IV. Groups and Group Actions, 117-210

DOI: 10.3792/euclid/9781429799980-4

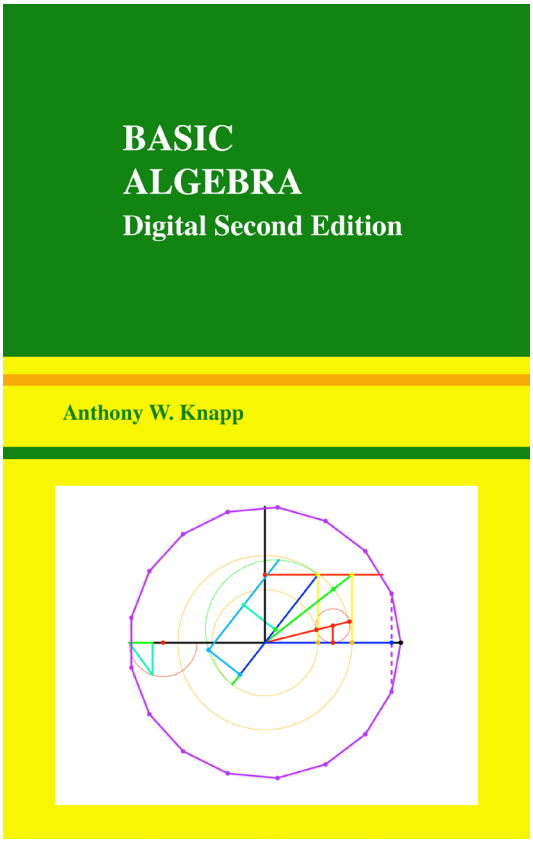
from

Basic Algebra

Digital Second Edition

Anthony W. Knapp

Full Book DOI: 10.3792/euclid/9781429799980
ISBN: 978-1-4297-9998-0



Anthony W. Knapp
81 Upper Sheep Pasture Road
East Setauket, N.Y. 11733–1729, U.S.A.
Email to: aknapp@math.stonybrook.edu
Homepage: www.math.stonybrook.edu/~aknapp

Title: Basic Algebra
Cover: Construction of a regular heptadecagon, the steps shown in color sequence; see page 505.

Mathematics Subject Classification (2010): 15–01, 20–01, 13–01, 12–01, 16–01, 08–01, 18A05, 68P30.

First Edition, ISBN-13 978-0-8176-3248-9

Published by Birkhäuser Boston

Digital Second Edition, not to be sold, no ISBN

Published by the Author

CHAPTER IV

Groups and Group Actions

Abstract. This chapter develops the basics of group theory, with particular attention to the role of group actions of various kinds. The emphasis is on groups in Sections 1–3 and on group actions starting in Section 6. In between is a two-section digression that introduces rings, fields, vector spaces over general fields, and polynomial rings over commutative rings with identity.

Section 1 introduces groups and a number of examples, and it establishes some easy results. Most of the examples arise either from number-theoretic settings or from geometric situations in which some auxiliary space plays a role. The direct product of two groups is discussed briefly so that it can be used in a table of some groups of low order.

Section 2 defines coset spaces, normal subgroups, homomorphisms, quotient groups, and quotient mappings. Lagrange's Theorem is a simple but key result. Another simple but key result is the construction of a homomorphism with domain a quotient group G/H when a given homomorphism is trivial on H. The section concludes with two standard isomorphism theorems.

Section 3 introduces general direct products of groups and direct sums of abelian groups, together with their concrete "external" versions and their universal mapping properties.

Sections 4–5 are a digression to define rings, fields, and ring homomorphisms, and to extend the theories concerning polynomials and vector spaces as presented in Chapters I–II. The immediate purpose of the digression is to make prime fields and the notion of characteristic available for the remainder of the chapter. The definitions of polynomials are extended to allow coefficients from any commutative ring with identity and to allow more than one indeterminate, and universal mapping properties for polynomial rings are proved.

Sections 6–7 introduce group actions. Section 6 gives some geometric examples beyond those in Section 1, it establishes a counting formula concerning orbits and isotropy subgroups, and it develops some structure theory of groups by examining specific group actions on the group and its coset spaces. Section 7 uses a group action by automorphisms to define the semidirect product of two groups. This construction, in combination with results from Sections 5–6, allows one to form several new finite groups of interest.

Section 8 defines simple groups, proves that alternating groups on five or more letters are simple, and then establishes the Jordan–Hölder Theorem concerning the consecutive quotients that arise from composition series.

Section 9 deals with finitely generated abelian groups. It is proved that "rank" is well defined for any finitely generated free abelian group, that a subgroup of a free abelian group of finite rank is always free abelian, and that any finitely generated abelian group is the direct sum of cyclic groups.

Section 10 returns to structure theory for finite groups. It begins with the Sylow Theorems, which produce subgroups of prime-power order, and it gives two sample applications. One of these classifies the groups of order pq, where p and q are distinct primes, and the other provides the information necessary to classify the groups of order 12.

Section 11 introduces the language of "categories" and "functors." The notion of category is a precise version of what is sometimes called a "context" at points in the book before this section,

and some of the "constructions" in the book are examples of "functors." The section treats in this language the notions of "product" and "coproduct," which are abstractions of "direct product" and "direct sum."

1. Groups and Subgroups

Linear algebra and group theory are two foundational subjects for all of algebra, indeed for much of mathematics. Chapters II and III have introduced the basics of linear algebra, and the present chapter introduces the basics of group theory. In this section we give the definition and notation for groups and provide examples that fit with the historical development of the notion of group. Many readers will already be familiar with some group theory, and therefore we can be brief at the start.

A **group** is a nonempty set G with an operation $G \times G \to G$ satisfying the three properties (i), (ii), and (iii) below. In the absence of any other information the operation is usually called **multiplication** and is written $(a, b) \mapsto ab$ with no symbol to indicate the multiplication. The defining properties of a group are

(i) $(ab)c = a(bc)$ for all a, b, c in G (**associative** law),
(ii) there exists an element 1 in G such that $a1 = 1a = a$ for all a in G (existence of **identity**),
(iii) for each a in G, there exists an element a^{-1} in G with $aa^{-1} = a^{-1}a = 1$ (existence of **inverses**).

It is immediate from these properties that

- 1 is unique (since $1' = 1'1 = 1$),
- a^{-1} is unique (since $(a^{-1})' = (a^{-1})'1 = (a^{-1})'(a(a^{-1})) = ((a^{-1})'a)(a^{-1}) = 1(a^{-1}) = (a^{-1})$),
- the existence of a left inverse for each element implies the existence of a right inverse for each element (since $ba = 1$ and $cb = 1$ together imply $c = c(ba) = (cb)a = a$ and hence also $ab = cb = 1$),
- 1 is its own inverse (since $11 = 1$),
- $ax = ay$ implies $x = y$, and $xa = ya$ implies $x = y$ (**cancellation** laws) (since $x = 1x = (a^{-1}a)x = a^{-1}(ax) = a^{-1}(ay) = (a^{-1}a)y = 1y = y$ and since a similar argument proves the second implication).

Problem 2 at the end of Chapter II shows that the associative law extends to products of any finite number of elements of G as follows: parentheses can be inserted in any fashion in such a product, and the value of the product is unchanged; hence any expression $a_1a_2 \cdots a_n$ in G is well defined without the use of parentheses.

The group whose only element is the identity 1 will be denoted by $\{1\}$. It is called the **trivial group**.

We come to other examples in a moment. First we make three more definitions and offer some comments. A **subgroup** H of a group G is a subset containing the identity that is closed under multiplication and inverses. Then H itself is a group because the associativity in G implies associativity in H. The intersection of any nonempty collection of subgroups of G is again a subgroup.

An **isomorphism** of a group G_1 with a group G_2 is a function $\varphi : G_1 \to G_2$ that is one-one onto and satisfies $\varphi(ab) = \varphi(a)\varphi(b)$ for all a and b in G_1. It is immediate that

- $\varphi(1) = 1$ (by taking $a = b = 1$),
- $\varphi(a^{-1}) = \varphi(a)^{-1}$ (by taking $b = a^{-1}$),
- $\varphi^{-1} : G_2 \to G_1$ satisfies $\varphi^{-1}(cd) = \varphi^{-1}(c)\varphi^{-1}(d)$ (by taking $c = \varphi(a)$ and $d = \varphi(b)$ on the right side and then observing that $\varphi\big(\varphi^{-1}(c)\varphi^{-1}(d)\big)$ $= \varphi(ab) = \varphi(a)\varphi(b) = cd = \varphi(\varphi^{-1}(cd))$).

The first and second of these properties show that an isomorphism respects all the structure of a group, not just products. The third property shows that the inverse of an isomorphism is an isomorphism, hence that the relation "is isomorphic to" is symmetric. Since the identity isomorphism exhibits this relation as reflexive and since the use of compositions shows that it is transitive, we see that "is isomorphic to" is an equivalence relation. Common notation for an isomorphism between G_1 and G_2 is $G_1 \cong G_2$; because of the symmetry, one can say that G_1 and G_2 are **isomorphic**.

An **abelian group** is a group G with the additional property

(iv) $ab = ba$ for all a and b in G (**commutative** law).

In an abelian group the operation is sometimes, but by no means always, called **addition** instead of "multiplication." Addition is typically written $(a, b) \mapsto a+b$, and then the identity is usually denoted by 0 and the inverse of a is denoted by $-a$, the **negative** of a. Depending on circumstances, the trivial abelian group may be denoted by $\{0\}$ or 0. Problem 3 at the end of Chapter II shows for an abelian group G with its operation written additively that n-fold sums of elements of G can be written in any order: $a_1 + a_2 + \cdots + a_n = a_{\sigma(1)} + a_{\sigma(2)} + \cdots + a_{\sigma(n)}$ for each permutation σ of $\{1, \ldots, n\}$.

Historically the original examples of groups arose from two distinct sources, and it took a while for the above definition of group to be distilled out as the essence of the matter.

One of the two sources involved number systems and vectors. Here are examples.

EXAMPLES.

(1) Additive groups of familiar number systems. The systems in question are the integers $\mathbb{Z}$, the rational numbers $\mathbb{Q}$, the real numbers $\mathbb{R}$, and the complex

numbers $\mathbb{C}$. In each case the set with its usual operation of addition forms an abelian group. The group properties of $\mathbb{Z}$ under addition are taken as known in advance in this book, as mentioned in Section A3 of the appendix, and the group properties of $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$ under addition are sketched in Sections A3 and A4 of the appendix as part of the development of these number systems.

(2) Multiplicative groups connected with familiar number systems. In the cases of $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$, the nonzero elements form a group under multiplication. These groups are denoted by $\mathbb{Q}^\times$, $\mathbb{R}^\times$, and $\mathbb{C}^\times$. Again the properties of a group for each of them are properties that are sketched during the development of each of these number systems in Sections A3 and A4 of the appendix. With $\mathbb{Z}$, the nonzero integers do not form a group under multiplication, because only the two units, i.e., the divisors $+1$ and -1 of 1, have inverses. The units do form a group, however, under multiplication, and the group of units is denoted by $\mathbb{Z}^\times$.

(3) Vector spaces under addition. Spaces such as $\mathbb{Q}^n$ and $\mathbb{R}^n$ and $\mathbb{C}^n$ provide us with further examples of abelian groups. In fact, the defining properties of addition in a vector space are exactly the defining properties of an abelian group. Thus every vector space provides us with an example of an abelian group if we simply ignore the scalar multiplication.

(4) Integers modulo m, under addition. Another example related to number systems is the additive group of integers modulo a positive integer m. Let us say that an integer n_1 is **congruent modulo** m to an integer n_2 if m divides $n_1 - n_2$. One writes $n_1 \equiv n_2$ or $n_1 \equiv n_2 \bmod m$ or $n_1 = n_2 \bmod m$ for this relation.[1] It is an equivalence relation, and we can write $[n]$ for the equivalence class of n when it is helpful to do so. The division algorithm (Proposition 1.1) tells us that each equivalence class has one and only one member between 0 and $m-1$. Thus there are exactly m equivalence classes, and we know a representative of each. The set of classes will be denoted by[2] $\mathbb{Z}/m\mathbb{Z}$. The point is that $\mathbb{Z}/m\mathbb{Z}$ inherits an abelian-group structure from the abelian-group structure of $\mathbb{Z}$. Namely, we attempt to define

$$[a] + [b] = [a + b].$$

To see that this formula actually defines an operation on $\mathbb{Z}/m\mathbb{Z}$, we need to check that the result is meaningful if the representatives of the classes $[a]$ and $[b]$ are changed. Thus let $[a] = [a']$ and $[b] = [b']$. Then m divides $a - a'$ and $b - b'$, and m must divide the sum $(a - a') + (b - b') = (a + b) - (a' + b')$; consequently $[a + b] = [a' + b']$, and addition is well defined. The same kind of

[1] This notation was anticipated in a remark explaining the classical form of the Chinese Remainder Theorem (Corollary 1.9).

[2] The notation $\mathbb{Z}/(m)$ is an allowable alternative. Some authors, particularly in topology, write $\mathbb{Z}_m$ for this set, but the notation $\mathbb{Z}_m$ can cause confusion since $\mathbb{Z}_p$ is the standard notation for the "p-adic integers" when p is prime. These are defined in Chapter VI of *Advanced Algebra*.

argument shows that the associativity and commutativity of addition in $\mathbb{Z}$ imply associativity and commutativity in $\mathbb{Z}/m\mathbb{Z}$. The identity element is $[0]$, and group inverses (negatives) are given by $-[a] = [-a]$. Therefore $\mathbb{Z}/m\mathbb{Z}$ is an abelian group under addition, and it has m elements. If x and y are members of $\mathbb{Z}/m\mathbb{Z}$, their sum is often denoted by $x + y \bmod m$.

The other source of early examples of groups historically has the members of the group operating as transformations of some auxiliary space. Before abstracting matters, let us consider some concrete examples, ignoring some of the details of verifying the defining properties of a group.

EXAMPLES, CONTINUED.

(5) Permutations. A **permutation** of a nonempty finite set E of n elements is a one-one function from E onto itself. Permutations were introduced in Section I.4. The product of two permutations is just the composition, defined by $(\sigma\tau)(x) = \sigma(\tau(x))$ for x in E, with the symbol $\circ$ for composition dropped. The resulting operation makes the set of permutations of E into a group: we already observed in Section I.4 that composition is associative, and it is plain that the identity permutation may be taken as the group identity and that the inverse function to a permutation is the group inverse. The group is called the **symmetric group** on the n **letters** of E. It has $n!$ members for $n \geq 1$. The notation $\mathfrak{S}_n$ is often used for this group, especially when $E = \{1, \dots, n\}$. Signs ± 1 were defined for permutations in Section I.4, and we say that a permutation is **even** or **odd** according as its sign is $+1$ or -1. The sign of a product is the product of the signs, according to Proposition 1.24, and it follows that the even permutations form a subgroup of $\mathfrak{S}_n$. This subgroup is called the **alternating group** on n letters and is denoted by $\mathfrak{A}_n$. It has $\frac{1}{2}(n!)$ members if $n \geq 2$.

(6) Symmetries of a regular polygon. Imagine a regular polygon in $\mathbb{R}^2$ centered at the origin. The plane-geometry rotations and reflections about the origin that carry the polygon to itself form a group. If the number of sides of the polygon is n, then the group always contains the rotations through all multiples of the angle $2\pi/n$. The rotations themselves form an n-element subgroup of the group of all symmetries. To consider what reflections give symmetries, we distinguish the cases n odd and n even. When n is odd, the reflection in the line that passes through any vertex and bisects the opposite side carries the polygon to itself, and no other reflections have this property. Thus the group of symmetries contains n reflections. When n is even, the reflection in the line passing through any vertex and the opposite vertex carries the polygon to itself, and so does the reflection in the line that bisects a side and also the opposite side. There are $n/2$ reflections of each kind, and hence the group of symmetries again contains n reflections. The group of symmetries thus has $2n$ elements in all cases. It is called the **dihedral**

group D_n. The group D_n is isomorphic to a certain subgroup of the permutation group $\mathfrak{S}_n$. Namely, we number the vertices of the polygon, and we associate to each member of D_n the permutation that moves the vertices the way the member of D_n does.

(7) General linear group. With $\mathbb{F}$ equal to $\mathbb{Q}$ or $\mathbb{R}$ or $\mathbb{C}$, consider any n-dimensional vector space V over $\mathbb{F}$. One possibility is $V = \mathbb{F}^n$, but we do not insist on this choice. Among all one-one functions carrying V onto itself, let G consist of the linear ones. The composition of two linear maps is linear, and the inverse of an invertible function is linear if the given function is linear. The result is a group known as the **general linear group** $\mathrm{GL}(V)$. When $V = \mathbb{F}^n$, we know from Chapter II that we can identify linear maps from $\mathbb{F}^n$ to itself with matrices in $M_{nn}(\mathbb{F})$ and that composition corresponds to matrix multiplication. It follows that the set of all invertible matrices in $M_{nn}(\mathbb{F})$ is a group, which is denoted by $\mathrm{GL}(n, \mathbb{F})$, and that this group is isomorphic to $\mathrm{GL}(\mathbb{F}^n)$. The set $\mathrm{SL}(V)$ or $\mathrm{SL}(n, \mathbb{F})$ of all members of $\mathrm{GL}(V)$ or $\mathrm{GL}(n, \mathbb{F})$ of determinant 1 is a group since the determinant of a product is the product of the determinants; it is called the **special linear group**. The dihedral group D_n is isomorphic to a subgroup of $\mathrm{GL}(2, \mathbb{R})$ since each rotation and reflection of $\mathbb{R}^2$ that fixes the origin is given by the operation of a 2-by-2 matrix.

(8) Orthogonal and unitary groups. If V is a finite-dimensional inner-product space over $\mathbb{R}$ or $\mathbb{C}$, Chapter III referred to the linear maps carrying the space to itself and preserving lengths of vectors as **orthogonal** in the real case and **unitary** in the complex case. Such linear maps are invertible. The condition of preserving lengths of vectors is maintained under composition and inverses, and it follows that the orthogonal or unitary linear maps form a subgroup $\mathrm{O}(V)$ or $\mathrm{U}(V)$ of the general linear group $\mathrm{GL}(V)$. One writes $\mathrm{O}(n)$ for $\mathrm{O}(\mathbb{R}^n)$ and $\mathrm{U}(n)$ for $\mathrm{U}(\mathbb{C}^n)$. The subgroup of members of $\mathrm{O}(V)$ or $\mathrm{O}(n)$ of determinant 1 is called the **rotation group** $\mathrm{SO}(V)$ or $\mathrm{SO}(n)$. The subgroup of members of $\mathrm{U}(V)$ or $\mathrm{U}(n)$ of determinant 1 is called the **special unitary group** $\mathrm{SU}(V)$ or $\mathrm{SU}(n)$.

Before coming to Example 9, let us establish a closure property under the arithmetic operations for certain subsets of $\mathbb{C}$. We are going to use the theories of polynomials as in Chapter I and of vector spaces as in Chapter II with the rationals $\mathbb{Q}$ as the scalars. Fix a complex number θ, and form the result of evaluating at θ every polynomial in one indeterminate with coefficients in $\mathbb{Q}$. The resulting set of complex numbers comes by substituting θ for X in the members of $\mathbb{Q}[X]$, and we denote this subset of $\mathbb{C}$ by $\mathbb{Q}[\theta]$.

Suppose that θ has the property that the set $\{1, \theta, \theta^2, \dots, \theta^n\}$ is linearly dependent over $\mathbb{Q}$ for some integer $n \geq 1$, i.e., has the property that $F_0(\theta) = 0$ for some nonzero member F_0 of $\mathbb{Q}[X]$ of degree $\leq n$. For example, if $\theta = \sqrt{2}$, then the set $\{1, \sqrt{2}, (\sqrt{2})^2\}$ is linearly dependent since $2 - (\sqrt{2})^2 = 0$; if $\theta = e^{2\pi i/5}$,

then $\{1, \theta, \theta^2, \theta^3, \theta^4, \theta^5\}$ is linearly dependent since $1 - \theta^5 = 0$, or alternatively since $1 + \theta + \theta^2 + \theta^3 + \theta^4 = 0$.

Returning to the general θ, we lose no generality if we assume that the polynomial F_0 has degree exactly n. If we divide the equation $F_0(\theta) = 0$ by the leading coefficient, we obtain an equality $\theta^n = G_0(\theta)$, where G_0 is the zero polynomial or is a nonzero polynomial of degree at most $n - 1$. Then $\theta^{n+m} = \theta^m G_0(\theta)$, and we see inductively that every power θ^r with $r \geq n$ is a linear combination of the members of the set $\{1, \theta, \theta^2, \dots, \theta^{n-1}\}$. This set is therefore a spanning set for the vector space $\mathbb{Q}[\theta]$, and we find that $\mathbb{Q}[\theta]$ is finite-dimensional, with dimension at most n. Since every positive integer power of θ lies in $\mathbb{Q}[\theta]$ and since these powers are closed under multiplication, the vector space $\mathbb{Q}[\theta]$ is closed under multiplication. More striking is that $\mathbb{Q}[\theta]$ is closed under division, as is asserted in the following proposition.

Proposition 4.1. Let θ be in $\mathbb{C}$, and suppose for some integer $n \geq 1$ that the set $\{1, \theta, \theta^2, \dots, \theta^n\}$ is linearly dependent over $\mathbb{Q}$. Then the finite-dimensional rational vector space $\mathbb{Q}[\theta]$ is closed under taking reciprocals (of nonzero elements), as well as multiplication, and hence is closed under division.

REMARKS. Under the hypotheses of Proposition 4.1, $\mathbb{Q}[\theta]$ is called an **algebraic number field**,[3] or simply a **number field**, and θ is called an **algebraic number**. The relevant properties of $\mathbb{C}$ that are used in proving the proposition are that $\mathbb{C}$ is closed under the usual arithmetic operations, that these satisfy the usual properties, and that $\mathbb{Q}$ is a subset of $\mathbb{C}$. The deeper closure properties of $\mathbb{C}$ that are developed in Sections A3 and A4 of the appendix play no role.

PROOF. We have seen that $\mathbb{Q}[\theta]$ is closed under multiplication. If x is a nonzero member of $\mathbb{Q}[\theta]$, then all positive powers of x must be in $\mathbb{Q}[\theta]$, and the fact that $\dim \mathbb{Q}[\theta] \leq n$ forces $\{1, x, x^2, \dots, x^n\}$ to be linearly dependent. Therefore there are integers j and k with $0 \leq j < k \leq n$ such that $c_j x^j + c_{j+1} x^{j+1} + \cdots + c_k x^k = 0$ for some rational numbers $c_j, \dots, c_k$ with $c_k \neq 0$. Since x is assumed nonzero, we can discard unnecessary terms and arrange that $c_j \neq 0$. Then

$$1 = x(-c_j^{-1} c_{j+1} - c_j^{-1} c_{j+2} x - c_j^{-1} c_k x^{k-j-1}),$$

and the reciprocal of x has been exhibited as in $\mathbb{Q}[\theta]$. □

EXAMPLES, CONTINUED.

(9) Galois's notion of automorphisms of number fields. Let θ be a complex number as in Proposition 4.1. The subject of Galois theory, whose details will

[3]The definition of "algebraic number field" that is given later in the book is ostensibly more general, but the Theorem of the Primitive Element in Chapter IX will show that it amounts to the same thing as this.

be discussed in Chapter IX and whose full utility will be glimpsed only later, works in an important special case with the "automorphisms" of $\mathbb{Q}[\theta]$ that fix $\mathbb{Q}$. The automorphisms are the one-one functions from $\mathbb{Q}[\theta]$ onto itself that respect addition and multiplication and carry every element of $\mathbb{Q}$ to itself. The identity is such a function, the composition of two such functions is again one, and the inverse of such a function is again one. Therefore the automorphisms of $\mathbb{Q}[\theta]$ form a group under composition. We call this group $\mathrm{Gal}(\mathbb{Q}[\theta]/\mathbb{Q})$. Let us see that it is finite. In fact, if σ is in $\mathrm{Gal}(\mathbb{Q}[\theta]/\mathbb{Q})$, then σ is determined by its effect on θ, since we must have $\sigma(F(\theta)) = F(\sigma(\theta))$ for every F in $\mathbb{Q}[X]$. We know that there is some nonzero polynomial $F_0(X)$ such that $F_0(\theta) = 0$. Applying σ to this equality, we see that $F_0(\sigma(\theta)) = 0$. Therefore $\sigma(\theta)$ has to be a root of F_0. Viewing F_0 as in $\mathbb{C}[X]$, we can apply Corollary 1.14 and see that F_0 has only finitely many complex roots. Therefore there are only finitely many possibilities for σ, and the group $\mathrm{Gal}(\mathbb{Q}[\theta]/\mathbb{Q})$ has to be finite. Galois theory shows that this group gives considerable insight into the structure of $\mathbb{Q}[\theta]$. For example it allows one to derive the Fundamental Theorem of Algebra (Theorem 1.18) just from algebra and the Intermediate Value Theorem (Section A3 of the appendix); it allows one to show the impossibility of certain constructions in plane geometry by straightedge and compass; and it allows one to show that a quintic polynomial with rational coefficients need not have a root that is expressible in terms of rational numbers, arithmetic operations, and the extraction of square roots, cube roots, and so on. We return to these matters in Chapter IX.

Examples 5–9, which all involve auxiliary spaces, fit the pattern that the members of the group are invertible transformations of the auxiliary space and the group operation is composition. This notion will be abstracted in Section 6 and will lead to the notion of a "group action." For now, let us see why we obtained groups in each case. If X is any nonempty set, then the set of invertible functions $f : X \to X$ forms a group under composition, composition being defined by $(fg)(x) = f(g(x))$ with the usual symbol $\circ$ dropped. The associative law is just a matter of unwinding this definition:

$$((fg)h)(x) = (fg)(h(x)) = f(g(h(x))) = f((gh)(x)) = (f(gh))(x).$$

The identity function is the identity of the group, and inverse functions provide the inverse elements in the group.

For our examples, the set X was E in Example 5, $\mathbb{R}^2$ in Example 6, V or $\mathbb{F}^n$ in Example 7, V or $\mathbb{Q}^n$ or $\mathbb{R}^n$ or $\mathbb{C}^n$ in Example 8, and $\mathbb{Q}[\theta]$ in Example 9. All that was needed in each case was to know that our set G of invertible functions from X to itself formed a subgroup of the set of all invertible functions from X to itself. In other words, we had only to check that G contained the identity and was closed under composition and inversion. Associativity was automatic for G because it was valid for the group of all invertible functions from X to itself.

Actually, any group can be realized in the fashion of Examples 5–9. This is the content of the next proposition.

Proposition 4.2 (Cayley's Theorem). Any group G is isomorphic to a subgroup of invertible functions on a set X. The set X can be taken to be G itself. In particular any finite group with n elements is isomorphic to a subgroup of the symmetric group $\mathfrak{S}_n$.

PROOF. Define $X = G$, put $f_a(x) = ax$ for a in G, and let $G' = \{f_a \mid a \in G\}$. To see that G' is a group, we need G' to contain the identity and to be closed under composition and inverses. Since f_1 is the identity, the identity is indeed in G'. Since $f_{ab}(x) = (ab)x = a(bx) = f_a(bx) = f_a(f_b(x)) = (f_a f_b)(x)$, G' is closed under composition. The formula $f_a f_{a^{-1}} = f_1 = f_{a^{-1}} f_a$ then shows that $f_{a^{-1}} = (f_a)^{-1}$ and that G' is closed under inverses. Thus G' is a group.

Define $\varphi : G \to G'$ by $\varphi(a) = f_a$. Certainly φ is onto G', and it is one-one because $\varphi(a) = \varphi(b)$ implies $f_a = f_b$, $f_a(1) = f_b(1)$, and $a = b$. Also, $\varphi(ab) = f_{ab} = f_a f_b = \varphi(a)\varphi(b)$, and hence φ is an isomorphism.

In the case that G is finite with n elements, G is exhibited as isomorphic to a subgroup of the group of permutations of the members of G. Hence it is isomorphic to a subgroup of $\mathfrak{S}_n$. □

It took the better part of a century for mathematicians to sort out that two distinct notions are involved here—that of a group, as defined above, and that of a group action, as will be defined in Section 6. In sorting out these matters, mathematicians realized that it is wise to study the abstract group first and then to study the group in the context of its possible group actions. This does not at all mean ignoring group actions until after the study of groups is complete; indeed, we shall see in Sections 6, 7, and 10 that group actions provide useful tools for the study of abstract groups.

We turn to a discussion of two general group-theoretic notions—cyclic group and the direct product of two or more groups. The second of these notions will be discussed only briefly now; more detail will come in Section 3.

If a is an element of a group, we define a^n for integers $n > 0$ inductively by $a^1 = a$ and $a^n = a^{n-1}a$. Then we can put $a^0 = 1$ and $a^{-n} = (a^{-1})^n$ for $n > 0$. A little checking, which we omit, shows that the ordinary rules of exponents apply: $a^{m+n} = a^m a^n$ and $a^{mn} = (a^m)^n$ for all integers m and n. If the underlying group is abelian and additive notation is being used, these formulas read $(m + n)a = ma + na$ and $(mn)a = n(ma)$.

A **cyclic group** is a group with an element a such that every element is a power of a. The element a is called a **generator** of the group, and the group is said to be **generated** by a.

Proposition 4.3. Each cyclic group G is isomorphic either to the additive group $\mathbb{Z}$ of integers or to the additive group $\mathbb{Z}/m\mathbb{Z}$ of integers modulo m for some positive integer m.

PROOF. If all a^n are distinct, then the rule $a^{m+n} = a^m a^n$ implies that the function $n \mapsto a^n$ is an isomorphism of $\mathbb{Z}$ with G. On the other hand, if $a^k = a^l$ with $k > l$, then $a^{k-l} = 1$ and there exists a positive integer n such that $a^n = 1$. Let m be the least positive integer with $a^m = 1$. For any integers q and r, we have $a^{qm+r} = (a^m)^q a^r = a^r$. Thus the function $\varphi : \mathbb{Z}/m\mathbb{Z} \to G$ given by $\varphi([n]) = a^n$ is well defined, is onto G, and carries sums in $\mathbb{Z}/m\mathbb{Z}$ to products in G. If $0 \le l < k < m$, then $a^k \ne a^l$ since otherwise a^{k-l} would be 1. Hence φ is one-one, and we conclude that $\varphi : \mathbb{Z}/m\mathbb{Z} \to G$ is an isomorphism. □

Let us denote abstract cyclic groups by C_∞ and C_m, the subscript indicating the number of elements. Finite cyclic groups arise in guises other than as $\mathbb{Z}/m\mathbb{Z}$. For example the set of all elements $e^{2\pi ik/m}$ in $\mathbb{C}$, with multiplication as operation, forms a group isomorphic to C_m. So does the set of all rotation matrices $\begin{pmatrix} \cos 2\pi k/m & -\sin 2\pi k/m \\ \sin 2\pi k/m & \cos 2\pi k/m \end{pmatrix}$ with matrix multiplication as operation.

Proposition 4.4. Any subgroup of a cyclic group is cyclic.

REMARK. The proof of Proposition 4.4 exhibits a one-one correspondence between the subgroups of $\mathbb{Z}/m\mathbb{Z}$ and the positive integers k dividing m.

PROOF. Let G be a cyclic group with generator a, and let H be a subgroup. We may assume that $H \ne \{1\}$. Then there exists a positive integer n such that a^n is in H, and we let k be the smallest such positive integer. If n is any integer such that a^n is in H, then Proposition 1.2 produces integers x and y such that $xk + yn = d$, where $d = \mathrm{GCD}(k, n)$. The equation $a^d = (a^k)^x (a^n)^y$ exhibits a^d as in H, and the minimality of k forces $d \ge k$. Since $\mathrm{GCD}(k, n) \le k$, we conclude that $d = k$. Hence k divides n. Consequently H consists of the powers of a^k and is cyclic. □

A notion of the direct product of two groups is definable in the same way as was done with vector spaces in Section II.6, except that a little care is needed in saying how this construction interacts with mappings. As with the corresponding construction for vector spaces, one can define an explicit "external" direct product, and one can recognize a given group as an "internal" direct product, i.e., as isomorphic to an external direct product. We postpone a fuller discussion of direct product, as well as all comments about direct sums and mappings associated with direct sums and direct products, to Section 3.

The **external direct product** $G_1 \times G_2$ of two groups G_1 and G_2 is a group whose underlying set is the set-theoretic product of G_1 and G_2 and whose group

law is $(g_1, g_2)(g_1', g_2') = (g_1g_1', g_2g_2')$. The identity is $(1, 1)$, and the formula for inverses is $(g_1, g_2)^{-1} = (g_1^{-1}, g_2^{-1})$. The two subgroups $G_1 \times \{1\}$ and $\{1\} \times G_2$ of $G_1 \times G_2$ commute with each other.

A group G is the **internal direct product** of two subgroups G_1 and G_2 if the function from the external direct product $G_1 \times G_2$ to G given by $(g_1, g_2) \mapsto g_1g_2$ is an isomorphism of groups. The literal analog of Proposition 2.30, which gave three equivalent definitions of internal direct product[4] of vector spaces, fails here. It is not sufficient that G_1 and G_2 be two subgroups such that $G_1 \cap G_2 = \{1\}$ and every element in G decomposes as a product g_1g_2 with $g_1 \in G_1$ and $g_2 \in G_2$. For example, with $G = \mathfrak{S}_3$, the two subgroups

$$G_1 = \{1, (1\ \ 2)\} \qquad \text{and} \qquad G_2 = \{1, (1\ \ 2\ \ 3), (1\ \ 3\ \ 2)\}$$

have these properties, but G is not isomorphic to $G_1 \times G_2$ because the elements of G_1 do not commute with the elements of G_2.

Proposition 4.5. If G is a group and G_1 and G_2 are subgroups, then the following conditions are equivalent:

(a) G is the internal direct product of G_1 and G_2,
(b) every element in G decomposes uniquely as a product g_1g_2 with $g_1 \in G_1$ and $g_2 \in G_2$, and every member of G_1 commutes with every member of G_2,
(c) $G_1 \cap G_2 = \{1\}$, every element in G decomposes as a product g_1g_2 with $g_1 \in G_1$ and $g_2 \in G_2$, and every member of G_1 commutes with every member of G_2.

PROOF. We have seen that (a) implies (b). If (b) holds and g is in $G_1 \cap G_2$, then the formula $1 = gg^{-1}$ and the uniqueness of the decomposition of 1 as a product together imply that $g = 1$. Hence (c) holds.

If (c) holds, define $\varphi : G_1 \times G_2 \to G$ by $\varphi(g_1, g_2) = g_1g_2$. This map is certainly onto G. To see that it is one-one, suppose that $\varphi(g_1, g_2) = \varphi(g_1', g_2')$. Then $g_1g_2 = g_1'g_2'$ and hence $g_1'^{-1}g_1 = g_2'g_2^{-1}$. Since $G_1 \cap G_2 = \{1\}$, $g_1'^{-1}g_1 = g_2'g_2^{-1} = 1$. Thus $(g_1, g_2) = (g_1', g_2')$, and φ is one-one. Finally the fact that elements of G_1 commute with elements of G_2 implies that $\varphi((g_1, g_2)(g_1', g_2')) = \varphi(g_1g_1', g_2g_2') = g_1g_1'g_2g_2' = g_1g_2g_1'g_2' = \varphi(g_1, g_2)\varphi(g_1', g_2')$. Therefore φ is an isomorphism, and (a) holds. □

Here are two examples of internal direct products of groups. In each let $\mathbb{R}^+$ be the multiplicative group of positive real numbers. The first example is

[4]The direct sum and direct product of two vector spaces were defined to be the same thing in Chapter II.

$\mathbb{R}^\times \cong C_2 \times \mathbb{R}^+$ with C_2 providing the sign. The second example is $\mathbb{C}^\times \cong S^1 \times \mathbb{R}^+$, where S^1 is the multiplicative group of complex numbers of absolute value 1; the isomorphism here is given by the polar-coordinate mapping $(e^{i\theta}, r) \mapsto e^{i\theta} r$.

We conclude this section by giving an example of a group that falls outside the pattern of the examples above and by summarizing what groups we have identified with ≤ 15 elements.

EXAMPLES, CONTINUED.

(10) Groups associated with the quaternions. The set $\mathbb{H}$ of **quaternions** is an object like $\mathbb{R}$ or $\mathbb{C}$ in that it has both an addition/subtraction and a multiplication/division, but $\mathbb{H}$ is unlike $\mathbb{R}$ and $\mathbb{C}$ in that multiplication is not commutative. We give two constructions. In one we start from $\mathbb{R}^4$ with the standard basis vectors written as $1, \mathbf{i}, \mathbf{j}, \mathbf{k}$. The multiplication table for these basis vectors is

$$\begin{aligned} 1\,1 &= 1, & 1\mathbf{i} &= \mathbf{i}, & 1\mathbf{j} &= \mathbf{j}, & 1\mathbf{k} &= \mathbf{k}, \\ \mathbf{i}\,1 &= \mathbf{i}, & \mathbf{i}\,\mathbf{i} &= -1, & \mathbf{i}\,\mathbf{j} &= \mathbf{k}, & \mathbf{i}\,\mathbf{k} &= -\mathbf{j}, \\ \mathbf{j}\,1 &= \mathbf{j}, & \mathbf{j}\,\mathbf{i} &= -\mathbf{k}, & \mathbf{j}\,\mathbf{j} &= -1, & \mathbf{j}\,\mathbf{k} &= \mathbf{i}, \\ \mathbf{k}\,1 &= \mathbf{k}, & \mathbf{k}\,\mathbf{i} &= \mathbf{j}, & \mathbf{k}\,\mathbf{j} &= -\mathbf{i}, & \mathbf{k}\,\mathbf{k} &= -1, \end{aligned}$$

and the multiplication is extended to general elements by the usual distributive laws. The multiplicative identity is 1, and multiplicative inverses of nonzero elements are given by

$$(a1 + b\mathbf{i} + c\mathbf{j} + \mathbf{k})^{-1} = s^{-1}a1 - s^{-1}b\mathbf{i} - s^{-1}c\mathbf{j} - s^{-1}d\mathbf{k}$$

with $s = \sqrt{a^2 + b^2 + c^2 + d^2}$. Since $\mathbf{i}\mathbf{j} = \mathbf{k}$ while $\mathbf{j}\mathbf{i} = -\mathbf{k}$, multiplication is not commutative. What takes work to see is that multiplication is associative. To see this, we give another construction, using $M_{22}(\mathbb{C})$. Within $M_{22}(\mathbb{C})$, take

$$1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad \mathbf{i} = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}, \qquad \mathbf{j} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \qquad \mathbf{k} = \begin{pmatrix} 0 & -i \\ -i & 0 \end{pmatrix},$$

and define $\mathbb{H}$ to be the linear span, with real coefficients, of these matrices. The operations are the usual matrix addition and multiplication. Then multiplication is associative, and we readily verify the multiplication table for $1, \mathbf{i}, \mathbf{j}, \mathbf{k}$. A little computation verifies also the formula for multiplicative inverses. The set $\mathbb{H}^\times$ of nonzero elements forms a group under multiplication, and it is isomorphic to $\mathbb{R}^+ \times \mathrm{SU}(2)$, where

$$\mathrm{SU}(2) = \left\{ \begin{pmatrix} \alpha & \beta \\ -\bar{\beta} & \bar{\alpha} \end{pmatrix} \;\middle|\; |\alpha|^2 + |\beta|^2 = 1 \right\}$$

is the 2-by-2 special unitary group defined in Example 8. Of interest for our current purposes is the 8-element subgroup $\pm 1, \pm\mathbf{i}, \pm\mathbf{j}, \pm\mathbf{k}$, which is called the **quaternion group** and will be denoted by H_8.

The **order** of a finite group is the number of elements in the group. Let us list some of the groups we have discussed that have order at most 15:

1	C_1
2	C_2
3	C_3
4	$C_4,\ C_2 \times C_2$
5	C_5
6	$C_6,\ D_3$
7	C_7
8	$C_8,\ C_4 \times C_2,\ C_2 \times C_2 \times C_2,\ D_4,\ H_8$
9	$C_9,\ C_3 \times C_3$
10	$C_{10},\ D_5$
11	C_{11}
12	$C_{12},\ C_6 \times C_2,\ D_6,\ \mathfrak{A}_4$
13	C_{13}
14	$C_{14},\ D_7$
15	C_{15}

No two groups in the above table are isomorphic, as one readily checks by counting elements of each "order" in the sense of the next section. We shall see in Section 10 and in the problems at the end of the chapter that the above table is complete through order 15 except for one group of order 12. Some groups that we have discussed have been omitted from the above table because of isomorphisms with the groups above. For example, $\mathfrak{S}_2 \cong C_2$, $\mathfrak{A}_3 \cong C_3$, $C_3 \times C_2 \cong C_6$, $\mathfrak{S}_3 \cong D_3$, $C_5 \times C_2 \cong C_{10}$, $C_4 \times C_3 \cong C_{12}$, $D_3 \times C_2 \cong D_6$, $C_7 \times C_2 \cong C_{14}$, and $C_5 \times C_3 \cong C_{15}$.

2. Quotient Spaces and Homomorphisms

Let G be a group, and let H be a subgroup. For purposes of this paragraph, say that g_1 in G is equivalent to g_2 in G if $g_1 = g_2 h$ for some h in H. The relation "equivalent" is an equivalence relation: it is reflexive because 1 is in H, it is symmetric since H is closed under inverses, and it is transitive since H is closed under products. The equivalence classes are called **left cosets** of H in G. The left coset containing an element g of G is the set $gH = \{gh \mid h \in H\}$.

EXAMPLES.

(1) When $G = \mathbb{Z}$ and $H = m\mathbb{Z}$, the left cosets are the sets $r + m\mathbb{Z}$, i.e., the sets $\{x \in \mathbb{Z} \mid x \equiv r \bmod m\}$ for the various values of r.

(2) When $G = \mathfrak{S}_3$ and $H = \{(1),\ (1\ \ 3)\}$, there are three left cosets: H, $(1\ \ 2)H = \{(1\ \ 2),\ (1\ \ 3\ \ 2)\}$, and $(2\ \ 3)H = \{(2\ \ 3),\ (1\ \ 2\ \ 3)\}$.

Similarly one can define the **right cosets** Hg of H in G. When G is nonabelian, these need not coincide with the left cosets; in Example 2 above with $G = \mathfrak{S}_3$ and $H = \{(1),\ (1\ \ 3)\}$, the right coset $H(1\ \ 2) = \{(1\ \ 2),\ (1\ \ 2\ \ 3)\}$ is not a left coset.

Lemma 4.6. If H is a subgroup of the group G, then any two left cosets of H in G have the same cardinality, namely card H.

REMARKS. We shall be especially interested in the case that card H is finite, and then we write $|H| = \text{card } H$ for the number of elements in H.

PROOF. If g_1H and g_2H are given, then the map $g \mapsto g_2g_1^{-1}g$ is one-one on G and carries g_1H onto g_2H. Hence g_1H and g_2H have the same cardinality. Taking $g_1 = 1$, we see that this common cardinality is card H. □

We write G/H for the set $\{gH\}$ of all left cosets of H in G, calling it the **quotient space** or **left-coset space** of G by H. The set $\{Hg\}$ of right cosets is denoted by $H\backslash G$.

Theorem 4.7 (Lagrange's Theorem). If G is a finite group, then $|G| = |G/H|\,|H|$. Consequently the order of any subgroup of G divides the order of G.

REMARK. Using the formula in Theorem 4.7 three times yields the conclusion that if H and K are subgroups of a finite group G with $K \subseteq H$, then $|G/K| = |G/H|\,|H/K|$.

PROOF. Lemma 4.6 shows that each left coset has $|H|$ elements. The left cosets are disjoint and exhaust G, and there are $|G/H|$ left cosets. Thus G has $|G/H|\,|H|$ elements. □

If a is an element of a group G, then we have seen that the powers a^n of a form a cyclic subgroup of G that is isomorphic either to $\mathbb{Z}$ or to some group $\mathbb{Z}/m\mathbb{Z}$ for a positive integer m. We say that a has **finite order** m when the cyclic group is isomorphic to $\mathbb{Z}/m\mathbb{Z}$. Otherwise a has **infinite order**. In the finite-order case the order of a is thus the least positive integer n such that $a^n = 1$.

Corollary 4.8. If G is a finite group, then each element a of G has finite order, and the order of a divides the order of G.

PROOF. The order of a equals $|H|$ if $H = \{a^n \mid n \in \mathbb{Z}\}$, and Corollary 4.8 is thus a special case of Theorem 4.7. □

Corollary 4.9. If p is a prime, then the only group of order p, up to isomorphism, is the cyclic group C_p, and it has no subgroups other than $\{1\}$ and C_p itself.

PROOF. Suppose that G is a finite group of order p and that $H \neq \{1\}$ is a subgroup of G. Let $a \neq 1$ be in H, and let $P = \{a^n \mid n \in \mathbb{Z}\}$. Since $a \neq 1$, Corollary 4.8 shows that the order of a is an integer > 1 that divides p. Since p is prime, the order of a must equal p. Then $|P| = p$. Since $P \subseteq H \subseteq G$ and $|G| = p$, we must have $P = G$. □

Let G_1 and G_2 be groups. We say that $\varphi : G_1 \to G_2$ is a **homomorphism** if $\varphi(ab) = \varphi(a)\varphi(b)$ for all a and b in G. In other words, φ is to respect products, but it is not assumed that φ is one-one or onto. Any homomorphism φ automatically respects the identity and inverses, in the sense that

- $\varphi(1) = 1$ (since $\varphi(1) = \varphi(11) = \varphi(1)\varphi(1)$),
- $\varphi(a^{-1}) = \varphi(a)^{-1}$ (since $1 = \varphi(1) = \varphi(aa^{-1}) = \varphi(a)\varphi(a^{-1})$ and similarly $1 = \varphi(a^{-1})\varphi(a)$).

EXAMPLES. The following functions are homomorphisms: any isomorphism, the function $\varphi : \mathbb{Z} \to \mathbb{Z}/m\mathbb{Z}$ given by $\varphi(k) = k \bmod m$, the function $\varphi : \mathfrak{S}_n \to \{\pm 1\}$ given by $\varphi(\sigma) = \operatorname{sgn}\sigma$, the function $\varphi : \mathbb{Z} \to G$ given for fixed a in G by $\varphi(n) = a^n$, and the function $\varphi : GL(n, \mathbb{F}) \to \mathbb{F}^\times$ given by $\varphi(A) = \det A$.

The **image** of a homomorphism $\varphi : G_1 \to G_2$ is just the image of φ considered as a function. It is denoted by $\operatorname{image}\varphi = \varphi(G_1)$ and is necessarily a subgroup of G_2 since if $\varphi(g_1) = g_2$ and $\varphi(g_1') = g_2'$, then $\varphi(g_1 g_1') = g_2 g_2'$ and $\varphi(g_1^{-1}) = g_2^{-1}$.

The **kernel** of a homomorphism $\varphi : G_1 \to G_2$ is the set $\ker\varphi = \varphi^{-1}(\{1\}) = \{x \in G_1 \mid \varphi(x) = 1\}$. This is a subgroup since if $\varphi(x) = 1$ and $\varphi(y) = 1$, then $\varphi(xy) = \varphi(x)\varphi(y) = 1$ and $\varphi(x^{-1}) = \varphi(x)^{-1} = 1$.

The homomorphism $\varphi : G_1 \to G_2$ is one-one if and only if $\ker\varphi$ *is the trivial group* $\{1\}$. The necessity follows since 1 is already in $\ker\varphi$, and the sufficiency follows since $\varphi(x) = \varphi(y)$ implies that $\varphi(xy^{-1}) = 1$ and therefore that xy^{-1} is in $\ker\varphi$.

The kernel H of a homomorphism $\varphi : G_1 \to G_2$ has the additional property of being a **normal subgroup** of G_1 in the sense that ghg^{-1} is in H whenever g is in G_1 and h is in H, i.e., $gHg^{-1} = H$. In fact, if h is in $\ker\varphi$ and g is in G_1, then $\varphi(ghg^{-1}) = \varphi(g)\varphi(h)\varphi(g)^{-1} = \varphi(g)\varphi(g)^{-1} = 1$ shows that ghg^{-1} is in $\ker\varphi$.

EXAMPLES.

(1) Any subgroup H of an abelian group G is normal since $ghg^{-1} = gg^{-1}h = h$. The alternating subgroup $\mathfrak{A}_n$ of the symmetric group $\mathfrak{S}_n$ is normal since $\mathfrak{A}_n$ is the kernel of the homomorphism $\sigma \mapsto \operatorname{sgn}\sigma$.

(2) The subgroup $H = \{1,\ (1\ 2)\}$ of $\mathfrak{S}_3$ is not normal since $(1\ 2)H(1\ 2)^{-1} = \{1,\ (2\ 3)\}$.

(3) If a subgroup H of a group G has just two left cosets, then H is normal even if G is an infinite group. In fact, suppose $G = H \cup g_0 H$ whenever g_0 is not in H. Taking inverses of all elements of G, we see that $G = H \cup Hg_1$ whenever g_1 is not in H. If g in G is given, then either g is in H and $gHg^{-1} = H$, or g is not in H and $gH = Hg$, so that $gHg^{-1} = H$ in this case as well.

Let H be a subgroup of G. Let us look for the circumstances under which G/H inherits a multiplication from G. The natural definition is

$$(g_1H)(g_2H) \stackrel{?}{=} g_1g_2H,$$

but we have to check that this definition makes sense. The question is whether we get the same left coset as product if we change the representatives of g_1H and g_2H from g_1 and g_2 to g_1h_1 and g_2h_2. Since our prospective definition makes $(g_1h_1H)(g_2h_2H) = g_1h_1g_2h_2H$, the question is whether $g_1h_1g_2h_2H$ equals g_1g_2H. That is, we ask whether $g_1h_1g_2h_2 = g_1g_2h$ for some h in H. If this equality holds, then $h_1g_2h_2 = g_2h$, and hence $g_2^{-1}h_2g_2$ equals hh_2^{-1}, which is an element of H. Conversely if every expression $g_2^{-1}h_2g_2$ is in H, then we can go backwards and see that $g_1h_1g_2h_2 = g_1g_2h$ for some h in H, hence see that G/H indeed inherits a multiplication from G. Thus *a necessary and sufficient condition for G/H to inherit a multiplication from G is that the subgroup H is normal.* According to the next proposition, the multiplication inherited by G/H when this condition is satisfied makes G/H into a group.

Proposition 4.10. If H is a normal subgroup of a group G, then G/H becomes a group under the inherited multiplication $(g_1H)(g_2H) = (g_1g_2)H$, and the function $q : G \to G/H$ given by $q(g) = gH$ is a homomorphism of G onto G/H with kernel H. Consequently every normal subgroup of G is the kernel of some homomorphism.

REMARKS. When H is normal, the group G/H is called a **quotient group** of G, and the homomorphism $q : G \to G/H$ is called the **quotient homomorphism**.[5] In the special case that $G = \mathbb{Z}$ and $H = m\mathbb{Z}$, the construction reduces to the construction of the additive group of integers modulo m and accounts for using the notation $\mathbb{Z}/m\mathbb{Z}$ for that group.

PROOF. The coset $1H$ is the identity, and $(gH)^{-1} = g^{-1}H$. Also, the computation $(g_1Hg_2H)g_3H = g_1g_2g_3H = g_1H(g_2Hg_3H)$ proves associativity. Certainly q is onto G/H. It is a homomorphism since $q(g_1g_2) = g_1g_2H = g_1Hg_2H = q(g_1)q(g_2)$. □

In analogy with what was shown for vector spaces in Proposition 2.25, quotients in the context of groups allow for the factorization of certain homomorphisms of groups. The appropriate result is stated as Proposition 4.11 and is pictured in Figure 4.1. We can continue from there along the lines of Section II.5.

[5]Some authors call G/H a "factor group." A "factor set," however, is something different.

Proposition 4.11. Let $\varphi : G_1 \to G_2$ be a homomorphism between groups, let $H_0 = \ker \varphi$, let H be a normal subgroup of G_1 contained in H_0, and define $q : G_1 \to G_1/H$ to be the quotient homomorphism. Then there exists a homomorphism $\overline{\varphi} : G_1/H \to G_2$ such that $\varphi = \overline{\varphi} \circ q$, i.e, $\overline{\varphi}(g_1 H) = \varphi(g_1)$. It has the same image as φ, and $\ker \overline{\varphi} = \{h_0 H \mid h_0 \in H_0\}$.

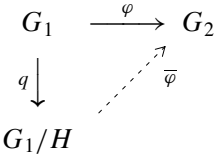


FIGURE 4.1. Factorization of homomorphisms of groups via the quotient of a group by a normal subgroup.

REMARK. One says that φ **factors through** G_1/H or **descends to** G_1/H. See Figure 4.1.

PROOF. We will have $\overline{\varphi} \circ q = \varphi$ if and only if $\overline{\varphi}$ satisfies $\overline{\varphi}(g_1 H) = \varphi(g_1)$. What needs proof is that $\overline{\varphi}$ is well defined. Thus suppose that g_1 and g_1' are in the same left coset, so that $g_1' = g_1 h$ with h in H. Then $\varphi(g_1') = \varphi(g_1)\varphi(h) = \varphi(g_1)$ since $H \subseteq \ker \varphi$, and $\overline{\varphi}$ is therefore well defined.

The computation $\overline{\varphi}(g_1 H g_2 H) = \overline{\varphi}(g_1 g_2 H) = \varphi(g_1 g_2) = \varphi(g_1)\varphi(g_2) = \overline{\varphi}(g_1 H)\overline{\varphi}(g_2 H)$ shows that $\overline{\varphi}$ is a homomorphism. Since image $\overline{\varphi}$ = image φ, $\overline{\varphi}$ is onto image φ. Finally $\ker \overline{\varphi}$ consists of all $g_1 H$ such that $\overline{\varphi}(g_1 H) = 1$. Since $\overline{\varphi}(g_1 H) = \varphi(g_1)$, the condition that g_1 is to satisfy is that g_1 be in $\ker \varphi = H_0$. Hence $\ker \overline{\varphi} = \{h_0 H \mid h_0 \in H_0\}$, as asserted. □

Corollary 4.12. Let $\varphi : G_1 \to G_2$ be a homomorphism between groups, and suppose that φ is onto G_2 and has kernel H. Then φ exhibits the group G_1/H as canonically isomorphic to G_2.

PROOF. Take $H = H_0$ in Proposition 4.11, and form $\overline{\varphi} : G_1/H \to G_2$ with $\varphi = \overline{\varphi} \circ q$. The proposition shows that $\overline{\varphi}$ is onto G_2 and has trivial kernel, i.e., the identity element of G_1/H. Having trivial kernel, $\overline{\varphi}$ is one-one. □

Theorem 4.13 (First Isomorphism Theorem). Let $\varphi : G_1 \to G_2$ be a homomorphism between groups, and suppose that φ is onto G_2 and has kernel K. Then the map $H_1 \mapsto \varphi(H_1)$ gives a one-one correspondence between

(a) the subgroups H_1 of G_1 containing K and
(b) the subgroups of G_2.

Under this correspondence normal subgroups correspond to normal subgroups. If H_1 is normal in G_1, then $gH_1 \mapsto \varphi(g)\varphi(H_1)$ is an isomorphism of G_1/H_1 onto $G_2/\varphi(H_1)$.

REMARK. In the special case of the last statement that $\varphi : G_1 \to G_2$ is a quotient map $q : G \to G/K$ and H is a normal subgroup of G containing K, the last statement of the theorem asserts the isomorphism

$$G/H \cong (G/K)\big/(H/K).$$

PROOF. The passage from (a) to (b) is by direct image under φ, and the passage from (b) to (a) will be by inverse image under φ^{-1}. Certainly the direct image of a subgroup as in (a) is a subgroup as in (b). To prove the one-one correspondence, we are to show that the inverse image of a subgroup as in (b) is a subgroup as in (a) and that these two constructions invert one another.

For any subgroup H_2 of G_2, $\varphi^{-1}(H_2)$ is a subgroup of G_1. In fact, if g_1 and g_1' are in $\varphi^{-1}(H_2)$, we can write $\varphi(g_1) = h_2$ and $\varphi(g_1') = h_2'$ with h_2 and h_2' in H_2. Then the equations $\varphi(g_1 g_1') = h_2 h_2'$ and $\varphi(g_1^{-1}) = \varphi(g_1)^{-1} = h_2^{-1}$ show that $g_1 g_1'$ and g_1^{-1} are in $\varphi^{-1}(H_2)$.

Moreover, the subgroup $\varphi^{-1}(H_2)$ contains $\varphi^{-1}(\{1\}) = K$. Therefore the inverse image under φ of a subgroup as in (b) is a subgroup as in (a). Since φ is a function, we have $\varphi(\varphi^{-1}(H_2)) = H_2$. Thus passing from (b) to (a) and back recovers the subgroup of G_2.

If H_1 is a subgroup of G_1 containing K, we still need to see that $H_1 = \varphi^{-1}(\varphi(H_1))$. Certainly $H_1 \subseteq \varphi^{-1}(\varphi(H_1))$. For the reverse inclusion let g_1 be in $\varphi^{-1}(\varphi(H_1))$. Then $\varphi(g_1)$ is in $\varphi(H_1)$, i.e., $\varphi(g_1) = \varphi(h_1)$ for some h_1 in H_1. Since φ is a homomorphism, $\varphi(g_1 h_1^{-1}) = 1$. Thus $g_1 h_1^{-1}$ is in $\ker \varphi = K$, which is contained in H_1 by assumption. Then h_1 and $g_1 h_1^{-1}$ are in H_1, and hence their product $(g_1 h_1^{-1})h_1 = g_1$ is in H_1. We conclude that $\varphi^{-1}(\varphi(H_1)) \subseteq H_1$, and thus passing from (a) to (b) and then back recovers the subgroup of G_1 containing K.

Next let us show that normal subgroups correspond to normal subgroups. If H_2 is normal in G_2, let H_1 be the subgroup $\varphi^{-1}(H_2)$ of G_1. For h_1 in H_1 and g_1 in G_1, we can write $\varphi(h_1) = h_2$ with h_2 in H_2, and then $\varphi(g_1 h_1 g_1^{-1}) = \varphi(g_1)h_2\varphi(g_1)^{-1}$ is in $\varphi(g_1)H_2\varphi(g_1)^{-1} = H_2$. Hence $g_1 h_1 g_1^{-1}$ is in $\varphi^{-1}(H_2) = H_1$. In the reverse direction let H_1 be normal in G_1, and let g_2 be in G_2. Since φ is onto G_2, we can write $g_2 = \varphi(g_1)$ for some g_1 in G_1. Then $g_2\varphi(H_1)g_2^{-1} = \varphi(g_1)\varphi(H_1)\varphi(g_1)^{-1} = \varphi(g_1 H_1 g_1^{-1}) = \varphi(H_1)$. Thus $\varphi(H_1)$ is normal.

For the final statement let $H_2 = \varphi(H_1)$. We have just proved that this image is normal, and hence G_2/H_2 is a group. The mapping $\Phi : G_1 \to G_2/H_2$ given by $\Phi(g_1) = \varphi(g_1)H_2$ is the composition of two homomorphisms and hence is a homomorphism. Its kernel is

$$\{g_1 \in G_1 \mid \varphi(g_1) \in H_2\} = \{g_1 \in G_1 \mid \varphi(g_1) \in \varphi(H_1)\} = \varphi^{-1}(\varphi(H_1)),$$

and this equals H_1 by the first conclusion of the theorem. Applying Corollary 4.12 to Φ, we obtain the required isomorphism $\overline{\Phi} : G_1/H_1 \to G_2/\varphi(H_1)$. □

Theorem 4.14 (Second Isomorphism Theorem). Let H_1 and H_2 be subgroups of a group G with H_2 normal in G. Then $H_1 \cap H_2$ is a normal subgroup of H_1, the set H_1H_2 of products is a subgroup of G with H_2 as a normal subgroup, and the map $h_1(H_1 \cap H_2) \mapsto h_1H_2$ is a well-defined canonical isomorphism of groups

$$H_1/(H_1 \cap H_2) \cong (H_1H_2)/H_2.$$

PROOF. The set $H_1 \cap H_2$ is a subgroup, being the intersection of two subgroups. For h_1 in H_1, we have $h_1(H_1 \cap H_2)h_1^{-1} \subseteq h_1H_1h_1^{-1} \subseteq H_1$ since H_1 is a subgroup and $h_1(H_1 \cap H_2)h_1^{-1} \subseteq h_1H_2h_1^{-1} \subseteq H_2$ since H_2 is normal in G. Therefore $h_1(H_1 \cap H_2)h_1^{-1} \subseteq H_1 \cap H_2$, and $H_1 \cap H_2$ is normal in H_1.

The set H_1H_2 of products is a subgroup since $h_1h_2h_1'h_2' = h_1h_1'(h_1'^{-1}h_2h_1')h_2'$ and since $(h_1h_2)^{-1} = h_1^{-1}(h_1h_2^{-1}h_1^{-1})$, and H_2 is normal in H_1H_2 since H_2 is normal in G.

The function $\varphi(h_1(H_1 \cap H_2)) = h_1H_2$ is well defined since $H_1 \cap H_2 \subseteq H_2$, and φ respects products. The domain of φ is $\{h_1(H_1 \cap H_2) \mid h_1 \in H_1\}$, and the kernel is the subset of this such that h_1 lies in H_2 as well as H_1. For this to happen, h_1 must be in $H_1 \cap H_2$, and thus the kernel is the identity coset of $H_1/(H_1 \cap H_2)$. Hence φ is one-one.

To see that φ is onto $(H_1H_2)/H_2$, let $h_1h_2H_2$ be given. Then $h_1(H_1 \cap H_2)$ maps to h_1H_2, which equals $h_1h_2H_2$. Hence φ is onto. □

3. Direct Products and Direct Sums

We return to the matter of direct products and direct sums of groups, direct products having been discussed briefly in Section 1. In a footnote in Section II.4 we mentioned a general principle in algebra that "whenever a new systematic construction appears for the objects under study, it is well to look for a corresponding construction with the functions relating these new objects." This principle will be made more precise in Section 11 of the present chapter with the aid of the language of "categories" and "functors."

Another principle that will be relevant for us is that constructions in one context in algebra often recur, sometimes in slightly different guise, in other contexts. One example of the operation of this principle occurs with quotients. The construction and properties of the quotient of a vector space by a vector subspace, as in Section II.5, is analogous in this sense to the construction and properties of the quotient of a group by a *normal* subgroup, as in Section 2 in the present chapter. The need for the subgroup to be normal is an example of what is meant by "slightly different guise." Anyway, this principle too will be made more precise in Section 11 of the present chapter using the language of categories and functors.

Let us proceed with an awareness of both these principles in connection with direct products and direct sums of groups, looking for analogies with what happened for vector spaces and expecting our work to involve constructions with homomorphisms as well as with groups.

The external direct product $G_1 \times G_2$ was defined as a group in Section 1 to be the set-theoretic product with coordinate-by-coordinate multiplication. There are four homomorphisms of interest connected with $G_1 \times G_2$, namely

$$\begin{aligned} i_1 &: G_1 \to G_1 \times G_2 \qquad \text{given by} \quad i_1(g_1) = (g_1, 1), \\ i_2 &: G_2 \to G_1 \times G_2 \qquad \text{given by} \quad i_2(g_2) = (1, g_2), \\ p_1 &: G_1 \times G_2 \to G_1 \qquad \text{given by} \quad p_1(g_1, g_2) = g_1, \\ p_2 &: G_1 \times G_2 \to G_2 \qquad \text{given by} \quad p_2(g_1, g_2) = g_2. \end{aligned}$$

Recall from the discussion before Proposition 4.5 that Proposition 2.30 for the direct product of two vector spaces does not translate directly into an analog for the direct product of groups; instead that proposition is replaced by Proposition 4.5, which involves some condition of commutativity.

Warned by this anomaly, let us work with mappings rather than with groups and subgroups, and let us use mappings in formulating a definition of the direct product of groups. As with the direct product of two vector spaces, the mappings to use are p_1 and p_2 but not i_1 and i_2. The way in which p_1 and p_2 enter is through the effect of the direct product on homomorphisms. If $\varphi_1 : H \to G_1$ and $\varphi_2 : H \to G_2$ are two homomorphisms, then $h \mapsto (\varphi_1(h), \varphi_2(h))$ is the corresponding homomorphism of H into $G_1 \times G_2$. In order to state matters fully, let us give the definition with an arbitrary number of factors.

Let S be an arbitrary nonempty set of groups, and let G_s be the group corresponding to the member s of S. The **external direct product** of the G_s's consists of a group $\prod_{s\in S} G_s$ and a system of group homomorphisms. The group as a set is $\bigtimes_{s\in S} G_s$, whose elements are arbitrary functions from S to $\bigcup_{s\in S} G_s$ such that the value of the function at s is in G_s, and the group law is $\big(\{g_s\}_{s\in S}\big)\big(\{g'_s\}_{s\in S}\big) = \{g_s g'_s\}_{s\in S}$. The group homomorphisms are the coordinate mappings $p_{s_0} : \prod_{s\in S} G_s \to G_{s_0}$ with $p_{s_0}\big(\{g_s\}_{s\in S}\big) = g_{s_0}$. The individual groups G_s are called the **factors**, and a direct product of n groups may be written as $G_1 \times \cdots \times G_n$ instead of with the symbol $\prod$. The group $\prod_{s\in S} G_s$ has the **universal mapping property** described in Proposition 4.15 and pictured in Figure 4.2.

Proposition 4.15 (universal mapping property of external direct product). Let $\{G_s \mid s \in S\}$ be a nonempty set of groups, and let $\prod_{s\in S} G_s$ be the external direct product, the associated group homomorphisms being the coordinate mappings $p_{s_0} : \prod_{s\in S} G_s \to G_{s_0}$. If H is any group and $\{\varphi_s \mid s \in S\}$ is a system of group homomorphisms $\varphi_s : H \to G_s$, then there exists a unique group homomorphism $\varphi : H \to \prod_{s\in S} G_s$ such that $p_{s_0} \circ \varphi = \varphi_{s_0}$ for all $s_0 \in S$.

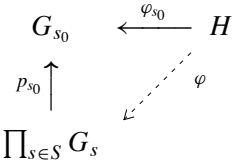


FIGURE 4.2. Universal mapping property of an external direct product of groups.

PROOF. Existence of φ is proved by taking $\varphi(h) = \{\varphi_s(h)\}_{s\in S}$. Then $p_{s_0}(\varphi(h))$ $= p_{s_0}\big(\{\varphi_s(h)\}_{s\in S}\big) = \varphi_{s_0}(h)$ as required. For uniqueness let $\varphi' : H \to \prod_{s\in S} G_s$ be a homomorphism with $p_{s_0} \circ \varphi' = \varphi_{s_0}$ for all $s_0 \in S$. For each h in H, we can write $\varphi'(h) = \{\varphi'(h)_s\}_{s\in S}$. For s_0 in S, we then have $\varphi_{s_0}(h) = (p_{s_0} \circ \varphi')(h) =$ $p_{s_0}(\varphi'(h)) = \varphi'(h)_{s_0}$, and we conclude that $\varphi' = \varphi$. □

Now we give an abstract definition of direct product that allows for the possibility that the direct product is "internal" in the sense that the various factors are identified as subgroups of a given group. The definition is by means of the above universal mapping property and will be seen to characterize the direct product up to canonical isomorphism. Let S be an arbitrary nonempty set of groups, and let G_s be the group corresponding to the member s of S. A **direct product** of the G_s's consists of a group G and a system of group homomorphisms $p_s : G \to G_s$ for $s \in S$ with the following **universal mapping property**: whenever H is a group and $\{\varphi_s \mid s \in S\}$ is a system of group homomorphisms $\varphi_s : H \to G_s$, then there exists a unique group homomorphism $\varphi : H \to G$ such that $p_s \circ \varphi = \varphi_s$ for all $s \in S$. Proposition 4.15 proves existence of a direct product, and the next proposition addresses uniqueness. A direct product is **internal** if each G_s is a subgroup of G and each restriction $p_s\big|_{G_s}$ is the identity map.

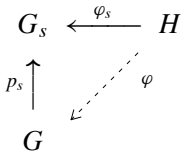


FIGURE 4.3. Universal mapping property of a direct product of groups.

Proposition 4.16. Let S be a nonempty set of groups, and let G_s be the group corresponding to the member s of S. If $(G, \{p_s\})$ and $(G', \{p'_s\})$ are two direct products, then the homomorphisms $p_s : G \to G_s$ and $p'_s : G' \to G_s$ are onto G_s, there exists a unique homomorphism $\Phi : G' \to G$ such that $p'_s = p_s \circ \Phi$ for all $s \in S$, and Φ is an isomorphism.

PROOF. In Figure 4.3 let $H = G'$ and $\varphi_s = p'_s$. If $\Phi : G' \to G$ is the homomorphism produced by the fact that G is a direct product, then we have

$p_s \circ \Phi = p'_s$ for all s. Reversing the roles of G and G', we obtain a homomorphism $\Phi' : G \to G'$ with $p'_s \circ \Phi' = p_s$ for all s. Therefore $p_s \circ (\Phi \circ \Phi') = p'_s \circ \Phi' = p_s$.

In Figure 4.3 we next let $H = G$ and $\varphi_s = p_s$ for all s. Then the identity 1_G on G has the same property $p_s \circ 1_G = p_s$ relative to all p_s that $\Phi \circ \Phi'$ has, and the uniqueness says that $\Phi \circ \Phi' = 1_G$. Reversing the roles of G and G', we obtain $\Phi' \circ \Phi = 1_{G'}$. Therefore Φ is an isomorphism.

For uniqueness suppose that $\Psi : G' \to G$ is another homomorphism with $p'_s = p_s \circ \Psi$ for all $s \in S$. Then the argument of the previous paragraph shows that $\Phi' \circ \Psi = 1_{G'}$. Applying Φ on the left gives $\Psi = (\Phi \circ \Phi') \circ \Psi = \Phi \circ (\Phi' \circ \Psi) = \Phi \circ 1_{G'} = \Phi$. Thus $\Psi = \Phi$.

Finally we have to show that the s^{th} mapping of a direct product is onto G_s. It is enough to show that p'_s is onto G_s. Taking G as the external direct product $\prod_{s \in S} G_s$ with p_s equal to the coordinate mapping, form the isomorphism $\Phi' : G \to G'$ that has just been proved to exist. This satisfies $p_s = p'_s \circ \Phi'$ for all $s \in S$. Since p_s is onto G_s, p'_s must be onto G_s. □

Let us turn to direct sums. Part of what we seek is a definition that allows for an abstract characterization of direct sums in the spirit of Proposition 4.16. In particular, the interaction with homomorphisms is to be central to the discussion. In the case of two factors, we use i_1 and i_2 rather than p_1 and p_2. If $\varphi_1 : G_1 \to H$ and $\varphi_2 : G_2 \to H$ are two homomorphisms, then the corresponding homomorphism φ of $G_1 \oplus G_2$ to H is to satisfy $\varphi_1 = \varphi \circ i_1$ and $\varphi_2 = \varphi \circ i_2$. With $G_1 \oplus G_2$ defined, as expected, to be the same group as $G_1 \times G_2$, we are led to the formula

$$\varphi(g_1, g_2) = \varphi(g_1, 1)\varphi(1, g_2) = \varphi_1(g_1)\varphi_2(g_2).$$

The images of commuting elements under a homomorphism have to commute, and hence H had better be abelian. Then in order to have an analog of Proposition 4.16, we will want to specialize H at some point to $G_1 \oplus G_2$, and therefore G_1 and G_2 had better be abelian. With these observations in place, we are ready for the general definition.

Let S be an arbitrary nonempty set of *abelian* groups, and let G_s be the group corresponding to the member s of S. We shall use additive notation for the group operation in each G_s. The **external direct sum** of the G_s's consists of an abelian group $\bigoplus_{s \in S} G_s$ and a system of group homomorphisms i_s for $s \in S$. The group is the subgroup of $\prod_{s \in S} G_s$ of all elements that are equal to 0 in all but finitely many coordinates. The group homomorphisms are the mappings $i_{s_0} : G_{s_0} \to \bigoplus_{s \in S} G_s$ carrying a member g_{s_0} of G_{s_0} to the element that is g_{s_0} in coordinate s_0 and is 0 at all other coordinates. The individual groups are called the **summands**, and a direct sum of n abelian groups may be written as $G_1 \oplus \cdots \oplus G_n$. The group $\bigoplus_{s \in S} G_s$ has the **universal mapping property** described in Proposition 4.17 and pictured in Figure 4.4.

Proposition 4.17 (universal mapping property of external direct sum). Let $\{G_s \mid s \in S\}$ be a nonempty set of abelian groups, and let $\bigoplus_{s\in S} G_s$ be the external direct sum, the associated group homomorphisms being the embedding mappings $i_{s_0} : G_{s_0} \to \bigoplus_{s\in S} G_s$. If H is any abelian group and $\{\varphi_s \mid s \in S\}$ is a system of group homomorphisms $\varphi_s : G_s \to H$, then there exists a unique group homomorphism $\varphi : \bigoplus_{s\in S} G_s \to H$ such that $\varphi \circ i_{s_0} = \varphi_{s_0}$ for all $s_0 \in S$.

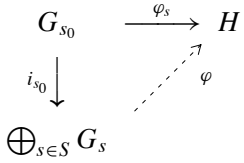


FIGURE 4.4. Universal mapping property of an external direct sum of abelian groups.

PROOF. Existence of φ is proved by taking $\varphi\big(\{g_s\}_{s\in S}\big) = \sum_s \varphi_s(g_s)$. The sum on the right side is meaningful since the element $\{g_s\}_{s\in S}$ of the direct sum has only finitely many nonzero coordinates. Since H is abelian, the computation

$$\begin{aligned}\varphi\big(\{g_s\}_{s\in S}\big) + \varphi\big(\{g'_s\}_{s\in S}\big) &= \textstyle\sum_s \varphi_s(g_s) + \sum_s \varphi_s(g'_s)\\ &= \textstyle\sum_s (\varphi_s(g_s) + \varphi_s(g'_s)) = \sum_s \varphi_s(g_s + g'_s)\\ &= \varphi\big(\{g_s + g'_s\}_{s\in S}\big) = \varphi\big(\{g_s\}_{s\in S} + \{g'_s\}_{s\in S}\big)\end{aligned}$$

shows that φ is a homomorphism. If g_{s_0} is given and $\{g_s\}_{s\in S}$ denotes the element that is g_{s_0} in the $s_0{}^{\text{th}}$ coordinate and is 0 elsewhere, then $\varphi(i_{s_0}(g_{s_0})) = \varphi\big(\{g_s\}_{s\in S}\big) = \sum_s \varphi_s(g_s)$, and the right side equals $\varphi_{s_0}(g_{s_0})$ since $g_s = 0$ for all other s's. Thus $\varphi \circ i_{s_0} = \varphi_{s_0}$.

For uniqueness let $\varphi' : \bigoplus_{s\in S} G_s \to H$ be a homomorphism with $\varphi' \circ i_{s_0} = \varphi_{s_0}$ for all $s_0 \in S$. Then the value of φ' is determined at all elements of $\bigoplus_{s\in S} G_s$ that are 0 in all but one coordinate. Since the most general member of $\bigoplus_{s\in S} G_s$ is a finite sum of such elements, φ' is determined on all of $\bigoplus_{s\in S} G_s$. □

Now we give an abstract definition of direct sum that allows for the possibility that the direct sum is "internal" in the sense that the various constituents are identified as subgroups of a given group. Again the definition is by means of a universal mapping property and will be seen to characterize the direct sum up to canonical isomorphism. Let S be an arbitrary nonempty set of *abelian* groups, and let G_s be the group corresponding to the member s of S. A **direct sum** of the G_s's consists of an abelian group G and a system of group homomorphisms $i_s : G_s \to G$ for $s \in S$ with the following **universal mapping property**: whenever H is an abelian group and $\{\varphi_s \mid s \in S\}$ is a system of group homomorphisms

$\varphi_s : G_s \to H$, then there exists a unique group homomorphism $\varphi : G \to H$ such that $\varphi \circ i_s = \varphi_s$ for all $s \in S$. Proposition 4.17 proves existence of a direct sum, and the next proposition addresses uniqueness. A direct sum is **internal** if each G_s is a subgroup of G and each mapping i_s is the inclusion mapping.

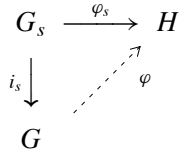


FIGURE 4.5. Universal mapping property of a direct sum of abelian groups.

Proposition 4.18. Let S be a nonempty set of abelian groups, and let G_s be the group corresponding to the member s of S. If $(G, \{i_s\})$ and $(G', \{i'_s\})$ are two direct sums, then the homomorphisms $i_s : G_s \to G$ and $i'_s : G_s \to G'$ are one-one, there exists a unique homomorphism $\Phi : G \to G'$ such that $i'_s = \Phi \circ i_s$ for all $s \in S$, and Φ is an isomorphism.

PROOF. In Figure 4.5 let $H = G'$ and $\varphi_s = i'_s$. If $\Phi : G \to G'$ is the homomorphism produced by the fact that G is a direct sum, then we have $\Phi \circ i_s = i'_s$ for all s. Reversing the roles of G and G', we obtain a homomorphism $\Phi' : G' \to G$ with $\Phi' \circ i'_s = i_s$ for all s. Therefore $(\Phi' \circ \Phi) \circ i_s = \Phi' \circ i'_s = i_s$.

In Figure 4.5 we next let $H = G$ and $\varphi_s = i_s$ for all s. Then the identity 1_G on G has the same property $1_G \circ i_s = i_s$ relative to all i_s that $\Phi' \circ \Phi$ has, and the uniqueness says that $\Phi' \circ \Phi = 1_G$. Reversing the roles of G and G', we obtain $\Phi \circ \Phi' = 1_{G'}$. Therefore Φ is an isomorphism.

For uniqueness suppose that $\Psi : G \to G'$ is another homomorphism with $i'_s = \Psi \circ i_s$ for all $s \in S$. Then the argument of the previous paragraph shows that $\Phi' \circ \Psi = 1_G$. Applying Φ on the left gives $\Psi = (\Phi \circ \Phi') \circ \Psi = \Phi \circ (\Phi' \circ \Psi) = \Phi \circ 1_G = \Phi$. Thus $\Psi = \Phi$.

Finally we have to show that the s^{th} mapping of a direct sum is one-one on G_s. It is enough to show that i'_s is one-one on G_s. Taking G as the external direct sum $\bigoplus_{s \in S} G_s$ with i_s equal to the embedding mapping, form the isomorphism $\Phi' : G' \to G$ that has just been proved to exist. This satisfies $i_s = \Phi' \circ i'_s$ for all $s \in S$. Since i_s is one-one, i'_s must be one-one. □

EXAMPLE. The group $\mathbb{Q}^\times$ is the direct sum of copies of $\mathbb{Z}$, one for each prime, plus one copy of $\mathbb{Z}/2\mathbb{Z}$. If p is a prime, the mapping $i_p : \mathbb{Z} \to \mathbb{Q}^\times$ is given by $i_p(n) = p^n$. The remaining coordinate gives the sign. The isomorphism results from unique factorization, only finitely many primes being involved for any particular nonzero rational number.

4. Rings and Fields

In this section we begin a two-section digression in order to develop some more number theory beyond what is in Chapter I and to make some definitions as new notions arise. In later sections of the present chapter, some of this material will yield further examples of concrete groups and tools for working with them.

We begin with the additive group $\mathbb{Z}/m\mathbb{Z}$ of integers modulo a positive integer m. We continue to write $[a]$ for the equivalence class of the integer a when it is helpful to do so. Our interest will be in the multiplication structure that $\mathbb{Z}/m\mathbb{Z}$ inherits from multiplication in $\mathbb{Z}$. Namely, we attempt to define

$$[a][b] = [ab].$$

To see that this formula is meaningful in $\mathbb{Z}/m\mathbb{Z}$, we need to check that the same equivalence class results on the right side if the representatives of $[a]$ and $[b]$ are changed. Thus let $[a] = [a']$ and $[b] = [b']$. Then m divides $a - a'$ and $b - b'$ and must divide the sum of products $(a - a')b + a'(b - b') = ab - a'b'$. Consequently $[ab] = [a'b']$, and multiplication is well defined. If x and y are in $\mathbb{Z}/m\mathbb{Z}$, their product is often denoted by $xy \bmod m$.

The same kind of argument as just given shows that the associativity of multiplication in $\mathbb{Z}$ and the distributive laws imply corresponding facts about $\mathbb{Z}/m\mathbb{Z}$. The result is that $\mathbb{Z}/m\mathbb{Z}$ is a "commutative ring with identity" in the sense of the following definitions.

A **ring** is a set R with two operations $R \times R \to R$, usually called **addition** and **multiplication** and often denoted by $(a, b) \mapsto a + b$ and $(a, b) \mapsto ab$, such that

(i) R is an abelian group under addition,
(ii) multiplication is associative in the sense that $a(bc) = (ab)c$ for all a, b, c in R,
(iii) the two distributive laws

$$a(b + c) = (ab) + (ac) \qquad \text{and} \qquad (b + c)a = (ba) + (ca)$$

hold for all a, b, c in R.

The additive identity is denoted by 0, and the additive inverse of a is denoted by $-a$. A sum $a + (-b)$ is often abbreviated $a - b$. By convention when parentheses are absent, multiplications are to be carried out before additions and subtractions. Thus the distributive laws may be rewritten as

$$a(b + c) = ab + ac \qquad \text{and} \qquad (b + c)a = ba + ca.$$

A ring R is called a **commutative ring** if multiplication satisfies the commutative law

(iv) $ab = ba$ for all a and b in R.

A ring R is called a **ring with identity**[6] if there exists an element 1 such that $1a = a1 = a$ for all a in R. It is immediate from the definitions that

- $0a = 0$ and $a0 = 0$ in any ring (since, in the case of the first formula, $0 = 0a - 0a = (0 + 0)a - 0a = 0a + 0a - 0a = 0a$),
- the multiplicative identity is unique in a ring with identity (since $1' = 1'1 = 1$),
- $(-1)a = -a = a(-1)$ in any ring with identity (partly since $0 = 0a = (1 + (-1))a = 1a + (-1)a = a + (-1)a$).

In a ring with identity, it will be convenient not to insist that the identity be different from the zero element 0. If 1 and 0 do happen to coincide in R, then it readily follows that 0 is the only element of R, and R is said to be the **zero ring**.

The set $\mathbb{Z}$ of integers is a basic example of a commutative ring with identity. Returning to $\mathbb{Z}/m\mathbb{Z}$, suppose now that m is a prime p. If $[a]$ is in $\mathbb{Z}/p\mathbb{Z}$ with a in $\{1, 2, \dots, p-1\}$, then $\mathrm{GCD}(a, p) = 1$ and Proposition 1.2 produces integers r and s with $ar + ps = 1$. Modulo p, this equation reads $[a][r] = [1]$. In other words, $[r]$ is a multiplicative inverse of $[a]$. The result is that $\mathbb{Z}/p\mathbb{Z}$, when p is a prime, is a "field" in the sense of the following definition.

A **field** $\mathbb{F}$ is a commutative ring with identity such that $\mathbb{F} \neq 0$ and such that

(v) to each $a \neq 0$ in $\mathbb{F}$ corresponds an element a^{-1} in $\mathbb{F}$ such that $aa^{-1} = 1$.

In other words, $\mathbb{F}^\times = \mathbb{F} - \{0\}$ is an abelian group under multiplication. Inverses are necessarily unique as a consequence of one of the properties of groups.

When p is prime, we shall write $\mathbb{F}_p$ for the field $\mathbb{Z}/p\mathbb{Z}$. Its multiplicative group $\mathbb{F}_p^\times$ has order $p - 1$, and Lagrange's Theorem (Corollary 4.8) immediately implies that $a^{p-1} \equiv 1 \bmod p$ whenever a and p are relatively prime. This result is known as **Fermat's Little Theorem**.[7]

For general m, certain members of $\mathbb{Z}/m\mathbb{Z}$ have multiplicative inverses. The product of two such elements is again one, and the inverse of one is again one. Thus, even though $\mathbb{Z}/m\mathbb{Z}$ need not be a field, the subset $(\mathbb{Z}/m\mathbb{Z})^\times$ of members of $\mathbb{Z}/m\mathbb{Z}$ with multiplicative inverses is a group. The same argument as when m is prime shows that the class of a has an inverse if and only if $\mathrm{GCD}(a, m) = 1$. The number of such classes was defined in Chapter I in terms of the Euler φ function as $\varphi(m)$, and a formula for $\varphi(m)$ was obtained in Corollary 1.10. The

[6]Some authors, particularly when discussing only algebra, find it convenient to incorporate the existence of an identity into the definition of a ring. However, in real analysis some important natural rings do not have an identity, and the theory is made more complicated by forcing an identity into the picture. For example the space of integrable functions on $\mathbb{R}$ forms a very natural ring, with convolution as multiplication, and there is no identity; forcing an identity into the picture in such a way that the space remains stable under translations makes the space large and unwieldy. The distinction between working with rings and working with rings with identity will be discussed further in Section 11.

[7]As opposed to Fermat's Last Theorem, which lies deeper.

conclusion is that $(\mathbb{Z}/m\mathbb{Z})^\times$ is an abelian group of order $\varphi(m)$. Application of Lagrange's Theorem yields Euler's generalization of Fermat's Little Theorem, namely that $a^{\varphi(m)} \equiv 1 \bmod m$ for every positive integer m and every integer a relatively prime to m.

More generally, in any ring R with identity, a **unit** is defined to be any element a such that there exists an element a^{-1} with $aa^{-1} = a^{-1}a = 1$. The element a^{-1} is unique if it exists[8] and is called the **multiplicative inverse** of a. The units of R form a group denoted by $R^\times$. For example the group $\mathbb{Z}^\times$ consists of $+1$ and -1, and the zero ring R has $R^\times = \{0\}$. If R is a nonzero ring, then 0 is not in $R^\times$.

Here are some further examples of fields.

EXAMPLES OF FIELDS.

(1) $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$. These are all fields.

(2) $\mathbb{Q}[\theta]$. This was introduced between Examples 8 and 9 of Section 1. It is assumed that θ is a complex number and that there exists an integer $n > 0$ such that the complex numbers $1, \theta, \theta^2, \dots, \theta^n$ are linearly dependent over $\mathbb{Q}$. The set $\mathbb{Q}[\theta]$ is defined to be the linear span over $\mathbb{Q}$ of all powers $1, \theta, \theta^2, \dots$ of θ, which is the same as the linear span of the finite set $1, \theta, \theta^2, \dots, \theta^{n-1}$. The set $\mathbb{Q}[\theta]$ was shown in Proposition 4.1 to be a subset of $\mathbb{C}$ that is closed under the arithmetic operations, including the passage to reciprocals in the case of the nonzero elements. It is therefore a field.

(3) A field of 4 elements. Let $\mathbb{F}_4 = \{0, 1, \theta, \theta+1\}$, where θ is some symbol not standing for 0 or 1. Define addition in $\mathbb{F}_4$ and multiplication in $\mathbb{F}_4^\times$ by requiring that $a + 0 = 0 + a = a$ for all a, that

$$\begin{aligned} 1+1&=0, & 1+\theta&=(\theta+1), & 1+(\theta+1)&=\theta,\\ \theta+1&=(\theta+1), & \theta+\theta&=0, & \theta+(\theta+1)&=1,\\ (\theta+1)+1&=\theta, & (\theta+1)+\theta&=1, & (\theta+1)+(\theta+1)&=0, \end{aligned}$$

and that

$$\begin{aligned} 11&=1, & 1\theta&=\theta, & 1(\theta+1)&=(\theta+1),\\ \theta 1&=\theta, & \theta\theta&=(\theta+1), & \theta(\theta+1)&=1,\\ (\theta+1)1&=(\theta+1), & (\theta+1)\theta&=1, & (\theta+1)(\theta+1)&=\theta. \end{aligned}$$

The result is a field. With this direct approach a certain amount of checking is necessary to verify all the properties of a field. We shall return to this matter in Chapter IX when we consider finite fields more generally, and we shall then have a way of constructing $\mathbb{F}_4$ that avoids tedious checking.

[8] In fact, if b and c exist with $ab = ca = 1$, then a is a unit with $a^{-1} = b = c$ because $b = 1b = (ca)b = c(ab) = c1 = c$.

In analogy with the theory of groups, we define a **subring** of a ring to be a nonempty subset that is closed under addition, negation, and multiplication. The set $2\mathbb{Z}$ of even integers is a subring of the ring $\mathbb{Z}$ of integers. A **subfield** of a field is a subset containing 0 and 1 that is closed under addition, negation, multiplication, and multiplicative inverses for its nonzero elements. The set $\mathbb{Q}$ of rationals is a subfield of the field $\mathbb{R}$ of reals.

Intermediate between rings and fields are two kinds of objects—integral domains and division rings—that arise frequently enough to merit their own names.

The setting for the first is a commutative ring R. A nonzero element a of R is called a **zero divisor** if there is some nonzero b in R with $ab = 0$. For example the element 2 in the ring $\mathbb{Z}/6\mathbb{Z}$ is a zero divisor because $2 \cdot 3 = 0$. An **integral domain** is a *nonzero* commutative ring with identity having no zero divisors. Fields have no zero divisors since if a and b are nonzero, then $ab = 0$ would force $b = 1b = (a^{-1}a)b = a^{-1}(ab) = a^{-1}0 = 0$ and would give a contradiction; therefore every field is an integral domain. The ring of integers $\mathbb{Z}$ is another example of an integral domain, and the polynomial rings $\mathbb{Q}[X]$ and $\mathbb{R}[X]$ and $\mathbb{C}[X]$ introduced in Section I.3 are further examples. A cancellation law for multiplication holds in any integral domain:

$$ab = ac \text{ with } a \neq 0 \qquad \text{implies} \qquad b = c.$$

In fact, $ab = ac$ implies $a(b - c) = 0$; since $a \neq 0$, $b - c$ must be 0.

The other object with its own name is a **division ring**, which is a nonzero ring with identity such that every nonzero element is a unit. The commutative division rings are the fields, and we have encountered only one noncommutative division ring so far. That is the set $\mathbb{H}$ of quaternions, which was introduced in Section 1. Division rings that are not fields will play only a minor role in this book but are of great interest in Chapters II and III of *Advanced Algebra*.

Let us turn to mappings. A function $\varphi : R \to R'$ between two rings is an **isomorphism** of rings if φ is one-one onto and satisfies $\varphi(a + b) = \varphi(a) + \varphi(b)$ and $\varphi(ab) = \varphi(a)\varphi(b)$ for all a and b in R. In other words, φ is to be an isomorphism of the additive groups and to satisfy $\varphi(ab) = \varphi(a)\varphi(b)$. Such a mapping carries the identity, if any, in R to the identity of R'. The relation "is isomorphic to" is an equivalence relation. Common notation for an isomorphism of rings is $R \cong R'$; because of the symmetry, one can say that R and R' are **isomorphic**.

A function $\varphi : R \to R'$ between two rings is a **homomorphism** of rings if φ satisfies $\varphi(a + b) = \varphi(a) + \varphi(b)$ and $\varphi(ab) = \varphi(a)\varphi(b)$ for all a and b in R. In other words, φ is to be a homomorphism of the additive groups and to satisfy $\varphi(ab) = \varphi(a)\varphi(b)$.

EXAMPLES OF HOMOMORPHISMS OF RINGS.

(1) The mapping $\varphi : \mathbb{Z} \to \mathbb{Z}/m\mathbb{Z}$ given by $\varphi(k) = k \bmod m$.

(2) The evaluation mapping $\varphi : \mathbb{R}[X] \to \mathbb{R}$ given by $P(X) \mapsto P(r)$ for some fixed r in $\mathbb{R}$.

(3) Mappings with the direct product $\mathbb{Z} \times \mathbb{Z}$. The additive group $\mathbb{Z} \times \mathbb{Z}$ becomes a commutative ring with identity under coordinate-by-coordinate multiplication, namely $(a, a') + (b, b') = (a + b, a' + b')$. The identity is $(1, 1)$. Projection $(a, a') \mapsto a$ to the first coordinate is a homomorphism of rings $\mathbb{Z} \times \mathbb{Z} \to \mathbb{Z}$ that carries identity to identity. Inclusion $a \mapsto (a, 0)$ of $\mathbb{Z}$ into the first coordinate is a homomorphism of rings $\mathbb{Z} \to \mathbb{Z} \times \mathbb{Z}$ that does not carry identity to identity.[9]

Proposition 4.19. If R is a ring with identity 1_R, then there exists a unique homomorphism of rings $\varphi_1 : \mathbb{Z} \to R$ such that $\varphi(1) = 1_R$.

PROOF. The formulas for manipulating exponents of an element in a group, when translated into the additive notation for addition in R, say that $n \mapsto nr$ satisfies $(m + n)r = mr + nr$ and $(mn)r = m(nr)$ for all r in R and all integers m and n. The first of these formulas implies, for any r in R, that $\varphi_r(n) = nr$ is a homomorphism between the additive groups of $\mathbb{Z}$ and R, and it is certainly uniquely determined by its value for $n = 1$. The distributive laws imply that $\psi_r(r') = r'r$ is another homomorphism of additive groups. Hence $\psi_r \circ \varphi_{r'}$ and $\varphi_{r'r}$ are homomorphisms between the additive groups of $\mathbb{Z}$ and R. Since $(\psi_r \circ \varphi_{r'})(1) = \psi_r(r') = r'r = \varphi_{r'r}(1)$, we must have $(\psi_r \circ \varphi_{r'})(m) = \varphi_{r'r}(m)$ for all integers m. Thus $(mr')r = m(r'r)$ for all m. Putting $r = n1_R$ and $r' = 1_R$ proves the fourth equality of the computation

$$\begin{aligned}\varphi_1(mn) &= (mn)1_R = m(n1_R) \\ &= m(1_R(n1_R)) = (m1_R)(n1_R) = \varphi_1(m)\varphi_1(n),\end{aligned}$$

and shows that φ_1 is in fact a homomorphism of rings. □

The image of a homomorphism $\varphi : R \to R'$ of rings is a subring of R', as is easily checked. The kernel turns out to be more than just of subring of R. If a is in the kernel and b is any element of R, then $\varphi(ab) = \varphi(a)\varphi(b) = 0\varphi(b) = 0$ and similarly $\varphi(ba) = 0$. Thus the kernel of a ring homomorphism is closed under products of members of the kernel with arbitrary members of R. Adapting a definition to this circumstance, one says that an **ideal** I of R (or **two-sided ideal** in case of ambiguity) is an additive subgroup such that ab and ba are in I whenever a is in I and b is in R. Briefly then, the kernel of a homomorphism of rings is an ideal.

Conversely suppose that I is an ideal in a ring R. Since I is certainly an additive subgroup of an abelian group, we can form the additive quotient group

[9]Sometimes authors who build the existence of an identity into the definition of "ring" insist as a matter of definition that homomorphisms of rings carry identity to identity. Such authors would then exclude this particular mapping from consideration as a homomorphism.

R/I. It is customary to write the individual cosets in additive notation, thus as $r + I$. In analogy with Proposition 4.10, we have the following result for the present context.

Proposition 4.20. If I is an ideal in a ring R, then a well-defined operation of multiplication is obtained within the additive group R/I by the definition $(r_1 + I)(r_2 + I) = r_1r_2 + I$, and R/I becomes a ring. If R has an identity 1, then $1 + I$ is an identity in R/I. With these definitions the function $q : R \to R/I$ given by $q(r) = r + I$ is a ring homomorphism of R onto R/I with kernel I. Consequently every ideal of R is the kernel of some homomorphism of rings.

REMARKS. When I is an ideal, the ring R/I is called a **quotient ring**[10] of R, and the homomorphism $q : R \to R/I$ is called the **quotient homomorphism**. In the special case that $R = \mathbb{Z}$ and $I = m\mathbb{Z}$, the construction of R/I reduces to the construction of $\mathbb{Z}/m\mathbb{Z}$ as a ring at the beginning of this section.

PROOF. If we change the representatives of the cosets from r_1 and r_2 to $r_1 + i_1$ and $r_2 + i_2$ with i_1 and i_2 in I, then $(r_1 + i_1)(r_2 + i_2) = r_1r_2 + (i_1r_2 + r_1i_2 + i_1i_2)$ is in $r_1r_2 + I$ by the closure properties of I. Hence multiplication is well defined.

The associativity of this multiplication follows from the associativity of multiplication in R because

$$\begin{aligned}\big((r_1 + I)(r_2 + I)\big)(r_3 + I) &= (r_1r_2 + I)(r_3 + I) = (r_1r_2)r_3 + I = r_1(r_2r_3) + I \\ &= (r_1 + I)(r_2r_3 + I) = (r_1 + I)\big((r_2 + I)(r_3 + I)\big).\end{aligned}$$

Similarly the computation

$$\begin{aligned}(r_1 + I)\big((r_2 + I) + (r_3 + I)\big) &= r_1(r_2 + r_3) + I = (r_1r_2 + r_1r_3) + I \\ &= (r_1 + I)(r_2 + I) + (r_1 + I)(r_3 + I)\end{aligned}$$

yields one distributive law, and the other distributive law is proved in the same way. If R has an identity 1, then $(1 + I)(r + I) = 1r + I = r + I$ and $(r + I)(1 + I) = r1 + I = r + I$ show that $1 + I$ is an identity in R/I.

Finally we know that the quotient map $q : R \to R/I$ is a homomorphism of additive groups, and the computation $q(r_1r_2) = r_1r_2 + I = (r_1 + I)(r_2 + I) = q(r_1)q(r_2)$ shows that q is a homomorphism of rings. □

EXAMPLES OF IDEALS.

(1) The ideals in the ring $\mathbb{Z}$ coincide with the additive subgroups and are the sets $m\mathbb{Z}$; the reason each $m\mathbb{Z}$ is an ideal is that if a and b are integers and m divides a, then m divides ab.

[10] Quotient rings are known also as "factor rings." A "ring of quotients," however, is something different.

(2) The ideals in a field $\mathbb{F}$ are 0 and $\mathbb{F}$ itself, no others; in fact, if $a \neq 0$ is in an ideal and b is in $\mathbb{F}$, then the equality $b = (ba^{-1})a$ shows that b is in the ideal and that the ideal therefore contains all elements of $\mathbb{F}$.

(3) If R is $\mathbb{Q}[X]$ or $\mathbb{R}[X]$ or $\mathbb{C}[X]$, then every ideal I is of the form $I = Rf(X)$ for some polynomial $f(X)$. In fact, we can take $f(X) = 0$ if $I = 0$. If $I \neq 0$, let $f(X)$ be a nonzero member of I of lowest possible degree. If $A(X)$ is in I, then Proposition 1.12 shows that $A(X) = f(X)B(X) + C(X)$ with $C(X) = 0$ or $\deg C < \deg f$. The equality $C(X) = A(X) - f(X)B(X)$ shows that $C(X)$ is in I, and the minimality of $\deg f$ implies that $C(X) = 0$. Thus $A(X) = f(X)B(X)$.

(4) In a ring R with identity 1, an ideal I is a proper subset of R if and only if 1 is not in I. In fact, I is certainly a proper subset if 1 is not in I. In the converse direction if 1 is in I, then every element $r = r1$, for r in R, lies in I. Hence $I = R$, and I is not a proper subset.

In analogy with what was shown for vector spaces in Proposition 2.25 and for groups in Proposition 4.11, quotients in the context of rings allow for the factorization of certain homomorphisms of rings. The appropriate result is stated as Proposition 4.21 and is pictured in Figure 4.6.

Proposition 4.21. Let $\varphi : R_1 \to R_2$ be a homomorphism of rings, let $I_0 = \ker \varphi$, let I be an ideal of R_1 contained in I_0, and let $q : R_1 \to R_1/I$ be the quotient homomorphism. Then there exists a homomorphism of rings $\overline{\varphi} : R_1/I \to R_2$ such that $\varphi = \overline{\varphi} \circ q$, i.e., $\overline{\varphi}(r_1 + I) = \varphi(r_1)$. It has the same image as φ, and $\ker \overline{\varphi} = \{r + I \mid r \in I_0\}$.

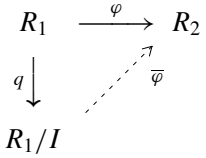


FIGURE 4.6. Factorization of homomorphisms of rings via the quotient of a ring by an ideal.

REMARK. One says that φ **factors through** R_1/I or **descends to** R_1/I.

PROOF. Proposition 4.11 shows that φ descends to a homomorphism $\overline{\varphi}$ of the additive group of R_1/I into the additive group of R_2 and that all the other conclusions hold except possibly for the fact that $\overline{\varphi}$ respects multiplication. To see that $\overline{\varphi}$ respects multiplication, we just compute that $\overline{\varphi}((r + I)(r' + I)) = \overline{\varphi}(rr' + I) = \varphi(rr') = \varphi(r)\varphi(r') = \overline{\varphi}(r + I)\overline{\varphi}(r' + I)$. □

An example of special interest occurs when φ is a homomorphism of rings $\varphi : \mathbb{Z} \to R$ and the ideal $m\mathbb{Z}$ of $\mathbb{Z}$ is contained in the kernel of φ. Then the proposition says that φ descends to a homomorphism of rings $\overline{\varphi} : \mathbb{Z}/m\mathbb{Z} \to R$. We shall make use of this result shortly. But first let us state a different special case as a corollary.

Corollary 4.22. Let $\varphi : R_1 \to R_2$ be a homomorphism of rings, and suppose that φ is onto R_2 and has kernel I. Then φ exhibits the ring R_1/I as canonically isomorphic to R_2.

PROOF. Take $I = I_0$ in Proposition 4.21, and form $\overline{\varphi} : R_1/I \to R_2$ with $\varphi = \overline{\varphi} \circ q$. The proposition shows that $\overline{\varphi}$ is onto R_2 and has trivial kernel, i.e., the identity element of R_1/I. Having trivial kernel, $\overline{\varphi}$ is one-one. □

Proposition 4.23. Any field $\mathbb{F}$ contains a subfield isomorphic to the rationals $\mathbb{Q}$ or to some field $\mathbb{F}_p$ with p prime.

REMARKS. The subfield in the proposition is called the **prime field** of $\mathbb{F}$. The **characteristic** of $\mathbb{F}$ is defined to be 0 if the prime field is isomorphic to $\mathbb{Q}$ and to be p if the prime field is isomorphic to $\mathbb{F}_p$.

PROOF. Proposition 4.19 produces a homomorphism of rings $\varphi_1 : \mathbb{Z} \to \mathbb{F}$ with $\varphi_1(1) = 1$. The kernel of φ_1 is an ideal, necessarily of the form $m\mathbb{Z}$ with m an integer ≥ 0, and the image of φ_1 is a commutative subring with identity in $\mathbb{F}$. Let $\overline{\varphi}_1 : \mathbb{Z}/m\mathbb{Z} \to \mathbb{F}$ be the descended homomorphism given by Proposition 4.21. The integer m cannot factor nontrivially, say as $m = rs$, because otherwise $\overline{\varphi}_1(r)$ and $\overline{\varphi}_1(s)$ would be nonzero members of $\mathbb{F}$ with $\overline{\varphi}_1(r)\overline{\varphi}_1(s) = \overline{\varphi}_1(rs) = \overline{\varphi}_1(0) = 0$, in contradiction to the fact that a field has no zero divisors.

Thus m is prime or m is 0. If m is a prime p, then $\mathbb{Z}/p\mathbb{Z}$ is a field, and the image of $\overline{\varphi}_1$ is the required subfield of $\mathbb{F}$. Thus suppose that $m = 0$. Then φ_1 is one-one, and $\mathbb{F}$ contains a subring with identity isomorphic to $\mathbb{Z}$. Define a function $\Phi_1 : \mathbb{Q} \to \mathbb{F}$ by saying that if k and l are integers with $l \neq 0$, then $\Phi_1(kl^{-1}) = \varphi_1(k)\varphi_1(l)^{-1}$. This is well defined because $\varphi_1(l) \neq 0$ and because $k_1l_1^{-1} = k_2l_2^{-1}$ implies $k_1l_2 = k_2l_1$ and hence $\varphi_1(k_1)\varphi_1(l_2) = \varphi_1(k_2)\varphi_1(l_1)$ and $\varphi_1(k_1)\varphi_1(l_1)^{-1} = \varphi_1(k_2)\varphi_1(l_2)^{-1}$. We readily check that Φ_1 is a homomorphism with kernel 0. Then $\mathbb{F}$ contains the subfield $\Phi_1(\mathbb{Q})$ isomorphic to $\mathbb{Q}$. □

5. Polynomials and Vector Spaces

In this section we complete the digression begun in Section 4. We shall be using the elementary notions of rings and fields established in Section 4 in order to

work with (i) polynomials over any commutative ring with identity and (ii) vector spaces over arbitrary fields.

It is an important observation that a good deal of what has been proved so far in this book concerning polynomials when $\mathbb{F}$ is $\mathbb{Q}$ or $\mathbb{R}$ or $\mathbb{C}$ remains valid when $\mathbb{F}$ is any field. Specifically all the results in Section I.3 through Theorem 1.17 on the topic of polynomials in one indeterminate remain valid as long as the coefficients are from a field. The theory breaks down somewhat when one tries to extend it by allowing coefficients that are not in a field or by allowing more than one indeterminate. Because of this circumstance and because we have not yet announced a universal mapping property for polynomial rings and because we have not yet addressed the several-variable case, we shall briefly review matters now while extending the reach of the theory that we have.

Let R be a nonzero commutative ring with identity, so that $1 \neq 0$. A polynomial in one indeterminate is to be an expression $P(X) = a_n X^n + \cdots + a_2 X^2 + a_1 X + a_0$ in which X is a symbol, not a variable. Nevertheless, the usual kinds of manipulations with polynomials are to be valid. This description lacks precision because X has not really been defined adequately. To make a precise definition, we remove X from the formalism and simply define the polynomial to be the tuple $(a_0, a_1, \ldots, a_n, 0, 0, \ldots)$ of its coefficients. Thus a **polynomial** in **one indeterminate** with **coefficients** in R is an infinite sequence of members of R such that all terms of the sequence are 0 from some point on. The indexing of the sequence is to begin with 0, and X is to refer to the polynomial $(0, 1, 0, 0, \ldots)$. We may refer to a polynomial P as $P(X)$ if we want to emphasize that the indeterminate is called X. Addition and negation of polynomials are defined in coordinate-by-coordinate fashion by

$$\begin{aligned}(a_0, a_1, \ldots, a_n, 0, 0, \ldots) + (b_0, b_1, \ldots, b_n, 0, 0, \ldots)& \\ = (a_0 + b_0, a_1 + b_1, \ldots, a_n + b_n, 0, 0, \ldots),& \\ -(a_0, a_1, \ldots, a_n, 0, 0, \ldots) = (-a_0, -a_1, \ldots, -a_n, 0, 0, \ldots),&\end{aligned}$$

and the set $R[X]$ of polynomials is then an abelian group isomorphic to the direct sum of infinitely many copies of the additive group of R. As in Section I.3, X^n is to be the polynomial whose coefficients are 1 in the n^{th} position, with $n \geq 0$, and 0 in all other positions. Polynomial multiplication is then defined so as to match multiplication of expressions $a_n X^n + \cdots + a_1 X + a_0$ if the product is expanded out, powers of X are added, and the terms containing like powers of X are collected. Thus the precise definition is that

$$(a_0, a_1, \ldots, 0, 0, \ldots)(b_0, b_1, \ldots, 0, 0, \ldots) = (c_0, c_1, \ldots, 0, 0, \ldots),$$

where $c_N = \sum_{k=0}^{N} a_k b_{N-k}$. It is a simple matter to check that this multiplication makes $R[X]$ into a commutative ring.

The polynomial with all entries 0 is denoted by 0 and is called the **zero polynomial**. For all polynomials $P = (a_0, \dots, a_n, 0, \dots)$ other than 0, the **degree** of P, denoted by $\deg P$, is defined to be the largest index n such that $a_n \neq 0$. In this case, a_n is called the **leading coefficient**, and $a_n X^n$ is called the **leading term**; if $a_n = 1$, the polynomial is called **monic**. The usual convention with the 0 polynomial is either to leave its degree undefined or to say that the degree is $-\infty$; let us follow the latter approach in this section in order not to have to separate certain formulas into cases.

There is a natural one-one homomorphism of rings $\iota : R \to R[X]$ given by $\iota(c) = (c, 0, 0, \dots)$ for c in R. This sends the identity of R to the identity of $R[X]$. Thus we can identify R with the **constant polynomials**, i.e., those of degree ≤ 0.

If P and Q are nonzero polynomials, then

$$\deg(P + Q) \leq \max(\deg P, \deg Q).$$

In this formula equality holds if $\deg P \neq \deg Q$. In the case of multiplication, let P and Q have respective leading terms $a_m X^m$ and $b_n X^n$. All the coefficients of PQ are 0 beyond the $(m+n)^{\text{th}}$, and the $(m+n)^{\text{th}}$ is $a_m b_n$. This in principle could be 0 but is nonzero if R is an integral domain. Thus P and Q nonzero implies

$$\deg(PQ) \begin{cases} \leq \deg P + \deg Q & \text{for general } R, \\ = \deg P + \deg Q & \text{if } R \text{ is an integral domain.} \end{cases}$$

It follows in particular that $R[X]$ is an integral domain if R is.

Normally we shall write out specific polynomials using the informal notation with powers of X, using the more precise notation with tuples only when some ambiguity might otherwise result.

In the special case that R is a field, Section I.3 introduced the notation of evaluation of a polynomial $P(X)$ at a point r in the field, thus providing a mapping $P(X) \mapsto P(r)$ from $R[X]$ to R for each r in R. We listed a number of properties of this mapping, and they can be summarized in our present language by the statement that the mapping is a homomorphism of rings. Evaluation is a special case of a more sweeping property of polynomials given in the next proposition as a **universal mapping property** of $R[X]$.

Proposition 4.24. Let R be a nonzero commutative ring with identity, and let $\iota : R \to R[X]$ be the identification of R with constant polynomials. If T is any commutative ring with identity, if $\varphi : R \to T$ is a homomorphism of rings sending 1 into 1, and if t is in T, then there exists a unique homomorphism of rings $\Phi : R[X] \to T$ carrying identity to identity such that $\Phi(\iota(r)) = \varphi(r)$ for all $r \in R$ and $\Phi(X) = t$.

REMARKS. The mapping Φ is called the **substitution homomorphism** extending φ and substituting t for X, and the mapping is written $P(X) \mapsto P^{\varphi}(t)$. The notation means that φ is to be applied to the coefficients of P and then X is to be replaced by t. A diagram of this homomorphism as a universal mapping property appears in Figure 4.7. In the special case that $T = R$ and φ is the identity, Φ reduces to **evaluation** at t, and the mapping is written $P(X) \mapsto P(t)$, just as in Section I.3.

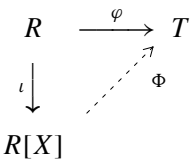


FIGURE 4.7. Substitution homomorphism for polynomials in one indeterminate.

PROOF. Define $\Phi(a_0, a_1, \dots, a_n, 0, \dots) = \varphi(a_0) + \varphi(a_1)t + \cdots + \varphi(a_n)t^n$. It is immediate that Φ is a homomorphism of rings sending the identity $\iota(1) = (1, 0, 0, \dots)$ of $R[X]$ to the identity $\varphi(1)$ of T. If r is in R, then $\Phi(\iota(r)) = \Phi(r, 0, 0, \dots) = \varphi(r)$. Also, $\Phi(X) = \Phi(0, 1, 0, 0, \dots) = t$. This proves existence. Uniqueness follows since $\iota(R)$ and X generate $R[X]$ and since a homomorphism defined on $R[X]$ is therefore determined by its values on $\iota(R)$ and X. □

The formulation of the proposition with the general $\varphi : R \to T$, rather than just the identity mapping on R, allows several kinds of applications besides the routine evaluation mapping. An example of one kind occurs when $R = \mathbb{C}$, $T = \mathbb{C}[X]$, and $\varphi : \mathbb{C} \to \mathbb{C}[X]$ is the composition of complex conjugation on $\mathbb{C}$ followed by the identification of complex numbers with constant polynomials in $\mathbb{C}[X]$; the proposition then says that complex conjugation of the coefficients of a member of $\mathbb{C}[X]$ is a ring homomorphism. This observation simplifies the solution of Problem 7 in Chapter I. Similarly one can set up matters so that the proposition shows the passage from $\mathbb{Z}[X]$ to $(\mathbb{Z}/m\mathbb{Z})[X]$ by reduction of coefficients modulo m to be a ring homomorphism.

Still a third kind of application is to take T in the proposition to be a ring with the same kind of universal mapping property that $R[X]$ has, and the consequence is an abstract characterization of $R[X]$. We carry out the details below as Proposition 4.25. This result will be applied later in this section to the several-indeterminate case to show that introducing several indeterminates at once yields the same ring, up to canonical isomorphism, as introducing them one at a time.

Proposition 4.25. Let R and S be nonzero commutative rings with identity, let X' be an element of S, and suppose that $\iota' : R \to S$ is a one-one ring

homomorphism of R into S carrying 1 to 1. Suppose further that (S, ι', X') has the following property: whenever T is a commutative ring with identity, $\varphi : R \to T$ is a homomorphism of rings sending 1 into 1, and t is in T, then there exists a unique homomorphism $\Phi' : S \to T$ carrying identity to identity such that $\Phi'(\iota'(r)) = \varphi(r)$ for all $r \in R$ and $\Phi'(X') = t$. Then there exists a unique homomorphism of rings $\Psi : R[X] \to S$ such that $\Psi \circ \iota = \iota'$ and $\Psi(X) = X'$, and Ψ is an isomorphism.

REMARK. A somewhat weaker conclusion than in the proposition is that any triple (S, ι', X') having the same universal mapping property as $(R[X], \iota, X)$ is isomorphic to (S, ι', X'), the isomorphism being unique.

PROOF. In the universal mapping property for S, take $T = R[X]$, $\varphi = \iota$, and $t = X$. The hypothesis gives us a ring homomorphism $\Phi' : S \to R[X]$ with $\Phi'(1) = 1$, $\Phi' \circ \iota' = \iota$, and $\Phi'(X') = X$. Next apply Proposition 4.24 with $T = S$, $\varphi = \iota'$, and $t = X'$. We obtain a ring homomorphism $\Phi : R[X] \to S$ with $\Phi(1) = 1$, $\Phi \circ \iota = \iota'$, and $\Phi(X) = X'$. Then $\Phi' \circ \Phi$ is a ring homomorphism from $R[X]$ to itself carrying 1 to 1, fixing X, and having $\Phi' \circ \Phi\big|_{\iota(R)} = \iota$. From the uniqueness in Proposition 4.24 when $T = R[X]$, $\varphi = \iota$, and $t = X$, we see that $\Phi' \circ \Phi$ is the identity on $R[X]$. Reversing the roles of Φ and Φ' and applying the uniqueness in the universal mapping property for S, we see that $\Phi \circ \Phi'$ is the identity on S. Therefore Φ may be taken as the isomorphism Ψ in the statement of the proposition. This proves existence for Ψ, and uniqueness follows since $\iota(R)$ and X together generate $R[X]$ and since Ψ is a homomorphism. □

If P is a polynomial over R in one indeterminate and r is in R, then r is a **root** of P if $P(r) = 0$. We know as a consequence of Corollary 1.14 that for any prime p, any polynomial in $\mathbb{F}_p[X]$ of degree $n \geq 1$ has at most n roots. This result does not extend to $\mathbb{Z}/m\mathbb{Z}$ for all positive integers m: when $m = 8$, the polynomial $X^2 - 1$ has 4 roots, namely 1, 3, 5, 7. This result about $\mathbb{F}_p[X]$ has the following consequence.

Proposition 4.26. If $\mathbb{F}$ is a field, then any finite subgroup of the multiplicative group $\mathbb{F}^\times$ is cyclic.

PROOF. Let C be a subgroup of $\mathbb{F}^\times$ of finite order n. Lagrange's Theorem (Corollary 4.8) shows that the order of each element of C divides n. With h defined as the maximum order of an element of C, it is enough to show that $h = n$. Let a be an element of order h. The polynomial $X^h - 1$ has at most h roots by Corollary 1.14, and a is one of them, by definition of "order." If $h < n$, then it follows that some member b of C is not a root of $X^h - 1$. The order h' of b is then a divisor of n but cannot be a divisor of h since otherwise we would have $b^h = (b^{h'})^{h/h'} = 1^{h/h'} = 1$. Consequently there exists a prime p such that

some power p^r of p divides h' but not h. Let $s < r$ be the exact power of p dividing h, and write $h = mp^s$, so that $\text{GCD}(m, p^r) = 1$ and $a' = a^{p^s}$ has order m. Put $q = h'/p^r$, so that $b' = b^q$ has order p^r. The proof will be completed by showing that $c = a'b'$ has order $mp^r = hp^{r-s} > h$, in contradiction to the maximality of h.

Let t be the order of c. On the one hand, from $c^{mp^r} = (a')^{mp^r}(b')^{mp^r} = a^{hp^{r+s}}b^{mp^rq} = a^{hp^{r+s}}b^{mh'} = (a^h)^{p^{r+s}}(b^{h'})^m = 1$, we see that t divides mp^r. On the other hand, $1 = c^t$ says that $(a')^t = (b')^{-t}$. Raising both sides to the p^r power gives $1 = ((b')^{p^r})^{-t} = (a')^{tp^r}$, and hence m divides tp^r; by Corollary 1.3, m divides t. Raising both sides of $(a')^t = (b')^{-t}$ to the m^{th} power gives $1 = ((a')^m)^t = (b')^{-tm}$, and hence p^r divides tm; by Corollary 1.3, p^r divides t. Applying Corollary 1.4, we conclude that mp^r divides t. Therefore $t = mp^r$, and the proof is complete. □

Corollary 4.27. The multiplicative group of a finite field is cyclic.

PROOF. This is a special case of Proposition 4.26. □

A finite field $\mathbb{F}$ can have a nonzero polynomial that is 0 at every element of $\mathbb{F}$. Indeed, every element of $\mathbb{F}_p$ is a root of $X^p - X$, as a consequence of Fermat's Little Theorem. It is for this reason that it is unwise to confuse a polynomial in an indeterminate with a "polynomial function."

Let us make the notion of a polynomial function of one variable rigorous. If $P(X)$ is a polynomial with coefficients in the commutative ring R with identity, then Proposition 4.24 gives us an evaluation homomorphism $P \mapsto P(r)$ for each r in R. The function $r \mapsto P(r)$ from R into R is the **polynomial function** associated to the polynomial P. This function is a member of the commutative ring of all R-valued functions on R, and the mapping $P \mapsto \big(r \mapsto P(r)\big)$ is a homomorphism of rings. What we know from Corollary 1.14 is that this homomorphism is one-one if R is an infinite field. A negative result is that if R is a finite commutative ring with identity, then $\prod_{r \in R}(X - r)$ is a polynomial that maps to the 0 function, and hence the homomorphism is not one-one. A more general positive result than the one above for infinite fields is the following.

Proposition 4.28.

(a) If R is a nonzero commutative ring with identity and $P(X)$ is a member of $R[X]$ with a root r, then $P(X) = (X - r)Q(X)$ for some $Q(X)$ in $R[X]$.

(b) If R is an integral domain, then a nonzero member of $R[X]$ of degree n has at most n roots.

(c) If R is an infinite integral domain, then the ring homomorphism of $R[X]$ to the ring of polynomial functions from R to R, given by evaluation, is one-one.

PROOF. For (a), we proceed by induction on the degree of P, the base case of the induction being degree ≤ 0. If the conclusion has been proved for degree $< n$ with $n \ge 1$, let the leading term of P be $a_n X^n$. Then $P(X) = a_n(X-r)^n + A(X)$ with $\deg A < n$. Evaluation at r gives, by virtue of Proposition 4.24, $0 = 0 + A(r)$. By the inductive hypothesis, $A(X) = (X-r)B(X)$. Then $P(X) = (X-r)Q(X)$ with $Q(X) = a_n(X-r)^{n-1} + B(X)$, and the induction is complete.

For (b), let $P(X)$ have degree n with at least $n+1$ distinct roots $r_1, \dots, r_{n+1}$. Part (a) shows that $P(X) = (X-r_1)P_1(X)$ with $\deg P_1 = n-1$. Also, $0 = P(r_2) = (r_2 - r_1)P_1(r_2)$. Since $r_2 - r_1 \ne 0$ and since R has no zero divisors, $P_1(r_2) = 0$. Part (a) then shows that $P_1(X) = (X-r_2)P_2(X)$, and substitution gives $P(X) = (X-r_1)(X-r_2)P_2(X)$. Continuing in this way, we obtain $P(X) = (X-r_1)\cdots(X-r_n)P_n(X)$ with $\deg P_n = 0$. Since $P \ne 0$, $P_n \ne 0$. So P_n is a nonzero constant polynomial $P_n(X) = c \ne 0$. Evaluating at r_{n+1}, we obtain $0 = (r_{n+1} - r_1)\cdots(r_{n+1} - r_n)c$ with each factor nonzero, in contradiction to the fact that R is an integral domain.

For (c), a polynomial in the kernel of the ring homomorphism has every member of R as a root. If R is infinite, (b) shows that such a polynomial is necessarily the zero polynomial. Thus the kernel is 0, and the ring homomorphism has to be one-one. □

Let us turn our attention to polynomials in several indeterminates. Fix the nonzero commutative ring R with identity, and let n be a positive integer. Informally a polynomial over R in n indeterminates is to be a finite sum

$$\sum_{j_1 \ge 0, \dots, j_n \ge 0} r_{j_1,\dots,j_n} X_1^{j_1} \cdots X_n^{j_n}$$

with each $r_{j_1,\dots,j_n}$ in R. To make matters precise, we work just with the system of coefficients, just as in the case of one indeterminate.

Let J be the set of integers ≥ 0, and let J^n be the set of n-tuples of elements of J. A member of J^n may be written as $j = (j_1, \dots, j_n)$. Addition of members of J^n is defined coordinate by coordinate. Thus $j + j' = (j_1 + j_1', \dots, j_n + j_n')$ if $j = (j_1, \dots, j_n)$ and $j' = (j_1', \dots, j_n')$. A **polynomial** in n **indeterminates** with **coefficients** in R is a function $f : J^n \to R$ such that $f(j) \ne 0$ for only finitely many $j \in J^n$. Temporarily let us write S for the set of all such polynomials for a particular n. If f and g are two such polynomials, their sum h and product k are the polynomials defined by

$$h(j) = f(j) + g(j),$$
$$k(i) = \sum_{j+j'=i} f(j)g(j').$$

Under these definitions, S is a commutative ring.

Define a mapping $\iota : R \to S$ by

$$\iota(r)(j) = \begin{cases} r & \text{if } j = (0, \ldots, 0), \\ 0 & \text{otherwise.} \end{cases}$$

Then ι is a one-one homomorphism of rings, $\iota(0)$ is the zero element of S and is called simply 0, and $\iota(1)$ is a multiplicative identity for S. The polynomials in the image of ι are called the **constant polynomials**.

For $1 \leq k \leq n$, let e_k be the member of J^n that is 1 in the k^{th} place and is 0 elsewhere. Define X_k to be the polynomial that assigns 1 to e_k and assigns 0 to all other members of J^n. We say that X_k is an **indeterminate**. If $j = (j_1, \ldots, j_n)$ is in J^n, define X^j to be the product

$$X^j = X_1^{j_1} \cdots X_n^{j_n}.$$

If r is in R, we allow ourselves to abbreviate $\iota(r)X^j$ as rX^j, and any such polynomial is called a **monomial**. The monomial rX^j is the polynomial that assigns r to j and assigns 0 to all other members of J^n. Then it follows immediately from the definitions that each polynomial has a unique expansion as a finite sum of nonzero monomials. Thus the most general member of S is of the form $\sum_{j \in J^n} r_j X^j$ with only finitely many nonzero terms. This is called the **monomial expansion** of the given polynomial.

We may now write $R[X_1, \ldots, X_n]$ for S. A polynomial $\sum_{j \in J^n} r_j X^j$ may be conveniently abbreviated as P or as $P(X)$ or as $P(X_1, \ldots, X_n)$ when its monomial expansion is either understood or irrelevant.

The **degree** of the 0 polynomial is defined for this section to be $-\infty$, and the degree of any monomial rX^j with $r \neq 0$ is defined to be the integer

$$|j| = j_1 + \cdots + j_n \qquad \text{if } j = (j_1, \ldots, j_n).$$

Finally the **degree** of any nonzero polynomial P, denoted by deg P, is defined to be the maximum of the degrees of the terms in its monomial expansion. If all the nonzero monomials in the monomial expansion of a polynomial P have the same degree d, then P is said to be **homogeneous** of degree d. Under these definitions the 0 polynomial has degree $-\infty$ but is homogeneous of every degree. If P and Q are homogeneous polynomials of degrees d and d', then PQ is homogeneous of degree dd' (and possibly equal to the 0 polynomial).

In any event, by grouping terms in the monomial expansion of a polynomial according to their degree, we see that every polynomial is uniquely the sum of nonzero homogeneous polynomials of distinct degrees. Let us call this the **homogeneous-polynomial expansion** of the given polynomial. Let us expand two such nonzero polynomials P and Q in this fashion, writing $P = P_{d_1} + \cdots + P_{d_k}$

and $Q = Q_{d'_1} + \cdots + Q_{d'_l}$ with $d_1 < \cdots < d_k$ and $d'_1 < \cdots < d'_l$. Then we see directly that

$$\deg(P + Q) \le \max(\deg P, \deg Q),$$

$$\deg(PQ) \le \deg P + \deg Q.$$

In the formula for $\deg(P + Q)$, the term that is potentially of largest degree is $P_{d_k} + Q_{d'_l}$, and it is of degree $\max(\deg P, \deg Q)$ if $\deg P \ne \deg Q$. In the formula for $\deg(PQ)$, the term that is potentially of largest degree is $P_{d_k} Q_{d'_l}$. It is homogeneous of degree $d_k + d'_l$, but it could be 0. Some proof is required that it is not 0 if R is an integral domain, as follows.

Proposition 4.29. If R is an integral domain, then $R[X_1, \ldots, X_n]$ is an integral domain.

PROOF. Let P and Q be nonzero homogeneous polynomials with $\deg P = d$ and $\deg Q = d'$. We are to prove that $PQ \ne 0$. We introduce an ordering on the set of all members j of J^n, saying $j = (j_1, \ldots, j_n) > j' = (j'_1, \ldots, j'_n)$ if there is some k such that $j_i = j'_i$ for $i < k$ and $j_k > j'_k$. In the monomial expansion of P as $P(X) = \sum_{|j|=d} a_j X^j$, let i be the largest n-tuple j in the ordering such that $a_j \ne 0$. Similarly with $Q(X) = \sum_{|j'|=d'} b_{j'} X^{j'}$, let i' be the largest n-tuple j' in the ordering such that $b_{j'} \ne 0$. Then

$$P(X)Q(X) = a_i b_{i'} X^{i+i'} + \sum_{\substack{j, j' \text{ with} \\ (j, j') \ne (i, i')}} a_j b_{j'} X^{j+j'},$$

and all terms in the sum $\sum_{j, j'}$ on the right side have $j + j' < i + i'$. Thus $a_i b_{i'} X^{i+i'}$ is the only term in the monomial expansion of $P(X)Q(X)$ involving the monomial $X^{i+i'}$. Since R is an integral domain and a_i and $b_{i'}$ are nonzero, $a_i b_{i'}$ is nonzero. Thus $P(X)Q(X)$ is nonzero. □

Proposition 4.30. Let R be a nonzero commutative ring with identity, let $R[X_1, \ldots, X_n]$ be the ring of polynomials in n indeterminates, and define $\iota : R \to R[X_1, \ldots, X_n]$ to be the identification of R with constant polynomials. If T is any commutative ring with identity, if $\varphi : R \to T$ is a homomorphism of rings sending 1 into 1, and if $t_1, \ldots, t_n$ are in T, then there exists a unique homomorphism $\Phi : R[X_1, \ldots, X_n] \to T$ carrying identity to identity such that $\Phi(\iota(r)) = \varphi(r)$ for all $r \in R$ and $\Phi(X_j) = t_j$ for $1 \le j \le n$.

REMARKS. The mapping Φ is called the **substitution homomorphism** extending φ and substituting t_j for X_j for $1 \le j \le n$, and the mapping is written $P(X_1, \ldots, X_n) \mapsto P^{\varphi}(t_1, \ldots, t_n)$. The notation means that φ is to be applied to each coefficient of P and then $X_1, \ldots, X_n$ are to be replaced by $t_1, \ldots, t_n$.

A diagram of this homomorphism as a **universal mapping property** appears in Figure 4.8. In the special case that $T = R \times \cdots \times R$ (cf. Example 3 of homomorphisms in Section 4) and φ is the identity, Φ reduces to **evaluation** at $(t_1, \ldots, t_n)$, and the mapping is written $P(X_1, \ldots, X_n) \mapsto P(t_1, \ldots, t_n)$.

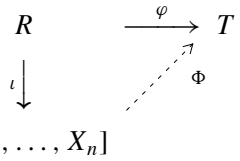


FIGURE 4.8. Substitution homomorphism for polynomials in n indeterminates.

PROOF. If $P(X_1, \ldots, X_n) = \sum_{j_1 \geq 0, \ldots, j_n \geq 0} a_{j_1, \ldots, j_n} X_1^{j_1} \cdots X_n^{j_n}$ is the monomial expansion of a member P of $R[X_1, \ldots, X_n]$, then $\Phi(P)$ is defined to be the corresponding finite sum $\sum_{j_1 \geq 0, \ldots, j_n \geq 0} a_{j_1, \ldots, j_n} t_1^{j_1} \cdots t_n^{j_n}$. Existence readily follows, and uniqueness follows since $\iota(R)$ and $X_1, \ldots, X_n$ generate $R[X_1, \ldots, X_n]$ and since Φ is a homomorphism. □

Corollary 4.31. If R is a nonzero commutative ring with identity, then $R[X_1, \ldots, X_{n-1}][X_n]$ is isomorphic as a ring to $R[X_1, \ldots, X_n]$.

REMARK. The proof will show that the isomorphism is the expected one.

PROOF. In the notation with n-tuples and J^n, any $(n-1)$-tuple may be identified with an n-tuple by adjoining 0 as its n^{th} coordinate, and in this way, every monomial in $R[X_1, \ldots, X_{n-1}]$ can be regarded as a monomial in $R[X_1, \ldots, X_n]$. The extension of this mapping to sums gives us a one-one homomorphism of rings $\iota' : R[X_1, \ldots, X_{n-1}] \to R[X_1, \ldots, X_n]$. We are going to use Proposition 4.25 to prove the isomorphism of rings $R[X_1, \ldots, X_{n-1}][X_n] \cong R[X_1, \ldots, X_n]$. In the notation of that proposition, the role of R is played by $R[X_1, \ldots, X_{n-1}]$, we take $S = R[X_1, \ldots, X_n]$, and we have constructed ι'. We are to show that (S, ι', X_n) satisfies a certain universal mapping property. Thus suppose that T is a commutative ring with identity, that t is in T, and that $\varphi' : R[X_1, \ldots, X_{n-1}] \to T$ is a homomorphism of rings carrying identity to identity.

We shall apply Proposition 4.30 in order to obtain the desired homomorphism $\Phi' : S \to T$. Let $\iota_{n-1} : R \to R[X_1, \ldots, X_{n-1}]$ be the identification of R with constant polynomials in $R[X_1, \ldots, X_{n-1}]$, and let $\iota_n = \iota' \circ \iota_{n-1}$ be the identification of R with constant polynomials in S. Define $\varphi : R \to T$ by $\varphi = \varphi' \circ \iota_{n-1}$, and take $t_n = t$ and $t_j = \varphi'(X_j)$ for $1 \leq j \leq n-1$. Then Proposition 4.30 produces a homomorphism of rings $\Phi' : S \to T$ with $\Phi'(\iota_n(r)) = \varphi(r)$ for $r \in R$, $\Phi'(\iota'(X_j)) = \varphi'(X_j)$ for $1 \leq j \leq n-1$, and $\Phi'(X_n) = t_n$. The equations

$$\Phi'(\iota'(\iota_{n-1}(r))) = \Phi'(\iota_n(r)) = \varphi(r) = \varphi'(\iota_{n-1}(r))$$

and

$$\Phi'(\iota'(X_j)) = \varphi'(X_j)$$

show that $\Phi' \circ \iota' = \varphi'$ on $R[X_1, \ldots, X_n]$. Also, $\Phi'(X_n) = t_n = t$. Thus the mapping Φ' sought by Proposition 4.25 exists. It is unique since $R[X_1, \ldots, X_{n-1}]$ and X_n together generate S. The conclusion from Proposition 4.25 is that S is isomorphic to $R[X_1, \ldots, X_{n-1}][X_n]$ via the expected isomorphism of rings. □

We conclude the discussion of polynomials in several variables by making the notion of a polynomial function of several variables rigorous. If $P(X_1, \ldots, X_n)$ is a polynomial in n indeterminates with coefficients in the commutative ring R with identity, then Proposition 4.30 gives us an evaluation homomorphism $P \mapsto P(r_1, \ldots, r_n)$ for each n-tuple $(r_1, \ldots, r_n)$ of members of R. The function $(r_1, \ldots, r_n) \mapsto P(r_1, \ldots, r_n)$ from $R \times \cdots \times R$ into R is the **polynomial function** associated to the polynomial P. This function is a member of the commutative ring of all R-valued functions on $R \times \cdots \times R$, and the mapping $P \mapsto \big((r_1, \ldots, r_n) \mapsto P(r_1, \ldots, r_n)\big)$ is a homomorphism of rings.

Corollary 4.32. If R is an infinite integral domain, then the ring homomorphism of $R[X_1, \ldots, X_n]$ to polynomial functions from $R \times \cdots \times R$ to R, given by evaluation, is one-one.

REMARK. This result extends Proposition 4.28 to several indeterminates.

PROOF. We proceed by induction on n, the case $n = 1$ being handled by Proposition 4.28. Assume the result for $n - 1$ indeterminates. If $P \neq 0$ is in $R[X_1, \ldots, X_n]$, Corollary 4.31 allows us to write

$$P(X_1, \ldots, X_n) = \sum_{i=1}^{k} P_i(X_1, \ldots, X_{n-1}) X_n^i$$

for some k, with each P_i in $R[X_1, \ldots, X_{n-1}]$ and with $P_k(X_1, \ldots, X_{n-1}) \neq 0$. By the inductive hypothesis, $P_k(r_1, \ldots, r_{n-1})$ is nonzero for some elements $r_1, \ldots, r_{n-1}$ of R. So the polynomial $\sum_{i=0}^{k} P_i(r_1, \ldots, r_{n-1}) X_n^i$ in $R[X_n]$ is not the 0 polynomial, and Proposition 4.28 shows that it is not 0 when evaluated at some r_n. Then $P(r_1, \ldots, r_n) \neq 0$. □

It is possible also to introduce polynomial rings in infinitely many variables. These will play roles only as counterexamples in this book, and thus we shall not stop to treat them in detail.

We complete this section with some remarks about vector spaces. The definition of a **vector space** over a general field $\mathbb{F}$ remains the same as in Section II.1, where $\mathbb{F}$ is assumed to be $\mathbb{Q}$ or $\mathbb{R}$ or $\mathbb{C}$. We shall make great use of the fact that all the results in Chapter II concerning vector spaces remain valid when $\mathbb{Q}$ or $\mathbb{R}$ or

$\mathbb{C}$ is replaced by a general field $\mathbb{F}$. The proofs need no adjustments, and it is not necessary to write out the details. For the moment we make only the following application of vector spaces over general fields, but the extended theory of vector spaces will play an important role in most of the remaining chapters of this book.

Proposition 4.33. If $\mathbb{F}$ is a finite field, then the number of elements in $\mathbb{F}$ is a power of a prime.

REMARK. We return to this matter in Chapter IX, showing at that time that for each prime power $p^n > 1$, there is one and only one field with p^n elements, up to isomorphism.

PROOF. The characteristic of $\mathbb{F}$ cannot be 0 since $\mathbb{F}$ is finite, and hence it is some prime p. Denote the prime field of $\mathbb{F}$ by $\mathbb{F}_p$. By restricting the multiplication so that it is defined only on $\mathbb{F}_p \times \mathbb{F}$, we make $\mathbb{F}$ into a vector space over $\mathbb{F}_p$, necessarily finite-dimensional. Proposition 2.18 shows that $\mathbb{F}$ is isomorphic as a vector space to the space $(\mathbb{F}_p)^n$ of n-dimensional column vectors for some n, and hence $\mathbb{F}$ must have p^n elements. □

6. Group Actions and Examples

Let X be a nonempty set, let $\mathcal{F}(X)$ be the group of invertible functions from X onto itself, the group operation being composition, and let G be a group. A **group action** of G on X is a homomorphism of G into $\mathcal{F}(X)$. When $X = \{1, \dots, n\}$, the group $\mathcal{F}(X)$ is just the symmetric group $\mathfrak{S}_n$. Thus Examples 5–9 of groups in Section 1 are all in fact subgroups of various groups $\mathcal{F}(X)$ and are therefore examples of group actions. Thus every group of permutations of $\{1, \dots, n\}$, every dihedral group acting on $\mathbb{R}^2$, and every general linear group or subgroup acting on a finite-dimensional vector space over $\mathbb{Q}$ or $\mathbb{R}$ or $\mathbb{C}$ or an arbitrary field $\mathbb{F}$ provides an example. So do the orthogonal and unitary groups acting on $\mathbb{R}^n$ and $\mathbb{C}^n$, as well as the automorphism group of any number field.

We saw an indication in Section 1 that many early examples of groups arose in this way. One source of examples that is of some importance and was not listed in Section 1 occurs in the geometry of $\mathbb{R}^2$. The translations in $\mathbb{R}^2$, together with the rotations about arbitrary points of $\mathbb{R}^2$ and the reflections about arbitrary lines in $\mathbb{R}^2$, form a group G of rigid motions of the plane.[11] This group G is a subgroup of $\mathcal{F}(\mathbb{R}^2)$, and thus G acts on $\mathbb{R}^2$. More generally, whenever a nonempty set X has a notion of distance, the set of **isometries** of X, i.e., the distance-preserving members of $\mathcal{F}(X)$, forms a subgroup of $\mathcal{F}(X)$, and thus the group of isometries of X acts on X.

[11] One can show that G is the full group of rigid motions of $\mathbb{R}^2$, but this fact will not concern us.

At any rate a group action τ of G on X, being a homomorphism of G into $\mathcal{F}(X)$, is of the form $g \mapsto \tau_g$, where τ_g is in $\mathcal{F}(X)$ and $\tau_{g_1 g_2} = \tau_{g_1}\tau_{g_2}$. There is an equivalent way of formulating matters that does not so obviously involve the notion of a homomorphism. Namely, we write $\tau_g(x) = gx$. In this notation the group action becomes a function $G \times X \to X$ with $(g, x) \mapsto gx$ such that

(i) $(g_1 g_2)x = g_1(g_2 x)$ for all g_1 and g_2 in G and for all x in X (from the fact that $\tau_{g_1 g_2} = \tau_{g_1}\tau_{g_2}$),
(ii) $1x = x$ for all x in X (from the fact that $\tau_1 = 1$).

Conversely if $G \times X \to X$ satisfies (i) and (ii), then the formulas $x = 1x = (gg^{-1})x = g(g^{-1}x)$ and $x = 1x = (g^{-1}g)x = g^{-1}(gx)$ show that the function $x \mapsto gx$ from X to itself is invertible with inverse $x \mapsto g^{-1}x$. Consequently the definition $\tau_g(x) = gx$ makes $g \mapsto \tau_g$ a function from G into $\mathcal{F}(X)$, and (i) shows that τ is a homomorphism. Thus (i) and (ii) indeed give us an equivalent formulation of the notion of a group action. Both formulations are useful.

Quite often the homomorphism $G \to \mathcal{F}(X)$ of a group action is one-one, and then G can be regarded as a subgroup of $\mathcal{F}(X)$. Here is an important geometric example in which the homomorphism is not one-one.

EXAMPLE. Linear fractional transformations. Let $X = \mathbb{C} \cup \{\infty\}$, a set that becomes the **Riemann sphere** in complex analysis. The group $G = \mathrm{GL}(2, \mathbb{C})$ acts on X by the **linear fractional transformations**

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}(z) = \frac{az + b}{cz + d},$$

the understanding being that the image of ∞ is ac^{-1} and the image of $-dc^{-1}$ is ∞, just as if we were to pass to a limit in each case. Property (ii) of a group action is clear. To verify (i), we simply calculate that

$$\begin{aligned}
\begin{pmatrix} a' & b' \\ c' & d' \end{pmatrix}\left(\begin{pmatrix} a & b \\ c & d \end{pmatrix}(z)\right) &= \frac{a'\left(\frac{az+b}{cz+d}\right) + b'}{c'\left(\frac{az+b}{cz+d}\right) + d'} \\
&= \frac{(a'a + b'c)z + (a'b + b'd)}{(c'a + d'c)z + (c'b + d'd)} \\
&= \left(\begin{pmatrix} a' & b' \\ c' & d' \end{pmatrix}\begin{pmatrix} a & b \\ c & d \end{pmatrix}\right)(z),
\end{aligned}$$

and indeed we have a group action. Let $\mathrm{SL}(2, \mathbb{R})$ be the subgroup of real matrices in $\mathrm{GL}(2, \mathbb{C})$ of determinant 1, and let Y be the subset of X where $\operatorname{Im} z > 0$, not

including ∞. The members of $\mathrm{SL}(2,\mathbb{R})$ carry the subset Y into itself, as we see from the computation

$$\begin{aligned}\operatorname{Im}\frac{az+b}{cz+d} &= \operatorname{Im}\frac{(az+b)(c\bar{z}+d)}{|cz+d|^2} = \operatorname{Im}\frac{adz+bc\bar{z}}{|cz+d|^2}\\ &= \frac{(ad-bc)\operatorname{Im} z}{|cz+d|^2} = \frac{\operatorname{Im} z}{|cz+d|^2}.\end{aligned}$$

Since the effect of a matrix g^{-1} is to invert the effect of g, and since both g and g^{-1} carry Y to itself, we conclude that $\mathrm{SL}(2,\mathbb{R})$ acts on $Y = \{z \in \mathbb{C} \mid \operatorname{Im} z > 0\}$ by linear fractional transformations. In similar fashion one can verify that the subgroup[12] of $\mathrm{GL}(2,\mathbb{C})$

$$\left\{ \begin{pmatrix} \alpha & \beta \\ \bar{\beta} & \bar{\alpha} \end{pmatrix} \;\middle|\; \alpha \in \mathbb{C},\ \beta \in \mathbb{C},\ |\alpha|^2 - |\beta|^2 = 1 \right\}$$

acts on $\{z \in \mathbb{C} \mid |z| < 1\}$ by linear fractional transformations.

One group action can yield many others. For example, from an action of G on X, we can construct an action on the space of all complex-valued functions on X. The definition is $(gf)(x) = f(g^{-1}x)$, the use of the inverse being necessary in order to verify property (i) of a group action:

$$\begin{aligned}((g_1g_2)f)(x) &= f((g_1g_2)^{-1}x) = f((g_2^{-1}g_1^{-1})x)\\ &= f(g_2^{-1}(g_1^{-1}x)) = (g_2f)(g_1^{-1}x) = (g_1(g_2f))(x).\end{aligned}$$

There is nothing special about the complex numbers as range for the functions here. We can allow any set as range, and we can even allow G to act on the range, as well as on the domain.[13] If G acts on X and Y, then the set of functions from X to Y inherits a group action under the definition

$$(gf)(x) = g(f(g^{-1}x)),$$

as is easily checked. In other words, we are to use g^{-1} where the domain enters the formula and we are to use g where the range enters the formula.

If V is a vector space over a field $\mathbb{F}$, a **representation** of G on V is a group action of G on V by *linear* functions. Specifically for each $g \in G$, τ_g is to be a

[12] This subgroup is commonly called $\mathrm{SU}(1,1)$ for reasons that are not relevant to the current discussion.

[13] When $\mathbb{C}$ was used as range in the previous display, the group action of G on $\mathbb{C}$ was understood to be **trivial** in the sense that $gz = z$ for every g in G and z in $\mathbb{C}$.

member of the group of linear maps from V into itself. Usually one writes $\tau(g)$ instead of τ_g in representation theory, and thus the condition is that $\tau(g)$ is to be linear for each $g \in G$ and we are to have $\tau(1) = 1$ and $\tau(g_1 g_2) = \tau(g_1)\tau(g_2)$ for all g_1 and g_2. There are interesting examples both when V is finite-dimensional and when V is infinite-dimensional.[14]

EXAMPLES OF REPRESENTATIONS.

(1) If $m \geq 1$, then the additive group $\mathbb{Z}/m\mathbb{Z}$ acts linearly on $\mathbb{R}^2$ by

$$\tau(k) = \begin{pmatrix} \cos\frac{2\pi k}{m} & -\sin\frac{2\pi k}{m} \\ \sin\frac{2\pi k}{m} & \cos\frac{2\pi k}{m} \end{pmatrix}, \qquad k \in \{0, 1, 2, \ldots, m-1\}.$$

Each $\tau(k)$ is a rotation matrix about the origin through an angle that is a multiple of $2\pi/m$. These transformations of $\mathbb{R}^2$ form a subgroup of the group of symmetries of a regular k-gon centered at the origin in $\mathbb{R}^2$.

(2) The dihedral group D_3 acts linearly on $\mathbb{R}^2$ with

$$\tau(1) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \ \tau(2\ 3) = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \ \tau(1\ 2) = \begin{pmatrix} -\frac{1}{2} & \frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & \frac{1}{2} \end{pmatrix}, \ \tau(1\ 3) = \begin{pmatrix} -\frac{1}{2} & -\frac{\sqrt{3}}{2} \\ -\frac{\sqrt{3}}{2} & \frac{1}{2} \end{pmatrix},$$

$$\tau(1\ 2\ 3) = \begin{pmatrix} -\frac{1}{2} & -\frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & -\frac{1}{2} \end{pmatrix}, \ \tau(1\ 3\ 2) = \begin{pmatrix} -\frac{1}{2} & \frac{\sqrt{3}}{2} \\ -\frac{\sqrt{3}}{2} & -\frac{1}{2} \end{pmatrix}.$$

Each of these matrices carries into itself the equilateral triangle with center at the origin and one vertex at $(1, 0)$. To obtain these matrices, we number the vertices #1, #2, #3 counterclockwise with the vertex at $(1, 0)$ as #1.

(3) The symmetric group $\mathfrak{S}_n$ acts linearly on $\mathbb{R}^n$ by permuting the indices of standard basis vectors. For example, with $n = 3$, we have $(1\ \ 3)e_1 = e_3$, $(1\ \ 3)e_2 = e_2$, etc. The matrices may be computed by the techniques of Section II.3. With $n = 3$, we obtain, for example,

$$(1\ \ 3) \mapsto \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} \qquad \text{and} \qquad (1\ \ 2\ \ 3) \mapsto \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}.$$

(4) If G acts on a set X, then the corresponding action $(gf)(x) = f(g^{-1}x)$ on complex-valued functions is a representation on the vector space of all complex-valued functions on X. This vector space is infinite-dimensional if X is an infinite set. The linearity of the action on functions follows from the definitions of addition

[14] In some settings a continuity assumption may be added to the definition of a representation, or the field $\mathbb{F}$ may be restricted in some way. We impose no such assumption here at this time.

and scalar multiplication of functions. In fact, let functions f_1 and f_2 be given, and let c be a scalar. Then

$$\begin{aligned}(g(f_1 + f_2))(x) &= (f_1 + f_2)(g^{-1}x) = f_1(g^{-1}x) + f_2(g^{-1}x)\\ &= (gf_1)(x) + (gf_2)(x) = (gf_1 + gf_2)(x)\end{aligned}$$

and

$$(g(cf_1))(x) = (cf_1)(g^{-1}x) = c(f_1(g^{-1}x)) = c((gf_1)(x)) = (c(gf_1))(x).$$

One more important class of group actions consists of those that are closely related to the structure of the group itself. Two simple ones are the action of G on itself by left translations $(g_1, g_2) \mapsto g_1g_2$ and the action of G on itself by right translations $(g_1, g_2) \mapsto g_2g_1^{-1}$. More useful is the action of G on a quotient space G/H, where H is a subgroup. This action is given by $(g_1, g_2H) \mapsto g_1g_2H$. There are still others, and some of them are particularly handy in analyzing finite groups. We give some applications in the present section and the next, and we postpone others to Section 10. Before describing some of these actions in detail, let us make some general definitions and establish two easy results.

Let $G \times X \to X$ be a group action. If p is in X, then $G_p = \{g \in G \mid gp = p\}$ is a subgroup of G called the **isotropy subgroup** at p or **stabilizer** of G at p. This is not always a normal subgroup; however, the subgroup $\bigcap_{p\in G} G_p$ that fixes all points of X is the kernel of the homomorphism $G \to \mathcal{F}(X)$ defining the group action, and such a kernel has to be normal.

Let p and q be in X. We say that p is equivalent to q for the purposes of this paragraph if $p = gq$ for some $g \in G$. The result is an equivalence relation: it is reflexive since $p = 1p$, it is symmetric since $p = gq$ implies $g^{-1}p = q$, and it is transitive since $p = gq$ and $q = g'r$ together imply $p = (gg')r$. The equivalence classes are called **orbits** of the group action. The orbit of a point p in X is $Gp = \{gp \mid g \in G\}$. If $Y = Gp$ is an orbit,[15] or more generally if Y is any subset of X carried to itself by every element of G, then $G \times Y \to Y$ is a group action. In fact, each function $y \mapsto gy$ is invertible on Y with $y \mapsto g^{-1}y$ as the inverse function, and properties (i) and (ii) of a group action follow from the same properties for X.

A group action $G \times X \to X$ is said to be **transitive** if there is just one orbit, hence if $X = Gp$ for each p in X. It is **simply transitive** if it is transitive and if for each p and q in X, there is just one element g of G with $gp = q$.

[15] Although the notation G_p for the isotropy subgroup and Gp for the orbit are quite distinct in print, it is easy to confuse the two in handwritten mathematics. Some readers may therefore prefer a different notation for one of them. The notation $Z_G(p)$ for the isotropy subgroup is one that is in common use; its use is consistent with the notation for the "centralizer" of an element in a group, which will be defined shortly. Another possibility, used by many mathematicians, is to write $G \cdot p$ for the orbit.

Proposition 4.34. Let $G \times X \to X$ be a group action, let p be in X, and let H be the isotropy subgroup at p. Then the map $G \to Gp$ given by $g \mapsto gp$ descends to a well-defined map $G/H \to Gp$ that is one-one from G/H onto the orbit Gp and respects the group actions.

REMARK. In other words, a group action of G on a single orbit is always **isomorphic as a group action** to the action of G on some quotient space G/H.

PROOF. Let $\varphi : G \to Gp$ be defined by $\varphi(g) = gp$. For h in $H = G_p$, $\varphi(gh) = (gh)p = g(hp) = gp = \varphi(g)$ shows that φ descends to a well-defined function $\overline{\varphi} : G/H \to Gp$, and $\overline{\varphi}$ is certainly onto Gp. If $\overline{\varphi}(g_1H) = \overline{\varphi}(g_2H)$, then $g_1p = \varphi(g_1p) = \varphi(g_2p) = g_2p$, and hence $g_2^{-1}g_1p = p$, $g_2^{-1}g_1$ is in H, g_1 is in g_2H, and $g_1H = g_2H$. Thus $\overline{\varphi}$ is one-one.

Respecting the group action means that $\overline{\varphi}(gg'H) = g\overline{\varphi}(g'H)$, and this identity holds since $g\overline{\varphi}(g'H) = g\varphi(g') = g(g'p) = (gg')p = \varphi(gg') = \overline{\varphi}(gg'H)$. □

A simple consequence is the following important **counting formula** in the case of a group action by a finite group.

Corollary 4.35. Let G be a finite group, let $G \times X \to X$ be a group action, let p be in X, and G_p be the isotropy group at p, and let Gp be the orbit of p. Then $|G| = |Gp|\,|G_p|$.

PROOF. Proposition 4.34 shows that the action of G on some G/G_p is the most general group action on a single orbit, G_p being the isotropy subgroup. Thus the corollary follows from Lagrange's Theorem (Theorem 4.7) with $H = G_p$ and $G/H = Gp$. □

We turn to applications of group actions to the structure of groups. If H is a subgroup of a group G, the **index** of H in G is the number of elements in G/H, finite or infinite. The first application notes a situation in which a subgroup of a finite group is automatically normal.

Proposition 4.36. Let G be a finite group, and let p be the smallest prime dividing the order of G. If H is a subgroup of G of index p, then H is normal.

REMARKS. The most important case is $p = 2$: any subgroup of index 2 is automatically normal, and this conclusion is valid even if G is infinite, as was already pointed out in Example 3 of Section 2. If G is finite and if 2 divides the order of G, there need not, however, be any subgroup of index 2; for example, the alternating group $\mathfrak{A}_4$ has order 12, and Problem 11 at the end of the chapter shows that $\mathfrak{A}_4$ has no subgroup of order 6.

PROOF. Let $X = G/H$, and restrict the group action $G \times X \to X$ to an action $H \times X \to X$. The subset $\{1H\}$ is a single orbit under H, and the remaining $p-1$ members of G/H form a union of orbits. Corollary 4.35 shows that the number of elements in an orbit has to be a divisor of $|H|$, and the smallest divisor of $|H|$ other than 1 is $\geq p$ since the smallest divisor of $|G|$ other than 1 equals p and since $|H|$ divides $|G|$. Hence any orbit of H containing more than one element has at least p elements. Since only $p-1$ elements are left under consideration, each orbit under H contains only one element. Therefore $hgH = gH$ for all h in H and g in G. Then $g^{-1}hg$ is in H, and we conclude that H is normal. □

If G is a group, the **center** Z_G of G is the set of all elements x such that $gx = xg$ for all g in G. The center of G is a subgroup (since $gx = xg$ and $gy = yg$ together imply $g(xy) = xgy = (xy)g$ and $xg^{-1} = g^{-1}(gx)g^{-1} = g^{-1}(xg)g^{-1} = g^{-1}x$), and every subgroup of the center is normal since $x \in Z_G$ and $g \in G$ together imply $gxg^{-1} = x$. Here are examples: the center of a group G is G itself if and only if G is abelian, the center of the quaternion group H_8 is $\{\pm 1\}$, and the center of any symmetric group $\mathfrak{S}_n$ with $n \geq 3$ is $\{1\}$.

If x is in G, the **centralizer** of x in G, denoted by $Z_G(x)$, is the set of all g such that $gx = xg$. This is a subgroup of G, and it equals G itself if and only if x is in the center of G. For example the centralizer of $\mathbf{i}$ in H_8 is the 4-element subgroup $\{\pm 1, \pm \mathbf{i}\}$.

Having made these definitions, we introduce a new group action of G on G, namely $(g, x) \mapsto gxg^{-1}$. The orbits are called the **conjugacy classes** of G. If x and y are two elements of G, we say that x is **conjugate** to y if x and y are in the same conjugacy class. In other words, x is conjugate to y if there is some g in G with $gxg^{-1} = y$. The result is an equivalence relation. Let us write $C\ell(x)$ for the conjugacy class of x. We can easily compute the isotropy subgroup G_x at x under this action; it consists of all $g \in G$ such that $gxg^{-1} = x$ and hence is exactly the centralizer $Z_G(x)$ of x in G. In particular, $C\ell(x) = \{x\}$ if and only if x is in the center Z_G. Applying Corollary 4.35, we immediately obtain the following result.

Proposition 4.37. If G is a finite group, then $|G| = |\,C\ell(x)|\,|Z_G(x)|$ for all x in G.

Thus $|\,C\ell(x)|$ is always a divisor of $|G|$, and it equals 1 if and only if x is in the center Z_G. Let us apply these considerations to groups whose order is a power of a prime.

Corollary 4.38. If G is a finite group whose order is a positive power of a prime, then the center Z_G is not $\{1\}$.

PROOF. Let $|G| = p^n$ with p prime and with $n > 0$. The conjugacy classes of G exhaust G, and thus the sum of all $|\,C\ell(x)|$'s equals $|G|$. Since $|\,C\ell(x)| = 1$ if and only if x is in Z_G, the sum of $|Z_G|$ and all the $|\,C\ell(x)|$'s that are not 1 is equal to $|G|$. All the terms $|\,C\ell(x)|$ that are not 1 are positive powers of p, by Proposition 4.37, and so is $|G|$. Therefore p divides $|Z_G|$. □

Corollary 4.39. If G is a finite group of order p^2 with p prime, then G is abelian.

PROOF. From Corollary 4.38 we see that either $|Z_G| = p^2$, in which case G is abelian, or $|Z_G| = p$. We show that the latter is impossible. If fact, if x is not in Z_G, then $Z_G(x)$ is a subgroup of G that contains Z_G and the element x. It must then have order p^2 and be all of G. Hence every element of G commutes with x, and x is in Z_G, contradiction. □

Corollary 4.40. If G is a finite group whose order is a positive power p^n of a prime p, then there exist normal subgroups G_k of G for $0 \le k \le n$ such that $|G_k| = p^k$ for all $k \le n$ and such that $G_k \subsetneqq G_{k+1}$ for all $k < n$.

PROOF. We proceed by induction on n. The base case of the induction is $n = 1$ and is handled by Corollary 4.9. Assume inductively that the result holds for n, and let G have order p^{n+1}. Corollary 4.38 shows that $Z_G \ne \{1\}$. Any element $\ne 1$ in Z_G must have order a power of p, and some power of it must therefore have order p. Thus let a be an element of Z_G of order p, and let H be the subgroup consisting of the powers of a. Then H is normal and has order p. Let $G' = G/H$ be the quotient group, and let $\varphi : G \to G'$ be the quotient homomorphism. The group G' has order p^n, and the inductive hypothesis shows that G' has normal subgroups G'_k for $0 \le k \le n$ such that $|G'_k| = p^k$ for $k \le n$ and $G'_k \subseteq G'_{k+1}$ for $k \le n-1$. For $1 \le k \le n+1$, define $G_k = \varphi^{-1}(G'_{k-1})$, and let $G_0 = \{1\}$. The First Isomorphism Theorem (Theorem 4.13) shows that each G_k for $k \ge 1$ is a normal subgroup of G containing H and that $\varphi(G_k) = G'_{k-1}$. Then $\varphi\big|_{G_k}$ is a homomorphism of G_k onto G'_{k-1} with kernel H, and hence $|G_k| = |G'_{k-1}|\,|H| = p^{k-1}p = p^k$. Therefore the G_k's will serve as the required subgroups of G. □

It is not always so easy to determine the conjugacy classes in a particular group. For example, in $\mathrm{GL}(n, \mathbb{C})$ the question of conjugacy is the question whether two matrices are similar in the sense of Section II.3; this will be one of the main problems addressed in Chapter V. By contrast, the problem of conjugacy in symmetric groups has a simple answer. Recall that every permutation is uniquely the product of disjoint cycles. The **cycle structure** of a permutation consists of the number of cycles of each length in this decomposition.

Lemma 4.41. Let σ and τ be members of the symmetric group $\mathfrak{S}_n$. If σ is expressed as the product of disjoint cycles, then $\tau\sigma\tau^{-1}$ has the same cycle structure as σ, and the expression for $\tau\sigma\tau^{-1}$ as the product of disjoint cycles is obtained from that for σ by substituting $\tau(k)$ for k throughout.

REMARK. For example, if $\sigma = (a\ \ b)(c\ \ d\ \ e)$, then $\tau\sigma\tau^{-1}$ decomposes as $\big(\tau(a)\ \ \tau(b)\big)\big(\tau(c)\ \ \tau(d)\ \ \tau(e)\big)$.

PROOF. Because the conjugate of a product equals the product of the conjugates, it is enough to handle a cycle $\gamma = (a_1\ \ a_2\ \ \cdots\ \ a_r)$ appearing in σ. The corresponding cycle $\gamma' = \tau\gamma\tau^{-1}$ is asserted to be $\gamma' = (\tau(a_1)\ \ \tau(a_2)\ \ \cdots\ \ \tau(a_r))$. Application of τ^{-1} to $\tau(a_j)$ yields a_j, application of σ to this yields a_{j+1} if $j < r$ and a_1 if $j = r$, and application of τ to the result yields $\tau(a_{j+1})$ or $\tau(a_1)$. For each of the symbols b not in the list $\{a_1, \ldots, a_r\}$, $\tau\gamma\tau^{-1}(\tau(b)) = \tau(b)$ since $\gamma(b) = b$. Thus $\tau\gamma\tau^{-1} = \gamma'$, as asserted. □

Proposition 4.42. Let H be a subgroup of a symmetric group $\mathfrak{S}_n$. If $C\ell(x)$ denotes a conjugacy class in H, then all members of $C\ell(x)$ have the same cycle structure. Conversely if $H = \mathfrak{S}_n$, then the conjugacy class of a permutation σ consists of all members of $\mathfrak{S}_n$ having the same cycle structure as σ.

PROOF. The first conclusion is immediate from Lemma 4.41. For the second conclusion, let σ and σ' have the same cycle structure, and let τ be the permutation that moves, for each k, the k^{th} symbol appearing in the disjoint-cycle expansion of σ into the k^{th} symbol in the corresponding expansion of σ'. Define τ on the remaining symbols in any fashion at all. Application of the lemma shows that $\tau\sigma\tau^{-1} = \sigma'$. Thus any two permutations with the same cycle structure are conjugate. □

7. Semidirect Products

One more application of group actions to the structure theory of groups will be to the construction of "semidirect products" of groups. If H is a group, then an isomorphism of H with itself is called an **automorphism**. The set of automorphisms of H is a group under composition, and we denote it by Aut H. We are going to be interested in "group actions by automorphisms," i.e., group actions of a group G on a space X when X is itself a group and the action by each member of G is an automorphism of the group structure of X; the group action is therefore a homomorphism of the form $\tau : G \to \text{Aut } X$.

EXAMPLE 1. In $\mathbb{R}^2$, we can identify the additive group of the underlying vector space with the group of translations $\ell_v(w) = v + w$; the identification

associates a translation ℓ with the member $\ell(0)$ of $\mathbb{R}^2$. Let H be the group of translations. The rotations about the origin in $\mathbb{R}^2$, namely the linear maps with matrices $\begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}$, form a group $G = \mathrm{SO}(2)$ that acts on $\mathbb{R}^2$, hence acts on the set H of translations. The linearity of the rotations says that the action of $G = \mathrm{SO}(2)$ on the translations is by automorphisms of H, i.e., that each rotation, in its effect on G, is in Aut H. Out of these data—the two groups G and H and a homomorphism of G into Aut H—we will construct below what amounts to the group of all rotations (about any point) and translations of $\mathbb{R}^2$. The construction is that of a "semidirect product."

EXAMPLE 2. Take any group G, and let G act on $X = G$ by conjugation. Each conjugation $x \mapsto gxg^{-1}$ is an automorphism of G, and thus the action of G on itself by conjugation is an action by automorphisms.

Let G and H be groups. Suppose that a group action $\tau : G \to \mathcal{F}(H)$ is given with G acting on H by automorphisms. That is, suppose that each map $h \mapsto \tau_g(h)$ is an automorphism of H. We define a group $G \times_\tau H$ whose underlying set will be the Cartesian product $G \times H$. The motivation for the definition of multiplication comes from Example 2, in which $\tau_g(h) = ghg^{-1}$. We want to write a product $g_1h_1g_2h_2$ in the form $g'h'$, and we can do so using the formula

$$g_1h_1g_2h_2 = g_1g_2(g_2^{-1}h_1g_2)h_2 = \big(g_1g_2\big)\big((\tau_{g_2^{-1}}(h_1))h_2\big).$$

Similarly the formula for inverses is motivated by the formula

$$(gh)^{-1} = h^{-1}g^{-1} = g^{-1}(gh^{-1}g^{-1}) = g^{-1}\tau_g(h^{-1}).$$

Proposition 4.43. Let G and H be groups, and let τ be a group action of G on H by automorphisms. Then the set-theoretic product $G \times H$ becomes a group $G \times_\tau H$ under the definitions

$$(g_1, h_1)(g_2, h_2) = (g_1g_2, (\tau_{g_2^{-1}}(h_1))h_2)$$

and

$$(g, h)^{-1} = (g^{-1}, \tau_g(h^{-1})).$$

The mappings $i_1 : G \to G \times_\tau H$ and $i_2 : H \to G \times_\tau H$ given by $i_1(g) = (g, 1)$ and $i_2(h) = (1, h)$ are one-one homomorphisms, and $p_1 : G \times_\tau H \to G$ given by $p_1(g, h) = g$ is a homomorphism onto G. The images $G' = i_1(G)$ and $H' = i_2(H)$ are subgroups of $G \times_\tau H$ with H' normal such that $G' \cap H' = \{1\}$, such that every element of $G \times_\tau H$ is the product of an element of G' and an element of H', and such that conjugation of G' on H' is given by $i_1(g)i_2(h)i_1(g)^{-1} = i_2(\tau_g(h))$.

REMARK. The group $G \times_\tau H$ is called the **external semidirect product**[16] of G and H with respect to τ.

PROOF. For associativity we compute directly that

$$\big((g_1, h_1)(g_2, h_2)\big)(g_3, h_3) = (g_1 g_2 g_3, \tau_{g_3^{-1}}(\tau_{g_2^{-1}}(h_1)h_2)h_3)$$

and $$(g_1, h_1)\big((g_2, h_2)(g_3, h_3)\big) = (g_1 g_2 g_3, \tau_{g_3^{-1}g_2^{-1}}(h_1)\tau_{g_3^{-1}}(h_2)h_3).$$

Since

$$\tau_{g_3^{-1}}(\tau_{g_2^{-1}}(h_1)h_2) = (\tau_{g_3^{-1}}\tau_{g_2^{-1}}(h_1))\tau_{g_3^{-1}}(h_2) = \tau_{g_3^{-1}g_2^{-1}}(h_1)\tau_{g_3^{-1}}(h_2),$$

we have a match. It is immediate that $(1, 1)$ is a two-sided identity. Since $(g, h)(g^{-1}, \tau_g(h^{-1})) = (1, \tau_g(h)\tau_g(h^{-1})) = (1, \tau_g(hh^{-1})) = (1, \tau_g(1)) = (1, 1)$ and $(g^{-1}, \tau_g(h^{-1}))(g, h) = (1, \tau_{g^{-1}}(\tau_g(h^{-1}))h) = (1, \tau_1(h^{-1})h) = (1, 1)$, $(g^{-1}, \tau_g(h^{-1}))$ is indeed a two-sided inverse of (g, h). It is immediate from the definition of multiplication that i_1, i_2, and p_1 are homomorphisms, that i_1 and i_2 are one-one, that p_1 is onto, that $G' \cap H' = \{1\}$, and that $G \times_\tau H = G'H'$. Since i_1 and i_2 are homomorphisms, G' and H' are subgroups. Since H' is the kernel of p_1, H' is normal. Finally the definition of multiplication gives $i_1(g)i_2(h)i_1(g)^{-1} = (g, h)(g, 1)^{-1} = (g, h)(g^{-1}, 1) = (1, (\tau_g(h))1) = i_2(\tau_g(h))$, and the proof is complete. □

Proposition 4.44. Let S be a group, let G and H be subgroups with H normal, and suppose that $G \cap H = \{1\}$ and that every element of S is the product of an element of G and an element of H. For each $g \in G$, define an automorphism τ_g of H by $\tau_g(h) = ghg^{-1}$. Then τ is a group action of G on H by automorphisms, and the mapping $G \times_\tau H \to S$ given by $(g, h) \mapsto gh$ is an isomorphism of groups.

REMARKS. In this case we call S an **internal semidirect product** of G and H with respect to τ. We shall not attempt to write down a universal mapping property that characterizes internal semidirect products.

PROOF. Since $\tau_{g_1g_2}(h) = g_1g_2hg_2^{-1}g_1^{-1} = g_1\tau_{g_2}(h)g_1^{-1} = \tau_{g_1}\tau_{g_2}(h)$ and since each τ_g is an automorphism of H, τ is an action by automorphisms. Proposition 4.43 therefore shows that $G \times_\tau H$ is a well-defined group. The function φ from $G \times_\tau H$ to S given by $\varphi(g, h) = gh$ is a homomorphism by the same computation that motivated the definition of multiplication in a semidirect product, and φ is onto S since every element of S lies in the set GH of products. If $gh = 1$, then $g = h^{-1}$ exhibits g as in $G \cap H = \{1\}$. Hence $g = 1$ and $h = 1$. Therefore φ is one-one and must be an isomorphism. □

[16]The notation $\ltimes$ is used by some authors in place of $\times_\tau$. The normal subgroup goes on the open side of the $\ltimes$ and on the side of the subscript τ in $\times_\tau$.

EXAMPLE 1. Dihedral groups D_n. We show that D_n is the internal semidirect product of a 2-element group and the rotation subgroup. Let H be the group of rotations about the origin through multiples of the angle $2\pi/n$. This group is cyclic of order n, and it is normal in D_n because it is of index 2. If s is any of the reflections in D_n, then $G = \{1, s\}$ is a subgroup of D_n of order 2 with $G \cap H = \{1\}$. Counting the elements, we see that every element of D_n is of the form r^k or sr^k, in other words that the set of products GH is all of D_n. Thus Proposition 4.44 shows that D_n is an (internal) semidirect product of G and H with respect to some $\tau : G \to \operatorname{Aut} H$. To understand the homomorphism τ, let us write the members of H as the powers of r, where r is rotation counterclockwise about the origin through the angle $2\pi/n$. For the reflection s (or indeed for any reflection in D_n), a look at the geometry shows that $sr^k s^{-1} = r^{-k}$ for all k. In other words, the automorphism $\tau(1)$ leaves each element of H fixed while $\tau(s)$ sends each k mod n to $-k$ mod n. The map that sends each element of a cyclic group to its group inverse is indeed an automorphism of the cyclic group, and thus τ is indeed a homomorphism of G into $\operatorname{Aut} H$.

EXAMPLE 2. Construction of a nonabelian group of order 21. Let $H = C_7$, written multiplicatively with generator a, and let $G = C_3$, written multiplicatively with generator b. To arrange for G to act on H by automorphisms, we make use of a nontrivial automorphism of H of order 3. Such a mapping is $a^k \mapsto a^{2k}$. In fact, there is no doubt that this mapping is an automorphism, and we have to see that it has order 3. The effect of applying it twice is $a^k \mapsto a^{4k}$, and the effect of applying it three times is $a^k \mapsto a^{8k}$. But $a^{8k} = a^k$ since $a^7 = 1$, and thus the mapping $a^k \mapsto a^{2k}$ indeed has order 3. We send b^n into the n^{th} power of this automorphism, and the result is a homomorphism $\tau : G \to \operatorname{Aut} H$. The semidirect product $G \times_\tau H$ is certainly a group of order $3 \times 7 = 21$. To see that it is nonabelian, we observe from the group law in Proposition 4.43 that $ab = b\tau_{b^{-1}}(a) = ba^4$. Thus $ab \neq ba$, and $G \times_\tau H$ is nonabelian.

It is instructive to generalize the construction in Example 2 a little bit. To do so, we need a lemma.

Lemma 4.45. If p is a prime, then the automorphisms of the additive group of the field $\mathbb{F}_p$ are the multiplications by the members of the multiplicative group $\mathbb{F}_p^\times$, and consequently $\operatorname{Aut} C_p$ is isomorphic to a cyclic group C_{p-1}.

PROOF. Let us write $\operatorname{Aut} \mathbb{F}_p$ for the automorphism group of the additive group of $\mathbb{F}_p$. Each function $\varphi_a : \mathbb{F}_p \to \mathbb{F}_p$ given by $\varphi_a(n) = na$, taken modulo p, is in $\operatorname{Aut} \mathbb{F}_p$ as a consequence of the distributive law. We define a function $\Phi : \operatorname{Aut} \mathbb{F}_p \to \mathbb{F}_p^\times$ by $\Phi(\varphi) = \varphi(1)$ for $\varphi \in \operatorname{Aut} \mathbb{F}_p$. Again by the distributive law $\varphi(n) = n\varphi(1)$ for every integer n. Thus if φ_1 and φ_2 are in $\operatorname{Aut} \mathbb{F}_p$, then

$\Phi(\varphi_1 \circ \varphi_2) = (\varphi_1 \circ \varphi_2)(1) = \varphi_1(\varphi_2(1)) = \varphi_2(1)\varphi_1(1)$, and consequently Φ is a homomorphism. If a member φ of $\operatorname{Aut}\mathbb{F}_p$ has $\Phi(\varphi) = 1$ in $\mathbb{F}_p^\times$, then $\varphi(1) = 1$ and therefore $\varphi(n) = n\varphi(1) = n$ for all n. Therefore φ is the identity in $\operatorname{Aut}\mathbb{F}_p$. We conclude that Φ is one-one. If a is given in $\mathbb{F}_p^\times$, then $\Phi(\varphi_a) = \varphi_a(1) = a$, and hence Φ is onto $\mathbb{F}_p^\times$. Therefore Φ is an isomorphism of $\operatorname{Aut}\mathbb{F}_p$ and $\mathbb{F}_p^\times$. By Corollary 4.27, Φ exhibits $\operatorname{Aut}\mathbb{F}_p$ as isomorphic to the cyclic group C_{p-1}. □

Proposition 4.46. If p and q are primes with $p < q$ such that p divides $q - 1$, then there exists a nonabelian group of order pq.

REMARKS. For $p = 2$, the divisibility condition is automatic, and the proof will yield the dihedral group D_q. For $p = 3$ and $q = 7$, the condition is that 3 divides $7 - 1$, and the constructed group will be the group in Example 2 above.

PROOF. Let $G = C_p$ with generator a, and let $H = C_q$. Lemma 4.45 shows that $\operatorname{Aut} C_q \cong C_{q-1}$. Let b be a generator of $\operatorname{Aut} C_q$. Since p divides $q - 1$, $b^{(q-1)/p}$ has order p. Then the map $a^k \mapsto b^{k(q-1)/p}$ is a well-defined homomorphism τ of G into $\operatorname{Aut} H$, and it determines a semidirect product $S = G \times_\tau H$, by Proposition 4.43. The order of S is pq, and the multiplication is nonabelian since for $h \in H$, we have $(a, 1)(1, h) = (a, h)$ and $(1, h)(a, 1) = (a, \tau_{a^{-1}}(h)) = (a, b^{-(q-1)/p}(h))$, but $b^{-(q-1)/p}$ is not the identity automorphism of H because it has order p. □

8. Simple Groups and Composition Series

A group $G \neq \{1\}$ is said to be **simple** if its only normal subgroups are $\{1\}$ and G.

Among abelian groups the simple ones are the cyclic groups of prime order. Indeed, a cyclic group C_p of prime order has no nontrivial subgroups at all, by Corollary 4.9. Conversely if G is abelian and simple, let $a \neq 1$ be in G. Then $\{a^n\}$ is a cyclic subgroup and is normal since G is abelian. Thus $\{a^n\}$ is all of G, and G is cyclic. The group $\mathbb{Z}$ is not simple, having the nontrivial subgroup $2\mathbb{Z}$, and the group $\mathbb{Z}/(rs)\mathbb{Z}$ with $r > 1$ and $s > 1$ is not simple, having the multiples of r as a nontrivial subgroup. Thus G has to be cyclic of prime order.

The interest is in nonabelian simple groups. We shall establish that the alternating groups $\mathfrak{A}_n$ are simple for $n \geq 5$, and some other simple groups will be considered in Problems 55–62 at the end of the chapter.

Theorem 4.47. The alternating group $\mathfrak{A}_n$ is simple if $n \geq 5$.

PROOF. Let $K \neq \{1\}$ be a normal subgroup of $\mathfrak{A}_n$. Choose σ in K with $\sigma \neq 1$ such that $\sigma(i) = i$ for the maximum possible number of integers i with $1 \leq i \leq n$.

The main step is to show that σ is a 3-cycle. Arguing by contradiction, suppose that σ is not a 3-cycle. Then there are two cases.

The first case is that the decomposition of σ as the product of disjoint cycles contains a k-cycle for some $k \geq 3$. Without loss of generality, we may take the cycle in question to be $\gamma = (1\ \ 2\ \ 3\ \ \cdots)$, and then $\sigma = \gamma\rho = (1\ \ 2\ \ 3\ \ \cdots)\rho$ with ρ equal to a product of disjoint cycles not containing the symbols appearing in γ. Being even and not being a 3-cycle, σ moves at least two other symbols besides the three listed ones, say 4 and 5. Put $\tau = (3\ \ 4\ \ 5)$. Lemma 4.41 shows that $\sigma' = \tau\sigma\tau^{-1} = \gamma'\rho' = (1\ \ 2\ \ 4\ \ \cdots)\rho'$ with ρ' not containing any of the symbols appearing in γ'. Thus $\sigma'\sigma^{-1}$ moves 3 into 4 and cannot be the identity. But $\sigma'\sigma^{-1}$ is in K and fixes all symbols other than 1, 2, 3, 4, 5 that are fixed by σ. In addition, $\sigma'\sigma^{-1}$ fixes 2, and none of 1, 2, 3, 4, 5 is fixed by σ. Thus $\sigma'\sigma^{-1}$ is a member of K other than the identity that fixes fewer symbols than σ, and we have arrived at a contradiction.

The second case is that σ is a product $\sigma = (1\ \ 2)(3\ \ 4)\cdots$ of disjoint transpositions. There must be at least two factors since σ is even. Put $\tau = (1\ \ 2)(4\ \ 5)$, the symbol 5 existing since the group $\mathfrak{A}_n$ in question has $n \geq 5$. Then $\sigma' = (1\ \ 2)(3\ \ 5)\cdots$. Since $\sigma'\sigma^{-1}$ carries 4 into 5, $\sigma'\sigma^{-1}$ is a member of K other than the identity. It fixes all symbols other than 1, 2, 3, 4, 5 that are fixed by σ, and in addition it fixes 1 and 2. Thus $\sigma'\sigma^{-1}$ fixes more symbols than σ does, and again we have arrived at a contradiction.

We conclude that K contains a 3-cycle, say $(1\ \ 2\ \ 3)$. If i, j, k, l, m are five arbitrary symbols, then we can construct a permutation τ with $\tau(1) = i, \tau(2) = j$, $\tau(3) = k$, $\tau(4) = l$, and $\tau(5) = m$. If τ is odd, we replace τ by $\tau(l\ \ m)$, and the result is even. Thus we may assume that τ is in $\mathfrak{A}_n$ and has $\tau(1) = i$, $\tau(2) = j$, and $\tau(3) = k$. Lemma 4.41 shows that $\tau\sigma\tau^{-1} = (i\ \ j\ \ k)$. Since K is normal, we conclude that K contains all 3-cycles.

To complete the proof, we show for $n \geq 3$ that every element of $\mathfrak{A}_n$ is a product of 3-cycles. If σ is in $\mathfrak{A}_n$, we use Corollary 1.22 to decompose σ as a product of transpositions. Since σ is even, we can group these in pairs. If the members of a pair of transpositions are not disjoint, then their product is a 3-cycle. If they are disjoint, then the identity $(1\ \ 2)(3\ \ 4) = (1\ \ 2\ \ 3)(2\ \ 3\ \ 4)$ shows that their product is a product of 3-cycles. This completes the proof. □

Let G be a group. A descending sequence

$$G_n \supseteq G_{n-1} \supseteq \cdots \supseteq G_1 \supseteq G_0$$

of subgroups of G with $G_n = G$, $G_0 = \{1\}$, and each G_{k-1} normal in G_k is called a **normal series** for G. The normal series is called a **composition series** if each inclusion $G_k \supseteq G_{k-1}$ is proper and if each consecutive quotient G_k/G_{k-1} is simple.

EXAMPLES.

(1) Let G be a cyclic group of order N. A normal series for G consists of certain subgroups of G, all necessarily cyclic by Proposition 4.4. Their respective orders $N_n, N_{n-1}, \ldots, N_1, N_0$ have $N_n = N$, $N_0 = 1$, and $N_{k-1} \mid N_k$ for all k. The series is a composition series if and only if each quotient N_k/N_{k-1} is prime. In this case the primes that occur are exactly the prime divisors of N, and a prime p occurs r times if p^r is the exact power of p that divides N. Thus the consecutive quotients from a composition series of this G, up to isomorphisms, are independent of the particular composition series—though they may arise in a different order.

(2) For $G = \mathbb{Z}$, a normal series is of the form

$$\mathbb{Z} \supseteq m_1\mathbb{Z} \supseteq m_1m_2\mathbb{Z} \supseteq m_1m_2m_3\mathbb{Z} \supseteq \cdots \supseteq 0.$$

The group $G = \mathbb{Z}$ has no composition series.

(3) For the symmetric group $G = \mathfrak{S}_4$, let $C_2 \times C_2$ refer to the 4-element subgroup $\{1,\ (1\ 2)(3\ 4),\ (1\ 3)(2\ 4),\ (1\ 4)(2\ 3)\}$. The series

$$\mathfrak{S}_4 \supseteq \mathfrak{A}_4 \supseteq C_2 \times C_2 \supseteq \{1,\ (1\ 2)(3\ 4)\} \supseteq \{1\}$$

is a composition series, the consecutive quotients being C_2, C_3, C_2, C_2. Each term in the composition series except for $\{1,\ (1\ 2)(3\ 4)\}$ is actually normal in the whole group G, but there is no way to choose the 2-element subgroup to make it normal in G. The other two possible choices of 2-element subgroup, which lead to different composition series but with isomorphic consecutive quotients, are obtained by replacing $\{1,\ (1\ 2)(3\ 4)\}$ by $\{1,\ (1\ 3)(2\ 4)\}$ and again by $\{1,\ (1\ 4)(2\ 3)\}$.

(4) For the symmetric group $G = \mathfrak{S}_5$, the series

$$\mathfrak{S}_5 \supseteq \mathfrak{A}_5 \supseteq \{1\}$$

is a composition series, the consecutive quotients being C_2 and $\mathfrak{A}_5$.

(5) Let G be a finite group of order p^n with p prime. Corollary 4.40 produces a composition series, and this time all the subgroups are normal in G. The successive normal subgroups have orders p^k for $k = n, n-1, \ldots, 0$, and each consecutive quotient is isomorphic to C_p.

Historically the Jordan–Hölder Theorem addressed composition series for groups, showing that the consecutive quotients, up to isomorphisms, are independent of the particular composition series. They can then consistently be called the **composition factors** of the group. Finding the composition factors of a particular

group may be regarded as a step toward understanding the structure of the group. A generalization of the Jordan–Hölder Theorem due to Zassenhaus and Schreier applies to normal series in situations in which composition series might not exist, such as Example 2 above. We prove the Zassenhaus–Schreier Theorem, and the Jordan–Hölder Theorem is then a special case.

Two normal series

$$G_m \supseteq G_{m-1} \supseteq \cdots \supseteq G_1 \supseteq G_0$$

and

$$H_n \supseteq H_{n-1} \supseteq \cdots \supseteq H_1 \supseteq H_0$$

for the same group G are said to be **equivalent normal series** if $m = n$ and the order of the consecutive quotients $G_m/G_{m-1}, G_{m-1}/G_{m-2}, \ldots, G_1/G_0$ may be rearranged so that they are respectively isomorphic to $H_m/H_{m-1}, H_{m-1}/H_{m-2}, \ldots, H_1/H_0$. One normal series is said to be a **refinement** of another if the subgroups appearing in the second normal series all appear as subgroups in the first normal series.

Lemma 4.48 (Zassenhaus). Let G_1, G_2, G_1', and G_2' be subgroups of a group G with $G_1' \subseteq G_1$ and $G_2' \subseteq G_2$, G_1' normal in G_1, and G_2' normal in G_2. Then $(G_1 \cap G_2')G_1'$ is normal in $(G_1 \cap G_2)G_1'$, $(G_1' \cap G_2)G_2'$ is normal in $(G_1 \cap G_2)G_2'$, and

$$((G_1 \cap G_2)G_1')/((G_1 \cap G_2')G_1') \cong ((G_1 \cap G_2)G_2')/((G_1' \cap G_2)G_2').$$

PROOF. Let us check that $(G_1 \cap G_2')G_1'$ is normal in $(G_1 \cap G_2)G_1'$. Handling conjugation by members of $G_1 \cap G_2$ is straightforward: If g is in $G_1 \cap G_2$, then $g(G_1 \cap G_2')g^{-1} = G_1 \cap G_2'$ since g is in G_1 and $gG_2'g^{-1} = G_2'$. Also, $gG_1'g^{-1} = G_1'$ since g is in G_1. Hence $g(G_1 \cap G_2')G_1'g^{-1} = (G_1 \cap G_2')G_1'$.

Handling conjugation by members of G_1' requires a little trick: Let g be in G_1' and let hg' be in $(G_1 \cap G_2')G_1'$. Then $g(hg')g^{-1} = h(h^{-1}gh)g'g^{-1}$. The left factor h is in $G_1 \cap G_2'$. The remaining factors are in G_1'; for g' and g^{-1}, this is a matter of definition, and for $h^{-1}gh$, it follows because h is in G_1 and g is in G_1'. Thus $g(G_1 \cap G_2')G_1'g^{-1} = (G_1 \cap G_2')G_1'$, and $(G_1 \cap G_2')G_1'$ is normal in $(G_1 \cap G_2)G_1'$. The other assertion about normal subgroups holds by symmetry in the indexes 1 and 2.

By the Second Isomorphism Theorem (Theorem 4.14),

$$\begin{aligned}(G_1 \cap G_2)/(((G_1 \cap G_2')G_1') \cap (G_1 \cap G_2)) &\\ \cong ((G_1 \cap G_2)(G_1 \cap G_2')G_1')/((G_1 \cap G_2')G_1') &\qquad (*)\\ = ((G_1 \cap G_2)G_1')/((G_1 \cap G_2')G_1').&\end{aligned}$$

Since we have

$$((G_1 \cap G_2')G_1') \cap (G_1 \cap G_2) = ((G_1 \cap G_2')G_1') \cap G_2 = (G_1 \cap G_2')(G_1' \cap G_2),$$

we can rewrite the conclusion of $(*)$ as

$$(G_1 \cap G_2)/((G_1 \cap G_2')(G_1' \cap G_2)) \cong ((G_1 \cap G_2)G_1')/((G_1 \cap G_2')G_1'). \quad (**)$$

The left side of $(**)$ is symmetric under interchange of the indices 1 and 2. Hence so is the right side, and the lemma follows. □

Theorem 4.49 (Schreier). Any two normal series of a group G have equivalent refinements.

PROOF. Let the two normal series be

$$\begin{aligned} G_m &\supseteq G_{m-1} \supseteq \cdots \supseteq G_1 \supseteq G_0, \\ H_n &\supseteq H_{n-1} \supseteq \cdots \supseteq H_1 \supseteq H_0, \end{aligned} \quad (*)$$

and define

$$\begin{aligned} G_{ij} &= (G_i \cap H_j)G_{i+1} \qquad \text{for } 0 \le j \le n, \\ H_{ji} &= (G_i \cap H_j)H_{j+1} \qquad \text{for } 0 \le i \le m. \end{aligned} \quad (**)$$

Then we obtain respective refinements of the two normal series $(*)$ given by

$$\begin{aligned} G = G_{00} &\supseteq G_{01} \supseteq \cdots \supseteq G_{0n} \\ &\supseteq G_{10} \supseteq G_{11} \supseteq \cdots \supseteq G_{1n} \cdots \supseteq G_{m-1,n} = \{1\}, \\ G = H_{00} &\supseteq H_{01} \supseteq \cdots \supseteq H_{0m} \\ &\supseteq H_{10} \supseteq H_{11} \supseteq \cdots \supseteq H_{1m} \cdots \supseteq H_{n-1,m} = \{1\}. \end{aligned} \quad (\dagger)$$

The containments $G_{in} \supseteq G_{i+1,0}$ and $H_{jm} \supseteq H_{j+1,0}$ are equalities in $(\dagger)$, and the only nonzero consecutive quotients are therefore of the form $G_{ij}/G_{i,j+1}$ and $H_{ji}/H_{j,i+1}$. For these we have

$$\begin{aligned} G_{ij}/G_{i,j+1} &= ((G_i \cap H_j)G_{i+1})/((G_i \cap H_{j+1})G_{i+1}) && \text{by } (**) \\ &\cong ((G_i \cap H_j)H_{j+1})/((G_{i+1} \cap H_j)H_{j+1}) && \text{by Lemma 4.48} \\ &= H_{ji}/H_{j,i+1} && \text{by } (**), \end{aligned}$$

and thus the refinements $(\dagger)$ are equivalent. □

Corollary 4.50 (Jordan–Hölder Theorem). Any two composition series of a group G are equivalent as normal series.

PROOF. Let two composition series be given. Theorem 4.49 says that we can insert terms in each so that the refined series have the same length and are equivalent. Since the given series are composition series, the only way to insert a new term is by repeating some term, and the repetition results in a consecutive quotient of $\{1\}$. Because of Theorem 4.49 we know that the quotients $\{1\}$ from the two refined series must match. Thus the number of terms added to each series is the same. Also, the quotients that are not $\{1\}$ must match in pairs. Thus the given composition series are equivalent. □

9. Structure of Finitely Generated Abelian Groups

A set of **generators** for a group G is a set such that each element of G is a finite product of generators and their inverses. (A generator and its inverse are allowed to occur multiple times in a product.)

In this section we shall study *abelian* groups having a finite set of generators. Such groups are said to be **finitely generated abelian groups**, and our goal is to classify them up to isomorphism. We use additive notation for all our abelian groups in this section. We begin by introducing an analog $\mathbb{Z}^n$ for the integers $\mathbb{Z}$ of the vector space $\mathbb{R}^n$ for the reals $\mathbb{R}$, and along with it a generalization.

A **free abelian group** is any abelian group isomorphic to a direct sum, finite or infinite, of copies of the additive group $\mathbb{Z}$ of integers. The external direct sum of n copies of $\mathbb{Z}$ will be denoted by $\mathbb{Z}^n$. Let us use Proposition 4.17 to see that we can recognize groups isomorphic to free abelian groups by means of the following condition: an abelian group G is isomorphic to a free abelian group if and only if it has a $\mathbb{Z}$ **basis**, i.e., a subset that generates G and is such that no nontrivial linear combination, with integer coefficients, of the members of the subset is equal to the 0 element of the group. It will be helpful to use terminology adapted from the theory of vector spaces for this latter condition—that the subset is to be **linearly independent** over $\mathbb{Z}$.

Let us give the proof that the condition is necessary and sufficient for G to be free abelian. In one direction if G is an external direct sum of copies of $\mathbb{Z}$, then the members of G that are 1 in one coordinate and are 0 elsewhere form a $\mathbb{Z}$ basis. Conversely if $\{g_s\}_{s\in S}$ is a $\mathbb{Z}$ basis, let G_{s_0} be the subgroup of multiples of g_{s_0}, and let φ_{s_0} be the inclusion homomorphism of G_{s_0} into G. Proposition 4.17 produces a unique group homomorphism $\varphi : \bigoplus_{s\in S} G_s \to G$ such that $\varphi \circ i_{s_0} = \varphi_{s_0}$ for all $s_0 \in S$. The spanning condition for the $\mathbb{Z}$ basis says that φ is onto G, and the linear independence condition for the $\mathbb{Z}$ basis says that φ has 0 kernel.

The similarity between vector-space bases and $\mathbb{Z}$ bases suggests further comparison of vector spaces and abelian groups. With vector spaces over a field, every vector space has a basis over the field. However, it is exceptional for an abelian group to have a $\mathbb{Z}$ basis. Two examples that hint at the difficulty are the additive group $\mathbb{Z}/m\mathbb{Z}$ with $m > 1$ and the additive group $\mathbb{Q}$. The group $\mathbb{Z}/m\mathbb{Z}$ has no nonempty linearly independent set, while the group $\mathbb{Q}$ has a linearly independent set of one element, no spanning set of one element, and no linearly independent set of more than one element. Here are two positive examples.

EXAMPLES.

(1) The additive group of all points in $\mathbb{R}^n$ whose coordinates are integers. The standard basis of $\mathbb{R}^n$ is a $\mathbb{Z}$ basis.

(2) The additive group of all points (x, y) in $\mathbb{R}^2$ with x and y both in $\mathbb{Z}$ or both in $\mathbb{Z} + \frac{1}{2}$. The set $\left\{(1, 0), \left(\frac{1}{2}, \frac{1}{2}\right)\right\}$ is a $\mathbb{Z}$ basis.

Next we take a small step that eliminates technical complications from the discussion, proving that any subgroup of a finitely generated abelian group is finitely generated.

Lemma 4.51. Let $\varphi : G \to H$ be a homomorphism of abelian groups. If $\ker \varphi$ and $\operatorname{image} \varphi$ are finitely generated, then G is finitely generated.

PROOF. Let $\{x_1, \dots, x_m\}$ and $\{y_1, \dots, y_n\}$ be respective finite sets of generators for $\ker \varphi$ and $\operatorname{image} \varphi$. For $1 \le j \le n$, choose x_j' in G with $\varphi(x_j') = y_j$. We shall prove that $\{x_1, \dots, x_m, x_1', \dots, x_n'\}$ is a set of generators for G. Thus let x be in G. Since $\varphi(x)$ is in $\operatorname{image} \varphi$, there exist integers $a_1, \dots, a_n$ with $\varphi(x) = a_1 y_1 + \cdots + a_n y_n$. The element $x' = a_1 x_1' + \cdots + a_n x_n'$ of G has $\varphi(x') = a_1 y_1 + \cdots + a_n y_n = \varphi(x)$. Therefore $\varphi(x - x') = 0$, and there exist integers $b_1, \dots, b_m$ with $x - x' = b_1 x_1 + \cdots + b_m x_m$. Hence

$$x = b_1 x_1 + \cdots + b_m x_m + x' = b_1 x_1 + \cdots + b_m x_m + a_1 x_1' + \cdots + a_n x_n'. \quad \square$$

Proposition 4.52. Any subgroup of a finitely generated abelian group is finitely generated.

PROOF. Let G be finitely generated with a set $\{g_1, \dots, g_n\}$ of n generators, and define $G_k = \mathbb{Z}g_1 + \cdots + \mathbb{Z}g_k$ for $1 \le k \le n$. If H is any subgroup of G, define $H_k = H \cap G_k$ for $1 \le k \le n$. We shall prove by induction on k that every H_k is finitely generated, and then the case $k = n$ gives the proposition. For $k = 1$, $G_1 = \mathbb{Z}g_1$ is a cyclic group, and any subgroup of it is cyclic by Proposition 4.4 and hence is finitely generated.

Assume inductively that every subgroup of G_k is known to be finitely generated. Let $q : G_{k+1} \to G_{k+1}/G_k$ be the quotient homomorphism, and let $\varphi = q\big|_{H_{k+1}}$,

mapping H_{k+1} into G_{k+1}/G_k. Then $\ker\varphi = H_{k+1} \cap G_k$ is a subgroup of G_k and is finitely generated by the inductive hypothesis. Also, image φ is a subgroup of G_{k+1}/G_k, which is a cyclic group with generator equal to the coset of g_{k+1}. Since a subgroup of a cyclic group is cyclic, image φ is finitely generated. Applying Lemma 4.51 to φ, we see that H_{k+1} is finitely generated. This completes the induction and the proof. □

A free abelian group has **finite rank** if it has a finite $\mathbb{Z}$ basis, hence if it is isomorphic to $\mathbb{Z}^n$ for some n. The first theorem is that the integer n is determined by the group.

Theorem 4.53. The number of $\mathbb{Z}$ summands in a free abelian group of finite rank is independent of the direct-sum decomposition of the group.

We define this number to be the **rank** of the free abelian group. Actually, "rank" is a well-defined cardinal in the infinite-rank case as well, because the rank coincides in that case with the cardinality of the group. In any event, Theorem 4.53 follows immediately by two applications of the following lemma.

Lemma 4.54. If G is a free abelian group with a finite $\mathbb{Z}$ basis $x_1, \dots, x_n$, then any linearly independent subset of G has $\le n$ elements.

PROOF. Let $\{y_1, \dots, y_m\}$ be a linearly independent set in G. Since $\{x_1, \dots, x_n\}$ is a $\mathbb{Z}$ basis, we can define an m-by-n matrix C of integers by $y_i = \sum_{j=1}^n C_{ij}x_j$. As a matrix in $M_{mn}(\mathbb{Q})$, C has rank $\le n$. Consequently if $m > n$, then the rows are linearly dependent over $\mathbb{Q}$, and we can find rational numbers $q_1, \dots, q_m$ not all 0 such that $\sum_{i=1}^m q_i C_{ij} = 0$ for all j. Multiplying by a suitable integer to clear fractions, we obtain integers $k_1, \dots, k_m$ not all 0 such that $\sum_{i=1}^m k_i C_{ij} = 0$ for all j. Then we have

$$\sum_{i=1}^m k_i y_i = \sum_{i=1}^m k_i \sum_{j=1}^n C_{ij}x_j = \sum_{j=1}^n \Big(\sum_{i=1}^m k_i C_{ij}\Big)x_j = \sum_{j=1}^n 0x_j = 0,$$

in contradiction to the linear independence of $\{y_1, \dots, y_m\}$ over $\mathbb{Z}$. Therefore $m \le n$. □

Now we come to the two main results of this section. The first is a special case of the second by Proposition 4.52 and Lemma 4.54. The two will be proved together, and it may help to regard the proof of the first as a part of the proof of the second.

Theorem 4.55. A subgroup H of a free abelian group G of finite rank n is free abelian of rank $\le n$.

REMARK. This result persists in the case of infinite rank, but we do not need the more general result and will not give a proof.

Theorem 4.56 (Fundamental Theorem of Finitely Generated Abelian Groups). Every finitely generated abelian group is a finite direct sum of cyclic groups. The cyclic groups may be taken to be copies of $\mathbb{Z}$ and various C_{p^k} with p prime, and in this case the cyclic groups are unique up to order and to isomorphism.

REMARKS. The main conclusion of the theorem is the decomposition of each finitely generated abelian group into the direct sum of cyclic groups. An alternative decomposition of the given group that forces uniqueness is as the direct sum of copies of $\mathbb{Z}$ and finite cyclic groups $C_{d_1}, \dots, C_{d_r}$ such that $d_1 \mid d_2$, $d_2 \mid d_3, \dots$, $d_{r-1} \mid d_r$. A proof of the additional statement appears in the problems at the end of Chapter VIII. The integers $d_1, \dots, d_r$ are sometimes called the **elementary divisors** of the group.

Let us establish the setting for the proof of Theorem 4.56. Let G be the given group, and say that it has a set of n generators. Proposition 4.17 produces a homomorphism $\varphi : \mathbb{Z}^n \to G$ that carries the standard generators $x_1, \dots, x_n$ of $\mathbb{Z}^n$ to the generators of G, and φ is onto G. Let H be the kernel of φ. As a subgroup of $\mathbb{Z}^n$, H is finitely generated, by Proposition 4.52. Let $y_1, \dots, y_m$ be generators. Theorem 4.55 predicts that H is in fact free abelian, hence that $\{y_1, \dots, y_m\}$ could be taken to be linearly independent over $\mathbb{Z}$ with $m \le n$, but we do not assume that knowledge in the proof of Theorem 4.56.

The motivation for the main part of the proof of Theorem 4.56 comes from the elementary theory of vector spaces, particularly from the method of using a basis for a finite-dimensional vector space to find a basis of a vector subspace when we know a finite spanning set for the vector subspace. Thus let V be a finite-dimensional vector space over $\mathbb{R}$, with basis $\{x_j\}_{j=1}^n$, and let U be a vector subspace with spanning set $\{y_i\}_{i=1}^m$. To produce a vector-space basis for U, we imagine expanding the y_i's as linear combinations of $x_1, \dots, x_n$. We can think symbolically of this expansion as expressing each y_i as the product of a row vector of real numbers times the formal "column vector" $\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$. The entries of this column vector are vectors, but there is no problem in working with it since this is all just a matter of notation anyway. Then the formal column vector $\begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix}$ of m members of U equals the product of an m-by-n matrix of real numbers times the formal column vector $\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$. We know from Chapter II that the procedure for finding a basis of U is to row reduce this matrix of real numbers. The nonzero rows of the result determine a basis of the span of the m vectors we have used, and this basis is related tidily to the given basis for V. We can compare the two bases

to understand the relationship between U and V. To prove Theorem 4.56, we would like to use the same procedure, but we have to work with an integer matrix and avoid division. This means that only two of the three usual row operations are fully available for the row reduction; division of a row by an integer is allowable only when the integer is ± 1. A partial substitute for division comes by using the steps of the Euclidean algorithm via the division algorithm (Proposition 1.1), but even that is not enough. For example, if the m-by-n matrix is $\begin{pmatrix} 2 & 1 & 1 \\ 0 & 0 & 3 \end{pmatrix}$, no further row reduction is possible with integer operations. However, the equations tell us that H is the subgroup of $\mathbb{Z}^3$ generated by $(2, 1, 1)$ and $(0, 0, 3)$, and it is not at all clear how to write $\mathbb{Z}^3/H$ as a direct sum of cyclic groups.

The row operations have the effect of changing the set of generators of H while maintaining the fact that they generate H. What is needed is to allow also column reduction with integer operations. Steps of this kind have the effect of changing the $\mathbb{Z}$ basis of $\mathbb{Z}^n$. When steps of this kind are allowed, we can produce new generators of H and a new basis of $\mathbb{Z}^n$ so that the two can be compared. With the example above, suitable column operations are

$$\begin{pmatrix} 2 & 1 & 1 \\ 0 & 0 & 3 \end{pmatrix} \mapsto \begin{pmatrix} 1 & 2 & 1 \\ 0 & 0 & 3 \end{pmatrix} \mapsto \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 3 \end{pmatrix} \mapsto \begin{pmatrix} 1 & 0 & 0 \\ 0 & 3 & 0 \end{pmatrix}.$$

The equations with the new generators say that $y_1' = x_1'$ and $y_2' = 3x_2'$. Thus H is the subgroup $\mathbb{Z} \oplus 3\mathbb{Z} \oplus 0\mathbb{Z}$, nicely aligned with $\mathbb{Z}^3 = \mathbb{Z} \oplus \mathbb{Z} \oplus \mathbb{Z}$. The quotient is $(\mathbb{Z}/\mathbb{Z}) \oplus (\mathbb{Z}/3\mathbb{Z}) \oplus (\mathbb{Z}/0\mathbb{Z}) \cong C_3 \oplus \mathbb{Z}$.

The proof of Theorem 4.56 will make use of an algorithm that uses row and column operations involving only allowable divisions and that converts the matrix C of coefficients so that its nonzero entries are the **diagonal entries** C_{ii} for $1 \le i \le r$ and no other entries. The algorithm in principle can be very slow, and it may be helpful to see what it does in an ordinary example.

EXAMPLE. Suppose that the relationship between generators y_1, y_2, y_3 of H and the standard $\mathbb{Z}$ basis $\{x_1, x_2\}$ of $\mathbb{Z}^2$ is

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = C \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \qquad \text{where } C = \begin{pmatrix} 3 & 5 \\ 7 & 13 \\ 5 & 9 \end{pmatrix}.$$

In row reduction in vector-space theory, we would start by dividing the first row of C by 3, but division by 3 is not available in the present context. Our target for the upper-left entry is $\mathrm{GCD}(3, 7, 5) = 1$, and we use the division algorithm one step at a time to get there. To begin with, it says that $7 = 2 \cdot 3 + 1$ and hence $7 - 2 \cdot 3 = 1$. The first step of row reduction is then to replace the second row by

the difference of it and 2 times the first row. The result can be achieved by left multiplication by

$$\begin{pmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \qquad \text{and is} \qquad \begin{pmatrix} 3 & 5 \\ 1 & 3 \\ 5 & 9 \end{pmatrix}.$$

We write this step as

$$\begin{pmatrix} 3 & 5 \\ 7 & 13 \\ 5 & 9 \end{pmatrix} \overset{\text{left by } \left(\begin{smallmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{smallmatrix}\right)}{\longmapsto} \begin{pmatrix} 3 & 5 \\ 1 & 3 \\ 5 & 9 \end{pmatrix}.$$

The entry 1 in the first column is our target for this stage since $\mathrm{GCD}(3, 7, 5) = 1$. The next step interchanges two rows to move the 1 to the upper left entry, and the subsequent step uses the 1 to eliminate the other entries of the first column:

$$\begin{pmatrix} 3 & 5 \\ 1 & 3 \\ 5 & 9 \end{pmatrix} \overset{\text{left by } \left(\begin{smallmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{smallmatrix}\right)}{\longmapsto} \begin{pmatrix} 1 & 3 \\ 3 & 5 \\ 5 & 9 \end{pmatrix} \overset{\text{left by } \left(\begin{smallmatrix} 1 & 0 & 0 \\ -3 & 1 & 0 \\ -5 & 0 & 1 \end{smallmatrix}\right)}{\longmapsto} \begin{pmatrix} 1 & 3 \\ 0 & -4 \\ 0 & -6 \end{pmatrix}.$$

The algorithm next seeks to eliminate the off-diagonal entry 3 in the first row. This is done by a column operation:

$$\begin{pmatrix} 1 & 3 \\ 0 & -4 \\ 0 & -6 \end{pmatrix} \overset{\text{right by } \left(\begin{smallmatrix} 1 & -3 \\ 0 & 1 \end{smallmatrix}\right)}{\longmapsto} \begin{pmatrix} 1 & 0 \\ 0 & -4 \\ 0 & -6 \end{pmatrix}.$$

With two further row operations we are done:

$$\begin{pmatrix} 1 & 0 \\ 0 & -4 \\ 0 & -6 \end{pmatrix} \overset{\text{left by } \left(\begin{smallmatrix} 1 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{smallmatrix}\right)}{\longmapsto} \begin{pmatrix} 1 & 0 \\ 0 & 2 \\ 0 & -6 \end{pmatrix} \overset{\text{left by } \left(\begin{smallmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 3 & 1 \end{smallmatrix}\right)}{\longmapsto} \begin{pmatrix} 1 & 0 \\ 0 & 2 \\ 0 & 0 \end{pmatrix}.$$

Our steps are summarized by the fact that the matrix A with

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 3 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ -3 & 1 & 0 \\ -5 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

has

$$AC \begin{pmatrix} 1 & -3 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 2 \\ 0 & 0 \end{pmatrix}$$

and by the fact that the integer matrices to the left and right of C have determinant ± 1. The determinant condition ensures that A^{-1} and $\left(\begin{smallmatrix} 1 & -3 \\ 0 & 1 \end{smallmatrix}\right)^{-1}$ have integer entries, according to Cramer's rule (Proposition 2.38).

Lemma 4.57. If C is an m-by-n matrix of integers, then there exist an m-by-m matrix A of integers with determinant ± 1 and an n-by-n matrix B of integers with determinant ± 1 such that for some $r \geq 0$, the nonzero entries of $D = ACB$ are exactly the diagonal entries $D_{11}, D_{22}, \ldots, D_{rr}$.

PROOF. Given C, choose (i, j) with $|C_{ij}| \neq 0$ but $|C_{ij}|$ as small as possible. (If $C = 0$, the algorithm terminates.) Possibly by interchanging two rows and/or then two columns (a left multiplication with determinant -1 and then a right multiplication with determinant -1), we may assume that $(i, j) = (1, 1)$. By the division algorithm write, for each i,

$$C_{i1} = q_i C_{11} + r_i \qquad \text{with } 0 \leq r_i < |C_{11}|,$$

and replace the i^{th} row by the difference of the i^{th} row and q_i times the first row (a left multiplication). If some r_i is not 0, the result will leave a nonzero entry in the first column that is $< |C_{11}|$ in absolute value. Permute the least such $r_i \neq 0$ to the upper left and repeat the process. Since the least absolute value is going down, this process at some point terminates with all r_i equal to 0. The first column then has a nonzero diagonal entry and is otherwise 0.

Now consider C_{1j} and apply the division algorithm and column operations in similar fashion in order to process the first row. If we get a smaller nonzero remainder, permute the smallest one to the first column. Repeat this process until the first row is 0 except for entry C_{11}. Continue alternately with row and column operations in this fashion until *both* $C_{1j} = 0$ for $j > 1$ and $C_{i1} = 0$ for $i > 1$.

Repeat the algorithm for the $(m-1)$-by-$(n-1)$ matrix consisting of rows 2 through m and columns 2 through n, and continue inductively. The algorithm terminates when either the reduced-in-size matrix is empty or is all 0. At this point the original matrix has been converted into the desired "diagonal form." □

Lemma 4.58. Let $G_1, \ldots, G_n$ be abelian groups, and for $1 \leq j \leq n$, let H_j be a subgroup of G_j. Then

$$(G_1 \oplus \cdots \oplus G_n)/(H_1 \oplus \cdots \oplus H_n) \cong (G_1/H_1) \oplus \cdots \oplus (G_n/H_n).$$

PROOF. Let $\varphi : G_1 \oplus \cdots \oplus G_n \to (G_1/H_1) \oplus \cdots \oplus (G_n/H_n)$ be the homomorphism defined by $\varphi(g_1, \ldots, g_n) = (g_1 H_1, \ldots, g_n H_n)$. The mapping φ is onto $(G_1/H_1) \oplus \cdots \oplus (G_n/H_n)$, and the kernel is $H_1 \oplus \cdots \oplus H_n$. Then Corollary 4.12 shows that φ descends to the required isomorphism. □

PROOF OF THEOREM 4.55 AND MAIN CONCLUSION OF THEOREM 4.56. Given G with n generators, we set up matters as indicated immediately after the statement of Theorem 4.56, writing

$$\begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} = C \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix},$$

where $x_1, \ldots, x_n$ are the standard generators of $\mathbb{Z}^n$, $y_1, \ldots, y_m$ are the generators of the kernel of the homomorphism from $\mathbb{Z}^n$ onto G, and C is a matrix of integers. Applying Lemma 4.57, let A and B be square integer matrices of determinant ± 1 such that $D = ACB$ is diagonal as in the statement of the lemma. Define

$$\begin{pmatrix} z_1 \\ \vdots \\ z_m \end{pmatrix} = A \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} \qquad \text{and} \qquad \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix} = B^{-1} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}.$$

Substitution gives

$$\begin{pmatrix} z_1 \\ \vdots \\ z_m \end{pmatrix} = A \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} = (ACB)B^{-1} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = ACB \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix}.$$

If $(c_1 \ \cdots \ c_n)$ and $(d_1 \ \cdots \ d_n) = (c_1 \ \cdots \ c_n)B^{-1}$ are row vectors, then the formula

$$\begin{aligned} c_1u_1 + \cdots + c_nu_n = (c_1 \ \cdots \ c_n) \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix} &= (d_1 \ \cdots \ d_n) \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \\ &= d_1x_1 + \cdots + d_nx_n \end{aligned} \qquad (*)$$

shows that $\{u_1, \ldots, u_n\}$ generates the same subset of $\mathbb{Z}^n$ as $\{x_1, \ldots, x_n\}$. Since $(c_1 \ \cdots \ c_n)$ is nonzero if and only if $(d_1 \ \cdots \ d_n)$ is nonzero, the formula $(*)$ shows also that the linear independence of $\{x_1, \ldots, x_n\}$ implies that of $\{u_1, \ldots, u_n\}$. Hence $\{u_1, \ldots, u_n\}$ is a $\mathbb{Z}$ basis of $\mathbb{Z}^n$. Similarly $\{y_1, \ldots, y_m\}$ and $\{z_1, \ldots, z_m\}$ generate the same subgroup H of $\mathbb{Z}^n$. Therefore we can compare H and $\mathbb{Z}^n$ using $\{z_1, \ldots, z_m\}$ and $\{u_1, \ldots, u_n\}$. Since D is diagonal, the equations relating $\{z_1, \ldots, z_m\}$ and $\{u_1, \ldots, u_n\}$ are $z_j = D_{jj}u_j$ for $j \leq \min(m, n)$ and $z_j = 0$ for $\min(m, n) < j \leq m$. If $q = \min(m, n)$, then we see that

$$H = \sum_{i=1}^{m} \mathbb{Z}z_i = \sum_{i=1}^{q} D_{ii}\mathbb{Z}u_i + \sum_{i=q+1}^{m} \mathbb{Z}z_i = \sum_{i=1}^{q} D_{ii}\mathbb{Z}u_i.$$

Since the set $\{u_1, \ldots, u_q\}$ is linearly independent over $\mathbb{Z}$, this sum exhibits H as given by

$$H = D_{11}\mathbb{Z} \oplus \cdots \oplus D_{qq}\mathbb{Z}$$

with $D_{11}u_1, \ldots, D_{qq}u_q$ as a $\mathbb{Z}$ basis. Consequently H has been exhibited as free abelian of rank $\leq q \leq n$. This proves Theorem 4.55. Applying Lemma 4.58 to the quotient $\mathbb{Z}^n/H$ and letting $D_{11}, \ldots, D_{rr}$ be the nonzero diagonal entries of D, we see that H has rank r, and we obtain an expansion of G in terms of cyclic groups as

$$G = C_{D_{11}} \oplus \cdots \oplus C_{D_{rr}} \oplus \mathbb{Z}^{n-r}.$$

This proves the main conclusion of Theorem 4.56. □

PROOF OF THE DECOMPOSITION WITH CYCLIC GROUPS OF PRIME-POWER ORDER. It is enough to prove that if $m = \prod_{j=1}^{N} p_j^{k_j}$ with the p_j equal to distinct primes, then $\mathbb{Z}/m\mathbb{Z} \cong (\mathbb{Z}/p_1^{k_1}\mathbb{Z}) \oplus \cdots \oplus (\mathbb{Z}/p_N^{k_N}\mathbb{Z})$. This is a variant of the Chinese Remainder Theorem (Corollary 1.9). For the proof let

$$\varphi : \mathbb{Z} \to (\mathbb{Z}/p_1^{k_1}\mathbb{Z}) \oplus \cdots \oplus (\mathbb{Z}/p_N^{k_N}\mathbb{Z})$$

be the homomorphism given by $\varphi(s) = \left(s \bmod p_1^{k_1}, \dots, s \bmod p_N^{k_N}\right)$ for $s \in \mathbb{Z}$. Since $\varphi(m) = (0, \dots, 0)$, φ descends to a homomorphism

$$\overline{\varphi} : \mathbb{Z}/m\mathbb{Z} \to (\mathbb{Z}/p_1^{k_1}\mathbb{Z}) \oplus \cdots \oplus (\mathbb{Z}/p_N^{k_N}\mathbb{Z}).$$

The map $\overline{\varphi}$ is one-one because if $\varphi(s) = 0$, then $p_j^{k_j}$ divides s for all j. Since the $p_j^{k_j}$ are relatively prime in pairs, their product m divides s. Since m divides s, $s \equiv 0 \bmod m$. The map $\overline{\varphi}$ is onto since it is one-one and since the finite sets $\mathbb{Z}/m\mathbb{Z}$ and $(\mathbb{Z}/p_1^{k_1}\mathbb{Z}) \oplus \cdots \oplus (\mathbb{Z}/p_N^{k_N}\mathbb{Z})$ both have m elements. □

PROOF OF UNIQUENESS OF THE DECOMPOSITION. Write $G = \mathbb{Z}^s \oplus T$, where

$$T = (\mathbb{Z}/p_1^{l_1}\mathbb{Z}) \oplus \cdots \oplus (\mathbb{Z}/p_M^{l_M}\mathbb{Z})$$

and the p_j's are not necessarily distinct. The subgroup T is the subgroup of elements of finite order in G, and it is well defined independently of the decomposition of G as the direct sum of cyclic groups. The quotient $G/T \cong \mathbb{Z}^s$ is free abelian of finite rank, and its rank s is well defined by Theorem 4.53. Thus the number s of factors of $\mathbb{Z}$ in the decomposition of G is uniquely determined, and we need only consider uniqueness of the decomposition of the finite abelian group T.

For p prime the elements of T of order p^a for some a are those in the sum of the groups $\mathbb{Z}/p_j^{l_j}\mathbb{Z}$ for which $p_j = p$, and we are reduced to considering a group

$$H = \mathbb{Z}/p^{l_1}\mathbb{Z} \oplus \cdots \oplus \mathbb{Z}/p^{l_{M'}}\mathbb{Z}$$

with p fixed and $l_1 \le \cdots \le l_{M'}$. The set of p^j powers of elements of H is a subgroup of H and is given by $\mathbb{Z}/p^{l_t - j}\mathbb{Z} \oplus \cdots \oplus \mathbb{Z}/p^{l_{M'} - j}\mathbb{Z}$ if l_t is the first index $\ge j$, while the set of p^{j+1} powers of elements of H is given by $\mathbb{Z}/p^{l_{t'} - j - 1}\mathbb{Z} \oplus \cdots \oplus \mathbb{Z}/p^{l_{M'} - j}\mathbb{Z}$ if $l_{t'}$ is the first index $\ge j+1$. Therefore Lemma 4.58 gives

$$p^j H/p^{j+1}H \cong (\mathbb{Z}/p^{l_{t'} - j - 1}\mathbb{Z})/(\mathbb{Z}/p^{l_{t'} - j}\mathbb{Z}) \oplus \cdots \oplus (\mathbb{Z}/p^{l_{M'} - j - 1}\mathbb{Z})/(\mathbb{Z}/p^{l_{M'} - j}\mathbb{Z}).$$

Each term of $p^j H/p^{j+1}H$ has order p, and thus

$$|p^j H/p^{j+1}H| = p^{|\{i \,|\, l_i > j\}|}.$$

Hence H determines the integers $l_1, \dots, l_{M'}$, and uniqueness is proved. □

10. Sylow Theorems

This section continues the use of group actions to obtain results concerning structure theory for abstract groups. We shall prove the three Sylow Theorems, which are a starting point for investigations of the structure of finite groups that are deeper than those in Sections 6 and 7. We state the three theorems as the parts of Theorem 4.59.

Theorem 4.59 (Sylow Theorems). Let G be a finite group of order $p^m r$, where p is prime and p does not divide r. Then

(a) G contains a subgroup of order p^m, and any subgroup of G of order p^l with $0 \le l < m$ is contained in a subgroup of order p^m,
(b) any two subgroups of order p^m in G are conjugate in G, i.e., any two such subgroups P_1 and P_2 have $P_2 = aP_1a^{-1}$ for some $a \in G$,
(c) the number of subgroups of order p^m is of the form $pk + 1$ and divides r.

REMARK. A subgroup of order p^m as in the theorem is called a **Sylow p-subgroup** of G. A consequence of (a) when $m \ge 1$ is that G has a subgroup of order p; this special case is sometimes called **Cauchy's Theorem in group theory**.

Before coming to the proof, let us carefully give two simple applications to structure theory. The applications combine Theorem 4.59, some results of Sections 6 and 7, and Problems 35–38 and 45–48 at the end of the chapter.

Proposition 4.60. If p and q are primes with $p < q$, then there exists a nonabelian group of order pq if and only if p divides $q - 1$, and in this case the nonabelian group is unique up to isomorphism. It may be taken to be a semidirect product of the cyclic groups C_p and C_q with C_q normal.

REMARK. It follows from Theorem 4.56 that the only abelian group of order pq, up to isomorphism, is $C_p \times C_q \cong C_{pq}$. If $p = 2$ in the proposition, then q is odd and p divides $q - 1$; the proposition yields the dihedral group D_q. For $p > 2$, the divisibility condition may or may not hold: For $pq = 15$, the condition does not hold, and hence every group of order 15 is cyclic. For $pq = 21$, the condition does hold, and there exists a nonabelian group of order 21; this group was constructed explicitly in Example 2 in Section 7.

PROOF. Existence of a nonabelian group of order pq, together with the semidirect-product structure, is established by Proposition 4.46 if p divides $q-1$. Let us see uniqueness and the necessity of the condition that p divide $q - 1$.

If G has order pq, Theorem 4.59a shows that G has a Sylow p-subgroup H_p and a Sylow q-subgroup H_q. Corollary 4.9 shows that these two groups are cyclic.

The conjugates of H_q are Sylow q-subgroups, and Theorem 4.59c shows that the number of such conjugates is of the form $kq + 1$ and divides p. Since $p < q$, $k = 0$. Therefore H_q is normal. (Alternatively, one can apply Proposition 4.36 to see that H_q is normal.)

Each element of G is uniquely a product ab with a in H_p and b in H_q. For the uniqueness, if $a_1b_1 = a_2b_2$, then $a_2^{-1}a_1 = b_2b_1^{-1}$ is an element of $H_p \cap H_q$. Its order must divide both p and q and hence must be 1. Thus the pq products ab with a in H_p and b in H_q are all different. Since the number of them equals the order of G, every member of G is such a product. By Proposition 4.44, G is a semidirect product of H_p and H_q.

If the action of H_p on H_q is nontrivial, then Problem 37 at the end of the chapter shows that p divides $q - 1$, and Problem 38 shows that the group is unique up to isomorphism. On the other hand, if the action is trivial, then G is certainly abelian. □

Proposition 4.61. If G is a group of order 12, then G contains a subgroup H of order 3 and a subgroup K of order 4, and at least one of them is normal. Consequently there are exactly five groups of order 12, up to isomorphism—two abelian and three nonabelian.

REMARK. The second statement follows from the first, as a consequence of Problems 45–48 at the end of the chapter. Those problems show how to construct the groups.

PROOF. Theorem 4.59a shows that H may be taken to be a Sylow 3-subgroup and K may be taken to be a Sylow 2-subgroup. We have to prove that either H or K is normal.

Suppose that H is not normal. Theorem 4.59c shows that the number of Sylow 3-subgroups is of the form $3k + 1$ and divides 4. The subgroup H, not being normal, fails to equal one of its conjugates, which will be another Sylow 3-subgroup; hence $k > 0$. Therefore $k = 1$, and there are four Sylow 3-subgroups. The intersection of any two such subgroups is a subgroup of both and must be trivial since 3 is prime. Thus the set-theoretic union of the Sylow 3-subgroups accounts for $4 \cdot 2 + 1$ elements. None of these elements apart from the identity lies in K, and thus K contributes 3 further elements, for a total of 12. Thus every element of G lies in K or a conjugate of H. Consequently K equals every conjugate of K, and K is normal. □

Let us see where we are with classifying finite groups of certain orders, up to isomorphism. A group of order p is cyclic by Corollary 4.9, and a group of order p^2 is abelian by Corollary 4.39. Groups of order pq are settled by Proposition 4.60. Thus for p and q prime, we know the structure of all groups of order p,

p^2, and pq. Problems 39–44 at the end of the chapter tell us the structure of the groups of order 8, and Proposition 4.61 and Problems 45–48 tell us the structure of the groups of order 12. In particular, the table at the end of Section 1, which gives examples of nonisomorphic groups of order at most 15, is complete except for the one group of order 12 that is discussed in Problem 48.

Problems 30–34 and 49–54 at the end of the chapter go in the direction of classifying finite groups of certain other orders.

Now we return to Theorem 4.59. The proof of the theorem makes use of the theory of group actions as in Section 6. In fact, the proof of existence of Sylow p-subgroups is just an elaboration of the argument used to prove Corollary 4.38, saying that a group of prime-power order has a nontrivial center. The relevant action for the existence part of the proof is the one $(g, x) \mapsto gxg^{-1}$ given by conjugation of the elements of the group, the orbit of x being the conjugacy class $C\ell(x)$. Proposition 4.37 shows that $|G| = |\,C\ell(x)||Z_G(x)|$, where $Z_G(x)$ is the centralizer of x. Since the disjoint union of the conjugacy classes is all of $|G|$, we have

$$\boxed{|G| = |Z_G| + \sum_{\substack{\text{representatives } x_j \\ \text{of each conjugacy class} \\ \text{with } |\,C\ell(x)| \neq 1}} |G|/|Z_G(x_j)|,}$$

a formula sometimes called the **class equation** of G.

PROOF OF EXISTENCE OF SYLOW p-SUBGROUPS IN THEOREM 4.59a. We induct on $|G|$, the base case being $|G| = 1$. Suppose that existence holds for groups of order $< |G|$. Without loss of generality suppose that $m > 0$, so that p divides $|G|$.

First suppose that p does not divide $|Z_G|$. Referring to the class equation of G, we see that p must fail to divide some integer $|G|/|Z_G(x_j)|$ for which $|Z_G(x_j)| < |G|$. Since p^m is the exact power of p dividing $|G|$, we conclude that p^m divides this $|Z_G(x_j)|$ and p^{m+1} does not. Since $|Z_G(x_j)| < |G|$, the inductive hypothesis shows that $Z_G(x_j)$ has a subgroup of order p^m, and this is a Sylow p-subgroup of G.

Now suppose that p divides $|Z_G|$. The group Z_G is finitely generated abelian, hence is a direct sum of cyclic groups by Theorem 4.56. Thus Z_G contains an element c of order p. The cyclic group C generated by c then has order p. Being a subgroup of Z_G, C is normal in G. The group G/C has order $p^{m-1}r$, and the inductive hypothesis implies that G/C has a subgroup H of order p^{m-1}. If $\varphi : G \to G/C$ denotes the quotient map, then $\varphi^{-1}(H)$ is a subgroup of G of order $|H||\ker\varphi| = p^{m-1}p = p^m$. □

For the remaining parts of Theorem 4.59, we make use of a different group action. If Γ denotes the set of all subgroups of G, then G acts on Γ by conjugation:

$(g, H) \mapsto gHg^{-1}$. The orbit of a subgroup of H consists of all subgroups conjugate to H in G, and the isotropy subgroup at the point H in Γ is

$$\{g \in G \mid gHg^{-1} = H\}.$$

This is a subgroup $N(H)$ of G known as the **normalizer** of H in G. It has the properties that $N(H) \supseteq H$ and that H is a normal subgroup of $N(H)$. The counting formula of Corollary 4.35 gives

$$\big|\{gHg^{-1} \mid g \in G\}\big| = |G/N(H)|.$$

Meanwhile, application of Lagrange's Theorem (Theorem 4.7) to the three quotients G/H, $G/N(H)$, and $N(H)/H$ shows that

$$|G/H| = |G/N(H)||N(H)/H|,$$

with all three factors being integers.

Now assume as in the statement of Theorem 4.59 that $|G| = p^m r$ with p prime and p not dividing r. In this setting we have the following lemma.

Lemma 4.62. If P is a Sylow p-subgroup of G and if H is a subgroup of the normalizer $N(P)$ whose order is a power of p, then $H \subseteq P$.

PROOF. Since $H \subseteq N(P)$ and P is normal in $N(P)$, the set HP of products is a group, by the same argument as used for $H_p H_q$ in the proof of Proposition 4.60. Then $HP/P \cong H/(H \cap P)$ by the Second Isomorphism Theorem (Theorem 4.14), and hence $|HP/P|$ is some power p^k of p. By Lagrange's Theorem (Theorem 4.7), $|HP| = p^{m+k}$ with $k \geq 0$. Since no subgroup of G can have order p^l with $l > m$, we must have $k = 0$. Thus $HP = P$ and $H \subseteq P$. □

PROOF OF THE REMAINDER OF THEOREM 4.59. Within the set Γ of all subgroups of G, let Π be the set of all subgroups of G of order p^m. We have seen that Π is not empty. Since the conjugate of a subgroup has the same order as the subgroup, Π is the union of orbits in Γ under conjugation by G. Thus we can restrict the group action by conjugation from $G \times \Gamma \to \Gamma$ to $G \times \Pi \to \Pi$.

Let P and P' be members of Π, and let Σ and Σ' be the G orbits of P and P' under conjugation. Suppose that Σ and Σ' are distinct orbits of G. Let us restrict the group action by conjugation from $G \times \Pi \to \Pi$ to $P \times \Pi \to \Pi$. The G orbits Σ and Σ' then break into P orbits, and the counting formula Corollary 4.35 says for each orbit that

$$p^m = |P| = \#\{\text{subgroups in a } P \text{ orbit}\} \times \big|\text{isotropy subgroup within } P\big|.$$

Hence the number of subgroups in a P orbit is of the form p^l for some $l \geq 0$.

Suppose that $l = 0$. Then the P orbit is some singleton set $\{P''\}$, and the corresponding isotropy subgroup within P is all of P:

$$P = \{p \in P \mid pP''p^{-1} = P''\} \subseteq N(P'').$$

Lemma 4.62 shows that $P \subseteq P''$, and therefore $P = P''$. Thus $l = 0$ only for the P orbit $\{P\}$. In other words, the number of elements in any P orbit other than $\{P\}$ is divisible by p. Consequently $|\Sigma| \equiv 1 \bmod p$ while $|\Sigma'| \equiv 0 \bmod p$, the latter because Σ and Σ' are assumed distinct. But this conclusion is asymmetric in the G orbits Σ and Σ', and we conclude that Σ and Σ' must coincide. Hence there is only one G orbit in Π, and it has $kp+1$ members for some k. This proves parts (b) and (c) except for the fact that $kp + 1$ divides r.

For this divisibility let us apply the counting formula Corollary 4.35 to the orbit Σ of G. The formula gives $|G| = |\Sigma|\,|\text{isotropy subgroup}|$, and hence $|\Sigma|$ divides $|G| = p^m r$. Since $|\Sigma| = kp + 1$, we have $\text{GCD}(|\Sigma|, p) = 1$ and also $\text{GCD}(|\Sigma|, p^m) = 1$. By Corollary 1.3, $kp + 1$ divides r.

Finally we prove that any subgroup H of G of order p^l lies in some Sylow p-subgroup. Let $\Sigma = \Pi$ again be the G orbit in Γ of subgroups of order p^m, and restrict the action by conjugation from $G \times \Sigma \to \Sigma$ to $H \times \Sigma \to \Sigma$. Each H orbit in Σ must have p^a elements for some a, by one more application of the counting formula Corollary 4.35. Since $|\Sigma| \equiv 1 \bmod p$, some H orbit has one element, say the H orbit of P. Then the isotropy subgroup of H at the point P is all of H, and $H \subseteq N(P)$. By Lemma 4.62, $H \subseteq P$. This completes the proof of Theorem 4.59. □

11. Categories and Functors

The mathematics thus far in the book has taken place in several different contexts, and we have seen that the same notions sometimes recur in more than one context, possibly with variations. For example we have worked with vector spaces, inner-product spaces, groups, rings, and fields, and we have seen that each of these areas has its own definition of isomorphism. In addition, the notion of direct product or direct sum has arisen in more than one of these contexts, and there are other similarities. In this section we introduce some terminology to make the notion of "context" precise and to provide a setting for discussing similarities between different contexts.

A **category** $\mathcal{C}$ consists of three things:

- a class of **objects**, denoted by $\text{Obj}(\mathcal{C})$,
- for any two objects A and B in the category, a set $\text{Morph}(A, B)$ of **morphisms**,

- for any three objects A, B, and C in the category, a **law of composition** for morphisms, i.e., a function carrying $\mathrm{Morph}(A, B) \times \mathrm{Morph}(B, C)$ into $\mathrm{Morph}(A, C)$, with the image of (f, g) under composition written as gf,

and these are to satisfy certain properties that we list in a moment. When more than one category is under discussion, we may use notation like $\mathrm{Morph}_{\mathcal{C}}(A, B)$ to distinguish between the categories.

We are to think initially of the objects as the sets we are studying with a particular kind of structure on them; the morphisms are then the functions from one object to another that respect this additional structure, and the law of composition is just composition of functions. Indeed, the defining conditions that are imposed on general categories are arranged to be obvious for this special kind of category, and this setting accounts for the order in which we write the composition of two morphisms. But the definition of a general category is not so restrictive, and it is important not to restrict the definition in this way.

The properties that are to be satisfied to have a category are as follows:

(i) the sets $\mathrm{Morph}(A_1, B_1)$ and $\mathrm{Morph}(A_2, B_2)$ are disjoint unless $A_1 = A_2$ and $B_1 = B_2$ (because two functions are declared to be different unless their domains match and their ranges match, as is underscored in Section A1 of the appendix),

(ii) the law of composition satisfies the associativity property $h(gf) = (hg)f$ for $f \in \mathrm{Morph}(A, B)$, $g \in \mathrm{Morph}(B, C)$, and $h \in \mathrm{Morph}(C, D)$,

(iii) for each object A, there is an **identity morphism** 1_A in $\mathrm{Morph}(A, A)$ such that $f1_A = f$ and $1_A g = g$ for $f \in \mathrm{Morph}(A, B)$ and $g \in \mathrm{Morph}(C, A)$.

A **subcategory** $\mathcal{S}$ of a category $\mathcal{C}$ by definition is a category with $\mathrm{Obj}(\mathcal{S}) \subseteq \mathrm{Obj}(\mathcal{C})$ and $\mathrm{Morph}_{\mathcal{S}}(A, B) \subseteq \mathrm{Morph}_{\mathcal{C}}(A, B)$ whenever A and B are in $\mathrm{Obj}(\mathcal{S})$, and it is assumed that the laws of composition in $\mathcal{S}$ and $\mathcal{C}$ are consistent when both are defined.

Here are several examples in which the morphisms are functions and the law of composition is ordinary composition of functions. They are usually identified in practice just by naming their objects, since the morphisms are understood to be all functions from one object to another respecting the additional structure on the objects.

EXAMPLES OF CATEGORIES.

(1) The category of all sets. An object A is a set, and a morphism in the set $\mathrm{Morph}(A, B)$ is a function from A into B.

(2) The category of all vector spaces over a field $\mathbb{F}$. The morphisms are linear maps.

(3) The category of all groups. The morphisms are group homomorphisms.

(4) The category of all abelian groups. The morphisms again are group homomorphisms. This is a subcategory of the previous example.

(5) The category of all rings. The morphisms are all ring homomorphisms. The kernel and the image of a morphism are necessarily objects of the category.

(6) The category of all rings with identity. The morphisms are all ring homomorphisms carrying identity to identity. This is a subcategory of the previous example. The image of a morphism is necessarily an object of the category, but the kernel of a morphism is usually not in the category.

(7) The category of all fields. The morphisms are as in Example 6, and the result is a subcategory of Example 6. In this case any morphism is necessarily one-one and carries inverses to inverses.

(8) The category of all group actions by a particular group G. If G acts on X and on Y, then a morphism from the one space to the other is a G **equivariant mapping** from X to Y, i.e., a function $\varphi : X \to Y$ such that $\varphi(gx) = g\varphi(x)$ for all x in X.

(9) The category of all representations by a particular group G on a vector space over a particular field $\mathbb{F}$. The morphisms are the linear G equivariant functions. This is a subcategory of the previous example.

Readers who are familiar with point-set topology will recognize that one can impose topologies on everything in the above examples, insisting that the functions be continuous, and again we obtain examples of categories. For example the category of all topological spaces consists of objects that are topological spaces and morphisms that are continuous functions. The category of all continuous group actions by a particular topological group has objects that are group actions $G \times X \to X$ that are continuous functions, and the morphisms are the equivariant functions that are continuous.

Readers who are familiar with manifolds will recognize that another example is the category of all smooth manifolds, which consists of objects that are smooth manifolds and morphisms that are smooth functions.

The morphisms in a category need not be functions in the usual sense. An important example is the "opposite category" $\mathcal{C}^{\text{opp}}$ to a category $\mathcal{C}$, which is a handy technical device and is discussed in Problems 78–80 at the end of the chapter.

In all of the above examples of categories, the class of objects fails to be a set. This behavior is typical. However, it does not cause problems in practice because in any particular argument involving categories, we can restrict to a subcategory for which the objects do form a set.[17]

[17] For the interested reader, a book that pays closer attention to the inherent set-theoretic difficulties in the theory is Mac Lane's *Categories for the Working Mathematician.*

If $\mathcal{C}$ is a category, a morphism $\varphi \in \text{Morph}(A, B)$ is said to be an **isomorphism** if there exists a morphism $\psi \in \text{Morph}(B, A)$ such that $\psi\varphi = 1_A$ and $\varphi\psi = 1_B$. In this case we say that A is **isomorphic** to B in the category $\mathcal{C}$. Let us check that the morphism ψ is unique if it exists. In fact, if ψ' is a member of $\text{Morph}(B, A)$ with $\psi'\varphi = 1_A$ and $\varphi\psi' = 1_B$, then $\psi = 1_A\psi = (\psi'\varphi)\psi = \psi'(\varphi\psi) = \psi' 1_B = \psi'$. We can therefore call ψ the **inverse** to φ.

The relation "is isomorphic to" is an equivalence relation.[18] In fact, the relation is symmetric by definition, and it is reflexive because $1_A \in \text{Morph}(A, A)$ has 1_A as inverse. For transitivity let $\varphi_1 \in \text{Morph}(A, B)$ and $\varphi_2 \in \text{Morph}(B, C)$ be isomorphisms, with respective inverses $\psi_1 \in \text{Morph}(B, A)$ and $\psi_2 \in \text{Morph}(C, B)$. Then $\varphi_2\varphi_1$ is in $\text{Morph}(A, C)$, and $\psi_1\psi_2$ is in $\text{Morph}(C, A)$. Calculation gives $(\psi_1\psi_2)(\varphi_2\varphi_1) = \psi_1(\psi_2(\varphi_2\varphi_1)) = \psi_1((\psi_2\varphi_2)\varphi_1) = \psi_1(1_B\varphi_1) = \psi_1\varphi_1 = 1_A$, and similarly $(\varphi_2\varphi_1)(\psi_1\psi_2) = 1_C$. Therefore $\varphi_2\varphi_1 \in \text{Morph}(A, C)$ is an isomorphism, and "is isomorphic to" is an equivalence relation. When A is isomorphic to B, it is permissible to say that A and B are **isomorphic**.

The next step is to abstract a frequent kind of construction that we have used with our categories. If $\mathcal{C}$ and $\mathcal{D}$ are two categories, a **covariant functor** $F : \mathcal{C} \to \mathcal{D}$ associates to each object A in $\text{Obj}(\mathcal{C})$ an object $F(A)$ in $\text{Obj}(\mathcal{D})$ and to each pair of objects A and B and morphism f in $\text{Morph}_{\mathcal{C}}(A, B)$ a morphism $F(f)$ in $\text{Morph}_{\mathcal{D}}(F(A), F(B))$ such that

(i) $F(gf) = F(g)F(f)$ for $f \in \text{Morph}_{\mathcal{C}}(A, B)$ and $g \in \text{Morph}_{\mathcal{C}}(B, C)$,
(ii) $F(1_A) = 1_{F(A)}$ for A in $\text{Obj}(\mathcal{C})$.

EXAMPLES OF COVARIANT FUNCTORS.

(1) Inclusion of a subcategory into a category is a covariant functor.

(2) Let $\mathcal{C}$ be the category of all sets. If F carries each set X to the set 2^X of all subsets of X, then F is a covariant functor as soon as its effect on functions between sets, i.e., its effect on morphisms, is defined in an appropriate way. Namely, if $f : X \to Y$ is a function, then $F(f)$ is to be a function from $F(X) = 2^X$ to $F(Y) = 2^Y$. That is, we need a definition of $F(f)(A)$ as a subset of Y whenever A is a subset of X. A natural way of making such a definition is to put $F(f)(A) = f(A)$, and then F is indeed a covariant functor.

(3) Let $\mathcal{C}$ be any of Examples 2 through 6 of categories above, and let $\mathcal{D}$ be the category of all sets, as in Example 1 of categories. If F carries an object A in $\mathcal{C}$ (i.e., a vector space, group, ring, etc.) into its underlying set and carries each morphism into its underlying function between two sets, then F is a covariant functor and furnishes an example of what is called a **forgetful functor**.

[18] Technically one considers relations only when they are defined on sets, and the class of objects in a category is typically not a set. However, just as with vector spaces, groups, and so on, we can restrict attention in any particular situation to a subcategory for which the objects do form a set, and then there is no difficulty.

(4) Let $\mathcal{C}$ be the category of all vector spaces over a field $\mathbb{F}$, let U be a vector space over $\mathbb{F}$, and let $F : \mathcal{C} \to \mathcal{C}$ be defined on a vector space to be the vector space of linear maps $F(V) = \mathrm{Hom}_{\mathbb{F}}(U, V)$. The set of morphisms $\mathrm{Morph}_{\mathcal{C}}(V_1, V_2)$ is $\mathrm{Hom}_{\mathbb{F}}(V_1, V_2)$. If f is in $\mathrm{Morph}_{\mathcal{C}}(V_1, V_2)$, then $F(f)$ is to be in $\mathrm{Morph}_{\mathcal{C}}\big(\mathrm{Hom}_{\mathbb{F}}(U, V_1), \mathrm{Hom}_{\mathbb{F}}(U, V_2)\big)$, and the definition is that $F(f)(L) = f \circ L$ for $L \in \mathrm{Hom}_{\mathbb{F}}(U, V_1)$. Then F is a covariant functor: to check that $F(gf) = F(g)F(f)$ when g is in $\mathrm{Morph}_{\mathcal{C}}(V_2, V_3)$, we write $F(gf)(L) = gf \circ L = g \circ fL = g \circ F(f) = F(g)F(f)$.

(5) Let $\mathcal{C}$ be the category of all groups, let $\mathcal{D}$ be the category of all sets, let G be a group, and let $F : \mathcal{C} \to \mathcal{D}$ be the functor defined as follows. For a group H, $F(H)$ is the set of all group homomorphisms from G into H. The set of morphisms $\mathrm{Morph}_{\mathcal{C}}(H_1, H_2)$ is the set of group homomorphisms from H_1 into H_2. If f is in $\mathrm{Morph}_{\mathcal{C}}(H_1, H_2)$, then $F(f)$ is to be a function with domain the set of homomorphisms from G into H_1 and with range the set of homomorphisms from G into H_2. Let $F(f)(\varphi) = \varphi \circ f$. Then F is a covariant functor.

(6) Let $\mathcal{C}$ be the category of all sets, and let $\mathcal{D}$ be the category of all abelian groups. To a set S, associate the free abelian group $F(S)$ with S as $\mathbb{Z}$ basis. If $f : S \to S'$ is a function, then the universal mapping property of external direct sums of abelian groups (Proposition 4.17) yields a corresponding group homomorphism from $F(S)$ to $F(S')$, and we define this group homomorphism to be $F(f)$. Then F is a covariant functor.

(7) Let $\mathcal{C}$ be the category of all finite sets, fix a commutative ring R with identity, and let $\mathcal{D}$ be the category of all commutative rings with identity. To a finite set S, associate the commutative ring $F(S) = R[\{X_s \mid s \in S\}]$. If $f : S \to S'$ is a function, then the properties of substitution homomorphisms give us a corresponding homomorphism of rings with identity carrying $F(S)$ to $F(S')$, and the result is a covariant functor.

There is a second kind of functor of interest to us. If $\mathcal{C}$ and $\mathcal{D}$ are two categories, a **contravariant functor** $F : \mathcal{C} \to \mathcal{D}$ associates to each object A in $\mathrm{Obj}(\mathcal{C})$ an object $F(A)$ in $\mathrm{Obj}(\mathcal{D})$ and to each pair of objects A and B and morphism f in $\mathrm{Morph}_{\mathcal{C}}(A, B)$ a morphism $F(f)$ in $\mathrm{Morph}_{\mathcal{D}}(F(B), F(A))$ such that

(i) $F(gf) = F(f)F(g)$ for $f \in \mathrm{Morph}_{\mathcal{C}}(A, B)$ and $g \in \mathrm{Morph}_{\mathcal{D}}(B, C)$,
(ii) $F(1_A) = 1_{F(A)}$ for A in $\mathrm{Obj}(\mathcal{C})$.

EXAMPLES OF CONTRAVARIANT FUNCTORS.

(1) Let $\mathcal{C}$ be the category of all vector spaces over a field $\mathbb{F}$, let W be a vector space over $\mathbb{F}$, and let $F : \mathcal{C} \to \mathcal{C}$ be defined on a vector space to be the vector space of linear maps $F(V) = \mathrm{Hom}_{\mathbb{F}}(V, W)$. The set of morphisms $\mathrm{Morph}_{\mathcal{C}}(V_1, V_2)$ is $\mathrm{Hom}_{\mathbb{F}}(V_1, V_2)$. If f is in $\mathrm{Morph}_{\mathcal{C}}(V_1, V_2)$, then $F(f)$ is to be in

$\mathrm{Morph}_{\mathcal{C}}\big(\mathrm{Hom}_{\mathbb{F}}(V_2, W), \mathrm{Hom}_{\mathbb{F}}(V_1, W)\big)$, and the definition is that $F(f)(L) = L \circ f$ for $L \in \mathrm{Hom}_{\mathbb{F}}(V_2, W)$. Then F is a contravariant functor: to check that $F(gf) = F(f)F(g)$ when g is in $\mathrm{Morph}_{\mathcal{C}}(V_2, V_3)$, we write $F(gf)(L) = L \circ gf = Lg \circ f = F(f)(Lg) = F(f)F(g)(L)$.

(2) Let $\mathcal{C}$ be the category of all vector spaces over a field $\mathbb{F}$, define F of a vector space V to be the dual vector space V', and define F of a linear mapping f between two vector spaces V and W to be the contragredient f^t carrying W' into V', defined by $f^t(w')(v) = w'(f(v))$. This is the special case of Example 1 of contravariant functors in which $W = \mathbb{F}$. Hence F is a contravariant functor.

(3) Let $\mathcal{C}$ be the category of all groups, let $\mathcal{D}$ be the category of all sets, let G be a group, and let $F : \mathcal{C} \to \mathcal{D}$ be the functor defined as follows. For a group H, $F(H)$ is the set of all group homomorphisms from H into G. The set of morphisms $\mathrm{Morph}_{\mathcal{C}}(H_1, H_2)$ is the set of group homomorphisms from H_1 into H_2. If f is in $\mathrm{Morph}_{\mathcal{C}}(H_1, H_2)$, then $F(f)$ is to be a function with domain the set of homomorphisms from H_2 into G and with range the set of homomorphisms from H_1 into G. The definition is $F(f)(\varphi) = \varphi \circ f$. Then F is a contravariant functor.

It is an important observation about functors that the composition of two functors is a functor. This is immediate from the definition. If the two functors are both covariant or both contravariant, then the composition is covariant. If one of them is covariant and the other is contravariant, then the composition is contravariant.

$$\begin{array}{ccc} A & \xrightarrow{\alpha} & B \\ {\scriptstyle\beta}\downarrow & & \downarrow{\scriptstyle\gamma} \\ C & \xrightarrow[\delta]{} & D \end{array}$$

FIGURE 4.9. A square diagram. The square commutes if $\gamma\alpha = \delta\beta$.

In the subject of category theory, a great deal of information is conveyed by "commutative diagrams" of objects and morphisms. By a **diagram** is meant a directed graph, usually but not necessarily planar, in which the vertices represent some relevant objects in a category and the arrows from one vertex to another represent morphisms of interest between pairs of these objects. Often the vertices and arrows are labeled, but in fact labels on the vertices can be deduced from the labels on the arrows since any morphism determines its "domain" and "range" as a consequence of defining property (i) of categories. A diagram is said to be **commutative** if for each pair of vertices A and B and each directed path from A to B, the compositions of the morphisms along each path are the same. For

example a square as in Figure 4.9 is commutative if $\gamma\alpha = \delta\beta$. The triangular diagrams in Figures 4.1 through 4.8 are all commutative.

$$\begin{array}{ccc} F(A) & \xrightarrow{F(\alpha)} & F(B) \\ {\scriptstyle F(\beta)}\downarrow & & \downarrow{\scriptstyle F(\gamma)} \\ F(C) & \xrightarrow[F(\delta)]{} & F(D) \end{array} \qquad \text{and} \qquad \begin{array}{ccc} G(A) & \xleftarrow{G(\alpha)} & G(B) \\ {\scriptstyle G(\beta)}\uparrow & & \uparrow{\scriptstyle G(\gamma)} \\ G(C) & \xleftarrow[G(\delta)]{} & G(D) \end{array}$$

FIGURE 4.10. Diagrams obtained by applying a covariant functor F and a contravariant functor G to the diagram in Figure 4.9.

Functors can be applied to diagrams, yielding new diagrams. For example, suppose that Figure 4.9 is a diagram in the category $\mathcal{C}$, that $F : \mathcal{C} \to \mathcal{D}$ is a covariant functor, and that $G : \mathcal{C} \to \mathcal{D}$ is a contravariant functor. Then we can apply F and G to the diagram in Figure 4.9, obtaining the two diagrams in the category $\mathcal{D}$ that are pictured in Figure 4.10. *If the diagram in Figure 4.9 is commutative, then so are the diagrams in Figure 4.10,* as a consequence of the effect of functors on compositions of morphisms.

The subject of category theory seeks to analyze functors that make sense for all categories, or at least all categories satisfying some additional properties. The most important investigation of this kind is concerned with homology and cohomology, as well as their ramifications, for "abelian categories," which include several important examples affecting algebra, topology, and several complex variables. The topic in question is called "homological algebra" and is discussed further in *Advanced Algebra*, particularly in Chapter IV.

There are a number of other functors that are investigated in category theory, and we mention four:

- products, including direct products,
- coproducts, including direct sums,
- direct limits, also called inductive limits,
- inverse limits, also called projective limits.

We discuss general products and coproducts in the present section, omitting a general discussion of direct limits and inverse limits. Inverse limits will arise in Section VII.6 of *Advanced Algebra* for one category in connection with Galois groups, but we shall handle that one situation on its own without attempting a generalization. An attempt in the 1960s to recast as much mathematics as possible in terms of category theory is now regarded by many mathematicians as having been overdone, and it seems wiser to cast bodies of mathematics in the framework of category theory only when doing so can be justified by the amount of time saved by eliminating redundant arguments.

When a category $\mathcal{C}$ and a nonempty set S are given, we can define a category $\mathcal{C}^S$. The objects of $\mathcal{C}^S$ are functions on S with the property that the value of the function at each s in S is in $\mathrm{Obj}(\mathcal{C})$, two such functions being regarded as the same if they consist of the same ordered pairs.[19] Let us refer to such a function as an S-**tuple** of members of $\mathrm{Obj}(\mathcal{C})$, denoting it by an expression like $\{X_s\}_{s\in S}$. A morphism in $\mathrm{Morph}_{\mathcal{C}^S}\left(\{X_s\}_{s\in S}, \{Y_s\}_{s\in S}\right)$ is an S-tuple $\{f_s\}_{s\in S}$ of morphisms of $\mathcal{C}$ such that f_s lies in $\mathrm{Morph}_{\mathcal{C}}(X_s, Y_s)$ for all s, and the law of composition of such morphisms takes place coordinate by coordinate.

Let $\{X_s\}_{s\in S}$ be an object in $\mathcal{C}^S$. A **product** of $\{X_s\}_{s\in S}$ is a pair $(X, \{p_s\}_{s\in S})$ such that X is in $\mathrm{Obj}(\mathcal{C})$ and each p_s is in $\mathrm{Morph}_{\mathcal{C}}(X, X_s)$ with the following **universal mapping property**: whenever A in $\mathrm{Obj}(\mathcal{C})$ is given and a morphism $\varphi_s \in \mathrm{Morph}_{\mathcal{C}}(A, X_s)$ is given for each s, then there exists a unique morphism $\varphi \in \mathrm{Morph}_{\mathcal{C}}(A, X)$ such that $p_s\varphi = \varphi_s$ for all s. The relevant diagram is pictured in Figure 4.11.

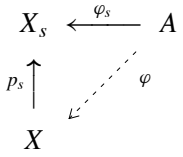


FIGURE 4.11. Universal mapping property of a product in a category.

EXAMPLES OF PRODUCTS.

(1) Products exist in the category of vector spaces over a field $\mathbb{F}$. If vector spaces V_s indexed by a nonempty set S are given, then their product exists in the category, and an example is their external direct product $\prod_{s\in S} V_s$, according to Figure 2.4 and the discussion around it.

(2) Products exist in the category of all groups. If groups G_s indexed by a nonempty set S are given, then their product exists in the category, and an example is their external direct product $\prod_{s\in S} G_s$, according to Figure 4.2 and Proposition 4.15. If the groups G_s are abelian, then $\prod_{s\in S} G_s$ is abelian, and it follows that products exist in the category of all abelian groups.

(3) Products exist in the category of all sets. If sets X_s indexed by a nonempty set S are given, then their product exists in the category, and an example is their Cartesian product $\bigtimes_{s\in S} X_s$, as one easily checks.

(4) Products exist in the category of all rings and in the category of all rings with identity. If objects R_s in the category indexed by a nonempty set S are given, then

[19] In other words, the range of such a function is considered as irrelevant. We might think of the range as $\mathrm{Obj}(\mathcal{C})$ except for the fact that a function is supposed to have a *set* as range and $\mathrm{Obj}(\mathcal{C})$ need not be a set.

their product may be taken as an abelian group to be the external direct product $\prod_{s\in S} R_s$, with multiplication defined coordinate by coordinate, and the group homomorphisms p_s are easily checked to be morphisms in the category.

A product of objects in a category need not exist in the category. An artificial example may be formed as follows: Let $\mathcal{C}$ be a category with one object G, namely a group of order 2, and let $\mathrm{Morph}(G, G) = \{0, 1_G\}$, the law of composition being the usual composition. Let S be a 2-element set, and let the corresponding objects be $X_1 = G$ and $X_2 = G$. The claim is that the product $X_1 \times X_2$ does not exist in $\mathcal{C}$. In fact, take $A = G$. There are four S-tuples of morphisms (φ_1, φ_2) meeting the conditions of the definition. Yet the only possibility for the product is $X = G$, and then there are only two possible φ's in $\mathrm{Morph}(A, X)$. Hence we cannot account for all possible S-tuples of morphisms, and the product cannot exist.

The thing that category theory addresses is the uniqueness. A product is always unique up to canonical isomorphism, according to Proposition 4.63. We proved uniqueness for products in the special cases of Examples 1 and 2 above in Propositions 2.32 and 4.16.

Proposition 4.63. Let $\mathcal{C}$ be a category, and let S be a nonempty set. If $\{X_s\}_{s\in S}$ is an object in $\mathcal{C}^S$ and if $(X, \{p_s\})$ and $(X', \{p'_s\})$ are two products, then there exists a unique morphism $\Phi : X' \to X$ such that $p'_s = p_s \circ \Phi$ for all $s \in S$, and Φ is an isomorphism.

REMARK. There is no assertion that p_s is onto X_s. In fact, "onto" has no meaning for a general category.

PROOF. In Figure 4.11 let $A = X'$ and $\varphi_s = p'_s$. If $\Phi \in \mathrm{Morph}(X', X)$ is the morphism produced by the fact that X is a direct product, then we have $p_s\Phi = p'_s$ for all s. Reversing the roles of X and X', we obtain a morphism $\Phi' \in \mathrm{Morph}(X, X')$ with $p'_s\Phi' = p_s$ for all s. Therefore $p_s(\Phi\Phi') = (p_s\Phi)\Phi' = p'_s\Phi' = p_s$.

In Figure 4.11 we next let $A = X$ and $\varphi_s = p_s$ for all s. Then the identity 1_X in $\mathrm{Morph}(X, X)$ has the same property $p_s 1_X = p_s$ relative to all p_s that $\Phi\Phi'$ has, and the uniqueness in the statement of the universal mapping property implies that $\Phi\Phi' = 1_X$. Reversing the roles of X and X', we obtain $\Phi'\Phi = 1_{X'}$. Therefore Φ is an isomorphism.

For uniqueness suppose that $\Psi \in \mathrm{Morph}(X', X)$ is another morphism with $p'_s = p_s\Psi$ for all $s \in S$. Then the argument of the previous paragraph shows that $\Phi'\Psi = 1_{X'}$. Consequently $\Psi = 1_X\Psi = (\Phi\Phi')\Psi = \Phi(\Phi'\Psi) = \Phi 1_{X'} = \Phi$, and $\Psi = \Phi$. □

If products always exist in a particular category, they are not unique, only unique up to canonical isomorphism. Such a product is commonly denoted by

$\prod_{s\in S} X_s$, even though it is not uniquely defined. *It is customary to treat the product over S as a covariant functor $F : \mathcal{C}^S \to \mathcal{C}$*, the effect of the functor on objects being given by $F(\{X_s\}_{s\in S}) = \prod_{s\in S} X_s$. For a well-defined functor we have to fix a choice of product for each object under consideration[20] in $\mathrm{Obj}(\mathcal{C}^S)$. For the effect of F on morphisms, we argue with the universal mapping property. Thus let $\{X_s\}_{s\in S}$ and $\{Y_s\}_{s\in S}$ be objects in $\mathcal{C}^S$, let f_s be in $\mathrm{Morph}_{\mathcal{C}}(X_s, Y_s)$ for all s, and let the products in question be $\left(\prod_{s\in S} X_s, \{p_s\}_{s\in S}\right)$ and $\left(\prod_{s\in S} Y_s, \{q_s\}_{s\in S}\right)$. Then $f_{s_0} p_{s_0}$ is in $\mathrm{Morph}_{\mathcal{C}}\left(\prod_{s\in S} X_s, Y_{s_0}\right)$ for each s_0, and the universal mapping property gives us f in $\mathrm{Morph}_{\mathcal{C}}\left(\prod_{s\in S} X_s, \prod_{s\in S} Y_s\right)$ such that $q_s f = f_s p_s$ for all s. We define this f to be $F(\{f_s\}_{s\in S})$, and we readily check that F is a functor.

We turn to coproducts, which include direct sums. Let $\{X_s\}_{s\in S}$ be an object in $\mathcal{C}^S$. A **coproduct** of $\{X_s\}_{s\in S}$ is a pair $(X, \{i_s\}_{s\in S})$ such that X is in $\mathrm{Obj}(\mathcal{C})$ and each i_s is in $\mathrm{Morph}_{\mathcal{C}}(X_s, X)$ with the following **universal mapping property**: whenever A in $\mathrm{Obj}(\mathcal{C})$ is given and a morphism $\varphi_s \in \mathrm{Morph}_{\mathcal{C}}(X_s, A)$ is given for each s, then there exists a unique morphism $\varphi \in \mathrm{Morph}_{\mathcal{C}}(X, A)$ such that $\varphi i_s = \varphi_s$ for all s. The relevant diagram is pictured in Figure 4.12.

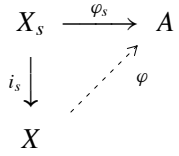


FIGURE 4.12. Universal mapping property of a coproduct in a category.

EXAMPLES OF COPRODUCTS.

(1) Coproducts exist in the category of vector spaces over a field $\mathbb{F}$. If vector spaces V_s indexed by a nonempty set S are given, then their coproduct exists in the category, and an example is their external direct sum $\bigoplus_{s\in S} V_s$, according to Figure 2.5 and the discussion around it.

(2) Coproducts exist in the category of all abelian groups. If abelian groups G_s indexed by a nonempty set S are given, then their coproduct exists in the category, and an example is their external direct sum $\bigoplus_{s\in S} G_s$, according to Figure 4.4 and Proposition 4.17.

(3) Coproducts exist in the category of all sets. If sets X_s indexed by a nonempty set S are given, then their coproduct exists in the category, and an example is their disjoint union $\bigcup_{s\in S} \{(x_s, s) \mid x_s \in X_s\}$. The verification appears as Problem 74 at the end of the chapter.

[20] Since $\mathrm{Obj}(\mathcal{C}^S)$ need not be a set, it is best to be wary of applying the Axiom of Choice when the indexing of sets is given by $\mathrm{Obj}(\mathcal{C}^S)$. Instead, one makes the choice only for all objects in some set of objects large enough for a particular application.

(4) Coproducts exist in the category of all groups. Suppose that groups G_s indexed by a nonempty set S are given. It will be shown in Chapter VII that the coproduct is the "free product" $\ast_{s\in S} G_s$ that is defined in that chapter. In the special case that each G_s is the group $\mathbb{Z}$ of integers, the free product coincides with the free group on S. Therefore, even if all the groups G_s are abelian, their coproduct need not be a subgroup of the direct product and need not even be abelian. In particular it need not coincide with the direct sum.

A coproduct of objects in a category need not exist in the category. Problem 76 at the end of the chapter offers an example that the reader is invited to check.

Proposition 4.64. Let $\mathcal{C}$ be a category, and let S be a nonempty set. If $\{X_s\}_{s\in S}$ is an object in $\mathcal{C}^S$ and if $(X, \{i_s\})$ and $(X', \{i'_s\})$ are two coproducts, then there exists a unique morphism $\Phi : X \to X'$ such that $i'_s = \Phi \circ i_s$ for all $s \in S$, and Φ is an isomorphism.

REMARKS. There is no assertion that i_s is one-one. In fact, "one-one" has no meaning for a general category. This proposition may be derived quickly from Proposition 4.63 by a certain duality argument that is discussed in Problems 78–80 at the end of the chapter. Here we give a direct argument without taking advantage of duality.

PROOF. In Figure 4.12 let $A = X'$ and $\varphi_s = i'_s$. If $\Phi \in \mathrm{Morph}(X, X')$ is the morphism produced by the fact that X is a coproduct, then we have $\Phi\, i_s = i'_s$ for all s. Reversing the roles of X and X', we obtain a morphism $\Phi' \in \mathrm{Morph}(X', X)$ with $\Phi' i'_s = i_s$ for all s. Therefore $(\Phi'\Phi) i_s = \Phi' i'_s = i_s$.

In Figure 4.12 we next let $A = X$ and $\varphi_s = i_s$ for all s. Then the identity 1_X in $\mathrm{Morph}(X, X)$ has the same property $1_X i_s = i_s$ relative to all i_s that $\Phi'\Phi$ has, and the uniqueness says that $\Phi'\Phi = 1_X$. Reversing the roles of X and X', we obtain $\Phi\Phi' = 1_{X'}$. Therefore Φ is an isomorphism.

For uniqueness suppose that $\Psi \in \mathrm{Morph}(X, X')$ is another morphism with $i'_s = \Psi\, i_s$ for all $s \in S$. Then the argument of the previous paragraph shows that $\Phi'\Psi = 1_X$. Consequently $\Psi = 1_{X'}\Psi = (\Phi\Phi')\Psi = \Phi(\Phi'\Psi) = \Phi 1_X = \Phi$, and $\Psi = \Phi$. □

If coproducts always exist in a particular category, they are not unique, only unique up to canonical isomorphism. Such a coproduct is commonly denoted by $\coprod_{s\in S} X_s$, even though it is not uniquely defined. As with product, *it is customary to treat the coproduct over S as a covariant functor* $F : \mathcal{C}^S \to \mathcal{C}$, the effect of the functor on objects being given by $F(\{X_s\}_{s\in S}) = \coprod_{s\in S} X_s$. For a well-defined functor we have to fix a choice of coproduct for each object under consideration in $\mathrm{Obj}(\mathcal{C}^S)$. For the effect of F on morphisms, we argue with the universal mapping property. Thus let $\{X_s\}_{s\in S}$ and $\{Y_s\}_{s\in S}$ be objects in $\mathcal{C}^S$, let f_s be in

$\mathrm{Morph}_{\mathcal{C}}(X_s, Y_s)$ for all s, and let the coproducts in question be $\left(\coprod_{s\in S} X_s, \{i_s\}_{s\in S}\right)$ and $\left(\coprod_{s\in S} Y_s, \{j_s\}_{s\in S}\right)$. Then $j_{s_0} f_{s_0}$ is in $\mathrm{Morph}_{\mathcal{C}}\left(X_{s_0}, \coprod_{s\in S} Y_s\right)$ for each s_0, and the universal mapping property gives us f in $\mathrm{Morph}_{\mathcal{C}}\left(\coprod_{s\in S} X_s, \coprod_{s\in S} Y_s\right)$ such that $f i_s = j_s f_s$ for all s. We define this f to be $F(\{f_s\}_{s\in S})$, and we readily check that F is a functor.

Universal mapping properties occur in other contexts than for products and coproducts. We have already seen them in connection with homomorphisms on free abelian groups and with substitution homomorphisms on polynomial rings, and more such properties will occur in the development of tensor products in Chapter VI. A general framework for discussing universal mapping properties appears in the problems at the end of Chapter VI.

12. Problems

1. Let G be a group in which all elements other than the identity have order 2. Prove that G is abelian.

2. The dihedral group D_4 of order 8 can be viewed as a subgroup of the symmetric group $\mathfrak{S}_4$ of order 8. Find 8 explicit permutations in $\mathfrak{S}_4$ forming a subgroup isomorphic to D_4.

2A. Let g be an element of finite order $\mathrm{ord}(g)$ in a group G. Prove that
   (a) g^{-1} has the same order as g.
   (b) $g^k = 1$ if and only if $\mathrm{ord}(g)$ divides k.
   (c) for each $r \in \mathbb{Z}$, the order of g^r is $\mathrm{ord}(g)/\mathrm{GCD}(\mathrm{ord}(g), r)$.

3. Suppose G is a finite group, H is a subgroup, and $a \in G$ is an element with a^l in H for some integer l with $\mathrm{GCD}(l, |G|) = 1$. Prove that a is in H.

4. Let G be a group, and define a new group G' to have the same underlying set as G but to have multiplication given by $a \circ b = ba$. Prove that G' is a group and that it is isomorphic to G.

5. Prove that if G is an abelian group and n is an integer, then $a \mapsto a^n$ is a homomorphism of G. Give an example of a nonabelian group for which $a \mapsto a^2$ is not a homomorphism.

6. Suppose that G is a group and that H and K are *normal* subgroups of G with $H \cap K = \{1\}$. Verify that the set HK of products is a subgroup and that this subgroup is isomorphic as a group to the external direct product $H \times K$.

7. Take as known that 8191 is prime, so that $\mathbb{F}_{8191}$ is a field. Without carrying through the computations and without advocating trial and error, describe what steps you would carry out to solve for x mod 8191 such that $1234x \equiv 1$ mod 8191.

8. **(Wilson's Theorem)** Let p be an odd prime. Starting from the fact that $1, \ldots, p-1$ are roots of the polynomial $X^{p-1} - 1 \equiv 0 \bmod p$ in $\mathbb{F}_p$, prove that $(p-1)! \equiv -1 \bmod p$.

9. Classify, up to isomorphism, all groups of order p^2 if p is a prime.

10. This problem concerns conjugacy classes in a group G.
    (a) Prove that all elements of a conjugacy class have the same order.
    (b) Prove that if ab is in a conjugacy class, so is ba.

11. (a) Find explicitly all the conjugacy classes in the alternating group $\mathfrak{A}_4$.
    (b) For each conjugacy class in $\mathfrak{A}_4$, find the centralizer of one element in the class.
    (c) Prove that $\mathfrak{A}_4$ has no subgroup isomorphic to C_6 or $\mathfrak{S}_3$.

12. Prove that the alternating group $\mathfrak{A}_5$ has no subgroup of order 30.

13. Let G be a nonabelian group of order p^n, where p is prime. Prove that any subgroup of order p^{n-1} is normal.

14. Let G be a finite group, and let H be a normal subgroup. If $|H| = p$ and p is the smallest prime dividing $|G|$, prove that H is contained in the center of G.

15. Let G be a group. An automorphism of G of the form $x \mapsto gxg^{-1}$ is called an **inner automorphism**. Prove that the set of inner automorphisms is a normal subgroup of the group $\operatorname{Aut} G$ of all automorphisms and is isomorphic to G/Z_G.

16. (a) Prove that $\operatorname{Aut} C_m$ is isomorphic to $(\mathbb{Z}/m\mathbb{Z})^\times$.
    (b) Find a value of m for which $\operatorname{Aut} C_m$ is not cyclic.

17. Fix $n \geq 2$. In the symmetric group $\mathfrak{S}_n$, for each integer k with $1 \leq k \leq n/2$, let C_k be the set of elements in $\mathfrak{S}_n$ that are products of k disjoint transpositions.
    (a) Prove that if τ is an automorphism of $\mathfrak{S}_n$, then $\tau(C_1) = C_k$ for some k.
    (b) Prove that $|C_k| = \dbinom{n}{2k} \dfrac{(2k)!}{2^k k!}$.
    (c) Prove that $|C_k| \neq |C_1|$ unless $k = 1$ or $n = 6$. (Educational note: From this, it follows that $\tau(C_1) = C_1$ except possibly when $n = 6$. One can deduce as a consequence that every automorphism of $\mathfrak{S}_n$ is inner except possibly when $n = 6$.)

18. Give an example: G is a group with a normal subgroup N, N has a subgroup M that is normal in N, yet M is not normal in G.

19. Show that the cyclic group C_{rs} is isomorphic to $C_r \times C_s$ if and only if $\operatorname{GCD}(r, s) = 1$.

20. How many abelian groups, up to isomorphism, are there of order 27?

21. Let G be the free abelian group with $\mathbb{Z}$ basis $\{x_1, x_2, x_3\}$. Let H be the subgroup of G generated by $\{u_1, u_2, u_3\}$, where
$$u_1 = 3x_1 + 2x_2 + 5x_3,$$
$$u_2 = x_2 + 3x_3,$$
$$u_3 = x_2 + 5x_3.$$
Express G/H as a direct sum of cyclic groups.

22. Let $\{e_1, e_2, e_3, e_4\}$ be the standard basis of $\mathbb{R}^4$. Let G be the additive subgroup of $\mathbb{R}^4$ generated by the four elements
$$e_1, \quad e_1 + e_2, \quad \tfrac{1}{2}(e_1 + e_2 + e_3 + e_4), \quad \tfrac{1}{2}(e_1 + e_2 + e_3 - e_4),$$
and let H be the subgroup of G generated by the four elements
$$e_1 - e_2, \quad e_2 - e_3, \quad e_3 - e_4, \quad e_3 + e_4.$$
Identify the abelian group G/H as a direct sum of cyclic groups.

23. Let G be the free abelian group with $\mathbb{Z}$ basis $\{x_1, \dots, x_n\}$, and let H be the subgroup generated by $\{u_1, \dots, u_m\}$, where $\begin{pmatrix} u_1 \\ \vdots \\ u_m \end{pmatrix} = C \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ for an m-by-n matrix C of integers. Prove that the number of summands $\mathbb{Z}$ in the decomposition of G/H into cyclic groups is equal to the rank of the matrix C when C is considered as in $M_{mn}(\mathbb{Q})$.

24. Prove that every abelian group is the homomorphic image of a free abelian group.

25. Let G be a group, and let H and K be subgroups.
    (a) For x and y in G, prove that $xH \cap yK$ is empty or is a coset of $H \cap K$.
    (b) Deduce from (a) that if H and K have finite index in G, then so does $H \cap K$.

26. Let G be a free abelian group of finite rank n, and let H be a free abelian subgroup of rank n. Prove that H has finite index in G.

27. Let $G = \mathfrak{S}_4$ be the symmetric group on four letters.
    (a) Find a Sylow 2-subgroup of G. How many Sylow 2-subgroups are there, and why?
    (b) Find a Sylow 3-subgroup of G. How many Sylow 3-subgroups are there, and why?

28. Let H be a subgroup of a group G. Prove or disprove that the normalizer $N(H)$ of H in G is a normal subgroup of G.

29. How many elements of order 7 are there in a simple group of order 168?

30. Let G be a group of order pq^2, where p and q are primes with $p < q$. Let S_p and S_q be Sylow subgroups for the primes p and q. Prove that G is a semidirect product of S_p and S_q with S_q normal.

31. Suppose that G is a finite group and that H is a subgroup whose index in G is a prime p. By considering the action of G on the set of subgroups conjugate to H and considering the possibilities for the normalizer $N(H)$, determine the possibilities for the number of subgroups conjugate to H.

32. Let G be a group of order 24, let H be a subgroup of order 8, and assume that H is not normal.
    (a) Using the Sylow Theorems, explain why H has exactly 3 conjugates in G, counting H itself as one.
    (b) Show how to use the conjugates in (a) to define a homomorphism of G into the symmetric group $\mathfrak{S}_3$ on three letters.
    (c) Use the homomorphism of (b) to conclude that G is not simple.

33. Let G be a group of order 36. Arguing in the style of the previous problem, show that there is a nontrivial homomorphism of G into the symmetric group $\mathfrak{S}_4$.

34. Let G be a group of order $2pq$, where p and q are primes with $2 < p < q$.
    (a) Prove that if $q + 1 \neq 2p$, then a Sylow q-subgroup is normal.
    (b) Suppose that $q + 1 = 2p$, let H be a Sylow p-subgroup, and let K be a Sylow q-subgroup. Prove that at least one of H and K is normal, that the set HK of products is a subgroup, and that the subgroup HK is cyclic of index 2 in G.

Problems 35–38 concern the detection of isomorphisms among semidirect products. For the first two of the problems, let H and K be groups, and let $\varphi_1 : H \to \operatorname{Aut} K$ and $\varphi_2 : H \to \operatorname{Aut} K$ be homomorphisms.

35. Suppose that $\varphi_2 = \varphi_1 \circ \varphi$ for some automorphism φ of H. Define $\psi : H \times_{\varphi_2} K \to H \times_{\varphi_1} K$ by $\psi(h, k) = (\varphi(h), k)$. Prove that ψ is an isomorphism.

36. Suppose that $\varphi_2 = \varphi \circ \varphi_1$ for some inner automorphism φ of $\operatorname{Aut} K$ in the sense of Problem 15, i.e., $\varphi : \operatorname{Aut} K \to \operatorname{Aut} K$ is to be given by $\varphi(x) = axa^{-1}$ with a in $\operatorname{Aut} K$. Define $\psi : H \times_{\varphi_1} K \to H \times_{\varphi_2} K$ by $\psi(h, k) = (h, a(k))$. Prove that ψ is an isomorphism.

37. Suppose that p and q are primes and that the cyclic group C_p acts on C_q by automorphisms with a nontrivial action. Prove that p divides $q - 1$.

38. Suppose that p and q are primes such that p divides $q - 1$. Let τ_1 and τ_2 be nontrivial homomorphisms from C_p to $\operatorname{Aut} C_q$. Prove that $C_p \times_{\tau_1} C_q \cong C_p \times_{\tau_2} C_q$, and conclude that there is only one nonabelian semidirect product $C_p \times_\tau C_q$ up to isomorphism.

Problems 39–44 discuss properties of groups of order 8, obtaining a classification of these groups up to isomorphism.

39. Prove that the five groups C_8, $C_4 \times C_2$, $C_2 \times C_2 \times C_2$, D_4, and H_8 are mutually nonisomorphic and that the first three exhaust the abelian groups of order 8, apart from isomorphisms.

40. (a) Find a composition series for the 8-element dihedral group D_4.
    (b) Find a composition series for the 8-element quaternion group H_8.
41. (a) Prove that every subgroup of the quaternion group H_8 is normal.
    (b) Identify the conjugacy classes in H_8.
    (c) Compute the order of $\mathrm{Aut}\, H_8$.
42. Suppose that G is a nonabelian group of order 8. Prove that G has an element of order 4 but no element of order 8.
43. Let G be a nonabelian group of order 8, and let K be the copy of C_4 generated by some element of order 4. If G has some element of order 2 that is not in K, prove that $G \cong D_4$.
44. Let G be a nonabelian group of order 8, and let K be the copy of C_4 generated by some element of order 4. If G has no element of order 2 that is not in K, prove that $G \cong H_8$.

Problems 45–48 classify groups of order 12, making use of Proposition 4.61, Problem 15, and Problems 35–38. Let G be a group of order 12, let H be a Sylow 3-subgroup, and let K be a Sylow 2-subgroup. Proposition 4.61 says that at least one of H and K is normal. Consequently there are three cases, and these are addressed by the first three of the problems.

45. Verify that there are only two possibilities for G up to isomorphism if G is abelian.
46. Suppose that K is normal, so that $G \cong H \times_\tau K$. Prove that either
    (i) τ is trivial or
    (ii) τ is nontrivial and $K \cong C_2 \times C_2$,
    and deduce that G is abelian if (i) holds and that $G \cong \mathfrak{A}_4$ if (ii) holds.
47. Suppose that H is normal, so that $G = K \times_\tau H$. Prove that one of the conditions
    (i) τ is trivial,
    (ii) $K \cong C_2 \times C_2$ and τ is nontrivial,
    (iii) $K \cong C_4$ and τ is nontrivial
    holds, and deduce that G is abelian if (i) holds, that $G \cong D_6$ if (ii) holds, and that G is nonabelian and is not isomorphic to $\mathfrak{A}_4$ or D_6 if (iii) holds.
48. In the setting of the previous problem, prove that there is one and only one group, up to isomorphism, satisfying condition (iii), and find the order of each of its elements.

Problems 49–52 assume that p and q are primes with $p < q$. The problems go in the direction of classifying finite groups of order p^2q.

49. If G is a group of order p^2q, prove that either $p^2q = 12$ or a Sylow q-subgroup is normal.
50. If p^2 divides $q-1$, exhibit three nonabelian groups of order p^2q that are mutually nonisomorphic.

51. If p divides $q-1$ but p^2 does not divide $q-1$, exhibit two nonabelian groups of order p^2q that are not isomorphic.

52. If p does not divide $q-1$, prove that any group of order p^2q is abelian.

Problems 53–54 concern nonabelian groups of order 27.

53. (a) Show that multiplication by the elements $1, 4, 7 \bmod 9$ defines a nontrivial action of $\mathbb{Z}/3\mathbb{Z}$ on $\mathbb{Z}/9\mathbb{Z}$ by automorphisms.
    (b) Show from (a) that there exists a nonabelian group of order 27.
    (c) Show that the group in (b) is generated by elements a and b that satisfy

$$a^9 = b^3 = b^{-1}aba^{-4} = 1.$$

54. Show that any nonabelian group of order 27 having a subgroup H isomorphic to C_9 and an element of order 3 not lying in H is isomorphic to the group constructed in the previous problem.

Problems 55–62 give a construction of infinitely many simple groups, some of them finite and some infinite. Let $\mathbb{F}$ be a field. For $n \geq 2$, let $\mathrm{SL}(n, \mathbb{F})$ be the special linear group for the space $\mathbb{F}^n$ of n-dimensional column vectors. The center Z of $\mathrm{SL}(n, \mathbb{F})$ consists of the scalar multiples of the identity, the scalar being an n^{th} root of 1. Let $\mathrm{PSL}(n, \mathbb{F}) = \mathrm{SL}(n, \mathbb{F})/Z$. It is known that $\mathrm{PSL}(n, \mathbb{F})$ is simple except for $\mathrm{PSL}(2, \mathbb{F}_2)$ and $\mathrm{PSL}(2, \mathbb{F}_3)$. These problems will show that $\mathrm{PSL}(2, \mathbb{F})$ is simple if $|\mathbb{F}| > 5$ and $\mathbb{F}$ is not of characteristic 2. Most of the argument will consider $\mathrm{SL}(2, \mathbb{F})$, and the passage to PSL will occur only at the very end. In Problems 56–61, G denotes a normal subgroup of $\mathrm{SL}(2, \mathbb{F})$ that is not contained in the center Z, and it is to be proved that $G = \mathrm{SL}(2, \mathbb{F})$.

55. Suppose that $\mathbb{F}$ is a finite field with q elements.
    (a) By considering the possibilities for the first column of a matrix and then considering the possibilities for the second column when the first column is fixed, compute $|\mathrm{GL}(2, \mathbb{F})|$ as a function of q.
    (b) By using the determinant homomorphism, compute $|\mathrm{SL}(2, \mathbb{F})|$ in terms of $|\mathrm{GL}(2, \mathbb{F})|$.
    (c) Taking into account that $\mathbb{F}$ does not have characteristic 2, prove that $|\mathrm{PSL}(2, \mathbb{F})| = \frac{1}{2}|\mathrm{SL}(2, \mathbb{F})|$.
    (d) Show for a suitable finite field $\mathbb{F}$ with more than 5 elements that $\mathrm{PSL}(2, \mathbb{F})$ has order 168.

56. Let M be a member of G that is not in Z. Since M is not scalar, there exists a column vector u with Mu not a multiple of u. Define $v = Mu$, so that (u, v) is an ordered basis of $\mathbb{F}^2$. By rewriting all matrices with the ordered basis (u, v), show that there is no loss in generality in assuming that G contains a matrix $A = \begin{pmatrix} 0 & -1 \\ 1 & c \end{pmatrix}$ if it is ultimately shown that $G = \mathrm{SL}(2, \mathbb{F})$.

57. Let a be a member of the multiplicative group $\mathbb{F}^\times$ to be chosen shortly, and let B be the member $\begin{pmatrix} ca & a^{-1} \\ -a & 0 \end{pmatrix}$ of $\mathrm{SL}(2, \mathbb{F})$. Prove that
    (a) $B^{-1}A^{-1}BA$ is upper triangular and is in G,
    (b) $B^{-1}A^{-1}BA$ has unequal diagonal entries if $a^4 \neq 1$,
    (c) the condition in (b) can be satisfied for a suitable choice of a under the assumption that $|\mathbb{F}| > 5$.
58. Suppose that $C = \begin{pmatrix} x & y \\ 0 & x^{-1} \end{pmatrix}$ is a member of G for some $x \neq \pm 1$ and some y. Taking $D = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ and forming $CDC^{-1}D^{-1}$, show that G contains a matrix $E = \begin{pmatrix} 1 & \lambda \\ 0 & 1 \end{pmatrix}$ with $\lambda \neq 0$.
59. By conjugating E by $\begin{pmatrix} \alpha & 0 \\ 0 & \alpha^{-1} \end{pmatrix}$, show that the set of λ in $\mathbb{F}$ such that $\begin{pmatrix} 1 & \lambda \\ 0 & 1 \end{pmatrix}$ is in G is closed under multiplication by squares and under addition and subtraction.
60. Using the identity $x = \frac{1}{4}(x+1)^2 - \frac{1}{4}(x-1)^2$, deduce from Problems 56–59 that G contains all matrices $\begin{pmatrix} 1 & \lambda \\ 0 & 1 \end{pmatrix}$ with $\lambda \in \mathbb{F}$.
61. Show that $\begin{pmatrix} 1 & \lambda \\ 0 & 1 \end{pmatrix}$ is conjugate to $\begin{pmatrix} 1 & 0 \\ -\lambda & 1 \end{pmatrix}$, and show that the set of all matrices $\begin{pmatrix} 1 & \lambda \\ 0 & 1 \end{pmatrix}$ and $\begin{pmatrix} 1 & 0 \\ \lambda' & 1 \end{pmatrix}$ generates $\mathrm{SL}(2, \mathbb{F})$. Conclude that $G = \mathrm{SL}(2, \mathbb{F})$.
62. Using the First Isomorphism Theorem, conclude that the only normal subgroup of $\mathrm{PSL}(2, \mathbb{F})$ other than $\{1\}$ is $\mathrm{PSL}(2, \mathbb{F})$ itself.

Problems 63–73 briefly introduce the theory of error-correcting codes. Let $\mathbb{F}$ be the finite field $\mathbb{Z}/2\mathbb{Z}$. The vector space $\mathbb{F}^n$ over $\mathbb{F}$ will be called **Hamming space**, and its members are regarded as "words" (potential messages consisting of 0's and 1's). The **weight** $\mathrm{wt}(c)$ of a word c is the number of nonzero entries in c. The **Hamming distance** $d(a, b)$ between words $a = (a_1, \ldots, a_n)$ and $b = (b_1, \ldots, b_n)$ is the weight of $a - b$, i.e., the number of indices i with $1 \leq i \leq n$ and $a_i \neq b_i$. A **code** is a nonempty subset C of $\mathbb{F}^n$, and the **minimal distance** $\delta(C)$ of a code is the smallest value of $d(a, b)$ for a and b in C with $a \neq b$. By convention if $|C| = 1$, take $\delta(C) = n + 1$. One imagines that members of C, which are called code words, are allowable messages, i.e., words that can be stored and retrieved, or transmitted and received. A code with minimal distance δ can then detect up to $\delta - 1$ errors in a word ostensibly from C that has been retrieved from storage or has been received in a transmission. The code can correct up to $(\delta - 1)/2$ errors because no word of $\mathbb{F}^n$ can be at distance $\leq (\delta - 1)/2$ from more than one word in C, by Problem 63 below. The interest is in **linear codes**, those for which C is a vector subspace. It is desirable that each message have a high percentage of content and a relatively low percentage of further information used for error correction; thus a fundamental theoretical problem for linear codes is to find the maximum dimension of a linear code if n and a lower bound on the minimal distance for the code are given. As a practical matter, information is likely to be processed in packets of a standard length,

such as some power of 2. In many situations packets can be reprocessed if they have been found to have errors. The initial interest is therefore in codes that can recognize and possibly correct a small number of errors. The problems in this set are continued at the ends of Chapters VII and IX.

63. Prove that the Hamming distance satisfies $d(a,b) \leq d(a,c) + d(c,b)$, and conclude that if a word w in $\mathbb{F}^n$ is at distance $\leq (D-1)/2$ from two distinct members of the linear code C, then $\delta(C) < D$.

64. Explain why the minimal distance $\delta(C)$ of a linear code $C \neq \{0\}$ is given by the minimal weight of the nonzero words in C.

65. Fix $n \geq 2$. List $\delta(C)$ and $\dim C$ for the following elementary linear codes:
    (a) $C = 0$.
    (b) $C = \mathbb{F}^n$.
    (c) **(Repetition code)** $C = \{0,\ (1,1,\ldots,1)\}$.
    (d) **(Parity-check code)** $C = \{c \in \mathbb{F}^n \mid \mathrm{wt}(c) \text{ is even}\}$. (Educational note: To use this code, one sends the message in the first $n-1$ bits and adjusts the last bit so that the word is in C. If there is at most one error in the word, this parity bit will tell when there is an error, but it will not tell where the error occurs.)

66. One way to get a sense of what members of a linear code C in $\mathbb{F}^n$ have small weight starts by making a basis for the code into the row vectors of a matrix and row reducing the matrix.
    (a) Taking into account the distinction between corner variables and independent variables in the process of row reduction, show that every basis vector of C has weight at most the sum of 1 and the number of independent variables. Conclude that $\dim C + \delta(C) \leq n+1$.
    (b) Give an example of a linear code with $\delta(C) = 2$ for which equality holds.
    (c) Examining the argument for (a) more closely, show that $2 \leq \dim C \leq n-2$ implies $\dim C + \delta(C) \leq n$.

67. Let C be a linear code with a basis consisting of the rows of $\begin{pmatrix} 1\,0\,0\,1\,1\,0 \\ 0\,1\,0\,1\,0\,1 \\ 0\,0\,1\,0\,1\,1 \end{pmatrix}$. Show that $\delta(C) = 3$. Educational note: Thus for $n = 6$ and $\delta(C) = 3$, we always have $\dim C \leq 3$, and equality is possible.

68. **(Hamming codes)** The Hamming code C_7 of order 7 is a certain linear code having $\dim C_7 = 4$ that will be seen to have $\delta(C_7) = 3$. The code words of a basis, with their commas removed, may be taken as

$$1110000,\ 1001100,\ 0101010,\ 1101001.$$

    The basis may be described as follows. Bits $1, 2, 4$ are used as checks. The remaining bits are used to form the standard basis of $\mathbb{F}^4$. What is put in bits $1, 2, 4$ is the binary representation of the position of the nonzero entry in

positions 3, 5, 6, 7. When all 16 members of C_7 are listed in the order dictated by the bits in positions 3, 5, 6, 7, the resulting list is

Decimal value in 3, 5, 6, 7	Code word	Decimal value in 3, 5, 6, 7	Code word
0	0000000	8	1110000
1	1101001	9	0011001
2	0101010	10	1011010
3	1000011	11	0110011
4	1001100	12	0111100
5	0100101	13	1010101
6	1100110	14	0010110
7	0001111	15	1111111

For the general members of C_7, not just the basis vectors, the check bits in positions 1, 2, 4 may be described as follows: the bit in position 1 is a parity bit for the positions among 3, 5, 6, 7 having a 1 in their binary expansions, the bit in position 2 is a parity bit for the positions among 3, 5, 6, 7 having a 2 in their binary expansions, and the bit in position 4 is a parity bit for the positions among 3, 5, 6, 7 having a 4 in their binary expansions. The Hamming code C_8 of order 8 is obtained from C_7 by adjoining a parity bit in position 8.

(a) Prove that $\delta(C_7) = 3$. (Educational note: Thus for $n = 7$ and $\delta(C) = 3$, we always have $\dim C \leq 4$, and equality is possible.)
(b) Prove that $\delta(C_8) = 4$.
(c) Describe how to form a generalization that replaces $n = 8$ by $n = 2^r$ with $r \geq 3$. The Hamming codes that are obtained will be called C_{2^r-1} and C_{2^r}.
(d) Prove that $\dim C_{2^r-1} = \dim C_{2^r} = 2^r - r - 1$, $\delta(C_{2^r-1}) = 3$, and $\delta(C_{2^r}) = 4$.

69. The matrix $H = \begin{pmatrix} 1\,0\,1\,0\,1\,0\,1 \\ 0\,1\,1\,0\,0\,1\,1 \\ 0\,0\,0\,1\,1\,1\,1 \end{pmatrix}$, when multiplied by any column vector c in the Hamming code C_7, performs the three parity checks done by bits 1, 2, 4 and described in the previous problem. Therefore such a c must have $Hc = 0$.
(a) Prove that the condition works in the reverse direction as well—that $Hc = 0$ only if c is in C_7.
(b) Deduce that if a received word r is not in C_7 and if r is assumed to match some word of C_7 except in the i^{th} position, then Hr matches the i^{th} column of H and this fact determines the integer i. (Educational note: Thus there is a simple procedure for testing whether a received word is a code word and for deciding, in the case that it is not a code word, what unique bit to change to convert it into a code word.)

70. Let $r \geq 4$. Prove for $2^{r-1} \leq n \leq 2^r - 1$ that any linear code C in $\mathbb{F}^n$ with $\delta(C) \geq 3$ has $\dim C \leq n - r$. Observe that equality holds for $C = C_{2^r-1}$.

71. The **weight enumerator polynomial** of a linear code C is the polynomial $W_C(X, Y)$ in $\mathbb{Z}[X, Y]$ given by $W_C(X, Y) = \sum_{k=0}^{n} N_k(C)X^{n-k}Y^k$, where $N_k(C)$ is the number of words of weight k in C.
    (a) Compute $W_C(X, Y)$ for the following linear codes C: the 0 code, the code $\mathbb{F}^n$, the repetition code, the parity code, the code in Problem 67, the Hamming code C_7, and the Hamming code C_8.
    (b) Why is the coefficient of X^n in $W_C(X, Y)$ necessarily equal to 1?
    (c) Show that $W_C(X, Y) = \sum_{c \in C} X^{n-\mathrm{wt}(c)}Y^{\mathrm{wt}(c)}$.

72. **(Cyclic redundancy codes)** Cyclic redundancy codes treat blocks of data as coefficients of polynomials in $\mathbb{F}[X]$. With the size n of data blocks fixed, one fixes a monic **generating polynomial** $G(X) = 1 + a_1X + \cdots + a_{g-1}X^{g-1} + X^g$ with a nonzero constant term and with degree g suitably less than n. Data to be transmitted are provided as members $(b_0, b_1, \ldots, b_{n-g-1})$ of $\mathbb{F}^{n-g}$ and are converted into polynomials $B(X) = b_0 + b_1X + \cdots + b_{n-g-1}X^{n-g-1}$. Then the n-tuple of coefficients of $G(X)B(X)$ is transmitted. To decode a polynomial $P(X)$ that is received, one writes $P(X) = G(X)Q(X) + R(X)$ via the division algorithm. If $R(X) = 0$, it is assumed that $P(X)$ is a code word. Otherwise $R(X)$ is definitely not a code word. Thus the code C amounts to the system of coefficients of all polynomials $G(X)B(X)$ with $B(X) = 0$ or $\deg B(X) \leq n - g - 1$. A basis of C is obtained by letting $B(X)$ run through the monomials $1, X, \ldots, X^{n-g-1}$, and therefore $\dim C = n-g$. Take $G(X) = 1+X+X^2+X^4$ and $n \geq 8$. Prove that $\delta(C) = 2$.

73. **(CRC-8)** The cyclic redundancy code C bearing the name CRC-8 has $G(X) = 1 + X + X^2 + X^8$. Prove that if $8 \leq n \leq 19$, then $\delta(C) = 4$. (Educational note: It will follow from the theory of finite fields in Chapter IX, together with the problems on coding theory at the end of that chapter, that $n = 255$ plays a special role for this code, and $\delta(C) = 4$ in that case.)

Problems 74–77 concern categories and functors. Problem 75 assumes knowledge of point-set topology.

74. Let $\mathcal{C}$ be the category of all sets, the morphisms being the functions between sets. Verify that the disjoint union of sets is a coproduct.

75. Let $\mathcal{C}$ be the category of all topological spaces, the morphisms being the continuous functions. Let S be a nonempty set, and let X_s be a topological space for each s in S.
    (a) Show that the Cartesian product of the spaces X_s, with the product topology, is a product of the X_s's.
    (b) Show that the disjoint union of the spaces X_s, topologized so that a set E is open if and only if its intersection with each X_s is open, is a coproduct of the X_s's.

76. Taking a cue from the example of a category in which products need not exist, exhibit a category in which coproducts need not exist.

77. Let $\mathcal{C}$ be a category having just one object, say X, and suppose that every member of $\mathrm{Morph}(X, X)$ is an isomorphism. Prove that $\mathrm{Morph}(X, X)$ is a group under the law of composition for the category. Can every group be realized in this way, up to isomorphism?

Problems 78–80 introduce a notion of **duality** in category theory and use it to derive Proposition 4.64 from Proposition 4.63. If $\mathcal{C}$ is a category, then the **opposite category** $\mathcal{C}^{\mathrm{opp}}$ is defined to have $\mathrm{Obj}(\mathcal{C}^{\mathrm{opp}}) = \mathrm{Obj}(\mathcal{C})$ and $\mathrm{Morph}_{\mathcal{C}^{\mathrm{opp}}}(A, B) = \mathrm{Morph}_{\mathcal{C}}(B, A)$. If $\circ$ denotes the law of composition in $\mathcal{C}$, then the law of composition $\circ^{\mathrm{opp}}$ in $\mathcal{C}^{\mathrm{opp}}$ is defined by $g \circ^{\mathrm{opp}} f = f \circ g$ for $f \in \mathrm{Morph}_{\mathcal{C}^{\mathrm{opp}}}(A, B)$ and $g \in \mathrm{Morph}_{\mathcal{C}^{\mathrm{opp}}}(B, C)$.

78. Verify that $\mathcal{C}^{\mathrm{opp}}$ is indeed a category, that $(\mathcal{C}^{\mathrm{opp}})^{\mathrm{opp}} = \mathcal{C}$, and that to pass from a diagram involving objects and morphisms in $\mathcal{C}$ to a corresponding diagram involving the same objects and morphisms considered as in $\mathcal{C}^{\mathrm{opp}}$, one leaves all the vertices and labels alone and reverses the directions of all the arrows. Verify also that the diagram of $\mathcal{C}$ commutes if and only if the diagram in $\mathcal{C}^{\mathrm{opp}}$ commutes.

79. Let $\mathcal{C}$ be the category of all sets, the morphisms in $\mathrm{Morph}_{\mathcal{C}}(A, B)$ being all functions from A to B. Show that the morphisms in $\mathrm{Morph}_{\mathcal{C}^{\mathrm{opp}}}(A, B)$ cannot necessarily all be regarded as functions from A to B.

80. Suppose that S is a nonempty set and that $\{X_s\}_{s\in S}$ is an object in $\mathcal{C}$.
    (a) Prove that if $(X, \{p_s\}_{s\in S})$ is a product of $\{X_s\}_{s\in S}$ in $\mathcal{C}$, then $(X, \{p_s\}_{s\in S})$ is a coproduct of $\{X_s\}_{s\in S}$ in $\mathcal{C}^{\mathrm{opp}}$, and that if $(X, \{p_s\}_{s\in S})$ is a coproduct of $\{X_s\}_{s\in S}$ in $\mathcal{C}$, then $(X, \{p_s\}_{s\in S})$ is a product of $\{X_s\}_{s\in S}$ in $\mathcal{C}^{\mathrm{opp}}$.
    (b) Show that Proposition 4.64 for $\mathcal{C}$ follows from the validity of Proposition 4.63 for $\mathcal{C}^{\mathrm{opp}}$.