Basic Algebra

Final Version, August, 2006
For Publication by Birkhäuser Boston
Along with a Companion Volume *Advanced Algebra*
In the Series

Cornerstones

Selected Pages from Chapter IV: pp. 116–134, 158–166, 188–198

Anthony W. Knapp

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 *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.

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 $(g_1, g_2)(g_1', g_2') = (g_1 g_1', g_2 g_2')$. 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 $(g_1, g_2) \mapsto g_1 g_2$ 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 $g_1 g_2$ 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 $g_1 g_2$ 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 $g_1 g_2$ 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 $\varphi(g_1, g_2) = g_1 g_2$. 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 $g_1 g_2 = g_1' g_2'$ 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 $\varphi((g_1, g_2)(g_1', g_2')) = \varphi(g_1 g_1', g_2 g_2') = g_1 g_1' g_2 g_2' = g_1 g_2 g_1' g_2 = \varphi(g_1, g_2)\varphi(g_1', g_2')$. 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 $\mathbb{R}^\times \cong C_2 \times \mathbb{R}^+$ 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 $(e^{i\theta}, r) \mapsto e^{i\theta} r$.

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

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

$$\begin{aligned} 11 &= 1, & 1\mathbf{i} &= \mathbf{i}, & 1\mathbf{j} &= \mathbf{j}, & 1\mathbf{k} &= \mathbf{k}, \\ \mathbf{i}1 &= \mathbf{i}, & \mathbf{ii} &= -1, & \mathbf{ij} &= \mathbf{k}, & \mathbf{ik} &= -\mathbf{j}, \\ \mathbf{j}1 &= \mathbf{j}, & \mathbf{ji} &= -\mathbf{k}, & \mathbf{jj} &= -1, & \mathbf{jk} &= \mathbf{i}, \\ \mathbf{k}1 &= \mathbf{k}, & \mathbf{ki} &= \mathbf{j}, & \mathbf{kj} &= -\mathbf{i}, & \mathbf{kk} &= -1, \end{aligned}$$

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 $\mathbf{ij} = \mathbf{k}$ while $\mathbf{ji} = -\mathbf{k}$, 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 $H = mZ$, 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.

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 $H = \{1,\ (1\ 3)\}$ 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 $\ker \overline{\varphi} = \{h_0 H \mid h_0 \in H_0\}$.

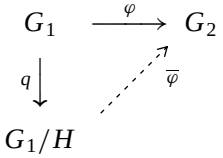

FIGURE 4.1. Factorization of homomorphisms of groups via the quotient of a group by a normal subgroup.

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 $h_2 h_2'$ and h_2^{-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 $(h_1h_2)^{-1} = (h_2^{-1}h_1^{-1}h_2)h_2^{-1}$, 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.

Pages 135–157 do not appear in this file.

$\mathbb{C}$ 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 $\mathcal{F}(X)$. Examples 5–9 of groups in Section 1 were 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 $\mathrm{Im}\, z > 0$, 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

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

of $\mathrm{GL}(2,\mathbb{C})$ 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.[12] 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 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 $\tau(g_1g_2) = \tau(g_1)\tau(g_2)$ for all g_1 and g_2. There are interesting examples both when V is finite-dimensional and when V is infinite-dimensional.[13]

[12] 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 $\mathbb{C}$.

[13] 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.

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 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 $(g_1, g_2) \mapsto g_1 g_2$ and the action of G on itself by right translations $(g_1, g_2) \mapsto g_2 g_1^{-1}$. More useful is the action of G on a quotient space G/H, where H is a subgroup. This action is given by $(g_1, g_2 H) \mapsto g_1 g_2 H$. 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. 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, 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 $gp = q$.

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 $\overline{\varphi}(g_1 H) = \overline{\varphi}(g_2 H)$, then $g_1 p = \varphi(g_1 p) = \varphi(g_2 p) = g_2 p$, and hence $g_2^{-1} g_1 p = p$, $g_2^{-1} g_1$ is in H, g_1 is in $g_2 H$, and $g_1 H = g_2 H$. 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 $|G| = p^k$ for all $k \le n$ and such that $G_k \subseteq G_{k+1}$ 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 4.39 shows that $Z_G \neq \{1\}$. Any element $\neq 1$ 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 $\gamma = (a_1\ \ a_2\ \ \cdots\ \ a_n)$ appearing in σ. The corresponding cycle $\gamma' = \tau\gamma\tau^{-1}$ is asserted to be $\gamma' = (\tau(a_1)\ \ \tau(a_2)\ \ \cdots\ \ \tau(a_n))$. Application of τ^{-1} to $\tau(a_j)$ yields a_j, application of σ to this yields a_{j+1} if $j < n$ and a_1 if $j = n$, 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 $\{a_1, \ldots, a_n\}$, $\tau\gamma\tau^{-1}(\tau(b)) = \tau(b)$ 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 $\operatorname{Aut} H$. 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 $\tau : G \to \operatorname{Aut} X$.

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 $\operatorname{Aut} H$. Out of these data—the two groups G and H and a homomorphism of G into $\operatorname{Aut} H$—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 $h \to \tau_g(h)$ 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

Pages 167–187 do not appear in this file.

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 $\mathrm{Obj}(\mathcal{C})$,
- for any two objects A and B in the category, a set $\mathrm{Morph}(A, B)$ 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 $f \in \text{Morph}(A, B)$, $g \in \text{Morph}(B, C)$, and $h \in \text{Morph}(C, D)$,

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

A **subcategory** $\mathcal{S}$ of a category $\mathcal{C}$ by definition is a category with $\text{Obj}(\mathcal{S}) \subseteq \text{Obj}(\mathcal{C})$ and $\text{Morph}_{\mathcal{S}}(A, B) \subseteq \text{Morph}_{\mathcal{C}}(A, B)$ whenever A and B are in $\text{Obj}(\mathcal{S})$, 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 $\text{Morph}(A, B)$ 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.[15]

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.[16] 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

[15] 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*.

[16] 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.

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 Obj($\mathcal{C}$) an object $F(A)$ in 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 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**.

(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 $F(V) = \text{Hom}_{\mathbb{F}}(U, V)$. The set of morphisms $\text{Morph}_{\mathcal{C}}(V_1, V_2)$ is $\text{Hom}_{\mathbb{F}}(V_1, V_2)$. If f is in $\text{Morph}_{\mathcal{C}}(V_1, V_2)$, then $F(f)$ is to be in $\text{Morph}_{\mathcal{C}}\big(\text{Hom}_{\mathbb{F}}(U, V_1), \text{Hom}_{\mathbb{F}}(U, V_2)\big)$, and the definition is that $F(f)(L) = f \circ L$ for $L \in \text{Hom}_{\mathbb{F}}(U, V_1)$. Then F is a covariant functor: to check that $F(gf) = F(g)F(f)$ when g is in $\text{Morph}_{\mathcal{C}}(V_2, V_3)$, 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 $\text{Morph}_{\mathcal{C}}(H_1, H_2)$ is the set of group homomorphisms from H_1 into H_2. If f is in $\text{Morph}_{\mathcal{C}}(H_1, H_2)$, 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 $L \in \mathrm{Hom}_{\mathbb{F}}(V_1, W)$. 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 $F(gf)(L) = L \circ gf = Lg \circ f = F(f)(Lg) = F(f)F(g)$.

(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 $F(f)(\varphi) = f \circ \varphi$. 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*.

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 *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.[17] 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.

[17] 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.

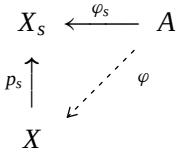

FIGURE 4.11. Universal mapping property of a product in a category.

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 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 $\Phi \in \text{Morph}(X', X)$ 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 $\Phi' \in \text{Morph}(X, X')$ 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 $\text{Morph}(X, X)$ 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 $\Psi \in \text{Morph}(X', X)$ 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[18] in $\text{Obj}(\mathcal{C}^S)$. 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 $\text{Morph}_{\mathcal{C}}(X_s, Y_s)$ 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 $\text{Morph}_{\mathcal{C}}\left(\prod_{s\in S} X_s, Y_{s_0}\right)$ for each s_0, and the universal mapping property gives us f in $\text{Morph}_{\mathcal{C}}\left(\prod_{s\in S} X_s, \prod_{s\in S} Y_s\right)$ 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 $\text{Obj}(\mathcal{C})$ and each i_s is in $\text{Morph}_{\mathcal{C}}(X_s, X)$ with the following **universal mapping property**: whenever A in $\text{Obj}(\mathcal{C})$ is given and a morphism $\varphi_s \in \text{Morph}_{\mathcal{C}}(X_s, A)$ is given

[18]Since $\text{Obj}(\mathcal{C}^S)$ 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 $\text{Obj}(\mathcal{C}^S)$. Instead, one makes the choice only for all objects in some set of objects large enough for a particular application.

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.

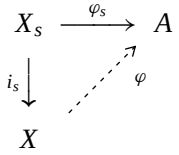

FIGURE 4.12. Universal mapping property of a coproduct in a category.

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.

(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.