-check matrix H ?

b) What is the rate of this code?

6. Compare the rates of the Hamming (7, 4) code and the (3, 1) triple-repetition code.

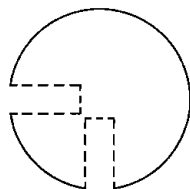
7. a) Let $p = 0.01$ be the probability of incorrect transmission for a binary symmetric channel. If the message 1011 is sent via the Hamming (7, 4) code, what is the probability of correct decoding?

b) Answer part (a) for a 20-bit message sent in five blocks of length 4.

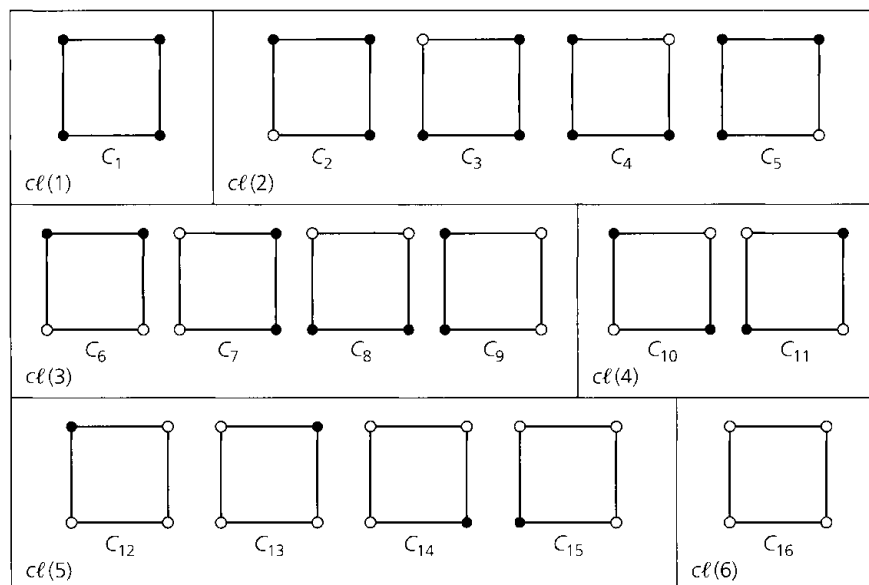
16.10

Counting and Equivalence: Burnside's Theorem

In this section and the next two we shall develop a counting technique known as Polya's Method of Enumeration. Our development will not be very rigorous. Often we shall only state the general results of the theory as seen in the solution of a specific problem. Our first encounter with the type of problem to which this counting technique applies is presented in the following example.

EXAMPLE 16.28**Figure 16.4**

We have a set of sticks, all of the same length and color, and a second set of round plastic disks. Each disk contains two holes, as shown in Fig. 16.4, into which the sticks can be inserted in order to form different shapes, such as a square. (See Fig. 16.5.) If each disk is either red or white, how many distinct squares can we form?

**Figure 16.5**

If the square is considered stationary, then the four disks are located at four distinct locations; a red or white disk is used at each location. Thus there are $2^4 = 16$ different configurations, as shown in Fig. 16.5, where a dark circle indicates a red disk. The configurations have been split into six classes, $cl(1)$, $cl(2)$, \dots , $cl(6)$, according to the number and relative location of the red disks.

Now suppose that the square is not fixed but can be moved about in space. Unless the vertices (disks) are marked somehow, certain configurations in Fig. 16.5 are indistinguishable when we move them about.

To place these notions in a more mathematical setting, we use the nonabelian group of three-dimensional rigid motions of a square to define an equivalence relation on the configurations in Fig. 16.5. Since this group will be used throughout this section and the next two sections, we now give a detailed description of its elements.

In Fig. 16.6 we have the group $G = \{\pi_0, \pi_1, \pi_2, \pi_3, r_1, r_2, r_3, r_4\}$ for the rigid motions of the square in part (a), where we have labeled the vertices with 1, 2, 3, and 4. Parts (b) through (i) of the figure show how each element of G is applied. We have expressed each group element as a permutation of $\{1, 2, 3, 4\}$ and in a new form called a *product of disjoint cycles*. For example, in part (b) we find $\pi_1 = (1234)$. The cycle (1234) indicates that if we start with the square in part (a), after applying π_1 , we find that 1 has moved to the position originally occupied by 2, 2 to that of 3, 3 to that of 4, and 4 to that of 1. In general, if xy appears in a cycle, then x moves to the position originally occupied by y . Also, for a cycle where x and y appear as $(x \dots y)$, y moves to the position originally occupied by x when the motion described by this cycle is applied. Note that $(1234) = (2341) = (3412) = (4123)$. We say that each of these cycles has *length* 4, the number of elements in the cycle. In the case of r_1 in part (f) of the figure, starting with 1 we find that r_1 sends 1 to 4, so we have

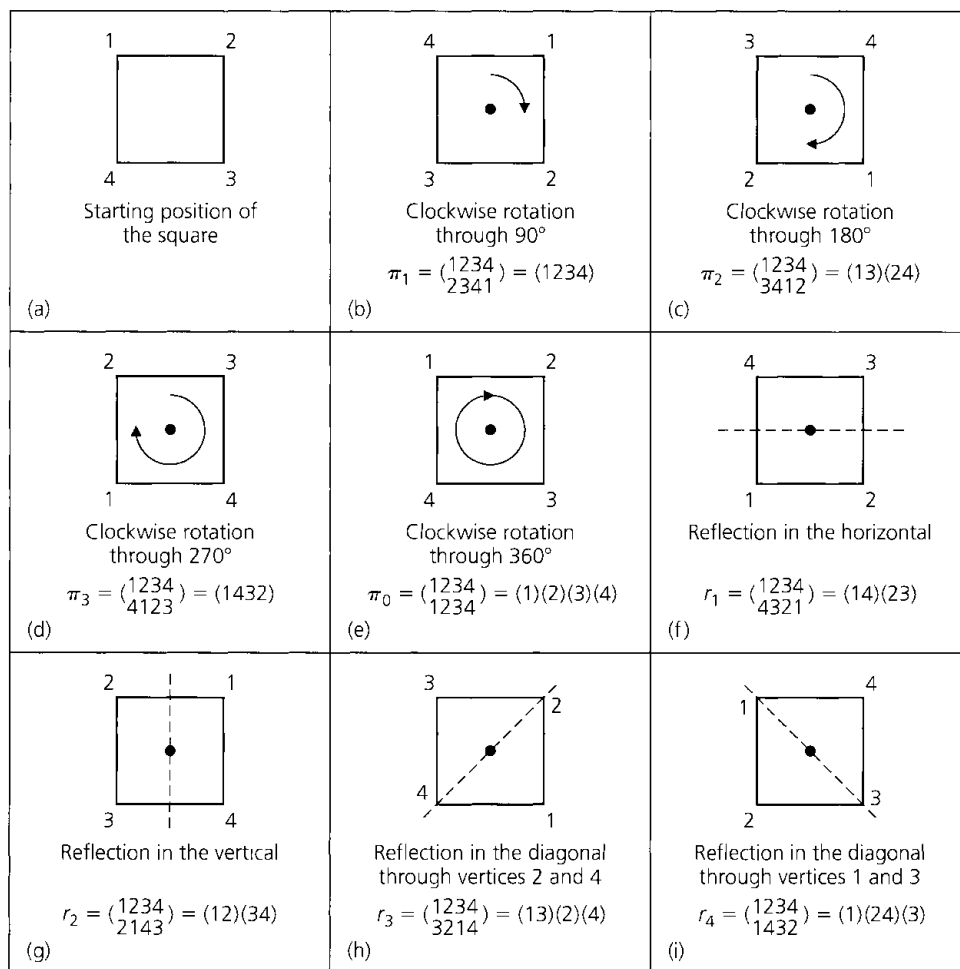


Figure 16.6

(14 . . .) as the start of our first cycle in this *decomposition* of r_1 . However, here r_1 sends 4 to 1, so we have completed a portion — namely, (14) — of the complete decomposition. We then select a vertex that has not yet appeared — for example, vertex 2. Since r_1 sends 2 to 3 and 3, in turn, to 2, we get a second cycle (23). This exhausts all vertices and so $(14)(23) = r_1$, where the cycles (14) and (23) have no vertex in common. Here $(14)(23) = (23)(14) = (23)(41) = (32)(41)$ all provide a representation of r_1 as a product of disjoint cycles, each of length 2. Last, for the group element $r_3 = (13)(2)(4)$, the cycle (2) indicates that 2 is fixed, or *invariant*, under the permutation r_3 . When the number of vertices involved is known, the permutation r_3 may also be written as $r_3 = (13)$, where the missing elements are understood to be fixed. However, we shall write all of the cycles in our decompositions, for this will be useful later in our discussion.

Before continuing with the main discussion concerning the disks and sticks, let us examine some further results on disjoint cycles.

In the group S_6 of all permutations of $\{1, 2, 3, 4, 5, 6\}$, let $\pi = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 2 & 3 & 1 & 4 & 6 & 5 \end{pmatrix}$. As a product of disjoint cycles,

$$\pi = (123)(4)(56) = (56)(4)(123) = (4)(231)(65).$$

If $\sigma \in S_6$, with $\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 2 & 4 & 5 & 1 & 6 & 3 \end{pmatrix}$, then

$$\sigma = (124)(356) = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 2 & 4 & 3 & 1 & 5 & 6 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 1 & 2 & 5 & 4 & 6 & 3 \end{pmatrix},$$

so each cycle can be thought of as an element of S_6 .

Finally, if $\alpha = (124)(3)(56)$ and $\beta = (13)(245)(6)$ are elements of S_6 , then

$$\alpha\beta = (124)(3)(56)(13)(245)(6) = (143)(256),$$

whereas

$$\beta\alpha = (13)(245)(6)(124)(3)(56) = (132)(465).$$

Returning to the 16 configurations, or colorings, in Fig. 16.5, we now examine how each element in the group G , in Fig. 16.6, acts upon these configurations. For example, $\pi_1 = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 1 \end{pmatrix}$ permutes the numbers $\{1, 2, 3, 4\}$ according to a 90° clockwise rotation for the square in Fig. 16.6(a), yielding the result in Fig. 16.6(b). How does such a rotation act on $S = \{C_1, C_2, \dots, C_{16}\}$, our set of colorings? We use π_1^* to distinguish between the 90° clockwise rotation for $\{1, 2, 3, 4\}$ and the same rotation when applied to $S = \{C_1, C_2, \dots, C_{16}\}$. We find that

$$\pi_1^* = \begin{pmatrix} C_1 & C_2 & C_3 & C_4 & C_5 & C_6 & C_7 & C_8 & C_9 & C_{10} & C_{11} & C_{12} & C_{13} & C_{14} & C_{15} & C_{16} \\ C_1 & C_3 & C_4 & C_5 & C_2 & C_7 & C_8 & C_9 & C_6 & C_{11} & C_{10} & C_{13} & C_{14} & C_{15} & C_{12} & C_{16} \end{pmatrix}.$$

As a product of disjoint cycles,

$$\pi_1^* = (C_1)(C_2C_3C_4C_5)(C_6C_7C_8C_9)(C_{10}C_{11})(C_{12}C_{13}C_{14}C_{15})(C_{16}).$$

We note that under the action of π_1^* , no configuration is changed into one that is in another class.

As a second example, consider the reflection r_3 in Fig. 16.6(h). The action of this rigid motion on S is given by

$$\begin{aligned} r_3^* &= \begin{pmatrix} C_1 & C_2 & C_3 & C_4 & C_5 & C_6 & C_7 & C_8 & C_9 & C_{10} & C_{11} & C_{12} & C_{13} & C_{14} & C_{15} & C_{16} \\ C_1 & C_2 & C_5 & C_4 & C_3 & C_7 & C_6 & C_9 & C_8 & C_{10} & C_{11} & C_{14} & C_{13} & C_{12} & C_{15} & C_{16} \end{pmatrix} \\ &= (C_1)(C_2)(C_3C_5)(C_4)(C_6C_7)(C_8C_9)(C_{10})(C_{11})(C_{12}C_{14})(C_{13})(C_{15})(C_{16}). \end{aligned}$$

Once again, no configuration is taken by r_3^* into one that is outside the class that it was in originally.

Using the idea of *the group G acting on the set S* , we define a relation \mathcal{R} on S as follows. For colorings $C_i, C_j \in S$, where $1 \leq i, j \leq 16$, we write $C_i \mathcal{R} C_j$ if there is a permutation $\sigma \in G$ such that $\sigma^*(C_i) = C_j$. That is, as σ^* acts on the 16 configurations in S , C_i is transformed into C_j . This relation \mathcal{R} is an equivalence relation, as we now verify.

- a) (Reflexive Property) For all $C_i \in S$, where $1 \leq i \leq 16$, it follows that $C_i \mathcal{R} C_i$ because G contains the identity permutation. [$\pi_0^*(C_i) = C_i$ for all $1 \leq i \leq 16$.]
- b) (Symmetric Property) If $C_i \mathcal{R} C_j$ for $C_i, C_j \in S$, then $\sigma^*(C_i) = C_j$, for some $\sigma \in G$. G is a group, so $\sigma^{-1} \in G$, and we find that $(\sigma^*)^{-1} = (\sigma^{-1})^*$. (Verify this for two choices of $\sigma \in G$.) Hence $C_i = (\sigma^{-1})^*(C_j)$, and $C_j \mathcal{R} C_i$.

c) (Transitive Property) Let $C_i, C_j, C_k \in S$ with $C_i \mathcal{R} C_j$ and $C_j \mathcal{R} C_k$. Then $C_j = \sigma^*(C_i)$ and $C_k = \tau^*(C_j)$, for some $\sigma, \tau \in G$. By closure in G , $\sigma\tau \in G$, and we find that $(\sigma\tau)^* = \sigma^*\tau^*$, where σ is applied first in $\sigma\tau$ and σ^* first in $\sigma^*\tau^*$. (Verify this for two specific permutations $\sigma, \tau \in G$.) Then $C_k = (\sigma\tau)^*(C_i)$ and \mathcal{R} is transitive. [The reader may have noticed that $C_k = \tau^*(C_j) = \tau^*(\sigma^*(C_i))$ and felt that we should have written $(\sigma\tau)^* = \tau^*\sigma^*$. Once again, there has been a change in the notation for the composite function as we first defined it in Chapter 5. Here we write $\sigma^*\tau^*$ for $(\sigma\tau)^*$, and σ^* is applied first.]

Since \mathcal{R} is an equivalence relation on S , \mathcal{R} partitions S into equivalence classes, which are precisely the classes $c\ell(1), c\ell(2), \dots, c\ell(6)$ of Fig. 16.5. Consequently, there are six nonequivalent configurations under the group action. So among the original 16 colorings only 6 are really distinct.

What has happened in this example generalizes as follows. With S a set of configurations, let G be a group (of permutations) that acts on S . If the relation \mathcal{R} is defined on S by $x \mathcal{R} y$ if $\pi^*(x) = y$, for some $\pi \in G$, then \mathcal{R} is an equivalence relation.

With only red and white disks to connect the sticks, the answer to this example could have been determined from the results in Fig. 16.5. However, we developed quite a bit of mathematical overkill to answer the question. Referring to S as the set of 2-colorings of the vertices of a square, we start to wonder about the role of 2 and seek the number of nonequivalent configurations if the disks come in three or more colors.

In addition, we might notice that the function $f(r, w) = r^4 + r^3w + 2r^2w^2 + rw^3 + w^4$ is the generating function (of two variables) for the number of nonequivalent configurations from S . Here the coefficient of $r^i w^{4-i}$, for $0 \leq i \leq 4$, yields the number of distinct 2-colorings that have i red disks and $(4-i)$ white ones. The coefficient of $r^2 w^2$ is 2 because of the two equivalence classes $c\ell(3)$ and $c\ell(4)$. Finally, $f(1, 1) = 6$, the number of equivalence classes. This generating function $f(r, w)$ is called the *pattern inventory* for the configurations. We shall examine it in more detail in the next two sections.

For now we record an extended version of our present results in the following theorem. (A proof of this result is given on pages 136–137 of C. L. Liu [17].)

THEOREM 16.18

Burnside's Theorem. Let S be a set of configurations on which a finite group G of permutations acts. The number of equivalence classes into which S is partitioned by the action of G is then given by

$$\frac{1}{|G|} \sum_{\pi \in G} \psi(\pi^*),$$

where $\psi(\pi^*)$ is the number of configurations in S fixed under π^* .

To better accept the validity of this theorem, we first examine two examples where we already know the answers.

EXAMPLE 16.29

In Example 16.28 we find that $\psi(\pi_1^*) = 2$ because only C_1 and C_{16} are fixed, or *invariant*, under π_1^* . For $r_3 \in G$, however, $\psi(r_3^*) = 8$ because $C_1, C_2, C_4, C_{10}, C_{11}, C_{13}, C_{15}$, and C_{16} remain fixed under this group action. In like manner $\psi(\pi_2^*) = 4$, $\psi(\pi_3^*) = 2$, $\psi(\pi_0^*) = 16$,

$\psi(r_1^*) = \psi(r_2^*) = 4$, and $\psi(r_4^*) = 8$. With $|G| = 8$, Burnside's Theorem implies that the number of equivalence classes, or nonequivalent configurations, is

$$(1/8)(16 + 2 + 4 + 2 + 4 + 4 + 8 + 8) = (1/8)(48) = 6,$$

the original answer.

EXAMPLE 16.30

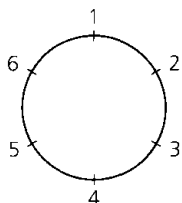


Figure 16.7

In how many ways can six people be arranged around a circular table if two arrangements are considered equivalent when one can be obtained from the other by means of a clockwise rotation through $i \cdot 60^\circ$, for $0 \leq i \leq 5$?

Here the six distinct people are to be placed in six chairs located at a table, as shown in Fig. 16.7. Our permutation group G consists of the clockwise rotations π_i through $i \cdot 60^\circ$, where $0 \leq i \leq 5$. Here reflections are not meaningful. The situation is two-dimensional, for we can rotate the circle (representing the table) only in the plane; the circle never lifts off the plane. The total number of possible configurations is $6!$. We find that $\psi(\pi_0^*) = 6!$ and that $\psi(\pi_i^*) = 0$, for $1 \leq i \leq 5$. (It's impossible to move different people and simultaneously have them stay in a fixed location.)

Consequently, the total number of nonequivalent seating arrangements is

$$\left(\frac{1}{|G|}\right) \sum_{\sigma \in G} \psi(\sigma^*) = \left(\frac{1}{6}\right) (6! + 0 + 0 + 0 + 0 + 0) = 5!,$$

as we found in Example 1.16 of Chapter 1.

We now examine a situation where the power of this theorem is made apparent.

EXAMPLE 16.31

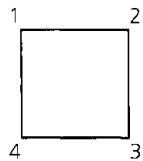


Figure 16.8

In how many ways can the vertices of a square be 3-colored, if the square can be moved about in three dimensions?

Now we have the sticks of Example 16.28, along with red, white, and blue disks. Considering the group in Fig. 16.6, we find the following:

$\psi(\pi_0^*) = 3^4$, because the identity fixes all 81 configurations in the set S of possible configurations.

$\psi(\pi_1^*) = \psi(\pi_3^*) = 3$, for each of π_1^* , π_3^* leaves invariant only those configurations with all vertices the same color.

$\psi(\pi_2^*) = 9$, for π_2^* can fix only those configurations where the opposite (diagonally) vertices have the same color. Consider a square like the one shown in Fig. 16.8. There are three choices for placing a colored disk at vertex 1 and then one choice for matching it at vertex 3. Likewise, there are three choices for colors at vertex 2 and then one for vertex 4. Consequently, there are nine configurations invariant under π_2^* .

$\psi(r_1^*) = \psi(r_2^*) = 9$. In the case of r_1^* , for the square shown in Fig. 16.8 we have three choices for coloring each of the vertices 1 and 2, and then we must match the color of vertex 4 with the color of vertex 1, and the color of vertex 3 with that of vertex 2.

Finally, $\psi(r_3^*) = \psi(r_4^*) = 27$. For r_3^* , we have nine choices for coloring the two vertices at 2 and 4, and three choices for vertex 1. Then there is only one choice for vertex 3 because we must match the color of vertex 1.

By Burnside's Theorem, the number of nonequivalent configurations is

$$(1/8)(3^4 + 3 + 3^2 + 3 + 3^2 + 3^3 + 3^3) = 21.$$

EXERCISES 16.10

- Consider the configurations shown in Fig. 16.5.
 - Determine π_2^* , π_3^* , r_2^* , and r_4^* .
 - Verify that $(\pi_1^{-1})^* = (\pi_1^*)^{-1}$ and $(r_3^{-1})^* = (r_3^*)^{-1}$.
 - Verify that $(\pi_1 r_1)^* = \pi_1^* r_1^*$ and $(\pi_3 r_4)^* = \pi_3^* r_4^*$.
- Express each of the following elements of S_7 as a product of disjoint cycles.

$$\alpha = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 2 & 4 & 6 & 7 & 1 & 5 & 3 \end{pmatrix}$$

$$\beta = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 3 & 6 & 5 & 2 & 1 & 7 & 4 \end{pmatrix}$$

$$\gamma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 2 & 3 & 1 & 7 & 5 & 4 & 6 \end{pmatrix}$$

$$\delta = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 4 & 2 & 7 & 1 & 3 & 6 & 5 \end{pmatrix}$$

- Determine the order of each of the elements in Exercise 2.
 - State a general result about the order of an element in S_n in terms of the lengths of the cycles in its decomposition as a product of disjoint cycles.
- Determine the number of distinct ways one can color the vertices of an equilateral triangle using the colors red and white, if the triangle is free to move in three dimensions.
 - Answer part (a) if the color blue is also available.
- Answer the questions in Exercise 4 for a regular pentagon.
- How many distinct ways are there to paint the *edges* of a square with three different colors?
 - Answer part (a) for the edges of a regular pentagon.
- We make a child's bracelet by symmetrically placing four beads about a circular wire. The colors of the beads are red, white, blue, and green, and there are at least four beads of each color. (a) How many distinct bracelets can we make in this way, if the bracelets can be rotated but not reflected? (b) Answer part (a) if the bracelets can be rotated and reflected.
- A baton is painted with three cylindrical bands of color (not necessarily distinct), with each band of the same length.

- How many distinct paintings can be made if there are three colors of paint available? How many for four colors?
- Answer part (a) for batons with four cylindrical bands.
- Answer part (a) for batons with n cylindrical bands.
- Answer parts (a) and (b) if adjacent cylindrical bands are to have different colors.

- In how many ways can we 2-color the vertices of the configurations shown in Fig. 16.9 if they are free to move in (a) two dimensions? (b) three dimensions?

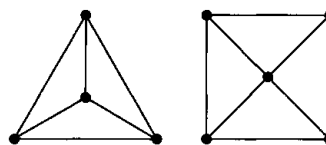


Figure 16.9

- A pyramid has a square base and four faces that are equilateral triangles. If we can move the pyramid about (in three dimensions), how many nonequivalent ways are there to paint its five faces if we have paint of four different colors? How many if the color of the base must be different from the color(s) of the triangular faces?
- In how many ways can we paint the cells of a 3×3 chessboard using red and blue paint? (The back of the chessboard is black.)
 - In how many ways can we construct a 3×3 chessboard by joining (with paste) the edges of nine 1×1 plastic squares that are transparent and tinted red or blue? (There are nine squares of each color available.)
- Answer Exercise 11 for a 4×4 chessboard. [Replace each "nine" in part (b) with "sixteen."]
- In how many ways can we paint the seven (identical) horses on a carousel using black, brown, and white paint?
- Let S be a set of configurations and G a group of permutations that acts on S . If $x \in S$, prove that $\{\pi \in G \mid \pi^*(x) = x\}$ is a subgroup of G (called the *stabilizer* of x).
 - Determine the respective stabilizer subgroups in part (a) for each of the configurations C_7 and C_{15} in Fig. 16.5.

16.11

The Cycle Index

In applying Burnside's Theorem we have been faced with computing $\psi(\pi^*)$ for each $\pi \in G$, where G is a permutation group acting on a set S of configurations. As the number of available colors increases and the configurations get more complex, such computations can get a bit involved. In addition, it seems that if we can determine the number of 2-colorings for a set S of configurations, we should be able to use some of the work in this case to determine the number of 3-colorings, 4-colorings, and so on. We shall now find

some assistance as we return to the solution of Example 16.28. This time more attention will be paid to the representation of each permutation $\pi \in G$ as a product of disjoint cycles. Our results are summarized in Table 16.10.

Table 16.10

Rigid Motions π (Elements of G)	Configurations in S that Are Invariant under π^*	Cycle Structure Representa- tion of π	Inventory of Configurations that Are Invariant under π^*
$\pi_0 = (1)(2)(3)(4)$	2^4 : All configurations in S	x_1^4	$(r + w)^4 = r^4 + 4r^3w + 6r^2w^2 + 4rw^3 + w^4$
$\pi_1 = (1234)$	2 : C_1, C_{16}	x_4	$r^4 + w^4 = r^4 + w^4$
$\pi_2 = (13)(24)$	2^2 : $C_1, C_{10}, C_{11}, C_{16}$	x_2^2	$(r^2 + w^2)^2 = r^4 + 2r^2w^2 + w^4$
$\pi_3 = (1432)$	2 : C_1, C_{16}	x_4	$r^4 + w^4 = r^4 + w^4$
$r_1 = (14)(23)$	2^2 : C_1, C_7, C_9, C_{16}	x_2^2	$(r^2 + w^2)^2 = r^4 + 2r^2w^2 + w^4$
$r_2 = (12)(34)$	2^2 : C_1, C_6, C_8, C_{16}	x_2^2	$(r^2 + w^2)^2 = r^4 + 2r^2w^2 + w^4$
$r_3 = (13)(2)(4)$	2^3 : $C_1, C_2, C_4, C_{10},$ $C_{11}, C_{12}, C_{15}, C_{16}$	$x_2x_1^2$	$(r^2 + w^2)(r + w)^2 = r^4 + 2r^3w + 2r^2w^2 + 2rw^3 + w^4$
$r_4 = (1)(24)(3)$	2^3 : $C_1, C_3, C_5, C_{10},$ $C_{11}, C_{12}, C_{14}, C_{16}$	$x_2x_1^2$	$(r^2 + w^2)(r + w)^2 = r^4 + 2r^3w + 2r^2w^2 + 2rw^3 + w^4$
	$P_G(x_1, x_2, x_3, x_4) =$ $\frac{1}{8}(x_1^4 + 2x_4 + 3x_2^2 + 2x_2x_1^2)$	Complete Inventory	$\left. \begin{array}{l} \\ \end{array} \right\} = 8r^4 + 8r^3w + 16r^2w^2 + 8rw^3 + 8w^4$

For π_0 , the identity of G , we write $\pi_0 = (1)(2)(3)(4)$, a product of four disjoint cycles. We shall represent this cycle structure algebraically by x_1^4 , where x_1 indicates a cycle of length 1. The term x_1^4 is called the *cycle structure representation* of π_0 . Here we interpret “disjoint” as “independent,” in the sense that whatever color is used to paint the vertices in one cycle has no bearing on the choice of color for the vertices in another cycle. As long as all the vertices in a given cycle have the same color, we shall find configurations that are invariant under π_0^* . (Admittedly, this seems like mathematical overkill again, inasmuch as π_0^* fixes all 2-colorings of the square.) In addition, since we can paint the vertices in each cycle either red or white, we have 2^4 configurations, and we find that $(r + w)^4 = r^4 + 4r^3w + 6r^2w^2 + 4rw^3 + w^4$ generates these 16 configurations. For example, from the term $6r^2w^2$ we find that there are six configurations with two red and two white vertices, as found in classes $c\ell(3)$ and $c\ell(4)$ of Fig. 16.5.

Turning to π_1 , we find $\pi_1 = (1234)$, a cycle of length 4. This cycle structure is represented by x_4 , and here there are only two invariant configurations. The fact that the cycle structure for π_1 has only one cycle tells us that for a configuration to be invariant under π_1^* , every vertex in this cycle must be painted the same color. With two colors to choose from, there are only two possible configurations, C_1 and C_{16} . In this case the term $r^4 + w^4$ generates these configurations.

Continuing with r_1 , we have $r_1 = (14)(23)$, a product of two disjoint cycles of length 2; the term x_2^2 represents this cycle structure. For a configuration to be invariant under r_1^* , the vertices at 2 and 3 must be the same color; that is, we have two choices for coloring the

vertices in (23). We also have two choices for coloring the vertices in (14). Consequently, we get 2^2 invariant configurations: $C_1(r^4)$, $C_7(r^2w^2)$, $C_9(r^2w^2)$, and $C_{16}(w^4)$. $[(r^2 + w^2)^2 = r^4 + 2r^2w^2 + w^4]$

Finally, in the case of $r_3 = (13)(2)(4)$, we find that $x_2x_1^2$ indicates its decomposition into one cycle of length 2 and two of length 1. The vertices at 1 and 3 must be painted the same color if the configuration is to be invariant under r_3^* . With three cycles and two choices of color for each cycle, we find 2^3 invariant configurations. They are $C_1(r^4)$, $C_2(r^3w)$, $C_4(r^3w)$, $C_{10}(r^2w^2)$, $C_{11}(r^2w^2)$, $C_{13}(rw^3)$, $C_{15}(rw^3)$, and $C_{16}(w^4)$. These configurations are generated by $(r^2 + w^2)(r + w)^2$, for when we consider the cycle (13) we have two choices: both vertices red (r^2) or both vertices white (w^2). This gives us $r^2 + w^2$. For each single vertex in the two cycles of length 1, $r + w$ provides the choices for each cycle, $(r + w)^2$ the choices for the two. By the independence of choice of colors as we go from one cycle to another, $(r^2 + w^2)(r + w)^2$ generates the 2^3 configurations that are invariant under r_3^* .

Similar arguments provide the information in Table 16.10 for the permutations π_2 , π_3 , r_2 , and r_4 .

At this point we see that what determines the number of configurations that are invariant under π^* , for $\pi \in G$, depends on the cycle structure of π . Within each cycle the same color must be used, but that color can be selected from the two or more choices made available. For r_1 , we had two cycles (of length 2) and 2^2 configurations. If three colors had been available, the number of invariant configurations would have been 3^2 . For m colors, the number is m^2 . Adding these terms for all the cycle structures that arise gives $\sum_{\pi \in G} \psi(\pi^*)$.

We now wish to place more emphasis on cycle structures, so we define the *cycle index*, P_G , for the *group* G (of permutations) as

$$P_G(x_1, x_2, x_3, x_4) = \frac{1}{|G|} \sum_{\pi \in G} (\text{cycle structure representation of } \pi).$$

In this example,

$$P_G(x_1, x_2, x_3, x_4) = (1/8)(x_1^4 + 2x_4 + 3x_2^2 + 2x_2x_1^2).$$

When each occurrence of x_1, x_2, x_3, x_4 is replaced by 2, we find that the number of non-equivalent 2-colorings is equal to

$$P_G(2, 2, 2, 2) = (1/8)(2^4 + 2(2) + 3(2^2) + 2(2)(2^2)) = 6.$$

We summarize our present findings in the following result.

THEOREM 16.19

Let S be a set of configurations that are acted upon by a permutation group G . [G is a subgroup of S_n , the group of all permutations of $\{1, 2, 3, \dots, n\}$, and the cycle index $P_G(x_1, x_2, x_3, \dots, x_n)$ of G is

$$(1/|G|) \sum_{\pi \in G} (\text{cycle structure representation of } \pi).]$$

The number of nonequivalent m -colorings of S is then $P_G(m, m, m, \dots, m)$.

We close this section with an example that uses this theorem.

EXAMPLE 16.32

In how many distinct ways can we 4-color the vertices of a regular hexagon that is free to move in space?

For a regular hexagon there are twelve rigid motions: (a) the six clockwise rotations through 0° , 60° , 120° , 180° , 240° , and 300° ; (b) the three reflections in diagonals through opposite vertices; and (c) the three reflections about lines passing through the midpoints of opposite edges.

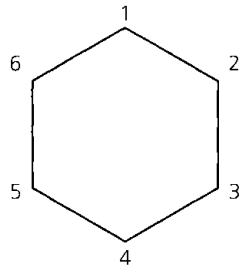
(1) (1)(2)(3)(4)(5)(6) x_1^6		(7) (1)(26)(35)(4) $x_1^2 x_2^2$
(2) (123456) x_6		(8) (13)(46)(2)(5) $x_1^2 x_2^2$
(3) (135)(246) x_3^2		(9) (15)(24)(3)(6) $x_1^2 x_2^2$
(4) (14)(25)(36) x_2^3		(10) (12)(36)(45) x_2^3
(5) (153)(264) x_3^2		(11) (14)(23)(56) x_2^3
(6) (165432) x_6		(12) (16)(25)(34) x_2^3

Figure 16.10

In Fig. 16.10 we have listed each group element as a product of disjoint cycles, together with its cycle structure representation. Here

$$P_G(x_1, x_2, x_3, x_4, x_5, x_6) = (1/12)(x_1^6 + 2x_6 + 2x_3^2 + 4x_2^3 + 3x_1^2 x_2^2),$$

and there are

$$P_G(4, 4, 4, 4, 4, 4) = (1/12)(4^6 + 2(4) + 2(4^2) + 4(4^3) + 3(4^2)(4^2)) = 430$$

nonequivalent 4-colorings of a regular hexagon. (Note: Even though neither x_4 nor x_5 occurs in a cycle structure representation, we may list these variables among the arguments of P_G .)

EXERCISES 16.11

1. In how many ways can we 5-color the vertices of a square that is free to move in (a) two dimensions? (b) three dimensions?
2. Answer Exercise 1 for a regular pentagon.
3. Find the number of nonequivalent 4-colorings of the vertices in the configurations shown in Fig. 16.11 when they are free to move in (a) two dimensions; (b) three dimensions.

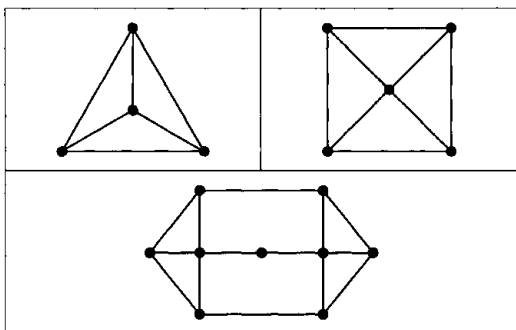


Figure 16.11

4. a) In how many ways can we 3-color the vertices of a regular hexagon that is free to move in space?
b) Give a combinatorial argument to show that for all $m \in \mathbb{Z}^+$, $(m^6 + 2m + 2m^2 + 4m^3 + 3m^4)$ is divisible by 12.
5. a) In how many ways can we 5-color the vertices of a regular hexagon that is free to move in two dimensions?
b) Answer part (a) if the hexagon is free to move in three dimensions.
c) Find two 5-colorings that are equivalent for case (b) but distinct for case (a).
6. In how many distinct ways can we 3-color the edges in the configurations shown in Fig. 16.11 if they are free to move in (a) two dimensions; (b) three dimensions?
7. a) In how many distinct ways can we 3-color the edges of a square that is free to move in three dimensions?
b) In how many distinct ways can we 3-color both the vertices and the edges of such a square?
c) For a square that can move in three dimensions, let k , m , and n denote the number of distinct ways in which we can 3-color its vertices (alone), its edges (alone), and both its vertices and edges, respectively. Does $n = km$? (Give a geometric explanation.)

16.12

The Pattern Inventory: Polya's Method of Enumeration

In this final section we return to Example 16.28 and its continued analysis in Section 16.11. At this time we introduce the pattern inventory and how it is derived from the cycle index.

For $\pi_0 \in G$, every configuration in S is invariant. The cycle structure (representation) for π_0 is given by x_1^4 , where for each cycle of length 1 we have a choice of coloring the vertex in that cycle red (r) or white (w). Using $+$ to represent *exclusive or*, we write $r + w$ to denote the two choices for that vertex (cycle of length 1). With four such cycles, $(r + w)^4$ generates the patterns of the 16 configurations.

In the case of $\pi_1 = (1234)$, x_4 denotes the cycle structure, and here all four vertices must be the same color for the configuration to remain fixed under π_1^* . Consequently, we have all four vertices red or all four vertices white, and we express this algebraically by $r^4 + w^4$.

At this point we notice that for each of the permutations we have considered, the number of factors in the expression used to generate the patterns fixed under a certain permutation equals the number of factors in the cycle structure (representation) of that permutation. Is this just a coincidence?

Continue now with $r_1 = (14)(23)$, whose cycle structure is x_2^2 . For the cycle (14) we must color both of the vertices 1 and 4 either red or white. These choices are represented by $r^2 + w^2$. Since there are two such cycles of length 2, we find that $(r^2 + w^2)^2$ will generate the patterns of the configurations in S fixed under r_1^* . Once again the number of factors in the cycle structure equals the number of factors in the corresponding term used to generate the patterns.

Last, for $r_3 = (13)(24)$, the cycle structure is $x_2 x_1^2 = x_1^2 x_2$. For each of the cycles (2) and (4), $r + w$ represents the choices for each of these vertices, so that $(r + w)^2$ accounts for all four colorings of the pair. The cycle (13) indicates that vertices 1 and 3 must have the same color; $r^2 + w^2$ accounts for the two possibilities. Therefore, $(r + w)^2(r^2 + w^2)$ generates the patterns of the configurations in S fixed under r_3^* , and we find three factors in both the cycle structure and the product $(r + w)^2(r^2 + w^2)$. But even more comes to light here.

Looking at the terms in the cycle structures, we see that, for $1 \leq i \leq n$, the factor x_i in the cycle structure corresponds with the term $r^i + w^i$ in the expression used to generate the patterns.

Continuing with the cycle structures for π_2 , π_3 , r_2 , and r_4 , we find that the *pattern inventory* can be obtained by replacing each x_i in $P_G(x_1, x_2, x_3, x_4)$ with $r^i + w^i$, for $1 \leq i \leq 4$. Consequently,

$$P_G(r + w, r^2 + w^2, r^3 + w^3, r^4 + w^4) = r^4 + r^3w + 2r^2w^2 + rw^3 + w^4.$$

(This result is (1/8)-th of the complete inventory listed in Table 16.10.)

If we had three colors (red, white, and blue), the replacement for x_i would be $r^i + w^i + b^i$, where $1 \leq i \leq 4$.

We generalize these observations in the following theorem.

THEOREM 16.20

Polya's Method of Enumeration. Let S be a set of configurations that are acted upon by a permutation group G , where G is a subgroup of S_n and G has cycle index $P_G(x_1, x_2, \dots, x_n)$.

Then the pattern inventory of nonequivalent m -colorings of S is given by

$$P_G \left(\sum_{i=1}^m c_i, \sum_{i=1}^m c_i^2, \dots, \sum_{i=1}^m c_i^n \right),$$

where c_1, c_2, \dots, c_m denote the m colors that are available.

One important point should be reiterated here before applying Theorem 16.20 — namely, the pattern inventory is another example of a generating function. Having made that point, we now apply this theorem in the following examples.

EXAMPLE 16.33

A child's bracelet is formed by placing three beads — red, white, and blue — on a circular piece of wire. Bracelets are considered equivalent if one can be obtained from the other by a (planar) rotation. Find the pattern inventory for these bracelets.

Here G is the group of rotations of an equilateral triangle, so $G = \{(1)(2)(3), (123), (132)\}$, where 1, 2, 3 denote the vertices of the triangle. Then $P_G(x_1, x_2, x_3) = (1/3) \cdot (x_1^3 + 2x_3)$, and the pattern inventory is given by $(1/3)[(r + w + b)^3 + 2(r^3 + w^3 + b^3)] = (1/3)[3r^3 + 3r^2w + 3r^2b + 3rw^2 + 6rwb + 3rb^2 + 3w^3 + 3w^2b + 3wb^2 + 3b^3] = r^3 + r^2w + r^2b + rw^2 + 2rwb + rb^2 + w^3 + w^2b + wb^2 + b^3$. We interpret this result as follows:

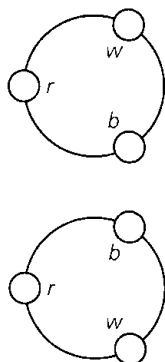


Figure 16.12

- 1) For each summand, other than $2rwb$, the coefficient is 1 because there is only one (distinct) bracelet of that type. That is, there is one bracelet with three red beads (for r^3), one with two red beads and one white bead (for r^2w), and so on for the other seven summands with coefficient 1.
- 2) The summand $2rwb$ has coefficient 2 because there are two nonequivalent bracelets with one red, one white, and one blue bead — as shown in Fig. 16.12.

If the bracelets can also be reflected, then G becomes $\{(1)(2)(3), (123), (132), (1)(23), (2)(13), (3)(12)\}$, and the pattern inventory here is the same as the one above, with one exception. Here we have $rw b$, instead of $2rwb$, because the nonequivalent (for rotations) patterns in Fig. 16.12 become equivalent when reflections are allowed.

EXAMPLE 16.34

Consider the 3-colorings of the configurations in Example 16.28. If the three colors are red, white, and blue, how many nonequivalent configurations have exactly two red vertices?

Given that $P_G(x_1, x_2, x_3, x_4) = (1/8)(x_1^4 + 2x_4 + 3x_2^2 + 2x_2x_1^2)$, the answer is the sum of the coefficients of r^2w^2 , r^2b^2 , and r^2wb in $(1/8)[(r + w + b)^4 + 2(r^4 + w^4 + b^4) + 3(r^2 + w^2 + b^2)^2 + 2(r^2 + w^2 + b^2)(r + w + b)^2]$.

In $(r + w + b)^4$, we find the term $6r^2w^2 + 6r^2b^2 + 12r^2wb$. For $3(r^2 + w^2 + b^2)^2$, we are interested in the term $6r^2w^2 + 6r^2b^2$, whereas $4r^2w^2 + 4r^2b^2 + 4r^2bw$ arises in $2(r^2 + w^2 + b^2)(r + w + b)^2$.

Then $(1/8)[6r^2w^2 + 6r^2b^2 + 12r^2wb + 6r^2w^2 + 6r^2b^2 + 4r^2w^2 + 4r^2b^2 + 4r^2bw] = 2r^2w^2 + 2r^2b^2 + 2r^2bw$, the inventory of the six nonequivalent configurations that contain exactly two red vertices.

Our next example deals with the pattern inventory for the 2-colorings of the vertices of a cube. (The colors are red and white.)

EXAMPLE 16.35

For the cube in Fig. 16.13, we find that its group G of rigid motions consists of the following.

- 1) The identity transformation with cycle structure x_1^8 .
- 2) Rotations through 90° , 180° , and 270° about an axis through the centers of two opposite faces: From Fig. 16.13(a) we have

$$90^\circ \text{ rotation: } (1234)(5678) \quad \text{Cycle structure: } x_4^2$$

$$180^\circ \text{ rotation: } (13)(24)(57)(68) \quad \text{Cycle structure: } x_2^4$$

$$270^\circ \text{ rotation: } (1432)(5876) \quad \text{Cycle structure: } x_4^2$$

Since there are two other pairs of opposite faces, these nine rotations account for the term $3x_2^4 + 6x_4^2$ in the cycle index.

- 3) Rotations through 180° about an axis through the midpoints of two opposite edges: As in Fig. 16.13(b), we have the permutation $(17)(28)(34)(56)$, whose cycle structure is given by x_2^4 . With six pairs of opposite edges, these rotations contribute the term $6x_2^4$ to the cycle index.

- 4) Rotations through 120° and 240° about an axis through two diagonally opposite vertices: From part (c) of the figure we have

$$120^\circ \text{ rotation: } (168)(274)(35) \quad \text{Cycle structure: } x_1^2 x_3^2$$

$$240^\circ \text{ rotation: } (186)(247)(35) \quad \text{Cycle structure: } x_1^2 x_3^2$$

Here there are four such pairs of vertices, and these give rise to the term $8x_1^2 x_3^2$ in the cycle index.

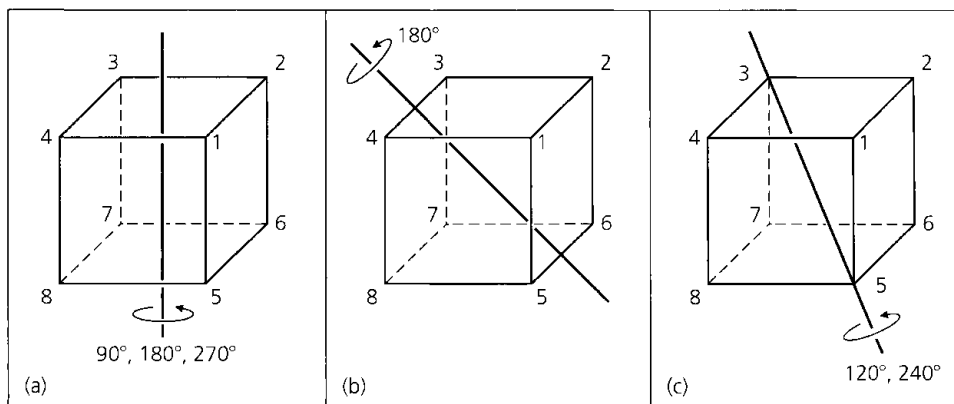


Figure 16.13

Therefore, $P_G(x_1, x_2, \dots, x_8) = (1/24)(x_1^8 + 9x_2^4 + 6x_4^2 + 8x_1^2 x_3^2)$, and the pattern inventory for these configurations is given by the generating function

$$\begin{aligned} f(r, w) &= (1/24)[(r + w)^8 + 9(r^2 + w^2)^4 + 6(r^4 + w^4)^2 + 8(r + w)^2(r^3 + w^3)^2] \\ &= r^8 + r^7w + 3r^6w^2 + 3r^5w^3 + 7r^4w^4 + 3r^3w^5 + 3r^2w^6 + rw^7 + w^8. \end{aligned}$$

Replacing r and w by 1, we find 23 nonequivalent configurations here.

Since Polya's Method of Enumeration was first developed in order to count isomers of organic compounds, we close this section with an application that deals with a certain class

of organic compounds. This is based on an example by C. L. Liu. (See pp. 152–154 of reference [17].)

EXAMPLE 16.36

Here we are concerned with organic molecules of the form shown in Fig. 16.14, where C is a carbon atom and X denotes any of the following components: Br (bromine), H (hydrogen), CH_3 (methyl), or C_2H_5 (ethyl). For example, if each X is replaced by H, the compound CH_4 (methane) results. Figure 16.14 should not be allowed to mislead us. The structure of these organic compounds is three-dimensional. Consequently, we turn to the regular tetrahedron in order to model this structure. We would place the carbon atom at the center of the tetrahedron and then place our selections for X at vertices 1, 2, 3, and 4 as shown in Fig. 16.15.

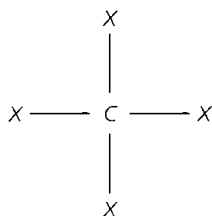


Figure 16.14

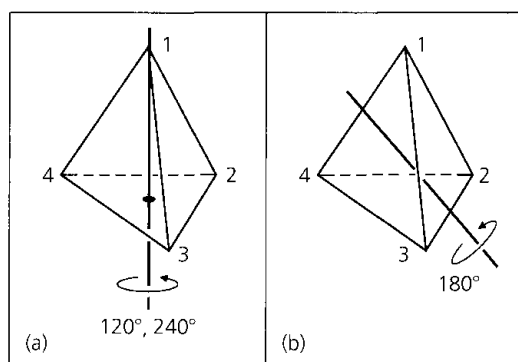


Figure 16.15

The group G acting on these configurations is given as follows:

- 1) The identity transformation $(1)(2)(3)(4)$ with cycle structure x_1^4 .
- 2) Rotations through 120° or 240° about an axis through a vertex and the center of the opposite face: As Fig. 16.15(a) shows, we have

120° rotation: $(1)(243)$ with cycle structure $x_1 x_3$

240° rotation: $(1)(234)$ with cycle structure $x_1 x_3$

By symmetry there are three other pairs of vertices and opposite faces, so these rigid motions account for the term $8x_1 x_3$ in $P_G(x_1, x_2, x_3, x_4)$.

- 3) Rotations of 180° about an axis through the midpoints of two opposite edges: The case shown in part (b) of the figure is given by the permutation $(14)(23)$ whose cycle structure is x_2^2 . With three pairs of opposite edges, we get the term $3x_2^2$ in $P_G(x_1, x_2, x_3, x_4)$.

Hence $P_G(x_1, x_2, x_3, x_4) = (1/12)[x_1^4 + 8x_1 x_3 + 3x_2^2]$ and $P_G(4, 4, 4, 4) = (1/12) \cdot [4^4 + 8(4^2) + 3(4^2)] = 36$, so there are 36 distinct organic compounds that can be formed in this way.

Last, if we wish to know how many of these compounds have exactly two bromine atoms, we let w, x, y , and z represent the “colors” Br, H, CH_3 , and C_2H_5 , respectively, and find the sum of the coefficients of $w^2 x^2$, $w^2 y^2$, $w^2 z^2$, $w^2 xy$, $w^2 xz$, and $w^2 yz$ in the pattern inventory

$$(1/12)[(w + x + y + z)^4 + 8(w + x + y + z)(w^3 + x^3 + y^3 + z^3) + 3(w^2 + x^2 + y^2 + z^2)^2].$$

For $(w + x + y + z)^4$ the relevant term is $6w^2x^2 + 6w^2y^2 + 6w^2z^2 + 12w^2xy + 12w^2xz + 12w^2yz$. The middle summand of the pattern inventory does not give rise to any of the desired configurations, whereas in $3(w^2 + x^2 + y^2 + z^2)^2$ we find $6w^2x^2 + 6w^2y^2 + 6w^2z^2$.

Consequently that part of the pattern inventory for the compounds containing exactly two bromine atoms is

$$(1/12)[12w^2x^2 + 12w^2y^2 + 12w^2z^2 + 12w^2xy + 12w^2xz + 12w^2yz]$$

and there are six such organic compounds.

EXERCISES 16.12

- Find the pattern inventory for the 2-colorings of the edges of a square that is free to move in (i) two dimensions; (ii) three dimensions. (Let the colors be red and white.)
 - Answer part (a) for 3-colorings, where the colors are red, white, and blue.
- If a regular pentagon is free to move in space and we can color its vertices with red, white, and blue paint, how many nonequivalent configurations have exactly three red vertices? How many have two red, one white, and two blue vertices?
- Suppose that in Example 16.35 we 2-color the faces of the cube, which is free to move in space.
 - How many distinct 2-colorings are there for this situation?
 - If the available colors are red and white, determine the pattern inventory.
 - How many nonequivalent colorings have three red and three white faces?
- For the organic compounds in Example 16.36, how many have at least one bromine atom? How many have exactly three hydrogen atoms?
- Find the pattern inventories for the 2-colorings of the vertices in the configurations in Fig. 16.11, when they are free to move in space. (Let the colors be green and gold.)
- In how many ways can the seven (identical) horses on a carousel be painted with black, brown, and white paint in such a way that there are three black, two brown, and two white horses?
 - In how many ways would there be equal numbers of black and brown horses?
 - Give a combinatorial argument to verify that for all $n \in \mathbb{Z}^+$, $n^7 + 6n$ is divisible by 7.
- In how many ways can we paint the eight squares of a 2×4 chessboard, using the colors red and white? (The back of the chessboard is black cardboard.)
 - Find the pattern inventory for the colorings in part (a).
 - How many of the colorings in part (a) have four red and four white squares? How many have six red and two white squares?
- In how many ways can we 2-color the eight regions of the pinwheel shown in Fig. 16.16, using the colors black and gold, if the back of each region remains grey?
 - Answer part (a) for the possible 3-colorings, using black, gold, and blue paints to color the regions.
 - For the colorings in part (b), how many have four black, two gold, and two blue regions?

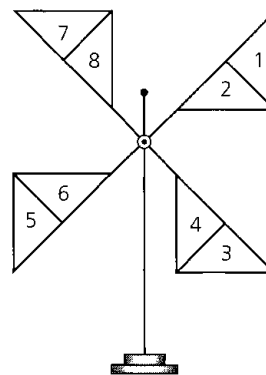


Figure 16.16

- Let $m, n \in \mathbb{Z}^+$ with $n \geq 3$. How many distinct summands appear in the pattern inventory for the m -colorings of the vertices of a regular polygon of n sides?

16.13

Summary and Historical Review

Although the notion of a group of transformations evolved gradually in the study of geometry, the major thrust in the development of the group concept came from the study of polynomial equations.

Methods for solving quadratic equations were known to the ancient Greeks. Then in the sixteenth century, advances were made toward solving cubic and quartic polynomial equations where the coefficients were rational numbers. Continuing with polynomials of fifth and higher degree, both Leonhard Euler (1707–1783) and Joseph-Louis Lagrange (1736–1813) attempted to solve the general quintic. Lagrange realized there had to be a connection between the degree n of a polynomial equation and the permutation group S_n . However, it was Niels Henrik Abel (1802–1829) who finally proved that it was not possible to find a formula for solving the general quintic using only addition, subtraction, multiplication, division, and root extraction. During this same period, the existence of a necessary and sufficient condition for when a polynomial of degree $n \geq 5$ with rational coefficients can be solved by radicals was investigated and solved by the illustrious French mathematician Evariste Galois (1811–1832). Since the work of Galois utilizes the structures of both groups and fields, we shall say more about him in the summary of Chapter 17.



Niels Henrik Abel (1802–1829)

Examining pages 278–280 of J. Stillwell [28], one finds that the group concept, and in fact the actual word “group,” first appears in Galois’ work *Mémoire sur les conditions de résolubilité des équations par radicaux*, published in 1831. Associativity, the group identity, and inverses were consequences of Galois’ assumptions, for he only dealt with a group of permutations of a finite set and his definition of a group required only the closure property. It was Arthur Cayley (1821–1895) (in 1854, in his paper *On the Theory of Groups, as Depending on the Symbolic Equation $\theta^n = 1$*) who first found the need to state the associative property for group elements. The first actual mention of inverses in the definition of a group occurs in the 1883 article *Gruppentheoretischen Studien II* by Walther Franz Anton von Dyck (1856–1934).

The concept of the coset, which we introduced in Section 16.3, was also developed by Evariste Galois (in 1832). The actual term was coined (in 1910) by George Abram Miller (1863–1951).

Following the accomplishments of Galois, group theory affected many areas of mathematics. During the late nineteenth century, for example, the German mathematician Felix Klein (1849–1929), in what has come to be known as the *Erlanger Programm*, attempted to codify all existing geometries according to the group of transformations under which the properties of the geometry were invariant.

Many other mathematicians, such as Augustin-Louis Cauchy (1789–1857), Arthur Cayley (1821–1895), Ludwig Sylow (1832–1918), Richard Dedekind (1831–1916), and Leopold Kronecker (1823–1891), contributed to the further development of certain types of groups. However, it was not until 1900 that lists of defining conditions were given for the general abstract group.

During the twentieth century a great deal of research took place in the attempt to analyze the structure of finite groups. For finite abelian groups, it is known that any such group is isomorphic to a direct product of cyclic groups of prime power order. However, the case of the finite nonabelian groups has turned out to be considerably more complex. Starting with the work of Galois, one finds particular attention paid to a special type of subgroup called a normal subgroup. For any group G , a subgroup H (of G) is called *normal* if, for all $g \in G$ and all $h \in H$, we have $ghg^{-1} \in H$. In an abelian group every subgroup is normal, but this is not the case for nonabelian groups. In every group G , both $\{e\}$ and G are normal subgroups, but if G has no other normal subgroups it is called *simple*. During the past six decades mathematicians have sought and determined all the finite simple groups and examined their role in the structure of all finite groups. Among the prime movers in the classification of the finite simple groups are Professors Walter Feit, John Thompson, Daniel Gorenstein, Michael Aschbacher, and Robert Griess, Jr. For more on the history and impact of this monumental work we refer the reader to the articles by J. A. Gallian [5], A. Gardiner [7], M. Gardner [9], R. Silvestri [27], and, especially, the one by D. Gorenstein [13].

There are many texts one can turn to for further study in the theory of groups. At the introductory level, the texts by J. A. Gallian [6] and V. H. Larney [16] provide further coverage beyond the introduction given in this chapter. The text by I. N. Herstein [15] is an excellent source and includes material on Galois theory.

More on the RSA public-key cryptosystem of Section 16.4 can be found in the references by T. H. Barr [2], P. Garrett [10], and W. Trappe and L. C. Washington [31]. An early description of the system is given in the article by M. Gardner [8], where a message is encrypted using, as the modulus n , the product of a 64-digit prime and a 65-digit prime. The article by G. Taubes [30] relates the effort set forth by Arjen Lenstra, Paul Leyland, Michael Graff, and Derek Atkins, along with 600 volunteers, in factoring n .

The beginnings of algebraic coding theory can be traced to 1941, when Claude Elwood Shannon began his investigations of problems in communications. These problems were prompted by the needs of the war effort. His research resulted in many new ideas and principles that were later published in 1948 in the journal article [26]. As a result of this work, Shannon is acknowledged as the founder of information theory. After this publication, results by M. J. E. Golay [11] and R. W. Hamming [14] soon followed, giving further impetus to research in this area. The 1478 references listed in the bibliography at the end of Volume II of the texts by F. J. MacWilliams and N. J. A. Sloane [18] should convey some idea of the activity in this area between 1950 and 1975.

Our coverage of coding theory followed the development in Chapter 5 of the text by L. L. Dornhoff and F. E. Hohn [4]. The texts by E. F. Assmus, Jr., and J. D. Key [1], S. W. Golomb, R. A. Scholtz, and R. E. Peile [12], V. Pless [20], and S. Roman [24] provide a nice coverage of topics at a fairly intermediate level. More advanced work in coding can be found in the books by F. J. MacWilliams and N. J. A. Sloane [18], S. Roman [25], and A. P. Street and W. D. Wallis [29]. An interesting application on the use of the pigeonhole principle in coding theory is given in Chapter XI of [29].

In Sections 10, 11, and 12 of the chapter, we came upon an enumeration technique whose development is attributed to the Hungarian mathematician George Polya (1887–1985). His article [21] provided the fundamental techniques for counting equivalence classes of chemical isomers, graphs, and trees. (To some extent, the ideas in this work were anticipated by J. H. Redfield [23].) Since then these techniques have been found invaluable for counting problems in such areas as the electronic realizations of Boolean functions. Polya's fundamental theorem was first generalized in the article by N. G. DeBruijn [3], and other extensions of these ideas can be found in the literature. The article by R. C. Read [22] relates the profound influence that Polya's Theorem has had on developments in combinatorial analysis. (The issue of the journal that contains this article also includes several other articles dealing with the life and work of George Polya.)

Our coverage of this topic follows the presentation given in the article by A. Tucker [32]. A more rigorous presentation of this method can be found in Chapter 5 of the text by C. L. Liu [17].

In dealing with Burnside's Theorem we have another instance of an inaccurate attribution. As we learn in the article by P. M. Neumann [19], the result appears in a paper by Georg Frobenius (1848–1917) that was published in 1887, as well as in some of Cauchy's work from 1845.

REFERENCES

1. Assmus, E. F., Jr., and Key, J. D. *Designs and Their Codes*. New York: Cambridge University Press, 1992.
2. Barr, Thomas H. *Invitation to Cryptology*. Upper Saddle River, N. J.: Prentice-Hall, 2002.
3. DeBruijn, Nicolaas Govert. "Polya's Theory of Counting." Chapter 5 in *Applied Combinatorial Mathematics*, ed. by Edwin F. Beckenbach. New York: Wiley, 1964.
4. Dornhoff, Larry L., and Hohn, Franz E. *Applied Modern Algebra*. New York: Macmillan, 1978.
5. Gallian, Joseph A. "The Search for Finite Simple Groups." *Mathematics Magazine* 49, 1976, pp. 163–179.
6. Gallian, Joseph A. *Contemporary Abstract Algebra*, 5th ed. Boston, Mass.: Houghton Mifflin, 2002.
7. Gardiner, Anthony. "Groups of Monsters." *New Scientist*, April 5, 1979, p. 34.
8. Gardner, Martin. "A New Kind of Cipher That Would Take Millions of Years to Break." *Scientific American* (August 1977): pp. 120–124.
9. Gardner, Martin. "The Capture of the Monster: A Mathematical Group with a Ridiculous Number of Elements." *Scientific American* 242 (6), 1980, pp. 20–32.
10. Garrett, Paul. *Making, Breaking Codes: An Introduction to Cryptology*. Upper Saddle River, N. J.: Prentice-Hall, 2001.
11. Golay, Marcel J. E. "Notes on Digital Coding." *Proceedings of the IRE* 37, 1949, p. 657.
12. Golomb, Solomon W., Scholtz, Robert A., and Peile, Robert E. *Basic Concepts in Information Theory and Coding*. New York: Plenum, 1994.
13. Gorenstein, Daniel. "The Enormous Theorem." *Scientific American* 253 (6), 1985, pp. 104–115.
14. Hamming, Richard Wesley. "Error Detecting and Error Correcting Codes." *Bell System Technical Journal* 29, 1950, pp. 147–160.

15. Herstein, Israel Nathan. *Topics in Algebra*, 2nd ed. Lexington, Mass.: Xerox College Publishing, 1975.
16. Larney, Violet H. *Abstract Algebra: A First Course*. Boston: Prindle, Weber & Schmidt, 1975.
17. Liu, C. L. *Introduction to Combinatorial Mathematics*. New York: McGraw-Hill, 1968.
18. MacWilliams, F. Jessie, and Sloane, Neil J. A. *The Theory of Error-Correcting Codes*, Volumes I and II. Amsterdam: North-Holland, 1977.
19. Neumann, Peter M. "A Lemma That Is Not Burnside's." *The Mathematical Scientist*, Vol. 4, 1979, pp. 133–141.
20. Pless, Vera. *Introduction to the Theory of Error-Correcting Codes*, 2nd ed. New York: Wiley, 1989.
21. Polya, George. "Kombinatorische Anzahlbestimmungen für Gruppen, Graphen und Chemische Verbindungen." *Acta Mathematica* 68, 1937, pp. 145–254.
22. Read, R. C. "Polya's Theorem and Its Progeny." *Mathematics Magazine* 60, 1987, pp. 275–282.
23. Redfield, J. Howard. "The Theory of Group Reduced Distributions." *American Journal of Mathematics* 49, 1927, pp. 433–455.
24. Roman, Steven. *Introduction to Coding and Information Theory*. New York: Springer-Verlag, 1997.
25. Roman, Steven. *Coding and Information Theory*. New York: Springer-Verlag, 1992.
26. Shannon, Claude E. "The Mathematical Theory of Communication." *Bell System Technical Journal* 27, 1948, pp. 379–423, 623–656. Reprinted in C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication* (Urbana: University of Illinois Press, 1949).
27. Silvestri, Richard. "Simple Groups of Finite Order." *Archive for the History of Exact Sciences* 20, 1979, pp. 313–356.
28. Stillwell, John. *Mathematics and Its History*. New York: Springer-Verlag, 1989.
29. Street, Anne Penfold, and Wallis, W. D. *Combinatorial Theory: An Introduction*. Winnipeg, Canada: The Charles Babbage Research Center, 1977.
30. Taubes, G. "Small Army of Code-breakers Conquers a 129-digit Giant." *Science* 264, 1994, pp. 776–777.
31. Trappe, Wade, and Washington, Lawrence C. *Introduction to Cryptography with Coding Theory*. Upper Saddle River, N. J.: Prentice-Hall, 2002.
32. Tucker, Alan. "Polya's Enumeration Formula by Example." *Mathematics Magazine* 47, 1974, pp. 248–256.

SUPPLEMENTARY EXERCISES

1. Let $f: G \rightarrow H$ be a group homomorphism with e_H the identity in H . Prove that
 - a) $K = \{x \in G \mid f(x) = e_H\}$ is a subgroup of G . (K is called the *kernel* of the homomorphism.)
 - b) if $g \in G$ and $x \in K$, then $gxg^{-1} \in K$.
2. If G , H , and K are groups and $G = H \times K$, prove that G contains subgroups that are isomorphic to H and K .
3. Let G be a group where $a^2 = e$ for all $a \in G$. Prove that G is abelian.
4. If G is a group of even order, prove that there is an element $a \in G$ with $a \neq e$ and $a = a^{-1}$.
5. Let $f: G \rightarrow H$ be a group homomorphism onto H . If G is a cyclic group, prove that H is also cyclic.
6. a) Consider the group $(\mathbf{Z}_2 \times \mathbf{Z}_2, \oplus)$ where, for $a, b, c, d \in \mathbf{Z}_2$, $(a, b) \oplus (c, d) = (a + c, b + d)$ —the sums $a + c$

and $b + d$ are computed using addition modulo 2. What is the value of $(1, 0) \oplus (0, 1) \oplus (1, 1)$ in this group?

b) Now consider the group $(\mathbf{Z}_2 \times \mathbf{Z}_2 \times \mathbf{Z}_2, \oplus)$ where $(a, b, c) \oplus (d, e, f) = (a + d, b + e, c + f)$. (Here the sums $a + d, b + e, c + f$ are computed using addition modulo 2.) What do we obtain when we add the seven nonzero (or nonidentity) elements of this group?

c) State and prove a generalization that includes the results in parts (a) and (b).

7. Let (G, \circ) be a group where

$$x \circ a \circ y = b \circ a \circ c \Rightarrow x \circ y = b \circ c,$$

for all $a, b, c, x, y \in G$. Prove that (G, \circ) is an abelian group.

8. For $k, n \in \mathbf{Z}^+$ with $n \geq k \geq 1$, let $Q(n, k)$ count the number of permutations $\pi \in S_n$ where any representation of π , as a product of disjoint cycles, contains no cycle of length greater than k . Verify that

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

9. For $k, n \in \mathbb{Z}^+$ where $n \geq 2$ and $1 \leq k \leq n$, let $P(n, k)$ denote the number of permutations $\pi \in S_n$ that have k cycles. [For example, $(1)(23)$ is counted in $P(3, 2)$, $(12)(34)$ is counted in $P(4, 2)$, and $(1)(23)(4)$ is counted in $P(4, 3)$.]

a) Verify that $P(n+1, k) = P(n, k-1) + nP(n, k)$.

b) Determine $\sum_{k=1}^n P(n, k)$.

10. For $n \geq 1$, if $\sigma, \tau \in S_n$, define the distance $d(\sigma, \tau)$ between σ and τ by

$$d(\sigma, \tau) = \max\{|\sigma(i) - \tau(i)| \mid 1 \leq i \leq n\}.$$

a) Prove that the following properties hold for d .

i) $d(\sigma, \tau) \geq 0$ for all $\sigma, \tau \in S_n$

ii) $d(\sigma, \tau) = 0$ if and only if $\sigma = \tau$

iii) $d(\sigma, \tau) = d(\tau, \sigma)$ for all $\sigma, \tau \in S_n$

iv) $d(\rho, \tau) \leq d(\rho, \sigma) + d(\sigma, \tau)$, for all $\rho, \sigma, \tau \in S_n$

b) Let ϵ denote the identity element of S_n (that is, $\epsilon(i) = i$ for all $1 \leq i \leq n$). If $\pi \in S_n$ and $d(\pi, \epsilon) \leq 1$, what can we say about $\pi(n)$?

c) For $n \geq 1$ let a_n count the number of permutations π in S_n , where $d(\pi, \epsilon) \leq 1$. Find and solve a recurrence relation for a_n .

11. Wilson's Theorem [in part (d) of Exercise 19 of Section 16.1] tells us that $(p-1)! \equiv -1 \pmod{p}$, for p a prime.

a) Is the converse of this theorem true or false—that is, if $n \in \mathbb{Z}^+$ and $n \geq 2$, does $(n-1)! \equiv -1 \pmod{n} \Rightarrow n$ is prime?

b) For p an odd prime, prove that

$$2(p-3)! \equiv -1 \pmod{p}.$$

12. In how many ways can Nicole paint the eight regions of the square shown in Fig. 16.17 if

a) five colors are available?

b) she actually uses exactly four of the five available colors?

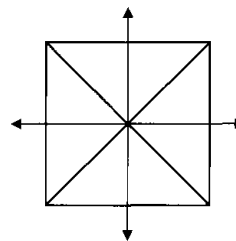


Figure 16.17

7. a) $x \leq y \Rightarrow x + \bar{x} \leq y + \bar{x} \Rightarrow 1 \leq y + \bar{x} \Rightarrow y + \bar{x} = \bar{x} + y = 1$. Conversely, $\bar{x} + y = 1 \Rightarrow x(\bar{x} + y) = x \cdot 1 \Rightarrow x\bar{x} (= 0) + xy = x \Rightarrow xy = x \Rightarrow x \leq y$.
 b) $x \leq \bar{y} \Rightarrow x\bar{y} = x \Rightarrow xy = (x\bar{y})y = x(\bar{y}y) = x \cdot 0 = 0$. Conversely, $xy = 0 \Rightarrow x = x \cdot 1 = x(y + \bar{y}) = xy + x\bar{y} = x\bar{y}$ and $x = x\bar{y} \Rightarrow x \leq \bar{y}$.
9. a) $f(w, x, y, z) = \bar{w}\bar{x} + xy$ b) $g(v, w, x, y, z) = \bar{v}\bar{w}yz + xz + w\bar{y}\bar{z} + \bar{x}\bar{y}\bar{z}$
11. a) $2^{(2^n-1)}$ b) $2^4; 2^{n+1}$
13. a) If $n = 60$, there are 12 divisors, and no Boolean algebra contains 12 elements since 12 is not a power of 2.
 b) If $n = 120$, there are 16 divisors. However, if $x = 4$, then $\bar{x} = 30$ and $x \cdot \bar{x} = \gcd(x, \bar{x}) = \gcd(4, 30) = 2$, which is not the zero element. So the Inverse Laws are not satisfied.

Chapter 16

Groups, Coding Theory, and Polya's Method of Enumeration

Section 16.1 – p. 751

1. a) Yes. The identity is 1 and each element is its own inverse.
 b) No. The set is not closed under addition and there is no identity.
 c) No. The set is not closed under addition.
 d) Yes. The identity is 0; the inverse of $10n$ is $10(-n)$ or $-10n$.
 e) Yes. The identity is 1_A and the inverse of $g: A \rightarrow A$ is $g^{-1}: A \rightarrow A$.
 f) Yes. The identity is 0; the inverse of $a/(2^n)$ is $(-a)/(2^n)$.
3. Subtraction is not an associative (closed) binary operation for \mathbf{Z} . For example, $(3 - 2) - 4 = -3 \neq 5 = 3 - (2 - 4)$.
5. Since $x, y \in \mathbf{Z} \Rightarrow x + y + 1 \in \mathbf{Z}$, the operation is a closed binary operation (or \mathbf{Z} is closed under \circ). For all $w, x, y \in \mathbf{Z}$, $w \circ (x \circ y) = w \circ (x + y + 1) = w + (x + y + 1) + 1 = (w + x + 1) + y + 1 = (w \circ x) \circ y$, so the binary operation is associative. Furthermore, $x \circ y = x + y + 1 = y + x + 1 = y \circ x$, for all $x, y \in \mathbf{Z}$, so \circ is also commutative. If $x \in \mathbf{Z}$, then $x \circ (-1) = x + (-1) + 1 = x = (-1) \circ x$, so -1 is the identity element for \circ . And finally, for each $x \in \mathbf{Z}$, we have $-x - 2 \in \mathbf{Z}$ and $x \circ (-x - 2) = x + (-x - 2) + 1 = -1 = (-x - 2) \circ x$, so $-x - 2$ is the inverse for x under \circ . Consequently, (\mathbf{Z}, \circ) is an abelian group.
7. $U_{20} = \{1, 3, 7, 9, 11, 13, 17, 19\}$ $U_{24} = \{1, 5, 7, 11, 13, 17, 19, 23\}$
9. a) The result follows from Theorem 16.1(b) because both $(a^{-1})^{-1}$ and a are inverses of a^{-1} .
 b)

$$(b^{-1}a^{-1})(ab) = b^{-1}(a^{-1}a)b = b^{-1}(e)b = b^{-1}b = e \text{ and}$$

$$(ab)(b^{-1}a^{-1}) = a(bb^{-1})a^{-1} = a(e)a^{-1} = aa^{-1} = e$$

So $b^{-1}a^{-1}$ is an inverse of ab , and by Theorem 16.1(b), $(ab)^{-1} = b^{-1}a^{-1}$.
11. a) $\{0\}; \{0, 6\}; \{0, 4, 8\}; \{0, 3, 6, 9\}; \{0, 2, 4, 6, 8, 10\}; \mathbf{Z}_{12}$
 b) $\{1\}; \{1, 10\}; \{1, 3, 4, 5, 9\}; \mathbf{Z}_{11}^*$
 c) $\{\pi_0\}; \{\pi_0, \pi_1, \pi_2\}; \{\pi_0, r_1\}; \{\pi_0, r_2\}; \{\pi_0, r_3\}; S_3$
13. a) There are 10: five rotations through $i(72^\circ)$, $0 \leq i \leq 4$, and five reflections about lines containing a vertex and the midpoint of the opposite side.
 b) For a regular n -gon ($n \geq 3$) there are $2n$ rigid motions. There are the n rotations through $i(360^\circ/n)$, $0 \leq i \leq n - 1$. There are n reflections. For n odd, each reflection is about a line through a vertex and the midpoint of the opposite side. For n even, there are $n/2$ reflections about lines through opposite vertices and $n/2$ reflections about lines through the midpoints of opposite sides.
15. Since $eg = ge$ for all $g \in G$, it follows that $e \in H$ and $H \neq \emptyset$. If $x, y \in H$, then $xg = gx$ and $yg = gy$ for all $g \in G$. Consequently, $(xy)g = x(yg) = x(gy) = (xg)y = (gx)y = g(xy)$ for all $g \in G$, and we have $xy \in H$. Finally, for each $x \in H$, $g \in G$, $xg^{-1} = g^{-1}x$. So $(xg^{-1})^{-1} = (g^{-1}x)^{-1}$, or $gx^{-1} = x^{-1}g$, and $x^{-1} \in H$. Therefore, H is a subgroup of G .
17. b) (i) 216

- (ii) $H_1 = \{(x, 0, 0) | x \in \mathbf{Z}_6\}$ is a subgroup of order 6
 $H_2 = \{(x, y, 0) | x, y \in \mathbf{Z}_6, y = 0, 3\}$ is a subgroup of order 12
 $H_3 = \{(x, y, 0) | x, y \in \mathbf{Z}_6\}$ has order 36
 (iii) $-(2, 3, 4) = (4, 3, 2)$; $-(4, 0, 2) = (2, 0, 4)$; $-(5, 1, 2) = (1, 5, 4)$
19. a) $x = 1, x = 4$ b) $x = 1, x = 10$
 c) $x = x^{-1} \Rightarrow x^2 \equiv 1 \pmod{p} \Rightarrow x^2 - 1 \equiv 0 \pmod{p} \Rightarrow (x - 1)(x + 1) \equiv 0 \pmod{p} \Rightarrow$
 $x - 1 \equiv 0 \pmod{p}$ or $x + 1 \equiv 0 \pmod{p} \Rightarrow x \equiv 1 \pmod{p}$ or $x \equiv -1 \equiv p - 1 \pmod{p}$.
 d) The result is true for $p = 2$, since $(2 - 1)! = 1! \equiv -1 \pmod{2}$. For $p \geq 3$, consider the elements $1, 2, \dots, p - 1$ in (\mathbf{Z}_p^*, \cdot) . The elements $2, 3, \dots, p - 2$ yield $(p - 3)/2$ pairs of the form x, x^{-1} . (For example, when $p = 11$ we find that $2, 3, 4, \dots, 9$ yield the four pairs $2, 6$; $3, 4$; $5, 9$; $7, 8$.) Consequently, $(p - 1)! \equiv (1)(1)^{(p-3)/2}(p - 1) \equiv p - 1 \equiv -1 \pmod{p}$.

Section 16.2—p. 756

1. b) $f(a^{-1}) \cdot f(a) = f(a^{-1} \cdot a) = f(e_G) = e_H$ and $f(a) \cdot f(a^{-1}) = f(a \cdot a^{-1}) = f(e_G) = e_H$, so $f(a^{-1})$ is an inverse of $f(a)$. By the uniqueness of inverses (Theorem 16.1b), it follows that $f(a^{-1}) = [f(a)]^{-1}$.
3. $f(0) = (0, 0)$ $f(1) = (1, 1)$ $f(2) = (2, 0)$
 $f(3) = (0, 1)$ $f(4) = (1, 0)$ $f(5) = (2, 1)$
5. $f(4, 6) = -5g_1 + 3g_2$
7. a) $\phi(\pi_0) = 1, \phi(\pi_1) = \phi(\pi_2) = 3, \phi(r_1) = \phi(r_2) = \phi(r_3) = 2$
 b) (See Fig. 16.6) $\phi(\pi_0) = 1, \phi(\pi_1) = \phi(\pi_3) = 4, \phi(\pi_2) = \phi(r_1) = \phi(r_2) = \phi(r_3) = \phi(r_4) = 2$
9. a) The elements of order 10 are 4, 12, 28, and 36.
11. $\mathbf{Z}_5^* = \langle 2 \rangle = \langle 3 \rangle$; $\mathbf{Z}_7^* = \langle 3 \rangle = \langle 5 \rangle$; $\mathbf{Z}_{11}^* = \langle 2 \rangle = \langle 6 \rangle = \langle 7 \rangle = \langle 8 \rangle$
13. Let $(G, +), (H, *), (K, \cdot)$ be the given groups. For all $x, y \in G$, $(g \circ f)(x + y) = g(f(x + y)) = g(f(x) * f(y)) = (g(f(x))) \cdot (g(f(y))) = ((g \circ f)(x)) \cdot ((g \circ f)(y))$, since f, g are homomorphisms. Hence, $g \circ f: G \rightarrow K$ is a group homomorphism.
15. a) $\langle \mathbf{Z}_{12}, + \rangle = \langle 1 \rangle = \langle 5 \rangle = \langle 7 \rangle = \langle 11 \rangle$
 $\langle \mathbf{Z}_{16}, + \rangle = \langle 1 \rangle = \langle 3 \rangle = \langle 5 \rangle = \langle 7 \rangle = \langle 9 \rangle = \langle 11 \rangle = \langle 13 \rangle = \langle 15 \rangle$
 $\langle \mathbf{Z}_{24}, + \rangle = \langle 1 \rangle = \langle 5 \rangle = \langle 7 \rangle = \langle 11 \rangle = \langle 13 \rangle = \langle 17 \rangle = \langle 19 \rangle = \langle 23 \rangle$
 b) Let $G = \langle a^k \rangle$. Since $G = \langle a \rangle$, we have $a = (a^k)^s$ for some $s \in \mathbf{Z}$. Then $a^{1-ks} = e$, so $1 - ks = tn$ since $\phi(a) = n$. $1 - ks = tn \Rightarrow 1 = ks + tn \Rightarrow \gcd(k, n) = 1$. Conversely, let $G = \langle a \rangle$ where $a^k \in G$ and $\gcd(k, n) = 1$. Then $\langle a^k \rangle \subseteq G$. $\gcd(k, n) = 1 \Rightarrow 1 = ks + tn$, for some $s, t \in \mathbf{Z} \Rightarrow a = a^1 = a^{ks+tn} = (a^k)^s (a^n)^t = (a^k)^s (e)^t = (a^k)^s \in \langle a^k \rangle$. Hence $G \subseteq \langle a^k \rangle$. So $G = \langle a^k \rangle$, or a^k generates G .
 c) $\phi(n)$.

Section 16.3—p. 758

1. a) $\left\{ \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 4 & 1 & 2 \end{pmatrix}, \begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 1 & 2 & 3 \end{pmatrix}, \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 2 & 3 & 4 \end{pmatrix} \right\}$
 b) $\begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 1 \end{pmatrix} H = \left\{ \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 4 & 1 & 2 \end{pmatrix}, \begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 1 & 2 & 3 \end{pmatrix}, \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 2 & 3 & 4 \end{pmatrix} \right\}$
 $\begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 2 & 3 & 1 \end{pmatrix} H = \left\{ \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 3 & 4 & 2 \end{pmatrix}, \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 1 & 3 \end{pmatrix}, \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 1 & 2 & 4 \end{pmatrix}, \begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 2 & 3 & 1 \end{pmatrix} \right\}$
 $\begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 2 & 4 & 3 \end{pmatrix} H = \left\{ \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 1 & 4 \end{pmatrix}, \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 4 & 2 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 1 & 3 & 2 \end{pmatrix}, \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 2 & 4 & 3 \end{pmatrix} \right\}$
 $\begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 3 & 2 & 4 \end{pmatrix} H = \left\{ \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 3 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 1 & 4 & 2 \end{pmatrix}, \begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 2 & 1 & 3 \end{pmatrix}, \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 3 & 2 & 4 \end{pmatrix} \right\}$
 $\begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 4 & 2 & 3 \end{pmatrix} H = \left\{ \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 1 & 3 & 4 \end{pmatrix}, \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 2 & 4 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 3 & 1 & 2 \end{pmatrix}, \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 4 & 2 & 3 \end{pmatrix} \right\}$
 $\begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 2 & 3 & 4 \end{pmatrix} H = H$
3. 12
5. From Lagrange's Theorem we know that $|K| = 66 (= 2 \cdot 3 \cdot 11)$ divides $|H|$ and that $|H|$ divides $|G| = 660 (= 2^2 \cdot 3 \cdot 5 \cdot 11)$. Consequently, since $K \neq H$ and $H \neq G$, it follows that $|H|$ is $2(2 \cdot 3 \cdot 11) = 132$ or $5(2 \cdot 3 \cdot 11) = 330$.
7. a) Let $\epsilon = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 2 & 3 & 4 \end{pmatrix}$, $\alpha = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 1 & 4 & 3 \end{pmatrix}$, $\beta = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 4 & 1 & 2 \end{pmatrix}$, and $\delta = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 3 & 2 & 1 \end{pmatrix}$.

\cdot	ϵ	α	β	δ
ϵ	ϵ	α	β	δ
α	α	ϵ	δ	β
β	β	δ	ϵ	α
δ	δ	β	α	ϵ

It follows from Theorem 16.3 that H is a subgroup of G . And since the entries in the accompanying table are symmetric about the diagonal from the upper left to the lower right, we have H an abelian subgroup of G .

b) Since $|G| = 4! = 24$ and $|H| = 4$, there are $24/4 = 6$ left cosets of H in G .

c) Consider the function $f: H \rightarrow \mathbf{Z}_2 \times \mathbf{Z}_2$ defined by

$$f(\epsilon) = (0, 0), \quad f(\alpha) = (1, 0), \quad f(\beta) = (0, 1), \quad f(\delta) = (1, 1).$$

This function f is one-to-one and onto, and for all $x, y \in H$ we find that

$$f(x \cdot y) = f(x) \oplus f(y).$$

Consequently, f is an isomorphism.

(Note: There are other possible answers that can be given here. In fact, there are six possible isomorphisms that one can define here.)

9. a) If H is a proper subgroup of G , then by Lagrange's Theorem, $|H|$ is 2 or p . If $|H| = 2$, then $H = \{e, x\}$ where $x^2 = e$, so $H = \langle x \rangle$. If $|H| = p$, let $y \in H$, $y \neq e$. Then $\mathfrak{e}(y) = p$, so $H = \langle y \rangle$.
- b) Let $x \in G$, $x \neq e$. Then $\mathfrak{e}(x) = p$ or $\mathfrak{e}(x) = p^2$. If $\mathfrak{e}(x) = p$, then $|\langle x \rangle| = p$. If $\mathfrak{e}(x) = p^2$, then $G = \langle x \rangle$ and $\langle x^p \rangle$ is a subgroup of G of order p .
11. b) Let $x \in H \cap K$. If the order of x is r , then r must divide both m and n . Since $\gcd(m, n) = 1$, it follows that $r = 1$, so $x = e$ and $H \cap K = \{e\}$.
13. a) In (\mathbf{Z}_p^*, \cdot) there are $p - 1$ elements, so by Exercise 8, for each $[x] \in (\mathbf{Z}_p^*, \cdot)$, $[x]^{p-1} = [1]$, or $x^{p-1} \equiv 1 \pmod{p}$, or $x^p \equiv x \pmod{p}$. For all $a \in \mathbf{Z}$, if $p \mid a$, then $a \equiv 0 \pmod{p}$ and $a^p \equiv 0 \equiv a \pmod{p}$. If $p \nmid a$, then $a \equiv b \pmod{p}$ where $1 \leq b \leq p - 1$, and $a^p \equiv b^p \equiv b \equiv a \pmod{p}$.
- b) In the group G of units of \mathbf{Z}_n , there are $\phi(n)$ elements. If $a \in \mathbf{Z}$ and $\gcd(a, n) = 1$, then $[a] \in G$ and $[a]^{\phi(n)} = [1]$ or $a^{\phi(n)} \equiv 1 \pmod{n}$.
- c) and d) These results follow from Exercises 6 and 8. They are special cases of Exercise 8.

Section 16.4—p. 761

1. 0462 0170 1809 0462 1809 1981 0305
3. DRIVESAFELYX 5. $p = 157, q = 773$

Section 16.5—p. 765

1. a) $e = 0001001$ b) $r = 1111011$ c) $c = 0101000$
3. a) (i) $D(111101100) = 101$ (ii) $D(000100011) = 000$
(iii) $D(010011111) = 011$
- b) 000000000, 000000001, 100000000 c) 64

Sections 16.6 and 16.7—p. 772

1. $S(101010, 1) = \{101010, 001010, 111010, 100010, 101110, 101000, 101011\}$
 $S(111111, 1) = \{111111, 011111, 101111, 110111, 111011, 111101, 111110\}$
3. a) $|S(x, 1)| = 11$; $|S(x, 2)| = 56$; $|S(x, 3)| = 176$
- b) $|S(x, k)| = 1 + \binom{n}{1} + \binom{n}{2} + \cdots + \binom{n}{k} = \sum_{i=0}^k \binom{n}{i}$
5. a) The minimum distance between code words is 3. The code can detect all errors of weight ≤ 2 or correct all single errors.
- b) The minimum distance between code words is 5. The code can detect all errors of weight ≤ 4 or correct all errors of weight ≤ 2 .

- c) The minimum distance is 2. The code detects all single errors but has no correction capability.
7. a) $C = \{00000, 10110, 01011, 11101\}$. The minimum distance between code words is 3, so the code can detect all errors of weight ≤ 2 or correct all single errors.
- b) $H = \begin{bmatrix} 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 \end{bmatrix}$
- c) (i) 01 (ii) 11 (v) 11 (vi) 10
 For (iii) and (iv) the syndrome is $(111)^T$, which is not a column of H . Assuming a double error, if $(111)^T = (110)^T + (001)^T$, then the decoded received word is 01 [for (iii)] and 10 [for (iv)]. If $(111)^T = (011)^T + (100)^T$, we get 10 [for (iii)] and 01 [for (iv)].
9. $G = [I_8 | A]$ where I_8 is the 8×8 multiplicative identity matrix and A is a column of eight 1's. $H = [A^T | 1] = [11111111 | 1]$.
11. Compare the generator (parity-check) matrix in Exercise 9 with the parity-check (generator) matrix in Exercise 10.

**Sections 16.8 and 16.9—
p. 779**

1. $\binom{256}{2}$; 255
3. a)

Syndrome	Coset Leader			
000	00000	10110	01011	11101
110	10000	00110	11011	01101
011	01000	11110	00011	10101
100	00100	10010	01111	11001
010	00010	10100	01001	11111
001	00001	10111	01010	11100
101	11000	01110	10011	00101
111	01100	11010	00111	10001

(The last two rows are not unique.)

b) Received Word	Code Word	Decoded Message
11110	10110	10
11101	11101	11
11011	01011	01
10100	10110	10
10011	01011	01
10101	11101	11
11111	11101	11
01100	00000	00

5. a) G is 57×63 ; H is 6×63 b) The rate is $\frac{57}{63}$.
7. a) $(0.99)^7 + \binom{7}{1}(0.99)^6(0.01)$ b) $[(0.99)^7 + \binom{7}{1}(0.99)^6(0.01)]^5$

Section 16.10—p. 784

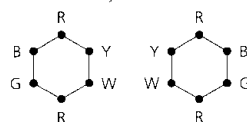
1. a) $\pi_2^* = \begin{pmatrix} C_1 & C_2 & C_3 & C_4 & C_5 & C_6 & C_7 & C_8 & C_9 & C_{10} & C_{11} & C_{12} & C_{13} & C_{14} & C_{15} & C_{16} \\ C_1 & C_4 & C_5 & C_2 & C_3 & C_8 & C_9 & C_6 & C_7 & C_{10} & C_{11} & C_{14} & C_{15} & C_{12} & C_{13} & C_{16} \end{pmatrix}$
- $r_4^* = \begin{pmatrix} C_1 & C_2 & C_3 & C_4 & C_5 & C_6 & C_7 & C_8 & C_9 & C_{10} & C_{11} & C_{12} & C_{13} & C_{14} & C_{15} & C_{16} \\ C_1 & C_4 & C_3 & C_2 & C_5 & C_9 & C_8 & C_7 & C_6 & C_{10} & C_{11} & C_{12} & C_{15} & C_{14} & C_{13} & C_{16} \end{pmatrix}$
- b) $(\pi_1^{-1})^* = \begin{pmatrix} C_1 & C_2 & C_3 & C_4 & C_5 & C_6 & C_7 & C_8 & C_9 & C_{10} & C_{11} & C_{12} & C_{13} & C_{14} & C_{15} & C_{16} \\ C_1 & C_5 & C_2 & C_3 & C_4 & C_9 & C_6 & C_7 & C_8 & C_{11} & C_{10} & C_{15} & C_{12} & C_{13} & C_{14} & C_{16} \end{pmatrix}$
 $= (\pi_1^*)^{-1}$
- c) $\pi_3^* r_4^* = \begin{pmatrix} C_1 & C_2 & C_3 & C_4 & C_5 & C_6 & C_7 & C_8 & C_9 & C_{10} & C_{11} & C_{12} & C_{13} & C_{14} & C_{15} & C_{16} \\ C_1 & C_5 & C_4 & C_3 & C_2 & C_6 & C_9 & C_8 & C_7 & C_{11} & C_{10} & C_{13} & C_{12} & C_{15} & C_{14} & C_{16} \end{pmatrix}$
 $= (\pi_3 r_4)^*$

3. a) $e(\alpha) = 7$; $e(\beta) = 12$; $e(\gamma) = 3$; $e(\delta) = 6$
 b) Let $\alpha \in S_n$, with $\alpha = c_1 c_2 \cdots c_k$, a product of disjoint cycles. Then $e(\alpha)$ is the lcm of $\ell(c_1), \ell(c_2), \dots, \ell(c_k)$, where $\ell(c_i)$ = length of c_i , for $1 \leq i \leq k$.
 5. a) 8 b) 39 7. a) 70 b) 55
 9. Triangular figure: a) 8 b) 8 Square figure: a) 12 b) 12
 11. a) 140 b) 102 13. 315

Section 16.11 – p. 788

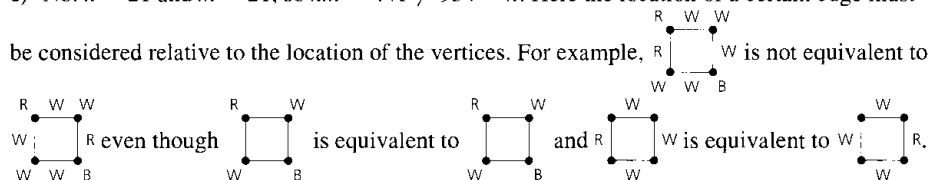
1. a) 165 b) 120
 3. Triangular figure: a) 96 b) 80
 Square figure: a) 280 b) 220
 Hexagonal figure: a) 131,584 b) 70,144
 5. a) 2635 b) 1505

c)



7. a) 21 b) 954

c) No: $k = 21$ and $m = 21$, so $km = 441 \neq 954 = n$. Here the location of a certain edge must be considered relative to the location of the vertices. For example,



Section 16.12 – p. 793

1. a) (i) and (ii) $r^4 + w^4 + r^3w + 2r^2w^2 + rw^3$
 b) (i) $(1/4)[(r+b+w)^4 + 2(r^4 + b^4 + w^4) + (r^2 + b^2 + w^2)^2]$
 (ii) $(1/8)[(r+b+w)^4 + 2(r^4 + b^4 + w^4) + 3(r^2 + b^2 + w^2)^2 + 2(r+b+w)^2(r^2 + b^2 + w^2)]$
 3. a) 10
 b) $(1/24)[(r+w)^6 + 6(r+w)^2(r^4 + w^4) + 3(r+w)^2(r^2 + w^2)^2 + 6(r^2 + w^2)^3 + 8(r^3 + w^3)^2]$
 c) 2
 5. Let g = green and y = gold.
 Triangular figure: $(1/6)[(g+y)^4 + 2(g+y)(g^3 + y^3) + 3(g+y)^2(g^2 + y^2)]$
 Square figure: $(1/8)[(g+y)^5 + 2(g+y)(g^4 + y^4) + 3(g+y)(g^2 + y^2)^2 + 2(g+y)^3(g^2 + y^2)]$
 Hexagonal figure: $(1/4)[(g+y)^9 + 2(g+y)(g^2 + y^2)^4 + (g+y)^5(g^2 + y^2)^2]$
 7. a) 136 b) $(1/2)[(r+w)^8 + (r^2 + w^2)^4]$ c) 38; 16 9. $\binom{m+n-1}{n}$

Supplementary Exercises – p. 797

1. a) Since $f(e_G) = e_H$, it follows that $e_G \in K$ and $K \neq \emptyset$. If $x, y \in K$, then $f(x) = f(y) = e_H$ and $f(xy) = f(x)f(y) = e_H e_H = e_H$, so $xy \in K$. Also, for $x \in K$, $f(x^{-1}) = [f(x)]^{-1} = e_H^{-1} = e_H$, so $x^{-1} \in K$. Hence K is subgroup of G .
 b) If $x \in K$, then $f(x) = e_H$. For all $g \in G$,

$$f(gxg^{-1}) = f(g)f(x)f(g^{-1}) = f(g)e_H f(g^{-1}) = f(g)f(g^{-1}) = f(gg^{-1}) = f(e_G) = e_H.$$

 Hence, for all $x \in K$, $g \in G$, we find that $gxg^{-1} \in K$.
 3. Let $a, b \in G$. Then $a^2b^2 = ee = e = (ab)^2 = abab$. But $a^2b^2 = abab \Rightarrow aabb = abab \Rightarrow ab = ba$, so G is abelian.
 5. Let $G = \langle g \rangle$ and let $h = f(g)$. If $h_1 \in H$, then $h_1 = f(g^n)$ for some $n \in \mathbb{Z}$, since f is onto and G is cyclic. Therefore, $h_1 = f(g^n) = [f(g)]^n = h^n$, and $H = \langle h \rangle$.

7. For all $a, b \in G$,

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

and so it follows that (G, \circ) is an abelian group.

9. a) Consider a permutation σ that is counted in $P(n+1, k)$. If $(n+1)$ is a cycle (of length 1) in σ , then σ (restricted to $\{1, 2, 3, \dots, n\}$) is counted in $P(n, k-1)$. Otherwise, consider each permutation τ that is counted in $P(n, k)$. For each cycle of τ , say $(a_1 a_2 \dots a_r)$, there are r locations in which to place $n+1$ — (1) between a_1 and a_2 ; (2) between a_2 and a_3 ; \dots ; $(r-1)$ between a_{r-1} and a_r ; and (r) between a_r and a_1 . Hence there are n locations, in total, to locate $n+1$ in τ . Consequently, $P(n+1, k) = P(n, k-1) + nP(n, k)$.

b) $\sum_{k=1}^n P(n, k)$ counts all of the permutations in S_n , which has $n!$ elements.

11. a) Suppose that n is composite. We consider two cases.

(1) $n = m \cdot r$, where $1 < m < r < n$: Here $(n-1)! = 1 \cdot 2 \cdot \dots \cdot (m-1) \cdot m \cdot (m+1) \cdot \dots \cdot (r-1) \cdot r \cdot (r+1) \cdot \dots \cdot (n-1) \equiv 0 \pmod{n}$. Hence $(n-1)! \not\equiv -1 \pmod{n}$.

(2) $n = q^2$, where q is a prime: If $(n-1)! \equiv -1 \pmod{n}$ then $0 \equiv q(n-1)! \equiv q(-1) \equiv n-q \not\equiv 0 \pmod{n}$. So in this case we also have $(n-1)! \not\equiv -1 \pmod{n}$.

b) From Wilson's Theorem, when p is an odd prime, we find that

$$-1 \equiv (p-1)! \equiv (p-3)!(p-2)(p-1) \equiv (p-3)!(p^2-3p+2) \equiv 2(p-3)! \pmod{p}.$$

Chapter 17

Finite Fields and Combinatorial Designs

Section 17.1 — p. 806

1. $f(x) + g(x) = 2x^4 + 5x^3 + x^2 + 5$
 $f(x)g(x) = 6x^7 + 2x^6 + 3x^5 + 4x^4 + 2x^3 + x^2 + 4x + 4$
3. $(10)(11)^2$; $(10)(11)^3$; $(10)(11)^4$; $(10)(11)^n$
7. a) and b) $f(x) = (x^2 + 4)(x - 2)(x + 2)$; the roots are ± 2 .
 c) $f(x) = (x + 2i)(x - 2i)(x - 2)(x + 2)$; the roots are $\pm 2, \pm 2i$.
 d) (a) $f(x) = (x^2 - 5)(x^2 + 5)$; there are no rational roots.
 (b) $f(x) = (x - \sqrt{5})(x + \sqrt{5})(x^2 + 5)$; the roots are $\pm \sqrt{5}$.
 (c) $f(x) = (x - \sqrt{5})(x + \sqrt{5})(x - \sqrt{5}i)(x + \sqrt{5}i)$; the roots are $\pm \sqrt{5}, \pm i\sqrt{5}$.
9. a) $f(3) = 8060$ b) $f(1) = 1$ c) $f(-9) = f(2) = 6$
11. 4; 6; $p-1$
13. Let $f(x) = \sum_{i=0}^m a_i x^i$ and $h(x) = \sum_{i=0}^k b_i x^i$, where $a_i \in R$ for $0 \leq i \leq m$, $b_i \in R$ for $0 \leq i \leq k$, and $m \leq k$. Then $f(x) + h(x) = \sum_{i=0}^k (a_i + b_i)x^i$, where $a_{m+1} = a_{m+2} = \dots = a_k = z$, the zero of R , so $G(f(x) + h(x)) = G(\sum_{i=0}^k (a_i + b_i)x^i) = \sum_{i=0}^k g(a_i + b_i)x^i = \sum_{i=0}^k [g(a_i) + g(b_i)]x^i = \sum_{i=0}^k g(a_i)x^i + \sum_{i=0}^k g(b_i)x^i = G(f(x)) + G(h(x))$. Also, $f(x)h(x) = \sum_{i=0}^{m+k} c_i x^i$, where $c_i = a_i b_0 + a_{i-1} b_1 + \dots + a_1 b_{i-1} + a_0 b_i$, and

$$G(f(x)h(x)) = G\left(\sum_{i=0}^{m+k} c_i x^i\right) = \sum_{i=0}^{m+k} g(c_i)x^i.$$

Since $g(c_i) = g(a_i)g(b_0) + g(a_{i-1})g(b_1) + \dots + g(a_1)g(b_{i-1}) + g(a_0)g(b_i)$, it follows that

$$\sum_{i=0}^{m+k} g(c_i)x^i = \left(\sum_{i=0}^m g(a_i)x^i\right)\left(\sum_{i=0}^k g(b_i)x^i\right) = G(f(x)) \cdot G(h(x)).$$

Consequently, $G: R[x] \rightarrow S[x]$ is a ring homomorphism.

15. In $\mathbb{Z}_4[x]$, $(2x+1)(2x+1) = 1$, so $(2x+1)$ is a unit. This does not contradict Exercise 14 because $(\mathbb{Z}_4, +, \cdot)$ is not an integral domain.
17. First note that for $f(x) = a_n x^n + a_{n-1} x^{n-1} + \dots + a_2 x^2 + a_1 x + a_0$, we have $a_n + a_{n-1} + \dots + a_2 + a_1 + a_0 = 0$ if and only if $f(1) = 0$. Since the zero polynomial is in S , the set S is