James B. Carrell

Groups, Matrices, and Vector Spaces

A Group Theoretic Approach to Linear Algebra

Springer

Groups, Matrices, and Vector Spaces

James B. Carrell

Groups, Matrices, and Vector Spaces

A Group Theoretic Approach to Linear Algebra

Springer

James B. Carrell
Department of Mathematics
University of British Columbia
Vancouver, BC
Canada

ISBN 978-0-387-79427-3 ISBN 978-0-387-79428-0 (eBook)
DOI 10.1007/978-0-387-79428-0

Library of Congress Control Number: 2017943222

Mathematics Subject Classification (2010): 15-01, 20-01

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer Science+Business Media, LLC
The registered company address is: 233 Spring Street, New York, NY 10013, U.S.A.

Foreword

This book is an introduction to group theory and linear algebra from a geometric viewpoint. It is intended for motivated students who want a solid foundation in both subjects and are curious about the geometric aspects of group theory that cannot be appreciated without linear algebra. Linear algebra and group theory are connected in very pretty ways, and so it seems that presenting them together is an appropriate goal. Group theory, founded by Galois to study the symmetries of roots of polynomial equations, was extended by many nineteenth-century mathematicians who were also leading figures in the development of linear algebra such as Cauchy, Cayley, Schur, and Lagrange. It is amazing that such a simple concept has touched so many rich areas of current research: algebraic geometry, number theory, invariant theory, representation theory, combinatorics, and cryptography, to name some. Matrix groups, which are part matrix theory, part linear algebra, and part group theory, have turned out to be richest source of finite simple groups and the basis for the theory of linear algebraic groups and for representation theory, two very active areas of current research that have linear algebra as their basis. The orthogonal and unitary groups are matrix groups that are fundamental tools for particle physicists and for quantum mechanics. And to bring linear algebra in, we should note that every student of physics also needs to know about eigentheory and Jordan canonical form.

For the curious reader, let me give a brief description of what is covered. After a brief preliminary chapter on combinatorics, mappings, binary operations, and relations, the first chapter covers the basics of group theory (cyclic groups, permutation groups, Lagrange's theorem, cosets, normal subgroups, homomorphisms, and quotient groups) and gives an introduction to the notion of a field. We define the basic fields $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$, and discuss the geometry of the complex plane. We also state the fundamental theorem of algebra and define algebraically closed fields. We then construct the prime fields $\mathbb{F}_p$ for all primes p and define Galois fields. It is especially nice to have finite fields, since computations involving matrices over $\mathbb{F}_2$ are delightfully easy. The lovely subject of linear coding theory, which requires $\mathbb{F}_2$, will be treated in due course. Finally, we define polynomial rings and prove the multiple root test.

We next turn to matrix theory, studying matrices over an arbitrary field. The standard results on Gaussian elimination are proven, and *LPDU* factorization is studied. We show that the reduced row echelon form of a matrix is unique, thereby enabling us to give a rigorous treatment of the rank of a matrix. (The uniqueness of the reduced row echelon form is a result that most linear algebra books curiously ignore.) After treating matrix inverses, we define matrix groups, and give examples such as the general linear group, the orthogonal group, and the $n \times n$ permutation matrices, which we show are isomorphic to the symmetric group $S(n)$. We conclude the chapter with the Birkhoff decomposition of the general linear group.

The next chapter treats the determinant. After defining the signature of a permutation and showing that it is a homomorphism, we define $\det(A)$ via the alternating sum over the symmetric group known as Leibniz's formula. The proofs of the product formula (that is, that det is a homomorphism) and the other basic results about the determinant are surprisingly simple consequences of the definition. The determinant is an important and rich topic, so we treat the standard applications such as the Laplace expansion and Cramer's rule, and we introduce the important special linear group. Finally, we consider a recent application of the determinant known as Dodgson condensation.

In the next chapter, finite-dimensional vector spaces, bases, and dimension are covered in succession, followed by more advanced topics such as direct sums, quotient spaces, and the Grassmann intersection formula. Inner product spaces over $\mathbb{R}$ and $\mathbb{C}$ are covered, and in the appendix, we give an introduction to linear coding theory and error-correcting codes, ending with perfect codes and the hat game, in which a player must guess the color of her hat based on the colors of the hats her teammates are wearing.

The next chapter moves us on to linear mappings. The basic properties such as the rank–nullity theorem are covered. We treat orthogonal linear mappings and the orthogonal and unitary groups, and we classify the finite subgroups of $SO(2, \mathbb{R})$. Using the $O(2, \mathbb{R})$ dichotomy, we also obtain Leonardo da Vinci's classification that all finite subgroups of $O(2, \mathbb{R})$ are cyclic or dihedral. This chapter also covers dual spaces, coordinates, and the change of basis formulas for matrices of linear mappings.

We then take up eigentheory: eigenvalues and eigenvectors, the characteristic polynomial of a linear mapping, and its matrix and diagonalization. We show how the Fibonacci sequence is obtained from the eigenvalue analysis of a certain dynamical system. Next, we consider eigenspace decompositions and prove that a linear mapping is semisimple—equivalently, that its matrix is diagonalizable—if and only if its minimal polynomial has simple roots. We give a geometric proof of the principal axis theorem for both Hermitian and real symmetric matrices and for self-adjoint linear mappings. Our proof of the Cayley–Hamilton theorem uses a simple inductive argument noticed by the author and Jochen Kuttler. Finally, returning to the geometry of $\mathbb{R}^3$, we show that $SO(3, \mathbb{R})$ is the group of rotations of $\mathbb{R}^3$ and conclude with the computation of the rotation groups of several of the Platonic solids.

Following eigentheory, we cover the normal matrix theorem and quadratic forms, including diagonalization and Sylvester's law of inertia. Then we classify linear mappings, proving the Jordan–Chevalley decomposition theorem and the existence of the Jordan canonical form for matrices over an algebraically closed field. The final two chapters concentrate on group theory. The penultimate chapter establishes the basic theorems of abstract group theory up to the Jordan-Schreier theorem and gives a treatment of finite group theory (e.g., Cauchy's theorem and the Sylow theorems) using the very efficient approach via group actions and the orbit-stabilizer theorem. We also classify the finite subgroups of $SO(3, \mathbb{R})$. The appendix to this chapter contains a description of how Polish mathematicians reconstructed the German Enigma machine before the Second World War via group theory. This was a milestone in abstract algebra and to this day is surely the most significant application of group theory ever made.

The final chapter is an informal introduction to the theory of linear algebraic groups. We give the basic definitions and discuss the basic concepts: maximal tori, the Weyl group, Borel subgroups, and the Bruhat decomposition. While these concepts were already introduced for the general linear group, the general notions came into use relatively recently. We also consider reductive groups and invariant theory, which are two topics of contemporary research involving both linear algebra and group theory.

Acknowledgements: The author is greatly indebted to the editors at Springer, Ann Kostant (now retired) and Elizabeth Loew, who, patiently, gave me the opportunity to publish this text. I would also like to thank Ann for suggesting the subtitle. I would like to thank several colleagues who made contributions and gave me valuable suggestions. They include Kai Behrend, Patrick Brosnan, Kiumars Kaveh, Hanspeter Kraft, Jochen Kuttler, David Lieberman, Vladimir Popov, Edward Richmond, and Zinovy Reichstein. I would also like to thank Cameron Howie for his very careful reading of the manuscript and many comments.

May 2017 Jim Carrell

Contents

Chapter 1
Preliminaries

In this brief chapter, we will introduce (or in many cases recall) some elementary concepts that will be used throughout the text. It will be convenient to state them at the beginning so that they will all be in the same place.

1.1 Sets and Mappings

A set is a collection of objects called the elements of X. If X is a set, the notation $x \in X$ will mean that x is an element of X. Sets are frequently defined in terms of a property. For example, if $\mathbb{R}$ denotes the set of all real numbers, then the set of all positive real numbers is denoted by the expression $\{r \in \mathbb{R} \mid r > 0\}$. One can also define a set by listing its elements, an example being the set consisting of the integers 1, 2, and 3, which could be denoted by either

$$\{1, 2, 3\},$$

or, more clumsily,

$$\{r \in \mathbb{R} \mid r \text{ is an integer and } 1 \leq r \leq 3\}.$$

A set with exactly one element is called a *singleton*.

The *union* of two sets X and Y is the set

$$X \cup Y = \{a \mid a \in X \text{ or } a \in Y\}$$

whose elements are the elements of X together with the elements of Y. The *intersection* of X and Y is the set

J.B. Carrell, *Groups, Matrices, and Vector Spaces*,
DOI 10.1007/978-0-387-79428-0_1

$$X \cap Y = \{a \mid a \in X \text{ and } a \in Y\}.$$

The *difference of X and Y* is the set

$$X \backslash Y = \{x \in X \mid x \notin Y\}.$$

Note that Y does not need to be a subset of X for one to speak about the difference $X \backslash Y$. Notice that set union and difference are analogous to addition and subtraction. Intersection is somewhat analogous to multiplication. For example, for any three sets X, Y, Z, one has

$$X \cap (Y \cup Z) = (X \cap Y) \cup (X \cap Z),$$

which is analogous to the distributive law $a(b + c) = ab + ac$ for real numbers a, b, c. The *product* of X and Y is the set

$$X \times Y = \{(x, y) \mid x \in X \text{ and } y \in Y\}.$$

We call (x, y) an *ordered pair*. That is, when $X = Y$, then $(x, y) \neq (y, x)$ unless $x = y$. The product $X \times X$ can be denoted by X^2. For example, $\mathbb{R} \times \mathbb{R}$ is the Cartesian plane, usually denoted by $\mathbb{R}^2$.

A *mapping* from X to Y is a rule F that assigns to every element $x \in X$ a unique element $F(x) \in Y$. The notation $F : X \to Y$ will be used to denote a mapping from X to Y; X is called the *domain* of F, and Y is called its *target*. The *image of F* is

$$F(X) = \{y \in Y \mid y = F(x) \text{ for some } x \in X\}.$$

For example, if $F : \mathbb{R} \to \mathbb{R}$ is the mapping $F(r) = r^2$, then $F(\mathbb{R})$ is the set of all nonnegative reals. The *composition* of two mappings $F : X \to Y$ and $G : W \to X$ is the mapping $F \circ G : W \to Y$ defined by $G \circ F(w) = F(G(w))$. The composition $F \circ G$ is defined whenever the domain of F is contained in the image of G.

1.1.1 Binary operations

The following notion will be used in the definition of a field.

Definition 1.1. A *binary operation* on a set X is a function

$$F : X \times X \to X.$$

Example 1.1. Let $\mathbb{Z}$ denote the set of integers. There are two binary operations on $\mathbb{Z}$ called addition and multiplication. They are defined, respectively, by $F_+(m,n) = m+n$ and $F_\cdot(m,n) = mn$. Note that division is not a binary operation on $\mathbb{Z}$.

We also need the notion of a subset being closed with respect to a binary operation.

Definition 1.2. Let F be a binary operation on a set A. A subset B of A such that $F(x,y) \in B$ whenever $x, y \in B$ is said to be *closed under the binary operation* F.

For example, let $A = \mathbb{Z}$ and let B be the set of all nonnegative integers. Then B is closed under both addition and multiplication. The odd integers are closed under multiplication, but not closed under addition, since, for instance, $1 + 1 = 2$.

If F is a mapping from X to Y and $y \in Y$, then the *inverse image of* y is

$$F^{-1}(y) = \{x \in X \mid F(x) = y\}.$$

Of course, $F^{-1}(y)$ may not have any elements; that is, it may be the empty set. For example, if $F : \mathbb{R} \to \mathbb{R}$ is the mapping $F(r) = r^2$, then $F^{-1}(-1)$ is empty. Notice that if $y \neq y'$, then $F^{-1}(y) \cap F^{-1}(y')$ is the empty set.

The notion of the inverse image of an element is useful in defining some further properties of mappings. For example, a mapping $F : X \to Y$ is *one to one*, or *injective*, if and only if $F^{-1}(y)$ is either empty or a single element of X for all $y \in Y$. In other words, F is injective if and only if $F(x) = F(x')$ implies $x = x'$. Similarly, F is *onto*, or equivalently, *surjective*, if $F^{-1}(y)$ is nonempty for all $y \in Y$. Alternatively, F is surjective if and only if $F(X) = Y$. A mapping $F : X \to Y$ that is both injective and surjective is said to be *bijective*. A mapping that is injective, surjective, or bijective will be called an injection, surjection, or bijection respectively. A bijective map $F : X \to Y$ has an *inverse mapping* F^{-1}, which is defined by putting $F^{-1}(y) = x$ if and only if $F(x) = y$. It follows directly from the definition that $F^{-1} \circ F(x) = x$ and $F \circ F^{-1}(y) = y$ for all $x \in X$ and $y \in Y$.

The following proposition gives criteria for injectivity and surjectivity.

Proposition 1.1. *Suppose $F : X \to Y$ is a mapping and suppose there exists a mapping $G : Y \to X$ such that $G \circ F(x) = x$ for all $x \in X$. Then F is injective and G is surjective. Moreover, F is bijective if and only if $G \circ F$ and $F \circ G$ are the identity mappings on X and Y respectively.*

1.1.2 Equivalence relations and equivalence classes

We now want to define an equivalence relation on a set. This will give us a way of partitioning a set into disjoint subsets called equivalence classes. The notion of equivalence is a generalization of the notion of equality. First, we need to recall what a relation on a set is.

Definition 1.3. Let S be a nonempty set. A subset E of $S \times S$ is called a *relation on* S. If E is a relation on S, and a and b are elements of S, we will say that a and b are *related* by E and write aEb if and only if $(a, b) \in E$. A relation E on S is called an *equivalence relation* when the following three conditions hold for all $a, b, c \in S$:

(i) (E is reflexive) aEa,

(ii) (E is symmetric) if aEb, then bEa, and

(iii) (E is transitive) if aEb and bEc, then aEc.

If E is an equivalence relation on S and $a \in S$, then the *equivalence class of* a is defined to be the set of all elements $b \in S$ such that bEa. An element of an equivalence class is called a *representative* of the class.

Before proving the main property of an equivalence relation, we will consider two basic examples.

Example 1.2. The model on which the notion of an equivalence relation is built is equality. On an arbitrary nonempty set S, let us say that sEt if and only if $s = t$. The equivalence classes consist of the singletons $\{s\}$, as s varies over S. ⊠

The second example is an equivalence relation on the integers frequently used in number theory.

Example 1.3 (The Integers Modulo m)**.** Let m denote an integer, and consider the pairs of integers (r, s) such that $r - s$ is divisible by m. That is, $r - s = km$ for some integer k. This defines a relation $\mathcal{C}_m$ on $\mathbb{Z} \times \mathbb{Z}$ called *congruence modulo* m. When $(r, s) \in \mathcal{C}_m$, one usually writes $r \equiv s \bmod(m)$. We claim that congruence modulo m is an equivalence relation. For example, $r\mathcal{C}_m r$, and if $r\mathcal{C}_m s$, then certainly $s\mathcal{C}_m r$. For transitivity, assume $r\mathcal{C}_m s$ and $s\mathcal{C}_m t$. Then $r - s = 2i$ and $s - t = 2j$ for some integers i and j. Hence $r - t = (r - s) + (s - t) = 2i + 2j = 2(i + j)$, so $r\mathcal{C}_m t$. Hence $\mathcal{C}_m$ is an equivalence relation on $\mathbb{Z}$. ⊠

Here is the basic result on equivalence relations.

Proposition 1.2. *Let E be an equivalence relation on a set S. Then every element $a \in S$ is in its own equivalence class, and two equivalence classes are either disjoint or equal. Therefore, S is the disjoint union of the equivalence classes of E.*

Proof. Every element is equivalent to itself, so S is the union of its equivalence classes. We have to show that two equivalence classes are either equal or disjoint. Suppose C_1 and C_2 are equivalence classes, and let $c \in C_1 \cap C_2$. If $a \in C_1$, then aEc. If C_2 is the equivalence class of b, then cEb, so aEb. Hence, $a \in C_2$, so $C_1 \subset C_2$. Similarly, $C_2 \subset C_1$, so $C_1 = C_2$. □

Definition 1.4. The set of equivalence classes of an equivalence relation E on S is called the *quotient of S by E*.

1.2 Some Elementary Combinatorics

Combinatorics deals with the properties of various kinds of finite sets. A nonempty set X is said to be *finite* if there exist an integer $n > 0$ and a bijection $\sigma : \{1, 2, \ldots, n\} \to X$. If X is finite, the number of elements of X is denoted by $|X|$. If X is the empty set, we define $|X| = 0$. The following result is an example of an elementary combinatorial result.

Proposition 1.3. *Let X be finite set that is the union of mutually disjoint (nonempty) subsets $X_1, \ldots, X_k$. Then*

$$|X| = \sum_{i=1}^{k} |X_i|.$$

In particular, if Y is a proper subset of X, then $|Y| < |X|$.

Proof. We will consider the case $k = 2$ first and then finish the proof by applying the principle of mathematical induction, which is introduced in the next section. Suppose $X = X_1 \cup X_2$, where $X_1 \cap X_2$ is empty and both X_1 and X_2 are nonempty. Let $|X_1| = j$ and $|X_2| = k$. By definition, there exist bijections $\sigma_1 : \{1, \ldots, i\} \to X_1$ and $\sigma_2 : \{1, \ldots, j\} \to X_2$. Define $\sigma : \{1, \ldots, i + j\} \to X$ by $\sigma(r) = \sigma_1(r)$ if $1 \le r \le i$ and $\sigma(r) = \sigma_2(r - i)$ if $i + 1 \le r \le i + j$. By construction, σ is a bijection from $\{1, \ldots, i + j\}$ to $X = X_1 \cup X_2$. Therefore, $|X| = i + j = |X_1| + |X_2|$. □

This identity applies, for example, to the equivalence classes of an equivalence relation on a finite set. Another application (left to the reader) is contained in the following proposition.

Proposition 1.4. *If X and Y are finite sets, then*

$$|X \cup Y| = |X| + |Y| - |X \cap Y|.$$

Here is another consequence.

Proposition 1.5. *Let X and Y be finite and $F : X \to Y$. Then*

$$|X| = \sum_{y \in Y} |F^{-1}(y)|. \tag{1.1}$$

In particular, if F is surjective, then $|X| \ge |Y|$.

Proof. Let $y \in Y$. By definition, $F^{-1}(y)$ is a nonempty subset of X if and only if $y \in F(X)$. Moreover, $F^{-1}(y) \cap F^{-1}(y')$ is empty if $y \ne y'$ for any $y' \in Y$. Thus, X can be written as the disjoint union of nonempty subsets

$$X = \bigcup_{y \in F(X)} F^{-1}(y).$$

Since $|F^{-1}(y)| = 0$ if $y \notin F(X)$, the identity (1.1) follows from Proposition 1.3. □

Proposition 1.6 (The Pigeonhole Principle)**.** *Let X and Y be finite sets with $|X| = |Y|$, and suppose $F : X \longrightarrow Y$ is a mapping. If F is either injective or surjective, then F is a bijection.*

Proof. If F is injective, then $|F^{-1}(y)|$ is either 0 or 1 for all $y \in Y$. But if $|F^{-1}(y)| = 0$ for some $y \in Y$, then

$$|X| = \sum_{y \in Y} |F^{-1}(y)| < |Y|,$$

which is impossible, since $|X| = |Y|$. Thus, F is surjective. On the other hand, if F is surjective, then $|F^{-1}(y)| \geq 1$ for all y. Thus,

$$|Y| \leq \sum_{y \in Y} |F^{-1}(y)| = |X|.$$

Since $|X| = |Y|$, $|F^{-1}(y)| = 1$ for all y, so F is a bijection. □

1.2.1 Mathematical induction

Mathematical induction is a method of proof that allows one to prove propositions that state that some property holds for the set of all positive integers. Here is an elementary example.

Proposition 1.7 (The Principle of Mathematical Induction)**.** *A proposition $P(n)$ defined for each positive integer n holds for all positive integers provided:*

(i) $P(n)$ holds for $n = 1$, and

(ii) $P(n+1)$ holds whenever $P(n)$ holds.

The proof is an application of the fact that every nonempty set of positive integers has a least element. □

Let us now finish the proof of Proposition 1.3. Let $P(k)$ be the conclusion of the proposition when X is any finite set that is the union of mutually disjoint (nonempty) subsets $X_1, \ldots, X_k$. The statement $P(1)$ is true, since $X = X_1$. Now assume that $P(i)$ holds for $i < k$, where $k > 1$. Let $Y_1 = X_1 \cup \cdots X_{k-1}$ and $Y_2 = X_k$. Now $X = Y_1 \cup Y_2$, and since Y_1 and Y_2 are disjoint, $|X| =$

$|Y_1| + |Y_2|$, as we already showed. Now apply the principle of mathematical induction: since $P(k-1)$ holds, $|Y_1| = |X_1| + \cdots |X_{k-1}|$. Thus,

$$|X| = |Y_1| + |Y_2| = |X_1| + \cdots |X_{k-1}| + |X_k|,$$

which is exactly the assertion that $P(k)$ holds. Hence Proposition 1.3 is proved for all k. □

Here is a less pedestrian application.

Proposition 1.8. *For every positive integer n,*

$$1 + 2 + \cdots + n = \frac{n(n+1)}{2}. \tag{1.2}$$

Proof. Equality certainly holds if $n = 1$. Suppose (1.2) holds for an integer $k > 0$. We have to show that it holds for $k+1$. Applying (1.2) for k, we see that

$$1 + 2 + \cdots + k + (k+1) = \frac{k(k+1)}{2} + (k+1).$$

But

$$\frac{k(k+1)}{2} + (k+1) = \frac{k(k+1) + 2(k+1)}{2} = \frac{(k+2)(k+1)}{2},$$

so indeed (1.2) holds for $(k+1)$. Hence, by the principle of mathematical induction, (1.2) holds for all positive integers n.

Induction proofs can often be avoided. For example, one can also see the identity (1.2) by observing that the sum of the integers in the array

$$\begin{array}{ccccc} 1 & 2 & \cdots & (n-1) & n \\ n & (n-1) & \cdots & 2 & 1 \end{array}$$

is $n(n+1)$, since there are n columns, and each column sum is $n+1$.

1.2.2 The Binomial Theorem

The binomial theorem is a formula for expanding $(a+b)^n$ for any positive integer n, where a and b are variables that commute. The formula uses the binomial coefficients. First we note that if n is a positive integer, then by definition, $n! = 1 \cdot 2 \cdots n$: we also define $0! = 1$. Then the binomial coefficients are the integers

$$\binom{n}{i} = \frac{n!}{i!\ (n-i)!}, \tag{1.3}$$

where $0 \leq i \leq n$.

One can show that the binomial coefficient (1.3) is exactly the number of subsets of $\{1, 2, \ldots, n\}$ with exactly i elements. The binomial theorem states that

$$(a+b)^n = \sum_{i=0}^{n} \binom{n}{i} a^{n-i} b^i. \tag{1.4}$$

A typical application of the binomial theorem is the formula

$$2^n = \sum_{i=0}^{n} \binom{n}{i}.$$

This shows that a set with n elements has exactly 2^n subsets.

The binomial theorem is typical of the kind of result that is most easily proven by induction. The multinomial theorem is a generalization of the binomial theorem that gives a formula for expanding quantities such as $(a + b + c)^3$. Let n be a positive integer and suppose $n_1, n_2, \ldots, n_k$ are nonnegative integers such that $n_1 + \cdots + n_k = n$. The associated multinomial coefficient is defined as

$$\binom{n}{n_1, n_2, \cdots, n_k} = \frac{n!}{n_1! n_2! \cdots n_k!}.$$

This multinomial coefficient is the number of ways of partitioning a set with n objects into k subsets, the first with n_1 elements, the second with n_2 elements, and so forth. The multinomial theorem goes as follows.

Theorem 1.9 (Multinomial theorem)**.** *Let $a_1, \ldots, a_k$ be commuting variables. Then*

$$(a_1 + a_2 + \cdots + a_k)^n = \sum_{n_1 + \cdots + n_k = n} \binom{n}{n_1, n_2, \cdots, n_k} a_1^{n_1} a_2^{n_2} \cdots a_k^{n_k}. \tag{1.5}$$

Chapter 2
Groups and Fields: The Two Fundamental Notions of Algebra

Algebra is the mathematical discipline that arose from the problem of solving equations. If one starts with the integers $\mathbb{Z}$, one knows that every equation $a + x = b$, where a and b are integers, has a unique solution. However, the equation $ax = b$ does not necessarily have a solution in $\mathbb{Z}$, or it might have infinitely many solutions (take $a = b = 0$). So let us enlarge $\mathbb{Z}$ to the rational numbers $\mathbb{Q}$, consisting of all fractions c/d, where $d \neq 0$. Then both equations have a unique solution in $\mathbb{Q}$, provided that $a \neq 0$ for the equation $ax = b$. So $\mathbb{Q}$ is a field. If, for example, one takes the solutions of an equation such as $x^2 - 5 = 0$ and forms the set of all numbers of the form $a + b\sqrt{5}$, where a and b are rational, we get a larger field, denoted by $\mathbb{Q}(\sqrt{5})$, called an algebraic number field. In the study of fields obtained by adjoining the roots of polynomial equations, a new notion arose, namely, the symmetries of the field that permute the roots of the equation. Évariste Galois (1811–1832) coined the term *group* for these symmetries, and now this group is called the Galois group of the field. While still a teenager, Galois showed that the roots of an equation are expressible by radicals if and only if the group of the equation has a property now called solvability. This stunning result solved the 350-year-old question whether the roots of every polynomial equation are expressible by radicals.

2.1 Groups and homomorphisms

We now justly celebrate the *Galois group* of a polynomial, and indeed, the Galois group is still an active participant in the fascinating theory of elliptic curves. It even played an important role in the solution of Fermat's last theorem. However, groups themselves have turned out to be central in all sorts of mathematical disciplines, particularly in geometry, where they allow us to classify the symmetries of a particular geometry. And they have also

J.B. Carrell, *Groups, Matrices, and Vector Spaces*,
DOI 10.1007/978-0-387-79428-0_2

become a staple in other disciplines such as chemistry (crystallography) and physics (quantum mechanics).

In this section we will define the basic concepts of group theory starting with the definition of a group itself and the most basic related concepts such as cyclic groups, the symmetric group. and group homomorphisms. We will also prove some of the beginning results in group theory such as Lagrange's theorem and Cayley's theorem.

2.1.1 The Definition of a Group

The notion of a group involves a set with a binary operation that satisfies three natural properties. Before stating the definition, let us mention some basic but very different examples to keep in mind. The first is the integers under the operation of addition. The second is the set of all bijections of a set, and the third is the set of all complex numbers ζ such that $\zeta^n = 1$. We now state the definition.

Definition 2.1. A *group* is a set G with a binary operation written $(x, y) \to xy$ such that

(i) $(xy)z = x(yz)$ for all $x, y, z \in G$;

(ii) G contains an *identity element* 1 such that $1x = x1 = x$ for all $x \in G$, and

(iii) if $x \in G$, then there exists $y \in G$ such that $xy = 1$. In this case, we say that every element of G has a *right inverse*.

Property (i) is called the *associative law*. In other words, the group operation is associative. In group theory, it is customary to use the letter e to denote an identity. But it is more convenient for us to use 1. Note that property (iii) involves being able to solve an equation that is a special case of the equation $ax = b$ considered above. There are several additional properties that one can impose to define special classes of groups. For example, we might require that the group operation be independent of the order in which we take the group elements. More precisely, we make the following definition.

Definition 2.2. A group G is said to be *commutative* or *abelian* if and only if for all $x, y \in G$, we have $xy = yx$. A group that is not abelian is said to be nonabelian.

Example 2.1 (The Integers)**.** The integers $\mathbb{Z}$ form a group under addition. The fact that addition is associative is well known. Zero is an additive identity. In fact, it is the only additive identity. An additive inverse of $m \in \mathbb{Z}$ is its negative $-m$: $m + (-m) = 0$. Moreover, $\mathbb{Z}$ is abelian: $m + n = n + m$ for all $m, n \in \mathbb{Z}$. ⊠

The group $G = \{1, -1\}$ under multiplication is an even simpler example of an abelian group. Before we consider some examples of groups that are nonabelian, we will introduce a much more interesting class of groups, namely the finite groups.

Definition 2.3. A group G is said to be finite if the number $|G|$ of elements in the set G is finite. We will call $|G|$ the *order* of G.

The order of $G = \{1, -1\}$ is two, while $\mathbb{Z}$ is an infinite group. Of course, it is not clear yet why we say that the abelian groups are not as interesting as the finite groups. But this will become evident later.

2.1.2 Some basic properties of groups

Before going on to more examples of groups, we would like to prove a proposition that gives some basic consequences of the definition of a group. In particular, we will show that there is only one identity element 1, and we will also show that every element x in a group has a unique two-sided inverse x^{-1}. The reader may find the proofs amusing. Before we state this proposition, the reader may want to recall that we already noticed these facts in $\mathbb{Z}$: there is only one additive identity, 0, and likewise only one right inverse, $-m$, for each m. Moreover, $-m$ is also a left inverse of m. These properties are usually stated as part of the definition of a group, but for reasons we cannot explain now, we have chosen to use a minimal set of group axioms.

Proposition 2.1. *In every group, there is exactly one identity element,* 1. *Furthermore, if y is a right inverse of x, then $xy = yx = 1$. Hence, a right inverse is also a left inverse. Therefore, each $x \in G$ has a two-sided inverse y, which is characterized by the property that either $xy = 1$ or $yx = 1$. Moreover, each two sided inverse is unique.*

Proof. To prove the uniqueness of 1, suppose the 1 and $1'$ are both identity elements. Then $1 = 1 \cdot 1' = 1'$. Thus the identity is unique. Now let x have a right inverse y and let w be a right inverse of y. Then

$$w = 1w = (xy)w = x(yw) = x1 = x.$$

Since $w = x$, it follows that if $xy = 1$, then $yx = 1$. Thus, every right inverse is a two-sided inverse. We will leave the assertion that inverses are unique as an exercise. □

From now on, we will refer to the unique left or right inverse of x as *the inverse* of x. The notation for the inverse of x is x^{-1}. The next result is the formula for the inverse of a product.

Proposition 2.2. *For all $x, y \in G$, we have $(xy)^{-1} = y^{-1}x^{-1}$.*

Proof. Let $w = y^{-1}x^{-1}$. Then it suffices to show that $w(xy) = 1$. But

$$w(xy) = (wx)y = ((y^{-1}x^{-1})x)y = (y^{-1}(x^{-1}x))y = (y^{-1}1)y = y^{-1}y = 1.$$

□

If $x_1, x_2, \ldots, x_n$ are arbitrary elements of a group G, then the expression $x_1x_2\cdots x_n$ will stand for $x_1(x_2\cdots x_n)$, where $x_2\cdots x_n = x_2(x_3\cdots x_n)$ and so on. This gives an inductive definition of the product of an arbitrary finite number of elements of G. Moreover, by associativity, pairs $(\cdots)$ of parentheses can be inserted or removed in the expression $x_1x_2\cdots x_n$ without making any change in the group element being represented, provided the new expression makes sense. (For example, you can't have an empty pair of parentheses, and the number of left parentheses has to be the same as the number of right parentheses.) Thus the calculation in the proof of Proposition 2.2 can be simplified to

$$(y^{-1}x^{-1})(xy) = y^{-1}(x^{-1}x)y^{-1} = y1y^{-1} = 1.$$

2.1.3 The symmetric groups $S(n)$

We now come to the symmetric groups, also known as the permutation groups. They form the single most important class of finite groups. All permutation groups of order greater than two are nonabelian, but more importantly, permutation groups are one of the foundational tools in the discipline of combinatorics. The symmetric group is undoubtedly the single most frequently encountered finite group in mathematics. In fact, as we shall soon see, all finite groups of order n can be realized inside the symmetric group $S(n)$. This fact, known as Cayley's theorem, will be proved at the end of this section.

Let X denote a set. A bijective mapping $\sigma : X \to X$ will be called a *permutation* of X. The set of all permutations of X is denoted by $\mathrm{Sym}(X)$ and (due to the next result) is called the *symmetric group of* X. When $X = \{1, 2, \ldots, n\}$, $\mathrm{Sym}(X)$ is denoted by $S(n)$ and called (somewhat inaccurately) the *symmetric group on n letters*.

Proposition 2.3. *The set* $\mathrm{Sym}(X)$ *of permutations of* X *is a group under composition whose identity element is the identity map* $id_X : X \to X$. *If* $|X| = n$, *then* $|\mathrm{Sym}(X)| = n!$.

Proof. That $\mathrm{Sym}(X)$ is a group follows from the fact that the composition of two bijections is a bijection, the inverse of a bijection is a bijection, and the identity map is a bijection. The associativity follows from the fact that composition of mappings is associative. Now suppose that $X = \{x_1, \ldots, x_n\}$. To define an element of $\mathrm{Sym}(X)$, it suffices, by the pigeonhole principle (see Chap. 1), to define an injection $\sigma : X \to X$. Note that

there are n choices for the image $\sigma(x_1)$. In order to ensure that σ is one to one, $\sigma(x_2)$ cannot be $\sigma(x_1)$. Hence there are $n-1$ possible choices for $\sigma(x_2)$. Similarly, there are $n-2$ possible choices for $\sigma(x_3)$, and so on. Thus the number of injective maps $\sigma : X \to X$ is $n(n-1)(n-2)\cdots 2\cdot 1 = n!$. Therefore, $|\mathrm{Sym}(X)| = n!$. □

In the following example, we will consider a scheme for writing down the elements of $S(3)$ that easily generalizes to $S(n)$ for all $n > 0$. We will also see that $S(3)$ is nonabelian.

Example 2.2. To write down the six elements σ of $S(3)$, we need a way to encode $\sigma(1)$, $\sigma(2)$, and $\sigma(3)$. To do so, we represent σ by the array

$$\begin{pmatrix} 1 & 2 & 3 \\ \sigma(1) & \sigma(2) & \sigma(3) \end{pmatrix}.$$

For example, if $\sigma(1) = 2$, $\sigma(2) = 3$, and $\sigma(3) = 1$, then

$$\sigma = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}.$$

We will leave it to the reader to complete the list of elements of $S(3)$. If $n > 2$, then $S(n)$ is nonabelian: the order in which two permutations are applied matters. For example, if $n = 3$ and

$$\tau = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix},$$

then

$$\sigma\tau = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix},$$

while

$$\tau\sigma = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix}.$$

Hence $\sigma\tau \neq \tau\sigma$. ⊠

2.1.4 Cyclic groups

The next class of groups we will consider consists of the cyclic groups. Before defining these groups, we need to explain how exponents work. If G is a group and $x \in G$, then if m is a positive integer, $x^m = x \cdots x$ (m factors). We define x^{-m} to be $(x^{-1})^m$. Also, $x^0 = 1$. Then the usual laws of exponentiation hold for all integers m, n:

(i) $x^m x^n = x^{m+n}$,

(ii) $(x^m)^n = x^{mn}$.

Definition 2.4. A group G is said to be *cyclic* if there exists an element $x \in G$ such that for every $y \in G$, there is an integer m such that $y = x^m$. Such an element x is said to *generate* G.

In particular, $\mathbb{Z}$ is an example of an infinite cyclic group in which z^m is interpreted as mz. The multiplicative group $G = \{1, -1\}$ is also cyclic. The additive groups $\mathbb{Z}_m$ consisting of integers modulo a positive integer m form an important class of cyclic groups, which we will define after discussing quotient groups. They are the building blocks of the finite (in fact, finitely generated) abelian groups. However, this will not be proven until Chap. 11. Notice that all cyclic groups are abelian. However, cyclic groups are not so common. For example, $S(3)$ is not cyclic, nor is $\mathbb{Q}$, the additive group of rational numbers. We will soon prove that all finite groups of prime order are cyclic.

When G is a finite cyclic group and $x \in G$ is a generator of G, one often writes $G =< x >$. It turns out, however, that a finite cyclic group can have several generators, so the expression $G =< x >$ is not necessarily unique. To see an example of this, consider the twenty-four hour clock as a finite cyclic group. This is a preview of the group $\mathbb{Z}_m$ of integers modulo m, where $m = 24$.

Example 2.3. Take a clock with 24 hours numbered 0 through 23. The group operation on this clock is time shift by some whole number n of hours. A forward time shift occurs when n is positive, and a backward time shift occurs when n is negative. When $n = 0$, no shift occurs, so hour 0 will be the identity. A one-hour time shift at 23 hours sends the time to 0 hours, while a two-hour time shift sends 23 hours to 1 hour, and so on. In other words, the group operation is addition modulo 24. The inverse of, say, the ninth hour is the fifteenth hour. Two hours are inverse to each other if shifting one by the other puts the time at 0 hour. This makes the 24-hour clock into a group of order 24, which is in fact cyclic, since repeatedly time shifting by one hour starting at 0 hours can put you at any hour. However, there are other generators, and we will leave it as an exercise to find all of them. ⊠

Another interesting finite cyclic group is the group C_n of nth roots of unity, that is, the solutions of the equation $\zeta^n = 1$. We will postpone the discussion of C_n until we consider the complex numbers $\mathbb{C}$.

2.1.5 Dihedral groups: generators and relations

In the next example, we give an illustration of a group G that is described by giving a set of its generators and the relations the generators satisfy. The

group we will study is called the dihedral group. We will see in due course that the dihedral groups are the symmetry groups of the regular polygons in the plane. As we will see in this example, defining a group by giving generators and relations does not necessarily reveal much information about the group.

Example 2.4. (Dihedral Groups) The dihedral groups are groups that are defined by specifying two generators a and b and also specifying the relations that the generators satisfy. When we define a group by generators and relations, we consider all words in the generators, in this case a and b: these are all the strings or products $x_1x_2\cdots x_n$, where each x_i is either a or b, and n is an arbitrary positive integer. For example, $abbaabbaabba$ is a word with $n = 16$. Two words are multiplied together by placing them side by side. Thus,

$$(x_1x_2\cdots x_n)(y_1y_2\cdots y_p) = x_1x_2\cdots x_ny_1y_2\cdots y_p.$$

This produces an associative binary operation on the set of words. The next step is to impose some relations that a and b satisfy. Suppose $m > 1$. The dihedral group $D(m)$ is defined to be the set of all words in a and b with the above multiplication that we assume is subject to the following relations:

$$a^m = b^2 = 1, \quad ab = ba^{m-1}. \tag{2.1}$$

It is understood that the cyclic groups $< a >$ and $< b >$ have orders m and two respectively. By (2.1), $a^{-1} = a^{m-1}$ and $b = b^{-1}$. For example, if $m = 3$, then $a^3 = b^2 = 1$, so

$$aaababababb = (aaa)(bab)(ab)(bb) = (a^2)(ab) = a^3b = b.$$

The reader can show that $D(3) = \{1, a, a^2, b, ab, ba\}$. For example, $a^2b = a(ab) = a(ba^2) = (ab)a^2 = ba^4 = ba$. Hence, $D(3)$ has order 6. We will give a more convincing argument in due course. ⊠

Example 2.5. Let us now verify that $D(2)$ is a group. Since the multiplication of words is associative, it follows from the requirement that $a^2 = b^2 = 1$ and $ab = ba$ that every word can be collapsed to one of $1, a, b, ab, ba$. But $ab = ba$, so $D(2) = \{1, a, b, ab\}$. To see that $D(2)$ is closed under multiplication, we observe that $a(ab) = a^2b = b$, $b(ab) = (ba)b = ab^2 = a$, $(ab)a = (ba)a = ba^2 = b$, and $(ab)(ab) = (ba)(ab) = ba^2b = b^2 = 1$. Therefore, $D(2)$ is closed under multiplication, so it follows from our other remarks that $D(2)$ a group. Note that the order of $D(2)$ is 4. ⊠

It turns out that $D(m)$ is a finite group of order $2m$ for all $m > 0$. This will be proved in Example 2.12. But first we must define subgroups and show that every subgroup H of a group G partitions G into disjoint subsets gH, called cosets, all of which have the same number of elements when G is finite.

Another computation of the order $|D(m)|$ uses the fact that $D(m)$ is the symmetry group of a regular m-gon and a principle called $O(2,\mathbb{R})$-dichotomy. The details of this are in Section 7.3.7.

2.1.6 Subgroups

We now single out the most important subsets of a group: namely those that are also groups.

Definition 2.5. A nonempty subset H of a group G is called a *subgroup* of G if whenever $x, y \in H$, we have $xy^{-1} \in H$.

In particular, since every subgroup is nonempty, every subgroup H of G contains the identity of G, hence also the inverses of all of its elements. Moreover, by definition, H is closed under the group operation of G. Finally, associativity of the group operation on H follows from its associativity in G. Consequently, every subgroup of G is also a group. Thus we have proved the following result.

Proposition 2.4. *A subset H of a group G is a subgroup if and only if H is a group under the group operations of G. That is, H is closed under the group operation and contains the identity of G, and the inverse of an element of H is its inverse in G.*

Example 2.6. Suppose G denotes the integers. The even integers make up a subgroup of G, since the difference of two even integers is even. On the other hand, the odd integers do not, since the difference of two odd integers is even. ⊠

Example 2.7. Here are some other examples of subgroups.

(*i*) Let $m \in \mathbb{Z}$. Then all integral multiples mn of m form the subgroup $m\mathbb{Z}$ of $\mathbb{Z}$.

(*ii*) In every group, the identity element 1 determines the trivial subgroup $\{1\}$.

(*iii*) If H and K are subgroups of a group G, then $H \cap K$ is also a subgroup of G.

(*iv*) If $G =< a >$ is a finite cyclic group and $k \in \mathbb{Z}$, then $H =< a^k >$ is a subgroup of G. Note that $< a^k >$ need not be a proper subgroup of $< a >$. If $|G| = n$, then $H = G$ if and only if the greatest common divisor of k and n is 1. ⊠

(*v*) The dihedral group $D(m)$ generated by a and b defined in Example 2.4 has a cyclic subgroup of order m, namely $< a >$. The subgroup $< b >$ is cyclic of order two. ⊠

Here is a nice criterion for a subgroup.

Proposition 2.5. *Let G be a group and suppose H is a nonempty finite subset of G such that for every $a, b \in H$, we have $ab \in H$. Then H is a subgroup of G.*

Proof. Consider the mapping $L_a : G \to G$ defined by left multiplication by a. That is, $L_a(g) = ag$. This mapping is injective; for if $ag = ag'$, then left multiplication by a^{-1} gives $g = g'$. By assumption, if $a \in H$, then $L_a(h) \in H$ for all $h \in H$. Hence, since H is finite, the pigeonhole principle implies that L_a is a bijection of H onto H. Now H contains an element a, so there exists an $h \in H$ such that $L_a(h) = ah = a$. But since G is a group, it follows that $h = 1$. Thus, H contains the identity of G. It follows that a has a right inverse, since $L_a(h) = 1$ for some $h \in H$. Hence every element $a \in H$ has an inverse in H, so H satisfies the property that for every $a, b \in H$, $ab^{-1} \in H$. Consequently, H is a subgroup, by definition. □

Remark. Note that for a group G and $a \in G$, the left-multiplication mapping $L_a : G \to G$ is a bijection. For L_a is injective by the proof of the above proposition. It is also surjective, since if $g \in G$, the equation $L_a(x) = ax = g$ has a solution, namely $x = a^{-1}g$. The same remark holds for right multiplication $R_a : G \to G$, which is defined by $R_a(g) = ga$.

2.1.7 Homomorphisms and Cayley's Theorem

One often needs to compare or relate groups. The basic tool for this is given by the notion of a homomorphism.

Definition 2.6. If G and H are groups, then a mapping $\varphi : G \to H$ is said to be a *homomorphism* if and only if $\varphi(gg') = \varphi(g)\varphi(g')$ for all $g, g' \in G$. A bijective homomorphism is called an *isomorphism*. If there exists an isomorphism $\varphi : G \to H$, we will say that G and H are *isomorphic* and write $G \cong H$. The *kernel* of a homomorphism $\varphi : G \to H$ is defined to be

$$\ker(\varphi) = \{g \in G \mid \varphi(g) = 1 \in H\}.$$

Here is an interesting example that we will generalize several times.

Example 2.8. Let $\mathbb{R}^*$ denote the nonzero real numbers. Multiplication by $a \in \mathbb{R}^*$ defines a bijection $\mu_a : \mathbb{R} \to \mathbb{R}$ given by $\mu_a(r) = ar$. The distributive law for $\mathbb{R}$ says that

$$\mu_a(r + s) = a(r + s) = ar + as = \mu_a(r) + \mu_a(s)$$

for all $r, s \in \mathbb{R}$. Thus $\mu_a : \mathbb{R} \to \mathbb{R}$ is a homomorphism for the additive group structure on $\mathbb{R}$. Since $a \neq 0$, μ_a is in fact an isomorphism. Furthermore, the

associative and commutative laws for $\mathbb{R}$ imply that $a(rs) = (ar)s = (ra)s = r(as)$. Hence,

$$\mu_a(rs) = r\mu_a(s).$$

Combining these two identities says that μ_a is a *linear mapping* of $\mathbb{R}$. (Linear mappings will be studied in great detail later.) The linear mappings μ_a form an important group, denoted by $GL(1,\mathbb{R})$ called the *general linear group of* $\mathbb{R}$. The corresponding general linear group $GL(n,\mathbb{R})$ of linear bijections of $\mathbb{R}^n$ will be introduced in Section 4.2. I claim that $\mathbb{R}^*$ and $GL(1,\mathbb{R})$ are isomorphic via the homomorphism $\mu : \mathbb{R}^* \to GL(1,\mathbb{R})$ defined by $\mu(a) = \mu_a$. We leave this claim as an exercise. The cyclic subgroup $< -1 >= \{-1, 1\}$ of $\mathbb{R}^*$ is the one-dimensional case of an important subgroup $O(n)$ of $GL(n,\mathbb{R})$ called the orthogonal group. We may thus denote $< -1 >$ by $O(1)$. ⊠

Example 2.9. If $g \in G$, the mapping $\sigma_g : G \to G$ defined by putting $\sigma_g(h) = ghg^{-1}$ is an isomorphism. The mapping σ_g is called *conjugation by* g. An isomorphism of the form σ_g is called an *inner automorphism of* G. An isomorphism $\sigma : G \to G$ that is not of the form σ_g for some $g \in G$ is said to be an *outer automorphism* of G. ⊠

Notice that if G is abelian, then its only inner automorphism is the identity map $I_G(g) = g$ for all $g \in G$.

Proposition 2.6. *The image of a homomorphism $\varphi : G \to G'$ is a subgroup of G', and its kernel is a subgroup of G.*

We leave this as an exercise. An example of an outer automorphism $\varphi : G \to G$ is given by letting G be any abelian group and putting $\varphi(g) = g^{-1}$. For example, if $G = \mathbb{Z}$, then $\varphi(m) = -m$.

The next result, known as Cayley's theorem, reveals a nontrivial fundamental property of the symmetric group.

Theorem 2.7. *Every finite group G is isomorphic to a subgroup of* $\mathrm{Sym}(G)$. *Hence if $|G| = n$, then G is isomorphic to a subgroup of $S(n)$.*

Proof. As already noted in the remark following Proposition 2.5, the mapping $L_a : G \to G$ defined by left multiplication by $a \in G$ is a bijection of G. Thus, by definition, $L_a \in \mathrm{Sym}(G)$. Now let $\varphi : G \to \mathrm{Sym}(G)$ be defined by $\varphi(a) = L_a$. I claim that φ is a homomorphism. That is, $\varphi(ab) = \varphi(a)\varphi(b)$. For if $a, b, c \in G$, then

$$\varphi(ab)(c) = L_{ab}(c) = (ab)c = a(bc) = L_a(L_b(c))$$

by associativity. But

$$\varphi(a)\varphi(b)(c) = \varphi(a)(L_b(c)) = L_a(L_b(c)),$$

so indeed φ is a homomorphism. Hence $H = \varphi(G)$ is a subgroup of $\mathrm{Sym}(G)$. To show that $\varphi : G \to H$ is an isomorphism, it suffices to show that φ is injective. But if $\varphi(a) = \varphi(b)$, then $ag = bg$ for all $g \in G$. This implies $a = b$, so φ is indeed injective. □

Exercises

Exercise 2.1.1. Let S be a finite set. Show that $\mathrm{Sym}(S)$ is a group.

Exercise 2.1.2. Show that $S(n)$ is nonabelian for all $n > 2$.

Exercise 2.1.3. Let $G = S(3)$ and let $\sigma_1, \sigma_2 \in G$ be given by

$$\sigma_1 = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix} \quad \text{and} \quad \sigma_2 = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix}.$$

(*i*) Compute $\sigma_1\sigma_2$ and $\sigma_2\sigma_1$ as arrays.

(*ii*) Likewise, compute $\sigma_1\sigma_2\sigma_1^{-1}$.

Exercise 2.1.4. The purpose of this exercise is to enumerate the elements of $S(3)$. Let $\sigma_1, \sigma_2 \in G$ be defined as in Exercise 2.1.3. Show that every element of $S(3)$ can be expressed as a product of σ_1 and σ_2.

Exercise 2.1.5. In part (*iii*) of Example 2.7, we asserted that if $G =< a >$ is a finite cyclic group of order n and if $m \in \mathbb{Z}$, then $H =< a^m >$ is a subgroup of G, and $H = G$ if and only if the greatest common divisor of m and n is 1. Prove this. Note: the greatest common divisor (gcd) of m and n is the largest integer dividing both m and n. The key property of the gcd is that the gcd of m and n is d if and only if there exist integers a and b such that $am + bn = d$.

Exercise 2.1.6. For the notation in this exercise, see Example 2.8. Prove that $\mathbb{R}^*$ and $GL(1, \mathbb{R})$ are isomorphic via the mapping $\mu : \mathbb{R}^* \to GL(1, \mathbb{R})$ defined by $\mu(a) = \mu_a$.

Exercise 2.1.7. Suppose Alice, Bob, Carol, and Ted are seated in the same row at a movie theater in the order $ABCT$. Suppose Alice and Ted switch places, then Ted and Carol switch, and finally Bob and Ted switch. Putting $A = 1$, $B = 2$, $C = 3$ and $T = 4$, Do the same if Ted and Carol switch first, then Bob and Ted switch, and Alice and Ted switch last. Compare the results.

Exercise 2.1.8. Suppose only Alice, Bob, and Carol are seated in a row. Find the new seating arrangement if Alice and Bob switch, then Bob and Carol switch, and finally Alice and Bob switch again. Now suppose Bob and Carol switch first, then Alice and Bob switch, and finally Bob and Carol switch again.

(*i*) Without computing the result, how do you think the seating arrangements differ?

(*ii*) Now compute the new arrangements and comment on the result.

Exercise 2.1.9. Prove Proposition 2.6. That is, show that if G and G' are groups and $\varphi : G \to G'$ is a homomorphism, then the image of φ is a subgroup of G', and its kernel is a subgroup of G.

Exercise 2.1.10. Let G be a group and $g \in G$. Prove that the inner automorphism $\sigma_g : G \to G$ defined by $\sigma_g(h) = ghg^{-1}$ is an isomorphism.

Exercise 2.1.11. Let $\mathrm{Aut}(G)$ denote the set of all automorphisms of a group G.

(*i*) Prove that $\mathrm{Aut}(G)$ is a group;

(*ii*) Prove that the inner automorphism $\sigma_g : G \to G$ defined by $\sigma_g(h) = ghg^{-1}$ is an element of $\mathrm{Aut}(G)$;

(*iii*) Prove that the mapping $\Phi : G \to \mathrm{Aut}(G)$ defined by $\Phi(g) = \sigma_g$ is a homomorphism; and

(*iv*) Describe the kernel of Φ. The kernel is called the *center* of G.

(*v*) Show that if G is finite, then $\mathrm{Aut}(G)$ is also finite.

2.2 The Cosets of a Subgroup and Lagrange's Theorem

Suppose G is a group and H is a subgroup of G. We are now going to use H to partition G into mutually disjoint subsets called cosets. In general, cosets are not subgroups; they are translates of H by elements of G. Hence there is a bijection of H onto each of its cosets. Actually, cosets come in two varieties: left cosets and right cosets. When G is a finite group, every pair of cosets of both types of a subgroup H have the same number of elements. This is the key fact in the proof of Lagrange's Theorem, which was one of the first theorems in group theory and is still an extremely useful result.

2.2.1 The definition of a coset

Suppose G is a group and H is a subgroup of G.

Definition 2.7. For every $x \in G$, the subset

$$xH = \{g \in G \mid g = xh \ \ \exists h \in H\}$$

is called the *left coset* of H containing x. Similarly, the subset

$$Hx = \{g \in G \mid g = hx \ \ \exists h \in H\}$$

of G is called the *right coset* of H containing x. Note, xH and Hx both contain x, since $1 \in H$. The set of left cosets of H is denoted by G/H, while the set of right cosets of H is denoted by $H\backslash G$. We will call x a *representative* of either coset xH or Hx.

Let us make some basic observations. Since $x \in xH$, it follows that G is the union of all the left cosets of H:

$$G = \bigcup_{x \in G} xH.$$

Of course, a similar assertion holds for the right cosets. Since two cosets xH and yH may intersect or even coincide, the above expression for G needs to be made more precise. To do so, we will consider how two cosets xH and yH intersect. The answer might be a little surprising.

Proposition 2.8. *Two left cosets xH and yH of the subgroup H of G are either equal or disjoint. That is, either $xH = yH$ or $xH \cap yH$ is empty. Furthermore, $xH = yH$ if and only if $x^{-1}y \in H$. Consequently, G is the disjoint union of all the left cosets of H.*

Proof. We will show that if xH and yH have at least one element in common, then they coincide. Suppose $z \in xH \cap yH$. Then $z = xh = yk$. Now we observe that $xH \subset yH$. For if $u \in xH$, then $u = xj$ for some $j \in H$. Since $x = ykh^{-1}$, $u = ykh^{-1}j$. But $kh^{-1}j \in H$, since H is a subgroup. Thus $u \in yH$. Similarly, $yH \subset xH$, so indeed, $xH = yH$. Thus, two left cosets that have a nonempty intersection coincide. Consequently, two cosets are either equal or disjoint. To prove the second statement, assume $xH = yH$. Then $xh = yk$ for some $h, k \in H$, so $x^{-1}y = hk^{-1} \in H$. On the other hand, if $x^{-1}y \in H$, then $y = xh$ for some $h \in H$. Thus xH and yH have y as a common element. Therefore, $xH = yH$ by the first statement of the proposition. The final statement follows from the fact that every element of G is in a coset. □

If G is abelian, then $xH = Hx$ for all $x \in G$ and all subgroups H. In nonabelian groups, it is often the case that $xH \neq Hx$. Left and right cosets may be different. We will see that subgroups H such that $xH = Hx$ for all $x \in G$ play a special role in group theory. They are called *normal subgroups.* Let us now consider some examples.

Example 2.10. Let us find the cosets of the subgroup $m\mathbb{Z}$ of $\mathbb{Z}$. By definition, every coset has the form $k + m\mathbb{Z} = \{k + mn \mid n \in \mathbb{Z}\}$. Suppose we begin with $m = 2$. In this case, $m\mathbb{Z} = 2\mathbb{Z}$ is the set of even integers. Then $2\mathbb{Z}$ and $1 + 2\mathbb{Z}$ are cosets. Similarly, $3 + 2\mathbb{Z}$ is also a coset, as is $4 + 2\mathbb{Z}$. We could continue, but it is better to check first whether all these cosets are distinct. For example, $4 + 2\mathbb{Z} = 0 + 2\mathbb{Z}$, since $4 - 0$ is even. Similarly, $3 - 1 \in 2\mathbb{Z}$, so $3 + 2\mathbb{Z} = 1 + 2\mathbb{Z}$. Thus when $m = 2$, there are only two cosets: the even integers and the odd integers. Now suppose m is an arbitrary positive integer. Then by similar reasoning, the cosets are

$$m\mathbb{Z},\ 1 + m\mathbb{Z},\ 2 + m\mathbb{Z}, \ldots,\ (m-1) + m\mathbb{Z}.$$

In other words, $m\mathbb{Z}$ has m distinct cosets. Note that since $m\mathbb{Z}$ is an abelian group,
the right cosets and the left cosets are the same: $k + m\mathbb{Z} = m\mathbb{Z} + k$ for all integers k. ⊠

In number theory, one says that two integers r and s are *congruent modulo* m if their difference is a multiple of m: $r - s = tm$. When r and s are congruent modulo m, one usually writes $r \equiv s \bmod(m)$. Interpreted in terms of congruence, Proposition 2.8 says that $r \equiv s \bmod(m)$ if and only if r and s are in the same coset of $m\mathbb{Z}$ if and only if $r + m\mathbb{Z} = s + m\mathbb{Z}$.

Example 2.11 (Cosets in $\mathbb{R}^2$)**.** Recall that $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$ denotes the set of all ordered pairs $\{(r, s) \mid r, s \in \mathbb{R}\}$. By the standard identification, $\mathbb{R}^2$ is the Cartesian plane. It is also an abelian group by componentwise addition: $(a, b) + (c, d) = (a + c, b + d)$. The solution set ℓ of a linear equation

$rx + sy = 0$ is a line in $\mathbb{R}^2$ through the origin $(0, 0)$. Note that $(0, 0)$ is the identity element of $\mathbb{R}^2$. Note also that every line ℓ through the origin is a subgroup of $\mathbb{R}^2$. By definition, every coset of ℓ has the form $(a, b) + \ell$, and thus the cosets of ℓ are lines in $\mathbb{R}^2$ parallel to ℓ. For if $(u, v) \in \ell$, then $ru + sv = 0$; hence $r(a + u) + s(b + v) = (ra + sb) + (ru + sv) = ra + sb$. Thus the coset $(a, b) + \ell$ is the line $rx + sy = d$, where $d = ra + sb$. If $d \neq 0$, this is a line in $\mathbb{R}^2$ parallel to ℓ, which verifies our claim. Consequently, the coset decomposition of $\mathbb{R}^2$ determined by a line ℓ through the origin consists of ℓ together with the set of all lines in $\mathbb{R}^2$ parallel to ℓ. Finally, we remark that every subgroup of $\mathbb{R}^2$ distinct from $\mathbb{R}^2$ and $\{(0, 0)\}$ is a line through the origin, so we have determined all possible cosets in $\mathbb{R}^2$: every coset is either $\mathbb{R}^2$, a point, or a line. In Exercise 2.2.7, the reader is asked to carefully supply the details of this example. ⊠

The final example of this section verifies our claim about the order of the dihedral group.

Example 2.12 (The order of the dihedral group $D(m)$ is $2m$)**.** Let m be a positive integer. In this example, we will compute a coset decomposition of the dihedral group $D(m)$ and use it to show that $|D(m)| = 2m$. Recall from Example 2.4 that $D(m)$ is generated by two elements a and b satisfying $a^m = 1$, $b^2 = 1$, and $ab = ba^{m-1}$. Let H denote the cyclic subgroup $< a >$, and let us compute the cosets of H. One is H, and another is bH, since $b \notin H$. Now consider abH. Since $ab = ba^{-1}$, it follows that $abH = bH$. Furthermore, $b^2H = H$. Thus the coset decomposition of $D(m)$ into left cosets of H has to be

$$D(m) = H \bigcup bH.$$

Since $|H| = m$, and $|bH| = |H|$ due to the fact that $L_b(h) = bh$ defines a bijection of H to bH, it follows that $|D(m)| = m + m = 2m$. ⊠

2.2.2 Lagrange's Theorem

The most important consequence of the fact that the cosets of a subgroup of a finite group partition the group into subsets all having the same number of elements is the famous theorem of Lagrange, which we will now prove. We first note the following result.

Proposition 2.9. *If G is a group and H is a finite subset of G, then for every $a \in G$, $|aH| = |Ha| = |H|$.*

Proof. Recall that the left multiplication mapping $L_a : G \to G$ defined by $L_a(g) = ag$ is injective. Since $L_a(H) = aH$, we conclude that $|aH| = |H|$. Similarly, $|Ha| = |H|$. □

We now prove Lagrange's theorem. It is undoubtedly the most frequently cited elementary result on finite groups.

Theorem 2.10. *Suppose G is a finite group and H is a subgroup of G. Then the order of H divides the order of G. In fact,*

$$|G|/|H| = |G/H|.$$

Proof. Every element of G is in a unique left coset of H, and by the previous lemma, any two cosets have the same number of elements. Since distinct cosets are disjoint, we may conclude that $|G| = |G/H||H|$. □

Definition 2.8. If $|G|$ is finite, the common value of $|G/H|$ and $|H\backslash G|$ is called the *index of H in G*.

Since the left and right cosets in an abelian group are the same, it is also possible to define the index of a subgroup of an infinite abelian group.

Definition 2.9. If a subgroup H of an arbitrary abelian group G has only finitely many cosets, then the *index of H in G* is defined to be $|G/H|$. If G/H is infinite, we say that H has *infinite index in G*.

For example, by Example 2.10, the index of $H = m\mathbb{Z}$ in $G = \mathbb{Z}$ is exactly m. Let us now consider the order of an element $x \neq 1$.

Definition 2.10. Let G be a group. An element $x \neq 1 \in G$ is said to have *finite order* if $x^m = 1$ for some $m > 0$. The *order* of an element x of finite order is the smallest integer $n > 0$ such that $x^n = 1$. By assumption, the identity has order one.

Notice that if x has order $n > 0$ and $x^m = 1$ for some $m > 0$, then n divides m. Reason: By definition, $m \geq n$. Dividing m by n, we can write $m = sn + r$, where s and r are nonnegative integers and $0 \leq r < n$. But then $x^m = x^{sn}x^r = x^r = 1$. Since $r < n$, it follows by the definition of n that $r = 0$. Of course, groups that are not finite need not have any elements of finite order (example?). Another remark is that an element and its inverse have the same order. Let us now derive the first of several applications of Lagrange's theorem.

Proposition 2.11. *Every element in a finite group G has finite order, and the order of every element divides the order of the group.*

Proof. If $x \in G$, the powers $x, x^2, x^3, \ldots$ cannot all be distinct. Hence there are positive integers $r < s$ such that $x^r = x^s$. Consequently, $x^{s-r} = 1$. Hence every element of G has finite order. Let x have order n. Then the cyclic group $< x >= \{1, x, \ldots, x^{n-2}, x^{n-1}\}$ is a subgroup of G of order n, so by Lagrange, n divides $|G|$. □

Recall that an integer $p > 1$ is said to be *prime* if its only positive integer divisors are 1 and p. The next result is an immediate consequence of Proposition 2.11.

Corollary 2.12. *A group of prime order is cyclic.*

Another useful result is the following.

Proposition 2.13. *In a finite group, the number of elements of prime order p is divisible by $p - 1$.*

Proof. By Lagrange, two distinct subgroups of prime order p meet exactly at 1, for their intersection is a subgroup of both. But in a group of order p, every element except the identity has order exactly p, since p is prime. Thus the number of elements of order p is $m(p-1)$ where m is the number of subgroups of order p. □

The reader may wonder whether Lagrange's theorem has a converse: does a group of order m have a subgroup of order k for every divisor k of m? It turns out that the answer is no. For example, a group of order 12 need not have a subgroup of order 4. However, there is a famous theorem of Cauchy that says that if a prime p divides $|G|$, then G contains an element of order p. What one can say in general about the converse of Lagrange's theorem is partially answered by the Sylow theorems, which describe the Sylow subgroups of G, namely those subgroups whose order is the highest power p^m of a prime p dividing $|G|$. These results are all proved in Chap. 11.

Exercises

Exercise 2.2.1. Suppose G is a group and H is a subgroup of G. Show that the relation on G given by $x \equiv y$ if and only if $x^{-1}y \in H$ is an equivalence relation whose equivalence classes are exactly the left cosets of H. Conclude that G is the disjoint union of its left cosets. This gives an alternative proof of Proposition 2.8.

Exercise 2.2.2. Let G have order 15 and let H have order 8. Does there exist a surjective homomorphism $\varphi : G \to H$?

Exercise 2.2.3. Let G have order 9 and let G' have order 20. Describe the set of all homomorphisms $\varphi : G \to G'$.

Exercise 2.2.4. Suppose G is a finite cyclic group.

(*i*) Show that every subgroup of G is also cyclic.

(*ii*) Show that for every divisor k of m, there exists a cyclic subgroup of G having order k.

Exercise 2.2.5. Show that if φ is a homomorphism on a group G and if H is the kernel of φ, then φ is constant on each coset of H, and φ takes different values on different cosets.

Exercise 2.2.6. Let G be an abelian group, and suppose x and y are elements of G of finite orders m and n respectively.

(*i*) Show that the order of xy is the least common multiple of m and n.

(*ii*) Give an explicit example of a finite nonabelian group G containing noncommuting elements x and y such that the order of xy is not the least common multiple of m and n. (Try $S(3)$.)

Exercise 2.2.7. Consider the plane $\mathbb{R}^2 = \{(x, y) \mid x, y \in \mathbb{R}\}$, and define addition on $\mathbb{R}^2$ by $(x, y) + (u, v) = (x + u, y + v)$.

(*i*) Show that $\mathbb{R}^2$ is an abelian group under addition.

(*ii*) Let $a, b \in \mathbb{R}$, and put $\ell = \{(x, y) \mid ax + by = 0\}$. Thus ℓ is a line in $\mathbb{R}^2$ through the origin $(0, 0)$. Show that ℓ is a subgroup of $\mathbb{R}^2$, and conclude that lines through the origin in $\mathbb{R}^2$ are subgroups.

(*iii*) Show that if $(r, s) \in \mathbb{R}^2$, then the coset $(r, s) + \ell$ of ℓ is a line in $\mathbb{R}^2$. In fact, show that $(r, s) + \ell$ is the line $ax + by = ar + bs$. Conclude from this that any two cosets of ℓ coincide or are disjoint.

Exercise 2.2.8. Consider the group $\mathbb{Z}_{24}$ as defined in the 24-hour clock example. Find the number of elements of order n for each divisor of 24.

2.3 Normal Subgroups and Quotient Groups

Suppose G is a group and H is a subgroup of G. Recall that G/H is the set of all left cosets of H in G. The plan in this section is to define what is called a *normal subgroup* of G and to show that when H is a normal subgroup of G, then G/H is also a group. The group G/H is called the *quotient of G by H*.

2.3.1 Normal subgroups

Suppose H is a subgroup of an arbitrary group G. Let us consider what it means to impose the condition that a left coset xH is also a right coset Hy.

Proposition 2.14. *Suppose the left coset xH coincides with the right coset Hy for a pair $x, y \in G$. Then $x = hy$ for some $h \in H$, so $Hy = Hx$. Thus $xH = Hx$, and consequently, $xHx^{-1} = H$.*

Proof. Left to the reader.

It follows that the subgroups H of G such that every left coset of H is a right coset are characterized by the property that $xHx^{-1} = H$ for all $x \in G$. From now on, such subgroups will be called *normal subgroups*.

Definition 2.11. Let H be a subgroup of G. The *normalizer* of H in G is defined to be $N_G(H) = \{g \in G \mid gHg^{-1} = H\}$.

We leave it as an exercise to show that $N_G(H)$ is a subgroup of G and H is a normal subgroup of $N_G(H)$. For example, if G is abelian, then every subgroup is normal. By the above proposition, $N_G(H)$ consists of those $x \in G$ such that the coset xH is also a right coset.

Example 2.13. Let m be a positive integer, and let G denote the dihedral group $D(m)$ with generators a and b satisfying $a^m = b^2 = 1$ and $ab = ba^{m-1}$. Let us show that the cyclic subgroup $H =< a >$ of order m is normal. To see this, note that since $b = b^{-1}$, we know that $bab^{-1} = a^{m-1}$. It follows that $b \in N_G(H)$. Moreover, $a \in N_G(H)$, so $G = N_G(H)$, since $N_G(H)$ contains the generators a, b of G. Therefore, $< a >$ is normal in $D(m)$. Now consider the subgroup $< b >$. Since $aba^{-1} = ba^{m-1}a^{-1} = ba^{m-2}$, $aba^{-1} \notin < b >$ unless $m = 2$. But in this case, as we have already seen, G is abelian. Therefore, for $m > 2$, $< b >$ is not normal in $D(m)$. ⊠

The following proposition gives a well-known condition for a subgroup to be normal.

Proposition 2.15. *Let G be a finite group and suppose H is a subgroup of index two. Then H is normal in G.*

Proof. Recall that G is the disjoint union of both its left cosets and its right cosets. Hence if $x \notin H$, then

$$G = H \cup xH = H \cup Hx.$$

This implies that $xH = Hx$ for every $x \in G$. Therefore, H is normal. □

Since $|D(m)| = 2m$ (by Example 2.12) and the cyclic subgroup $< a >$ has order m, it follows that $< a >$ has index two. This gives another proof that $< a >$ is normal.

The next proposition gives a method for producing normal subgroups.

Proposition 2.16. *Let G and G' be groups, and let $\varphi : G \to G'$ be a homomorphism with kernel H. Then H is normal in G.*

Proof. Let $h \in H$ and take any $g \in G$. Then

$$\varphi(ghg^{-1}) = \varphi(g)\varphi(h)\varphi(g^{-1}) = \varphi(g)1\varphi(g^{-1}) = \varphi(gg^{-1}) = 1.$$

Thus, $ghg^{-1} \in H$ for all $g \in G$ and $h \in H$. Therefore, $G = N_G(H)$, so H is normal. □

2.3.2 Constructing the quotient group G/H

We will now construct the *quotient group* G/H of G by a normal subgroup H.

Proposition 2.17. *Suppose H is a normal subgroup of G. Then the set G/H of left cosets of H has a natural group structure under the binary operation $G/H \times G/H \to G/H$ defined by $(xH, yH) \to xyH$. Moreover, the natural mapping $\pi : G \to G/H$ sending $g \in G$ to its coset gH is a homomorphism. Similarly, the set of right cosets of H admits a group structure defined analogously.*

Proof. The first step is to show that the rule $(xH, yH) \to xyH$ doesn't depend on the way the cosets are represented. Thus, suppose $xH = uH$ and $yH = vH$. I claim that $xyH = uvH$. To see this, we must show that $(xy)^{-1}(uv) \in H$. Since $xH = uH$ and $vH = yH$, $x^{-1}u = h_1 \in H$ and $y^{-1}v = h_2 \in H$, so

$$(xy)^{-1}(uv) = y^{-1}x^{-1}uv = y^{-1}x^{-1}uyy^{-1}v = y^{-1}h_1yh_2 \in H,$$

since $y^{-1}h_1 y \in H$, due to the fact that H is a normal subgroup. Therefore, $xyH = uvH$, as claimed. We leave it as an exercise to verify that the group operation is associative. The other group properties are routine: $1 \in G/H$ is the identity coset H and the inverse $(xH)^{-1}$ is equal to $x^{-1}H$. Thus G/H satisfies the group axioms. Finally, it is evident that π is a homomorphism. □

The map $\pi : G \to G/H$ is called the *quotient map*. The quotient group G/H is often referred to as G modulo H, or simply G mod H. The terminology for the group of right cosets is the same.

Remark. It is worth noting that when H is normal in G, the product of the two cosets xH and yH is the coset xyH, where by the product of xH and yH, we mean the set $(xH)(yH) \subset G$ consisting of all elements of the form $xhyk$, where $h, k \in H$. We will leave the verification of this claim as an exercise.

Example 2.14 (Example 2.8 continued)**.** Since $\mathbb{R}^*$ is abelian, $O(1) = \{\pm 1\}$ is a normal subgroup. Each element of the quotient group $\mathbb{R}^*/O(1)$ is a coset that can be uniquely written $rO(1)$, where r is a positive real number. Note that the set $\mathbb{R}_{>0}$ of all positive reals is also a subgroup of $\mathbb{R}^*$. In fact, the mapping $\varphi : \mathbb{R}_{>0} \to \mathbb{R}^*/O(1)$ given by $\varphi(r) = rO(1)$ is an isomorphism, so $\mathbb{R}_{>0} \cong \mathbb{R}^*/O(1)$. ⊠

Example 2.15. One of the most important examples of a quotient group is the group $\mathbb{Z}_m = \mathbb{Z}/m\mathbb{Z}$ of integers modulo m, where m is a positive integer. Since $\mathbb{Z}$ is an abelian group under addition, the subgroup $m\mathbb{Z}$ is normal in $\mathbb{Z}$, so it follows that $\mathbb{Z}_m$ is indeed a group. As we noted after Example 2.10, the index of $m\mathbb{Z}$ in $\mathbb{Z}$ is m. Thus $|\mathbb{Z}/m\mathbb{Z}| = |\mathbb{Z}_m| = m$. ⊠

Here are two more interesting examples.

Example 2.16. Viewing $\mathbb{R}$ as an additive abelian group and $\mathbb{Z}$ as a normal subgroup, the quotient $S = \mathbb{R}/\mathbb{Z}$ of $\mathbb{R}$ modulo $\mathbb{Z}$ can be pictured by taking the unit interval $[0, 1]$ in $\mathbb{R}$ and identifying 0 and 1, thus producing a circle that represents S. The quotient map $\pi : \mathbb{R} \to \mathbb{R}/\mathbb{Z}$ wraps the real line $\mathbb{R}$ around the circle S an infinite number of times so that each interval $[n, n + 1]$ repeats what happened on $[0, 1]$. We will see below that the group operation on S can be explicitly realized using complex numbers and the complex exponential. In fact, S is then realized concretely as the unit circle in the complex plane. ⊠

Example 2.17. The plane $\mathbb{R}^2$, which consists of all pairs (x, y) of real numbers, is an abelian group under the addition $(u, v) + (x, y) = (u + x, v + y)$. Let $\mathbb{Z}^2$ denote the subgroup consisting of all pairs (m, n) of integers m and n. The quotient group $T = \mathbb{R}^2/\mathbb{Z}^2$ is called a *torus*. One can visualize T by taking the unit square S in $\mathbb{R}^2$ with vertices $(0, 0)$, $(1, 0)$, $(0, 1)$, and $(1, 1)$

and first identifying the horizontal edge from $(0,0)$ to $(1,0)$ with the horizontal edge from $(0,1)$ to $(1,1)$ placing $(0,0)$ on $(0,1)$ and $(1,0)$ on $(1,1)$. The result is a horizontal cylinder of length one. Next, identify the left edge circle with the right circle in the same way, placing the point $(0,0)$ on $(1,1)$. This produces a curved surface that looks like the surface of a doughnut or, in mathematical terminology, a torus. ⊠

When G is a finite group and H is normal in G, Lagrange's theorem gives us the following.

Proposition 2.18. *The quotient group G/H has order $|G/H| = \frac{|G|}{|H|}$. Hence, $|H|$ and $|G/H|$ both divide $|G|$.*

By Cauchy's theorem, which we mentioned above, if a prime p divides $|G|$, then G contains an element of order exactly p. This tells us something interesting about a group G of order $2p$: G has an element a of order p and an element b of order two. Moreover, the cyclic subgroup $< a >$ is normal, since its index in G is two. Thus, $bab^{-1} = a^r$. Then $G = D(2p)$ when $r = p - 1$. But there are other possibilities. For example, if $r = 1$, then $ab = ba$, so this implies that G is abelian. In fact, in the abelian case, the element ab has order $2p$, provided $p \neq 2$. Hence $G =< ab >$.

2.3.3 Euler's Theorem via quotient groups

There are several beautiful applications of Lagrange's theorem and quotient groups to number theory. One of the nicest is the proof of Euler's theorem that we now give. First, let us recall the *greatest common divisor*, or gcd, (m,n) of two integers m and n, already defined in Exercise 2.1.5. By definition, (m,n) is the largest positive integer d dividing both m and n, and $(m,n) = d$ if and only if there exist integers r and s such that $rm + sn = d$. When $(m,n) = 1$, we say that m and n are *relatively prime*. The proof of this characterization of (m,n) can be found in any book on elementary number theory. Euler's *phi function* $\phi : \mathbb{Z}_{>0} \to \mathbb{Z}_{>0}$ is defined as follows: if $m \in \mathbb{Z}_{>0}$, then $\phi(m)$ is the number of integers $a \in [1,m]$ such that $(a,m) = 1$.

Theorem 2.19 (Euler's theorem)**.** *Let m and a be positive integers such that $(a,m) = 1$. Then*

$$a^{\phi(m)} \equiv 1 \bmod(m).$$

Euler's theorem was published in 1736 and hence preceded groups and Lagrange's theorem. Thus the proof given here cannot be Euler's original proof. Despite its antiquity, the theorem has a modern application: it turns out to be one of the ideas that RSA encryption is based on. After defining a certain finite abelian group U_m known as the group of units modulo m, we

will apply Lagrange's theorem to deduce the result. The key turns out to be the fact that the order of U_m is $\phi(m)$.

Before defining U_m, we need to observe that the additive group $\mathbb{Z}_m$ of integers modulo m also has an associative and commutative multiplication such that $1 + m\mathbb{Z}$ is a multiplicative identity. For those already familiar with the notion of a ring, these claims follow because $\mathbb{Z}_m$ is also a ring. The construction of the multiplication will be repeated (in more detail) in Section 2.6.1, so we may omit some details. The product on $\mathbb{Z}_m$ is given by the rule

$$(a + m\mathbb{Z})(b + m\mathbb{Z}) = ab + m\mathbb{Z}.$$

This definition is independent of the integers a and b representing their corresponding cosets. Furthermore, this multiplication is associative and commutative, and $1 + m\mathbb{Z}$ is a multiplicative identity. We will say that the coset $a + m\mathbb{Z}$ in $\mathbb{Z}_m$ is a *unit* if there exists a coset $b + m\mathbb{Z}$ such that $(a + m\mathbb{Z})(b + m\mathbb{Z}) = 1 + m\mathbb{Z}$. This is saying that there is a solution to the equation $rx = 1$ in $\mathbb{Z}_m$. Now let $U_m \subset \mathbb{Z}_m$ denote the set of unit cosets. Then U_m contains a multiplicative identity $1 + m\mathbb{Z}$, and each element $a + m\mathbb{Z}$ of U_m is invertible. That is, there exists $b + m\mathbb{Z}$ such that $(a + m\mathbb{Z})(b + m\mathbb{Z}) = 1 + m\mathbb{Z}$. To finish showing that U_m is a group, in fact an abelian group, we must prove the following assertion.

Claim 1: *The product of two units is a unit.*

Proof. Let $a + m\mathbb{Z}$ and $b + m\mathbb{Z}$ be units with inverses $c + m\mathbb{Z}$ and $d + m\mathbb{Z}$ respectively. Then, omitting parentheses, we have

$$(a + m\mathbb{Z})(b + m\mathbb{Z})(c + m\mathbb{Z})(d + m\mathbb{Z}) = (a + m\mathbb{Z})(c + m\mathbb{Z})(b + m\mathbb{Z})(d + m\mathbb{Z})$$

$$= (1 + m\mathbb{Z})(1 + m\mathbb{Z}) = 1 + m\mathbb{Z}.$$

Consequently, U_m is a group.

Claim 2: $|U_m| = \phi(m)$.

Proof. Every element of U_m is a coset $a + m\mathbb{Z}$ represented by an integer a such that $0 < a < m$. It will suffice to show that $a + m\mathbb{Z}$ is a unit if and only if $(a, m) = 1$. Suppose first that $a + m\mathbb{Z}$ is a unit. Then there exists an integer b such that $(a + m\mathbb{Z})(b + m\mathbb{Z}) = 1 + m\mathbb{Z}$. Thus $ab - 1 = mn$ for some $n \in \mathbb{Z}$, so $ab - mn = 1$. We conclude that $(a, m) = 1$. Conversely, if $(a, m) = 1$, then there exist integers b, c such that $ab + cm = 1$, from which it follows that $(a + m\mathbb{Z})(b + m\mathbb{Z}) = 1 + m\mathbb{Z}$. Thus, $a + m\mathbb{Z} \in U_m$, finishing the proof of Claim 2.

To prove Euler's theorem, we must show that for every integer a such that $(a, m) = 1$, $a^{\phi(m)} \equiv 1 \bmod(m)$. Let $\mathbf{a}$ denote $a + m\mathbb{Z}$. Then $\mathbf{a} \in U_m$, so by Lagrange's theorem, $\mathbf{a}^{\phi(m)} = a^{\phi(m)} + m\mathbb{Z} = 1 + m\mathbb{Z}$. This is equivalent to the conclusion of the theorem, so we are done. □

The groups of units U_m are interesting in themselves, because they give nontrivial examples of finite abelian groups. Let us consider a couple of examples.

Example 2.18. Let us calculate U_m for $m = 8$ and 10. As above, let $\mathbf{a}$ denote $a + m\mathbb{Z}$. Then $U_8 = \{\mathbf{1},\ \mathbf{3},\ \mathbf{5},\ \mathbf{7}\}$. A simple check shows that every element of U_8 has order two. For example, $\mathbf{3}^2 = \mathbf{9} = \mathbf{1}$. In particular, U_8 cannot be cyclic. On the other hand, $U_{10} = \{\mathbf{1},\ \mathbf{3},\ \mathbf{7},\ \mathbf{9}\}$ has at least one element of order 4 (find one!). Consequently, U_{10} is cyclic.

2.3.4 The First Isomorphism Theorem

Finally, let us prove one of the most fundamental results in group theory.

Theorem 2.20. *(The first isomorphism theorem) Let $\varphi : G \to G'$ be a surjective group homomorphism, and let $H = \ker(\varphi)$. Then φ induces an isomorphism $\Phi : G/H \to G'$ for which $\Phi(gH) = \varphi(g)$.*

Proof. Since $H = \ker(\varphi)$, Proposition 2.16 implies that H is a normal subgroup. To see that Φ is well defined, we must show that its definition is independent of the representative g of gH. But all representatives have the form gh for some $h \in H$, while $\varphi(gh) = \varphi(g)\varphi(h) = \varphi(g)$, so Φ is well defined. Suppose $\Phi(gH) = \Phi(kH)$. Then $\varphi(g) = \varphi(k)$. Since $\varphi(k^{-1}) = \varphi(k)^{-1}$ (due to the fact that $\varphi(k)\varphi(k^{-1}) = \varphi(kk^{-1}) = 1$), we thus have $\varphi(k^{-1}g) = \varphi(k^{-1})$ $\varphi(g) = 1$. Therefore, $k^{-1}g \in H$, and so $kH = gH$. Therefore, Φ is injective. It is clearly surjective, so Φ is an isomorphism. □

There is a second isomorphism theorem, but it is not as useful, and we will not mention it here.

Exercises

Exercise 2.3.1. Show that if H is a subgroup of G, then $N_G(H)$ is a subgroup of G and H is a normal subgroup of $N_G(H)$.

Exercise 2.3.2. Show that $S(3)$ has both normal subgroups and subgroups that aren't normal by showing that the subgroup $H_1 = \{e, \sigma_1\sigma_2, \sigma_2\sigma_1\}$ is normal, while the subgroup $H_2 = \{e, \sigma_1\}$ is not normal.

Exercise 2.3.3. The center of a group G is defined as $Z(G) = \{g \in G \mid gh = hg \ \ \forall \ \ h \in G\}$. Show that $Z(G)$ is a normal subgroup.

Exercise 2.3.4. Show that the center of $S(3)$ is the trivial subgroup.

Exercise 2.3.5. Show that $S(3)$ contains a proper nontrivial normal subgroup H.

Exercise 2.3.6. Is U_{12} cyclic? What about U_{16}?

Exercise 2.3.7. A *primitive element* of a cyclic group G is an element $x \in G$ such that $G = \langle x \rangle$. Let G be a cyclic group of order $m > 1$. The purpose of this exercise is to study the number of primitive elements of G. First apply the fundamental theorem of arithmetic (Theorem 2.27) to factor $m = p_1^{a_1} \cdots p_k^{a_k}$, where $p_1, \ldots, p_k$ are the prime factors of m and $a_i \geq 1$ for all i.

(*i*) Show that the number of primitive elements of G is $\phi(m)$, where ϕ is Euler's ϕ-function.

(*ii*) Next, prove that for every prime p and integer $a > 0$,

$$\phi(p^a) = p^a - p^{a-1} = p^a(1 - 1/p).$$

(Just stare hard at $1, 2, \ldots, p^a$.)

(*iii*) Now show that if m and n are relatively prime positive integers, then $\phi(mn) = \phi(m)\phi(n)$. (This is elementary.)

(*iv*) Finally, conclude that the number of primitive elements of G is

$$\phi(m) = m \prod_{i=1}^{k} (1 - 1/p_i).$$

This is a well-known expression for the Euler function.

Exercise 2.3.8. Define a mapping $\varphi : \mathbb{Z}_4 \to \mathbb{Z}_{10}$ by $\varphi(x) = 5x$. That is, for every $a \in \mathbb{Z}$, $\varphi(a + 4\mathbb{Z}) = (5a + 10\mathbb{Z})$.

(*i*) Show that φ is a well-defined homomorphism.

(*ii*) Find the kernel and image of φ.

Exercise 2.3.9. Prove that if G is a finite group such that $G/Z(G)$ is cyclic, then G is abelian. (Recall that $Z(G)$ is the center of G.)

2.4 Fields

We now come to the second fundamental topic in this introduction to abstract algebra, the notion of a field. Roughly speaking, a field is a set $\mathbb{F}$ with two binary operations, addition and multiplication. The first requirement is that $\mathbb{F}$ be an abelian group under addition; the second is that if 0 denotes the additive identity of $\mathbb{F}$, then $\mathbb{F}^* = \mathbb{F} - \{0\}$ is an abelian group under multiplication. In addition, addition and multiplication are related by the distributive laws. The reader is undoubtedly already familiar with examples of fields. For example, the rational numbers, $\mathbb{Q}$, consisting of all quotients m/n with $m, n \in \mathbb{Z}$ and $n \neq 0$, form a field, as do the real numbers $\mathbb{R}$. A third very important field is the complex numbers $\mathbb{C}$, which allow one to solve an equation like $x^2 + 1 = 0$ by adjoining the imaginary numbers to $\mathbb{R}$. The complex numbers are an indispensable tool for physicists, chemists, and engineers as well as mathematicians. We will also consider another class of fields called Galois fields. Galois fields are finite: as we will show, the rings $\mathbb{Z}_p$, where p is a prime, form a class of Galois fields called the prime fields. The simplest prime field is $\mathbb{Z}_2$, in which 0 represents off, 1 represents on, and addition by 1 changes off to on and on to off. Galois fields are a tool of coding theory and computer science.

2.4.1 The definition of a field

The definition of a field is long but not difficult, due to the fact that every condition in the definition is a familiar arithmetic property of the real numbers.

Definition 2.12. A *field* is a set $\mathbb{F}$ that has two binary operations $F_+(a, b) = a + b$ and $F.(a, b) = ab$. These operations are called addition and multiplication respectively. They are required to satisfy the following conditions:

(*i*) $\mathbb{F}$ is an abelian group under addition.

(*ii*) Let $0 \in \mathbb{F}$ be the additive identity, and put

$$\mathbb{F}^* = \{x \in \mathbb{F} \mid x \neq 0\}.$$

Then $\mathbb{F}^*$ is an abelian group under multiplication. In particular, $\mathbb{F}^*$ contains a multiplicative identity 1, and $0 \neq 1$.

(*iii*) The distributive law holds: $a(b + c) = ab + ac$ for all $a, b, c \in \mathbb{F}$.

The associativity, commutativity, and distributivity of addition and multiplication are called the arithmetic properties of $\mathbb{F}$. One of the consequences of these properties is that the identity

$$a0 = 0 \tag{2.2}$$

holds for all $a \in \mathbb{F}$. To see this, use the distributive law to infer $a0 = a(0+0) = a0 + a0$. Then $a0 + (-a0) = (a0 + a0) + (-a0)$. Now applying associativity of addition gives the identity, since

$$0 = a0 + (a0 + (-a0)) = a0 + 0 = a0.$$

In the above definition, we referred to *the* additive and multiplicative identities as well as to *the* additive and multiplicative inverses. The uniqueness follows from the fact that identities and inverses in groups are unique.

From now on, the additive inverse of $a \in \mathbb{F}$ will be denoted by $-a$, and its multiplicative inverse will be denoted by a^{-1}, provided $a \neq 0$. We next prove a fundamental and useful property that all fields possess.

Proposition 2.21. *Let $\mathbb{F}$ be a field, and suppose $a, b \in \mathbb{F}$ satisfy $ab = 0$. Then either $a = 0$ or $b = 0$. Put another way, if neither a nor b is zero, then $ab \neq 0$.*

Proof. Suppose $ab = 0$ but $a \neq 0$. Then

$$0 = a^{-1}0 = a^{-1}(ab) = (a^{-1}a)b = 1b = b.$$

Therefore, $b = 0$. □

We will consider the rational numbers $\mathbb{Q}$, the real numbers $\mathbb{R}$, and the complex numbers $\mathbb{C}$ in the next section. After that, we will construct the prime fields $\mathbb{F}_p$ having p elements, where p is a prime. To conclude this section, let us describe the simplest prime field $\mathbb{F}_2$ in more concrete terms than in the introduction.

Example 2.19. Computers use two states O and I that interact. We will call this setup $O|I$. We describe it as follows: let $O|I = \{0, 1\}$, with 0 denoting the additive identity and 1 the multiplicative identity. To make $O|I$ a field, we have to completely define the two binary operations. Multiplication is forced on us by the definition of a field and (2.2). That is, $0 \cdot 0 = 0$, $0 \cdot 1 = 0$, and $1 \cdot 1 = 1$. (Here we are denoting multiplication ab by $a \cdot b$ for clarity.) Furthermore, addition by 0 is also already determined. Thus it remains to define $1 + 1$. But if one puts $1 + 1 = 1$, then necessarily $0 = 1$, so we are forced to set $1 + 1 = 0$. With this stipulation, 1 is its own additive inverse. Thus addition corresponds to the operation of changing the state from 0 to 1 and 1 to 0. When $p > 2$, however, multiplication is nontrivial. We leave it to the reader to show that in fact, $O|I$ satisfies all the field axioms (see Exercise 2.4.2). ⊠

Finally, we make a useful though obvious definition.

Definition 2.13. Let $\mathbb{F}$ be a field. A subset $\mathbb{F}'$ of $\mathbb{F}$ is said to be a *subfield of* $\mathbb{F}$ if $\mathbb{F}'$ is a field under the addition and multiplication of $\mathbb{F}$.

2.4.2 Arbitrary sums and products

Just as for groups, one frequently has to express the sums and products of more than three elements in a field. For $a, b, c, d \in \mathbb{F}$, we can put $a + b + c = a + (b + c)$ and then put $a + b + c + d = a + (b + c + d)$. But is it the case that $a + b + c + d = ((a + b) + c) + d$? The answer is yes. Moreover, $a + b + c + d = (a + b) + (c + d)$ too. For, if $e = a + b$, then $(a + b) + (c + d) = e + (c + d) = (e + c) + d = ((a + b) + c) + d$. In fact, no matter how one associates the terms a, b, c, d, their sum will always have the same value.

More generally, defining the sum of any n elements $a_1, \ldots, a_n \in \mathbb{F}$ inductively by

$$a_1 + a_2 + \cdots + a_n = (a_1 + a_2 + \cdots + a_{n-1}) + a_n$$

as in Section 2.1.2 allows us to ignore parentheses. Moreover, if $a'_1, \ldots, a'_n$ are the same elements as $a_1, \ldots, a_n$, but taken in a different order, then

$$\sum_{i=1}^{n} a_i = \sum_{i=1}^{n} a'_i.$$

In other words, arbitrary finite sums of elements in a field are well defined and may be computed by associating them in any manner or rearranging them in any way. These claims can be proved via mathematical induction, though their proofs are tedious (that is, no fun). Thus we will not attempt them. The analogous results for products are also true and proved in exactly the same way.

Exercises

Exercise 2.4.1. Show that in a field, the additive identity cannot have a multiplicative inverse.

Exercise 2.4.2. Finish Example 2.19 by showing that if $O|I$ denotes the set $\{0, 1\}$ with addition $0 + 0 = 0$, $0 + 1 = 1 + 0 = 1$, $1 + 1 = 0$ and multiplication $0 \cdot 0 = 0$, $0 \cdot 1 = 1 \cdot 0 = 0$, and $1 \cdot 1 = 1$, then $O|I$ is a field.

Exercise 2.4.3. Let $a_1, \ldots, a_n$ be elements of a field $\mathbb{F}$. Let $a'_1, \ldots, a'_n$ be the same elements, but taken in a different order. Use induction to show that

$$\sum_{i=1}^{n} a_i = \sum_{i=1}^{n} a'_i$$

and

$$\prod_{i=1}^{n} a_i = \prod_{i=1}^{n} a'_i.$$

Exercise 2.4.4. Consider the set $\mathcal{Q}$ of all pairs (a, b), where $a, b \in \mathbb{Z}$ and $b \neq 0$. Consider two pairs (a, b) and (c, d) to be the same if $ad = bc$. Now define operations of addition and multiplication on $\mathcal{Q}$ as follows:

$$(a, b) + (c, d) = (ad + bc, bd) \text{ and } (a, b)(c, d) = (ac, bd).$$

Show that $\mathcal{Q}$ is a field. Can you identify $\mathcal{Q}$?

Exercise 2.4.5. Let $\mathbb{F} = \{a + b\sqrt{2} \mid a, b \in \mathbb{Q}\}$.

(i) Show that $\mathbb{F}$ is a subfield of the real numbers $\mathbb{R}$.

(ii) Find $(1 - \sqrt{2})^{-1}$ and $(3 - 4\sqrt{2})^{-1}$.

2.5 The Basic Number Fields $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$

We now describe the most familiar examples of fields: the rational numbers $\mathbb{Q}$, the real numbers $\mathbb{R}$, and the complex numbers $\mathbb{C}$. We will assume that the real numbers exist and will not attempt to construct them.

2.5.1 The rational numbers $\mathbb{Q}$

The set of rational numbers $\mathbb{Q}$ consists of all fractions a/b, where $a, b \in \mathbb{Z}$ and $b \neq 0$. By assumption, $b/b = 1$, provided $b \neq 0$. Addition is defined by the rule

$$a/b + c/d = (ad + bc)/bd,$$

and multiplication is defined by

$$(a/b)(c/d) = ac/bd.$$

In particular, if $c \neq 0$, then $ac/bc = a/b$. Moreover, $a/b = 0$ if and only if $a = 0$. One easily checks that $-(a/b) = -a/b$. The multiplicative inverse of $a/b \neq 0$ is b/a. That is,

$$(a/b)^{-1} = b/a,$$

provided $a, b \neq 0$.

That the rationals satisfy all the field axioms follows from the arithmetic properties of the integers. Of course, $\mathbb{Z}$ is not a field, since the only nonzero integers with multiplicative inverses are ± 1. In fact, one can show that every field containing $\mathbb{Z}$ also contains a subfield indistinguishable from $\mathbb{Q}$. In other words, $\mathbb{Q}$ is the smallest field containing $\mathbb{Z}$. The integers form a structure known as a *ring*. In other words, a field is a ring in which every nonzero element has a multiplicative inverse. We have already encountered examples of rings: for example, the integers modulo a composite number (see Example 2.15). The $n \times n$ matrices over a field defined in the next chapter will give other examples of rings.

2.5.2 The real numbers $\mathbb{R}$

The construction of the real numbers involves some technical mathematics that would require a lengthy digression. Thus we will simply view $\mathbb{R}$ as the set of all decimal expansions

$$\pm a_1 a_2 \cdots a_r . b_1 b_2 \cdots ,$$

where all a_i and b_j are integers between 0 and 9 and $a_1 \neq 0$ unless r=1. Note that there can be infinitely many b_j to the right of the decimal point but only finitely many a_j to the left. It is also necessary to identify certain decimal expansions. For example, $1 = .999999\ldots$. In these terms, $\mathbb{Q}$ is the set of real numbers $\pm a_1 a_2 \cdots a_r . b_1 b_2 \cdots$ such that the decimal part $b_1 b_2 \cdots$ is either finite (that is, all $b_i = 0$ for i sufficiently large) or eventually repeats itself ad infinitum. Examples are $1 = 1.000\ldots$ or $1/3 = .333\ldots$.

The real numbers have the useful property of having an ordering; every real number x is either positive, negative, or 0, and the product of two numbers with the same sign is positive. This makes it possible to solve linear inequalities such as $a_1 x_1 + a_2 x_2 + \cdots + a_n x_n > 0$, although we will not need to treat such ideas in this text. The reals also have the *Archimedean property*: if $a, b > 0$, then there exists an $x > 0$ such that $ax > b$. In other words, inequalities in $\mathbb{R}$ always have solutions.

2.5.3 The complex numbers $\mathbb{C}$

The set of complex numbers $\mathbb{C}$ is a field containing $\mathbb{R}$ and the square roots of negative real numbers. Put another way, if a is a positive real number, then the equation $x^2 + a = 0$ has two complex roots. It is therefore possible to give a meaning to the square root of a negative number such as $\sqrt{-a}$. These numbers are said to be *imaginary*.

The definition of $\mathbb{C}$ starts with $\mathbb{R}^2$, namely the set of all ordered pairs (a, b) of real numbers a and b with the usual componentwise addition:

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

The interesting feature is the definition of multiplication:

$$(a, b)(c, d) = (ac - bd, ad + bc). \tag{2.3}$$

Then we have the following proposition.

Proposition 2.22. *The set $\mathbb{R}^2$ with addition and multiplication defined as above is a field. The zero element 0 is $(0, 0)$ and the multiplicative identity 1 is $(1, 0)$. The additive inverse of (a, b) is*

$$-(a, b) = (-a, -b),$$

and the multiplicative inverse of $(a, b) \neq (0, 0)$ is

$$(a,b)^{-1} = \frac{1}{a^2+b^2}(a,-b).$$

The proof is a straightforward calculation and will be omitted. We will now simplify the notation by identifying the pair $(a,0)$ with the real number a. (Note: in fact, the mapping $\mathbb{R} \to \mathbb{R}^2$ sending $a \to (a,0)$ is an injective field homomorphism.) Since $(a,0)(r,s) = (ar,as)$, we obtain after this identification that

$$a(r,s) = (ar,as).$$

This operation is the usual *scalar multiplication* on $\mathbb{R}^2$. Later will we will say that $\mathbb{C}$ is a vector space over $\mathbb{R}$.

Next, denote $(0,1)$ by i. Then (a,b) may be written

$$(a,b) = a(1,0) + b(0,1) = a + bi.$$

Addition and multiplication are then given by the rules

$$(a+bi)+(c+di) = (a+c)+(b+d)i \text{ and } (a+bi)(c+di) = (ac-bd)+(ad+bc)i.$$

In particular, $i^2 = -1$, and if $a + bi \neq 0$, then

$$(a+bi)^{-1} = \frac{a-bi}{a^2+b^2}.$$

We now make the following definition.

Definition 2.14. We will call $a + bi$ as defined above a *complex number* and let $\mathbb{C}$ denote the set of all complex numbers $a + bi$ with a and b arbitrary real numbers.

Summarizing the above discussion, we have the following assertion.

Proposition 2.23. *The set of complex numbers* $\mathbb{C} = \{a + bi \mid a, b \in \mathbb{R}\}$ *with the definitions of addition and multiplication given above is a field containing* $\mathbb{R}$ *as a subfield in which every element* $a \in \mathbb{R}$ *different from zero has two distinct square roots* $\pm\sqrt{a}$.

The first component a of $a + bi$ is called its *real part*, and the second component b is called its *imaginary part*. The points of $\mathbb{C}$ with imaginary part zero are called *real points*, and the points of the form bi with real part zero are called *imaginary points*.

2.5.4 The geometry of $\mathbb{C}$

We now make some definitions for complex numbers that lead to some beautiful connections with the geometry of $\mathbb{R}^2$. First of all, the *complex conjugate* ζ of $\zeta = a + ib$ is defined by $\zeta = a - ib$. It is straightforward to check the following identities:

$$\overline{\omega + \zeta} = \overline{\omega} + \overline{\zeta} \tag{2.4}$$

and

$$\overline{\omega\zeta} = \overline{\omega}\ \overline{\zeta}. \tag{2.5}$$

The real numbers are obviously the numbers $\zeta \in \mathbb{C}$ for which $\zeta = \overline{\zeta}$. Geometrically speaking, complex conjugation is a mapping with domain and target $\mathbb{R}^2$ sending a point to its reflection through the real axis.

The *length* of the complex number $\zeta = a + ib$ is defined as the length of the point $(a, b) \in \mathbb{R}^2$. That is, $|\zeta| = (a^2 + b^2)^{1/2}$. One calls $|\zeta|$ the *modulus* of ζ. Since $\zeta\overline{\zeta} = (a + bi)(a - bi) = a^2 + b^2$,

$$|\zeta| = (\zeta\overline{\zeta})^{1/2}.$$

Applying this to the formula for an inverse, we get the lovely formula

$$\zeta^{-1} = \frac{\overline{\zeta}}{|\zeta|^2},$$

if $\zeta \neq 0$. This gives a nice geometric fact about inversion: if $|\zeta| = 1$, then $\zeta^{-1} = \overline{\zeta}$. In other words, the inverse of a complex number of modulus one is its reflection through the real axis.

The complex numbers $\zeta = x + yi$ of unit length are the points of $\mathbb{R}^2$ on the unit circle

$$S^1 = \{(x, y) \mid x^2 + y^2 = 1\}.$$

Since every point of S^1 can be expressed in the form $(\cos\theta, \sin\theta)$ for a unique angle θ such that $0 \leq \theta < 2\pi$, we can parameterize the unit circle by introducing the *complex exponential* function

$$e^{i\theta} := \cos\theta + i\sin\theta. \tag{2.6}$$

The following proposition uses this observation.

Proposition 2.24. *Every $\zeta \in \mathbb{C}$ can be represented as $\zeta = |\zeta|e^{i\theta}$ for some $\theta \in \mathbb{R}$. Two values of θ parameterize the same point of $\mathbb{C}$ if and only if their difference is a multiple of 2π.*

The unique value of θ in $[0, 2\pi)$ such that $\zeta = |\zeta|e^{i\theta}$ is called the *argument* of ζ. A key property of the complex exponential is the identity

$$e^{i(\theta+\mu)} = e^{i\theta}e^{i\mu}, \tag{2.7}$$

which follows from the trigonometric formulas for the sine and cosine of the sum of two angles. (We will give a simple geometric proof of this identity using rotations in the plane.) The identity (2.7) can be interpreted group-theoretically too.

Proposition 2.25. *The complex exponential defines a homomorphism π from the additive group $\mathbb{R}$ to the multiplicative group $\mathbb{C}^*$ of nonzero complex numbers. The image $\pi(\mathbb{R})$ is the subgroup S^1 of $\mathbb{C}^*$ consisting of the complex numbers of length one.*

The identity (2.7) implies De Moivre's identity: for every integer $n > 0$,

$$(\cos\theta + i\sin\theta)^n = \cos n\theta + i\sin n\theta.$$

It also gives a geometric interpretation of complex multiplication: if $\zeta = |\zeta|e^{i\theta}$ and $\omega = |\omega|e^{i\mu}$, then

$$\omega\zeta = (|\omega|e^{i\mu})(|\zeta|e^{i\theta}) = (|\omega||\zeta|)e^{i(\mu+\theta)}. \tag{2.8}$$

In other words, the product $\omega\zeta$ is obtained by multiplying the lengths of ω and ζ and adding their arguments (modulo 2π).

Example 2.20 (The mth roots of unity)**.** Suppose m is a positive integer and put $\theta = 2\pi/m$. Let

$$C_m = \{1, e^{i\theta}, \ldots, e^{i(m-1)\theta}\}.$$

The elements of C_m are distinct solutions of the polynomial equation $z^m - 1 = 0$. We call C_m the set of mth *roots of unity*. We leave it as an exercise to show that C_m is a cyclic subgroup of the multiplicative subgroup $\mathbb{C}^*$ of $\mathbb{C}$ of order m. Since $z^m - 1 = 0$ has at most m solutions in $\mathbb{C}$ (see the discussion in Section 2.5.5), C_m gives all mth roots of unity. The points of C_m are the m equally spaced points on the unit circle S^1 including 1. ⊠

Here is a surprisingly nice consequence of Proposition 2.9 (cf. Lagrange's theorem).

Proposition 2.26. *Suppose G is a finite subgroup of $\mathbb{C}^*$ of order m. Then $G = C_m$. Therefore, the only finite subgroups of $\mathbb{C}^*$ are the cyclic groups C_m of mth roots of unity.*

Proof. Since G is finite of order m, it follows that for all $z \in G$, $z^m = 1$, by Lagrange. In particular, $G \subset C_m$. It follows immediately that $G = C_m$. □

2.5.5 The Fundamental Theorem of Algebra

Suppose z denotes an arbitrary element of $\mathbb{C}$ and let n be a positive integer. A function $f : \mathbb{C} \to \mathbb{C}$ of the form $f(z) = \alpha_0 z^n + \alpha_1 z^{n-1} + \cdots + \alpha_{n-1} z + \alpha_n$, where the coefficients $\alpha_0, \alpha_1, \ldots, \alpha_n \in \mathbb{C}$, is called a *complex polynomial*. If $\alpha_0 \neq 0$, we say that f has *degree* n. The roots of f are the elements $\zeta \in \mathbb{C}$ such that $f(\zeta) = 0$. As noted earlier, the polynomial $z^2 + 1$ has real coefficients but no real roots, However, $\zeta = \pm i$ are a pair of roots in $\mathbb{C}$. Consequently, $z^2 + 1 = (z + i)(z - i)$. The fundamental theorem of algebra is the following remarkable generalization of this example.

Theorem 2.27. *Every complex polynomial*

$$f(z) = z^n + \alpha_1 z^{n-1} + \cdots + \alpha_{n-1} z + \alpha_n$$

with $n > 0$ has a complex root. That is, there exists $\zeta \in \mathbb{C}$ such that $f(\zeta) = 0$.

The fundamental theorem of algebra poses a conundrum to algebraists: the only known algebraic proof is long and complicated. The best proofs, in the sense of explaining why in particular the theorem is true, come from complex analysis and topology. The proof using complex analysis is extremely elementary and elegant, so it is invariably included in complex analysis courses.

It follows from division with remainders (see Proposition 2.36) that if $f(z)$ is a complex polynomial such that $f(\zeta) = 0$, then there exists a complex polynomial $g(z)$ such that

$$f(z) = (z - \zeta)g(z).$$

Applying the fundamental theorem of algebra to $g(z)$ and so forth, it follows that there are not necessarily distinct $\zeta_1, \ldots, \zeta_n \in \mathbb{C}$ such that

$$f(z) = (z - \zeta_1)(z - \zeta_2) \cdots (z - \zeta_n).$$

Thus every complex polynomial $f(z)$ can be expressed as a product of linear polynomials. A field $\mathbb{F}$ that has the property that every polynomial with coefficients in $\mathbb{F}$ of positive degree has a root in $\mathbb{F}$ is said to be *algebraically closed*. For example, $\mathbb{C}$ is algebraically closed, but $\mathbb{R}$ isn't. A fundamental result in algebra, which is well beyond the scope of this discussion, says that every field is a subfield of an algebraically closed field.

Exercises

Exercise 2.5.1. Express the following complex numbers in the form $a + ib$:

$$(i)\ \frac{2-3i}{1+2i},\ \text{and}\ (ii)\ \frac{(2+i)(3-2i)}{4-2i}.$$

Exercise 2.5.2. Find the inverse of

$$\frac{(2+i)(3-2i)}{4-2i}$$

without explicitly computing the fraction.

Exercise 2.5.3. Express all solutions of the equation $x^3 + 8 = 0$ in the form $re^{i\theta}$ and graph them as elements of $\mathbb{C} = \mathbb{R}^2$.

Exercise 2.5.4. Find $\alpha_1, \ldots, \alpha_4$ such that

$$x^4 - 1 = (x - \alpha_1)(x - \alpha_2)(x - \alpha_3)(x - \alpha_4).$$

Exercise 2.5.5. Find all $(x_1, x_2, x_3) \in \mathbb{C}^3$ satisfying the equations

$$\begin{aligned} ix_1 + 2x_2 + (1-i)x_3 &= 0, \\ -x_1 + ix_2 - (2+i)x_3 &= 0. \end{aligned}$$

Exercise 2.5.6. If necessary, look up formulas for $\cos(\theta + \mu)$ and $\sin(\theta + \mu)$, and use them to prove formula De Moivre's identity. That is, show that $e^{i(\theta+\mu)} = e^{i\theta}e^{i\mu}$.

Exercise 2.5.7. Suppose $p(x)$ is a polynomial with real coefficients. Show that all the roots of $p(x) = 0$ occur in conjugate pairs $\lambda, \overline{\lambda} \in \mathbb{C}$. Conclude that a real polynomial of odd degree has a real root.

Exercise 2.5.8. An nth root of unity α is called *primitive* if for every nth root of unity β, there is an integer $r \geq 0$ such that $\beta = \alpha^r$. Prove that for every $n > 0$, there exists a primitive nth root of unity.

Exercise 2.5.9. Let a, b, c, d be arbitrary integers. Show that there exist integers m and n such that $(a^2 + b^2)(c^2 + d^2) = m^2 + n^2$.

2.6 Galois fields

A field $\mathbb{F}$ which is finite is called a *Galois field.* We have already encountered an example of a Galois field in Example 2.19, namely the two-element field $O|I$, or $\mathbb{F}_2$ as it is usually called. In this section, we will construct a Galois field with a prime number of elements for every prime.

2.6.1 The prime fields $\mathbb{F}_p$

Let p denote an arbitrary prime. Our goal is to construct a field $\mathbb{F}_p$, called a *prime field*, having exactly p elements. We have already defined a field with two elements, and since every field has to contain at least two elements, (namely the additive and multiplicative identities), $\mathbb{F}_2$ is the smallest field. Let us now define $\mathbb{F}_p$. Putting $\mathbb{F}_p = \mathbb{Z}_p = \mathbb{Z}/p\mathbb{Z}$ gives us an additive abelian group with p elements. It remains to define a multiplication on $\mathbb{F}_p$ such that $\mathbb{F}_p - \{0\}$ is an abelian group. Given two cosets $\mathbf{a} = a + p\mathbb{Z}$ and $\mathbf{b} = b + p\mathbb{Z}$, define

$$\mathbf{ab} = (a + p\mathbb{Z})(b + p\mathbb{Z}) = ab + p\mathbb{Z}.$$

Since

$$(a + mp)(b + np) = ab + p(an + bm + mnp),$$

this coset multiplication is well defined. It is immediate that the coset $1 + p\mathbb{Z}$ is a multiplicative identity. Associativity of multiplication and the distributive law follow easily from the arithmetic of $\mathbb{Z}$.

Now recall the Euler group $U_p \subset \mathbb{Z}/p\mathbb{Z} = \mathbb{F}_p$ of units modulo p. We showed that U_p is an abelian group whose order is $\phi(p) = p - 1$. This says that in fact, $\mathbb{F}_p - \{0\} = U_p$. Therefore, $\mathbb{F}_p - \{0\}$ is an abelian group, and hence we have proved the following theorem.

Theorem 2.28. *If p is a prime, then $\mathbb{F}_p$ is a field.*

There is another proof that doesn't use Euler groups, which although longer, is also instructive. First, one notes that a property of fields we already proved holds for $\mathbb{F}_p$. **Claim.** If p be a prime number and $\mathbf{ab} = \mathbf{0}$ in $\mathbb{F}_p$, then either $\mathbf{a} = \mathbf{0}$ or $\mathbf{b} = \mathbf{0}$ (or both). For since $\mathbf{ab} = \mathbf{0}$ in $\mathbb{F}_p$ is the same thing as saying that p divides the product ab in $\mathbb{Z}$, the claim follows from the fact that if the prime number p divides ab, then either p divides a or p divides b. This is immediate from the following theorem.

Theorem 2.29 (Fundamental theorem of arithmetic)**.** *Every integer $m > 1$ can be factored $m = p_1 p_2 \cdots p_k$, where $p_1, p_2, \ldots, p_k$ are not necessarily distinct primes. Moreover, the number of times each prime p_i occurs in this factorization is unique.*

To show that every nonzero element in $\mathbb{F}_p$ has an inverse, we will show that multiplication by any $\mathbf{a} \in \mathbb{F}_p - 0$ induces an injective map

$$\phi_{\mathbf{a}} : \mathbb{F}_p - 0 \longrightarrow \mathbb{F}_p - 0$$

defined by $\phi_{\mathbf{a}}(\mathbf{x}) = \mathbf{ax}$. To see that $\phi_{\mathbf{a}}$ is indeed injective, let $\phi_{\mathbf{a}}(\mathbf{x}) = \phi_{\mathbf{a}}(\mathbf{y})$, that is, $\mathbf{ax} = \mathbf{ay}$. Then $\mathbf{a}(\mathbf{x} - \mathbf{y}) = \mathbf{0}$, so $\mathbf{x} - \mathbf{y} = \mathbf{0}$, since $\mathbf{a} \neq \mathbf{0}$ (Proposition ??). Therefore, $\phi_{\mathbf{a}}$ is indeed injective. Since $\mathbb{F}_p - 0$ is finite, the pigeonhole principle says that $\phi_{\mathbf{a}}$ is a bijection. In particular, there exists an $\mathbf{x} \in \mathbb{F}_p - 0$ such that $\phi_{\mathbf{a}}(\mathbf{x}) = \mathbf{ax} = \mathbf{1}$. Hence $\mathbf{x}$ is the required inverse of $\mathbf{a}$. This completes the second proof that $\mathbb{F}_p$ is a field. □

Here is an explicit example. As we did earlier, we will use a dot to denote multiplication.

Example 2.21 Consider $\mathbb{F}_3 = \{\mathbf{0}, \mathbf{1}, \mathbf{2}\}$. Addition is given by $\mathbf{1} + \mathbf{1} = \mathbf{2}$, $\mathbf{1} + \mathbf{2} = \mathbf{0}$, and $\mathbf{2} + \mathbf{2} = \mathbf{1}$, the latter sum because $2 + 2 = 4$ in $\mathbb{Z}$, and 4 is in the coset $1 + 3\mathbb{Z}$. Finding products is similar. For example, $\mathbf{2} \cdot \mathbf{2} = \mathbf{4} = \mathbf{1}$. Thus $\mathbf{2^{-1}} = \mathbf{2}$ in $\mathbb{F}_3$. A good way to picture addition and multiplication is to construct tables. The addition table for $\mathbb{F}_3$ is

$+$	**0**	**1**	**2**
0	**0**	**0**	**0**
1	**1**	**2**	**0**
2	**2**	**0**	**1**

.

Similarly, the multiplication table is

$\cdot$	**0**	**1**	**2**
0	**0**	**0**	**0**
1	**0**	**1**	**2**
2	**0**	**2**	**1**

.

⊠

2.6.2 A four-element field

We will now give an example of a Galois field that is not an $\mathbb{F}_p$ by constructing a field $\mathbb{F}$ with four elements. We will define addition and multiplication by tables and leave it to the reader to verify that $\mathbb{F}$ is a field. Let $0, 1, \alpha$, and β denote the elements of $\mathbb{F}$, with 0 and 1 the usual identities. Ignoring addition by 0, the addition table is defined as follows:

$+$	1	α	β
1	0	β	α
α	β	0	1
β	α	1	0

.

The multiplication table (omitting the obvious cases 0 and 1) is

$\cdot$	α	β
α	β	1
β	1	α

.

Proposition 2.30. *The set $\mathbb{F} = \{0, 1, \alpha, \beta\}$ having 0 and 1 as identities and addition and multiplication defined as above is a field. Moreover, $\mathbb{F}_2$ is a subfield of $\mathbb{F}$.*

Notice that since $\alpha^2 = \beta$ and $\beta^2 = \alpha$, it follows that $\alpha^3 = \beta^3 = 1$. This isn't surprising, since $\mathbb{F}^*$ has order three and hence must be cyclic. Since $\alpha^4 = \alpha$ and $\beta^4 = \beta$, all elements of $\mathbb{F}$ satisfy the equation $x^4 - x = 0$, since 0 and 1 trivially do. Using the definition of a polynomial over an arbitrary field given in Section 2.6.4, we can view $x^4 - x$ as a polynomial in the variable x over the subfield $\mathbb{F}_2$ of $\mathbb{F}$, where we have the identity $x^4 - x = x^4 + x$, since $1 = -1$. Thus,

$$x^4 - x = x(x + 1)(x^2 + x + 1).$$

(Recall that $1 + 1 = 0$, so $2x = 2x^2 = 0$.) It can be verified directly from the tables that the elements α and β are the two roots of $x^2 + x + 1 = 0$. Since $x^4 - x$ has distinct roots (by the multiple root test in Section 2.6.4), we have shown that all the elements of $\mathbb{F}$ are roots of a polynomial over a subfield of $\mathbb{F}$, namely $x^4 - x$.

We will eventually show, using a theorem about finite-dimensional vector spaces, that the number of elements in a Galois field $\mathbb{F}$ is always a power p^n of a prime p. This prime is called the characteristic of the field, the topic of the next section. The integer n turns out to be interpreted as the dimension of $\mathbb{F}$ as a vector space over $\mathbb{F}_p$. It is a fundamental result in the theory of fields that for every prime p and integer $n > 0$, there exists a Galois field with p^n elements, and two Galois fields $\mathbb{F}$ and $\mathbb{F}'$ with the same number of elements are isomorphic.

2.6.3 The characteristic of a field

If $\mathbb{F}$ is a Galois field, then some multiple $r1$ of the identity $1 \in \mathbb{F}$ has to be 0. (Note: by $r1$, we mean $1 + \cdots + 1$ with r summands.) The reason for this is

that since $\mathbb{F}$ is finite, the multiples $r1$ of 1 cannot all be distinct. Hence there have to be two distinct positive integers m and n such that $m1 = n1$ in $\mathbb{F}$. This implies $m1 - n1 = 0$. But by associativity of addition, $m1 - n1 = (m - n)1$. Assuming without loss of generality that $m > n$, it follows that there exists a positive integer r such that $r1 = 0$ in $\mathbb{F}$.

I claim that the least positive integer r such that $r1 = 0$ is a prime. For if r can be expressed as a product $r = ab$, where a, b are positive integers, then $r1 = (ab)1 = (a1)(b1) = 0$. Thus, by Proposition 2.21, $a1 = 0$ or $b1 = 0$. But by the minimality of r, one of a and b is r, so r is a prime, say $r = p$. The prime p is the *characteristic of* $\mathbb{F}$. In general, one makes the following definition:

Definition 2.15. Let $\mathbb{F}$ be an arbitrary field. If some nonzero multiple $q1$ of 1 equals 0, we say that $\mathbb{F}$ has *positive characteristic*. In that case, the *characteristic of* $\mathbb{F}$ is defined to be the smallest positive integer q such that $q1 = 0$. If all multiples $q1$ are nonzero, we say that $\mathbb{F}$ has *characteristic* 0.

Example 2.22. The characteristic of the field $\mathbb{F}_4$ defined above is two. The characteristic of $\mathbb{F}_p$ is p. ⊠

Summarizing the above discussion, we state the following proposition.

Proposition 2.31. *If a field $\mathbb{F}$ has positive characteristic, then its characteristic is a prime p, and $pa = 0$ for all $a \in \mathbb{F}$.*

Proof. We already proved that if the characteristic of $\mathbb{F}$ is positive, then it has to be a prime. If $p1 = 0$, then by the distributive law,

$$pa = a + \cdots + a = 1a + \cdots + 1a = (1 + \cdots + 1)a = (p1)a = 0a = 0$$

for all $a \in \mathbb{F}$. □

Proposition 2.32. *The characteristics of $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$ are all 0. Moreover, the characteristic of every subfield of a field of characteristic 0 is also 0.*

The notion of the characteristic has a nice application.

Proposition 2.33. *Suppose $\mathbb{F}$ is a field of characteristic $p > 0$. Then for all $a_1, \ldots, a_n \in \mathbb{F}$,*

$$(a_1 + \cdots + a_n)^p = a_1^p + \cdots + a_n^p.$$

This is an application of the binomial theorem. We will leave the proof as an exercise. A consequence of the previous proposition is that $a^p = a$ for every $a \in \mathbb{F}_p$. Thus $a^{p-1} = 1$. This also gives a formula for a^{-1} for every $a \in \mathbb{F}_p$, namely, $a^{-1} = a^{p-2}$.

Example 2.23. The formula $a^{-1} = a^{p-2}$ for the inverse of a nonzero element a of $\mathbb{F}_p$ actually isn't so easy to apply without a computer. For example, to compute the inverse of 5 in $\mathbb{F}_{23}$, one needs to find 5^{21}, which is 476837158203125. After that, one has to reduce 476837158203125 modulo 23. The result is 14. Thus $5^{-1} = 14$ in $\mathbb{F}_{23}$. But this can be seen easily without having to do a long computation, since $5 \cdot 14 = 70 = 69 + 1$. ⊠

The result $a^{p-1} = 1$ in $\mathbb{F}_p$ translates into a well-known result from elementary number theory known as Fermat's little theorem.

Proposition 2.34. *Let p be prime. Then for every integer $a \not\equiv 0 \mod p$, $a^{(p-1)} \equiv 1 \mod p$.*

We leave the proof as an exercise. Notice that Fermat's little theorem is a special case of Euler's theorem, since if p is prime, then $\phi(p) = p - 1$. Here is another interesting property of $\mathbb{F}_p$.

Proposition 2.35. *The product of all the nonzero elements of $\mathbb{F}_p$ is -1.*

Proof. This follows by noting that $\{\mathbf{2}, \mathbf{3}, \ldots, \mathbf{p-2}\}$ can be partitioned into pairs $\{\mathbf{a}, \mathbf{b}\}$ such that $\mathbf{a}^{-1} = \mathbf{b}$. We will leave the proof of this to the reader. Therefore, $\mathbf{2 \cdot 3 \cdots (p-2)} = 1$ in $\mathbb{F}_p$. The result follows by multiplying by $\mathbf{-1 = p-1}$. □

Proposition 2.35 stated in number-theoretic terms is one of the assertions of Wilson's theorem: $(q-1)! \equiv -1 \mod q$ if and only if q is prime. Wilson's theorem gives a test for determining whether a number is prime, but the problem is that implementing this test requires knowing $(q-1)!$. It turns out that Fermat's little theorem gives a much easier test, known as the Fermat primality test. In this test, one checks whether $a^{q-1} \equiv 1 \mod q$ for some "random" values of a. If the congruence fails for any value of a, then q isn't prime, while if it holds for several values, then the probability that q is a prime is very high. Knowing large primes is useful, for example, in employing RSA encryption.

2.6.4 Appendix: polynomials over a field

The purpose of this appendix is to define the polynomials over a field $\mathbb{F}$. For each integer $i > 0$, let x^i denote a symbol, and suppose that every pair of these symbols x^i and x^j can be multiplied with the result $x^i x^j = x^{i+j}$ for all $i, j > 0$. We will denote x^1 simply by x. We will put $x^0 = 1 \in \mathbb{F}$. The symbol x is sometimes called an indeterminate. Note that $x^i = x \cdots x$ (with i factors). We assume that each symbol x^i can be multiplied by an arbitrary element of $\mathbb{F}$, so that $0x^i = 0$, $1x^i = x^i$, and $(ax^i)(bx^j) = (ab)x^{i+j}$. Let $\mathbb{F}[x]$ denote the set of all expressions $f(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n$,

$a_0, a_1, \ldots, a_n \in \mathbb{F}$ and $n \geq 0$. Then $f(x)$ is called a *polynomial over* $\mathbb{F}$. We agree that two polynomials $f(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n$ and $g(x) = b_0 + a_1 x + b_2 x^2 + \cdots + b_m x^m$ are equal if $a_i = b_i$ for every value of i. If $a_n \neq 0$, we say that f has *degree* n and write $\deg(f) = n$. The value of $f(x)$ at $r \in \mathbb{F}$ is by definition $f(r) = a_0 + a_1 r + a_2 r^2 + \cdots + a_n r^n$. Thus, a polynomial $f(x)$ determines a function $f : \mathbb{F} \to \mathbb{F}$. However, if $\mathbb{F}$ is a Galois field, then there exist polynomials $f(x) \in \mathbb{F}[x]$ such that the function $f : \mathbb{F} \to \mathbb{F}$ is identically zero, so polynomials should not be thought of as the same thing as functions.

Example 2.24. For example, if $\mathbb{F} = \mathbb{F}_2$ and $f(x) = x^2 + x$, then $f(1) = f(0) = 0$. Hence the function corresponding to $f(x)$ is identically zero on $\mathbb{F}_2$. However, in the field $\mathbb{F}_4$ containing $\mathbb{F}_2$ defined in Section 2.6.2, the polynomial $f(x)$ satisfies $f(\alpha) = f(\beta) = 1$. (Recall that α and β satisfy $x^2 + x + 1 = 0$. ⊠

Addition and multiplication of polynomials are defined as follows. Addition amounts to adding together the coefficients of each corresponding power x^i of x. For example,

$$(3x^2 - 2x + 1) + (x^4 - x^3 - 3x^2 + x) = x^4 - x^3 - x + 1.$$

Multiplication of two polynomials uses the above rules and the distributive law. Thus,

$$(3x^2 - 2x + 1)(x^4 - x^3 - 2x^2 + x) = 3x^6 - 5x^5 - 7x^4 + 6x^3 - 4x^2 + x.$$

When the field $\mathbb{F}$ has characteristic zero, for example $\mathbb{F} = \mathbb{Q}$, $\mathbb{R}$, or $\mathbb{C}$, there is a more natural formulation of $\mathbb{F}[x]$ that avoids the problem encountered in $\mathbb{F}_2[x]$, where polynomials can define the zero function. In that case, let $x : \mathbb{F} \to \mathbb{F}$ denote the identity function $x(r) = r$ for all $r \in \mathbb{F}$. If $i > 0$, then x^i denotes the function $x^i(r) = r^i$. Then $\mathbb{F}[x]$ may be defined as the set of all functions $f : \mathbb{F} \to \mathbb{F}$ of the form

$$f(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n,$$

where $a_0, a_1, \ldots, a_n \in \mathbb{F}$. Addition and multiplication are defined pointwise as above.

In the remainder of this section, we are going to consider two properties of polynomials over an arbitrary field. The first is that $\mathbb{F}[x]$ admits division with remainder.

Proposition 2.36 (Division with remainder for polynomials)**.** *Suppose $f(x)$ and $g(x)$ are in $\mathbb{F}[x]$. Then there exists a unique expression*

$$f(x) = q(x)g(x) + r(x),$$

where $q(x), r(x) \in \mathbb{F}[x]$ *and* $\deg(r) < \deg(g)$. *In particular, if* $a \in \mathbb{F}$ *is a root of* f, *i.e.,* $f(a) = 0$, *then* $f(x) = (x - a)g(x)$ *for some* $g(x) \in \mathbb{F}[x]$.

Proof. This can be proved by induction on the degree of $f(x)$. We will omit the details.

The second property is a test for when a polynomial has a multiple root. We say that a is a multiple root of $p(x) \in \mathbb{F}[x]$ if there exists a polynomial $q(x) \in \mathbb{F}[x]$ such that $p(x) = q(x)(x - a)^2$. To state the second property, we need to define the *derivative* of a polynomial. Suppose $n > 0$ and $p(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$. Then the derivative of $p(x)$ is defined to be the polynomial

$$p'(x) = na_n x^{n-1} + (n-1)a_{n-1}x^{n-2} + \cdots + 2a_2 x + a_1.$$

The derivative of a constant polynomial is defined to be zero. This definition agrees with the classical formula for the derivative of a polynomial. The derivative of a sum is the sum of the derivatives, and Leibniz's formula for the derivative of a product holds: $(pq)' = p'q + pq'$.

Proposition 2.37. *Let* $p(x) \in \mathbb{F}[x]$. *Then* $p(x) = (x - a)^2 q(x)$ *for some* $q(x) \in \mathbb{F}[x]$ *if and only if* $p(a) = p'(a) = 0$.

Proof. Suppose $p(x) = (x - a)^2 q(x)$. Then it is obvious from Leibniz that $p(a) = p'(a) = 0$. Suppose, conversely, that $p(a) = p'(a) = 0$. Then $p(x) = (x - a)s(x)$, where $s(x) \in \mathbb{F}[x]$, so $p'(x) = (x - a)'s(x) + (x - a)s'(x) = s(x) + (x - a)s'(x)$. Hence, $0 = p'(a) = s(a)$. This means that $s(x) = (x - a)t(x)$ for some $t(x) \in \mathbb{F}[x]$, and thus $p(x) = (x - a)^2 q(x)$, as claimed. □

One says that $a \in \mathbb{F}$ is a *simple root* of $p(x) \in \mathbb{F}[x]$ if and only if $(x - a)$ divides $p(x)$ but $(x - a)^2$ does not. A root that is not simple is called a *multiple* root. The next result formulates Proposition 2.37 as a test for whether a root of $p(x)$ is simple.

Corollary 2.38 (The simple root test)**.** *Let* $p(x) \in \mathbb{F}[x]$. *Then* a *is a simple root if and only if* $p(a) = 0$ *but* $p'(a) \neq 0$, *and* a *is a multiple root if and only if* $p(a) = p'(a) = 0$.

Exercises

Exercise 2.6.1. Write out the addition and multiplication tables for the field $\mathbb{F}_7$. Also, indicate the location of the multiplicative inverse for each nonzero element.

Exercise 2.6.2. Find both $-(\mathbf{6} + \mathbf{6})$ and $(\mathbf{6} + \mathbf{6})^{-1}$ in $\mathbb{F}_7$.

Exercise 2.6.3. Show that $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$ all have characteristic zero.

Exercise 2.6.4. Let $\mathbb{F}$ be a field and suppose that $\mathbb{F}' \subset \mathbb{F}$ is a subfield. Show that $\mathbb{F}$ and $\mathbb{F}'$ have the same characteristic.

Exercise 2.6.5. Show that the characteristic of $\mathbb{F}_p$ is p.

Exercise 2.6.6. Let $\mathbb{F}$ and $\mathbb{F}'$ be fields. Show that every field homomorphism $\varphi : \mathbb{F} \to \mathbb{F}'$ is injective.

Exercise 2.6.7. Suppose that the field $\mathbb{F}$ contains $\mathbb{F}_p$ as a subfield. Show that the characteristic of $\mathbb{F}$ is p.

Exercise 2.6.8. Suppose that $\mathbb{F}$ is a field of characteristic $p > 0$. Show that all multiples of 1 including 0 form a subfield of $\mathbb{F}$ with p elements. (This subfield is in fact a copy of $\mathbb{F}_p$.)

Exercise 2.6.9. Prove Proposition 2.33. That is, show that if $\mathbb{F}$ is a finite field of characteristic p, then for all $a_1, \ldots, a_n \in \mathbb{F}$,

$$(a_1 + \cdots + a_n)^p = a_1^p + \cdots + a_n^p.$$

Exercise 2.6.10. Use Proposition 2.33 to show that $a^p = a$ in $\mathbb{F}_p$. Use this to deduce Fermat's little theorem.

Exercise 2.6.11. In the definition of the field $\mathbb{F}_4$ in Section 2.6.2, can we alter the definition of multiplication by putting $\alpha^2 = \beta^2 = 1$ and still get a field?

Exercise 2.6.12. Suppose $\mathbb{F}$ is a field of characteristic p. Show that if $a, b \in \mathbb{F}$ and $a^p = b^p$, then $a = b$.

Exercise 2.6.13. A field of characteristic p is said to be *perfect* if every element is a pth power. Show that every Galois field is perfect. (Hint: use the pigeonhole principle.)

Exercise 2.6.14. Use Fermat's little theorem to find $\mathbf{9}^{-1}$ in $\mathbb{F}_p$ for $p = \mathbf{11}$, $\mathbf{13}$, $\mathbf{23}$, and $\mathbf{29}$. Use these results to solve the congruence equation $9x \equiv 15 \mod p$ for the above values of p.

Exercise 2.6.15. A *primitive element* of $\mathbb{F}_p$ is an element β such that

$$\mathbb{F}_p = \{0, 1, \beta, \beta^2, \ldots, \beta^{p-2}\}.$$

It can be shown that for every prime p, $\mathbb{F}_p$ contains a primitive element. Find at least one primitive element β for $\mathbb{F}_p$ when $p = 5$, 7, and 11.

Exercise 2.6.16. Write out the addition and multiplication tables for $\mathbb{Z}_6$. Is $\mathbb{Z}_6$ a field? If not, why not?

Exercise 2.6.17. Let p be a prime. Let $\phi_p : \mathbb{Z} \to \mathbb{F}_p$ be the quotient mapping $a \mapsto \mathbf{a}$, where $\mathbf{a} = a + p\mathbb{Z}$. Show that for all $a, b \in \mathbb{Z}$,

(1) $\phi_p(a + b) = \phi_p(a) + \phi_p(b)$, and

(2) $\phi_p(ab) = \phi_p(a)\phi_p(b)$.

Thus, ϕ is a homomorphism of rings. Use these two facts to deduce that addition and multiplication in $\mathbb{F}_p$ are associative from the fact that they are associative in $\mathbb{Z}$.

Exercise 2.6.18. Using the definition of the derivative of a polynomial $f(x) \in \mathbb{F}[x]$ in Section 2.6.4, show that the product rule for differentiation holds. That is, if $f(x), g(x) \in \mathbb{F}[x]$, then $(fg)' = f'g + fg'$.

Exercise 2.6.19. Let $\mathbb{F}$ be a field of characteristic p. Show that for every $n > 0$, $x^{p^n} - x = 0$ has only simple roots in $\mathbb{F}$.

Chapter 3
Matrices

Matrix theory is deeply embedded in the foundations of algebra. The idea of a matrix is very simple, and useful examples and ideas present themselves immediately, as we shall soon see. So it is surprising that their structure turns out to be subtle. Matrices also represent abstract objects called linear mappings, which we will treat after vector spaces. The main results in the theory of linear mappings, including the Cayley–Hamilton theorem, Jordan decomposition, and Jordan canonical form, are strikingly beautiful.

More down to earth, matrices represent systems of linear equations in several variables. Such a linear system has the form $A\mathbf{x} = \mathbf{b}$, where A, $\mathbf{x}$, and $\mathbf{b}$ are all matrices, and $A\mathbf{x}$ is the product of A and $\mathbf{x}$. In this form, a linear system is a special case of the basic algebraic equation $ax = b$. Matrix algebra arises from addition and multiplication in a field, and it satisfies some of the field axioms: for example, the associative and distributive laws. However, matrices usually do not have multiplicative inverses. When they do is a topic that will be taken up in the next chapter. The first goal of this chapter is matrix algebra and the procedure known as row reduction, which amounts to replacing a matrix A by a unique matrix in what is called reduced row echelon form. The reduced row echelon form of a matrix is used to solve linear systems. It also tells us the rank of the matrix and gives the inverse of A when one exists. Prove the existence of an $LPDU$ factorization, which is how one understands how the matrix is constructed.

3.1 Introduction to matrices and matrix algebra

The purpose of this section is to introduce the notion of a matrix, give some motivation, and make the basic definitions used in matrix algebra.

J.B. Carrell, *Groups, Matrices, and Vector Spaces*,
DOI 10.1007/978-0-387-79428-0_3

3.1.1 What is a matrix?

Matrices arise from linear systems. Let $\mathbb{F}$ be a field. A system of m linear equations in n variables $x_1, \ldots, x_n$ with coefficients a_{ij} and constants b_i, all of which lie in F, is a family of equations of the form

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1, \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2, \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m. \end{aligned} \tag{3.1}$$

Concentrating on the mn coefficients a_{ij}, let us form the rectangular array

$$A = \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{23} & \ldots & a_{2n} \\ \vdots & \vdots & \ldots & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{pmatrix}. \tag{3.2}$$

This array is called a matrix.

Definition 3.1. Let $\mathbb{F}$ be a field and m and n positive integers. An $m \times n$ *matrix over* $\mathbb{F}$ is a rectangular array of the form (3.2), where all the entries a_{ij} are in $\mathbb{F}$.

Notice that the elements in the ith row all have first subscript i, and those in the jth column have second subscript j. The set of all $m \times n$ matrices over $\mathbb{F}$ will be denoted by $\mathbb{F}^{m\times n}$. In particular, those with real entries are denoted by $\mathbb{R}^{m\times n}$, and those with complex entries are denoted by $\mathbb{C}^{m\times n}$.

Now let

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \quad \text{and} \quad \mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}.$$

Then the system (3.1) will turn out to be represented by the compact equation $A\mathbf{x} = \mathbf{b}$. Here $A\mathbf{x}$ is the column matrix on the left-hand side of the equalities, and $\mathbf{b}$ is the column matrix on the right-hand side. We should think of $A\mathbf{x}$ as a matrix product. The general definition of a matrix product will be given below.

3.1.2 Matrix addition

Our immediate goal is to define the algebraic operations for matrices: addition, multiplication, and scalar multiplication. We will begin with addition and scalar multiplication.

Definition 3.2. Let $A, B \in \mathbb{F}^{m\times n}$. The *matrix sum* (or simply the *sum*) $A + B$ is defined as the matrix $C \in \mathbb{F}^{m\times n}$ such that $c_{ij} = a_{ij} + b_{ij}$ for all pairs of indices (i, j). The *scalar multiple* of A by $\alpha \in \mathbb{F}$ is the matrix αA of $\mathbb{F}^{m\times n}$ whose (i, j) entry is αa_{ij}.

Thus addition is a binary operation on $\mathbb{F}^{m\times n}$. It is clearly associative, since addition is associative in $\mathbb{F}$. The matrix $O \in \mathbb{F}^{m\times n}$ all of whose entries are zero is called the *zero matrix.* The zero matrix is the *additive identity* for $\mathbb{F}^{m\times n}$. That is, $A + O = A$ for all $A \in \mathbb{F}^{m\times n}$. The matrix $-A = (-1)A$ is an *additive inverse* of A, since $A + (-A) = (-A) + A = O$. Thus we have the following result.

Proposition 3.1. *$\mathbb{F}^{m\times n}$ is an abelian group under matrix addition.*

Example 3.1. Here are some examples with $\mathbb{F} = \mathbb{R}$:

$$A = \begin{pmatrix} 1 & 0 & 0 & 2 \\ 0 & 1 & 0 & 3 \\ 0 & 0 & 1 & 5 \end{pmatrix}, \quad \text{and} \quad B = \begin{pmatrix} 1 & 2 & 3 & 0 \\ 0 & 0 & 3 & 1 \\ 1 & 2 & 3 & 0 \end{pmatrix}.$$

Then

$$A + B = \begin{pmatrix} 2 & 2 & 3 & 2 \\ 0 & 1 & 3 & 4 \\ 1 & 2 & 4 & 5 \end{pmatrix}.$$

Doubling A, that is multiplying A by the scalar 2, gives

$$2A = \begin{pmatrix} 2 & 0 & 0 & 4 \\ 0 & 2 & 0 & 6 \\ 0 & 0 & 2 & 10 \end{pmatrix}.$$

⊠

A matrix in $\mathbb{F}^{1\times n}$ is called a *row vector.* Similarly, a matrix in $\mathbb{F}^{n\times 1}$ is called a *column vector.* As long as the context is clear, we will often use $\mathbb{F}^n$ to denote either column vectors $\mathbb{F}^{n\times 1}$ or row vectors $\mathbb{F}^{1\times n}$. Vectors will be written as boldface letters like $\mathbf{x}$ and their components by the same letter in ordinary type. Thus, the ith component of $\mathbf{x}$ is x_i. To conserve space, we may write a column vector

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = (x_1, x_2, \ldots, x_n)^T. \tag{3.3}$$

The superscript T, which is called the transpose operator, changes rows into columns. A full discussion of the transpose operator is given below in Section 3.1.6.

Several matrices $A_1, \ldots, A_m$ in $\mathbb{F}^{m\times n}$ can be combined using addition and scalar multiplication. The result is called a *linear combination*. That is, given scalars $a_1, a_2, \ldots, a_m \in \mathbb{F}$, the matrix

$$a_1A_1 + a_2A_2 + \cdots + a_mA_m \in \mathbb{F}^{m\times n}$$

is the linear combination of $A_1, A_2, \ldots, A_m$ with coefficients $a_1, a_2, \ldots, a_m$. The set of all linear combinations $a_1A_1 + a_2A_2 + \cdots + a_mA_m$ is called the *span* of $A_1, \ldots, A_m$. The set of $m \times n$ matrices over $\mathbb{F}$ with the above addition and scalar multiplication is an important example of a vector space. The theory of vector spaces is, of course, one of the basic topics of this text.

3.1.3 Examples: matrices over $\mathbb{F}_2$

Matrices over the field $\mathbb{F}_2$ are themselves quite interesting. For example, they are easy to enumerate: since $\mathbb{F}_2$ has only two elements, there are precisely 2^{mn} $m \times n$ matrices. Addition of such matrices has some interesting features, which the following example illustrates.

Example 3.2. Let

$$A = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix} \quad \text{and} \quad E = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}.$$

Then

$$A + E = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}.$$

Thus, the parity of every element of A is reversed by adding E. Adding A to itself gives

$$A + A = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix} + \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} = O.$$

Thus a matrix over $\mathbb{F}_2$ is its own additive inverse. ⊠

Example 3.3. Here is how one might design a scanner to analyze black-and-white photographs. A photo is a rectangular array consisting of many black and white dots. By giving the white dots the value 0 and the black dots the value 1, a photo is therefore transformed into a (very large) matrix over $\mathbb{F}_2$. Now suppose one wants to compare two black-and-white photographs to see whether they are in fact identical. Suppose they have each been encoded as $m \times n$ matrices A and B. It turns out to be very inefficient for a computer to scan the two matrices to see in how many positions they agree. However, when A and B are added, the sum $A + B$ has a 1 in every component where A and B differ, and a 0 wherever they coincide. For example, the sum of two identical photographs is the zero matrix, and the sum of a photograph and its negative is the all-ones matrix. An obvious measure of how similar the two matrices A and B are is the number of nonzero entries of $A + B$, i.e., $\Sigma_{i,j}(a_{ij} + b_{ij})$. This is an easily tabulated number, which is known as the *Hamming distance* between A and B. ⊠

Example 3.4. Random Key Cryptography. Suppose Rocky the flying squirrel wants to send a message to his sidekick, Bullwinkle the moose, and he wants to make sure that the notorious villains Boris and Natasha won't be able to learn what it says. Here is what the ever resourceful squirrel might do. First he could assign the numbers 1 to a, 2 to b, and so forth up to 26 to z. He then computes the binary expansion of each integer between 1 and 26. Thus $1 = 1$, $2 = 10$, $3 = 11$, $4 = 100$, $\ldots$, $26 = 11010$. He now converts his message into a sequence of five-digit strings ($1 = 00001$, etc.). The result is an encoding of the message, which is referred to as the *plaintext*. To make things more compact, he arranges the plaintext into an $m \times n$ matrix, call it P. Now the deception begins. Rocky and Bullwinkle have already selected an $m \times n$ matrix of five-digit strings of 0's and 1's, which we call Q. This matrix is what cryptographers sometimes call a *key*. Rocky will send the matrix $P + Q$ to Bullwinkle, where the addition of strings is performed in $(\mathbb{F}_2)^{5\times 1}$. Bullwinkle will be able to recover P easily by adding Q. Indeed, $(P + Q) + Q = P + (Q + Q) = P + 2Q = P$. This is good, since Bullwinkle, being a moose, is somewhat mathematically challenged and finds subtraction difficult. Even if Boris and Natasha manage to intercept the ciphertext $P + Q$, they need to know the key Q to recover P. However, the key Q must be sufficiently random so that neither Boris nor Natasha can guess it. The squirrel's encryption scheme is extremely secure if the key Q is a *one-time pad*, that is, it is used only once. ⊠

3.1.4 Matrix multiplication

Matrix addition and scalar multiplication are simple and natural operations. The product of two matrices, on the other hand, is a little more

complicated. We already mentioned that multiplication can be used to represent the unwieldy linear system (3.1) as $A\mathbf{x} = \mathbf{b}$. Let us look at this more carefully. Suppose A is a $1 \times n$ row matrix and B is an $n \times 1$ column matrix. The product AB is then defined as follows:

$$AB = \begin{pmatrix} a_1 & \cdots & a_n \end{pmatrix} \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix} = \sum_{i=1}^{n} a_i b_i. \tag{3.4}$$

We will call this product the *dot product* of A and B. Note that it is very important that the number of columns of A and the number of rows of B agree. We next define the product of an $m \times n$ matrix A and an $n \times p$ matrix B by generalizing the dot product.

Definition 3.3. Let $A \in \mathbb{F}^{m\times n}$ and $B \in \mathbb{F}^{n\times p}$. Then the *product* AB of A and B is defined as the matrix $C \in \mathbb{F}^{m\times p}$ whose entry in the ith row and kth column is the dot product of the ith row of A and the kth column of B. That is,

$$c_{ik} = \sum_{j=1}^{n} a_{ij} b_{jk}.$$

Thus

$$AB = \Big(\sum_{j=1}^{n} a_{ij} b_{jk} \Big).$$

One can therefore express the effect of multiplication as

$$\mathbb{F}^{m\times n}\ \mathbb{F}^{n\times p} \subset \mathbb{F}^{m\times p}.$$

One can also formulate the product in terms of linear combinations. Namely, if the columns of A are $\mathbf{a}_1, \ldots, \mathbf{a}_n$, then, since the scalar in the ith row of each column of B multiplies an element in the ith column of A, we see that the rth column of AB is

$$b_{1r}\mathbf{a}_1 + b_{2r}\mathbf{a}_2 + \cdots + b_{nr}\mathbf{a}_n. \tag{3.5}$$

Hence the rth column of AB is a linear combination of all n columns of A. The entries in the rth column of B are the scalars. Similarly, one can express AB as a linear combination of the rows of B. We will leave this as an exercise.

Example 3.5. Here are two examples in which $\mathbb{F} = \mathbb{Q}$:

$$\begin{pmatrix} 1 & 3 \\ 2 & 4 \end{pmatrix} \begin{pmatrix} 6 & 0 \\ -2 & 7 \end{pmatrix} = \begin{pmatrix} 1\cdot 6 + 3\cdot(-2) & 1\cdot 0 + 3\cdot 7 \\ 2\cdot 6 + 4\cdot(-2) & 2\cdot 0 + 4\cdot 7 \end{pmatrix} = \begin{pmatrix} 0 & 21 \\ 4 & 28 \end{pmatrix}.$$

Note how the columns of the product are linear combinations. Computing the product in the opposite order gives a different result:

$$\begin{pmatrix} 6 & 0 \\ -2 & 7 \end{pmatrix}\begin{pmatrix} 1 & 3 \\ 2 & 4 \end{pmatrix} = \begin{pmatrix} 6\cdot 1+0\cdot 2 & 6\cdot 3+0\cdot 4 \\ -2\cdot 1+7\cdot 2 & -2\cdot 3+7\cdot 4 \end{pmatrix} = \begin{pmatrix} 6 & 18 \\ 12 & 22 \end{pmatrix}.$$

⊠

This example points out that matrix multiplication on $\mathbb{F}^{n\times n}$ is in general not commutative: if $A, B \in \mathbb{F}^{n\times n}$, then in general, $AB \neq BA$. In fact, if you randomly choose two 2×2 matrices over $\mathbb{Q}$, it is a safe bet that they won't commute.

3.1.5 The Algebra of Matrix Multiplication

We have now defined the three basic algebraic operations on matrices. Let us next see how they interact. Although matrix multiplication isn't commutative, matrix addition and multiplication behave as expected.

Proposition 3.2. *Assuming that all the sums and products below are defined, matrix addition and multiplication satisfy the following conditions.*

(*i*) **the associative laws.** *addition and multiplication are associative:*

$$\big(A+B\big)+C = A+\big(B+C\big) \qquad \textit{and} \qquad \big(AB\big)C = A\big(BC\big);$$

(*ii*) **the distributive laws.** *addition and multiplication are distributive:*

$$A\big(B+C\big) = AB+AC \qquad \textit{and} \qquad \big(A+B\big)C = AC+BC;$$

(*iii*) **the scalar multiplication law.** *for every scalar r,*

$$\big(rA\big)B = A\big(rB\big) = r\big(AB\big);$$

(*iv*) **the commutative law for addition.** *addition is commutative:*

$$A+B = B+A.$$

Verifying these properties is a routine exercise. We already commented that addition is associative and commutative while showing that $\mathbb{F}^{m\times n}$ is an abelian group under addition. The reader should note that the validity of the associative law for multiplication will be extremely useful. We will see this, for example, when we consider matrix inverses.

The $n \times n$ matrix I_n having ones everywhere on its diagonal and zeros everywhere off the diagonal is called the *identity matrix*. One sometimes writes $I_n = (\delta_{ij})$, where δ_{ij} is the *Kronecker delta*, which is defined by the rule $\delta_{ii} = 1$ while $\delta_{ij} = 0$ if $i \neq j$. For example,

$$I_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad I_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

A convenient way to define I_n is to write out its columns. Let $\mathbf{e}_i$ denote the column vector whose ith component is 1 and whose other components are 0. Then

$$I_n = (\mathbf{e}_1 \ \mathbf{e}_2 \ \cdots \ \mathbf{e}_n).$$

The reason I_n is called the identity matrix is due to the following fact.

Proposition 3.3. *Let A be an $m \times n$ matrix over $\mathbb{F}$. Then $I_m A = A$ and $AI_n = A$. In particular, the ith column of A is $A\mathbf{e}_i$.*

Proof. This follows immediately from the definition of matrix multiplication. □

3.1.6 The transpose of a matrix

The *transpose* of an $m \times n$ matrix A is the $n \times m$ matrix A^T whose ith row is the ith column of A. That is, if $A = (a_{ij})$, then $A^T = (c_{rs})$, where $c_{rs} = a_{sr}$. The definition becomes clearer after working an example.

Example 3.6. If

$$A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix},$$

then

$$A^T = \begin{pmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{pmatrix}.$$

⊠

For another example, note that $\mathbf{e}_i = (0, \ldots, 0, 1, 0, \ldots, 0)^T$, where 1 is in the ith component. Note that A and A^T have the same entries on the diagonal. Also note that the transpose of A^T is A:

$$(A^T)^T = A.$$

Definition 3.4. A matrix A that is equal to its own transpose is called *symmetric*.

An example of a 2×2 symmetric matrix is

$$\begin{pmatrix} 1 & 3 \\ 3 & 5 \end{pmatrix}.$$

Clearly, a symmetric matrix is square (but not uninteresting). For example, the symmetric matrices over $\mathbb{R}$ are all turn out to be diagonalizable due to a fundamental result called the principal axes theorem. But this will not be explained for quite a while.

The transpose of a sum is as expected, but the transpose of a product has a twist, as we note in the next proposition.

Proposition 3.4. *For every $m \times n$ matrices A and B,*

$$(A + B)^T = A^T + B^T.$$

Furthermore, if A is $m \times n$ and B is $n \times p$, then

$$\big(AB\big)^T = B^T A^T.$$

Proof. The first identity is immediate. To prove the product identity, note that the (i, j) entry of $B^T A^T$ is the dot product of the ith row of B^T and the jth column of A^T. This is the same thing as the dot product of the jth row of A and the ith column of B, which is the (j, i) entry of AB and hence the (i, j) entry of $(AB)^T$. Thus $(AB)^T = B^T A^T$. □

It is suggested that the reader try this proof on an example.

3.1.7 Matrices and linear mappings

One way to look at matrix multiplication is that it defines a mapping. Let $\mathbb{F}^n = \mathbb{F}^{n\times 1}$ and $\mathbb{F}^m = \mathbb{F}^{m\times 1}$. Then for every $A \in \mathbb{F}^{m\times n}$, say $A = (\mathbf{a}_1 \ \cdots \ \mathbf{a}_n)$, we obtain a mapping

$$T_A : \mathbb{F}^n \to \mathbb{F}^m, \quad \text{where} \quad T_A(\mathbf{x}) = A\mathbf{x} = x_1\mathbf{a}_1 + \cdots + x_n\mathbf{a}_n.$$

Note that $T_A(\mathbf{e}_i) = \mathbf{a}_i$ for each index i, and T_A is uniquely determined by the $T_A(\mathbf{e}_i)$.

Proposition 3.5. *For every $A \in \mathbb{F}^{m\times n}$, the mapping $T_A : \mathbb{F}^n \to \mathbb{F}^m$ associated to A is in fact a homomorphism with domain $\mathbb{F}^n$ and target $\mathbb{F}^m$. That is, $A(\mathbf{x} + \mathbf{y}) = A\mathbf{x} + A\mathbf{y}$ for every $\mathbf{x}, \mathbf{y} \in \mathbb{F}^n$.*

Proof. This is simply an application of the distributive law for multiplication. □

We will frequently use this proposition in geometric examples.

Example 3.7. (Rotations of $\mathbb{R}^2$). One such example is a rotation of $\mathbb{R}^2$ given by the mapping

$$\mathcal{R}_\theta \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x\cos\theta - y\sin\theta \\ x\sin\theta + y\cos\theta \end{pmatrix}.$$

Here, $\mathcal{R}_\theta$ is the matrix mapping associated to the matrix

$$R_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}. \tag{3.6}$$

Recalling that $\mathbb{C} = \mathbb{R}^2$, multiplication by the complex exponential $e^{i\theta} = \cos\theta + i\sin\theta$ sends $z = x + iy$ to

$$e^{i\theta} z = (\cos\theta + i\sin\theta)(x + iy) = (x\cos\theta - y\sin\theta) + i(x\sin\theta + y\cos\theta).$$

Therefore, the matrix mapping given by R_θ is the same as multiplication by the complex exponential $e^{i\theta}$. ⊠

Now suppose $A \in \mathbb{F}^{m\times n}$ and $B \in \mathbb{F}^{n\times p}$. Then T_{AB} is the mapping with domain $\mathbb{F}^p$ and target $\mathbb{F}^m$ defined by $T_{AB}(\mathbf{u}) = (AB)\mathbf{u}$ for every $\mathbf{u} \in \mathbb{F}^p$. Since by associativity $(AB)\mathbf{u} = A(B\mathbf{u})$, it follows that $T_A \circ T_B = T_{AB}$. Therefore, we have the following.

Proposition 3.6. *If $A \in \mathbb{F}^{m\times n}$ and $B \in \mathbb{F}^{n\times p}$, then the matrix mapping $T_{AB} : \mathbb{F}^p \to \mathbb{F}^m$ satisfies*

$$T_{AB} = T_A \circ T_B.$$

That is, T_{AB} is the composition of the matrix mappings $T_B : \mathbb{F}^p \to \mathbb{F}^n$ and $T_A : \mathbb{F}^n \to \mathbb{F}^m$. Hence matrix multiplication corresponds to the composition of matrix mappings.

Exercises

Exercises 3.1.1. As an experiment, construct three matrices A, B, C of dimensions such that AB and BC are defined. Then compute AB and $(AB)C$. Next compute BC and $A(BC)$ and compare your results. If A and B are also square, do AB and BA coincide?

Exercises 3.1.2. Prove the assertion $(A + B)^T = A^T + B^T$ in Proposition 3.4 without writing down matrix entries.

Exercises 3.1.3. Suppose A and B are symmetric $n \times n$ matrices. (You can even assume $n = 2$.)

(i) Must AB be symmetric? That is, are the $n \times n$ symmetric matrices closed under multiplication?

(ii) If the answer to (i) is no, find a condition that ensures that AB is symmetric.

Exercises 3.1.4. Suppose B has a column of zeros. How does this affect a product of the form AB? What if A has a row of zeros?

Exercises 3.1.5. State a rule for expressing the rows of AB as linear combinations of the rows of B. (Suggestion: use the transpose identity and the result expressing the columns of AB).

Exercises 3.1.6. Verify Proposition 3.3 for all A in $\mathbb{F}^{m\times n}$.

Exercises 3.1.7. Let $\mathbb{F} = \mathbb{R}$. Find all 2×2 matrices $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ such that $AB = BA$, where $B = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$.

Exercises 3.1.8. Assume here that $\mathbb{F} = \mathbb{F}_3$. Find all 2×2 matrices $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ such that $AC = CA$, where $C = \begin{pmatrix} 1 & 2 \\ 1 & 2 \end{pmatrix}$. Is your result here any different from the result you obtained in Exercise 3.1.7.

Exercises 3.1.9. Let $\mathbb{F}$ be a field. Prove that if $S \in \mathbb{F}^{2\times 2}$ commutes with every matrix $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathbb{F}^{2\times 2}$, then $S = aI_2$ for some $a \in \mathbb{F}$. Matrices of the form aI_n are called *scalar matrices.*

Exercises 3.1.10. Let p be a prime, and let A be the 2×2 matrix over $\mathbb{F}_p$ such that $a_{ij} = 1$ for each i, j. Compute A^m for every integer $m \geq 1$. (Note that A^m stands for the mth power of A.)

Exercises 3.1.11. Let A be the $n \times n$ matrix over $\mathbb{Q}$ such that $a_{ij} = 2$ for all i, j. Find a formula for A^j for every positive integer j.

Exercises 3.1.12. Give an example of a 2×2 matrix A such that every entry of A is either 0 or 1 and $A^2 = I_2$ as a matrix over $\mathbb{F}_2$, but $A^2 \neq I_2$ as a matrix over $\mathbb{Q}$.

3.2 Reduced Row Echelon Form

The standard procedure for solving a linear system is to use a process called Gaussian elimination to put the system in a standard form in which it is ready to solve. This process involves using row operations to put the coefficient matrix into a standard form called reduced row echelon form. It will turn out that every matrix can be put into this standard form by a premultiplication. We will later see that row operations preserve two important quantities associated to A: the row space of A and its null space.

3.2.1 Reduced row echelon form and row operations

Definition 3.5. A matrix A is said to be in *row echelon form* if

(i) the first nonzero entry in each row of A is to the right of the first nonzero entry in the preceding row (and hence in all preceding rows), and

(ii) every entry above a first nonzero entry is zero.

The first nonzero entry in a row is called its *pivot entry* or *corner entry*. A matrix A in row echelon form is said to be in *reduced row echelon form*, or simply, to be *reduced*, if each corner entry is 1. Here are some examples of reduced matrices:

$$\begin{pmatrix} 1 & 0 & 0 & 2 \\ 0 & 1 & 0 & 3 \\ 0 & 0 & 1 & 5 \end{pmatrix}, \quad \begin{pmatrix} 1 & 2 & 3 & 0 & 9 \\ 0 & 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \quad \begin{pmatrix} 0 & 1 & 3 & 0 & 9 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}.$$

The zero matrix $O \in \mathbb{F}^{m\times n}$ and the identity matrix I_n are reduced also.

A row operation on a matrix A replaces one of the rows of A by a new row. The three row operations we will now define are called the *elementary row operations*.

Definition 3.6. Let A be a matrix over $\mathbb{F}$ with rows $\mathbf{a}_1, \ldots, \mathbf{a}_m$. The elementary row operations over $\mathbb{F}$ on A are as follows:

(I) interchange the ith row $\mathbf{a}_i$ and the jth row $\mathbf{a}_j$, where $i \neq j$;

(II) replace the ith row $\mathbf{a}_i$ with a nonzero scalar multiple $r\mathbf{a}_i$, where $r \in \mathbb{F}$;

(III) replace the ith row $\mathbf{a}_i$ by $\mathbf{a}_i + r\mathbf{a}_j$, where $r \in \mathbb{F}$, $r \neq 0$, and $i \neq j$. That is, replace $\mathbf{a}_i$ by itself plus a nonzero multiple of some other row.

Row operations of type I are called *row swaps*. The type II operations are called *row dilations*, and operations of type III are called *transvections*. The next proposition gives the basic property of row operations.

Proposition 3.7. *Every matrix over a field $\mathbb{F}$ can be put into reduced row echelon form by a (not unique) sequence of elementary row operations over $\mathbb{F}$.*

Before giving a proof, let us work an example. Each arrow below indicates a single row operation, and the notation over the arrows indicates the row operation in an obvious way.

Example 3.8. Consider the counting matrix

$$C = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}$$

as a matrix over $\mathbb{Q}$. We can row reduce C as follows:

$$C \xrightarrow{R_2-4R_1} \begin{pmatrix} 1 & 2 & 3 \\ 0 & -3 & -6 \\ 7 & 8 & 9 \end{pmatrix} \xrightarrow{R_3-7R_1} \begin{pmatrix} 1 & 2 & 3 \\ 0 & -3 & -6 \\ 0 & -6 & -12 \end{pmatrix}$$

$$\xrightarrow{R_3-2R_2} \begin{pmatrix} 1 & 2 & 3 \\ 0 & -3 & -6 \\ 0 & 0 & 0 \end{pmatrix} \xrightarrow{(-1/3)R_2} \begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{pmatrix} \xrightarrow{R_1-2R_2} \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{pmatrix}.$$

⊠

Let us now prove Proposition 3.7.

Proof. We will induct on the number of columns of A. Assume first that A has just one column. If A is the zero matrix O, it is already in reduced row echelon form, so suppose $A \neq O$. If $a_{11} \neq 0$, multiply the first row by a_{11}^{-1} to produce a 1 in the $(1,1)$ position. (This is why we require row operations over a field.) We can then use row operation s of type III to make all other entries in the first column zero. If $a_{11} = 0$, but the first column has a nonzero entry somewhere, say the ith row, then swap the first row with the ith row to create a nonzero entry in the $(1,1)$ position. Next, proceed as before, dividing the new first row by the inverse of its $(1,1)$ entry, getting a one in the $(1,1)$ position. Then use row operations of type III to make all the other elements in the first column 0. Thus we have put A into reduced row echelon form by row operations.

Now suppose that every $m \times n$ matrix over $\mathbb{F}$ can be put into reduced row echelon form by row operations, and let A be an $m \times (n+1)$ matrix over $\mathbb{F}$. Let A' be the $m \times n$ matrix consisting of the first n columns of A. By the induction hypothesis, there exists a sequence of row operations that puts A' in reduced row echelon form, say B'. After performing this sequence of row operations on A, we obtain a matrix B whose the first n columns have been

put into reduced row echelon form. Assume that the lowest corner entry in B' is in the kth row. If the last column of B has only zeros below the kth row, then B is already in reduced row echelon form, completing the proof. Otherwise, B has a nonzero entry below the kth row, and we can repeat the steps used in the case of one column to get a corner entry in the $(k+1, n+1)$ of B and zeros below. Since these row operations affect only the rows below the kth row, B' is not changed. Thus, A can be put into reduced row echelon form. This completes the induction step, so every matrix over $\mathbb{F}$ can be put into reduced row echelon form by elementary row operations. □

Remark. One can refer to performing a sequence of row operations on A as row surgery. For every matrix A, there are many different row surgeries leading to a matrix in reduced row echelon form. An interesting and important point, often completely ignored, is that a matrix has only one reduced row echelon form. That is, the reduced row echelon form of an arbitrary $A \in \mathbb{F}^{m\times n}$ is unique. We will use A_{red} to denote this matrix, even though we have not yet proved its uniqueness. The proof, which is not obvious, will be given in Proposition 3.12. This result will let us assert that the number of nonzero rows in A_{red} is unique. The term commonly used for this number is the row rank of A. The row rank is important, because it gives us information such as when the solution of a linear system is unique.

First we will introduce elementary matrices in order to get an efficient algorithm for row reducing a matrix.

3.2.2 Elementary matrices and row operations

We now introduce elementary matrices and explain their role in reduced row echelon form.

Definition 3.7. An $n \times n$ matrix that is obtained by performing a single row operation on I_n is called an *elementary matrix.*

Example 3.9. The elementary 2×2 matrices are illustrated as follows:

$$E_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad E_2 = \begin{pmatrix} r & 0 \\ 0 & 1 \end{pmatrix} \text{ or } \begin{pmatrix} 1 & 0 \\ 0 & r \end{pmatrix}, \qquad E_3 = \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix} \text{ or } \begin{pmatrix} 1 & 0 \\ s & 1 \end{pmatrix}.$$

Here, r and s are nonzero scalars. ⊠

The following computations show that premultiplication by one of the above 2×2 elementary matrices E_i performs the same row operation on $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ that is performed on the identity I_2 to get E_i:

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} c & d \\ a & b \end{pmatrix} \quad \text{(row swap)},$$

$$\begin{pmatrix} r & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} ra & rb \\ c & d \end{pmatrix} \quad \text{(row dilation)},$$

$$\begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix}\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} a+sc & b+sd \\ c & d \end{pmatrix} \quad \text{(row transvection)}.$$

One might call the next proposition the golden rule of matrix multiplication.

Proposition 3.8. *Let A be of size $m \times n$, and assume that E is an elementary $m \times m$ matrix. Then EA is the matrix obtained by performing the row operation corresponding to E on A.*

Proof. Recall that for every $B \in \mathbb{F}^{m\times m}$, the rows of BA are linear combinations of the rows of A using the entries of $B = (b_{ij})$ as scalars. In fact, if $\mathbf{a}_i$ is the ith row of A, then the ith row of BA is

$$b_{i1}\mathbf{a}_1 + b_{i2}\mathbf{a}_2 + \cdots + b_{im}\mathbf{a}_m.$$

Thus, if B is the elementary matrix obtained by multiplying the ith row of I_n by r, then the ith row of BA becomes $r\mathbf{a}_i$, and all other rows are unchanged. Likewise, if B is obtained by interchanging the ith and jth rows of I_n, then the ith row of BA is $\mathbf{a}_j$, since $b_{ik} = \delta_{jk}$, and similarly, the jth row is $\mathbf{a}_i$. The argument for the third type of row operation is analogous, so it is left as an exercise. □

In fact, since $EI_n = E$, the matrix E performing the desired row operation is unique.

Thus row reduction can be expressed as follows: starting with A and replacing it by $A_1 = E_1A$, $A_2 = E_2(E_1A)$, and so forth, we get the sequence

$$A \to A_1 = E_1A \to A_2 = E_2(E_1A) \to \cdots \to A_k = E_k(E_{k-1}(\cdots(E_1A)\cdots)).$$

Assuming that the right-hand matrix A_k is A_{red}, we obtain by this process a product of elementary matrices

$$B = E_kE_{k-1}\cdots E_1$$

such that $BA = A_{red}$. Note: this assertion uses the associativity of matrix multiplication. It should be emphasized that the way we choose the E_i isn't unique. Yet it will turn out that in certain cases, their product B is unique. This seems to be a remarkable fact.

Example 3.10. Let's compute the matrix B produced by the sequence of row operations in Example 3.8, which puts the counting matrix C in reduced form. Examining the sequence of row operations, we see that B is the product

$$\begin{pmatrix} 1 & -2 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1/3 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -2 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -7 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ -4 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

In computing the matrix B, the easy way is to start at the right and apply the sequence of row operations working to the left. A convenient way of doing this is to begin with the 3×6 matrix $(A \mid I_3)$ and carry out the sequence of row operations; the final result will be $(A_{red} \mid B)$. Thus if we start with

$$(C \mid I_3) = \begin{pmatrix} 1 & 2 & 3 & 1 & 0 & 0 \\ 4 & 5 & 6 & 0 & 1 & 0 \\ 7 & 8 & 9 & 0 & 0 & 1 \end{pmatrix},$$

we end with

$$(C_{red} \mid B) = \begin{pmatrix} 1 & 0 & -1 & -5/3 & 2/3 & 0 \\ 0 & 1 & 2 & 4/3 & -1/3 & 0 \\ 0 & 0 & 0 & 1 & -2 & 1 \end{pmatrix}.$$

⊠

We may summarize this section by observing that elementary matrices perform row operations, so Proposition 3.7 can be restated as follows.

Proposition 3.9. *An arbitrary $m \times n$ matrix A over a field $\mathbb{F}$ can be put into reduced row echelon form by performing a sequence of left multiplications on A using $m \times m$ elementary matrices over $\mathbb{F}$. In other words, there exist elementary matrices $E_1, \ldots, E_k \in \mathbb{F}^{m \times m}$ such that*

$$A_{red} = E_k E_{k-1} \cdots E_1 A.$$

Proof. By Proposition 3.7, every matrix can be put into reduced form by a sequence of row operations. But row operations are performed by left multiplication by elementary matrices. □

3.2.3 The row space and uniqueness of reduced row echelon form

After doing several row reductions, the reader may wonder whether A_{red} is unique. Could choosing a different sequence of row operations lead to a

different reduced matrix? As we remarked above, the answer is no. We now prove this. We begin by noticing that every elementary row operation replaces a row of A by a linear combination of its other rows. This leads us to consider a fundamental notion, namely the *row space* of A.

Definition 3.8. Suppose $A \in \mathbb{F}^{m\times n}$. The *row space* of A is defined as the span of the rows of A.

The row space of A is denoted by $\mathrm{row}(A)$. Thus, $\mathrm{row}(A)$ is the set of all linear combinations of the rows of A. The basic result about row spaces and the key to understanding the role of row operations is the following:

Proposition 3.10. *If A' is obtained from A by applying an elementary row operation to one of the rows of A, then A' and A have the same row space. That is,* $\mathrm{row}(A') = \mathrm{row}(A)$.

Proof. Let A and A' be of size $m \times n$, and let $\mathbf{a}_1, \ldots, \mathbf{a}_m$ be the rows of A. If the row operation is of type I, then A and A' have the same row space, because a type I operation just reorders the rows, hence does not change the set of all linear combinations. Likewise, a type II row operation replaces some $\mathbf{a}_i$ by $r\mathbf{a}_i$, where r is a nonzero scalar. Thus

$$c_1\mathbf{a}_1 + \cdots + c_i\mathbf{a}_i + \cdots + c_m\mathbf{a}_m = c_1\mathbf{a}'_1 + \cdots + \frac{c_i}{r}\mathbf{a}'_i + \cdots + c_m\mathbf{a}'_m.$$

Hence A and A' also have the same row space. Finally, if A' is obtained by replacing $\mathbf{a}_i$ by $\mathbf{a}_i + r\mathbf{a}_j$, where $j > i$, then

$$\begin{aligned} c_1\mathbf{a}_1 + \cdots + c_i\mathbf{a}_i + \cdots + c_m\mathbf{a}_m = c_1\mathbf{a}_1 + \cdots + c_i(\mathbf{a}_i + r\mathbf{a}_j) \\ + \cdots + (c_j - rc_i)\mathbf{a}_j + \cdots + c_m\mathbf{a}_m. \end{aligned}$$

Again, it follows that A and A' have the same row space. □

We now prove the key step.

Proposition 3.11. *Suppose A and B are two $m \times n$ matrices over $\mathbb{F}$ in reduced row echelon form. Then $A = B$ if and only if their row spaces are equal.*

Proof. If $A = B$, then obviously their row spaces are equal. The proof in the other direction isn't hard, but it's a little hard to write out. First, suppose that A has k nonzero rows and B has ℓ of them. Since we don't know the relationship between k and ℓ, we may harmlessly suppose that $k \leq \ell$. Let $\mathbf{a}_1, \ldots, \mathbf{a}_k$ denote the nonzero rows of A. We will assume that they are labeled so that the first nonzero component of $\mathbf{a}_i$ is to the right of that of $\mathbf{a}_{i+1}$. Hence $\mathbf{a}_1$ is the last nonzero row of A. Assume that the nonzero rows $\mathbf{b}_1, \ldots, \mathbf{b}_\ell$ of

B are labeled in the same way. The first step is to show that $\mathbf{a}_1 = \mathbf{b}_1$. This is done as follows. Suppose the corner entries in $\mathbf{a}_1$ and $\mathbf{b}_1$ aren't in the same component; say the corner entry of $\mathbf{a}_1$ is to the left of the corner entry of $\mathbf{b}_1$. Then obviously $\mathbf{a}_1$ can't be a linear combination of the $\mathbf{b}_i$. By symmetry, the corner entries of $\mathbf{a}_1$ and $\mathbf{b}_1$ are in the same component. Since $\mathbf{b}_1$ is a linear combination of the $\mathbf{a}_i$, the only possibility is $\mathbf{b}_1 = r_1\mathbf{a}_1$, and in fact, $r_1 = 1$, due to the fact that the corner entries of $\mathbf{a}_1$ and $\mathbf{b}_1$ are both one. Now, the matrices

$$A_1 = \begin{pmatrix} \mathbf{a}_2 \\ \mathbf{a}_1 \end{pmatrix} \quad \text{and} \quad B_1 = \begin{pmatrix} \mathbf{b}_2 \\ \mathbf{b}_1 \end{pmatrix}$$

have the same second row and are, by assumption, in reduced row echelon form. Suppose the corner entry in $\mathbf{b}_2$ is to the right of that of $\mathbf{a}_2$. Then $\mathbf{b}_2$ cannot be a linear combination of the $\mathbf{a}_i$, so once again $\mathbf{a}_2$ and $\mathbf{b}_2$ have to have their corner entries in the same component. Hence, $\mathbf{b}_2 = \mathbf{a}_2 + r_1\mathbf{a}_1$. But since A_1 is in reduced row echelon form, $\mathbf{a}_2 + r_1\mathbf{a}_1$ has r_1 in the component corresponding to the corner entry of $\mathbf{a}_1$. But since B_1 is also in reduced row echelon form and $\mathbf{b}_1 = \mathbf{a}_1$, it follows that $\mathbf{b}_2$ has a zero in that component. Consequently, $r_1 = 0$. Thus, $\mathbf{b}_2 = \mathbf{a}_2$. One can now repeat this argument to show that $\mathbf{b}_3 = \mathbf{a}_3$, $\mathbf{b}_4 = \mathbf{a}_4$, and so on, eventually arriving at the conclusion $\mathbf{b}_i = \mathbf{a}_i$ for $1 \leq i \leq k$. Finally, we have to show that $\ell = k$. If $k < \ell$, then $\mathbf{b}_{k+1} \neq \mathbf{0}$, and it lies in $\mathrm{row}(B) = \mathrm{row}(A)$. Hence, $\mathbf{b}_{k+1}$ is a linear combination of $\sum r_i\mathbf{a}_i$ with some $r_i \neq 0$. But this is impossible, because the component of $\mathbf{b}_{k+1}$ corresponding to the corner entry is to the left of that of $\mathbf{a}_k$ due to the fact that $\mathbf{a}_k = \mathbf{b}_k$. This shows that $k = \ell$ and hence finishes the proof that $A = B$. □

Finally, we answer the basic question.

Proposition 3.12. *The reduced row echelon form of an arbitrary matrix $A \in \mathbb{F}^{m\times n}$ is unique.*

Proof. If two different sequences of row operations yield two different reduced row echelon forms B and C for A, then by the previous proposition, we obtain a contradiction to $\mathrm{row}(A) = \mathrm{row}(B) = \mathrm{row}(C)$. Hence $B = C$. □

The fact that the reduced row echelon form of a matrix A is unique means that the number of nonzero rows of A_{red} depends only on A. This leads to the following important definition.

Definition 3.9. Suppose $A \in \mathbb{F}^{m\times n}$. Then the number of nonzero rows in A_{red} is called the *rank* of A.

When we study vector spaces, we will show that the rank of A is actually the dimension of the row space of A. For square matrices, it turns out that rank is a measure of how far the matrix is from being invertible.

Finally, note that Proposition 3.9 and the above discussion imply the following.

Proposition 3.13. *A matrix $A \in \mathbb{F}^{n\times n}$ has rank n if and only if there exist elementary matrices $E_1, \ldots, E_k \in \mathbb{F}^{n\times n}$ such that*

$$E_k \cdots E_1 A = I_n.$$

Proof. By definition, A has rank n if and only if A_{red} has n nonzero rows. But since $A \in \mathbb{F}^{n\times n}$, A_{red} has n nonzero rows if and only if $A_{red} = I_n$. Thus, A has rank n implies by Proposition 3.9 that $E_k \cdots E_1 A = I_n$ for suitable elementary matrices $E_1, \ldots, E_k$. Conversely, if there exist elementary matrices $E_1, \ldots, E_k$ such that $E_k \cdots E_1 A = I_n$, then $A_{red} = I_n$ by the uniqueness of the reduced form. □

Definition 3.10. When the rank of an $n \times n$ matrix A over a field $\mathbb{F}$ is n, we will say that A is *nonsingular*. Otherwise, A is said to be *singular*.

Exercises

Exercises 3.2.1. Make a list of all the row reduced 2×3 matrices over $\mathbb{F}_2$.

Exercises 3.2.2. Let $\mathbb{F} = \mathbb{F}_3$. Find the reduced row echelon form of the following matrices:

$$\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}, \quad \begin{pmatrix} 0 & 1 & 2 & 1 \\ 1 & 2 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{pmatrix}, \quad \begin{pmatrix} 1 & 2 & 0 \\ 2 & 1 & 1 \\ 0 & 2 & 1 \end{pmatrix}.$$

Exercises 3.2.3. True or False (give a brief reason to justify your answer).

(*i*) Two square matrices that can be row reduced to the same reduced row echelon form are equal.

(*ii*) If a 3×3 matrix A has the property that each row contains two zeros and a one, and each column contains two zeros and a one, then A is nonsingular.

Exercises 3.2.4. If an $n \times n$ matrix has the property that each row and each column has exactly one nonzero entry, is it nonsingular?

Exercises 3.2.5. The field is $\mathbb{F}_2$. Consider the following matrices:

$$A_1 = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \end{pmatrix}, \quad A_2 = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}.$$

Find matrices B_1 and B_2 that are products of elementary matrices such that B_iA_i is reduced for $i = 1, 2$.

Exercises 3.2.6. Prove that if E is an elementary $n \times n$ matrix and F is the elementary matrix that performs the inverse (i.e., reverse) operation, then $FE = EF = I_n$.

Exercises 3.2.7. Write down all the 3×3 elementary matrices E over $\mathbb{F}_2$, and for each E, find the matrix F defined in the previous exercise such that $FE = EF = I_3$.

Exercises 3.2.8. In this exercise, we will introduce column operations.

(*i*) Define the notion of reduced column echelon form for an $m \times n$ matrix.

(*ii*) Next, define the three types of column operations.

(*iii*) Show how to perform column operations using elementary matrices.

3.3 Linear Systems

So far, we have defined matrices, introduced matrix algebra, and studied row operations and the reduced form of a matrix. Let us now go back and consider a problem that motivated the introduction of matrices: how to find the *solution set* of a system of linear equations.

3.3.1 The coefficient matrix of a linear system

Recall from (3.1) that a linear system of m equations in n variables with coefficients and constants in a field F has the form

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1, \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2, \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m. \end{aligned} \tag{3.7}$$

These equations can be expressed compactly in an array called the *coefficient matrix*, consisting of the coefficients and constants as follows:

$$(a_{ij} \mid \mathbf{b}) = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{pmatrix}. \tag{3.8}$$

The *solution set* of the system (3.7) is the set of all $\mathbf{x} = (x_1, x_2, \ldots, x_n)^T \in \mathbb{F}^n$ that satisfy each equation.

The procedure for finding the solution set called Gaussian elimination is to use row operations to bring $(a_{ij} \mid \mathbf{b})$ into reduced form. Since row operations can be performed by premultiplication using elementary matrices, it is natural to use the matrix equation form of the system, namely $A\mathbf{x} = \mathbf{b}$, where $A = (a_{ij})$ is the coefficient matrix and $\mathbf{b}$ is the column of constants. The key point is the following: for every elementary matrix $E \in \mathbb{F}^{m\times m}$, the matrix equations $A\mathbf{x} = \mathbf{b}$ and $EA\mathbf{x} = E\mathbf{b}$ are *equivalent* in the sense that they have exactly the same solution sets. For if $A\mathbf{x} = \mathbf{b}$, then $EA\mathbf{x} = E\mathbf{b}$. Conversely, if $EA\mathbf{x} = E\mathbf{b}$, then multiplying by the elementary matrix F such that $FE = I_n$ recovers the original system; that is,

$$A\mathbf{x} = (FE)A\mathbf{x} = F(EA\mathbf{x}) = F(E\mathbf{b}) = (FE)\mathbf{b} = \mathbf{b}.$$

Thus we have the following statement.

Proposition 3.14. *Let $A \in \mathbb{F}^{m\times n}$ and suppose B is a product of elementary matrices such that $BA = A_{red}$. Then the linear system $A\mathbf{x} = \mathbf{b}$ is equivalent to the reduced system $A_{red}\mathbf{x} = \mathbf{c}$, where $\mathbf{c} = B\mathbf{b}$*

3.3.2 Writing the solutions: the homogeneous case

We now describe how to express the solutions. For this, we need to distinguish between two types of systems. A linear system of the form $A\mathbf{x} = \mathbf{0}$ is said to be *homogeneous*, while a linear system $A\mathbf{x} = \mathbf{b}$, where $\mathbf{b} \neq \mathbf{0}$, is said to be *inhomogeneous*. The solution set of a homogeneous system $A\mathbf{x} = \mathbf{0}$ is called the *null space* of A and denoted by $\mathcal{N}(A)$. The previous proposition shows that $\mathcal{N}(A) = \mathcal{N}(A_{red})$. That is,

$$A\mathbf{x} = \mathbf{0} \text{ if and only if } A_{red}\mathbf{x} = \mathbf{0}.$$

The coefficient matrix of a homogeneous system $A\mathbf{x} = \mathbf{0}$ is simply defined as A. Thus, the coefficient matrix will be of size $m \times n$ instead of $m \times (n + 1)$. Let us now consider a homogeneous example.

Example 3.11. The homogeneous system

$$\begin{aligned} x_1 + x_2 + 2x_3 + 0x_4 + 3x_5 - x_6 &= 0,\\ x_4 + 2x_5 + 0x_6 &= 0, \end{aligned}$$

with coefficients in $\mathbb{Q}$ has coefficient matrix

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

Note that A is already reduced and has corners in the first and fourth columns. Thus we can solve for the corresponding corner variables x_1 and x_4 in terms of the noncorner variables x_2, x_3, x_5, x_6. This gives

$$\begin{aligned} x_1 &= -2x_3 - 3x_5 + x_6,\\ x_4 &= -2x_5. \end{aligned}$$

Notice that x_2 doesn't appear in these equations, but it will appear in the solution. The upshot is that the variables corresponding to A's corner columns are functions of the remaining variables, which we will call the *free variables*. Since there are six variables, we form the vector

$$\mathbf{s} = (-2x_3 - 3x_5 + x_6,\ x_2,\ x_3, -2x_5,\ x_5,\ x_6)^T \in \mathbb{Q}^6,$$

where we have replaced x_1 and x_4 in $\mathbf{x}$ by their expressions in terms of the free variables. Thus $\mathbf{s}$ contains only expressions in the free variables x_2, x_3, x_5, x_6. We call $\mathbf{s}$ the *general solution vector*. We now define *basic null vectors* $\mathbf{f}_1, \mathbf{f}_2, \mathbf{f}_3, \mathbf{f}_4$ such that

$$\mathbf{s} = x_2\mathbf{f}_1 + x_3\mathbf{f}_2 + x_5\mathbf{f}_3 + x_6\mathbf{f}_4. \tag{3.9}$$

The basic null vectors are found by inspection. To get $\mathbf{f}_1$, set $x_2 = 1$ and $x_3 = x_5 = x_6 = 0$. Repeating this for the other $\mathbf{f}_i$, we see that

$$\mathbf{f}_1 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \mathbf{f}_2 = \begin{pmatrix} -2 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \mathbf{f}_3 = \begin{pmatrix} -3 \\ 0 \\ 0 \\ -2 \\ 1 \\ 0 \end{pmatrix}, \mathbf{f}_4 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}.$$

Then (3.9) shows that every solution of $A\mathbf{x} = \mathbf{0}$ (as a vector in $\mathbb{Q}^6$) is a linear combination of the basic null vectors $\mathbf{f}_i$ with coefficients in $\mathbb{Q}$. Moreover, the coefficients x_2, x_3, x_5, x_6 are unique, since each of them occurs by itself in one of the components of $\mathbf{s}$. ⊠

The method of the above example can be used to find basic null vectors of any homogeneous linear system $A\mathbf{x} = \mathbf{0}$. In fact, since A_{red} is unique, the basic solutions as defined are also unique. Note, however, that basic null vectors exist only when the rank of A is less than n. The next result summarizes how a linear system $A\mathbf{x} = \mathbf{0}$ is solved. The proof imitates the previous example and will be omitted.

Proposition 3.15. *Suppose $A \in \mathbb{F}^{m\times n}$ has rank k, and assume $\ell = n - k > 0$. Then every $\mathbf{x} \in \mathcal{N}(A)$ can be expressed in exactly one way as a linear combination of basic null vectors $\mathbf{f}_1, \ldots, \mathbf{f}_\ell$ of A. If $k = n$, then $A\mathbf{x} = \mathbf{0}$ has the unique solution $\mathbf{x} = \mathbf{0}$ (hence there are no basic null vectors).*

3.3.3 The inhomogeneous case

The solution in the inhomogeneous case is described in the next proposition. We first note that an inhomogeneous linear system needn't have any solutions at all. The equation $0x = 1$ is such an example. More generally, if A is the $m \times n$ zero matrix, then $A\mathbf{x} = \mathbf{0}$ for every $\mathbf{x} \in \mathbb{F}^n$, so if $\mathbf{b} \neq \mathbf{0}$, then $A\mathbf{x} = \mathbf{b}$ cannot have a solution. A system $A\mathbf{x} = \mathbf{b}$ with no solutions is said to be *inconsistent*. To take a less obvious example of an inconsistent system, note that the equation $ax + by = c$ represents a line in $\mathbb{R}^2$ (assuming that a, b, c are real). Thus the system

$$\begin{aligned} ax + by &= c \\ dx + ey &= f \end{aligned}$$

represents the points where two lines in $\mathbb{R}^2$ intersect. If the lines are distinct parallel lines (e.g., $a = d,\ b = e,\ c \neq f$), then they have empty intersection, so the system is inconsistent. If the lines are distinct and nonparallel, they meet in a unique point. There is another way in which a system can be inconsistent. If it is *overdetermined*, that is, if there are more equations than variables, then it may be inconsistent. For example, three lines in $\mathbb{R}^2$ which are mutually nonparallel will meet only if the third passes through the unique intersection point of the first two. We now state the criterion for an inhomogeneous linear system to be consistent.

Proposition 3.16. *The inhomogeneous linear system $A\mathbf{x} = \mathbf{b}$ is consistent if and only if the rank of its coefficient matrix $(A \mid \mathbf{b})$ is the same as the rank of A. In particular, if A is of size $m \times n$ and has rank m, then the system $A\mathbf{x} = \mathbf{b}$ is always consistent.*

Proof. If the ranks are different, then the rank of $(A \mid \mathbf{b})$ is larger than the rank of A. This implies that if $BA = A_{red}$, then the equivalent system $BA\mathbf{x} = A_{red}\mathbf{x} = A_{red}\mathbf{b}$ contains an equation of the form $0x_1 + \cdots + 0x_n = c$, where $c \neq 0$. Such an equation is clearly inconsistent. Therefore, if the inhomogeneous linear system is consistent, the ranks of A and $(A \mid \mathbf{b})$ are the same. We leave the argument for the converse to the reader. If A has rank m, then so does $(A \mid \mathbf{b})$, so $A\mathbf{x} = \mathbf{b}$ is consistent. □

Example 3.12. The system

$$\begin{aligned} 3x + 3y &= 1 \\ x - y &= 2 \\ x + 2y &= 0 \end{aligned} .$$

has coefficient matrix

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

This matrix has rank three, so the system is inconsistent. ⊠

Let us now summarize our discussion.

Proposition 3.17. *Suppose that $A \in \mathbb{F}^{m \times n}$ has rank k, and consider an inhomogeneous linear system $A\mathbf{x} = \mathbf{b}$. This system is consistent if and only if the rank of $(A \mid \mathbf{b})$ is the same as the rank of A. Assume so and let $\mathbf{p}_0 \in \mathbb{F}^n$ be a particular solution. Then every solution can be written in the form $\mathbf{x} = \mathbf{p}_0 + \mathbf{w}$, where $\mathbf{w} \in \mathcal{N}(A)$. If $k = n$, then $\mathcal{N}(A) = \{\mathbf{0}\}$, and consequently*

$\mathbf{p}_0$ *is the unique solution. If* $k < n$*, let* $\ell = n - k$*, and suppose* $\mathbf{f}_1, \ldots, \mathbf{f}_\ell$ *are* A*'s basic null vectors. Thus, every* $\mathbf{x} \in \mathbb{F}^n$ *of the form*

$$\mathbf{x} = \mathbf{p}_0 + \sum_{i=1}^{\ell} a_i \mathbf{f}_i \quad (\text{all } a_i \in \mathbb{F}) \tag{3.10}$$

is a solution, and every solution can be written in the form (3.10) for a unique choice of the a_i*.*

Proof. Suppose $\mathbf{x}$ satisfies $A\mathbf{x} = \mathbf{b}$. Then

$$A(\mathbf{x} - \mathbf{p}_0) = A\mathbf{x} - A\mathbf{p}_0 = \mathbf{b} - \mathbf{b} = \mathbf{0}.$$

Thus, $\mathbf{w} = \mathbf{x} - \mathbf{p}_0 \in \mathcal{N}(A)$, so $\mathbf{x} = \mathbf{p}_0 + \mathbf{w}$, as claimed. If $k = n$, we know from Proposition 3.15 that $\mathcal{N}(A) = \{\mathbf{0}\}$, so $\mathbf{w} = \mathbf{0}$. Thus $\mathbf{p}_0$ is the unique solution. Now suppose $k < n$. There exist fundamental solutions $\mathbf{f}_1, \ldots, \mathbf{f}_\ell$ with $\ell = n - k$, and $\mathbf{w}$ can be uniquely written $\mathbf{w} = \sum_{i=1}^{\ell} a_i \mathbf{f}_i$. This implies the final assertion. □

Example 3.13. To illustrate this result, consider the inhomogeneous linear system

$$\begin{aligned} x_1 + x_2 + 2x_3 + 0x_4 + 3x_5 - x_6 &= 1, \\ x_4 + 2x_5 + 0x_6 &= -1. \end{aligned}$$

Note that the coefficient matrix A is taken from Example 3.11, so solving as above, we see that

$$\begin{aligned} x_1 &= 1 - 2x_3 - 3x_5 + x_6, \\ x_4 &= -1 - 2x_5. \end{aligned}$$

Therefore, $\mathbf{p}_0 = (1, 0, 0, -1, 0, 0)^T$ is a particular solution of the inhomogeneous linear system. Hence, every solution has the form $\mathbf{p}_0 + \mathbf{s}$, where $\mathbf{s}$ is the general solution vector of the homogeneous linear system $A\mathbf{x} = \mathbf{0}$. ⊠

Remark. Notice that if $A \in \mathbb{F}^{m \times n}$, then $\mathcal{N}(A)$ is an abelian group. In fact, it is a subgroup of $\mathbb{F}^n$. Proposition 3.17 says that the set of solutions of a consistent linear system $A\mathbf{x} = \mathbf{b}$ is a coset $\mathbf{p}_0 + \mathcal{N}(A)$ of $\mathcal{N}(A)$. The coset representative $\mathbf{p}_0$ is a particular solution.

Remark. Our final remark is that $A\mathbf{x} = \mathbf{b}$ is consistent if and only if $\mathbf{b}$ is a linear combination of the columns of A. This follows from the identity (3.5). We will also discuss this fact in more detail when the column space of a matrix is introduced.

3.3.4 A useful identity

Suppose $A \in \mathbb{F}^{m\times n}$. Since each corner of A_{red} occupies a unique row and a unique column, the rank k of A satisfies both $k \leq m$ and $k \leq n$. If $k < n$, then there exists at least one free variable, so A will always have a basic null vector. There is a simple but important relationship, already pointed out, between the rank k and the number of basic null vectors, or equivalently, the number of free variables:

$$\text{rank}(A) + \#\text{ free variables} = n. \tag{3.11}$$

We will restate this identity several times, each time in a more general form. In its most general form, the identity (3.11) is the rank–nullity theorem. For example, if $m < n$, then there exists at least one free variable. In particular, a system $A\mathbf{x} = \mathbf{b}$ cannot have a unique solution.

The identity (3.11) has the following useful consequence for $n \times n$ matrices.

Proposition 3.18. *Let $A \in \mathbb{F}^{n\times n}$. Then A is nonsingular if and only if $\mathcal{N}(A) = \{\mathbf{0}\}$.*

Proof. Suppose A is nonsingular. By definition, A has rank n, so $A_{red} = I_n$. This implies $\mathcal{N}(A) = \mathcal{N}(A_{red}) = \{\mathbf{0}\}$. Conversely, if $\mathcal{N}(A) = \{\mathbf{0}\}$, then $A\mathbf{x} = \mathbf{0}$ implies $\mathbf{x} = \mathbf{0}$. Hence there cannot be any free variables, so by the identity (3.11), the rank of A is n. Thus A is nonsingular. □

Remark. In the next chapter, we will develop some techniques for bringing a matrix into reduced row echelon form that are useful for solving linear systems.

Exercises

Exercises 3.3.1. Show that the null space of a matrix over a field is an abelian group.

Exercises 3.3.2. Let C denote the counting matrix of Example 3.8. Find an equation in a, b, c that determines when the system $C\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} a \\ b \\ c \end{pmatrix}$ is consistent. (Hint: row reduce.)

Exercises 3.3.3. The field is $\mathbb{F}_2$. Consider the following matrices:

$$A_1 = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix}, \quad A_2 = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}.$$

Find basic null vectors for both A_1 and A_2.

Exercises 3.3.4. Which of the following can happen and which cannot? Provide explanations.

(*i*) A is of size 15×24 and has exactly seven basic null vectors;

(*ii*) A is of size 3×3, and $A\mathbf{x} = \mathbf{b}$ is consistent for all $\mathbf{b}$.

(*iii*) A is of size 12×12, three of A's columns are zero, and A has 10 basic null vectors.

Chapter 4
Matrix Inverses, Matrix Groups and the *LPDU* Decomposition

In this chapter we continue our introduction to matrix theory beginning with the notion of a matrix inverse and the definition of a matrix group. For now, the main example of a matrix group is the group $GL(n, \mathbb{F})$ of invertible $n \times n$ matrices over a field $\mathbb{F}$ and its subgroups. We will also show that every matrix $A \in \mathbb{F}^{n \times n}$ can be factored as a product $LPDU$, where each of L, P, D, and U is a matrix in an explicit subset of $\mathbb{F}^{n \times n}$. For example, P is a partial permutation matrix, D is diagonal, and L and U are lower and upper triangular respectively. The expression $A = LPDU$ tells us a lot about A: for example, D contains the pivots d_{ii} of A and also tells us the rank of A by counting the number of nonzero pivots. In addition, A's determinant (which will be defined in the next chapter) is determined by PD as $\pm d_{11} \cdots d_{nn}$. When A is invertible, each of L, P, D, and U lies in a certain subgroup of $GL(n, \mathbb{F})$. For example, P lies in the group of $n \times n$ permutation matrices, which is a matrix group isomorphic to $S(n)$.

4.1 The Inverse of a Square Matrix

We now come to the interesting question of when an $n \times n$ matrix over a field $\mathbb{F}$ has an inverse under matrix multiplication. Since every nonzero element of a field has an inverse, the question for 1×1 matrices is already settled. The main fact turns out to be that an $n \times n$ matrix over $\mathbb{F}$ is invertible if and only if its rank is n.

4.1.1 The definition of the inverse

We begin with an essential definition:

J.B. Carrell, *Groups, Matrices, and Vector Spaces*,
DOI 10.1007/978-0-387-79428-0_4

Definition 4.1. Let $\mathbb{F}$ be a field and suppose $A \in \mathbb{F}^{n\times n}$. We say that A is *invertible* if there exists $B \in \mathbb{F}^{n\times n}$ such that $AB = BA = I_n$. We will say that B is an *inverse* of A, and likewise, that A is an inverse of B.

Example 4.1. All 2×2 elementary matrices are invertible. Their inverses are given as follows:

$$E_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \Rightarrow E_1^{-1} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

$$E_2 = \begin{pmatrix} r & 0 \\ 0 & 1 \end{pmatrix} \Rightarrow E_2^{-1} = \begin{pmatrix} r^{-1} & 0 \\ 0 & 1 \end{pmatrix},$$

and

$$E_3 = \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix} \Rightarrow E_3^{-1} = \begin{pmatrix} 1 & -s \\ 0 & 1 \end{pmatrix}.$$

⊠

In fact, we already noted that if $E \in \mathbb{F}^{n\times n}$ is an elementary matrix and F is the elementary matrix that reverses the row operation that E effects, then $FE = EF = I_n$. In other words, doing a row operation and then undoing it produces the same result as first undoing it and then doing it, and the result is that nothing changes. Thus every elementary matrix is invertible. Note that the definition can apply only to square matrices, but if A is square, we can consider both left and right inverses of A. A *left inverse* of $A \in \mathbb{F}^{n\times n}$ is a matrix $B \in \mathbb{F}^{n\times n}$ such that $BA = I_n$, and a *right inverse* of A is a matrix $B \in \mathbb{F}^{n\times n}$ such that $AB = I_n$. It is useful to note (as was already noted for groups) that when an inverse exists, it is unique. This is one of the nice consequences of the associativity of matrix multiplication we have previously mentioned.

Proposition 4.1. *An invertible $n \times n$ matrix has a unique inverse.*

Proof. Suppose $A \in \mathbb{F}^{n\times n}$ has two inverses B and C. Then $B = BI_n = B(AC) = (BA)C = I_nC = C$. Thus $B = C$. □

4.1.2 Results on Inverses

Elementary matrices are the key to understanding when inverses exist and how to find them. As we already noted above, if $E \in \mathbb{F}^{n\times n}$ is elementary, then there exists an elementary matrix $F \in \mathbb{F}^{n\times n}$ such that $FE = EF = I_n$. Thus elementary matrices are invertible. In fact, the inverse of an elementary matrix has the same type. Our first result about inverses is that a nonsingular

matrix (that is, an element of $\mathbb{F}^{n\times n}$ of rank n) is invertible. Recall that by Proposition 3.13, $A \in \mathbb{F}^{n\times n}$ is nonsingular if and only if there exist elementary matrices $E_1, \ldots, E_k \in \mathbb{F}^{n\times n}$ such that $E_k \cdots E_1 A = I_n$. In particular, if A is nonsingular, we can assert that A has a left inverse.

Proposition 4.2. *Let A in $\mathbb{F}^{n\times n}$. If A is nonsingular, then A is invertible. In fact, let $E_1, \ldots, E_k \in \mathbb{F}^{n\times n}$ be elementary matrices such that $E_k \cdots E_1 A = I_n$, and let $F_1, \ldots, F_k \in \mathbb{F}^{n\times n}$ denote the elementary matrices such that $F_i E_i = I_n$. Then $A = F_1 \cdots F_k$ and $A^{-1} = E_k \cdots E_1$. Consequently, if A is nonsingular, then A^{-1} is also nonsingular.*

Proof. Let $B = E_k \cdots E_1$ and $C = F_1 \cdots F_k$. Then by associativity of multiplication,

$$CB = (F_1 \cdots F_k)(E_k \cdots E_1) = I_n,$$

since $F_i E_i = I_n$ for each i. Now $BA = I_n$ by assumption, so

$$A = I_n A = (CB)A = C(BA) = CI_n = C.$$

Thus $A = F_1 \cdots F_k$, as claimed. Since $CB = I_n$ and $A = C$, we have $AB = I_n$, which proves that A is invertible. We leave it to the reader to show that if A is nonsingular, then so is A^{-1}. □

Therefore, nonsingular matrices are invertible. In fact, the converse is also true, which will be important in our discussion of matrix groups.

Theorem 4.3. *Let $A \in \mathbb{F}^{n\times n}$. Then A is nonsingular if and only if A is invertible if and only if A is a product of elementary matrices.*

Proof. By Proposition 4.2, a nonsingular matrix is invertible. To show that an invertible matrix is nonsingular, let A be invertible, and suppose $BA = I_n$. Now suppose $A\mathbf{x} = \mathbf{0}$. Then $\mathbf{x} = (BA)\mathbf{x} = B(A\mathbf{x}) = B\mathbf{0} = \mathbf{0}$, so by Proposition 3.18, it follows that A is nonsingular. If A is nonsingular, then by Proposition 4.2, A is a product of elementary matrices. Conversely, a product of elementary matrices is invertible, so the proof is finished. □

Notice that all that is required in the proof of the above theorem is that A have a left inverse B. To expand on this observation, we will prove one last result.

Proposition 4.4. *Let $A \in \mathbb{F}^{n\times n}$. Then if A has either a left inverse B (that is, $BA = I_n$) or a right inverse C (that is, $AC = I_n$), then A is invertible.*

Proof. Suppose A has a left inverse B. Arguing as in the proof of Theorem 4.3, we conclude that $A\mathbf{x} = \mathbf{0}$ implies $\mathbf{x} = \mathbf{0}$, so A is nonsingular. Therefore, A is invertible. Now suppose A has a right inverse C. Then C has a left inverse A, so C is invertible. Thus, A is invertible too. □

Consequently, the invertible $n \times n$ matrices over $\mathbb{F}$ are exactly those for which there is either a left or right inverse. This means that in an actual calculation of A^{-1}, it is necessary only to find a B such that either AB or BA is I_n. One final result is the following.

Proposition 4.5. *If $A, B \in \mathbb{F}^{n\times n}$ are both invertible, then so is AB. Moreover, $(AB)^{-1} = B^{-1}A^{-1}$. In addition, A^T is also invertible, and $(A^T)^{-1} = (A^{-1})^T$.*

Proof. Both A and B are products of elementary matrices, so it follows that AB is too. Therefore AB is also invertible. We leave it to the reader to check the formula for $(AB)^{-1}$. The assertion about A^T is left as an exercise. □

4.1.3 Computing inverses

We now consider the problem of actually inverting an $n \times n$ matrix A. One method is the row reduction procedure used in Example 3.10. This starts with the $n \times 2n$ matrix $(A \mid I_n)$ and applies row reduction until the matrix $(I_n \mid B)$ is obtained. Then $BA = I_n$, so $B = A^{-1}$ by Proposition 4.4. Otherwise, A is singular and A^{-1} doesn't exist. Alternatively, one can also find B by multiplying out a sequence of elementary matrices that row reduces A. This is actually not as bad as it sounds, since multiplying elementary matrices is elementary. In the following example, we will assume that the field is $\mathbb{F}_2$. Not surprisingly, this assumption makes the calculations a lot easier.

Example 4.2. Suppose the field is $\mathbb{F}_2$. Let us find the inverse of

$$A = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix},$$

if one exists. Now,

$$(A \mid I_3) = \begin{pmatrix} 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \end{pmatrix} \to \begin{pmatrix} 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \end{pmatrix} \to$$

$$\begin{pmatrix} 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 1 & 0 \end{pmatrix} \to \begin{pmatrix} 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 & 0 \end{pmatrix} \to$$

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

Hence

$$A^{-1} = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 0 \end{pmatrix}.$$

⊠

Example 4.3. For a slightly less simple example, let $\mathbb{F} = \mathbb{F}_2$, but put

$$A = \begin{pmatrix} 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix}.$$

Following the above procedure, we obtain that

$$A^{-1} = \begin{pmatrix} 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \end{pmatrix}.$$

Note that the correctness of this result should be checked by computing directly that

$$I_4 = \begin{pmatrix} 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix}.$$

⊠

There is a slightly less obvious third technique for inverting a matrix. If A is of size $n \times n$ and we form the matrix $(A \mid \mathbf{x})$, where $\mathbf{x}$ represents a variable column vector with components $x_1, x_2, \ldots, x_n$, then row reducing will produce a result of the form $(A_{red} \mid \mathbf{c})$, where the components of $\mathbf{c}$ are certain linear combinations of the components of $\mathbf{x}$. The coefficients in these linear combinations turn out to be the entries of the matrix B such that $BA = A_{red}$. Here is an example.

Example 4.4. Let the field be $\mathbb{Q}$ and

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

Now form

$$\begin{pmatrix} 1 & 2 & 0 & a \\ 1 & 3 & 1 & b \\ 0 & 1 & 2 & c \end{pmatrix}$$

and row reduce. The result is

$$\begin{pmatrix} 1 & 0 & 0 & 5a-4b+2c \\ 0 & 1 & 0 & -2a+2b-c \\ 0 & 0 & 1 & a-b+c \end{pmatrix}.$$

Thus,

$$A^{-1} = \begin{pmatrix} 5 & -4 & 2 \\ -2 & 2 & -1 \\ 1 & -1 & 1 \end{pmatrix}. \qquad \boxtimes$$

Exercises

Exercise 4.1.1. Find the inverse of each of the following matrices over $\mathbb{Q}$, or show that the inverse does not exist:

(a) $\begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & -1 \\ 1 & 1 & 0 \end{pmatrix}$; (b) $\begin{pmatrix} 1 & 0 & -2 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \end{pmatrix}$; (c) $\begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & -1 \\ 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix}$.

Exercise 4.1.2. If possible, invert

$$B = \begin{pmatrix} 1 & 2 & -1 & -1 \\ -2 & -1 & 3 & 1 \\ -1 & 4 & 3 & -1 \\ 0 & 3 & 1 & -1 \end{pmatrix}.$$

Exercise 4.1.3. We saw that the 3×3 counting matrix is singular. Determine whether the 2×2 and 4×4 counting matrices are nonsingular.

Exercise 4.1.4. Consider the matrix of Exercise 4.1.2 as a matrix A over $\mathbb{F}_5$. If possible, find A^{-1}.

Exercise 4.1.5. The following matrices are over $\mathbb{F}_2$. Determine which have inverses and find the inverses when they exist.

(a) $\begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \end{pmatrix}$; (b) $\begin{pmatrix} 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix}$; (c) $\begin{pmatrix} 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix}$.

Exercise 4.1.6. If possible, invert the following 5×5 matrix A over $\mathbb{F}_2$:

$$A = \begin{pmatrix} 1 & 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 & 1 \end{pmatrix}.$$

Exercise 4.1.7. Suppose $A = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$, and assume that $\Delta = ad - bc \neq 0$. Show that $A^{-1} = \frac{1}{\Delta}\left(\begin{smallmatrix} d & -b \\ -c & a \end{smallmatrix}\right)$. What does the condition $\Delta \neq 0$ mean in terms of the rows of A?

Exercise 4.1.8. Verify that if A is invertible, then A^T is also invertible. Find a formula for $(A^T)^{-1}$ and verify it.

Exercise 4.1.9. Suppose A is invertible.

(*i*) Show that A^{-1} is also invertible.

(*ii*) Find a formula for the inverse of A^m, where m is a positive integer.

(*iii*) True or false: $A + A^T$ is invertible. Include brief reasoning.

Exercise 4.1.10. True or false: If A is square and AA^T is invertible, then A is invertible. What if A isn't square?

Exercise 4.1.11. True or false: Suppose A is of size $n \times n$ and $A^3 + 2A - I_n = O$. Then A invertible. Include brief reasoning.

Exercise 4.1.12. True or false: if an $n \times n$ matrix with integer entries has an inverse, then the inverse also has integer entries. As usual, include brief reasoning.

Exercise 4.1.13. Let $C = \left(\begin{smallmatrix} 1 & a & b \\ 0 & 1 & c \\ 0 & 0 & 1 \end{smallmatrix}\right)$. Find a general formula for C^{-1}.

Exercise 4.1.14. Show that if A and B are of size $n \times n$ and have inverses, then $(AB)^{-1} = B^{-1}A^{-1}$. What is $(ABCD)^{-1}$ if all four matrices are invertible?

Exercise 4.1.15. Suppose A is an invertible $m \times m$ matrix and B is an $m \times n$ matrix. Solve the equation $AX = B$.

Exercise 4.1.16. Suppose A and B are both of size $n \times n$ and AB is invertible. Show that both A and B are invertible.

Exercise 4.1.17. Let A and B be two $n \times n$ matrices over $\mathbb{R}$. Suppose $A^3 = B^3$ and $A^2B = B^2A$. Show that if $A^2 + B^2$ is invertible, then $A = B$. (Hint: consider $(A^2 + B^2)A$.)

Exercise 4.1.18. Let A and B be $n \times n$ matrices over $\mathbb{F}$.

(i) Show that if the inverse of A^2 is B, then the inverse of A is AB.

(ii) Suppose A, B, and $A + B$ are all invertible. Find the inverse of $A^{-1} + B^{-1}$ in terms of A, B, and $A + B$.

Exercise 4.1.19. You are a cryptographer assigned to crack the clever cipher constructed as follows. First, let the sequence 01 represent A, 02 represent B, and so forth up to 26, which represents Z. For clarity, a space between words is indicated by inserting 00. A text can thus be unambiguously encoded as a sequence. For example, 1908040002090700041507 is the encoding of the phrase "the big dog." Since this string of integers has length 22, we will think of it as a vector in $\mathbb{Q}^{22}$. Suppose a text has been encoded as a sequence of length 14,212. Now, $14{,}212 = 44 \times 323$, so the sequence can be broken into 323 consecutive intervals of length 44. Next, suppose each subinterval is transposed and multiplied on the left by a single 44×44 matrix C. The new sequence obtained by transposing again and laying the products end to end will be the enciphered message, and it is your job to decipher it. Discuss the following questions.

(i) How does one produce an invertible 44×44 matrix in an efficient way, and how does one find its inverse?

(ii) How many of the subintervals will you need to decipher to break the whole cipher by deducing the matrix C?

4.2 Matrix Groups

The purpose of this section is to introduce the concept of a matrix group and produce a number of examples. A matrix group is essentially a collection of matrices that lie in some $\mathbb{F}^{n\times n}$ that forms a group under matrix multiplication, where the identity is I_n. Thus the axioms defining a matrix group are modest, but they are enough to ensure that matrix groups form an interesting central part of algebra. In the final chapter, we will outline the theory of the structure of matrix groups that are defined by some equations. These groups are known as linear algebraic groups. The matrix group structure will be useful in stating and proving some of the main results on matrices in later chapters.

4.2.1 The definition of a matrix group

We begin with the main definition.

Definition 4.2. Let $\mathbb{F}$ be a field. A subset G of $\mathbb{F}^{n\times n}$ is called a *matrix group* if it satisfies the following three conditions:

(*i*) if $A, B \in G$, then $AB \in G$ (that is, G is closed under multiplication);

(*ii*) $I_n \in G$; and

(*iii*) if $A \in G$, then A is invertible, and $A^{-1} \in G$.

Remark. Since matrix multiplication is associative, every matrix group G is a group under matrix multiplication with identity the identity matrix, and the inverse of every $A \in G$ given by A^{-1}.

The most basic example of a matrix group $G \subset \mathbb{F}^{n\times n}$ is the *general linear group over* $\mathbb{F}$, which is denoted by $GL(n, \mathbb{F})$. By definition,

$$GL(n, \mathbb{F}) = \{A \in \mathbb{F}^{n\times n} \mid A^{-1} \text{ exists}\}. \tag{4.1}$$

Thus, $GL(n, \mathbb{F})$ is the set of all invertible elements of $\mathbb{F}^{n\times n}$.

Proposition 4.6. *The set $GL(n, \mathbb{F})$ is a matrix group. Moreover, every matrix group $G \subset \mathbb{F}^{n\times n}$ is a subgroup of $GL(n, \mathbb{F})$.*

Proof. By Proposition 4.5, $GL(n, \mathbb{F})$ is closed under multiplication. Moreover, $I_n \in GL(n, \mathbb{F})$. Finally, if A is invertible, so is A^{-1}. Therefore, $GL(n, \mathbb{F})$ is a matrix group. □

A subgroup of a matrix group is called a *matrix subgroup*. For example, $\{I_n\}$ is a subgroup of every matrix group $G \subset GL(n, \mathbb{F})$. One usually calls

$\{I_n\}$ the trivial subgroup. The other basic definitions that apply to groups such as normal subgroups, cosets, order (in the finite group case) all apply without change to matrix groups.

4.2.2 Examples of matrix groups

We now give some examples of matrix groups.

Example 4.5 (Rotations). The first example is the group Rot(2) consisting of the 2×2 rotation matrices R_θ, $0 \leq \theta \leq 2\pi$. As we mentioned in Example 3.7, the mapping $\mathcal{R}_\theta : \mathbb{R}^2 \to \mathbb{R}^2$ defined by

$$\mathcal{R}_\theta \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x\cos\theta - y\sin\theta \\ x\sin\theta + y\cos\theta \end{pmatrix}$$

is the rotation of $\mathbb{R}^2$ through θ. Note that $\mathcal{R}_\theta$ is the matrix mapping associated to the matrix

$$R_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}, \tag{4.2}$$

which we will call a rotation matrix. Recall that in Example 3.7 we showed that $R_\theta \begin{pmatrix} x \\ y \end{pmatrix}$ is the same action on $\mathbb{R}^2$ as multiplication by the complex exponential $e^{i\theta}$ is on $\mathbb{C}$. That is,

$$R_\theta \begin{pmatrix} x \\ y \end{pmatrix} = e^{i\theta} z,$$

where $z = x + yi$. Since $e^{i(\theta+\psi)} = e^{i\theta}e^{i\psi}$, it follows that $R_{\theta+\psi} = R_\theta R_\psi$ for all θ, ψ. Consequently, Rot(2) is closed under multiplication. It is also closed under taking inverses. For using the formula for the inverse of a 2×2 matrix in Exercise 4.1.7, we have

$$(R_\theta)^{-1} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} = \begin{pmatrix} \cos(-\theta) & -\sin(-\theta) \\ \sin(-\theta) & \cos(-\theta) \end{pmatrix} = R_{-\theta}.$$

Finally, note that $I_2 = R_0$. Hence, Rot(2) is a matrix group. In addition, Rot(2) is abelian. For

$$R_\theta R_\psi = \begin{pmatrix} \cos\theta\cos\psi - \sin\theta\sin\psi & -\cos\theta\sin\psi - \sin\theta\cos\psi \\ \cos\theta\sin\psi + \sin\theta\cos\psi & \cos\theta\cos\psi - \sin\theta\sin\psi \end{pmatrix} = R_\psi R_\theta,$$

as can be seen by noticing that each term in the matrix stays the same after flipping θ and ψ. This also follows from the observation that $e^{i\theta}e^{i\psi} = e^{i\theta}e^{i\theta}$, since $\mathbb{C}^*$ is an abelian group. This is worth noting, since among other reasons, abelian matrix groups are rather rare. Another fact to notice is that there is an explicit group isomorphism $\phi : S^1 \to \mathrm{Rot}(2)$, which is defined by first setting $z = e^{i\theta}$ and then putting $\phi(z) = R_\theta$. We leave it as an exercise to show that ϕ is, as claimed, a well-defined isomorphism. ⊠

The next example of a matrix group plays a fundamental role in the Euclidean geometry of $\mathbb{R}^n$.

Definition 4.3 (The orthogonal group). Let $Q \in \mathbb{R}^{n\times n}$. We say that Q is *orthogonal* if $Q^TQ = I_n$ (i.e., $Q^{-1} = Q^T$). The set of all $n \times n$ orthogonal matrices is called the *orthogonal group of degree n*. The orthogonal group is denoted by $O(n, \mathbb{R})$.

Thus, $Q \in \mathbb{R}^{n\times n}$ is orthogonal if and only if when $Q = (\mathbf{q}_1\ \mathbf{q}_2\ \cdots\ \mathbf{q}_n)$, then $\mathbf{q}_i^T\mathbf{q}_j = \delta_{ij}$ for all $1 \le i, j \le n$.

Proposition 4.7. *$O(n, \mathbb{R})$ is a subgroup of $GL(n, \mathbb{R})$.*

Proof. We first show that if $Q, R \in O(n, \mathbb{R})$, then $QR \in O(n, \mathbb{R})$. We have to check that $(QR)^T(QR) = I_n$. But

$$(QR)^T(QR) = (R^TQ^T)(QR) = R^T(Q^TQ)R = R^TI_nR = I_n,$$

so $O(n, \mathbb{R})$ is closed under multiplication. Clearly $I_n \in O(n, \mathbb{R})$, so it remains to verify that $Q \in O(n, \mathbb{R})$ implies $Q^{-1} \in O(n, \mathbb{R})$. But $Q^{-1} = Q^T$, so we have to check that $(Q^T)^TQ^T = I_n$. This amounts to showing that $QQ^T = I_n$, which holds due to the fact that $Q^TQ = I_n$. This completes the verification. □

The alert reader may have noticed that the definition of the orthogonal group $O(n, \mathbb{R})$ actually had nothing to do with $\mathbb{R}$. That is, we could just as easily defined $O(n, \mathbb{F})$ in exactly the same way for any field $\mathbb{F}$. In fact, the matrix groups of the type $O(n, \mathbb{F})$ form an important class known as the *orthogonal groups over* $\mathbb{F}$. One frequently concentrates on $O(n, \mathbb{R})$ because its properties are related to the geometry of n-dimensional Euclidean space.

In the following section, we will give an example of a matrix group closely related to the symmetric group.

4.2.3 The group of permutation matrices

Recall that $\mathbf{e}_i$ is the column vector such that $I_n = (\mathbf{e}_1\ \cdots\ \mathbf{e}_n)$. If $\sigma \in S(n)$, put $P_\sigma = (\mathbf{e}_{\sigma(1)}\ \cdots\ \mathbf{e}_{\sigma(n)})$. Thus,

$$P_\sigma \mathbf{e}_k = \mathbf{e}_{\sigma(k)}.$$

Matrices of the form P_σ are called $n \times n$ *permutation matrices*. Let $P(n)$ denote the set of all $n \times n$ permutation matrices. Note that different permutations in $S(n)$ give rise to different permutation matrices. That is, if $\sigma \neq \mu$, then $P_\sigma \neq P_\mu$. Thus there are exactly $n!$ $n \times n$ permutation matrices.

Let us first show that row-swap matrices are the permutation matrices corresponding to transpositions. Suppose S is the $n \times n$ row-swap matrix that interchanges rows i and j, where $i < j$. Let $\tau \in S(n)$ be the transposition interchanging i and j. Thus $\tau(i) = j$, $\tau(j) = i$, and $\tau(k) = k$ for $k \neq i, j$. Then $S\mathbf{e}_i = \mathbf{e}_j = \mathbf{e}_{\tau(i)}$ and $S\mathbf{e}_j = \mathbf{e}_i = \mathbf{e}_{\tau(j)}$. Also, for $k \neq i, j$, $S\mathbf{e}_k = \mathbf{e}_k$. Therefore, $S = P_\tau$, as asserted. Now we want to study the relationship between the permutation matrices and $S(n)$. The key fact is contained in the following proposition.

Proposition 4.8. *Let $\sigma, \tau \in S(n)$. Then we have*

$$P_\tau P_\sigma = P_{\tau\sigma}.$$

Therefore, $P(n)$ is closed under multiplication. Moreover, I_n and $(P_\sigma)^{-1}$ are permutation matrices for all $\sigma \in S(n)$. Consequently, $P(n)$ is a matrix group.

Proof. For each index k, $1 \leq k \leq n$, we have

$$P_\tau P_\sigma \mathbf{e}_k = P_\tau \mathbf{e}_{\sigma(k)} = \mathbf{e}_{\tau(\sigma(k))} = P_{\tau\sigma}\mathbf{e}_k.$$

Hence $P_\tau P_\sigma = P_{\tau\sigma}$, so $P(n)$ is closed under multiplication. It also follows from this formula that $(P_\sigma)^{-1} = P_\tau$, where $\tau = \sigma^{-1}$, so every $P \in P(n)$ has its inverse in $P(n)$. Since $P_{id_n} = I_n$, where id_n is the identity, it follows that $P(n)$ is a matrix group. □

Now let $\varphi : S(n) \to P(n)$ be the mapping defined by $\varphi(\sigma) = P_\sigma$.

Proposition 4.9. *φ is an isomorphism.*

Proof. We already noted that φ is injective. Since $S(n)$ and $P(n)$ both have order $n!$, it follows that φ is a bijection. Therefore, φ is an isomorphism. □

Every permutation matrix P can be put into reduced row echelon form. But it is clear that type II row operations suffice for this, so P can be written as a product of row swaps. This implies the following result.

Corollary 4.10. *Every $\sigma \in S(n)$ can be written as a product of transpositions. These transpositions may be found by putting P_σ into reduced row echelon form using row swaps.*

Note, however, that there are in general many ways of writing a permutation matrix as a product of row swaps, and correspondingly, there are many ways of writing a permutation as a product of transpositions. In fact, one can always represent a permutation matrix as a product of row swaps that interchange adjacent rows. These correspond to transpositions of the form $(i\ i+1)$ and are called *simple transpositions*. The simple transpositions are fundamental in the study of the combinatorial properties of $S(n)$. The upshot is the following.

Proposition 4.11. *Every $\sigma \in S(n)$ can be expressed as a product of simple transpositions.*

Permutation matrices have the following beautiful property.

Proposition 4.12. *If $P \in P(n)$, then $P^{-1} = P^T$. In other words, every permutation matrix is orthogonal.*

We will leave the proof as an exercise. Thus, $P(n)$ is a subgroup of $O(n, \mathbb{R})$. This gives an interesting example of an infinite group, namely $O(n, \mathbb{R})$, containing a finite subgroup, which generalizes the example $\{\pm 1\} \subset O(1, \mathbb{R})$.

Example 4.6. For instance, $P(2)$ consists of two matrices, I_2 and

$$S = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

while $P(3)$ consists of the following six 3×3 matrices;

$$I_3,\quad \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix},\quad \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix},\quad \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix},\quad \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix},\quad \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}.$$

The second, third, and sixth matrices are row swaps, and the others are products of two row swaps. ⊠

Exercises

Exercise 4.2.1. Find the center of $GL(n, \mathbb{F})$ for an arbitrary field $\mathbb{F}$.

Exercise 4.2.2. Find the center of $O(n, \mathbb{F})$ for an arbitrary field $\mathbb{F}$.

Exercise 4.2.3. Show that $\mathrm{Rot}(2)$ is a subgroup of $O(2, \mathbb{R})$. Is $O(2, \mathbb{R}) = \mathrm{Rot}(2)$? If not, find an element of $O(2, \mathbb{R})$ that isn't a rotation.

Exercise 4.2.4. Let $G \subset GL(2, \mathbb{R})$ denote the set of all matrices $\begin{pmatrix} a & b \\ -b & a \end{pmatrix}$, where $a^2 + b^2 \neq 0$. Is G an abelian subgroup of $GL(2, \mathbb{R})$?

Exercise 4.2.5. Show that $Q \in \mathbb{R}^{n\times n}$ is orthogonal if and only if its rows $\mathbf{q}_1, \ldots, \mathbf{q}_n$ satisfy the condition $\mathbf{q}_i\mathbf{q}_j^T = \delta_{ij}$. That is, $\mathbf{q}_i\mathbf{q}_i^T = 1$, while $\mathbf{q}_i\mathbf{q}_j^T = 0$ if $i \neq j$.

Exercise 4.2.6. Show that if $Q \in \mathbb{R}^{n\times n}$ is orthogonal, then so is Q^T. Conclude that a symmetric orthogonal matrix satisfies $Q^2 = I_n$; hence Q is its own inverse. Give an example of a 2×2 symmetric orthogonal matrix different from I_2.

Exercise 4.2.7. Without computing, try to guess the inverse of the matrix

$$A = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & -1 \\ 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix}.$$

(Hint: consider the columns.)

Exercise 4.2.8. Let $S_1 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ and $S_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$. Show that every 3×3 permutation matrix is a product involving only S_1 and S_2.

Exercise 4.2.9. Let S_1 and S_2 be the permutation matrices defined in Exercise 4.2.8. Show that $(S_1S_2)^3 = I_3$.

Exercise 4.2.10. Show that every permutation matrix P satisfies the identity $P^{-1} = P^T$, and conclude that $P(n)$ is a subgroup of $O(n, \mathbb{R})$.

Exercise 4.2.11. Show that the following two matrices are permutation matrices and find their inverses:

$$\begin{pmatrix} 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{pmatrix}, \qquad \begin{pmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}.$$

Exercise 4.2.12. A *signed permutation matrix* is a square matrix Q of the form

$$Q = P(\pm\mathbf{e}_1 \pm \mathbf{e}_2 \cdots \pm \mathbf{e}_n),$$

where $P \in P(n)$.

(*i*) Show that the set $SP(n)$ of $n \times n$ signed permutation matrices is a subgroup of $O(n, \mathbb{R})$ of order $2^n n!$.

(*ii*) Prove that $P(n)$ is a normal subgroup of $SP(n)$.

(*iii*) Describe the quotient group $SP(n)/P(n)$.

Exercise 4.2.13. Prove that every finite group G of order n is isomorphic to a subgroup of $O(n, \mathbb{R})$.

4.3 The *LPDU* Factorization

Recall that an invertible $n \times n$ matrix A can be expressed as a product of elementary $n \times n$ matrices. In this section, we will prove a much more explicit result: every $n \times n$ matrix A over a field $\mathbb{F}$ can be expressed in the form $A = LPDU$, where each of the matrices L, P, D, and U is built up from a single type of elementary matrix. This $LPDU$ factorization is one of the most basic tools in the theory of matrices. One nice application is a well-known but nontrivial relationship between the ranks of A and A^T for any matrix A, which we prove below. The $LPDU$ decomposition is widely used in applied linear algebra for solving large systems of linear equations.

4.3.1 The basic ingredients: L, P, D, and U

The list of characters in the $LPDU$ decomposition consists of matrices that we have met before and matrices that we need to introduce. Let us begin with L and U.

Definition 4.4. An $n \times n$ matrix L is called *lower triangular* if $l_{ij} = 0$ for all $j > i$. That is, the nonzero entries of L are below or on the diagonal of L. Similarly, an $n \times n$ matrix U is *upper triangular* if $u_{ij} = 0$ for $i > j$. A matrix that is either lower or upper triangular is said to be *unipotent* if all its diagonal entries are 1.

Clearly, the transpose of a lower triangular matrix is upper triangular. The transpose of a unipotent matrix is also unipotent. In our discussion, the matrices L and U will always be lower and upper triangular and both will be unipotent.

Example 4.7. A lower triangular 3×3 unipotent matrix has the form

$$L = \begin{pmatrix} 1 & 0 & 0 \\ a & 1 & 0 \\ b & c & 1 \end{pmatrix}.$$

The transpose U of L is, of course,

$$U = \begin{pmatrix} 1 & a & b \\ 0 & 1 & c \\ 0 & 0 & 1 \end{pmatrix},$$

which is upper triangular. One can easily check that

$$L^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ -a & 1 & 0 \\ ac-b & -c & 1 \end{pmatrix}.$$

Thus L^{-1} is also lower triangular unipotent. ⊠

Recall that type III row operations are also called transvections. When a lower row is replaced by itself plus a multiple of some higher row, a transvection is said to be downward. Downward transvections are performed via left multiplication by lower triangular unipotent matrices. Similarly, rightward transvections are performed by right multiplication by upper triangular matrices. Here is a basic fact.

Proposition 4.13. *Let $\mathcal{L}_n(\mathbb{F})$ and $\mathcal{U}_n(\mathbb{F})$ denote respectively the set of all lower triangular unipotent and upper triangular unipotent $n \times n$ matrices over $\mathbb{F}$. Then $\mathcal{L}_n(\mathbb{F})$ and $\mathcal{U}_n(\mathbb{F})$ are matrix subgroups of $GL(n, \mathbb{F})$ for all $n > 0$.*

Proof. It follows from the definition of matrix multiplication that the product of two lower triangular matrices is also lower triangular. If A and B are lower triangular unipotent, then the diagonal entries of AB are also all 1. Indeed, if $AB = (c_{ij})$, then

$$c_{ii} = \sum_{k=1}^{n} a_{ik} b_{ki} = a_{ii} b_{ii} = 1,$$

since $a_{ij} = b_{ij} = 0$ if $i < j$. The identity I_n is also lower triangular unipotent, so to show that $\mathcal{L}_n(\mathbb{F})$ is a subgroup of $GL(n, \mathbb{F})$, it remains to show that the inverse of an element A of $\mathcal{L}_n(\mathbb{F})$ is also in $\mathcal{L}_n(\mathbb{F})$. But this follows from the explicit technique for inverting a matrix using row operations. Row swaps are never needed, since A is already lower triangular. Row dilations are never needed, since A is already unipotent. Thus, A^{-1} is obtained by a sequence of downward transvections. But these correspond to taking products in $\mathcal{L}_n(\mathbb{F})$, so $A^{-1} \in \mathcal{L}_n(\mathbb{F})$. The result for $\mathcal{U}_n(\mathbb{F})$ is proved in an analogous way. In fact, one can simply transpose the above proof. □

Continuing with the introduction of the basic ingredients, we now concentrate on P and D. We next describe P.

Definition 4.5. A *partial permutation matrix* is a matrix that is obtained from a permutation matrix P by setting some of the rows of P equal to zero.

Let Π_n denote the set of $n \times n$ partial permutation matrix matrices. To be explicit, every element of Π_n is either a permutation matrix or obtained from a permutation matrix P by replacing some of the ones in P by zeros. A matrix with a row of zeros is, of course, singular, so Π_n is not a matrix group. Nevertheless, it remains true that the product of two $n \times n$ partial permutation matrices is also a partial permutation matrix.

Lastly, we describe D.

Definition 4.6. A *diagonal matrix* is a square matrix $D = (d_{ij})$ all of whose off-diagonal entries are zero; that is, $d_{ij} = 0$ for all $i \neq j$.

Since a diagonal matrix D is invertible if and only if its diagonal entries d_{ii} are all different from 0 and the product of two diagonal matrices is also diagonal, we obtain the following result.

Proposition 4.14. *The set $\mathcal{D}_n(\mathbb{F})$ of all $n \times n$ invertible diagonal matrices over $\mathbb{F}$ is a matrix subgroup of $GL(n, \mathbb{F})$.*

4.3.2 The main result

We now derive the *LPDU* decomposition.

Theorem 4.15. *Every $n \times n$ matrix A over a field $\mathbb{F}$ can be expressed in the form $A = LPDU$, where $L \in \mathcal{L}_n(\mathbb{F})$, $P \in \Pi_n$, $D \in \mathcal{D}_n(\mathbb{F})$, and $U \in \mathcal{U}_n(\mathbb{F})$. The partial permutation matrix P is always unique. If A is invertible, then P is a full permutation matrix, and in that case, P and D are both unique.*

This result can be expressed in the form of a product of sets, three of which are matrix groups:

$$\mathbb{F}^{n\times n} = \mathcal{L}_n(\mathbb{F}) \cdot \Pi_n \cdot \mathcal{D}_n(\mathbb{F}) \cdot \mathcal{U}_n(\mathbb{F}).$$

Thus every $n \times n$ matrix over $\mathbb{F}$ is the product of the four types of matrices $\mathcal{L}_n(\mathbb{F})$, Π_n, $\mathcal{D}_n(\mathbb{F})$, and $\mathcal{U}_n(\mathbb{F})$. Specializing to invertible matrices, we have

$$GL(n, \mathbb{F}) = \mathcal{L}_n(\mathbb{F}) \cdot P(n) \cdot \mathcal{D}_n(\mathbb{F}) \cdot \mathcal{U}_n(\mathbb{F}).$$

In particular, $GL(n, \mathbb{F})$ is the product of four of its subgroups: $\mathcal{L}_n(\mathbb{F})$, $P(n)$, $\mathcal{D}_n(\mathbb{F})$, and $\mathcal{U}_n(\mathbb{F})$. This is a fundamental result in the theory of matrix groups.

Remark. It is also possible to define an *LPDU* decomposition for nonsquare matrices. If $A \in \mathbb{F}^{m\times n}$, where say $m < n$, then we can augment A by adding $n - m$ rows of zeros to make A an $n \times n$ matrix. Of course, in this situation, the last $n - m$ rows of P will also be zero.

The proof of the theorem is in the same spirit as the proof of the result that every matrix can be put into reduced row echelon form by a sequence of row operations, but it is somewhat more complicated, since both row and column operations are used. The reader may wish to look first at the example following the proof to get an idea of what is going on in an explicit case.

Proof of Theorem 4.15. We will first prove the existence of the *LPDU* decomposition. Let $A \in \mathbb{F}^{n\times n}$ be given. If $A = O$, put $P = O$ and $L = D = U = I_n$.

Thus suppose $A \neq O$. If A's first column is zero, go to the right until you reach the first nonzero column, say it's the jth. Let δ_j be the first nonzero entry (from the top), and suppose δ_j occurs in the ith row. That is, $\delta_j = a_{ij}$. Perform a sequence of downward transvections to make the entries below δ_j equal to zero. This transforms the jth column of A into

$$(0 \cdots 0 \ \delta_j \ 0 \cdots 0)^T. \tag{4.3}$$

Thus, we can premultiply A by a lower triangular unipotent matrix L_1 to bring the jth column of A into the form (4.3). (Of course, this requires that the matrix entries lie in a field.) The next step is to use δ_j to annihilate all the entries in the ith row of A to the right of the jth column. Since postmultiplying by elementary matrices performs column operations, this amounts to multiplying L_1A on the right by a sequence of unipotent upper triangular elementary matrices. This produces an upper triangular unipotent matrix U_1 such that the first $j-1$ columns of $(L_1A)U_1$ are zero, the jth has the form (4.3), and the ith row is

$$(0 \ \cdots \ 0 \ \delta_j \ 0 \ \cdots \ 0), \tag{4.4}$$

where $\delta_j \neq 0$. We now have the first j columns and ith row of A in the desired form, and from now on, they won't change.

To continue, scan to the right until we find the first nonzero column in L_1AU_1 to the right of the jth column, and suppose this column is the mth. Let b_{km} be its first nonzero entry, and put $\delta_m = b_{km}$. Of course, $k \neq i$. Now repeat the previous process by forming $L_2L_1AU_1U_2$ with suitable lower and upper triangular unipotent matrices L_2 and U_2. Continuing the process, we eventually obtain a lower triangular unipotent matrix L' and an upper triangular unipotent matrix U' such that each row and column of $L'AU'$ has at most one nonzero entry. Thus we may write $L'AU' = PD$ for some partial permutation matrix P, where D is a diagonal matrix. The matrix D is not unique, since some of the columns of PD may be zero. But the entry of D in the nonzero columns is, of course, the corresponding entry in PD. Since we can take any entries we want in the other columns of D, we can assume that D is nonsingular. Unraveling, we get $A = LPDU$, where $L = (L')^{-1}$ and $U = (U')^{-1}$. This proves the existence of the $LPDU$ decomposition.

We must now show that P is unique. So suppose $A = LPDU = L'P'D'U'$ are two $LPDU$ decompositions of A. We have to show $P = P'$. But we can write $L^{-1}L'P' = PDU(D'U')^{-1}$, so $P = P'$ will follow from the next lemma.

Lemma 4.16. *Suppose P and Q are $n \times n$ partial permutation matrices such that $PN = MQ$, where M is lower triangular unipotent and N is nonsingular and upper triangular. Then $P = Q$.*

Proof. We claim that P and Q have the same zero columns. For if the jth column of Q is zero, then the jth column of MQ, and hence of PN, is also

zero. If $p_{ij} \neq 0$, then $p_{ir} = 0$ if $r \neq j$. Thus the (i, j) entry of PN is $p_{ij}n_{jj}$. But $n_{jj} \neq 0$, since N is nonsingular and upper triangular. This implies that the jth column of PN has a nonzero entry, which is impossible. Thus the jth column of P is also zero. But since $QN^{-1} = M^{-1}P$, it follows from the same argument that every zero column of P is also a zero column of Q. This gives us the claim. Now suppose the jth columns of P and Q are nonzero. Then $p_{rj} = q_{sj} = 1$ for exactly one r and exactly one s. We have to show that $r = s$. If $r \neq s$, then since M does downward tranvections and the (r, j) entry $p_{rj}n_{jj}$ of PN is nonzero, it follows that $s \leq r$. (Otherwise, $q_{kj} = 0$ if $k \leq r$, which implies $p_{rj}n_{jj} = 0$.) But since $QN^{-1} = M^{-1}P$, this argument also shows that $r \leq s$. Consequently, $r = s$, and therefore $P = Q$. □

The existence part of the proof of Theorem 4.15 in fact gives an algorithm for finding the *LPDU* factorization. Let's examine it in an example.

Example 4.8. To illustrate the proof, assume that the field is $\mathbb{Q}$ and put

$$A = \begin{pmatrix} 0 & 2 & -2 \\ 0 & 4 & -5 \\ -1 & -2 & -1 \end{pmatrix}.$$

Since the first nonzero entry in the first column of A is $a_{13} = -1$, we can start by subtracting the first column twice from the second and subtracting it once from the third. The result is

$$AU_1 = \begin{pmatrix} 0 & 2 & -2 \\ 0 & 4 & -5 \\ -1 & 0 & 0 \end{pmatrix},$$

where

$$U_1 = \begin{pmatrix} 1 & -2 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Next we subtract twice the first row of AU_1 from the second, which gives

$$L_1AU_1 = \begin{pmatrix} 0 & 2 & -2 \\ 0 & 0 & -1 \\ -1 & 0 & 0 \end{pmatrix},$$

where

$$L_1 = \begin{pmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Finally, we add the second column to the third, getting

$$L_1AU_1U_2 = \begin{pmatrix} 0 & 2 & -2 \\ 0 & 0 & -1 \\ -1 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 2 & 0 \\ 0 & 0 & -1 \\ -1 & 0 & 0 \end{pmatrix}$$

$$= \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} -1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & -1 \end{pmatrix} = PD.$$

Now

$$U = U_1U_2 = \begin{pmatrix} 1 & -2 & -3 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}.$$

After computing $L = (L_1)^{-1}$ and $U = (U_1U_2)^{-1}$, we obtain the *LPDU* factorization

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

⊠

Notice that if $A = LPDU$ is nonsingular, then $A^{-1} = U^{-1}D^{-1}P^{-1}L^{-1}$. In theory, it is simpler to invert each of L, P, D, and U and to multiply them than to compute A^{-1} directly. Indeed, D^{-1} is easy to find, and $P^{-1} = P^T$, so it boils down to computing L^{-1} and U^{-1}. But the inverse of an upper or lower triangular unipotent matrix can be expressed by an explicit formula, although it is too complicated to write down here.

4.3.3 Matrices with an LDU decomposition

If A is invertible, we now know that in the expression $A = LPDU$, D and P are unique. The diagonal entries of D also turn out to be important: they are called the *pivots* of A. The purpose of this subsection is to determine when A admits an LDU decomposition, that is, an $LPDU$ decomposition in which P is the identity matrix. This is answered by considering the matrices A_k consisting of the $k \times k$ blocks in the upper left-hand corner of A. We first observe that if $A = LBU$, where $B \in \mathbb{F}^{n\times n}$, then as long as L and U are lower and upper triangular respectively, then $A_k = L_kB_kU_k$. We leave this as an exercise (see Exercise 4.3.15). We now determine when we can decompose A as $A = LDU$.

Proposition 4.17. *Let $A \in \mathbb{F}^{n\times n}$ be invertible. Then A can be written in the form LDU if and only if A_k is invertible for all $k = 1, \ldots, n$.*

Proof. If $A = LDU$, then $A_k = L_k D_k U_k$ for each index k. Since L_k, D_k, and U_k are invertible for each k, each A_k is invertible. Conversely, suppose each A_k is invertible, and write $A = LPDU$. Then $A_k = L_k(PD)_k U_k$ for each k, so if A_k is invertible for all k, it follows, in particular, that $(PD)_k$ is invertible for each k. However, since P is a permutation matrix and D is an invertible diagonal matrix, the only way this can happen is if each P_k is equal to I_k. In particular, $P = P_n = I_n$. □

We now prove an interesting result for matrices with an LDU decomposition and show that it doesn't always hold if $P \neq I_n$.

Proposition 4.18. *If an invertible matrix A admits an LDU decomposition, then L, D, and U are unique.*

Proof. We already know that D is unique. Assume that A has two LDU decompositions, say

$$A = L_1 D U_1 = L_2 D U_2.$$

Then

$$L_1^{-1} L_2 D = D U_1 U_2^{-1}. \tag{4.5}$$

But in (4.5), the left-hand side is lower triangular, and the right-hand side is upper triangular. Thus, both sides are diagonal. Multiplying by D^{-1} on the right immediately tells us that $L_1^{-1} L_2$ is diagonal, since $D U_1 U_2^{-1} D^{-1}$ is diagonal. But $L_1^{-1} L_2$ is also unipotent; hence $L_1^{-1} L_2 = I_n$. Therefore, $L_1 = L_2$. Canceling $L_1 D$ on both sides, we also see that $U_1 = U_2$. □

Thus, a natural question is whether L and U are unique for all $LPDU$ decompositions. We can answer this by considering the 2×2 case.

Example 4.9. Let $A = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ be invertible. That is, suppose $ad - bc \neq 0$. If $a \neq 0$, then the $LPDU$ decomposition of A is

$$A = \begin{pmatrix} 1 & 0 \\ -c/a & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a & 0 \\ 0 & (ad - bc)/a \end{pmatrix} \begin{pmatrix} 1 & -b/a \\ 0 & 1 \end{pmatrix}.$$

This is, of course, an LDU decomposition. If $a = 0$, then $bc \neq 0$, and A can be expressed either as

$$LPD = \begin{pmatrix} 1 & 0 \\ d/b & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} c & 0 \\ 0 & b \end{pmatrix}$$

or as

$$PDU = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} c & 0 \\ 0 & b \end{pmatrix} \begin{pmatrix} 1 & d/c \\ 0 & 1 \end{pmatrix}.$$

This tells us that if $P \neq I_2$, then L and U aren't necessarily unique.

⊠

4.3.4 The Symmetric LDU Decomposition

Suppose A is an invertible symmetric matrix that has an LDU decomposition. Then it turns out that L and U are not only unique, but they are related. In fact, $U = L^T$. This makes finding the LDU decomposition very simple. The reasoning for this goes as follows. If $A = A^T$ and $A = LDU$, then

$$LDU = (LDU)^T = U^T D^T L^T = U^T D L^T,$$

since $D = D^T$. Therefore, the uniqueness of L, D, and U implies that $U = L^T$.

The upshot is that to factor $A = LDU$ in the general symmetric case, all one needs to do is perform downward row operations on A until A is upper triangular. This is expressed by the equality $L'A = B$, where B is upper triangular. Then $B = DU$, where D is the diagonal matrix such that $d_{ii} = b_{ii}$ for all indices i, and (since all the b_{ii} are nonzero) $U = D^{-1}B$. Thus by construction, U is upper triangular unipotent, and we have $A = LDU$, where $L = U^T$ by the result proved in the previous paragraph.

Example 4.10. Consider the symmetric matrix

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

First bring A into upper triangular form, which is our DU. On doing so, we find that A reduces to

$$DU = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 2 & -2 \\ 0 & 0 & 1 \end{pmatrix}.$$

Hence

$$D = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \text{and} \quad U = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{pmatrix}.$$

Thus $A = LDU$, where U is as above, $L = U^T$, and $D = \operatorname{diag}(1, 2, 1)$.

Summarizing, we state the following result.

Proposition 4.19. *If A is an invertible $n \times n$ symmetric matrix admitting an LDU decomposition, then A can be written in the form $A = LDL^T$ for a unique lower triangular unipotent matrix L. This factorization exists if and only if each symmetric submatrix A_k, $k = 1, \ldots, n$, is invertible.*

The interested reader may wish to consider what happens when an invertible symmetric matrix A has zero pivots (see Exercise 4.3.16).

4.3.5 The Ranks of A and A^T

The *LPDU* decomposition turns out to tell us something a little surprising about the rank of a matrix and the rank of its transpose. Recall that the rank of A is the number of nonzero rows in A_{red} and is well defined, since the reduced form of A is unique. We shall now apply the *LPDU* decomposition to find another description of the rank.

Suppose $A \in \mathbb{F}^{m\times n}$. We may as well assume the A is square (i.e., $m = n$), since otherwise, we may add either rows or columns of zeros to make A square without having any effect on its rank. Thus assume that $A \in \mathbb{F}^{n\times n}$. First of all, we make the following assertion.

Proposition 4.20. *If we write $A = LPDU$, the rank of A is the number of nonzero rows in the partial permutation matrix P. Put another way, the rank of A is the number of ones in P.*

Proof. In fact, the proof of Theorem 4.15 shows that when we put A in *LPDU* form, the nonzero entries of PD are the pivots of A, and they become the corner entries of A_{red} after row permutations. □

We will now prove a nice fact.

Theorem 4.21. *For every matrix A over a field $\mathbb{F}$, A and A^T have the same rank.*

Proof. As usual, we may assume that A is $n \times n$. Writing $A = LPDU$, we see that $A^T = U^T D P^T L^T$. This is almost the *LPDU* decomposition of A^T. All we need to do is notice that there exists another nonsingular diagonal matrix D' such that $DP^T = P^T D'$. But P and P^T obviously have the same number of ones, so A and A^T have the same rank. □

This result is indeed a little surprising, because without knowledge of *LPDU*, there isn't any obvious reason that A and A^T are related in this way. In fact, we needed to do quite a bit of work to obtain *LPDU*. In particular, we first needed to prove the uniqueness of A_{red} to define rank, then we needed to show the existence of the *LPDU* decomposition, and finally, we had to establish the uniqueness of the partial permutation matrix P. Recall that the rank of a matrix tells us the number k of basic null vectors of a linear system $A\mathbf{x} = \mathbf{0}$. Namely, if $A \in \mathbb{F}^{m\times n}$, then $k = n - \text{rank}(A)$.

The result on the rank of A^T gives a simple proof of a well-known result in the theory of matrices.

Corollary 4.22. *Suppose $A \in \mathbb{F}^{m\times n}$ and assume that $M \in \mathbb{F}^{m\times m}$ and $N \in \mathbb{F}^{n\times n}$ are both nonsingular. Then the rank of MAN equals the rank of A.*

Proof. First, notice that A and MA have the same rank. Indeed, since M is nonsingular, it is a product of elementary matrices, so A and MA have the same reduced forms. Thus A, MA, and $A^T M^T$ have the same rank. Similarly, $A^T M^T$ and $N^T A^T M^T$ have the same rank, since N^T is nonsingular. Therefore A and MAN also have the same rank. □

Exercises

Exercise 4.3.1. Find the $LPDU$ decompositions and ranks of the following matrices over $\mathbb{Q}$:

$$\begin{pmatrix} 0 & 1 & 1 \\ 2 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}, \quad \begin{pmatrix} 0 & 0 & 3 \\ 0 & 2 & 1 \\ 1 & 1 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 & 1 \\ 0 & 2 & -1 \\ 1 & -1 & 0 \end{pmatrix}.$$

Exercise 4.3.2. Find the $LPDU$ decompositions and ranks of the transposes of the matrices in Exercise 4.3.1.

Exercise 4.3.3. Suppose $A = LPDU$. What can you say about the $LPDU$ decomposition of A^{-1} if it exists? What about $(A^{-1})^T$?

Exercise 4.3.4. Let

$$A = \begin{pmatrix} 1 & a & b \\ 0 & 1 & c \\ 0 & 0 & 1 \end{pmatrix}.$$

Find a formula expressing A as a product of upper triangular transvections (i.e., elementary matrices of type III).

Exercise 4.3.5. Find a general formula for the inverse of the general 4×4 upper triangular unipotent matrix

$$U = \begin{pmatrix} 1 & a & b & c \\ 0 & 1 & d & e \\ 0 & 0 & 1 & f \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Exercise 4.3.6. Show directly that an invertible upper triangular matrix B can be expressed as $B = DU$, where D is a diagonal matrix with nonzero diagonal entries and U is an upper triangular matrix all of whose diagonal entries are ones. Is this still true if B is singular?

Exercise 4.3.7. Find the LDU decomposition of the matrix

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

Exercise 4.3.8. Let $\mathbb{F} = \mathbb{F}_3$. Find the $LPDU$ decomposition of

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

Exercise 4.3.9. Find a 3×3 matrix A such that the matrix L in the $A = LPDU$ decomposition isn't unique.

Exercise 4.3.10. Assume that $A \in \mathbb{R}^{n \times n}$ is symmetric and has an LDU decomposition. Show that if all the diagonal entries of D are nonnegative, then A can be written $A = CC^T$, where C is lower triangular. This expression is called the Cholesky decomposition of A.

Exercise 4.3.11. Find the number of 2×2 matrices over $\mathbb{F}^2$ having rank 2, and do the same for 3×3 matrices of rank 3.

Exercise 4.3.12. Suppose p is prime. Find a formula for $|GL(2, \mathbb{F}_p)|$. (We will find a formula for $|GL(n, \mathbb{F}_p)|$ for every n in Chap. 6.)

Exercise 4.3.13. Find the rank of each of the following matrices:

$$\begin{pmatrix} 1 & 2 & 3 \\ 1 & 4 & 9 \\ 1 & 8 & 27 \end{pmatrix}, \quad \begin{pmatrix} 1 & 2 & 2 \\ 1 & 4 & 4 \\ 1 & 8 & 8 \end{pmatrix}.$$

Can you see a general result?

Exercise 4.3.14. Write each of the matrices

$$\begin{pmatrix} 1 & 1 & 2 & 1 \\ 1 & -1 & 0 & 2 \\ 2 & 0 & 0 & 1 \\ 1 & 2 & 1 & -1 \end{pmatrix} \text{ and } \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 2 & 2 \\ 0 & 2 & 4 & 4 \\ 1 & 2 & 4 & -3 \end{pmatrix}$$

in the form $LPDU$, where $U = L^T$.

Exercise 4.3.15. This is an exercise in matrix multiplication. Let $A \in \mathbb{F}^{n \times n}$ be expressed as $A = LBU$, where L and U are lower and upper triangular elements of $\mathbb{F}^{n \times n}$ respectively and $B \in \mathbb{F}^{n \times n}$ is arbitrary. Show that $A_k = L_k B_k U_k$ for each $k = 1, \ldots, n$. (Recall that A_k is the $k \times k$ matrix in the upper left-hand corner of A.)

Exercise 4.3.16. Prove the following result.

Proposition 4.23. *Let A be a symmetric invertible matrix. Then there exists an expression $A = LPDU$ with L, P, D, U as usual such that :*

(i) $U = L^T$,

(ii) $P = P^T = P^{-1}$, *and*

(iii) $PD = DP$.
Conversely, if L, P, D, U satisfy the above three conditions, then $LPDU$ is symmetric.

Exercise 4.3.17. Suppose p is prime. Find a formula for

$$|\{A \in GL(n, \mathbb{F}_p) \mid A = LDU\}|.$$

Chapter 5
An Introduction to the Theory of Determinants

In this chapter, we will introduce and study a remarkable function called the determinant, which assigns to an $n \times n$ matrix A over a field $\mathbb{F}$ a scalar $\det(A) \in \mathbb{F}$ having two remarkable properties: $\det(A) \neq 0$ if and only if A is invertible, and if B is also in $\mathbb{F}^{n\times n}$, then $\det(AB) = \det(A)\det(B)$. The latter property is referred to as the product formula. From a group-theoretic standpoint, the determinant is a group homomorphism $\det : GL(n, \mathbb{F}) \to \mathbb{F}^*$. In particular, $\det(I_n) = 1$. A further remarkable property of the determinant is that $\det(A^T) = \det(A)$. This implies, for example, that if A is orthogonal, that is, $A^T A = I_n$, then $\det(A)^2 = 1$. Thus the determinant of an orthogonal matrix A satisfies $\det(A) = \pm 1$. As can be imagined, the definition of a function of n^2 variables having all the properties claimed above is nontrivial. We will define $\det(A)$ via the classical formula attributed to Leibniz in 1683. This is not intended to imply that the notion of the determinant of a matrix preceded the notion of a matrix, since in fact, Leibniz's definition was applied to the coefficients of a linear system. If $A \in \mathbb{F}^{n\times n}$, the *determinant* of A is defined to be

$$\det(A) = \sum_{\pi \in S(n)} \operatorname{sgn}(\pi)\ a_{\pi(1)1} a_{\pi(2)2} \cdots a_{\pi(n)n}.$$

For example, $\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc$. Note that since there is a term for every element π of the symmetric group $S(n)$, the determinant of A contains $n!$ terms. The symbol $\operatorname{sgn}(\pi)$, which is known as the *signature* of π, is either 1 or -1, depending on whether π is an even or odd permutation. (One definition is that π is even (respectively odd) if it can be expressed as a product of an even number (respectively odd number) of transpositions. However, it isn't clear that permutations cannot be expressed both ways.) Thus, $\operatorname{sgn}(\pi)$ must be defined precisely. In fact, we will prove that the signature is a homomorphism from the group $S(n)$ to the multiplicative cyclic group $\{\pm 1\}$ such

J.B. Carrell, *Groups, Matrices, and Vector Spaces*,
DOI 10.1007/978-0-387-79428-0_5

that $\operatorname{sgn}(\tau) = -1$ if τ is a transposition. This justifies the ad hoc definition we gave above.

Of course, properties such as the product formula were not proved until the introduction of matrices. The determinant function has proved to be such a rich topic of research that between 1890 and 1929, Thomas Muir published a five-volume treatise on it entitled *The History of the Determinant*. We will discuss Charles Dodgson's fascinating formula generalizing the formula for a 2×2 determinant. Interest in Dodgson's formula has recently been revived, and it is now known as Dodgson condensation. The reader will undoubtedly recall that Charles Dodgson wrote *Alice in Wonderland* under the pen name Lewis Carroll.

Before diving into the signature and the proofs of the properties of the determinant mentioned above, we will state the main theorem (Theorem 5.1) on determinants, which lists its important properties. Then we will derive an efficient method for computing a determinant. A reader who wants to know only how to compute a determinant can skip the technical details of its definition. After those remarks, we will introduce the signature of a permutation, and finally, we will prove the main theorem.

Some further basic properties of the determinant such as the Laplace expansion and Cramer's rule are derived in the appendix. The Laplace expansion is frequently used to define the determinant by induction. This approach has the drawback that one is required to show that the Laplace expansions along any two rows or columns give the same result (which is true, but messy to prove). The appendix also contains a proof of Cramer's rule and a characterization of matrices with integer entries whose inverses have only integer entries.

5.1 An Introduction to the Determinant Function

This section is an exposition on the determinant function, stating its properties and describing how the determinant is calculated. It is intended for the reader who needs to know only the basic facts about the determinant. For such readers, we recommend also perusing Section 5.3.1 for further computational techniques.

5.1.1 The main theorem

We will assume that all matrices are defined over the same arbitrary field $\mathbb{F}$. The essential properties of the determinant function are captured in the following theorem.

Theorem 5.1. *There exists a unique function* $\det : \mathbb{F}^{n\times n} \to \mathbb{F}$*, called the* determinant*, with the following properties.*

(*i*) *(The product formula) For all* $A,\ B \in \mathbb{F}^{n\times n}$*,*

$$\det(AB) = \det(A)\det(B). \tag{5.1}$$

(*ii*) *If* $A = (a_{ij})$ *is either upper or lower triangular, then*

$$\det(A) = \prod_{i=1}^{n} a_{ii}. \tag{5.2}$$

(*iii*) *If* E *is a row swap matrix, then*

$$\det(E) = -1. \tag{5.3}$$

(*iv*) A *is nonsingular if and only if* $\det(A) \neq 0$*.*

One can deduce (*iv*) directly from (*i*), (*ii*), and (*iii*) (Exercise 5.1.2).

Example 5.1. Since $A \in \mathbb{F}^{n\times n}$ is nonsingular if and only if $\det(A) \neq 0$ and $\det(AB) = \det(A)\det(B)$, it follows that det defines a group homomorphism $\det : GL(n,\mathbb{F}) \to \mathbb{F}^*$. The kernel of this homomorphism, which is by definition $\{A \in \mathbb{F}^{n\times n} \mid \det(A) = 1\}$, is an important normal subgroup of the matrix group $GL(n,\mathbb{F})$ called the *special linear group*. The special linear group is denoted by $SL(n,\mathbb{F})$. For example,

$$SL(2,\mathbb{C}) = \{\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathbb{C}^{2\times 2} \mid ad - bc = 1\}.$$

⊠

Parts (*ii*) and (*iii*) tell us the values of $\det(E)$ for all elementary matrices. If E is obtained by multiplying the ith row of I_n by r, then $\det(E) = r$, and if E is obtained from I_n by adding a multiple of its jth row to its ith row, then $\det(E) = 1$. The method for computing $\det(A)$ is thus to row reduce A, making use of the identity $\det(EA) = \det(E)\det(A)$.

5.1.2 The computation of a determinant

We will assume that the determinant function exists and has the properties listed in Theorem 5.1. Let us now see how to compute it. The 1×1 case is easy. If $A = (a)$, then $\det(A) = a$. In the 2×2 case, the formula is given by

$$\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc, \tag{5.4}$$

which can be proved by applying Theorem 5.1 (see Exercise 5.1.3). Notice that this formula says that if A is 2×2, then $\det(A) = 0$ if and only if its rows are proportional.

The most efficient general technique for computing $\det(A)$ is to use row operations. Given $A \in \mathbb{F}^{n \times n}$, one can always find elementary matrices $E_1, \ \ldots \ , E_k$ such that $E_k \cdots E_1 A$ is an upper triangular matrix, say U. Repeated application of Theorem 5.1 then gives

$$\det(U) = \det(E_1) \cdots \det(E_k) \det(A).$$

Since $\det(U) = u_{11} u_{22} \cdots u_{nn}$, where the u_{ii} are the diagonal entries of U,

$$\det(A) = \frac{u_{11} u_{22} \cdots u_{nn}}{\det(E_1) \cdots \det(E_k)}, \tag{5.5}$$

which can be evaluated by applying Theorem 5.1.

Example 5.2. Let us compute $\det(A)$, where

$$A = \begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 \end{pmatrix},$$

taking the field of coefficients to be $\mathbb{Q}$. We can make the following sequence of row operations, all of type III except for the last, which is a row swap.

$$A \to \begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 \end{pmatrix} \to \begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & -1 & 0 \end{pmatrix} \to \begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & -1 \\ 0 & 1 & -1 & 0 \end{pmatrix} \to$$

$$\begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & -1 & 0 \end{pmatrix} \to \begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}.$$

Thus (5.5) implies $\det(A) = -1$.

One curiosity is that if the field $\mathbb{F}$ has characteristic 2, then row swaps have determinant 1. Let us rework the previous example with $\mathbb{F} = \mathbb{F}_2$ with this in mind.

Example 5.3. First add the first row to the third and fourth rows successively. Then we get

$$\det(A) = \det \begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \end{pmatrix}.$$

Since the field is $\mathbb{F}_2$, row swaps also leave $\det(A)$ unchanged. Thus

$$\det(A) = \det \begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{pmatrix}.$$

Adding the second row to the third row and the fourth row successively, we get

$$\det(A) = \det \begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$

Finally, switching the last two rows, we get

$$\det(A) = \det \begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} = 1.$$

⊠

Exercises

Exercise 5.1.1. Assuming Theorem 5.1, show that if $A \in \mathbb{F}^{n \times n}$ has two equal rows and the characteristic of $\mathbb{F}$ is not two, then $\det(A) = 0$.

Exercise 5.1.2. Show that (iv) of Theorem 5.1 is a consequence of (i), (ii), and (iii).

Exercise 5.1.3. The purpose of this exercise is to prove the identity (5.4). Define a function $F : \mathbb{F}^{2\times 2} \to \mathbb{F}$ by

$$F \begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc.$$

Show that F satisfies the conditions in Theorem 5.1, and conclude that $F(A) = \det(A)$.

Exercise 5.1.4. Let $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. Without appealing to Theorem 5.1, show the following:

(*i*) $\det(A) = 0$ if and only if the rows of A are proportional. Conclude that A has rank 2 if and only if $ad - bc \neq 0$.

(*ii*) A is invertible if and only if $\det(A) \neq 0$. In that case,

$$A^{-1} = \frac{1}{\det(A)} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

Exercise 5.1.5. Use Theorem 5.1 to prove that if B and C are square matrices over $\mathbb{F}$, then

$$\det \begin{pmatrix} B & * \\ O & C \end{pmatrix} = \det(B) \det(C).$$

Exercise 5.1.6. Compute

$$\det \begin{pmatrix} 1 & 1 & -1 & 0 \\ 2 & 1 & 1 & 1 \\ 0 & -1 & 2 & 0 \\ 1 & 1 & -1 & 1 \end{pmatrix}$$

in two cases: first when the field is $\mathbb{Q}$ and second, when the field is $\mathbb{F}_3$.

Exercise 5.1.7. Express

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

in the form $A = LPDU$, and use your result to compute $\det(A)$ in the following cases:

(a) $\mathbb{F} = \mathbb{Q}$;

(b) $\mathbb{F} = \mathbb{F}_2$; and

(c) $\mathbb{F} = \mathbb{F}_3$.

Exercise 5.1.8. Show that the image of $det : \mathbb{F}^{n \times n} \to \mathbb{F}^*$ is, in fact, all of $\mathbb{F}^*$.

5.2 The Definition of the Determinant

Before formally defining the determinant function, we have to define the *signature* $\operatorname{sgn}(\sigma)$ of a permutation $\sigma \in S(n)$. As already mentioned, the signature is a homomorphism from $S(n)$ to the multiplicative group $\{+1, -1\}$.

5.2.1 The signature of a permutation

Recall the isomorphism $\varphi : S(n) \to P(n)$ defined by $\varphi(\sigma) = P_\sigma$, where P_σ is the permutation matrix whose ith column is $\mathbf{e}_{\sigma(i)}$. The signature $\operatorname{sgn}(\sigma)$ of the permutation $\sigma \in S(n)$ will tell us whether one requires an even or odd number of row swaps S_i to write $P_\sigma = S_1 \cdots S_k$. Correspondingly, we will call π *even* or *odd*.

Definition 5.1. Suppose $\sigma \in S(n)$. If $n > 1$, define the *signature* $\operatorname{sgn}(\sigma)$ of σ to be

$$\operatorname{sgn}(\sigma) = \prod_{i<j} \frac{\sigma(i) - \sigma(j)}{i - j}. \tag{5.6}$$

If $n = 1$, put $\operatorname{sgn}(\sigma) = 1$.

Clearly, $\operatorname{sgn}(\sigma)$ is a nonzero rational number. There is at least one example in which the value of $\operatorname{sgn}(\sigma)$ is clear.

Example 5.4. The identity permutation $id_n \in S(n)$ has signature $\operatorname{sgn}(id_n) = 1$. ⊠

As usual, we will denote id_n by 1. We now establish the properties of sgn. We first show that $\operatorname{sgn}(\sigma) \in \{\pm 1\}$.

Proposition 5.2. *For every* $\sigma \in S(n)$, $\operatorname{sgn}(\sigma) = \pm 1$.

Proof. The case $n = 1$ is clear, so suppose $n > 1$. Since σ is a bijection of $\{1, 2, \ldots, n\}$ and

$$\frac{\sigma(i) - \sigma(j)}{i - j} = \frac{\sigma(j) - \sigma(j)}{j - i},$$

it follows that

$$(\operatorname{sgn}(\sigma))^2 = \prod_{i \neq j} \frac{\sigma(i) - \sigma(j)}{i - j}.$$

Moreover, since

$$\sigma(\sigma^{-1}(i)) - \sigma(\sigma^{-1}(j)) = i - j,$$

each possible value of $(i - j)$ occurs the same number of times in the numerator and denominator. Hence $\operatorname{sgn}(\sigma)^2 = 1$, so the proof is done. □

Let $N(\sigma) = \{(i,j) \mid i < j,\ \sigma(i) > \sigma(j)\}$, and put $n(\sigma) = |N(\sigma)|$. By the definition of $\operatorname{sgn}(\sigma)$ and Proposition 5.2,

$$\operatorname{sgn}(\sigma) = (-1)^{n(\sigma)}.$$

Here is an an example.

Example 5.5. Recall that $\sigma_{ij} \in S(n)$ denotes the transposition that switches i and j while leaving all other k, $1 \le k \le n$, unchanged. Let us compute $\operatorname{sgn}(\sigma_{12})$. Now σ_{12} interchanges 1 and 2 and leaves every k between 3 and n unchanged. Thus, $(1,2)$ is the only pair (i,j) such that $i < j$ for which $\sigma_{12}(i) > \sigma_{12}(j)$. Hence $n(\sigma_{12}) = 1$, so $\operatorname{sgn}(\sigma_{12}) = -1$. ⊠

We now establish the main properties of the signature.

Proposition 5.3. *The signature mapping* $\operatorname{sgn} : S(n) \to \{\pm 1\}$ *has the following properties:*

(i) sgn *is a group homomorphism; that is, for all* $\sigma, \tau \in S(n)$,

$$\operatorname{sgn}(\tau\sigma) = \operatorname{sgn}(\tau)\operatorname{sgn}(\sigma);$$

(ii) if σ *is any transposition, then* $\operatorname{sgn}(\sigma) = -1$*; and*

(iii) for all $\sigma \in S(n)$, $\operatorname{sgn}(\sigma^{-1}) = \operatorname{sgn}(\sigma)$.

Proof. To prove (i), it will suffice to show that $n(\tau\sigma) = n(\tau) + n(\sigma)$ for an arbitrary pair $\tau, \sigma \in S(n)$. Suppose $i < j$ and write

$$\frac{\tau(\sigma(i)) - \tau(\sigma(j))}{i - j} = \frac{\tau(\sigma(i)) - \tau(\sigma(j))}{\sigma(i) - \sigma(j)} \cdot \frac{\sigma(i) - \sigma(j)}{i - j}.$$

The left-hand side is negative if and only if $\dfrac{\tau(\sigma(i)) - \tau(\sigma(j))}{\sigma(i) - \sigma(j)}$ and $\dfrac{\sigma(i) - \sigma(j)}{i - j}$ have different signs. Thus, $(i,j) \in N(\tau\sigma)$ if and only if $(i,j) \in N(\sigma)$ or $(\sigma(i), \sigma(j)) \in N(\tau)$. Since there is no pair $(i,j) \in N(\tau\sigma)$ such that $(i,j) \in N(\sigma)$ and $(\sigma(i), \sigma(j)) \in N(\tau)$, it follows that $n(\tau\sigma) = n(\tau) + n(\sigma)$. This completes the proof of (i). For (ii), we can use the result of Example 5.5. Let σ_{ab} denote the transposition interchanging $a \neq b$. As an exercise, the reader should check that

$$\sigma_{ab} = \sigma_{1b}\sigma_{2a}\sigma_{12}\sigma_{2a}\sigma_{1b}. \tag{5.7}$$

Hence, by (i) and the result that $\operatorname{sgn}(\sigma_{12}) = -1$ (by Example 5.5),

$$\operatorname{sgn}(\sigma_{ab}) = \operatorname{sgn}(\sigma_{1b})\operatorname{sgn}(\sigma_{2a})\operatorname{sgn}(\sigma_{12})\operatorname{sgn}(\sigma_{2a})\operatorname{sgn}(\sigma_{1b}) = \operatorname{sgn}(\sigma_{12}) = -1.$$

This gives (*ii*). For (*iii*), just note that for every $\sigma \in S(n)$, we have $\sigma^{-1}\sigma = 1$, and apply (*i*) and the fact that $\operatorname{sgn}(id_n) = 1$. □

A permutation σ is said to be *even* if $\operatorname{sgn}(\sigma) = 1$ and odd otherwise. In particular, all transpositions are odd. Let $A(n)$ denote the kernel of sgn. Then $A(n)$ consists of the even permutations, and since the kernel of a homomorphism is a normal subgroup, we see that $A(n)$ is normal in $S(n)$. The subgroup $A(n)$ is called the *alternating group*. It is a well-known classical result that if $n > 4$, then $A(n)$ has no nontrivial normal subgroups.

5.2.2 The determinant via Leibniz's Formula

We now have everything needed to give the Leibniz definition of $\det(A)$.

Definition 5.2. Let $A \in \mathbb{F}^{n\times n}$. Then the *determinant of* A, $\det(A)$, is the scalar defined by the identity

$$\det(A) := \sum_{\pi\in S(n)} \operatorname{sgn}(\pi)\ a_{\pi(1)1}a_{\pi(2)2}\cdots a_{\pi(n)n}. \tag{5.8}$$

The above sum contains $n!$ terms, so one is hardly ever going to use the definition to compute a determinant except when n is small. The cases $n = 1$, 2, and 3 can be worked out, but even $n = 4$ is too difficult to do by hand without row operations. Fortunately, one seldom needs to actually compute $\det(A)$.

Example 5.6. (1×1 and 2×2 determinants). If $A = (a)$ is of size 1×1, then since $S(1) = \{id_1\}$ and $\operatorname{sgn}(id_1) = 1$, it follows that $\det(A) = a$. Now suppose $A = (a_{ij})$ is 2×2. There are exactly two elements in $S(2)$, namely the identity id_2 and the transposition σ_{12}. Denoting id_2 by σ and σ_{12} by τ, we have, by definition,

$$\begin{aligned}\det\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22}\end{pmatrix} &= \operatorname{sgn}(\sigma)a_{\sigma(1)1}a_{\sigma(2)2} + \operatorname{sgn}(\tau)a_{\tau(1)1}a_{\tau(2)2} \\ &= a_{11}a_{22} - a_{21}a_{12}.\end{aligned}$$

The properties of Theorem 5.1 are easy to check in these two cases. ⊠

Example 5.7. (3×3 determinants). For the 3×3 case, let us first list the elements $\sigma \in S(3)$ and their signatures. We will use the triple $[\sigma(1), \sigma(2), \sigma(3)]$ to represent each $\sigma \in S(3)$. Then the signatures for $S(3)$ are given in the following table:

σ	$[1,2,3]$	$[1,3,2]$	$[2,3,1]$	$[2,1,3]$	$[3,1,2]$	$[3,2,1]$
$\operatorname{sgn}(\sigma)$	1	-1	$+1$	-1	$+1$	-1

.

Hence,

$$\begin{aligned}\det(A) = a_{11}a_{22}a_{33} - a_{11}a_{32}a_{23} + a_{21}a_{32}a_{13}\\ -a_{21}a_{12}a_{33} + a_{31}a_{12}a_{23} - a_{31}a_{22}a_{13}.\end{aligned}$$

Rewriting this as

$$\begin{aligned}\det(A) = a_{11}a_{22}a_{33} + a_{21}a_{32}a_{13} + a_{31}a_{12}a_{23} +\\ -a_{11}a_{32}a_{23} - a_{21}a_{12}a_{33} - a_{31}a_{22}a_{13},\end{aligned}$$

one sees that $\det(A)$ is the sum of the products of three entries, either two or three of which are on a diagonal of A parallel to its main diagonal, less the sum of the products of three entries, either two or three of which are on a diagonal of A parallel to its antidiagonal, that is, the diagonal from the northeast corner of A to its southwest corner. This is a standard expression for a 3×3 determinant known as the rule of Sarrus. Warning: the rule of Sarrus does not generalize to the 4×4 case. ⊠

5.2.3 Consequences of the definition

Several parts of Theorem 5.1 follow directly from the definition. Let us first prove part (*ii*).

Proposition 5.4. *Suppose A is of size $n \times n$ and either upper or lower triangular. Then*

$$\det(A) = a_{11}a_{22}\cdots a_{nn}.$$

Proof. We will suppose that A is upper triangular and leave the lower triangular case as an exercise (not a very hard one at that). The point is that in this case, the only nonzero term in (5.2) is $a_{11}a_{22}\cdots a_{nn}$, which corresponds to the identity permutation. For if $\sigma \in S(n)$ is different from id_n, then $\sigma(i) > i$ for some i; hence $a_{\sigma(i)i} = 0$, since A is upper triangular. Thus, $\operatorname{sgn}(\sigma)a_{\sigma(1)1}a_{\sigma(2)2}\cdots a_{\sigma(n)n} = 0$. □

Suppose now that $P \in \mathbb{F}^{n\times n}$ is a permutation matrix. In this case, $\det(P)$ has a very pretty interpretation.

Proposition 5.5. *Assume that $P \in \mathbb{F}^{n\times n}$ is a permutation matrix, say $P = P_\mu$. Then $\det(P) = \operatorname{sgn}(\mu)$.*

Proof. Recall that $P_\mu = \begin{pmatrix}\mathbf{e}_{\mu(1)} & \mathbf{e}_{\mu(2)} & \cdots & \mathbf{e}_{\mu(n)}\end{pmatrix}$. Writing $P = (p_{ij})$, we have $p_{\mu(i)i} = 1$ for all i, while all other p_{rs} are equal to 0. Therefore, the only nonzero term in the expression for $\det(P)$ is

$$\operatorname{sgn}(\mu)p_{\mu(1)1}p_{\mu(2)2}\cdots p_{\mu(n)n} = \operatorname{sgn}(\mu).$$

□

The following result about the determinant of the transpose is going to be used in several places in the rest of this section.

Proposition 5.6. *For every $A \in \mathbb{F}^{n\times n}$,*

$$\det(A^T) = \det(A).$$

Proof. Since $A^T = (b_{ij})$, where $b_{ij} = a_{ji}$, formula (5.2) implies

$$\det(A^T) := \sum_{\sigma\in S(n)} \operatorname{sgn}(\sigma)\ a_{1\sigma(1)}a_{2\sigma(2)}\cdots a_{n\sigma(n)}. \tag{5.9}$$

Now if $\sigma(i) = j$, then $\sigma^{-1}(j) = i$, so $a_{i\sigma(i)} = a_{\sigma^{-1}(j)j}$. Thus,

$$\begin{aligned}\operatorname{sgn}(\sigma)\ a_{1\sigma(1)}a_{2\sigma(2)}\cdots a_{n\sigma(n)} &= \operatorname{sgn}(\sigma)\ a_{\sigma^{-1}(1)1}a_{\sigma^{-1}(2)2}\cdots a_{\sigma^{-1}(n)n}\\ &= \operatorname{sgn}(\sigma^{-1})\ a_{\sigma^{-1}(1)1}a_{\sigma^{-1}(2)2}\cdots a_{\sigma^{-1}(n)n},\end{aligned}$$

since by Proposition 5.3, $\operatorname{sgn}(\sigma^{-1}) = \operatorname{sgn}(\sigma)$. But the correspondence $\sigma \to \sigma^{-1}$ is a bijection of $S(n)$, so we can conclude that

$$\sum_{\sigma\in S(n)} \operatorname{sgn}(\sigma)\ a_{1\sigma(1)}a_{2\sigma(2)}\cdots a_{n\sigma(n)} = \sum_{\tau\in S(n)} \operatorname{sgn}(\tau)\ a_{\tau(1)1}a_{\tau(2)2}\cdots a_{\tau(n)n}.$$

Thus $\det(A^T) = \det(A)$. □

5.2.4 The effect of row operations on the determinant

In order to prove the product formula $\det(AB) = \det(A)\det(B)$ for all $A, B \in \mathbb{F}^{n\times n}$, we will demonstrate how a row operation changes the determinant. In fact, we will show that for every elementary matrix E, $\det(EA) = \det(E)\det(A)$. Since $\det(A) = \det(A^T)$, it follows that the product formula holds for AB if either A or B is a product of elementary matrices. First we prove the following result.

Proposition 5.7. *Suppose that $E \in \mathbb{F}^{n\times n}$ is the elementary matrix obtained from I_n by multiplying the ith row of I_n by $r \in \mathbb{F}$. Then for every $A \in \mathbb{F}^{n\times n}$,*

$$\det(EA) = r\det(A) = \det(E)\det(A).$$

Proof. Certainly $\det(E) = r$, while $\det(EA) = r\det(A)$, since every term in the expansion of $\det(A)$ is multiplied by r. □

We next prove a result for swap matrices.

Proposition 5.8. *Suppose $A \in \mathbb{F}^{n\times n}$ and $S \in \mathbb{F}^{n\times n}$ is a row swap matrix. Then*

$$\det(SA) = -\det(A) = \det(S)\det(A).$$

Proof. Suppose S is a row swap matrix. Put $SA = B = (b_{ij})$ and let τ denote the transposition such that $S = P_\tau$. Then $b_{ij} = a_{\tau(i)j}$. We will compute $\det(B)$ using the result that $\det(B) = \det(B^T)$. First of all, for every $\sigma \in S(n)$,

$$\begin{aligned} b_{1\sigma(1)}b_{2\sigma(2)}\cdots b_{n\sigma(n)} &= a_{\tau(1)\sigma(1)}a_{\tau(2)\sigma(2)}\cdots a_{\tau(n)\sigma(n)} \\ &= a_{\tau(1)\mu\tau(1)}a_{\tau(2)\mu\tau(2)}\cdots a_{\tau(n)\mu\tau(n)} \\ &= a_{1\mu(1)}a_{2\mu(2)}\cdots a_{n\mu(n)}, \end{aligned}$$

where $\mu = \sigma\tau$. Thus,

$$\begin{aligned} \det(B) &= \sum_{\sigma\in S(n)} \operatorname{sgn}(\sigma) b_{1\sigma(1)}b_{2\sigma(2)}\cdots b_{n\sigma(n)} \\ &= \sum_{\sigma\in S(n)} \operatorname{sgn}(\sigma) a_{1\mu(1)}a_{2\mu(2)}\cdots a_{n\mu(n)} \\ &= -\sum_{\mu\in S(n)} \operatorname{sgn}(\mu) a_{1\mu(1)}a_{2\mu(2)}\cdots a_{n\mu(n)} \\ &= -\det(A). \end{aligned}$$

The third equality uses two facts; first, if $\mu = \sigma\tau$, then $\operatorname{sgn}(\mu) = -\operatorname{sgn}(\sigma)$, and second, if σ varies through all of $S(n)$, then so does $\mu = \sigma\tau$. Thus, $\det(SA) = -\det(A)$. To finish the proof, we need to show that $\det(S) = -1$. But $\det(S) = \det(SI_n) = -\det(I_n) = -1$, since $\det(I_n) = 1$. Alternatively, since $S = P_\tau$, it follows that $\det(S) = \operatorname{sgn}(\tau) = -1$ by Proposition 5.5. □

The next step is to show that $\det(EA) = \det(A)$ if E is a transvection, that is, an elementary matrix of type III. Suppose $A \in \mathbb{F}^{n\times n}$ and let $\mathbf{a}_1, \ldots, \mathbf{a}_n$ denote its rows. Let E be of type III, say E is obtained from I_n by replacing the ith row of I_n by itself plus r times the jth row. Then the ith row of EA is $\mathbf{a}_i + r\mathbf{a}_j$, and the other rows are unchanged. Hence by (5.9), each term in the expansion of $\det(EA)$ has the form

$$\operatorname{sgn}(\sigma)a_{1\sigma(1)}\ldots a_{(i-1)\sigma(i-1)}(a_{i\sigma(i)} + ra_{j\sigma(i)})a_{i+1\sigma(i+1}\ldots a_{n\sigma(n)}.$$

Thus, $\det(EA)$ is of the form $\det(A) + r\det(C)$, where $C \in \mathbb{F}^{n\times n}$ has the property that its ith and jth rows both coincide with $\mathbf{a}_j$. If we apply the

fact that $\det(SB) = -\det(B)$ whenever S is a row swap matrix, then we see that if S swaps the ith and jth rows, we get $C = SC$, so $\det(C) = \det(SC) = -\det(C)$. Thus $2\det(C) = 0$. It follows that $\det(C) = 0$ as long as the characteristic of the field $\mathbb{F}$ is different from two. The formula $\det(C) = 0$ when $\mathbb{F}$ has characteristic two and two rows of C coincide requires a special argument, which is given in the appendix to this chapter. □

The following proposition summarizes what was proved in this section modulo showing that $\det(C) = 0$ in characteristic two if C has two equal rows.

Proposition 5.9. *If $E \in \mathbb{F}^{n\times n}$ is any elementary matrix, then $\det(EA) = \det(E)\det(A)$ for all $A \in \mathbb{F}^{n\times n}$. Moreover, $\det(E) = r$ if E multiplies the ith row of I_n by r, $\det(E) = -1$ if E is a row swap, and $\det(E) = 1$ if E is a transvection.*

5.2.5 The proof of the main theorem

We are now ready to complete the proof of the main theorem. Parts (*ii*) and (*iii*) of the main theorem have already been verified. Let $A \in \mathbb{F}^{n\times n}$, and let us first show that $\det(A) \neq 0$ if and only if A is nonsingular. There exist elementary matrices $E_1, \ldots, E_k$ such that $E_k \cdots E_1 A = A_{red}$, so

$$\det(A) = \frac{\det(A_{red})}{\prod_i \det(E_i)}.$$

Hence $\det(A) \neq 0$ if and only if $\det(A_{red}) \neq 0$. But since A_{red} is upper triangular, $\det(A_{red}) \neq 0$ if and only if $A_{red} = I_n$ if and only if A is nonsingular. It remains to prove the product formula. If A and B are both nonsingular, then the validity of the product formula is clear, for both A and B and hence AB are products of elementary matrices. On the other hand, if either A or B is singular, I claim that AB is singular. If not, then $(AB)^{-1}A$ is a left inverse of B, so B is nonsingular. Similarly, $B(AB)^{-1}$ is a right inverse of A, so A is also nonsingular. Hence, if $\det(A)\det(B) = 0$, then $\det(AB) = 0$. This finishes the proof of the product formula. □

5.2.6 Determinants and LPDU

Recall from Chap. 4 that every $A \in \mathbb{F}^{n\times n}$ can be written $A = LPDU$, where L and U are respectively lower and upper triangular unipotent matrices, P is a unique partial permutation matrix, and D is an invertible diagonal matrix.

Clearly, $\det(A) = 0$ unless P is a full permutation matrix. Furthermore, we have the following proposition.

Proposition 5.10. *If $A = LPDU$ is nonsingular, then*

$$\det(A) = \det(P)\det(D) = \pm\det(D).$$

Thus, up to sign, the determinant of an invertible matrix is the product of its pivots. If $A = LDU$, then $\det(A) = \det(D)$.

Proof. Just use the product rule and the fact that $\det(L) = \det(U) = 1$, since L and U are triangular and have ones on their diagonals. □

Recall from Proposition 4.17 that an invertible $A \in \mathbb{F}^{n\times n}$ has an LDU decomposition if and only if each A_k is also invertible, where A_k is the $k \times k$ submatrix in the upper left-hand corner of A, and that the pivots of A are the diagonal entries of D. Let $d_1, \ldots, d_n$ be these pivots. Then $A_k = L_k D_k U_k$, so

$$\det(A_k) = \det(D_k) = d_1 \cdots d_k.$$

This gives us the following result.

Proposition 5.11. *If $A \in \mathbb{F}^{n\times n}$ satisfies the condition that each A_k, $1 \leq k \leq n$, is invertible, then A has an LDU decomposition, and the kth pivot of A in its LDU decomposition is*

$$d_k = \frac{\det(A_k)}{\det(A_{k-1})}. \tag{5.10}$$

5.2.7 A beautiful formula: Lewis Carroll's identity

The theory of determinants is a remarkably rich topic, and we have barely scratched its surface. Many of the foremost mathematicians of the nineteenth century, among them Gauss, Laplace, Lagrange, Cayley, Sylvester, and Jacobi, discovered some of its important properties and applications. An example is the Jacobian of a mapping and the change of variables formula. In this section, we will consider a determinantal curiosity from the nineteenth century that has recently been found to have connections to contemporary mathematics. The formula $\det\begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc$ turns out to have a generalization to $n \times n$ matrices that was discovered by Charles Dodgson, a professor of mathematics at Oxford who, as is well known, wrote *Alice's Adventures in Wonderland* and *Through the Looking Glass* under the pseudonym Lewis Carroll. Dodgson found an identity involving the determinant of an $n \times n$ matrix A that is analogous to the formula in the 2×2 case. Let A_C be the

$(n-2)\times(n-2)$ submatrix of A obtained by deleting the first and last rows and the first and last columns. If $n=2$, put $\det(A_C)=1$. Also, let A_{NW} denote the $(n-1)\times(n-1)$ submatrix in the upper left-hand corner of A, and define A_{NE}, A_{SW}, and A_{SE} to be the $(n-1)\times(n-1)$ submatrices in the other three corners of A. Dodgson's formula asserts that

$$\det(A_C)\det(A)=\det(A_{NW})\det(A_{SE})-\det(A_{NE})\det(A_{SW}) \tag{5.11}$$

(see C.L. Dodgson, *Proc. Royal Soc. London* **17**, 555–560 (1860)). If $\det(A_C)\neq 0$, this substantially cuts down on the difficulty of finding $\det(A)$. Of course, if $\det(A_C)=0$, then the identity says nothing about $\det(A)$, although it does say something about the four determinants at the corners. This formula is referred to as Dodgson condensation. Dodgson himself supplied the term condensation. In the 1980s, it was noticed that Dodgson condensation is related to the problem of counting alternating sign matrices (ASMs). Eventually, the problem of enumerating the ASMs was given an elegant solution using statistical mechanics.

Exercises

Exercise 5.2.1. Two matrices $A, B \in \mathbb{F}^{n\times n}$ are said to be *similar* if $A = CBC^{-1}$ for some $C \in GL(n, \mathbb{F})$. Show that similar matrices have the same determinant.

Exercise 5.2.2. Suppose P is an $n \times n$ matrix over $\mathbb{C}$ such that $PP = P$. What is $\det(P)$? What is $\det(Q)$ if $Q^4 = Q^{-1}$?

Exercise 5.2.3. Which of the following statements are true. Give your reasoning or supply a counter example.

(*i*) The determinant of a real symmetric matrix is always nonnegative.

(*ii*) If A is a 2×3 real matrix, then $\det(AA^T)\mathrm{Gal}(\mathbb{E}/\mathbb{Q})0$.

(*iii*) If A is a square real matrix, then $\det(AA^T)\mathrm{Gal}(\mathbb{E}/\mathbb{Q})0$.

Exercise 5.2.4. An $n \times n$ matrix A over $\mathbb{R}$ is called *skew-symmetric* if $A^T = -A$. Show that if A is a skew-symmetric $n \times n$ matrix and n is odd, then A cannot be invertible.

Exercise 5.2.5. Let $H \in \mathbb{R}^{n\times n}$ be a Hadamard matrix. That is, suppose $HH^T = nI_n$. Find $\det(H)$.

Exercise 5.2.6. Recall that $SL(n, \mathbb{F})$ denotes the set of all $A \in \mathbb{F}^{n\times n}$ such that $\det(A) = 1$. Prove that $SL(n, \mathbb{F})$ is a matrix group and a proper normal subgroup of $GL(n, \mathbb{F})$ if the characteristic of $\mathbb{F}$ is different from 2.

Exercise 5.2.7. Find the 3×3 permutation matrices that lie in $SL(3, \mathbb{R})$.

Exercise 5.2.8. Let $SO(n, \mathbb{R})$ denote the set of all $n \times n$ orthogonal matrices Q such that $\det(Q) = 1$. Show that $SO(n, \mathbb{R})$ is a matrix group and a proper normal subgroup of $O(n, \mathbb{R})$.

Exercise 5.2.9. If $A \in \mathbb{C}^{m\times n}$, define $A^H = \overline{A}^T$, where $\overline{A}$ is the matrix obtained by conjugating each entry of A. An $n \times n$ matrix U over $\mathbb{C}$ is called *unitary* if $U^{-1} = U^H$. What are the possible values of the determinant of $\det(U)$ of a unitary matrix U?

Exercise 5.2.10. An $n \times n$ matrix K over $\mathbb{C}$ is called *Hermitian* if $K = K^H$. (See the previous exercise for the definition of K^H.) Show that if K is Hermitian, then $\det(K)$ is a real number.

Exercise 5.2.11. Determine whether

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

has an LDU decomposition over both $\mathbb{Q}$ and $\mathbb{F}_3$. If it does in either case, find its pivots.

Exercise 5.2.12. Suppose A is a square matrix over $\mathbb{F}$ such that each row sums to zero. Find $\det(A)$.

Exercise 5.2.13. Show that the condition in Proposition 7.27 that $F(A) = 0$ if two rows of A are equal is equivalent to the condition that $F(SA) = -F(A)$ if S is a row swap matrix.

Exercise 5.2.14. Find all values $x \in \mathbb{R}$ for which $\det(A(x)) = 0$ when

$$A(x) = \begin{pmatrix} 1 & x & 2 \\ x & 1 & x \\ 2 & 3 & 1 \end{pmatrix}.$$

Exercise 5.2.15. Repeat the previous exercise for the matrix

$$B(x) = \begin{pmatrix} 1 & x & 1 & x \\ 1 & 0 & x & 1 \\ 0 & x & 1 & 1 \\ 1 & 0 & 1 & 0 \end{pmatrix}.$$

Exercise 5.2.16. Why does condition (ii) in Theorem 7.27 imply that the determinant changes sign under a row swap?

Exercise 5.2.17. Recall that $SL(n, \mathbb{F}) = \{A \in \mathbb{F}^{n \times n} \text{ such that } \det(A) = 1\}$. Let $SL(n, \mathbb{Z})$ denote $\mathbb{Z}^{n \times n} \cap SL(n, \mathbb{Q})$. Show that $SL(n, \mathbb{Z})$ is a matrix group.

Exercise 5.2.18. Show that if $P \in \mathbb{Z}^{n \times n}$ is a row swap, then $GL(n, \mathbb{Z}) = SL(n, \mathbb{Z}) \cup PSL(n, \mathbb{Z})$.

5.3 Appendix: Further Results on Determinants

The purpose of this appendix is to briefly introduce the Laplace expansion of a determinant and explain Cramer's rule.

5.3.1 The Laplace expansion

In this section, we will obtain some further properties of the determinant, beginning with the Laplace expansion, which is the classical way of calculating an $n \times n$ determinant as a sum of $(n-1) \times (n-1)$ determinants. After the Laplace expansion, we will state Cramer's rule, which gives a closed form for inverting a nonsingular matrix. The Laplace expansion also allows one to show by induction that the determinant of a matrix with two equal rows is zero.

Suppose A is of size $n \times n$, and let A_{ij} denote the $(n-1) \times (n-1)$ submatrix obtained from A by deleting its ith row and jth column.

Theorem 5.12. *For every $A \in \mathbb{F}^{n \times n}$, we have*

$$\det(A) = \sum_{i=1}^{n} (-1)^{i+j} a_{ij} \det(A_{ij}). \tag{5.12}$$

This is the Laplace expansion along the jth column. The corresponding Laplace expansion along the ith row is

$$\det(A) = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} \det(A_{ij}). \tag{5.13}$$

Proof. Since $\det(A) = \det(A^T)$, it suffices to prove (5.12). For simplicity, we will assume $j = 1$, the other cases being similar. Now,

$$\begin{aligned}
\det(A) &= \sum_{\sigma \in S(n)} \operatorname{sgn}(\sigma)\ a_{\sigma(1)1} a_{\sigma(2)2} \cdots a_{\sigma(n)n} \\
&= a_{11} \sum_{\sigma(1)=1} \operatorname{sgn}(\sigma) a_{\sigma(2)2} \cdots a_{\sigma(n)n} + \\
&\quad a_{21} \sum_{\sigma(1)=2} \operatorname{sgn}(\sigma) a_{\sigma(2)2} \cdots a_{\sigma(n)n} + \\
&\quad + \cdots + a_{n1} \sum_{\sigma(1)=n} \operatorname{sgn}(\sigma) a_{\sigma(2)2} \cdots a_{\sigma(n)n}.
\end{aligned}$$

Suppose $\sigma(1) = r$. Let us evaluate

$$a_{r1} \sum_{\sigma(1)=r} \operatorname{sgn}(\sigma) a_{\sigma(2)2} \cdots a_{\sigma(n)n}.$$

Let P'_σ denote the element of $\mathbb{F}^{(n-1)\times(n-1)}$ obtained from P_σ by deleting the first column and the rth row. Then $P'_\sigma \in P(n-1)$, so $P'_\sigma = P_{\sigma'}$ for a unique $\sigma' \in S(n-1)$. Note that $\det(P_\sigma) = (-1)^{(r-1)} \det(P'_\sigma)$, since if bringing P'_σ to I_{n-1} by row swaps uses t steps, one needs $t + \sigma(1) - 1$ row swaps to bring P_σ to the identity. Thus,

$$\begin{aligned}\sum_{\sigma(1)=r} \operatorname{sgn}(\sigma) a_{\sigma(2)2} \cdots a_{\sigma(n)n} &= \sum_{\sigma(1)=r} (-1)^{r-1} \operatorname{sgn}(\sigma') a_{\sigma(2)2} \cdots a_{\sigma(n)n} \\ &= (-1)^{r+1} \det(A_{r1}).\end{aligned}$$

Substituting this into the above calculation for $r = 1, \ldots, n$ gives the result. □

Example 5.8. If A is of size 3×3, expanding $\det(A)$ along the first column gives

$$\det(A) = a_{11}(a_{22}a_{33} - a_{32}a_{23}) - a_{21}(a_{12}a_{23} - a_{13}a_{32}) + a_{31}(a_{12}a_{23} - a_{13}a_{22}).$$

This is the well-known formula for the triple product $\mathbf{a}_1 \cdot (\mathbf{a}_2 \times \mathbf{a}_3)$ of the rows of A.

Example 5.9. The Laplace expansion is useful for evaluating $\det(A)$ when A has entries that are functions. In fact, this situation will arise when we consider the characteristic polynomial of a square matrix A. Consider the matrix

$$C_x = \begin{pmatrix} 1-x & 2 & 0 \\ 2 & 1-x & -1 \\ 0 & -1 & 2-x \end{pmatrix}.$$

Suppose we want to find all values of $x \in \mathbb{C}$ such that C_x has rank less that 3, i.e., C_x is singular. The obvious way to proceed is to solve the equation $\det(C_x) = 0$ for x. Clearly, row operations aren't going to be of much help in finding $\det(C_x)$, so we will use Laplace, as in the previous example. Expanding along the first column gives

$$\begin{aligned}\det(C_x) &= (1-x)\big((1-x)(2-x) - (-1)(-1)\big) - 2\big(2(2-x) - 0(-1)\big) \\ &= -x^3 + 4x - 7.\end{aligned}$$

Hence C_x is singular at the three complex roots of $x^3 - 4x + 7 = 0$.

We can now finish the proof that $\det(C) = 0$ when two rows of C are identical. This will complete the proof of Theorem 5.1.

Proposition 5.13. *Suppose* $C \in \mathbb{F}^{n\times n}$ *has two equal rows. Then* $\det(C) = 0$.

Proof. Suppose the ith and jth rows of C are equal and S is the matrix that swaps these rows. Then $SC = C$, so $\det(SC) = \det(C)$, while $\det(SC) = -\det(C)$, since $\det(S) = -1$. Hence $2\det(C) = 0$. Thus, as long as the characteristic of $\mathbb{F}$ is different from 2, $\det(C) = 0$. The hard case is that in which the characteristic of $\mathbb{F}$ is 2. In this case, the signatures do not contribute to the determinant, since $1 = -1$ in $\mathbb{F}$, so

$$\det(C) = \sum_{\pi\in S(n)} c_{\pi(1)1}c_{\pi(2)2}\cdots c_{\pi(n)n}. \tag{5.14}$$

Let us now assume that the characteristic of $\mathbb{F}$ is two. By the usual formula, $\det(A) = ad - bc$ for $A \in \mathbb{F}^{2\times 2}$, it follows that if $a = c$ and $b = d$, then $\det(A) = ad - ad = 0$. Hence we may assume as our induction hypothesis that proposition holds for $n \geq 2$. Let $A \in \mathbb{F}^{(n+1)\times(n+1)}$ have two equal rows, say the first two rows. Using the Laplace expansion for $\det(A)$ along the first column, we get that $a_{11}\det(A_{11}) = a_{21}\det(A_{21})$, and the other terms are zero, since the induction hypothesis implies $\det(B) = 0$ if $B \in \mathbb{F}^{n\times n}$ has two equal rows. Consequently, $\det(A) = 2a_{11}\det(A_{11}) = 0$. This finishes the induction step, so the proposition is proved. □

Remark. As mentioned above, algebra texts often define the determinant inductively via the Laplace expansion. This avoids the problem of introducing the signature of a permutation. Leibniz was aware of the Laplace expansion, for example, well before the time of Laplace (1749–1827). The drawback of this approach is that in order for it to be of use, one needs to know that all possible row and column expansions have the same value. The only way to show that fact is to appeal to a formula such as the Leibniz formula, which avoids using rows and columns. The Laplace expansion is useful as a computational tool for matrices with few nonzero entries or when row operations are impractical, such as for a characteristic polynomial. But it would be useless to attempt to use Laplace for calculating even a 5×5 determinant, which could be done easily with row operations.

5.3.2 Cramer's Rule

Cramer's rule is a closed formula for the inverse of a square matrix with nonzero determinant. Recall that if A is of size 2×2, then

$$A^{-1} = \frac{1}{\det(A)} \begin{pmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{pmatrix}.$$

Inspecting this formula may suggest the correct formula for A^{-1} in the general case.

Definition 5.3. Suppose $A \in \mathbb{F}^{n\times n}$, and let A_{ij} denote the $(n-1)\times(n-1)$ submatrix of A obtained by deleting A's ith row and jth column. Then the matrix

$$\mathrm{Adj}(A) = \big((-1)^{i+j} \det(A_{ji})\big) \tag{5.15}$$

is called the *adjoint* of A.

Proposition 5.14. *Suppose $A \in \mathbb{F}^{n\times n}$. Then $Adj(A)A = \det(A)I_n$. Thus if $\det(A) \neq 0$, then*

$$A^{-1} = \frac{1}{\det(A)} Adj(A).$$

Proof. The essential ideas are all contained in the 3×3 case, so for simplicity, we will let $n = 3$. By definition,

$$\mathrm{Adj}(A) = \begin{pmatrix} \det(A_{11}) & -\det(A_{21}) & \det(A_{31}) \\ -\det(A_{12}) & \det(A_{22}) & -\det(A_{23}) \\ \det(A_{13}) & -\det(A_{23}) & \det(A_{33}) \end{pmatrix}.$$

Put

$$C = \begin{pmatrix} \det(A_{11}) & -\det(A_{21}) & \det(A_{31}) \\ -\det(A_{12}) & \det(A_{22}) & -\det(A_{23}) \\ \det(A_{13}) & -\det(A_{23}) & \det(A_{33}) \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}.$$

We have to show that $C = \det(A)I_n$. But it follows immediately from Theorem 5.12 that each diagonal entry of C is $\det(A)$. On the other hand, consider one of C's off-diagonal entries, say c_{21}. Expanding the above product gives

$$c_{21} = -a_{11}\det(A_{12}) + a_{21}\det(A_{22}) - a_{31}\det(A_{32}).$$

But this is exactly the Laplace expansion along the first column for the determinant of the matrix

$$-\begin{pmatrix} a_{11} & a_{11} & a_{13} \\ a_{21} & a_{21} & a_{23} \\ a_{31} & a_{31} & a_{33} \end{pmatrix}.$$

The determinant of this matrix is 0, since it has equal columns, so $c_{21} = 0$. Similarly, all c_{ij} with $i \neq j$ vanish, so $C = \det(A)I_3$. □

5.3.3 The inverse of a matrix over $\mathbb{Z}$

In most of our examples of inverting a matrix A, the entries of A are integers. But experience tells us that usually, at some time during the row operations, denominators appear. In fact, Cramer's rule, Proposition 5.14, tells us that when $\det(A) = \pm 1$, then A^{-1} also has integral entries. The question whether this is the whole story will now be answered.

Proposition 5.15. *Suppose A is an invertible matrix with integral entries. Then A^{-1} has integral entries if and only if* $\det(A) = \pm 1$.

Proof. The if statement follows from Cramer's rule. Conversely, suppose A^{-1} is integral. Then $\det(A)$ and $\det(A^{-1})$ both are integers. But

$$\det(AA^{-1}) = \det(A)\det(A^{-1}) = \det(I_n) = 1,$$

so the only possibility is that $\det(A) = \det(A^{-1}) = \pm 1$. □

A somewhat deeper fact is the following result.

Proposition 5.16. *An $n \times n$ matrix over $\mathbb{Z}$ is invertible over $\mathbb{Z}$ if and only if it can be expressed as a product of elementary matrices all of which are defined over $\mathbb{Z}$.*

We will skip the proof. Of course, row swap matrices are always integral. The restriction of sticking to elementary matrices over $\mathbb{Z}$ means that one can multiply a row only by ± 1 and replace it by itself plus an integral multiple of another row. Let $GL(n, \mathbb{Z}) = \{A \in \mathbb{Z}^{n \times n} \mid \det(A) = \pm 1\}$ and $SL(n, Z) = GL(n, \mathbb{Z}) \cap SL(n, \mathbb{R})$. The latter groups, especially $SL(2, \mathbb{Z})$, are examples of *modular groups* and are very important in number theory, topology, and complex analysis, to name a few areas where they are used.

Proposition 5.17. *$GL(n, \mathbb{Z})$ and $SL(n, \mathbb{Z})$ are matrix groups.*

Chapter 6
Vector Spaces

A vector space is a set V whose elements, called vectors, can be added and subtracted: in fact, a vector space is an abelian group under addition. A vector space also has an operation called scalar multiplication whereby the elements of a field $\mathbb{F}$ act on vectors. When we speak of a vector space, we also mention the scalars by saying that V is a vector space over $\mathbb{F}$. For example, $\mathbb{F}^n$, $\mathbb{F}^{m\times n}$ are two examples of vector spaces over $\mathbb{F}$, and the row space and null space of a matrix over $\mathbb{F}$ are two more examples. Another example is the set $C([a,b])$ of all continuous real-valued functions on a closed interval $[a,b]$, which is a vector space over $\mathbb{R}$. Here, one needs the theorem that the sum of two continuous real-valued functions on $[a,b]$ is continuous in order to speak of (vector) addition in this vector space.

One of the most important concepts associated with a vector space is its dimension. The definition of dimension requires quite a bit of preliminary groundwork. One must first introduce linear independence and the notion of a basis and then prove the fundamental result about bases: if a vector space has a finite basis, then every two bases B and B' have the same number of elements; that is, $|B| = |B'|$. We then call $|B|$ the dimension of V. This definition turns out to coincide with one's intuitive notion that a line is one-dimensional, a plane is two-dimensional, space is three-dimensional, and after that, you have to deal with objects such as spacetime, which you cannot actually picture. After covering the basic topics, we will investigate some special topics such as direct sums of subspaces, the Grassmann intersection formula, and quotient vector spaces. The reader is also advised to look at the appendix, which is an exposition of linear coding theory. This is an interesting contemporary topic involving vector spaces over finite fields.

J.B. Carrell, *Groups, Matrices, and Vector Spaces*,
DOI 10.1007/978-0-387-79428-0_6

6.1 The Definition of a Vector Space and Examples

The purpose of this section is to introduce the definition of an abstract vector space and to recall a few old examples as well as to add a few new ones. In this section, it will be useful (but not absolutely necessary) to have studied the material in Chap. 2 on groups and fields.

6.1.1 The vector space axioms

Put as succinctly as possible, a vector space over a field $\mathbb{F}$ consists of an abelian group V under addition whose elements are called vectors. Vectors admit a multiplication by the elements of $\mathbb{F}$, the field of scalars. Addition and scalar multiplication interact in a way that we will state precisely below. The definition below of a vector space is given in full detail so that it is not necessary to know the definition of an abelian group.

Definition 6.1. Let $\mathbb{F}$ be a field and suppose V is a set with a binary operation $+$ called addition assigning to any two elements $\mathbf{a}$ and $\mathbf{b}$ of V a unique sum $\mathbf{a}+\mathbf{b} \in V$. Suppose also that there is a second operation, called scalar multiplication, assigning to every $r \in \mathbb{F}$ and $\mathbf{a} \in V$ a unique scalar multiple $r\mathbf{a} \in V$. Addition and scalar multiplication have the following properties.

(1) Addition is commutative: $\mathbf{a}+\mathbf{b} = \mathbf{b}+\mathbf{a}$ for all $\mathbf{a}, \mathbf{b} \in V$.

(2) Addition is also associative: $(\mathbf{a}+\mathbf{b})+\mathbf{c} = \mathbf{a}+(\mathbf{b}+\mathbf{c})$ for all $\mathbf{a}, \mathbf{b}, \mathbf{c} \in V$.

(3) V contains an additive identity $\mathbf{0}$: that is, $\mathbf{0}+\mathbf{a} = \mathbf{a}$ for all $\mathbf{a} \in V$.

(4) For every element $\mathbf{v}$ of V, there is an element $-\mathbf{v}$ such that

$$\mathbf{v} + (-\mathbf{v}) = \mathbf{0}.$$

Thus $-\mathbf{v}$ is an additive inverse of $\mathbf{v}$.

(5) For all $\mathbf{a} \in V$, $1\mathbf{a} = \mathbf{a}$, where 1 is the multiplicative identity of $\mathbb{F}$.

(6) Scalar multiplication is associative: if $r, s \in \mathbb{F}$ and $\mathbf{a} \in V$, then

$$(rs)\mathbf{a} = r(s\mathbf{a}).$$

(7) Scalar multiplication is distributive: if $r, s \in \mathbb{F}$ and $\mathbf{a}, \mathbf{b} \in V$, then

$$r(\mathbf{a}+\mathbf{b}) = r\mathbf{a}+r\mathbf{b} \text{ and } (r+s)\mathbf{a} = r\mathbf{a}+s\mathbf{a}.$$

Then V is called a *vector space over* $\mathbb{F}$.

The first four axioms say that a vector space is an abelian group V under addition with identity $\mathbf{0}$. The properties that scalar multiplication must satisfy along with addition are specified by properties (5), (6) and (7). As proved in the group setting, the additive identity $\mathbf{0}$ and additive inverses are unique. To remind the reader, we will repeat the proofs (word for word) for vector spaces.

Proposition 6.1. *In a vector space, there can be only one zero vector. Furthermore, the additive inverse of a vector is always unique.*

Proof. Let $\mathbf{0}$ and $\mathbf{0}'$ both be additive identities. Then

$$\mathbf{0} = \mathbf{0} + \mathbf{0}' = \mathbf{0}',$$

by the definition of an additive identity. Hence the zero vector is unique. Now suppose $-\mathbf{v}$ and $-\mathbf{v}'$ are both additive inverses of $\mathbf{v} \in V$. Then

$$-\mathbf{v} = -\mathbf{v} + \mathbf{0} = -\mathbf{v} + (\mathbf{v} - \mathbf{v}') = (-\mathbf{v} + \mathbf{v}) + (-\mathbf{v}') = \mathbf{0} + (-\mathbf{v}') = -\mathbf{v}'.$$

Hence, additive inverses are also unique. □

Proposition 6.2. *In a vector space V, $0\mathbf{v} = \mathbf{0}$ for all $\mathbf{v} \in V$, and $r\mathbf{0} = \mathbf{0}$ for every scalar r. Moreover, $-\mathbf{v} = (-1)\mathbf{v}$.*

Proof. Let $\mathbf{v}$ be arbitrary. Now, by properties (4) and (7) of the definition,

$$\mathbf{v} = 1\mathbf{v} = (1 + 0)\mathbf{v} = 1\mathbf{v} + 0\mathbf{v} = \mathbf{v} + 0\mathbf{v}.$$

Adding $-\mathbf{v}$ to both sides and using associativity gives $0\mathbf{v} = \mathbf{0}$. For the second assertion, note that

$$\mathbf{0} = 0\mathbf{v} = (1 + (-1))\mathbf{v} = 1\mathbf{v} + (-1)\mathbf{v} = \mathbf{v} + (-1)\mathbf{v}.$$

Hence, $(-1)\mathbf{v}$ is an additive inverse of $\mathbf{v}$, so $(-1)\mathbf{v} = -\mathbf{v}$ for all $\mathbf{v} \in V$. □

If $\mathbf{v}_1, \ldots, \mathbf{v}_k \in V$, then one can define the sum

$$\mathbf{v}_1 + \cdots + \mathbf{v}_k = \sum_{i=1}^{k} \mathbf{v}_i$$

inductively as $(\mathbf{v}_1 + \cdots + \mathbf{v}_{k-1}) + \mathbf{v}_k$. Just as for sums in a field, the terms in this sum can be associated in any convenient way. Similarly, the summands $\mathbf{v}_i$ can be taken in any order without changing the sum, since addition is commutative. Recall that an expression $\sum_{i=1}^{k} r_i\mathbf{v}_i$, where $r_1, \ldots, r_k \in \mathbb{F}$, is called a *linear combination* of $\mathbf{v}_1, \ldots, \mathbf{v}_k \in V$.

6.1.2 Examples

Example 6.1. As mentioned above, the basic example of a vector space over the field $\mathbb{F}$ is the set $\mathbb{F}^n$ of all column n-tuples of elements of $\mathbb{F}$, where addition and scalar multiplication are carried out componentwise:

$$\mathbf{a}+\mathbf{b}=\begin{pmatrix}a_1\\a_2\\\vdots\\a_n\end{pmatrix}+\begin{pmatrix}b_1\\b_2\\\vdots\\b_n\end{pmatrix}=\begin{pmatrix}a_1+b_1\\a_2+b_2\\\vdots\\a_n+b_n\end{pmatrix}$$

and

$$r\mathbf{a}=r\begin{pmatrix}a_1\\a_2\\\vdots\\a_n\end{pmatrix}=\begin{pmatrix}ra_1\\ra_2\\\vdots\\ra_n\end{pmatrix}.$$

The elements of $\mathbb{F}^n$ are called column vectors. ⊠

Example 6.2. Let $\mathbb{F}^{m\times n}$ denote the $m\times n$ matrices over $\mathbb{F}$. In the previous example, we considered $\mathbb{F}^n$, which is the same as $\mathbb{F}^{n\times 1}$. The elements of $\mathbb{F}^{1\times n}$ are called row vectors. We have already defined matrix addition and scalar multiplication for $m\times n$ in an analogous way in Chap. 3, so we refer the reader there. These operations make $\mathbb{F}^{m\times n}$ a vector space over $\mathbb{F}$. One can express the elements of $\mathbb{F}^{mn}$ as matrices, so as a vector space, $\mathbb{F}^{m\times n}$ is indistinguishable from $\mathbb{F}^{mn}$. ⊠

Example 6.3. (See Example 3.2.) When $\mathbb{F}=\mathbb{F}_2$, the elements of $\mathbb{F}^n$ are binary strings, which are called *n-bit strings*. Binary strings are usually written as rows instead of columns, and the commas between components are omitted. For example, there are 2^3 3-bit strings, 000, 100, 010, 001, 110, 101, 011, and 111. Binary strings are the fundamental objects of coding theory, but one can just as well consider p-ary strings of length n, namely elements of the vector space $(\mathbb{F}_p)^n$ written as row vectors as in the binary case. A common notation for $(\mathbb{F}_p)^n$ is $V(n,p)$. Thus $V(n,p)$ consists of the p^n strings $a_1a_2\ldots a_n$ where each $a_i\in\mathbb{F}_p$. ⊠

Example 6.4. Let S be any set and define $\mathbb{F}^S$ to be the set of all $\mathbb{F}$-valued functions whose domain is S. We define addition and scalar multiplication pointwise as follows. If $\mu,\phi\in\mathbb{F}^S$, then $\mu+\phi\in\mathbb{F}^S$ is defined by the condition

$$(\mu+\phi)(s)=\mu(s)+\phi(s)$$

for all $s\in S$. Also, if $a\in\mathbb{F}$, then $a\mu$ is defined by

$$(a\mu)(s) = a\mu(s)$$

for all $s \in S$. These operations make $\mathbb{F}^S$ a vector space over $\mathbb{F}$. Notice that $\mathbb{F}^n$ is nothing but $\mathbb{F}^S$, where $S = \{1, 2, \ldots, n\}$. Indeed, specifying the n-tuple $\mathbf{a} = (a_1, a_2, \ldots, a_n)^T \in \mathbb{F}^n$ is the same as defining the function $f_{\mathbf{a}} : S \to \mathbb{F}$ by $f_{\mathbf{a}}(i) = a_i$. ⊠

Example 6.5. The set $\mathcal{P}_n$ of all polynomial functions with domain $\mathbb{R}$ and degree at most n consists of all functions $p : \mathbb{R} \to \mathbb{R}$ such that for every $r \in \mathbb{R}$,

$$p(r) = a_n r^n + a_{n-1} r^{n-1} + \cdots + a_1 r + a_0,$$

where the coefficients $a_0, a_1, \ldots, a_n$ are fixed elements of $\mathbb{R}$. If we let $x : \mathbb{R} \to \mathbb{R}$ be the identity function defined by $x(r) = r$ for all $r \in \mathbb{R}$, then the above polynomial can be written $p(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0$. Then $\mathcal{P}_n$ is a real vector space. ⊠

Example 6.6. (Polynomials over $\mathbb{F}$) For an arbitrary field $\mathbb{F}$, let x denote a quantity that admits multiplication by itself so that $x^i x^j = x^{i+j}$ and scalar multiples ax^k for all $a \in \mathbb{F}$ and all integers $i, j, k \geq 0$. Let $\mathbb{F}[x]$ denote the set of all polynomial expressions

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

where the coefficients $a_0, a_1, \ldots, a_n \in \mathbb{F}$, and n is an arbitrary nonnegative integer. The polynomials of degree at most n can be identified with $\mathbb{F}^{n+1}$, so there is a well defined addition and scalar multiplication on $\mathbb{F}[x]$ so that polynomials are added by adding the coefficients of x^i for all i. ⊠

Example 6.7. The set $C[a, b]$ of all continuous real-valued functions on $[a, b]$ with the usual pointwise addition and scalar multiplication of Example 6.4 is a slightly more exotic example of a vector space. To see that $C[a, b]$ is closed under addition and scalar multiplication requires knowing a basic theorem from calculus: the sum of two continuous functions is continuous, and any scalar multiple of a continuous function is continuous. Hence $f + g$ and rf belong to $C[a, b]$ for all f and g in $C[a, b]$ and real scalars r. ⊠

Example 6.8. Consider the differential equation

$$y'' + ay' + by = 0, \tag{6.1}$$

where a and b are real constants. This is an example of a homogeneous linear second-order differential equation with constant coefficients. The set of twice differentiable functions on $\mathbb{R}$ that satisfy (6.1) is a real vector space. ⊠

Exercises

Exercise 6.1.1. Suppose V is a vector space over the field $\mathbb{F}$. Show that if $\mathbf{v}$ is a nonzero element of V and a is a scalar such that $a\mathbf{v} = \mathbf{0}$, then $a = 0$. Conclude that if $a\mathbf{v} = b\mathbf{v}$, where $a, b \in \mathbb{F}$, then $a = b$.

Exercise 6.1.2. Suppose V is a vector space over the field $\mathbb{F}$, and $a \in \mathbb{F}$ is nonzero. Let $\mu_a : V \to V$ be defined by $\mu_a(\mathbf{v}) = a\mathbf{v}$. Show that μ_a is a bijection.

Exercise 6.1.3. Consider the set S of all $A \in \mathbb{F}^{n\times n}$ for which $A = A^T$. True or false: S is a vector space.

Exercise 6.1.4. Let $A \in \mathbb{F}^{m\times n}$.

(*i*) Show that $\mathcal{N}(A) = \{\mathbf{v} \in \mathbb{F}^n \mid A\mathbf{v} = \mathbf{0}\}$ is a vector space.

(*ii*) Suppose $\mathbf{b} \in \mathbb{F}^n$. When is the solution set $\{\mathbf{x} \mid A\mathbf{x} = \mathbf{b}\}$ a vector space?

Exercise 6.1.5. Is the unit circle $x^2 + y^2 = 1$ in $\mathbb{R}^2$ a vector space?

Exercise 6.1.6. When is a line $ax + by = c$ in $\mathbb{R}^2$ a vector space?

6.2 Subspaces and Spanning Sets

The purpose of this section is to introduce, study, and give examples of subspaces and spanning sets. Throughout this section, let V be a vector space over an arbitrary field $\mathbb{F}$.

Definition 6.2. A nonempty subset W of V is called a *linear subspace of* V, or simply a *subspace*, provided the following two conditions hold for all $\mathbf{a}, \mathbf{b} \in W$:

(i) $\mathbf{a} + \mathbf{b} \in W$, and
(ii) $r\mathbf{a} \in W$ whenever $r \in \mathbb{F}$.

In particular, every subspace of a vector space contains the zero vector $\mathbf{0}$. In fact, $\{\mathbf{0}\}$ is itself a subspace, called the *trivial subspace*. The following proposition is an immediate consequence of this definition.

Proposition 6.3. *A subspace W of V is also a vector space over $\mathbb{F}$ under the addition and scalar multiplication induced from V.*

We leave the proof to the reader.

Example 6.9. Suppose $A \in \mathbb{F}^{m\times n}$. Then the solution set $\mathcal{N}(A)$ of the homogeneous equation $A\mathbf{x} = \mathbf{0}$ is a subspace of $\mathbb{F}^n$. For if $A\mathbf{x}_i = \mathbf{0}$ for $i = 1, 2$, then

$$A(\mathbf{x}_1 + \mathbf{x}_2) = A\mathbf{x}_1 + A\mathbf{x}_2 = \mathbf{0} + \mathbf{0} = \mathbf{0},$$

while

$$A(r\mathbf{x}) = rA\mathbf{x} = r\mathbf{0} = \mathbf{0}.$$

Therefore, $\mathcal{N}(A)$ is indeed a subspace. ⊠

Example 6.10. The subspaces of $\mathbb{R}^2$ are easily described. They are $\{\mathbf{0}\}$, every line through $\mathbf{0}$, and $\mathbb{R}^2$ itself. We will consider the subspaces of $\mathbb{R}^3$ below. This example shows that subspaces when viewed as geometric objects need to be linear; that is, they need to be lines, planes, etc., through the zero vector $\mathbf{0}$. ⊠

Example 6.11. Let $\mathbb{F}$ be a field and suppose $\mathbb{F}'$ is a subfield of $\mathbb{F}$. Then $\mathbb{F}$ is a vector space over $\mathbb{F}'$, vectors being the elements of $\mathbb{F}$ and scalars being the elements of $\mathbb{F}'$. Note that elements of $\mathbb{F}'$ are also vectors, but that is irrelevant. In particular, a field is a vector space over itself. For example, $\mathbb{R}$ is a vector space over $\mathbb{Q}$, and $\mathbb{C}$ is a vector space over $\mathbb{R}$ and also a vector space over $\mathbb{Q}$. Note that $\mathbb{C}$ as a vector space over $\mathbb{R}$ is very different from $\mathbb{C}$ as a vector space over $\mathbb{Q}$. It will turn out that the dimension of $\mathbb{C}$ as a vector space over $\mathbb{R}$ is two, while its dimension as a vector space over $\mathbb{Q}$ is not finite (although we have yet to define what the dimension of a vector space is). ⊠

6.2.1 Spanning sets

We will now consider the most basic method for constructing subspaces.

Definition 6.3. Let $\mathbf{v}_1, \dots, \mathbf{v}_k$ be arbitrary elements of V. The *span* of $\mathbf{v}_1, \dots, \mathbf{v}_k$ is by definition the subset of V consisting of all linear combinations

$$\sum_{i=1}^{k} a_i \mathbf{v}_i,$$

where $a_1, \dots, a_k$ are arbitrary elements of $\mathbb{F}$. The span of $\mathbf{v}_1, \dots, \mathbf{v}_k$ is denoted by $\operatorname{span}\{\mathbf{v}_1, \dots, \mathbf{v}_k\}$.

Proposition 6.4. *For all $\mathbf{v}_1, \dots, \mathbf{v}_k$ in V, $\operatorname{span}\{\mathbf{v}_1, \dots, \mathbf{v}_k\}$ is a subspace of V.*

This follows readily from the definitions, so we will skip the details. However, the reader might benefit from writing them all out.

Example 6.12 (lines and planes). The subspace spanned by a single nonzero vector $\mathbf{v}$ is called the *line spanned by* $\mathbf{v}$. Thus, the line spanned by $\mathbf{v}$ consists of all scalar multiples $a\mathbf{v}$ with $a \in \mathbb{F}$, so we will frequently denote it by $\mathbb{F}\mathbf{v}$ instead of $\operatorname{span}\{\mathbf{v}\}$. A pair of vectors in V are said to be *noncollinear* if neither lies on the line spanned by the other; in other words, their span isn't a line. A subspace spanned by two noncollinear vectors $\mathbf{u}, \mathbf{v}$ is called a *plane*. Thus a plane always has the form $\operatorname{span}\{\mathbf{u}, \mathbf{v}\}$, but $\operatorname{span}\{\mathbf{u}, \mathbf{v}\}$ isn't necessarily a plane. It may be a line or even $\{\mathbf{0}\}$. ⊠

Example 6.13. Let $\mathbb{F} = \mathbb{F}_p$. If $\mathbf{v} \in \mathbb{F}^n$, one can ask how many elements $\mathbb{F}\mathbf{v}$ has. If $\mathbf{v} = \mathbf{0}$, the answer is clearly one. Otherwise, recall from Exercise 6.1.1 that if $a, b \in \mathbb{F}$ and $a \neq b$, then $a\mathbf{v} \neq b\mathbf{v}$. Consequently, the multiples of $\mathbf{v}$ are all distinct, and therefore $|\mathbb{F}\mathbf{v}| = |\mathbb{F}| = p$. If $\mathbf{a}$ and $\mathbf{b}$ are noncollinear elements of $\mathbb{F}_p$, then the plane they span has p^2 elements. More generally, we can ask how many elements an arbitrary subspace of $\mathbb{F}^n$ has. ⊠

Example 6.14. We saw in Example 6.9 that the null space $\mathcal{N}(A)$ of a matrix $A \in \mathbb{F}^{m \times n}$ is a subspace of $\mathbb{F}^n$. Recall from Example 3.11 that $\mathcal{N}(A)$ is spanned by the basic null vectors $\mathbf{f}_1, \dots, \mathbf{f}_k$ obtained from the reduced row echelon form A_{red} of A. Recall also that $k = n - \operatorname{rank}(A)$, where $\operatorname{rank}(A)$ is the number of nonzero rows in A_{red}. Now suppose $\mathbb{F} = \mathbb{F}_p$. Since every $\mathbf{v} \in \mathcal{N}(A)$ has an expansion $\mathbf{v} = a_1\mathbf{f}_1 + \cdots + a_k\mathbf{f}_k$ with all $a_i \in \mathbb{F}$, it follows that $|\mathcal{N}(A)| \leq p^k$. ⊠

The next two examples illustrate the other two subspaces associated with a matrix.

Example 6.15. Recall from Section 3.2.3 that if $A \in \mathbb{F}^{m\times n}$, then $\text{row}(A)$ is defined as the set of all linear combinations of the rows of A. Hence $\text{row}(A)$ is a subspace of $\mathbb{F}^n = \mathbb{F}^{1\times n}$. Recall that if B is obtained from A by row operations, then $\text{row}(A) = \text{row}(B)$. One of our main results on matrix theory is that two matrices in reduced row echelon form are equal if and only if they have the same row space. Because of this fact, we called the number of nonzero rows in A's reduced row echelon form the rank of A. ⊠

Example 6.16. A matrix $A \in \mathbb{F}^{m\times n}$ also has a column space $\text{col}(A)$: namely, the set of all linear combinations of the columns of A. The column space is a subspace of $\mathbb{F}^m = \mathbb{F}^{m\times 1}$. The column space has an important interpretation in terms of linear systems. If A has columns $\mathbf{a}_1, \ldots, \mathbf{a}_n$, then by definition, $\mathbf{b} \in \text{col}(A)$ if and only if there are scalars $c_1, \ldots, c_n$ such that $\mathbf{b} = c_1\mathbf{a}_1 + \cdots + c_n\mathbf{a}_n$. Thus the column space of A consists of all $\mathbf{b} \in \mathbb{F}^m$ for which the linear system $A\mathbf{x} = \mathbf{b}$ has a solution. ⊠

Example 6.17. Assume that $\mathbf{a}$ and $\mathbf{b}$ are vectors in $\mathbb{F}^3$. The *cross product* $\mathbf{a} \times \mathbf{b}$ is defined to be

$$\mathbf{a} \times \mathbf{b} = (a_2b_3 - a_3b_2, -(a_1b_3 - a_3b_1), a_1b_2 - a_2b_1)^T. \tag{6.2}$$

By direct calculation, $\mathbf{a}^T(\mathbf{a}\times\mathbf{b}) = \mathbf{b}^T(\mathbf{a}\times\mathbf{b}) = 0$. Furthermore, if $\mathbf{a}$ and $\mathbf{b}$ are noncollinear, it can be seen that $\mathbf{a} \times \mathbf{b} \neq \mathbf{0}$. We thus obtain a homogeneous equation satisfied by all vectors in $\text{span}\{\mathbf{a}, \mathbf{b}\}$: if $\mathbf{a} \times \mathbf{b} = (r, s, t)^T$, then such an equation is $rx + sy + tz = 0$. In the case $\mathbb{F} = \mathbb{R}$, this is interpreted as meaning that $\mathbf{a} \times \mathbf{b}$ is orthogonal to the plane spanned by $\mathbf{a}$ and $\mathbf{b}$. ⊠

Exercises

Exercise 6.2.1. Which of the following subsets of $\mathbb{R}^2$ are not subspaces?

(*i*) The line $x = y$;

(*ii*) The unit circle;

(*iii*) The line $2x + y = 1$;

(*iv*) The first octant $x, y \geq 0$.

Exercise 6.2.2. Prove that all lines through the origin and planes through the origin in $\mathbb{R}^3$ are subspaces.

Exercise 6.2.3. Let p be a prime. Show that a subset of $V(n, p)$ that is closed under addition is a subspace.

Exercise 6.2.4. Assume $n > 1$. Find a subspace of $V(n, p)$ that contains p points. Next find a subspace that contains p^2 points.

Exercise 6.2.5. Show that the plane $x + y + z = 0$ in $V(3, 2)$ has four elements. Find a spanning set for this plane.

Exercise 6.2.6. Find an equation for the plane in $V(3, 2)$ through the origin containing both $(1, 1, 1)$ and $(0, 1, 1)$.

Exercise 6.2.7. Find a spanning set in $V(4, 2)$ for the solution space of the equation $w + x + y + z = 0$. How many solutions in $V(4, 2)$ does this equation have?

Exercise 6.2.8. Find a spanning set for the plane $3ix - y + (2 - i)z = 0$ in $\mathbb{C}^3$.

Exercise 6.2.9. Find an equation for the plane in $\mathbb{R}^3$ through the origin containing both $(1, 2, -1)^T$ and $(3, 0, 1)^T$.

Exercise 6.2.10. Describe all subspaces of $\mathbb{C}^3$. What about $\mathbb{C}^4$?

Exercise 6.2.11. Find the number of subspaces of the vector space $V(n, p)$ in the following cases:

(i) $n = p = 2$;

(ii) $n = 2$, $p = 3$; and

(iii) $n = 3$, $p = 2$.

Exercise 6.2.12. Let $\mathbb{F}$ be an arbitrary field. Show that if $\mathbf{a}, \mathbf{b} \in \mathbb{F}^3$, then $\mathbf{a} \times \mathbf{b} \neq \mathbf{0}$ if and only if $\mathbf{a}$ and $\mathbf{b}$ are not collinear.

6.3 Linear Independence and Bases

As usual, let V denote a vector space over an arbitrary field $\mathbb{F}$. In order to understand the structure of V, especially its dimension, we need to introduce two new ideas: linear independence and bases. Linear independence is about uniquely representing the elements of V as linear combinations, and the notion of a basis concerns both linear independence and spanning the whole of V. As we mentioned in the introduction, the notion of the dimension of V depends on the existence of a basis.

6.3.1 The definition of linear independence

To put it informally, a nonempty set of vectors is *linearly independent* if no one of them is a linear combination of the others. For example, two vectors are linearly independent if they aren't collinear, and three vectors are linearly independent if they don't all lie in a plane through the origin. That is, they aren't coplanar. In general, two, three, or any finite number of vectors fail to be linearly independent when they are subject to a linear constraint. Let us now state this formally.

Definition 6.4. Let $\mathbf{v}_1, \ldots, \mathbf{v}_k$ be in V. Then we say that $\mathbf{v}_1, \ldots, \mathbf{v}_k$ are *linearly independent* (or simply *independent*) if the equation

$$a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_k\mathbf{v}_k = \mathbf{0}, \tag{6.3}$$

with $a_1, a_2, \ldots, a_k \in \mathbb{F}$, is satisfied only when $a_1 = a_2 = \cdots = a_k = 0$. If (6.3) has a nontrivial solution (i.e., some $a_i \neq 0$), we say that $\mathbf{v}_1, \ldots, \mathbf{v}_k$ are *linearly dependent* (or simply *dependent*).

We will also say that a nonempty finite subset S of V is independent or dependent if the vectors that are its elements are respectively independent or dependent. Notice that if two of the $\mathbf{v}_i$ coincide or if one of them is zero, then $\mathbf{v}_1, \ldots, \mathbf{v}_k$ are dependent. Notice also that we are defining linear independence only for a finite number of vectors. The reader might want to contemplate how to do this for infinite sets. Another way to think about independence is pointed out by the following proposition.

Proposition 6.5. *A set of vectors is linearly dependent if and only if one of them can be expressed as a linear combination of the others.*

Proof. Let $\mathbf{v}_1, \ldots, \mathbf{v}_k$ be the vectors, and suppose one of the vectors, say $\mathbf{v}_1$, is a linear combination of the others. Then

$$\mathbf{v}_1 = a_2\mathbf{v}_2 + \cdots + a_k\mathbf{v}_k.$$

Thus

$$\mathbf{v}_1 - a_2\mathbf{v}_2 - \cdots - a_k\mathbf{v}_k = \mathbf{0},$$

so (6.3) has a nontrivial solution with $a_1 = 1$. Therefore, $\mathbf{v}_1, \ldots, \mathbf{v}_k$ are dependent. Conversely, suppose $\mathbf{v}_1, \ldots, \mathbf{v}_k$ are dependent. This means that there is a nontrivial solution $a_1, a_2, \ldots, a_k$ of (6.3). We can assume (by reindexing the vectors) that $a_1 \neq 0$. Thus

$$\mathbf{v}_1 = b_2\mathbf{v}_2 + \cdots + b_k\mathbf{v}_k,$$

where $b_i = -a_i/a_1$, for $i \geq 2$, so the proof is done. □

Note how the fact that $\mathbb{F}$ is a field was used in the above proof. The next proposition gives one of the important properties of linearly independent sets.

Proposition 6.6. *Assume that $\mathbf{v}_1, \ldots, \mathbf{v}_k \in V$ are linearly independent, and suppose $\mathbf{v}$ is in their span. Then $\mathbf{v} = \sum_{i=1}^{k} a_i\mathbf{v}_i$ for exactly one linear combination of $\mathbf{v}_1, \ldots, \mathbf{v}_k$.*

Proof. By assumption, there exists an expression

$$\mathbf{v} = r_1\mathbf{v}_1 + r_2\mathbf{v}_2 + \cdots + r_k\mathbf{v}_k,$$

where $r_1, \ldots, r_k \in \mathbb{F}$. Suppose there is another expression, say

$$\mathbf{v} = s_1\mathbf{v}_1 + s_2\mathbf{v}_2 + \cdots + s_k\mathbf{v}_k,$$

where the s_i are also elements of $\mathbb{F}$. By subtracting the second expression from the first and collecting terms, we get that

$$\mathbf{0} = \mathbf{v} - \mathbf{v} = (r_1 - s_1)\mathbf{v}_1 + (r_2 - s_2)\mathbf{v}_2 + \cdots + (r_k - s_k)\mathbf{v}_k.$$

Since the $\mathbf{v}_i$ are independent, every coefficient $r_i - s_i$ is equal to 0. □

When $V = \mathbb{F}^m$, checking linear independence involves solving a homogeneous linear system. Viewing vectors in $\mathbb{F}^m$ as column vectors, consider the $m \times n$ matrix

$$A = (\mathbf{a}_1 \ \cdots \ \mathbf{a}_n).$$

By the theory of linear systems, we have the following result.

Proposition 6.7. *The vectors $\mathbf{a}_1, \ldots, \mathbf{a}_n$ in $\mathbb{F}^m$ are linearly independent exactly when the system $A\mathbf{x} = \mathbf{0}$ has only the trivial solution, that is, when $\mathcal{N}(A) = \{\mathbf{0}\}$. In particular, the columns $\mathbf{a}_1, \ldots, \mathbf{a}_n$ of A are independent if and only if rank$(A) = n$, so more than m vectors in $\mathbb{F}^m$ are linearly dependent.*

Proof. The first statement follows from the definitions. The second follows from the identity $\operatorname{rank}(A) + \#$ free variables $= n$. Since $\operatorname{rank}(A) \leq m$, there exist free variables whenever $n > m$, so $\mathcal{N}(A) \neq \{\mathbf{0}\}$. Thus $\mathbf{a}_1, \ldots, \mathbf{a}_n$ are dependent when $n > m$. □

6.3.2 The definition of a basis

We will now explain the second main ingredient in the notion of dimension. A basis combines the notions of independence and spanning. From now on, we will restrict our attention to vector spaces that have a finite spanning set. Thus we make the following definition.

Definition 6.5. A vector space V is said to be *finite-dimensional* if V has a finite spanning set.

Note that the trivial vector space is finite-dimensional, since it is spanned by $\mathbf{0}$. Vector spaces such as the space of all continuous functions on $[a, b]$ are therefore excluded from our considerations. Here, finally, is the definition of a basis.

Definition 6.6. A collection of vectors in V that is linearly independent and spans V is called a *basis* of V.

One of the main results we will prove is that every finite-dimensional vector space has a basis. In fact, every vector space has a basis, but the proof that a spanning set exists in the infinite-dimensional case is beyond our scope. For the remainder of this section, we will consider examples of bases.

Example 6.18 (The standard basis of $\mathbb{F}^n$). Recall that $\mathbf{e}_i$ denotes the ith column of I_n. Then $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ is called the *standard basis* of $\mathbb{F}^n$. Since

$$\begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} = a_1\mathbf{e}_1 + \cdots + a_n\mathbf{e}_n$$

and I_n has rank n, it follows that $\mathbf{e}_1, \ldots, \mathbf{e}_n$ indeed give a basis of $\mathbb{F}^n$. ⊠

Example 6.19 (Lines and planes). A nonzero vector in $\mathbb{R}^n$ spans a line through $\mathbf{0}$, and clearly a single nonzero vector is linearly independent. Hence a line through $\mathbf{0}$ has a basis consisting of a single element. (The choice of basis is not unique unless $\mathbb{F} = \mathbb{F}_2$.) A plane P containing the origin is spanned by any pair of noncollinear vectors in P, and two noncollinear vectors in P are linearly independent. Thus P has a basis consisting of two vectors. (The choice of basis is again not unique unless $\mathbb{F} = \mathbb{F}_2$.) ⊠

It should be noted that the trivial vector space $\{\mathbf{0}\}$ does not have a basis, since in order to contain a linearly independent subset it has to contain a nonzero vector. The next result gives an elementary but useful property of bases.

Proposition 6.8. *The vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ in V form a basis of V if and only if every vector $\mathbf{v}$ in V admits a unique expression*

$$\mathbf{v} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_n\mathbf{v}_n,$$

where $a_1, a_2, \ldots, a_n$ are elements of $\mathbb{F}$.

Proof. We leave this as an exercise. □

Remark. Proposition 6.8 shows that a basis $\mathcal{B}$ of V sets up a one-to-one correspondence, or bijection, $\Phi_{\mathcal{B}} : V \to \mathbb{F}^n$ by $\Phi_{\mathcal{B}}(\mathbf{v}) = (a_1, a_2, \ldots, a_n)^T$. The mapping $\Phi_{\mathcal{B}}$ also preserves linear combinations; that is, $\Phi_{\mathcal{B}}(a\mathbf{v} + b\mathbf{w}) = a\Phi_{\mathcal{B}}(\mathbf{v}) + b\Phi_{\mathcal{B}}(\mathbf{w})$ for all $a, b \in \mathbb{F}$ and $\mathbf{v}, \mathbf{w} \in V$. In particular, $\Phi_{\mathcal{B}}$ is an isomorphism in the sense of abelian groups between V and $\mathbb{F}^n$. The n-tuple $(a_1, a_2, \ldots, a_n)^T$ assigns coordinates to $\mathbf{v}$ with respect to the basis $\mathcal{B}$. We will discuss coordinates in detail in the next chapter.

Proposition 6.9. *Let $A \in \mathbb{F}^{m\times n}$. If A has rank m, its rows are a basis of* $\mathrm{row}(A)$. *If A has rank n, its columns are a basis of* $\mathrm{col}(A)$.

Proof. We leave the proof to the reader. □

If the rank of A is less than n, Proposition 6.7 tells us that its columns are dependent. However, the columns still span $\mathrm{col}(A)$, so it is natural to ask whether there is a basis of $\mathrm{col}(A)$ consisting of some of A's columns. This is the problem of finding a basis of V contained in a spanning set. The solution is treated below.

Example 6.20 (Basis of the null space). Let A be an $m \times n$ matrix over $\mathbb{F}$. The basic null vectors span the null space $\mathcal{N}(A)$. They are also independent, as can be seen by writing out the equation for independence and looking at the corner components. Thus the basic null vectors determine a basis of $\mathcal{N}(A)$. ⊠

Exercises

Exercise 6.3.1. Determine whether $(0, 2, 1, 0)^T$, $(1, 0, 0, 1)^T$, and $(1, 0, 1, 1)^T$ are linearly independent as vectors in $\mathbb{R}^4$. If so, do they form a basis of $\mathbb{R}^4$?

Exercise 6.3.2. Are $(0, 0, 1, 0)^T$, $(1, 0, 0, 1)^T$, and $(1, 0, 1, 1)^T$ independent in $V(4, 2) = (\mathbb{F}_2)^4$?

Exercise 6.3.3. Show that every nonempty subset of a linearly independent set is linearly independent.

Exercise 6.3.4. We say that $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k \in \mathbb{R}^n$ are mutually orthogonal unit vectors if $\mathbf{u}_i^T \mathbf{u}_j = 0$ whenever $i \neq j$ and $\mathbf{u}_i^T \mathbf{u}_i = 1$ for all i. Show that if $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k$ are mutually orthogonal unit vectors, then they are linearly independent.

Exercise 6.3.5. Show that m linearly independent vectors in $\mathbb{F}^m$ are a basis.

Exercise 6.3.6. Prove the assertion made in Example 6.20 that the basic null vectors are a basis of $\mathcal{N}(A)$.

Exercise 6.3.7. Use the theory of linear systems to show the following:

(*i*) More than m vectors in $\mathbb{F}^m$ are dependent.

(*ii*) Fewer than m vectors in $\mathbb{F}^m$ cannot span $\mathbb{F}^m$.

Exercise 6.3.8. Let $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ be a basis of $\mathbb{R}^3$.

(*i*) Determine whether $3\mathbf{u} + 2\mathbf{v} + \mathbf{w}$, $\mathbf{u} + \mathbf{v} + 0\mathbf{w}$, and $-\mathbf{u} + 2\mathbf{v} - 3\mathbf{w}$ are independent.

(*ii*) Do the vectors in part (*i*) span $\mathbb{R}^3$? Supply reasoning.

(*iii*) Find a general necessary and sufficient condition for the vectors $a_1\mathbf{u} + a_2\mathbf{v} + a_3\mathbf{w}$, $b_1\mathbf{u} + b_2\mathbf{v} + b_3\mathbf{w}$, and $c_1\mathbf{u} + c_2\mathbf{v} + c_3\mathbf{w}$ to be independent, where the a_i, b_j, c_k are arbitrary scalars.

Exercise 6.3.9. Suppose V is a vector space that contains an infinite subset S such that every finite nonempty subset of S is linearly independent. Show that V cannot be a finite-dimensional vector space.

Exercise 6.3.10. Recall that $\mathbb{R}[x]$ denotes the space of all polynomials with coefficients in $\mathbb{R}$. For each positive integer m, let $\mathbb{R}[x]_m \subset \mathbb{R}[x]$ denote the subset of all polynomials of degree at most m.

(*i*) Show that $\mathbb{R}[x]_m$ is a subspace of $\mathbb{R}[x]$.

(*ii*) Show that the powers $1, x, \ldots, x^m$ are linearly independent for all positive integers m. (Hint: use induction.)

(*iii*) Find a basis for each $\mathbb{R}[x]_m$.

(*iv*) Show that $\mathbb{R}[x]$ is not finite-dimensional.

(*v*) Exhibit (without proof) a basis for $\mathbb{R}[x]$.

Exercise 6.3.11. Suggest a definition for the notion of a basis of a vector space that isn't finite-dimensional.

Exercise 6.3.12. True or false: $\mathbf{v}_1, \ldots, \mathbf{v}_r \in V(n, 2)$ are linearly independent if and only if $\mathbf{v}_1 + \cdots + \mathbf{v}_r \neq \mathbf{0}$.

Exercise 6.3.13. Suppose $\mathcal{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is a basis of V. Consider the mapping $\Phi_{\mathcal{B}} : V \to \mathbb{F}^n$ defined by $\Phi_{\mathcal{B}}(\mathbf{v}) = (a_1, a_2, \ldots, a_n)^T$ if $\mathbf{v} = a_1\mathbf{v}_1 + \cdots + a_n\mathbf{v}_n$. Show that $\Phi_{\mathcal{B}}$ is a bijection that preserves linear combinations; that is, $\Phi_{\mathcal{B}}(a\mathbf{v} + b\mathbf{w}) = a\Phi_{\mathcal{B}}(\mathbf{v}) + b\Phi_{\mathcal{B}}(\mathbf{w})$ for all $a, b \in \mathbb{F}$ and $\mathbf{v}, \mathbf{w} \in V$. Conclude that $\Phi_{\mathcal{B}}$ is an isomorphism in the sense of abelian groups between V and $\mathbb{F}^n$.

6.4 Bases and Dimension

Much of the groundwork for the definition of the dimension of a finite-dimensional vector space has now been laid, and the ever alert reader has undoubtedly guessed that the dimension is the number of vectors in a basis. However, there is still the question whether a basis exists and whether all bases have the same number of elements. It turns out that answering this question is nontrivial.

6.4.1 The definition of dimension

Throughout this section, V denotes a finite-dimensional vector space over a field $\mathbb{F}$. Here is one of the most important definitions in the theory of vector spaces.

Definition 6.7. Suppose $V \neq \{\mathbf{0}\}$. Then the *dimension* of V is defined to be the number of elements in a basis of V. The dimension of V will be denoted by $\dim V$ or by $\dim_{\mathbb{F}} V$ in case there is a chance of confusion about which field is being considered. When $V = \{\mathbf{0}\}$, we will define the dimension of V to be 0.

As already mentioned, this definition is based on two assertions: every nontrivial finite-dimensional vector space V has a basis, and any two bases have the same number of elements. These claims, being far from obvious, need to be proved. They comprise the dimension theorem, which will be stated and proved below.

As we already noted above (also see Exercise 6.3.7), $\mathbb{F}^n$ can't contain more than n independent vectors, and fewer than n vectors can't span. This implies that every basis of $\mathbb{F}^n$ has n elements. We also know that there is a basis with n vectors, namely the standard basis. This implies the following.

Proposition 6.10. *For every field $\mathbb{F}$, $\mathbb{F}^n$ has a basis, and every basis has n elements. Thus,* $\dim \mathbb{F}^n = n$.

An intuitive interpretation of $\dim V$ is either the maximal number of independent vectors in V or the minimal number of spanning vectors, provided these are the same. This definition certainly gives the correct result for $\mathbb{F}^n$ as just noted.

There is a subtlety in the definition of dimension that is worth pointing out. Namely, V can often be viewed as a vector space over different fields, so the field has to be specified when talking about V's dimension. In fact, suppose V is a finite-dimensional vector space over $\mathbb{F}$. When $\mathbb{F}'$ is a subfield of $\mathbb{F}$, then V is automatically also a vector space over $\mathbb{F}'$. In this case, $\dim_{\mathbb{F}} V$

and $\dim_{\mathbb{F}'} V$ will be different if $\mathbb{F} \neq \mathbb{F}'$. For example, let $\mathbb{F} = \mathbb{C}$ and $V = \mathbb{C}^n$. By the previous proposition, $\dim_{\mathbb{C}} V = n$. But $\mathbb{R}$ is a subfield of $\mathbb{C}$, and in fact $\mathbb{C}^n$ is still a finite-dimensional vector space over $\mathbb{R}$. Indeed, since $\mathbb{C} = \mathbb{R}^2$, $\mathbb{C}^n = \mathbb{R}^{2n}$. Thus, $\dim_{\mathbb{R}} \mathbb{C}^n = 2n$. In general, if V is a finite-dimensional vector space over $\mathbb{C}$ with $\dim_{\mathbb{C}} V = n$, then $\dim_{\mathbb{R}} V = 2n$.

6.4.2 Some examples

Before proving the dimension theorem, let's consider some examples.

Example 6.21 (Lines and Planes). Let $V = \mathbb{F}^n$. If $\mathbf{v} \neq \mathbf{0}$, then a line $\mathbb{F}\mathbf{v}$ in V has a basis consisting of $\mathbf{v}$ and hence has dimension one. If $P = \text{span}\{\mathbf{a}_1, \mathbf{a}_2\}$, where $\mathbf{a}_1$ and $\mathbf{a}_2$ are noncollinear, then $\dim V = 2$. ⊠

Example 6.22 (Dimension of a Hyperplane). Again, let $V = \mathbb{F}^n$. Then we saw that $\dim V = n$. The dimension of the hyperplane $a_1x_1 + \cdots + a_nx_n = 0$ in V is $n-1$, provided some a_i in nonzero, since the $n-1$ basic null vectors form a basis of the hyperplane (see Example 6.20). ⊠

Example 6.23 (Dimension of $\mathbb{F}^{m\times n}$). As noted earlier, the vector space $\mathbb{F}^{m\times n}$ of $m \times n$ matrices over $\mathbb{F}$ is indistinguishable from $\mathbb{F}^{mn}$, so we would expect that $\dim \mathbb{F}^{m\times n} = mn$. Now, the matrix analogue of the standard basis of $\mathbb{F}^n$ is the set of $m \times n$ matrices E_{ij} that have a 1 in the ith row and jth column and a zero everywhere else. We leave the proof that they form a basis as an exercise. Therefore, $\dim \mathbb{F}^{m\times n} = mn$, as expected. ⊠

Example 6.24. By Exercise 6.3.10, we see that if $\dim \mathbb{R}[x]_m$ denotes the space of real polynomials of degree at most m. then $\dim \mathbb{R}[x]_m = m + 1$ for all $m \geq 0$, a basis being $1, x, \ldots, x^m$. ⊠

Example 6.25. Let $a_1, \ldots, a_m$ be real constants. Then the solution space of the homogeneous linear differential equation

$$y^{(m)} + a_1y^{(m-1)} + \cdots + a_{m-1}y' + a_my = 0$$

is a vector space over $\mathbb{R}$. It turns out, by a theorem on differential equations, that the dimension of this space is m. For example, when $m = 4$ and $a_i = 0$ for $1 \leq i \leq 4$, then the solution space is the vector space $\mathbb{R}[x]_3$ of the previous example. The solution space W of the equation $y'' + y = 0$ consists of all linear combinations of the functions $\sin x$ and $\cos x$. We leave it as an exercise to show that $\sin x$ and $\cos x$ are linearly independent, so $\dim W = 2$. ⊠

Example 6.26 (Symmetric $n \times n$ matrices). Let $\mathbb{F}_s^{n\times n}$ denote the set of symmetric $n \times n$ matrices over $\mathbb{F}$. Now, $\mathbb{F}_s^{n\times n}$ is certainly a subspace of $\mathbb{F}^{n\times n}$ (exercise). The basis $\{E_{ij} \mid 1 \leq i, j \leq n\}$ of $\mathbb{F}^{n\times n}$ doesn't contain a basis of

$\mathbb{F}_s^{n\times n}$, however, since E_{ij} isn't symmetric if $i \neq j$. To repair this problem, we put $S_{ij} = E_{ij} + E_{ji}$ when $i \neq j$. Then $S_{ij} \in \mathbb{F}_s^{n\times n}$, and I claim that the S_{ij} $(1 \leq i < j \leq n)$ together with the E_{ii} $(1 \leq i \leq n)$ are a basis of $\mathbb{F}_s^{n\times n}$. They certainly span $\mathbb{F}_s^{n\times n}$, since if $A = (a_{ij})$ is symmetric, then

$$A = \sum_{i<j} a_{ij}(E_{ij} + E_{ji}) + \sum_i a_{ii}E_{ii}.$$

We leave it as an exercise to verify that this spanning set is also independent. In particular, counting the number of basis vectors, we see that

$$\dim \mathbb{F}_s^{n\times n} = (n-1) + (n-2) + \cdots + 2 + 1 + n = n(n+1)/2,$$

by the well-known formula for the sum of the first n positive integers. ⊠

Example 6.27 (Skew-symmetric matrices). A square matrix $A \in \mathbb{F}^{n\times n}$ is called *skew-symmetric* if $A^T = -A$. The set $\mathbb{F}_{ss}^{n\times n}$ of skew-symmetric $n \times n$ matrices over $\mathbb{F}$ is another interesting subspace of $\mathbb{F}^{n\times n}$. If the characteristic of the field $\mathbb{F}$ is two, then skew-symmetric and symmetric matrices are the same thing, so for the rest of this example suppose $\mathrm{char}(\mathbb{F}) \neq 2$. For example, if

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^T = -\begin{pmatrix} a & b \\ c & d \end{pmatrix},$$

then $a = -a$, $d = -d$, $b = -c$, and $c = -b$. Thus a 2×2 skew-symmetric matrix has the form

$$\begin{pmatrix} 0 & b \\ -b & 0 \end{pmatrix},$$

so $E_{12} - E_{21}$ is a basis. We leave it as an exercise to show that $\dim \mathbb{F}_{ss}^{n\times n} = n(n-1)/2$ for all n. ⊠

6.4.3 The Dimension Theorem

We will now prove the dimension theorem. This settles the question whether a basis exists and the definition of dimension makes sense.

Theorem 6.11 (The dimension theorem). *Let V denote a finite-dimensional vector space with at least one nonzero element. Then V has a basis. In fact, every spanning set for V contains a basis, and every linearly independent subset of V is contained in a basis. Moreover, any two bases of V have the same number of elements.*

Proof. We'll begin by showing that every finite spanning set contains a basis. Let $\mathbf{w}_1, \ldots, \mathbf{w}_k$ span V, and consider the set of all subsets of $\{\mathbf{w}_1, \ldots, \mathbf{w}_k\}$ that also span V. Let $\{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$ be any such subset where r is minimal. There is no problem showing that such a subset exists, since $\{\mathbf{w}_1, \ldots, \mathbf{w}_k\}$ has only finitely many subsets. We now show that $\mathbf{v}_1, \ldots, \mathbf{v}_r$ are independent. So suppose

$$a_1\mathbf{v}_1 + \cdots + a_r\mathbf{v}_r = \mathbf{0},$$

but $a_i \neq 0$. Then

$$\mathbf{v}_i = \frac{-1}{a_i} \sum_{j \neq i} a_j \mathbf{v}_j,$$

so if $\mathbf{v}_i$ is deleted from $\{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$, we still have a spanning set, contradicting the minimality of r. Thus $\mathbf{v}_1, \ldots, \mathbf{v}_r$ are independent, so every spanning set contains a basis. In particular, since V has a finite spanning set, it has a basis.

We next show that every linearly independent set in V can be extended to a basis. Let $\mathbf{w}_1, \ldots, \mathbf{w}_m$ be independent, and put $W = \text{span}\{\mathbf{w}_1, \ldots, \mathbf{w}_m\}$. I claim that if $\mathbf{v} \notin W$, then $\mathbf{w}_1, \ldots, \mathbf{w}_m, \mathbf{v}$ are independent. To see this, suppose

$$a_1\mathbf{w}_1 + \cdots + a_m\mathbf{w}_m + b\mathbf{v} = \mathbf{0}.$$

If $b \neq 0$, it follows (as in the last argument) that $\mathbf{v} \in W$, contrary to the choice of $\mathbf{v}$. Thus $b = 0$. But then each a_k is equal to zero as well, since the $\mathbf{w}_i$ are independent. Now suppose $W \neq V$. We will use the basis $\mathbf{v}_1, \ldots, \mathbf{v}_r$ of V obtained above to obtain a basis containing $\mathbf{w}_1, \ldots, \mathbf{w}_m$. If each $\mathbf{v}_i$ is in W, then $W = V$, contrary to assumption. So let i be the first index such that $\mathbf{v}_i \notin W$. By the previous paragraph, $\mathbf{w}_1, \ldots, \mathbf{w}_m, \mathbf{v}_i$ are independent. Hence they form a basis for $W_1 = \text{span}\{\mathbf{w}_1, \ldots, \mathbf{w}_m, \mathbf{v}_i\}$. Repeating this construction with W_1 replacing W and so on, we eventually obtain a subspace W_k that contains all $\mathbf{v}_j$. Thus $W_k = V$, so the basis of W_k just constructed is a basis of V containing $\mathbf{w}_1, \ldots, \mathbf{w}_m$. This proves that every linearly independent set is contained in a basis.

It remains to show that any two bases of V have the same number of elements. Suppose $\mathbf{u}_1, \ldots, \mathbf{u}_m$ and $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are two bases of V. If $m \neq n$, we may assume without loss of generality that $m \leq n$. By definition, we can certainly write

$$\mathbf{v}_1 = r_1\mathbf{u}_1 + r_2\mathbf{u}_2 \cdots + r_m\mathbf{u}_m. \tag{6.4}$$

Since $\mathbf{v}_1 \neq \mathbf{0}$, some r_i is nonzero, so we may suppose, by renumbering the indices of the $\mathbf{u}_j$ if necessary, that $r_1 \neq 0$. I claim that $\mathbf{v}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m$ is also a basis of V. To see this, we must show that $\mathbf{v}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m$ are independent and span. Suppose that

$$x_1\mathbf{v}_1 + x_2\mathbf{u}_2 \cdots + x_m\mathbf{u}_m = \mathbf{0}.$$

If $x_1 \neq 0$, then

$$\mathbf{v}_1 = y_2\mathbf{u}_2 + \cdots + y_j\mathbf{u}_m,$$

where $y_i = -x_i/x_1$. Since $r_1 \neq 0$, we get two distinct ways of expanding $\mathbf{v}_1$ in terms of the first basis, which contradicts the uniqueness statement in Proposition 6.8. Hence $x_1 = 0$. It follows immediately that all x_i are equal to zero (why?), so $\mathbf{v}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m$ are independent. The fact that $\mathbf{v}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m$ span V follows from (6.4), since $r_1 \neq 0$. Hence we have produced a new basis of V in which $\mathbf{v}_1$ replaces $\mathbf{u}_1$. Now write

$$\mathbf{v}_2 = x_1\mathbf{v}_1 + x_2\mathbf{u}_2 + \cdots + \mathbf{u}_m.$$

Since $\mathbf{v}_1, \ldots, \mathbf{v}_m$ are independent, there exists $j \geq 2$ such that $x_j \neq 0$. Thus, after reindexing again, we can assume that $x_2 \neq 0$. Repeating the above argument, we see that $\mathbf{u}_2$ can be replaced by $\mathbf{v}_2$, giving another new basis $\mathbf{v}_1, \mathbf{v}_2, \mathbf{u}_3 \ldots, \mathbf{u}_m$ of V. Continuing this process, we will eventually replace all the $\mathbf{u}_i$, which implies that $\mathbf{v}_1, \ldots, \mathbf{v}_m$ must be a basis of V. But if $m < n$, it then follows that $\mathbf{v}_n$ is a linear combination of $\mathbf{v}_1, \ldots, \mathbf{v}_m$, contradicting the linear independence of $\mathbf{v}_1, \ldots, \mathbf{v}_n$. Thus $m = n$, and the proof of the dimension theorem is finished. □

The dimension theorem has several useful consequences.

Corollary 6.12. *If W is a subspace of a finite-dimensional vector space V, then W is finite-dimensional, and* $\dim W \leq \dim V$ *with equality exactly when $W = V$. In particular, every subset of V containing more that* $\dim V$ *elements is dependent.*

Proof. This is an exercise. □

Corollary 6.13. *If* $\dim V = m$*, then every set of m linearly independent vectors in V forms a basis. Similarly, every set of m vectors that span V is also a basis.*

Proof. This is also an exercise. □

Concentrating on linear systems, we may consider the following example.

Example 6.28 (Linear systems). By Example 6.20, we know that $\dim \mathcal{N}(A)$ is the number of free variables in the system $A\mathbf{x} = \mathbf{0}$. Thus the fundamental identity (3.11) for an $m \times n$ homogeneous linear system $A\mathbf{x} = \mathbf{0}$ can now be expressed in terms of dimension as follows:

$$\dim \mathcal{N}(A) + \operatorname{rank}(A) = n. \tag{6.5}$$

However, we can say even more. The rows of A_{red} are certainly linearly independent (why?), so they form a basis of $\operatorname{row}(A)$, the span of the rows of A. Thus, $\operatorname{rank}(A) = \dim \operatorname{row}(A)$, so for every $A \in \mathbb{F}^{m\times n}$, we also have

$$\dim \mathcal{N}(A) + \dim \operatorname{row}(A) = n. \tag{6.6}$$

This seems to be a more elegant statement than that of (3.11). There is yet another improvement from knowing $\operatorname{rank}(A) = \operatorname{rank}(A^T)$. Since $\operatorname{rank}(A) = \dim \operatorname{row}(A)$, we have $\dim \operatorname{row}(A) = \dim \operatorname{row}(A^T) = \dim \operatorname{col}(A)$. Thus,

$$\dim \mathcal{N}(A) + \dim \operatorname{col}(A) = n. \tag{6.7}$$

Since $\operatorname{col}(A) = \{\mathbf{b} \in \mathbb{F}^m \mid A\mathbf{x} = \mathbf{b}\ \exists\ \mathbf{x} \in \mathbb{F}^n\}$, $\dim \mathcal{N}(A)$ and $\dim \operatorname{col}(A)$ refer to the linear system $A\mathbf{x} = \mathbf{b}$. (Recall that understanding this equation is one of our main motivations.) The identity (6.7) is sometimes called the rank–nullity identity. ⊠

6.4.4 Finding a basis of the column space

Suppose $\{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$ is a spanning set for a subspace W of $\mathbb{F}^m$. We know that some subset of this set is a basis. Is there is an efficient procedure for extracting a basis of W? If we view the $\mathbf{a}_i$ as column vectors and form the $m \times n$ matrix $A = \begin{pmatrix}\mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n\end{pmatrix}$, then $W = \operatorname{col}(A)$. Thus we know the dimension of W: $\dim W = \operatorname{rank}(A)$. Hence by Corollary 6.13, it suffices to find $\operatorname{rank}(A)$ independent columns. Somewhat surprisingly, one can proceed by finding A_{red}. This follows from the next proposition.

Proposition 6.14 *The columns of a matrix $A \in \mathbb{F}^{m\times n}$ that correspond to a corner entry in A_{red} are a basis of* $\operatorname{col}(A)$.

Proof We can assume that A has rank m. As shown above, it suffices to show that the columns of A that correspond to a corner entry in A_{red} are independent. Since $A_{red} = BA$ with B invertible, it follows that $A\mathbf{x} = \mathbf{0}$ if and only if $A_{red}\mathbf{x} = \mathbf{0}$. In particular, every expression of linear dependence among a subset of the columns of A is also an expression of linear dependence among those columns of A_{red}. For example, if the fifth column of A is the sum of the first four columns of A, this also holds for the columns of A_{red}. of A_{red} containing a corner entry are standard basis vectors of $\mathbb{F}^m$, hence

are certainly independent. Therefore, these columns of A are also linearly independent. Since $\dim \operatorname{col}(A) = \operatorname{rank}(A)$, Corollary 6.13 says that these columns of A are a basis of W. □

Example 6.29 For example, suppose

$$A = \begin{pmatrix} 1 & 2 & 2 \\ 4 & 5 & 8 \\ 7 & 8 & 14 \end{pmatrix}.$$

Then

$$A_{red} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Proposition 6.14 implies that the first two columns are a basis of $\operatorname{col}(A)$. Notice that the first and third columns are dependent in both A and A_{red}, as the proof shows must happen. Proposition 6.14 says that the first two columns are a basis of the column space, but it makes no assertion about the second and third columns, which in fact are also a basis. ⊠

6.4.5 A Galois field application

Let p be a prime and consider a vector space V over the prime field $\mathbb{F}_p$. Then the dimension of V determines the number of elements of V as follows.

Proposition 6.15 *If V is finite-dimensional, then the number of elements of V is exactly $p^{\dim V}$.*

Proof Let $k = \dim V$ and choose a basis $\mathbf{w}_1, \ldots, \mathbf{w}_k$ of V. By Proposition 6.8, every $\mathbf{v} \in W$ has a unique expression

$$\mathbf{v} = a_1\mathbf{w}_1 + a_2\mathbf{w}_2 + \cdots + a_k\mathbf{w}_k,$$

where $a_1, a_2, \ldots, a_k \in \mathbb{F}_p$. Now it is simply a matter of counting such expressions. In fact, since $\mathbb{F}_p$ has p elements, there are p choices for each a_i, and since uniqueness says that different choices of the a_i give different elements of V, it follows that there are exactly $p \cdot p \cdots p = p^k$ distinct linear combinations. Thus V contains exactly p^k elements. □

For example, a line in $(\mathbb{F}_p)^n$ has p elements, a plane has p^2, and so forth. We can apply the last result to deduce a beautiful fundamental fact about Galois fields. Let $\mathbb{F}$ be a Galois field. Then the characteristic of $\mathbb{F}$ is a prime, say p. By Exercise 2.6.8, the multiples of 1 together with 0 form a subfield

with p elements. This subfield is indistinguishable from $\mathbb{F}_p$, but we will denote it by $\mathbb{F}'$. It follows from the field axioms that $\mathbb{F}$ is a vector space over $\mathbb{F}'$ (see Example 6.11). Moreover, since $\mathbb{F}$ itself is finite, it follows by definition that $\mathbb{F}$ is finite-dimensional (over $\mathbb{F}'$), since every $a \in \mathbb{F}$ has the expression $a = 1a$, so $\mathbb{F}$ spans itself over $\mathbb{F}'$, since $1 \in \mathbb{F}'$. Applying Proposition 6.15, we get the following result.

Proposition 6.16 *Let $\mathbb{F}$ be a finite field of characteristic p. Then $|\mathbb{F}| = p^n$, where n is the vector space dimension of $\mathbb{F}$ over the subfield $\mathbb{F}'$ of $\mathbb{F}$ consisting of all multiples of 1: that is, $n = \dim_{\mathbb{F}'} \mathbb{F}$.*

Recall that in Section 2.6.2, we considered a field $\mathbb{F} = \{0, 1, \alpha, \beta\}$, where $\alpha + \beta = 1$. Thus, for example, $\{1, \alpha\}$ is a basis, and $\mathbb{F}$ has $4 = 2^2$ elements, in agreement with the above result. It can be shown that for every prime p and every $n > 0$, there is a unique Galois field of order p^n. Combining the previous two results, we get the following corollary.

Corollary 6.17 *Let $\mathbb{F}$ be a Galois field of characteristic p, and let $q = p^n$ denote $|\mathbb{F}|$. Then if V is a finite-dimensional vector space over $\mathbb{F}$, we have*

$$|V| = q^{dim_{\mathbb{F}} V} = p^{n \dim_{\mathbb{F}} V}.$$

Here is another corollary.

Corollary 6.18 *Let $\mathbb{F}$ be a Galois field with $q = p^m$ elements. Then the general linear group $GL(n, \mathbb{F})$ has order*

$$|GL(n, \mathbb{F})| = (q^n - 1)(q^n - q)(q^n - q^2) \cdots (q^n - q^{n-1}). \tag{6.8}$$

Proof The elements of $GL(n, \mathbb{F})$ may be described as the elements A of $\mathbb{F}^{n \times n}$ whose columns form a basis of $\mathbb{F}^n$. Consequently, the first i columns of A have to form a basis of an i-dimensional subspace of $\mathbb{F}^n$. Thus, there are $q^n - 1$ choices for A's first column, $q^n - q$ choices for its second column, since the only restriction on the second column is that it not lie on the line spanned by the first column, and in general, there are $q^n - q^{i-1}$ choices for the ith column, since the span of the first $i - 1$ columns has q^{i-1} elements. □

Exercises

Exercise 6.4.1 Find a basis for the subspace of $\mathbb{R}^4$ spanned by

$$(1, 0, -2, 1)^T,\ (2, -1, 2, 1)^T,\ (1, 1, 1, 1)^T,\ (0, 1, 0, 1)^T,\ (0, 1, 1, 0)^T$$

that contains the first and fifth vectors.

Exercise 6.4.2 Consider the matrix $A = \begin{pmatrix} 1 & 2 & 0 & 1 & 2 \\ 2 & 0 & 1 & -1 & 2 \\ 1 & 1 & -1 & 1 & 0 \end{pmatrix}$ as an element of $\mathbb{R}^{3\times 5}$.

(*i*) Show that the basic null vectors of $A\mathbf{x} = \mathbf{0}$ are a basis of $\mathcal{N}(A)$.

(*ii*) Find a basis of $\text{col}(A)$.

(*iii*) Repeat (*i*) and (*ii*) when A is considered as a matrix over $\mathbb{F}_3$.

Exercise 6.4.3 Prove that $\cos x$ and $\sin x$ are linearly independent on the open interval $(0, 2\pi)$.

Exercise 6.4.4 Let W be a subspace of V, and let $\mathbf{w}_1, \ldots, \mathbf{w}_k \in W$. Show that if $\mathbf{w}_1, \ldots, \mathbf{w}_k$ are independent, then $\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_k, \mathbf{v}$ are also independent for every $\mathbf{v} \in V$ such that $\mathbf{v} \notin W$.

Exercise 6.4.5 Prove Corollary 6.12. That is, suppose V is a finite-dimensional vector space over a field $\mathbb{F}$, and let W be a subspace of V. Show the following:

(*i*) W is finite-dimensional.

(*ii*) In fact, $\dim W \leq \dim V$.

(*iii*) If $\dim W = \dim V$, then $W = V$.

Exercise 6.4.6 Prove Corollary 6.13. That is, show that if $\dim V = m$, then every set of m linearly independent vectors in V forms a basis, and similarly, every set of m vectors that span V is also a basis.

Exercise 6.4.7 True or false: For every $A \in \mathbb{F}^{n\times n}$, $rank(A) \geq rank(A^2)$. Prove your answer.

Exercise 6.4.8 Consider the subspace W of $V(4, 2) = (\mathbb{F}_2)^4$ spanned by 1011, 0110, and 1001.

(*i*) Find a basis of W and compute $|W|$.

(*ii*) Extend your basis to a basis of $(\mathbb{F}_2)^4$.

Exercise 6.4.9 Construct at least two proofs of the statement that for every matrix A, $\dim \text{col}(A) = \dim \text{row}(A)$.

Exercise 6.4.10 Let A and B be $n \times n$ matrices.

(*i*) Show that $\mathcal{N}(A) \subset \mathcal{N}(BA)$. When is $\mathcal{N}(BA) = \mathcal{N}(A)$?

(*ii*) Show that $\text{col}(A) \supset \text{col}(AB)$. When is $\text{col}(AB) = \text{col}(A)$?

(*iii*) Show that $AB = O$ if and only if $\text{col}(B) \subset \mathcal{N}(A)$.

Exercise 6.4.11 Consider the set $\mathbb{F}_s^{n\times n}$ of symmetric $n \times n$ matrices over $\mathbb{F}$.

(*i*) Show that $\mathbb{F}_s^{n\times n}$ is a subspace of $\mathbb{F}^{n\times n}$.

(*ii*) Show that the set of matrices S_{ij} with $i < j$ defined in Example 6.26 together with the E_{ii} make up a basis of $\mathbb{F}_s^{n\times n}$.

Exercise 6.4.12 Let $\mathbb{F}_{ss}^{n\times n}$ be the $n \times n$ skew-symmetric matrices over $\mathbb{F}$.

(*i*) Show that $\mathbb{F}_{ss}^{n\times n}$ is a subspace of $\mathbb{F}^{n\times n}$.

(*ii*) Find a basis of $\mathbb{F}_{ss}^{n\times n}$ and compute its dimension.

(*iii*) Find a basis of $\mathbb{F}^{n\times n}$ that uses only symmetric and skew-symmetric matrices.

Exercise 6.4.13 Let $\mathbb{F}$ be a field, and suppose V and W are subspaces of $\mathbb{F}^n$.

(*i*) Show that $V \cap W$ is a subspace of $\mathbb{F}^n$.

(*ii*) Let $V + W = \{\mathbf{u} \in \mathbb{F}^n \mid \mathbf{u} = \mathbf{v} + \mathbf{w}, \text{ where } v \in V, w \in W\}$. Show that $V + W$ is a subspace of $\mathbb{F}^n$.

(*iii*) Show that $\dim(V + W) \leq \dim V + \dim W$.

Exercise 6.4.14 Let W and Y be subspaces of a vector space V of dimension n. What are the minimum and maximum dimensions that $W \cap X$ can have? Discuss the case that W is a hyperplane (i.e., $\dim W = n - 1$) and X is a plane (i.e., $\dim X = 2$).

Exercise 6.4.15 Suppose X is a finite set, say $|X| = n$, and let $\mathbb{F} = \mathbb{F}_p$. Let $V = \mathbb{F}^X$. That is, V is the set of all maps with domain X and target $\mathbb{F}$. Show that V is an $\mathbb{F}$-vector space, find a basis, and compute $\dim V$.

Exercise 6.4.16 Show that the set of $n \times n$ upper triangular matrices over $\mathbb{F}$ is a subspace of $\mathbb{F}^{n\times n}$. Find a basis and its dimension.

Exercise 6.4.17 Let $\mathbb{F}$ be a Galois field of characteristic p, and let $\mathbb{F}'$ be the subfield of $\mathbb{F}$ consisting of all multiples of 1. If V is a finite-dimensional vector space over $\mathbb{F}$, show that $\dim_{\mathbb{F}'} V = \dim_{\mathbb{F}'} \mathbb{F} \dim_{\mathbb{F}} V$. Conclude that

$$|V| = p^{\dim_{\mathbb{F}'} \mathbb{F} \dim_{\mathbb{F}} V}.$$

Exercise 6.4.18 Let $|\mathbb{F}| = q$, where $q = p^n$, p a prime. Show that every element of $\mathbb{F}$ is a root of the polynomial $x^q - x \in \mathbb{F}_p[x]$.

Exercise 6.4.19 Let $V = \mathbb{R}$ as a vector space over $\mathbb{Q}$. Is $\dim V$ finite or infinite? Discuss.

Exercise 6.4.20 Let V be a vector space over $\mathbb{F}_p$ of dimension n. A linearly independent subset of V with m elements is called an *m-frame* in V. Show that the number of m-frames in V is exactly

$$(p^n - 1)(p^n - p) \cdots (p^n - p^{m-2})(p^n - p^{m-1}).$$

(Use Proposition 6.15 and part of the proof of the dimension theorem.)

Exercise 6.4.21 Use Exercise 6.4.20 to show that the number of subspaces of dimension m in an n-dimensional vector space V over $\mathbb{F}_p$ is

$$\frac{(p^n - 1)(p^n - p) \cdots (p^n - p^{m-2})(p^n - p^{m-1})}{(p^m - 1)(p^m - p) \cdots (p^n - p^{m-2})(p^m - p^{m-1})}.$$

(The set of m-dimensional subspaces of a finite-dimensional vector space is an important object, called a Grassmann variety.)

Exercise 6.4.22 Let V be a vector space. A *complete flag* in V is a sequence of subspaces

$$V_1 \subset V_2 \subset \cdots \subset V_{n-1} \subset V$$

such that $\dim V_i = i$. The set $\mathrm{Flag}(V)$ of all complete flags in V is called the *flag variety of V*.

(*i*) Let $B \subset GL(n, \mathbb{F})$ denote the set of all invertible upper triangular elements of $\mathbb{F}^{n \times n}$. Show that there exists a bijection

$$\Phi : GL(n, \mathbb{F})/B \to \mathrm{Flag}(\mathbb{F}^n).$$

(*ii*) Suppose $\mathbb{F}$ is Galois. Find a formula for the number of flags in $\mathbb{F}^n$. That is, find $|\mathrm{Flag}(\mathbb{F}^n)|$.

Exercise 6.4.23 (This is more challenging.) Consider a square array of lights numbered 1 to 9 inclusively: that is,

$$\begin{array}{ccc} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{array}$$

Turning a light on or off also changes all the on–off states of the lights directly above and below and to the left and right. For instance, if the center light (light number 5) is turned on, then lights 2, 4, 5, 6, and 8 also change their state.

(*i*) If all the lights are off, can they all be turned on?

(*ii*) How can you turn on just light number 1?

(*iii*) Is it possible to turn all the lights off from every starting configuration?

6.5 The Grassmann Intersection Formula

The intersection of a pair of nonparallel planes P_1 and P_2 in $\mathbb{R}^3$ through the origin is a line through the origin. This may seem obvious, but what would the answer be if instead of being planes in $\mathbb{R}^3$, P_1 and P_2 were three-dimensional subspaces of $\mathbb{R}^4$? The purpose of this section is to answer this question; that is, if W and Y are arbitrary subspaces of a finite-dimensional vector space V, what is $\dim(W \cap Y)$? The answer is given by the Grassmann intersection formula, which gives $\dim(W \cap Y)$ in terms of the dimensions of W, Y, and a subspace $W + Y$ known as the sum of W and Y. After that, we will introduce direct sums of subspaces and will derive conditions for sums of subspaces to be direct. These results on direct sums will be used when we study eigentheory in Chap. 8.

6.5.1 Intersections and sums of subspaces

Let V be a vector space over a field $\mathbb{F}$ with subspaces W and Y. The simplest way of building a new subspace is by taking the intersection $W \cap Y$.

Proposition 6.19 *The intersection $W \cap Y$ of the subspaces W and Y of V is also a subspace of V. More generally, the intersection of an arbitrary collection of subspaces of V is also a subspace.*

Proof This is an exercise. □

Proposition 6.19 is a generalization of the fact that the solution space of a homogeneous linear system is a subspace of $\mathbb{F}^n$. The solution space of a single homogeneous linear equation is a hyperplane in $\mathbb{F}^n$, and so the solution space of a homogeneous linear system is the intersection of a finite number of hyperplanes in $\mathbb{F}^n$.

Another simple way of forming a new subspace is to take the subspace spanned by W and Y. This is defined as follows.

Definition 6.8 The subspace spanned by W and Y or alternatively, the *sum* of W and Y, is defined to be the set of sums

$$W + Y = \{\mathbf{w} + \mathbf{y} \mid \mathbf{w} \in W, \mathbf{y} \in Y\}.$$

More generally, we can form the sum $V_1 + \cdots + V_k$ of an arbitrary (finite) number of subspaces $V_1, V_2, \ldots, V_k$ of V. The sum $V_1 + \cdots + V_k$ is also written as $\sum_{i=1}^{k} V_i$, or more simply $\sum V_i$.

Proposition 6.20 *The sum $\sum_{i=1}^{k} V_i$ of the subspaces $V_1, V_2, \ldots, V_k$ of V is also a subspace of V. It is, in fact, the smallest subspace of V containing every V_i.*

Proof We leave the proof as another exercise. □

6.5.2 Proof of the Grassmann intersection formula

We now return to the question of what one can say about the dimension of the intersection of two subspaces. For example, what is the dimension of the intersection of two three-dimensional subspaces of $\mathbb{R}^4$? The answer is given by the Grassmann intersection formula, which relates the dimensions of W, Y, $W + Y$, and $W \cap Y$. Before looking at the formula, the reader can try to guess the answer.

Theorem 6.21 *If W and Y are finite-dimensional subspaces of a vector space V, then $W + Y$ is finite-dimensional, and*

$$\dim(W + Y) = \dim W + \dim Y - \dim(W \cap Y). \tag{6.9}$$

Proof Since $W \cap Y$ is a subspace of W, it is finite-dimensional. Hence, we know that $W \cap Y$ has a basis, say $\mathbf{x}_1, \ldots, \mathbf{x}_k$. The dimension theorem allows us to extend this basis to a basis of W, say

$$\mathbf{x}_1, \ldots, \mathbf{x}_k, \mathbf{w}_{k+1}, \ldots, \mathbf{w}_{k+r}.$$

Likewise, since Y is also finite-dimensional, we can extend the basis of $W \cap Y$ to a basis of Y, say

$$\mathbf{x}_1, \ldots, \mathbf{x}_k, \mathbf{y}_{k+1}, \ldots, \mathbf{y}_{k+s}.$$

I claim that

$$\mathcal{B} = \{\mathbf{x}_1, \ldots, \mathbf{x}_k, \mathbf{w}_{k+1}, \ldots, \mathbf{w}_{k+r}, \mathbf{y}_{k+1}, \ldots, \mathbf{y}_{k+s}\}$$

is a basis of $W + Y$. It is not hard to see that $\mathcal{B}$ is a spanning set, so $W + Y$ is finite-dimensional even though V is an arbitrary vector space. To see that $\mathcal{B}$ is independent, suppose

$$\sum_{i=1}^{k} \alpha_i \mathbf{x}_i + \sum_{j=k+1}^{k+r} \beta_j \mathbf{w}_j + \sum_{m=k+1}^{k+s} \gamma_m \mathbf{y}_m = \mathbf{0}. \tag{6.10}$$

Thus,

$$\sum \gamma_m \mathbf{y}_m = -\big(\sum \alpha_i \mathbf{x}_i + \sum \beta_j \mathbf{w}_j\big).$$

Since the left-hand term is in Y and the one on the right is in W, we have

$$\sum \gamma_m \mathbf{y}_m \in Y \cap W.$$

Thus

$$\sum \gamma_m \mathbf{y}_m = \sum \delta_i \mathbf{x}_i$$

for some $\delta_i \in \mathbb{F}$. Hence

$$\sum \delta_i \mathbf{x}_i + \sum (-\gamma_m) \mathbf{y}_m = \mathbf{0}.$$

Therefore, all the δ_i and γ_m are zero. In particular, (6.10) becomes the expression

$$\sum \alpha_i \mathbf{x}_i + \sum \beta_j \mathbf{w}_j = \mathbf{0}.$$

But this implies that all the α_i and β_j are 0 also. Consequently, $\mathcal{B}$ is independent. Since $\mathcal{B}$ spans $W + Y$, it forms a basis of $W + Y$, so $\dim(W + Y) = k + r + s$. It remains to count dimensions. We have

$$\dim(W + Y) = k + r + s = (k + r) + (k + s) - k,$$

and hence $\dim(W + Y) = \dim W + \dim Y - \dim(W \cap Y)$. □

Notice that the Grassmann intersection formula doesn't mention $\dim V$, which is why we don't need to require that V be finite-dimensional. However, if it is, then $\dim(W + Y) \leq \dim V$, so we get a nice corollary.

Corollary 6.22 *If W and Y are subspaces of a finite-dimensional vector space V, then*

$$\dim(W \cap Y) \geq \dim W + \dim Y - \dim V. \tag{6.11}$$

In particular, if $\dim W + \dim Y > \dim V$, then $\dim(Y \cap W) > 0$.

Now let us see what can be said about two three-dimensional subspaces W and Y of a five-dimensional space V. Since the sum of the dimensions of W and Y is 6, they meet in at least a line, although the intersection can also have dimension two or three. If $\dim V$ decreases, then intuitively, $\dim(Y \cap W)$ should increase, since there is less room to maneuver. This is exactly what the inequality tells us. If $\dim V = 4$, then $\dim(W \cap Y) \geq 3 + 3 - 4 = 2$, so W and Y contain a common plane. However, once $\dim V$ is at least 6, the inequality no longer tells us anything.

Example 6.30 (Intersection of hyperplanes). Let H_1 and H_2 be distinct hyperplanes in $\mathbb{F}^n$. Then

$$\dim(H_1 \cap H_2) \geq (n - 1) + (n - 1) - n = n - 2.$$

But since the hyperplanes are distinct, $\dim(H_1 \cap H_2) < n-1$, so $\dim(H_1 \cap H_2) = n-2$ exactly. ⊠

Here is a nice example.

Example 6.31. Recall that $\mathbb{F}_s^{n\times n}$ and $\mathbb{F}_{ss}^{n\times n}$ denote, respectively, the spaces of of $n \times n$ symmetric and $n \times n$ skew-symmetric matrices (see Example 6.26). Let's assume now that the characteristic of $\mathbb{F}$ is not equal to 2. Then an arbitrary $A \in \mathbb{F}^{n\times n}$ can be expressed as the sum of a symmetric matrix and a skew-symmetric matrix. Namely,

$$A = \frac{1}{2}(A + A^T) + \frac{1}{2}(A - A^T). \tag{6.12}$$

Thus $\mathbb{F}^{n\times n} = \mathbb{F}_s^{n\times n} + \mathbb{F}_{ss}^{n\times n}$. Moreover, since $\text{char}(\mathbb{F}) \neq 2$, the only matrix that is both symmetric and skew-symmetric is the zero matrix. That is, $\mathbb{F}_s^{n\times n} \cap \mathbb{F}_{ss}^{n\times n} = \{\mathbf{0}\}$. Hence by the Grassmann intersection formula,

$$\dim \mathbb{F}^{n\times n} = \dim(\mathbb{F}_s^{n\times n}) + \dim(\mathbb{F}_{ss}^{n\times n}).$$

Note that we already knew this result from Section 6.4.2, where we showed that $\dim(\mathbb{F}_s^{n\times n}) = n(n+1)/2$ and $\dim(\mathbb{F}_{ss}^{n\times n}) = n(n-1)/2$. ⊠

As we will see in the next section, this example shows that $\mathbb{F}^{n\times n}$ is the direct sum of $\mathbb{F}_s^{n\times n}$ and $\mathbb{F}_{ss}^{n\times n}$. This is a stronger assertion than simply saying that $\mathbb{F}^{n\times n} = \mathbb{F}_s^{n\times n} + \mathbb{F}_{ss}^{n\times n}$, which is a consequence of (6.12). In particular, it implies that every square matrix can be uniquely expressed via (6.12) as the sum of a symmetric matrix and a skew-symmetric matrix, except when the characteristic is two.

6.5.3 Direct sums of subspaces

By the Grassmann intersection formula, two subspaces W and Y of V such that $\dim(W \cap Y) = 0$ have the property that $\dim(W + Y) = \dim W + \dim Y$, and conversely. An explicit example of this was considered above. This observation is related to the following definition.

Definition 6.9. We say that V is the *direct sum* of two subspaces W and Y if $V = W + Y$ and for every $\mathbf{v} \in V$, the expression $\mathbf{v} = \mathbf{w} + \mathbf{y}$ with $\mathbf{w} \in W$ and $\mathbf{y} \in Y$ is unique. If V is the direct sum of W and Y, we write $V = W \oplus Y$. More generally, we say that V is the *direct sum of the subspaces* $V_1, \ldots, V_k$ if $V = \sum V_i$ and for every $\mathbf{v} \in V$, the expression $\mathbf{v} = \sum \mathbf{v}_i$, where each $\mathbf{v}_i \in V_i$, is unique. (Equivalently, if $\mathbf{0} = \sum \mathbf{v}_i$, where each $\mathbf{v}_i \in V_i$, then each $\mathbf{v}_i = \mathbf{0}$.) In this case, we write $V = \bigoplus_{i=1}^{k} V_i$.

Proposition 6.23. *Suppose V is a finite-dimensional vector space with subspaces W and Y. Then the following conditions are all equivalent to the assertion that $V = W \oplus Y$:*

(i) $V = W + Y$ *and* $W \cap Y = \{\mathbf{0}\}$.

(ii) $V = W + Y$ *and* $\dim V = \dim W + \dim Y$.

(iii) $\dim V = \dim W + \dim Y$ *and* $W \cap Y = \{\mathbf{0}\}$.

Proof. That conditions (i), (ii), and (iii) are equivalent follows from the Grassmann intersection formula. Thus it suffices to show that (i) is equivalent to $V = W \oplus Y$. Assume (i), and suppose $\mathbf{v} = \mathbf{w} + \mathbf{y} = \mathbf{w}' + \mathbf{y}'$. Then $\mathbf{w} - \mathbf{w}' = \mathbf{y}' - \mathbf{y}$ is an element of $W \cap Y = \{\mathbf{0}\}$. Thus $\mathbf{w} = \mathbf{w}'$ and $\mathbf{y}' = \mathbf{y}$. Hence $V = W \oplus Y$. On the other hand, if $V = W \oplus Y$ and $W \cap Y \neq \{\mathbf{0}\}$, then every nonzero $\mathbf{v} \in W \cap Y$ has two expressions $\mathbf{v} = \mathbf{v} + \mathbf{0} = \mathbf{0} + \mathbf{v}$. This violates the definition of a direct sum, so $W \cap Y = \{\mathbf{0}\}$. □

Corollary 6.24. *Suppose* $\dim V = \dim(W + Y)$ *and* $\dim(W \cap Y) = 0$. *Then* $V = W \oplus Y$.

Proof. If $\dim V = \dim(W + Y)$, then $W + Y = V$. Therefore, the result follows from Proposition 6.23. □

Referring back to Example 6.31, we get the following assertion.

Proposition 6.25. *Assume char*$(\mathbb{F}) \neq 2$. *Then* $\mathbb{F}^{n\times n} = \mathbb{F}_s^{n\times n} \oplus \mathbb{F}_{ss}^{n\times n}$. *Thus, every square matrix over* $\mathbb{F}$ *can be uniquely expressed as in (6.12) as the sum of a symmetric matrix and a skew-symmetric matrix, both over* $\mathbb{F}$.

We will need the following extended version of Proposition 6.23.

Proposition 6.26. *Suppose V is finite-dimensional and $V_1, \ldots, V_k$ are subspaces of V such that $V = \sum_{i=1}^k V_i$. Then $V = \bigoplus_{i=1}^k V_i$ if and only if* $\dim V = \sum_{i=1}^k \dim V_i$.

Proof. Suppose $V = \sum_{i=1}^k V_i$, and $\dim V = \sum_{i=1}^k \dim V_i$. Choose a basis of each V_i and consider the union $\mathcal{B}$ of these bases. Then $\mathcal{B}$ clearly spans V, since the part in V_i spans V_i. Hence we get a spanning set in V with $\dim V$ elements. It follows that $\mathcal{B}$ is a basis of V. Now suppose the sum isn't direct. Then there exists an element $\mathbf{v}$ with two different decompositions

$$\mathbf{v} = \sum_{i=1}^k \mathbf{x}_i = \sum_{i=1}^k \mathbf{y}_i,$$

where the $\mathbf{x}_i$ and $\mathbf{y}_i$ are in V_i for each i. Thus $\mathbf{x}_j \neq \mathbf{y}_j$ for some index j. Hence $\mathbf{v}$ has to have two different expansions in terms of $\mathcal{B}$. This contradicts

the uniqueness of the expansion in a basis; hence the sum must be direct. Conversely, suppose $V = \bigoplus_{i=1}^{k} V_i$. Forming $\mathcal{B}$ as in the previous case, it follows that $\mathcal{B}$ is independent, since the sum is direct. Thus $\mathcal{B}$ is a basis, and therefore, $\dim V = |\mathcal{B}| = \sum_{i=1}^{k} \dim V_i$.

□

6.5.4 External direct sums

Let V and W be arbitrary vector spaces over the same field $\mathbb{F}$. Then we can form a new vector space $V \times W$ containing both V and W as subspaces. Recall from Chap. 1 that $V \times W$ denotes the Cartesian product of V and W, namely the set of all ordered pairs $(\mathbf{v}, \mathbf{w})$, where $\mathbf{v} \in V$ and $\mathbf{w} \in W$.

Definition 6.10. The *external direct sum* of V and W is the Cartesian product $V \times W$ with addition defined componentwise by

$$(\mathbf{v}_1, \mathbf{w}_1) + (\mathbf{v}_2, \mathbf{w}_2) = (\mathbf{v}_1 + \mathbf{v}_2, \mathbf{w}_1 + \mathbf{w}_2),$$

and scalar multiplication defined similarly by

$$r(\mathbf{v}, \mathbf{w}) = (r\mathbf{v}, r\mathbf{w}).$$

The alert reader will have noted that $\mathbb{F}^k \times \mathbb{F}^m = \mathbb{F}^{k+m}$. Thus the external direct sum is a generalization of the construction of $\mathbb{F}^n$. This operation can also be extended (inductively) to any finite number of vector spaces over $\mathbb{F}$. In fact, $\mathbb{F}^n$ is just the n-fold external direct sum of $\mathbb{F}$.

We leave it to the reader to show that V and W can both be considered subspaces of $V \times W$.

Proposition 6.27. *If V and W are finite-dimensional vector spaces over $\mathbb{F}$, then so is their external direct sum, and* $\dim(V \times W) = \dim V + \dim W$.

Proof. We leave this as an exercise. □

Exercises

Exercise 6.5.1. Suppose W and Y are two subspaces of a finite-dimensional vector space V such that $W \cap Y = \{\mathbf{0}\}$. Show that $\dim W + \dim Y \leq \dim V$.

Exercise 6.5.2. Prove Proposition 6.27.

Exercise 6.5.3. If two 22-dimensional subspaces of $\mathbb{R}^n$ always meet in at least a line, what can you say about n?

Exercise 6.5.4. Do two subspaces of $\mathbb{R}^{26}$ of dimensions 5 and 17 have to meet in more than $\mathbf{0}$? What about a subspace of dimension 22 and a subspace of dimension 13?

Exercise 6.5.5. Suppose W and Y are two subspaces of $(\mathbb{F}_p)^n$. Find expressions for $|W \cap Y|$ and $|W + Y|$.

6.6 Inner Product Spaces

In this section, we will study the notion of an inner product on a real vector space. We will also study the notion of a Hermitian inner product on a complex vector space, though not in as much detail. An inner product on a real vector space V allows one to measure distances and angles between vectors. Similarly, a Hermitian inner product permits one to do the same for vectors in a complex vector space. The main result of this section is that every finite-dimensional inner product space has a special type of basis known as an orthonormal basis. This is a basis having the properties of the standard basis of $\mathbb{R}^n$.

6.6.1 The definition of an inner product

We will first treat the real case.

Definition 6.11. Let V be a real vector space. An *inner product* on V is a rule that associates to every $\mathbf{a}, \mathbf{b} \in V$ a unique scalar $(\mathbf{a}, \mathbf{b}) \in \mathbb{R}$ having the following properties for all $\mathbf{a}, \mathbf{b}, \mathbf{c} \in V$ and $r \in \mathbb{R}$:

(i) $(\mathbf{a}, \mathbf{b}) = (\mathbf{b}, \mathbf{a})$,

(ii) $(\mathbf{a} + \mathbf{b}, \mathbf{c}) = (\mathbf{a}, \mathbf{c}) + (\mathbf{b}, \mathbf{c})$,

(iii) $(r\mathbf{a}, \mathbf{b}) = (\mathbf{a}, r\mathbf{b}) = r(\mathbf{a}, \mathbf{b})$, and

(iv) if $\mathbf{a} \neq \mathbf{0}$, then $(\mathbf{a}, \mathbf{a}) > 0$.

The property in condition (iv) is called positive definiteness. A real vector space V with an inner product is called an *inner product space*. Clearly, $(\mathbf{0}, \mathbf{0}) = 0$, so $(\mathbf{a}, \mathbf{a}) \geq 0$ for all $\mathbf{a} \in V$. Thus the following definition makes sense.

Definition 6.12. Let V be an inner product space as in the previous definition. Then the *length* of $\mathbf{a} \in V$ is defined by

$$|\mathbf{a}| = \sqrt{(\mathbf{a}, \mathbf{a})}, \tag{6.13}$$

and the *distance* between $\mathbf{a}$ and $\mathbf{b}$ in V is defined by

$$d(\mathbf{a}, \mathbf{b}) = |\mathbf{a} - \mathbf{b}|. \tag{6.14}$$

The basic example of an inner product space is $\mathbb{R}^n$ with the Euclidean inner product defined in the next example.

Example 6.32 (Euclidean n-space). The dot product on $\mathbb{R}^n$ defined in Section 3.1.6 by

$$\mathbf{a} \cdot \mathbf{b} = \mathbf{a}^T \mathbf{b} = \sum_{i=1}^{n} a_i b_i \tag{6.15}$$

defines an inner product by setting $(\mathbf{a}, \mathbf{b}) = \mathbf{a} \cdot \mathbf{b}$. The dot product is referred to as the Euclidean inner product on $\mathbb{R}^n$. ⊠

The next example gives an important inner product on $\mathbb{R}^{n\times n}$. First, we have to define the trace of an $n \times n$ matrix. If $A \in \mathbb{F}^{n\times n}$, define the *trace* of A to be $\mathrm{Tr(A)} = \sum_{i=1}^{n} a_{ii}$.

Example 6.33. Let $V = \mathbb{R}^{n\times n}$. For $A, B \in V$, put

$$(A, B) = \mathrm{Tr(AB^T)}.$$

This defines an inner product on $\mathbb{R}^{n\times n}$ known as the *Killing form*. The verification of the axioms for an inner product is left as an exercise. Notice, for example, that if A and B are diagonal matrices, say $A = \mathrm{diag}(a_1, \ldots, a_n)$ and $B = \mathrm{diag}(b_1, \ldots, b_n)$, then

$$(A, B) = \mathrm{Tr}(AB^T) = \mathrm{Tr}(\mathrm{diag}(a_1 b_1, \ldots, a_n b_n)) = \sum_{i=1}^{n} a_i b_i.$$

Thus the Killing form coincides with the Euclidean inner product on diagonal matrices. ⊠

6.6.2 Orthogonality

As we just saw, an inner product on a real vector space V has natural length and distance functions. We will now see that it allows one to imitate other Euclidean properties of $\mathbb{R}^n$, namely the notion of angles. The starting point is orthogonality.

Definition 6.13. We say that a pair of vectors $\mathbf{a}$, and $\mathbf{b}$ in an inner product space V are *orthogonal* if

$$(\mathbf{a}, \mathbf{b}) = 0. \tag{6.16}$$

Here are a couple of simple properties.

Proposition 6.28. *In an inner product space V, the zero vector is orthogonal to every vector. In fact, $\mathbf{0}$ is the only vector orthogonal to itself. Two vectors $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$ are orthogonal if and only if $\sum_{i=1}^{n} a_i b_i = 0$.*

Proof. Left to the reader. □

Orthogonality is a generalization of the notion of perpendicularity in $\mathbb{R}^2$. Two vectors $\mathbf{a} = (a_1, a_2)^T$ and $\mathbf{b} = (b_1, b_2)^T$ in $\mathbb{R}^2$ are perpendicular exactly when the triangle with sides $\mathbf{a}$ and $\mathbf{b}$ is a right triangle. This is the case exactly when $|\mathbf{a}+\mathbf{b}|^2 = |\mathbf{a}|^2 + |\mathbf{b}|^2$, by the Pythagorean theorem. The reader should check that this identity holds if and only if $a_1b_1 + a_2b_2 = 0$, that is, if $\mathbf{a} \cdot \mathbf{b} = 0$. On an inner product space V, orthogonality can be characterized in a similar manner.

Proposition 6.29. *Two vectors* $\mathbf{a}$ *and* $\mathbf{b}$ *in an inner product space* V *are orthogonal if and only if* $|\mathbf{a}+\mathbf{b}|^2 = |\mathbf{a}|^2 + |\mathbf{b}|^2$.

Proof. Since $|\mathbf{a}+\mathbf{b}|^2 = (\mathbf{a}+\mathbf{b}, \mathbf{a}+\mathbf{b})$, one gets that

$$|\mathbf{a}+\mathbf{b}|^2 = (\mathbf{a},\mathbf{a}) + 2(\mathbf{a},\mathbf{b}) + (\mathbf{b},\mathbf{b}) = |\mathbf{a}|^2 + 2(\mathbf{a},\mathbf{b}) + |\mathbf{b}|^2.$$

Hence $(\mathbf{a},\mathbf{b}) = 0$ if and only if $|\mathbf{a}+\mathbf{b}|^2 = |\mathbf{a}|^2 + |\mathbf{b}|^2$. □

We now discuss orthogonal decomposition. If $\mathbf{a}, \mathbf{b} \in V$ and $\mathbf{b} \neq \mathbf{0}$, we claim that there exists a unique $\lambda \in \mathbb{R}$ such that $\mathbf{a} = \lambda\mathbf{b} + \mathbf{c}$ and $(\mathbf{b},\mathbf{c}) = 0$. To see this, write $\mathbf{c} = \mathbf{a} - \lambda\mathbf{b}$. Using the properties of the inner product, we see that $(\mathbf{b},\mathbf{c}) = 0$ if and only if $\lambda = (\mathbf{a},\mathbf{b})/(\mathbf{b},\mathbf{b})$. By the previous proposition, this value of λ gives $|\mathbf{a}|^2 = \lambda^2|\mathbf{b}|^2 + |\mathbf{c}|^2$. Hence we get the following result.

Proposition 6.30. *Let* $\mathbf{a}$ *and* $\mathbf{b}$ *be elements of an inner product space* V*, and suppose* $\mathbf{b} \neq \mathbf{0}$*. Then* $\mathbf{a}$ *can be uniquely decomposed as a linear combination of two orthogonal vectors* $\mathbf{b}$ *and* $\mathbf{c}$ *as*

$$\mathbf{a} = \lambda\mathbf{b} + \mathbf{c}, \tag{6.17}$$

where $\lambda = (\mathbf{a},\mathbf{b})/(\mathbf{b},\mathbf{b})$ *and* $\mathbf{c} = \mathbf{a} - \lambda\mathbf{b}$.

In particular, since $|\mathbf{c}|^2 \geq 0$, it follows that

$$|\mathbf{a}|^2 \geq ((\mathbf{a},\mathbf{b})/(\mathbf{b},\mathbf{b}))^2|\mathbf{b}|^2.$$

Taking square roots gives a famous inequality.

Proposition 6.31 (Cauchy–Schwarz inequality). *For every* $\mathbf{a}, \mathbf{b} \in V$*,*

$$|(\mathbf{a},\mathbf{b})| \leq |\mathbf{a}||\mathbf{b}|, \tag{6.18}$$

with equality if and only if $\mathbf{a}$ *is a multiple of* $\mathbf{b}$.

Proof. The inequality surely holds when $\mathbf{b} = \mathbf{0}$. Thus suppose $\mathbf{b} \neq \mathbf{0}$. Then (6.18) follows from the inequality before the proposition, and equality holds if and only if $\mathbf{c} = \mathbf{0}$, or equivalently, if and only if $\mathbf{a}$ is a multiple of $\mathbf{b}$. □

The Cauchy–Schwarz inequality for $\mathbb{R}^n$ says that

$$|\sum_{i=1}^{n} a_i b_i| \leq \Big(\sum_{i=1}^{n} a_i^2\Big)^{1/2} \Big(\sum_{i=1}^{n} b_i^2\Big)^{1/2}.$$

The vector

$$P_{\mathbf{b}}(\mathbf{a}) = \big((\mathbf{a},\mathbf{b})/(\mathbf{b},\mathbf{b})\big)\mathbf{b}$$

is called the *orthogonal projection of* $\mathbf{a}$ *on* $\mathbf{b}$. Our final definition is the angle between two nonzero vectors. By the Cauchy–Schwarz inequality, $-1 \leq \frac{(\mathbf{a},\mathbf{b})}{|\mathbf{a}||\mathbf{b}|} \leq 1$. Hence there exists a unique $\theta \in [0,\pi]$ such that

$$\cos\theta = \frac{(\mathbf{a},\mathbf{b})}{|\mathbf{a}||\mathbf{b}|}. \tag{6.19}$$

Definition 6.14. The *angle* between two nonzero vectors $\mathbf{a}$ and $\mathbf{b}$ in an inner product space V is defined to be the unique angle $\theta \in [0,\pi]$ such that (6.19) holds.

In particular, $(\mathbf{a},\mathbf{b}) = 0$ if and only if the angle between $\mathbf{a}$ and $\mathbf{b}$ is $\pi/2$. In physics books, the identity

$$\mathbf{a}\cdot\mathbf{b} = |\mathbf{a}||\mathbf{b}|\cos\theta$$

is sometimes taken as a definition of the dot product. But this definition is not as easy to work with as the usual one.

Inner products can exist on infinite-dimensional vector spaces. This gives a method for extending properties of $\mathbb{R}^n$ to the infinite-dimensional setting. A particularly important example is the following.

Example 6.34. (An inner product on $C[a,b]$) Let $C[a,b]$ be the space of all continuous real-valued functions on the closed interval $[a,b]$ in $\mathbb{R}$. The inner product of $f, g \in C[a,b]$ is defined by

$$(f,g) = \int_a^b f(t)g(t)\mathrm{d}t.$$

The first three axioms for the inner product on $C[a,b]$ are verified by applying standard facts about integration proved (or at least stated) in calculus. The positive definiteness property requires that we verify that $(f,f) > 0$ if $f \neq 0$.

This involves a little bit of knowledge of how the Riemann integral is defined, and so we will skip the details. ⊠

6.6.3 Hermitian inner products

When we take up the principal axis theorem for Hermitian matrices in Chap. 9, we will need the notion of a Hermitian inner product space. Hermitian inner products are also very important in physics. We will first introduce the standard Hermitian inner product on $\mathbb{C}^n$, and then proceed to the general definition.

Example 6.35 (Hermitian n-space). The *Hermitian inner product* of a pair of vectors $\mathbf{w}, \mathbf{z} \in \mathbb{C}^n$ is defined to be the complex scalar

$$\mathbf{w} \bullet \mathbf{z} = \overline{\mathbf{w}}^T \mathbf{z} = \begin{pmatrix} \overline{w_1} & \overline{w_2} & \cdots & \overline{w_n} \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{pmatrix} = \sum_{i=1}^{n} \overline{w_i} z_i. \tag{6.20}$$

Let us put $\mathbf{w}^H = \overline{\mathbf{w}}^T$; $\mathbf{w}^H$ is called the *Hermitian transpose* of $\mathbf{w}$. Thus,

$$\mathbf{w} \bullet \mathbf{z} = \mathbf{w}^H \mathbf{z}.$$

Although $\mathbf{w} \bullet \mathbf{z}$ is not necessarily real, $\mathbf{w} \bullet \mathbf{w}$ is real, and in fact, $\mathbf{w} \bullet \mathbf{w} \geq 0$ for all $\mathbf{w}$. Thus we may define the Hermitian length function by

$$|\mathbf{w}| = (\mathbf{w} \bullet \mathbf{w})^{1/2} = (\mathbf{w}^H \mathbf{w})^{1/2} = \Big(\sum_{i=1}^{n} |w_i|^2 \Big)^{1/2}.$$

⊠

We now make a general definition.

Definition 6.15. Let V be a complex vector space. A *Hermitian inner product* on V is a rule assigning a scalar $(\mathbf{w}, \mathbf{z}) \in \mathbb{C}$ to every pair of vectors $\mathbf{w}, \mathbf{z} \in V$ such that

(i) $(\mathbf{w} + \mathbf{w}', \mathbf{z}) = (\mathbf{w}, \mathbf{z}) + (\mathbf{w}', \mathbf{z})$ and $(\mathbf{w}, \mathbf{z} + \mathbf{z}') = (\mathbf{w}, \mathbf{z}) + (\mathbf{w}, \mathbf{z}')$,

(ii) $(\mathbf{z}, \mathbf{w}) = \overline{(\mathbf{w}, \mathbf{z})}$,

(iii) $(\alpha \mathbf{w}, \mathbf{z}) = \overline{\alpha}(\mathbf{w}, \mathbf{z})$ and $(\mathbf{w}, \alpha \mathbf{z}) = \alpha(\mathbf{w}, \mathbf{z})$, and finally,

(iv) if $\mathbf{w} \neq \mathbf{0}$, $(\mathbf{w}, \mathbf{w}) > 0$.

A complex vector space endowed with a Hermitian inner product is called a *Hermitian inner product space.*

6.6.4 Orthonormal bases

In this section we will show that every finite-dimensional inner product space admits a type of basis called an orthonormal basis, which is analogous to the standard basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$ of Euclidean n-space $\mathbb{R}^n$. Let V denote a real inner product space.

Definition 6.16. A set $\mathcal{U}$ of unit vectors in V is called *orthonormal* if every pair of distinct elements of $\mathcal{U}$ are orthogonal to each other.

Example 6.36. The standard basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$ determines an orthonormal set in Euclidean n-space $\mathbb{R}^n$. ⊠

Proposition 6.32. *Every orthonormal subset $\mathcal{U}$ of V is linearly independent. (That is, every finite subset of $\mathcal{U}$ is independent.) In particular, if V is finite-dimensional, then every orthonormal set in V having* $\dim V$ *elements is a basis of V.*

Proof. Suppose $\mathbf{u}_1, \ldots, \mathbf{u}_m$ are orthonormal, and assume that

$$\sum_{i=1}^{m} a_i \mathbf{u}_i = \mathbf{0}.$$

Then for every index j,

$$\Big(\sum_{i=1}^{m} a_i \mathbf{u}_i, \mathbf{u}_j\Big) = (\mathbf{0}, \mathbf{u}_j) = 0.$$

Since $(\mathbf{u}_i, \mathbf{u}_j)$ equals 0 if $i \neq j$ and equals 1 if $i = j$, the left-hand side is

$$\sum_{i=1}^{m} a_i (\mathbf{u}_i, \mathbf{u}_j) = a_j,$$

so $a_j = 0$ for all j. Therefore, $\mathbf{u}_1, \ldots, \mathbf{u}_m$ are independent. Hence if $|\mathcal{U}| = \dim V$, then $\mathcal{U}$ is a basis of V. □

Definition 6.17. An *orthonormal basis* of V is a basis that is an orthonormal set.

Proposition 6.33. *A collection of vectors $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n$ in $\mathbb{R}^n$ is an orthonormal basis of $\mathbb{R}^n$ if and only if the matrix $U = (\mathbf{u}_1\ \mathbf{u}_2 \ldots \mathbf{u}_n)$ is orthogonal. Thus the set of $U \in \mathbb{F}^{n\times n}$ whose columns are an orthonormal basis of $\mathbb{R}^n$ is exactly the orthogonal group $O(n, \mathbb{R})$.*

Proof. Recall that U is orthogonal if and only if $U^TU = I_n$. But $U^TU = (\mathbf{u}_i^T\mathbf{u}_j) = (\mathbf{u}_i \cdot \mathbf{u}_j) = I_n$ if and only if $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n$ are orthonormal. □

For example, the standard basis vectors $\mathbf{e}_1, \ldots, \mathbf{e}_n$ in $\mathbb{R}^n$ are the columns of I_n. Here are some more examples.

Example 6.37. The vectors

$$\mathbf{u}_1 = \frac{1}{\sqrt{3}}(1,1,1)^T,\ \mathbf{u}_2 = \frac{1}{\sqrt{6}}(1,-2,1)^T,\ \mathbf{u}_3 = \frac{1}{\sqrt{2}}(1,0,-1)^T$$

form an orthonormal basis of $\mathbb{R}^3$. Moreover, $\mathbf{u}_1$ and $\mathbf{u}_2$ constitute an orthonormal basis of the plane $x - z = 0$. ⊠

Example 6.38. The matrix

$$Q = \frac{1}{2}\begin{pmatrix} 1 & 1 & 1 & 1 \\ -1 & 1 & -1 & 1 \\ 1 & -1 & -1 & 1 \\ 1 & 1 & -1 & -1 \end{pmatrix}$$

is orthogonal. Hence its columns form an orthonormal basis of $\mathbb{R}^4$. Since Q^T is also orthogonal, the rows of B form another orthonormal basis of $\mathbb{R}^4$. ⊠

6.6.5 The existence of orthonormal bases

We now show that every finite-dimensional inner product space has an orthonormal basis. In fact, we show a little more.

Proposition 6.34. *Let V be an inner product space. Then every nontrivial finite-dimensional subspace W of V admits an orthonormal basis.*

Proof. We prove this by induction on $\dim W$. Note that every subspace of V is also an inner product space via restricting the inner product on V to W. If $\dim W = 1$, the result is true, since a unit vector in W is an orthonormal basis. Thus suppose $\dim W = m > 1$ and that the result is true for every subspace of W of dimension at most $m - 1$. Let $\mathbf{u}$ be a unit vector in W and let $H = \{\mathbf{x} \in W \mid (\mathbf{x}, \mathbf{u}) = 0\}$. Then H is a subspace of W. Since $|\mathbf{u}| = 1$, $\mathbf{u} \notin H$. It follows from Corollary 6.12 that $\dim H < m$. Thus, by the induction hypothesis, H admits an orthonormal basis, say $\mathcal{U}$. Now I claim that $\mathcal{U}$ and $\mathbf{u}$ combine to give an orthonormal basis of W. Clearly, $\mathcal{U}$ and $\mathbf{u}$ form an orthonormal set, so it suffices to check that they span W. Let $\mathbf{x}$ be an arbitrary element of W, and let $\mathbf{y} = \mathbf{x} - (\mathbf{x}, \mathbf{u})\mathbf{u}$. Then

$$(\mathbf{y}, \mathbf{u}) = (\mathbf{x}, \mathbf{u}) - (\mathbf{x}, \mathbf{u})(\mathbf{u}, \mathbf{u}) = 0,$$

since $(\mathbf{u}, \mathbf{u}) = 1$. Thus $\mathbf{y} \in H$, so $\mathbf{y}$ is a linear combination of the elements of $\mathcal{U}$. Since $\mathbf{x} = \mathbf{y} + (\mathbf{x}, \mathbf{u})\mathbf{u}$, $\mathbf{x}$ is in the span of $\mathbf{u}$ and $\mathcal{U}$. Therefore W has an orthonormal basis, so the result is proven. □

The above proof gives us the following

Corollary 6.35. *If* $\mathbf{u}$ *is a unit vector in a finite-dimensional inner product space* V*, then*

$$H = \{\mathbf{x} \in V \mid (\mathbf{x}, \mathbf{u}) = 0\}$$

is a hyperplane in V*. That is,* $\dim H = \dim V - 1$.

Here is an example not involving $\mathbb{R}^n$.

Example 6.39. Let V denote the set of functions $f(x) = ax^2 + bx + c$ on $[-1, 1]$, where a, b, c are arbitrary real numbers. For $f, g \in V$, let $(f, g) = \int_{-1}^{1} f(x)g(x)dx$. In other words, V is a three-dimensional subspace of the inner product space $C[-1, 1]$. The functions 1, x, and x^2 are a basis of V, but unfortunately they aren't orthonormal. For although x is orthogonal to 1 and x^2, 1 and x^2 aren't orthogonal to each other: $\int_{-1}^{1} 1x^2 dx = 2/3$. To correct this, we replace x^2 with $x^2 - r$, where r is chosen such that $(1, x^2 - r) = 0$. Since $(1, 1) = 2$, it is easy to see that we should let $r = 1/3$. Thus 1, x, $x^2 - 1/3$ are orthogonal on $[-1, 1]$. We therefore obtain the orthonormal basis by normalizing. The result is $\mathbf{u}_1 = 1/\sqrt{2}$, $\mathbf{u}_2 = \sqrt{3/2}\, x$, and $\mathbf{u}_3 = c(x^2 - 1/3)$ where $c = \left(\int_{-1}^{1} (x^2 - 1/3)^2 dx\right)^{-1/2}$. ⊠

6.6.6 Fourier coefficients

We are now going to see one of the reasons that an orthonormal basis is very useful. If $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ is a basis of a vector space V, how does one express an arbitrary element $\mathbf{v}$ of V as a linear combination of these basis vectors? If $V = \mathbb{F}^n$, then this involves solving the linear system $A\mathbf{x} = \mathbf{v}$, where $A = \begin{pmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n \end{pmatrix}$. That is, $\mathbf{x} = A^{-1}\mathbf{v}$. But if V is arbitrary, we don't yet have a general method. On the other hand, if V has an orthonormal basis, there is a simple elegant solution.

Proposition 6.36. *Let* $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n$ *be an orthonormal basis of* V*. Then if* $\mathbf{w} \in V$*, we have*

$$\mathbf{w} = \sum_{1=i}^{n} (\mathbf{w}, \mathbf{u}_i)\mathbf{u}_i. \tag{6.21}$$

Proof. Let $\mathbf{w} = \sum_{1=i}^{n} x_i \mathbf{u}_i$. Then $(\mathbf{w}, \mathbf{u}_j) = \sum_{1=i}^{n} x_i(\mathbf{u}_i, \mathbf{u}_j) = x_j$, since the $\mathbf{u}_i$ are orthonormal. □

The coefficients $(\mathbf{w}, \mathbf{u}_i)$ in (6.21) are called the *Fourier coefficients* of $\mathbf{w}$ with respect to the orthonormal basis $\mathbf{u}_1, \ldots, \mathbf{u}_n$. We may also refer to (6.21) as the *Fourier expansion* of $\mathbf{w}$ in terms of the given orthonormal basis.

If $V = \mathbb{R}^n$, then (6.21) can be expressed in matrix form $I_n = QQ^T$; that is,

$$I_n = \sum_{1=i}^{n} \mathbf{u}_i \mathbf{u}_i^T. \tag{6.22}$$

Example 6.40. In terms of the orthonormal basis of Example 6.38, one gets, for example, that

$$\begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} = \frac{1}{4}\begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} - \frac{1}{4}\begin{pmatrix} -1 \\ 1 \\ -1 \\ 1 \end{pmatrix} + \frac{1}{4}\begin{pmatrix} 1 \\ -1 \\ -1 \\ 1 \end{pmatrix} + \frac{1}{4}\begin{pmatrix} 1 \\ 1 \\ -1 \\ -1 \end{pmatrix}.$$

⊠

6.6.7 The orthogonal complement of a subspace

Let U be a subset of an inner product space V.

Definition 6.18. The *orthogonal complement* of U is defined to be the set $U^\perp$ consisting of all vectors $\mathbf{v} \in V$ orthogonal to every vector in U.

Thus,

$$U^\perp = \{\mathbf{v} \in V \mid (\mathbf{v}, \mathbf{u}) = 0 \ \forall \ \mathbf{u} \in U\}. \tag{6.23}$$

Proposition 6.37. *For every subset U of an inner product space V, $U^\perp$ is a subspace of V. Moreover, if $W = \operatorname{span} U$, then $W \cap U^\perp = \{\mathbf{0}\}$.*

Proof. This is an exercise.

It may be instructive to visualize $W^\perp$ in matrix terms. Let $A \in \mathbb{R}^{m\times n}$, and let U denote its columns. Then $W = \operatorname{span} U = \operatorname{col}(A)$, the column space of A, and $W^\perp = \mathcal{N}(A^T)$. By (6.6), $\dim \mathcal{N}(A^T) + \dim \operatorname{row}(A^T) = m$. But $\operatorname{row}(A^T)$ and $W = \operatorname{col}(A)$ certainly have the same dimension, so we get the identity

$$\dim W + \dim W^\perp = m. \tag{6.24}$$

Thus, the column space of a matrix and the null space of its transpose are each the orthogonal complement of the other. We now prove a more general version of this.

Proposition 6.38. *Let W be a subspace of a finite-dimensional inner product space V and $W^\perp$ its orthogonal complement. Then $V = W \oplus W^\perp$. Thus,* $\dim V = \dim W + \dim W^\perp$. *In particular, every $\mathbf{v} \in V$ can be orthogonally decomposed in exactly one way as $\mathbf{v} = \mathbf{w} + \mathbf{y}$, where $\mathbf{w} \in W$ and $\mathbf{y} \in W^\perp$.*

Proof. By Proposition 6.34, we may choose an orthonormal basis of W, say $\mathbf{u}_1, \ldots, \mathbf{u}_k$. Let $\mathbf{v} \in V$ and put

$$\mathbf{y} = \mathbf{v} - \sum_{i=1}^{k} (\mathbf{v}, \mathbf{u}_i)\mathbf{u}_i. \tag{6.25}$$

Since the $\mathbf{u}_i$ are orthonormal, we have $(\mathbf{y}, \mathbf{u}_i) = 0$ for all i. Thus, by definition, $\mathbf{y} \in W^\perp$. But this says that if $\mathbf{w} = \sum_{i=1}^{k} (\mathbf{v}, \mathbf{u}_i)\mathbf{u}_i$, then $\mathbf{v} = \mathbf{w} + \mathbf{y}$. Therefore, $V = W + W^\perp$. Since $W \cap W^\perp = \{\mathbf{0}\}$, we get $V = W \oplus W^\perp$ by Proposition 6.23. Hence, $\dim W + \dim W^\perp = \dim V$. □

Definition 6.19. Let $\mathbf{v} = \mathbf{w} + \mathbf{y}$ be the above decomposition of $\mathbf{v} \in V$ with $\mathbf{w} \in W$. Then $\mathbf{w}$ is called the *component* of $\mathbf{v}$ in W.

Thus if $\mathbf{u}_1, \ldots, \mathbf{u}_k$ is an orthonormal basis of W, the component of an arbitrary vector $\mathbf{v} \in V$ is

$$\mathbf{w} = \sum_{i=1}^{k} (\mathbf{v}, \mathbf{u}_i)\mathbf{u}_i. \tag{6.26}$$

In particular, if W is a line, say $W = \mathbb{R}\mathbf{w}$, then the component of an arbitrary $\mathbf{v}$ in V can be easily worked out, since

$$\mathbf{u} = \frac{\mathbf{w}}{(\mathbf{w}, \mathbf{w})^{1/2}}$$

is an orthonormal basis for W. In particular,

$$\mathbf{v} = (\mathbf{v}, \mathbf{u})\mathbf{u} + \big(\mathbf{v} - (\mathbf{v}, \mathbf{u})\mathbf{u}\big).$$

6.6.8 Hermitian inner product spaces

The results about orthonormal bases in the case of an inner product space all have analogues for the Hermitian inner product spaces that were introduced in Section 6.6.3. Recall that the main example of a Hermitian inner product space is $\mathbb{C}^n$ with the Hermitian inner product $(\mathbf{w}, \mathbf{z}) = \mathbf{w}^H \mathbf{z}$. A basis $\mathbf{w}_1, \ldots, \mathbf{w}_n$ of a Hermitian inner product space V is said to be a *Hermitian orthonormal basis*, provided each $|\mathbf{w}_i|$ is equal to 1 and $(\mathbf{w}_i, \mathbf{w}_j) = 0$ if $i \neq j$. By imitating the proof of Proposition 6.34, one can prove the following result.

Proposition 6.39. *Every finite-dimensional Hermitian inner product space has a Hermitian orthonormal basis. Moreover, the identity (6.21) holds for every Hermitian orthonormal basis.*

Orthogonal complements are defined in the Hermitian case in exactly the same way as in the real case, and the Hermitian version of Proposition 6.38 goes through without any change. Hermitian orthonormal bases of $\mathbb{C}^n$ are related to unitary matrices in the same way that orthonormal bases of $\mathbb{R}^n$ are related to orthogonal matrices. Recall that if $U \in \mathbb{C}^{n\times n}$, then $U^H = (\overline{U})^T$.

Definition 6.20. A matrix $U \in \mathbb{C}^{n\times n}$ is said to be *unitary* if $U^H U = I_n$. The set of all $n \times n$ unitary matrices is denoted by $U(n)$.

Thus, unitary matrices are to the Hermitian inner product on $\mathbb{C}^n$ as orthogonal matrices are to the Euclidean inner product on $\mathbb{R}^n$.

Proposition 6.40. *$U(n)$ is a matrix group that is a subgroup of $GL(n, \mathbb{C})$.*

Proof. Exercise. □

Thus $U(n)$ is called the unitary group.

Exercises

Exercise 6.6.1. A nice application of Cauchy–Schwarz is the following fact: if $\mathbf{a}$ and $\mathbf{b}$ are unit vectors in $\mathbb{R}^n$ such that $\mathbf{a} \cdot \mathbf{b} = 1$, then $\mathbf{a} = \mathbf{b}$. Prove this.

Exercise 6.6.2. Prove the law of cosines: if a triangle has sides with lengths a, b, c, and θ is the angle opposite the side of length c, then $c^2 = a^2 + b^2 - 2ab\cos\theta$. (Hint: Consider $\mathbf{c} = \mathbf{b} - \mathbf{a}$.)

Exercise 6.6.3. Show that the Killing form $(A, B) = \mathrm{Tr}(AB^T)$ introduced in Example 6.33 is an inner product on the space $\mathbb{R}^{2\times 2}$ of real 2×2 matrices and find an orthonormal basis.

Exercise 6.6.4. Show that the orthogonal complement with respect to the Killing form of the space of 2×2 symmetric real matrices is the space of 2×2 skew symmetric real matrices. Conclude $\mathbb{R}^{2\times 2} = \mathbb{R}^{2\times 2}_s \oplus \mathbb{R}^{2\times 2}_{ss}$.

Exercise 6.6.5. The proof that the Killing form on $\mathbb{R}^{n\times n}$ is an inner product requires showing that $(A, B) = (B, A)$. Show this by proving the following statements.

(*i*) For all $A, B \in \mathbb{R}^{n\times n}$, $\mathrm{Tr}(AB) = \mathrm{Tr}(BA)$;

(*ii*) For all $A, B \in \mathbb{R}^{n\times n}$, $\mathrm{Tr}(AB^T) = \mathrm{Tr}(BA^T)$

Exercise 6.6.6. Orthogonally decompose the vector $(1, 2, 2)^T$ in $\mathbb{R}^3$ as $\mathbf{p}+\mathbf{q}$, where $\mathbf{p}$ is required to be a multiple of $(3, 1, 2)^T$ and $\mathbf{q}$ is orthogonal to $\mathbf{p}$.

Exercise 6.6.7. In this exercise, we consider the inner product space $V = C[-1, 1]$ of continuous real-valued functions on $[-1, 1]$ with inner product defined by $(f, g) = \int_{-1}^{1} f(t)g(t)dt$.

(i) Show that the functions 1 and x are orthogonal. In fact, show that x^k and x^m are orthogonal if k is even and m is odd.

(ii) Find the projection of x^2 on the constant function 1.

(iii) Use this to obtain the orthogonal decomposition of x^2 on $[-1, 1]$ in which one of the components has the form $r1$.

Exercise 6.6.8. Consider the real vector space $V = C[0, 2\pi]$ with the inner product defined by $(f, g) = \int_0^{2\pi} f(t)g(t)dt$.

(i) Find the length of $\sin^2 x$ in V.

(ii) Compute the inner product $(\cos x, \sin^2 x)$.

(iii) Find the projection of $\sin^2 x$ on each of the functions 1, $\cos x$, and $\sin x$ in V.

(iv) Are 1, $\cos x$, and $\sin x$ mutually orthogonal as elements of V?

(v) Compute the orthogonal projection of $\sin^2 x$ onto the subspace W of V spanned by 1, $\cos x$, and $\sin x$.

Exercise 6.6.9. Assume $f \in C[a, b]$. The average value of f over $[a, b]$ is defined to be

$$\frac{1}{b-a}\int_a^b f(t)dt.$$

Show that the average value of f over $[a, b]$ is the projection of f on 1. Does this suggest an interpretation of the average value?

Exercise 6.6.10. Let $f, g \in C[a, b]$. Give a formula for the scalar t that minimizes

$$||f - tg||^2 = \int_a^b (f(x) - tg(x))^2 dx.$$

Exercise 6.6.11. Show that the Hermitian inner product on $\mathbb{C}^n$ satisfies all the conditions listed in Definition 6.15.

Exercise 6.6.12. Consider the plane P in $\mathbb{R}^3$ given by the equation $x - y + 2z = 0$.

(i) Find an orthonormal basis of P.

(ii) Find the expansion of $(1, 1, 0)^T$ in terms of this orthonormal basis.

Exercise 6.6.13. Find the expansion of $(2,0,0)^T$ in terms of the orthonormal basis of Example 6.37.

Exercise 6.6.14. The Gram–Schmidt method gives an algorithm for producing an orthonormal basis of an inner product space starting from a general basis. Here is how it works for $\mathbb{R}^3$. Let $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ be a basis. First put

$$\mathbf{u}_1 = \frac{\mathbf{v}_1}{|\mathbf{v}_1|}.$$

Next, put

$$\mathbf{v}_2' = \mathbf{v}_2 - (\mathbf{v}_2 \cdot \mathbf{u}_1)\mathbf{u}_1 \quad \text{and} \quad \mathbf{u}_2 = \frac{\mathbf{v}_2'}{|\mathbf{v}_2'|}.$$

Finally, put

$$\mathbf{v}_3' = \mathbf{v}_3 - (\mathbf{v}_3 \cdot \mathbf{u}_1)\mathbf{u}_1 - (\mathbf{v}_3 \cdot \mathbf{u}_2)\mathbf{u}_2 \quad \text{and} \quad \mathbf{u}_3 = \frac{\mathbf{v}_3'}{|\mathbf{v}_3'|}.$$

Verify that $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3$ form an orthonormal basis of $\mathbb{R}^3$ having the property that $\text{span}\{\mathbf{u}_1\} = \text{span}\{\mathbf{v}_1\}$ and $\text{span}\{\mathbf{u}_1, \mathbf{u}_2\} = \text{span}\{\mathbf{v}_1, \mathbf{v}_2\}$. Why are $\mathbf{v}_2'$ and $\mathbf{v}_3'$ both nonzero?

Exercise 6.6.15. Generalize the Gram–Schmidt method from $\mathbb{R}^3$ to $\mathbb{R}^4$.

Exercise 6.6.16. Let W denote the hyperplane $w + x - y + z = 0$ in $\mathbb{R}^4$.

(*i*) Find the component of $(1,1,1,1)^T$ in W.

(*ii*) Find an orthonormal basis of W.

(*iii*) Find an orthonormal basis of $\mathbb{R}^4$ containing the orthonormal basis of part (*ii*).

(*iv*) Expand $(1,1,1,1)^T$ in terms of the basis in part (*iii*).

Exercise 6.6.17. Show that if W is a subspace of a finite-dimensional inner product space V, then $(W^\perp)^\perp = W$.

Exercise 6.6.18. Recall that the group $P(n)$ of $n \times n$ permutation matrices is a subgroup of $O(n, \mathbb{R})$. Show that the set of all left cosets $O(n, \mathbb{R})/P(n)$ is in one-to-one correspondence with the set of all orthonormal bases of $\mathbb{R}^n$.

Exercise 6.6.19. Let V be an inner product space. Show that the distance function $d(\mathbf{a}, \mathbf{b}) = |\mathbf{a} - \mathbf{b}|$ on $V \times V$ Defines a *metric* on V in the sense that for all $\mathbf{a}, \mathbf{b}, \mathbf{c} \in V$, we have the following properties:

(*i*) $d(\mathbf{a}, \mathbf{b}) \geq 0$ and $d(\mathbf{a}, \mathbf{b}) = 0$ if and only if $\mathbf{a} = \mathbf{b}$.

(*ii*) $d(\mathbf{a}, \mathbf{b}) = d(\mathbf{b}, \mathbf{a})$, and

(*iii*) $d(\mathbf{a}, \mathbf{c}) \leq d(\mathbf{a}, \mathbf{b}) + d(\mathbf{b}, \mathbf{c})$.
Another quite different example of a metric is given in the Appendix.

Exercise 6.6.20. Let V be a finite-dimensional inner product space and W a subspace. The distance $d(\mathbf{v}, W)$ from an arbitrary vector $\mathbf{v} \in V$ to the subspace W is defined to be the minimum distance $d(\mathbf{v}, \mathbf{w})$ as $\mathbf{w}$ ranges over W.

(i) If $\mathbf{v} \in W$, show that the distance from $\mathbf{v}$ to W is 0.

(ii) Suppose $\mathbf{v} = \mathbf{w} + \mathbf{y}$ is the orthogonal decomposition of $\mathbf{v}$ with $\mathbf{w} \in W$ and $\mathbf{y} \in W^{\perp}$. Show that $d(\mathbf{v}, \mathbf{w}) \leq d(\mathbf{v}, \mathbf{w}')$ for all $\mathbf{w}' \in W$ and conclude that $d(\mathbf{v}, W) = |\mathbf{y}|$. (This fact is called the principle of least squares.)

Exercise 6.6.21. Find the distance from $(1, 1, 1, 1)^T$ to the subspace of $\mathbb{R}^4$ spanned by $(2, 0, -1, 1)^T$ and $(0, 0, 1, 1)^T$.

Exercise 6.6.22. If K is a subspace of $\mathbb{C}^n$ with the Hermitian inner product, show that $\dim K + \dim K^{\perp} = n$, where $K^{\perp}$ is the orthogonal complement of K with respect to the Hermitian inner product.

Exercise 6.6.23. Find a Hermitian orthonormal basis of the subspace of $\mathbb{C}^3$ spanned by $(1, i, 0)^T$ and $(2, -i, 1)^T$, and then extend this basis to a Hermitian orthonormal basis of $\mathbb{C}^3$.

Exercise 6.6.24. Prove Proposition 6.40. That is, show that $U(n)$ is a matrix group.

Exercise 6.6.25. The complex analogue of the Killing form on $\mathbb{C}^{n \times n}$ is $(J, K) = \mathrm{Tr}(JK^H)$. Show that (J, K) defines a Hermitian inner product on $\mathbb{C}^{2 \times 2}$.

6.7 Vector Space Quotients

In the penultimate section of this chapter, we will construct the quotient vector space V/W of a vector space V by a subspace W. Since V is an abelian group and every subspace is a normal subgroup, this construction is an application of the construction of the quotient group G/H of a group G by a normal subgroup H (see Proposition 2.17). Thus, the space of cosets V/W is also an abelian group, and the natural map $\pi : V \to V/W$ is a group homomorphism.

Since we are not formally using results from group theory in this chapter, we will instead construct V/W from scratch. We will then show that if V is finite-dimensional, then V/W is a finite-dimensional vector space and $\dim(V/W) = \dim V - \dim W$. Unfortunately, the vector space V/W doesn't admit a useful geometric interpretation; one must think of V/W as an abstract vector space.

6.7.1 Cosets of a subspace

Let V be a vector space over $\mathbb{F}$ and let W be a subspace. The cosets of W were introduced in Section 2.2 in the setting of groups. Redefining them from scratch in the vector space setting goes as follows.

Definition 6.21. A *coset of* W is a subset of V of the form

$$\mathbf{v} + W = \{\mathbf{v} + \mathbf{w} \mid \mathbf{w} \in W\}, \tag{6.27}$$

where $\mathbf{v} \in V$.

The cosets of W are the equivalence classes of an equivalence relation E_W on V. Namely, if $\mathbf{u}, \mathbf{v} \in V$, let us write $\mathbf{u} E_W \mathbf{v}$ if $\mathbf{v} - \mathbf{u} \in W$. If $\mathbf{u} E_W \mathbf{v}$, we will say that $\mathbf{u}$ and $\mathbf{v}$ are *equivalent modulo* W.

Proposition 6.41. *Let W be a subspace of a vector space V. Then E_W is an equivalence relation on V, and the equivalence classes of this equivalence relation are exactly the cosets of W.*

Proof. Clearly $\mathbf{v} E_W \mathbf{v}$, since $\mathbf{v} - \mathbf{v} = \mathbf{0} \in W$. If $\mathbf{u} E_W \mathbf{v}$, then $\mathbf{v} E_W \mathbf{u}$, since W is closed under scalar multiplication, and $(\mathbf{u} - \mathbf{v}) = (-1)(\mathbf{v} - \mathbf{u})$. Finally, if $\mathbf{u} E_W \mathbf{v}$ and $\mathbf{v} E_W \mathbf{w}$, then $\mathbf{u} E_W \mathbf{w}$, since $\mathbf{w} - \mathbf{u} = (\mathbf{w} - \mathbf{v}) + (\mathbf{v} - \mathbf{u})$, and W is closed under addition. Hence E_W is an equivalence relation on V. Let C denote the equivalence class of $\mathbf{v}$ and consider $\mathbf{v} + W$. If $\mathbf{y} \in C$, then $\mathbf{y} - \mathbf{v} = \mathbf{w} \in W$. Hence $\mathbf{y} = \mathbf{v} + \mathbf{w}$, so $\mathbf{y} \in \mathbf{v} + W$. Therefore, $C \subset \mathbf{v} + W$. On the other hand, suppose $\mathbf{y} \in \mathbf{v} + W$. Then $\mathbf{y} = \mathbf{v} + \mathbf{w}$ for some $\mathbf{w} \in W$. But then $\mathbf{y} - \mathbf{v} \in W$, so $\mathbf{y} E_W \mathbf{v}$. Therefore, $\mathbf{v} + W \subset C$. Hence the equivalence classes are exactly the cosets of W. □

Example 6.41. For example, suppose $V = \mathbb{R}^3$ and W is a plane through $\mathbf{0}$. Then the coset $\mathbf{v} + W$ is simply the plane through $\mathbf{v}$ parallel to W. The cosets are all the planes in $\mathbb{R}^3$ parallel to W. ⊠

6.7.2 The quotient V/W and the dimension formula

We will refer to the set of cosets V/W as the quotient space of V modulo W. Two cosets $(\mathbf{v} + W)$ and $(\mathbf{y} + W)$ may be added by putting

$$(\mathbf{v} + W) + (\mathbf{y} + W) = (\mathbf{v} + \mathbf{y}) + W. \tag{6.28}$$

To make V/W into a vector space over $\mathbb{F}$, we also have to define scalar multiplication, which we do in a natural way: for $a \in \mathbb{F}$ and $\mathbf{v} \in V$, put

$$a(\mathbf{v} + W) = a\mathbf{v} + W. \tag{6.29}$$

The proof that addition is a well-defined binary operation uses the same reasoning as for the quotient group. We need to show that the rule (6.28) is independent of the way we write a coset. That is, suppose $\mathbf{v} + W = \mathbf{v}' + W$ and $\mathbf{y} + W = \mathbf{y}' + W$. Then we have to show that $(\mathbf{v}+\mathbf{y})+W = (\mathbf{v}'+\mathbf{y}')+W$. But this is so if and only if

$$(\mathbf{v} + \mathbf{y}) - (\mathbf{v}' + \mathbf{y}') \in W,$$

which indeed holds, since

$$(\mathbf{v} + \mathbf{y}) - (\mathbf{v}' + \mathbf{y}') = (\mathbf{v} - \mathbf{v}') + (\mathbf{y} - \mathbf{y}') \in W,$$

due to the fact that W is a subspace and both $\mathbf{v} - \mathbf{v}'$ and $\mathbf{y} - \mathbf{y}'$ are in W. Therefore, addition on V/W is well defined. The proof for scalar multiplication is analogous. The zero element is $\mathbf{0} + W$, and the additive inverse $-(\mathbf{v}+W)$ of $\mathbf{v}+W$ is $-\mathbf{v}+W$. Properties such as associativity and commutativity of addition follow from corresponding properties in V; we will omit all the details. Hence V/W is an $\mathbb{F}$-vector space, which is the first assertion of the following proposition. The second assertion gives a formula for $\dim V/W$ in the finite-dimensional setting.

Proposition 6.42. *Let V be a vector space over a field $\mathbb{F}$ and suppose W is a subspace of V. Then the set V/W of cosets of W in V with addition and scalar multiplication defined as in (6.28) and (6.29) is a vector space over $\mathbb{F}$. If V is finite-dimensional, then*

$$\dim V/W = \dim V - \dim W. \tag{6.30}$$

Proof. To check the dimension formula (6.30), let $\mathbf{w}_1, \ldots, \mathbf{w}_k$ be a basis of W, and extend this to a basis

$$\mathbf{w}_1, \ldots, \mathbf{w}_k, \mathbf{v}_1, \ldots, \mathbf{v}_{n-k}$$

of V. Then I claim that the cosets $\mathbf{v}_1 + W, \ldots, \mathbf{v}_{n-k} + W$ give a basis of V/W. To see that they are independent, put $\mathbf{v}_i + W = \alpha_i$ if $1 \leq i \leq n-k$, and suppose there exist $a_1, \ldots, a_{n-k} \in \mathbb{F}$ such that $\sum_{i=1}^{n-k} a_i \alpha_i = \mathbf{0} + W$. This means that $\sum_{i=1}^{n-k} a_i \mathbf{v}_i \in W$. Hence there exist $b_1, \ldots, b_k \in \mathbb{F}$ such that

$$\sum_{i=1}^{n-k} a_i \mathbf{v}_i = \sum_{j=1}^{k} b_j \mathbf{w}_j.$$

But the fact that the $\mathbf{v}_i$ and $\mathbf{w}_j$ constitute a basis of V implies that all a_i and b_j are zero. Therefore, $\alpha_1, \ldots, \alpha_{n-k}$ are linearly independent. We leave the fact that they span V/W as an exercise. □

Here is an example that shows how V/W can interpreted.

Example 6.42. Suppose $A \in \mathbb{F}^{m \times n}$, and recall that $\mathcal{N}(A) \subset \mathbb{F}^n$ is the null space of A. By Proposition 3.17, the elements of the quotient space $\mathbb{F}^n/\mathcal{N}(A)$ are the solution sets of the linear systems $A\mathbf{x} = \mathbf{b}$, where $\mathbf{b}$ varies through $\mathbb{F}^m$. The zero element $\mathcal{N}(A)$ corresponds to the homogeneous linear system $A\mathbf{x} = \mathbf{0}$, while the element $\mathbf{p}_0 + \mathcal{N}(A)$ corresponds to the inhomogeneous linear system $A\mathbf{x} = \mathbf{b}$, where $A(\mathbf{p}_0) = \mathbf{b}$. Suppose $\mathbf{p}_0, \mathbf{q}_0 \in \mathbb{F}^n$, and suppose $A(\mathbf{p}_0) = \mathbf{b}$ and $A(\mathbf{q}_0) = \mathbf{c}$. By definition, $(\mathbf{p}_0 + \mathcal{N}(A)) + (\mathbf{q}_0 + \mathcal{N}(A)) = (\mathbf{p}_0 + \mathbf{q}_0) + \mathcal{N}(A)$. But $A(\mathbf{p}_0 + \mathbf{q}_0) = \mathbf{b} + \mathbf{c}$, so here the addition of cosets of $\mathcal{N}(A)$ corresponds to the addition of linear systems. ⊠

Exercises

Exercise 6.7.1. Prove that the cosets $\alpha_1, \ldots, \alpha_{n-k}$ defined in the proof of Proposition 6.42 span V/W.

Exercise 6.7.2. Let W be the subspace of $V = (\mathbb{F}_2)^4$ spanned by 1001, 1101, and 0110. Write down all elements of W, and find a complete set of coset representatives for V/W. That is, find an element in each coset.

Exercise 6.7.3. Let A and B be arbitrary subsets of a vector space V over $\mathbb{F}$. Define their Minkowski sum to be

$$A + B = \{\mathbf{x} + \mathbf{y} \mid \mathbf{x} \in A, \quad \mathbf{y} \in B\}.$$

Show that if A and B are cosets of a subspace W of V, then so is $A + B$.

Exercise 6.7.4. Let V and W be any two subspaces of $\mathbb{F}^n$.

(*i*) Find a formula for $\dim(V + W)/W$.

(*ii*) Are the dimensions of $(V + W)/W$ and $V/(V \cap W)$ the same?

Exercise 6.7.5. Find a basis of the quotient $\mathbb{R}^4/W$, where W is the subspace of $\mathbb{R}^4$ spanned by $(1, 2, 0, 1)$ and $(0, 1, 1, 0)$.

Exercise 6.7.6. Let V be a vector space over $\mathbb{F}_p$ of dimension n, and let W be a subspace of dimension k.

(*i*) Show that every coset of W has p^k elements. (Find a bijection from W to $\mathbf{x} + W$.)

(*ii*) Show that the number of cosets of W is $p^{(n-k)}$.

6.8 Appendix: Linear Coding Theory

In 1948, a mathematician and electrical engineer named Claude Shannon published a fundamental paper entitled "A Mathematical Theory of Communication." This was followed in 1950 by a groundbreaking paper by Richard Hamming on error-detecting codes. These papers laid the foundations for coding theory and the theory of error-detecting codes, both of which have been crucial components of the electronic revolution brought about by computers and the Internet. The purpose of this appendix is to give a brief introduction to linear coding theory and to explain how error detection operates. The ideas we will introduce here depend on concepts from the theory of finite-dimensional vector spaces over a Galois field. In the final section, we will play the hat game.

6.8.1 The notion of a code

In mathematics, a *code* is just a subset of a vector space over a Galois field. Let p be a prime, and recall that $V(n,p)$ denotes the vector space $(\mathbb{F}_p)^n$ over the field $\mathbb{F}_p$. Subsets of $V(n,p)$ are called p-ary codes of length n. The elements of a code C are called its *codewords*, and the number of codewords is denoted by $|C|$.

A p-ary *linear code of length* n is a code $C \subset V(n,p)$ that is also a subspace. By definition, all linear codes are finite-dimensional. For example, a code $C \subset V(n,2)$ consists of binary strings, i.e., strings of 0's and 1's, of length n. Such a code C is linear if and only if the sum of two codewords is again a codeword. Having the structure of a vector space gives a linear code some advantages over nonlinear codes, one of them being that a linear code is completely determined once any set of codewords that spans it is given. A basis of a linear code is called a set of *basic codewords*. Recall from the dimension theorem that two bases of a finite-dimensional vector space always have the same number of elements, namely the dimension of C. If $C \subset V(n,p)$, then $\dim C$ determines the number of codewords by the formula $|C| = p^{\dim C}$.

Example 6.43. The equation $x_1 + x_2 + x_3 + x_4 = 0$ over $\mathbb{F}_2$ defines a linear code $C = \mathcal{N}(M)$, where $M = \begin{pmatrix} 1 & 1 & 1 & 1 \end{pmatrix}$. Since $\dim(C) = 3$, there are $8 = 2^3$ codewords. Rewriting the defining equation as $x_1 + x_2 + x_3 = x_4$ shows that x_4 can be viewed as a check digit, since it is uniquely determined by x_1, x_2, x_3, which can be arbitrarily given. Here, the codewords are the 4-bit strings with an even number of 1's. A set of basic codewords is $\{1001, 0101, 0011\}$, although there are also other choices. (How many?) ⊠

6.8.2 Generating matrices

A *generating matrix* for a linear code C is a matrix of the form $M = (I_m \mid A)$ whose row space is C. Notice that a generating matrix is in reduced row echelon form. We know from Proposition 3.12 that the reduced row echelon form of a matrix is unique, so the generating matrix for a linear code is unique.

Example 6.44. Let $p = 2$. If

$$M = \begin{pmatrix} 1 & 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 \end{pmatrix},$$

then $C = \text{row}(M)$ has 8 elements. Besides the rows of M and the null word, the elements of C are

$$110010,\ 101100,\ 011110,\ 111001.$$

⊠

Codes defined by an $m \times n$ generating matrix have the following property: every element of $C = \text{row}(M)$ can be expressed as a matrix product of the form $(x_1 \ \dots \ x_m)M$. (To see this, transpose the fact that the column space of M^T consists of all vectors of the form $M^T(x_1 \ \dots \ x_m)^T$.) Thus, to every $\mathbf{x} = (x_1 \dots x_m) \in \mathbb{F}^m$, there corresponds a unique codeword $\mathbf{c}(\mathbf{x}) = (x_1 \ \dots \ x_m)M \in C$. For a generating matrix M as above,

$$\mathbf{c}(\mathbf{x}) = x_1 \dots x_m \sum_{i=1}^{m} a_{i1}x_i \cdots \sum_{i=1}^{m} a_{i(n-m)}x_i \in \mathbb{F}^n.$$

Since $x_1, \dots, x_m \in \mathbb{F}$ are arbitrary, the first m entries $x_1 \dots x_m$ are called the *message digits*, and the last $n - m$ digits are called the *check digits*.

6.8.3 Hamming distance

Hamming distance is a natural distance function on $V(n, p)$ meant to solve the following problem. If $\mathbf{c}$ is a codeword for a code $C \subset V(n, p)$, and another string $\mathbf{c}'$ that differs from $\mathbf{c}$ only by a transposition is received during a transmission, can one determine whether $\mathbf{c}'$ is the result of an error in sending or receiving $\mathbf{c}$? Before making the definition, let us first define the weight $\omega(\mathbf{v})$ of an arbitrary $\mathbf{v} \in V(n, p)$.

Definition 6.22. Suppose $\mathbf{v} = v_1 \dots v_n \in V(n,p)$. Define the *weight* $\omega(\mathbf{v})$ of $\mathbf{v}$ to be the number of nonzero components of $\mathbf{v}$. That is,

$$\omega(\mathbf{v}) = |\{i \mid v_i \neq 0\}|.$$

The *Hamming distance* $d(\mathbf{u},\mathbf{v})$ between any pair $\mathbf{u},\mathbf{v} \in V(n,p)$ is defined as

$$d(\mathbf{u},\mathbf{v}) = \omega(\mathbf{u}-\mathbf{v}).$$

For example, $\omega(1010111) = 5$. The only vector of weight zero is the zero vector. Therefore, $\omega(\mathbf{u}-\mathbf{v}) = 0$ exactly when $\mathbf{u} = \mathbf{v}$. The reason $d(\mathbf{u},\mathbf{v})$ is called the distance is due to the following result.

Proposition 6.43. *Suppose* $\mathbf{u},\mathbf{v},\mathbf{w} \in V(n,p)$. *Then:*

(*i*) $d(\mathbf{u},\mathbf{v}) \geq 0$, *and* $d(\mathbf{u},\mathbf{v}) = 0$ *if and only if* $\mathbf{u} \neq \mathbf{v}$;

(*ii*) $d(\mathbf{u},\mathbf{v}) = d(\mathbf{v},\mathbf{u})$; *and*

(*iii*) $d(\mathbf{u},\mathbf{w}) \leq d(\mathbf{u},\mathbf{v}) + d(\mathbf{v},\mathbf{w})$.

Property (*iii*) is called the *triangle inequality*. It says that the length of one side of a triangle cannot exceed the sum of the lengths of the other two sides. In general, if S is any set, then a function $d : S \times S \to \mathbb{R}$ satisfying (*i*), (*ii*), and (*iii*) is called a *metric* on S, and $d(s,t)$ is defined to be the distance between $s,t \in S$. The notion of a metric is a natural generalization of the metric on a finite-dimensional inner product space. See Exercise 6.6.19 for this.

Proof. Properties (*i*) and (*ii*) are clear, but the triangle inequality requires proof. For the triangle inequality, first consider the case that $\mathbf{u}$ and $\mathbf{v}$ differ in every component. Thus $d(\mathbf{u},\mathbf{v}) = n$. Let $\mathbf{w}$ be any vector in $V(n,p)$, and suppose $d(\mathbf{u},\mathbf{w}) = k$. Then $\mathbf{u}$ and $\mathbf{w}$ agree in $n-k$ components, which tells us that $\mathbf{v}$ and $\mathbf{w}$ cannot agree in those $n-k$ components, so $d(\mathbf{v},\mathbf{w}) \geq n-k$. Thus

$$d(\mathbf{u},\mathbf{v}) = n = k + (n-k) \leq d(\mathbf{u},\mathbf{w}) + d(\mathbf{v},\mathbf{w}).$$

In the general case, let $\mathbf{u},\mathbf{v},\mathbf{w}$ be given, and let $\mathbf{u}',\mathbf{v}'$ and $\mathbf{w}'$ denote the strings obtained by dropping the components where $\mathbf{u}$ and $\mathbf{v}$ agree. Thus we are in the previous case, so

$$d(\mathbf{u},\mathbf{v}) = d(\mathbf{u}',\mathbf{v}') \leq d(\mathbf{u}',\mathbf{w}') + d(\mathbf{u},\mathbf{w}').$$

But $d(\mathbf{u}',\mathbf{w}') \leq d(\mathbf{u},\mathbf{w})$ and $d(\mathbf{v}',\mathbf{w}') \leq d(\mathbf{v},\mathbf{w})$, since dropping components decreases the Hamming distance. Therefore,

$$d(\mathbf{u}, \mathbf{v}) \leq d(\mathbf{u}, \mathbf{w}) + d(\mathbf{v}, \mathbf{w}),$$

and the triangle inequality is established. □

For each $C \subset V(n, p)$, let $d(C)$ denote the minimum value of $d(\mathbf{u}, \mathbf{v})$ as $\mathbf{u}, \mathbf{v}$ vary over C. As we will see below, one wants to maximize $d(C)$ for a given value of $|C|$. When $C \subset V(n, p)$ is linear, then $d(C)$ is the minimum of the weights of all the nonzero codewords. That is,

$$d(C) = \min\{\omega(\mathbf{c}) \mid \mathbf{c} \in C,\ \mathbf{c} \neq \mathbf{0}\}.$$

(The proof is an exercise.) This demonstrates one of the nice properties of linear codes: the minimum distance $d(C)$ requires only $|C|$ computations, which is considerably fewer than the number needed for an arbitrary code.

6.8.4 Error-correcting codes

In the terminology of coding theory, a code $C \subset V(n, p)$ such that $|C| = M$ and $d(C) = d$ is known as a p-ary (n, M, d)-code. As mentioned above, the game is to make the minimal distance $d(C)$ as large as possible for a given M. The reason for this is the next result.

Proposition 6.44. *An (n, M, d)-code C detects up to $d - 1$ errors and corrects up to $e = (d - 1)/2$ errors. That is, if $\mathbf{c} \in C$ and $d(\mathbf{v}, \mathbf{c}) \leq d - 1$, then either $\mathbf{v} = \mathbf{c}$ or $\mathbf{v} \notin C$. Moreover, if $\mathbf{v}$ is not a codeword, then there exists at most one codeword $\mathbf{c}$ such that $d(\mathbf{v}, \mathbf{c}) \leq e$.*

Thus, if $\mathbf{v} \notin C$, but $d(\mathbf{v}, \mathbf{c}) \leq e$, then we say that $\mathbf{c}$ is c error-correcting for $\mathbf{v}$. The conclusion about error-correction implies that if all but e digits of a codeword $\mathbf{c}$ are known, then every digit of $\mathbf{c}$ is known. Note that if $d(C) \geq 3$, then two codewords cannot differ by a transposition.

Example 6.45. Suppose C is a 6-bit code with $d = 3$. Then $e = 1$. If $\mathbf{c} = 100110$ is a codeword, then $\mathbf{v} = 010110$, which differs from $\mathbf{c}$ by a transposition, cannot be a codeword. Also, $\mathbf{w} = 000110$ can't be in C, since $d(\mathbf{c}, \mathbf{w}) = 1$, but $\mathbf{c}$ is the unique codeword within Hamming distance 1 of $\mathbf{w}$. If $\mathbf{x} = 000010$, then $d(\mathbf{c}, \mathbf{x}) = 2$, so there could be other codewords $\mathbf{c}'$ such that $d(\mathbf{c}', \mathbf{x}) = 2$. Note, however, that the triangle inequality says that $d(\mathbf{c}, \mathbf{c}') \leq d(\mathbf{c}, \mathbf{x}) + d(\mathbf{x}, \mathbf{c}') = 4$, so if $d(C) = 5$, then in fact, $\mathbf{c} = \mathbf{c}'$. ⊠

Let us now prove the proposition.

Proof. We will leave the first assertion as an exercise. So assume $d(\mathbf{v}, \mathbf{c}) \leq (d-1)/2$, and suppose for some $\mathbf{c}' \in C$ that we have $d(\mathbf{v}, \mathbf{c}') \leq d(\mathbf{v}, \mathbf{c})$. Then $d(\mathbf{v}, \mathbf{c}') \leq (d-1)/2$ too. By the triangle inequality,

$$d(\mathbf{c}, \mathbf{c}') \leq d(\mathbf{c}, \mathbf{v}) + d(\mathbf{v}, \mathbf{c}') \leq (d-1)/2 + (d-1)/2 = d - 1,$$

so indeed $\mathbf{c} = \mathbf{c}'$. □

Example 6.46. For the binary 4-bit code of Example 6.43 given by $x_1 + x_2 + x_3 + x_3 + x_4 = 0$, one can check that $d(C) = 2$. Thus C detects a single error, but $e = 1/2$, so there is no error correction. However, some additional information, such as the component where an error occurs, may allow error correction. Here, the linear equation defining the code enables that to be possible.

Designing codes that maximize $d(C)$ given $|C|$ is a basic problem in coding theory. The next example is a binary $(n, M, d) = (8, 16, 4)$ linear code C that maximizes $d(C)$ for $M = 16$. This code is sometimes denoted by C_8 and called the *extended Hamming code*.

Example 6.47. Let

$$A = \begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 & 1 \end{pmatrix}.$$

Then C_8 is defined to be $\mathrm{row}(A)$. Since every row of A has weight 4, the minimum distance $d(C_8)$ is at most 4. But every sum of the rows of A can be seen to have weight at most 4, so $d(C_8) = 4$. Since A is a generating matrix, $M = |C_8| = 2^4 = 16$. ⊠

We will now show that $d(C_8)$ is maximal for $M = 16$ and $n = 8$.

Proposition 6.45. *The code C_8 with 16 codewords maximizes $|C|$ among all 8-bit binary linear codes with $d(C) \geq 4$.*

Proof. Since $\dim C_8 = 4$, we have to show that there are no 8-bit binary linear codes C with $d(C) \geq 4$ and $\dim C \geq 5$, hence 32 codewords. This turns out to be a routine argument involving row reduction. Suppose C is such a code. By taking a spanning set for C as the rows of a $k \times 8$ matrix A, we can use row operations to put A into reduced row echelon form A_{red} without affecting C. Note that $k \geq 5$. By reordering the columns, we can suppose that A_{red} is a generating matrix $(I_r \mid M)$, where $r \geq 5$. Hence M has at most three columns. But the requirement $d(C) \geq 4$ implies that all entries of M are 1. This shows that there must exist codewords of weight two, a contradiction. Thus $\dim C < 5$, so $|C| = 16$ is the maximum. □

6.8.5 Cosets and perfect codes

A code $C \subset V(n,p)$ with minimum distance $d(C) = d \geq 3$ and $e = (d-1)/2 \geq 1$ is called *perfect* if every $\mathbf{x} \in V(n,p)$ is within e of some codeword $\mathbf{c}$. Since Proposition 6.44 says that every $\mathbf{x} \in V(n,p)$ is within distance e of at most one codeword, every element in $V(n,p)$ is within e of exactly one codeword. It is convenient to state this condition in a geometric way by bringing in the notion of a ball. Assume $r > 0$. The *ball of radius r centered at $\mathbf{v} \in V(n,p)$* is defined to be

$$B_r(\mathbf{v}) = \{\mathbf{w} \in V(n,p) \mid d(\mathbf{w},\mathbf{v}) \leq r\}. \tag{6.31}$$

Thus, a code $C \subset V(n,p)$ with $d(C) \geq 3$ and $e \geq 1$ is perfect if and only if $V(n,p)$ is the disjoint union of the balls $B_e(\mathbf{c})$ as $\mathbf{c}$ varies over C. That is,

$$V(n,p) = \bigcup_{\mathbf{c} \in C} B_e(\mathbf{c}) \quad \text{(disjoint union)}. \tag{6.32}$$

Example 6.48. Consider the binary linear code $C = \{000, 111\}$. Note that $d = 3$, so $e = 1$. Now

$$V(3,2) = \{000, 100, 010, 001, 110, 101, 011, 111\}.$$

The first four elements are within 1 of 000, and the last four are within 1 of 111. Therefore, C is perfect. ⊠

What makes perfect codes so nice is that there is a simple numerical criterion for deciding whether C is perfect. First note that for every $\mathbf{v} \in V(n,p)$, we have $|B_e(\mathbf{v})| = |B_e(\mathbf{0})|$. Hence we have the following.

Proposition 6.46. *A code $C \subset V(n,p)$ is perfect if and only if*

$$|C||B_e(\mathbf{0})| = p^n.$$

Thus, if C is linear of dimension k, then C is perfect if and only if $|B_e(\mathbf{0})| = p^{n-k}$.

One can check this criterion in the previous example, since

$$B_e(\mathbf{0}) = \{000, 100, 010, 001\},$$

while $|C| = 2$. Notice that a necessary condition for a binary code C of length n with $e = 1$ to be perfect is that $n + 1 = 2^m$ for some m, since $B_1(\mathbf{0}) = \{\mathbf{0}, \mathbf{e}_1, \ldots, \mathbf{e}_n\}$. Thus perfect codes with $e = 1$ must have length 3, 7, 15, and so on. Here is an example with $n = 7$.

Example 6.49. The linear code $C_7 \subset V(7,2)$ with generating matrix

$$A = \begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 \end{pmatrix}$$

is perfect. Indeed, enumerating the 16 elements of C_7, one sees that $d(C_7) = 3$, so $e = 1$. Clearly, $|B_e(\mathbf{0})| = 2^3$, while $|C_7| = 2^4$. Thus the criterion for perfection, $|C_7||B_e(\mathbf{0})| = 2^7$, holds. ⊠

There is also a connection between cosets and perfect linear codes. Recall that a linear code $C \subset V(n,p)$ of dimension k has exactly p^{n-k} cosets. Thus, C is perfect exactly when the number of cosets of C is $|B_e(\mathbf{0})|$. Furthermore, we have the following proposition.

Proposition 6.47. *A linear code $C \subset V(n,p)$ is perfect if and only if every coset $\mathbf{x} + C$ of C meets $B_e(\mathbf{0})$.*

Proof. Assume $\dim(C) = k$ and suppose that every coset $\mathbf{x}+C$ meets $B_e(\mathbf{0})$. We claim that $\mathbf{x} + C$ cannot contain more than one element of $B_e(\mathbf{0})$. For if $\mathbf{x} + \mathbf{c}$ and $\mathbf{x} + \mathbf{c}'$ lie in $B_e(\mathbf{0})$, then

$$d(\mathbf{x} + \mathbf{c}, \mathbf{x} + \mathbf{c}') = d(\mathbf{c}, \mathbf{c}') \leq d(\mathbf{c}, \mathbf{0}) + d(\mathbf{0}, \mathbf{c}') \leq 2e = d - 1,$$

so $\mathbf{c} = \mathbf{c}'$, since $d(C) = d$. Thus, $p^{n-k} \leq |B_e(\mathbf{0})|$. But since $V(n,p)$ is the union of the cosets of C, every element of $B_e(\mathbf{0})$ lies in a coset. Thus $|B_e(\mathbf{0})| = p^{n-k}$, so C is perfect. On the other hand, suppose C is perfect and consider a coset $\mathbf{x}+C$. By definition, $\mathbf{x}+C$ meets some ball $B_e(\mathbf{c})$, where $\mathbf{c} \in C$. Hence there exists $\mathbf{c}' \in C$ such that $d(\mathbf{x} + \mathbf{c}', \mathbf{c}) \leq e$. But $d(\mathbf{x} + \mathbf{c}', \mathbf{c}) = d(\mathbf{x} + (\mathbf{c}' - \mathbf{c}), \mathbf{0})$, so $\mathbf{x} + C$ meets $B_e(\mathbf{0})$, since C is linear. □

6.8.6 The hat problem

The solution of the hat problem is an example of a surprising application of mathematics, in this case coding theory. Consider the following problem: Three people are each wearing either a white hat or a black hat. Each player can see the other two hats but not their own. Although the players are not allowed to communicate with each other, they are allowed to discuss before getting hats what strategy they would use. Each person has a buzzer with three buttons marked B, W, and A (for black, white, or abstain). At the same time, each person presses B, W, or A according to whether they want to guess their hat color or abstain from guessing. If at least one player guesses their color correctly, and nobody guesses incorrectly, they win a huge prize.

A pretty good strategy would be to agree that two players abstain and the third makes a random guess. The probability of winning is 0.5. But this strategy doesn't make any use of fact that each player can see the other two hats. Instead, consider the following strategy. Suppose the players agree that they will play under the assumption that the array is not either BBB or WWW. Why does this help? First of all, there are eight possible arrays, so if the players can find an algorithm to guarantee that they avoid the set $F = \{\text{BBB}, \text{WWW}\}$ and do not make a mistake, then they will have a probability of $6/8 = 0.75$ of winning. Now let's analyze what happens if the hat array is BWB, for example. The first player sees WB, and knows that both BWB and WWB lie outside F but can't take the chance of an incorrect guess. So the first player must hit A. The same is true of the third player. The second player is the key. That player sees BB, so is forced to press W or otherwise land in F. If the array is outside of F, this strategy is guaranteed to produce a victory. The question is why, and the answer is that the 3-bit code $C_2 = \{000, 111\}$ in $V(3, 2)$ is a perfect code with $e = 1$.

Let us now see whether this strategy can be extended to seven players using the perfect linear code C_7 in the previous example. The players agree in the strategy session to proceed as if the hat array is not in C_7. Since $|C_7| = 2^4$, the probability that the hat array is in C_7 is $2^4/2^7 = 1/8$, so the probability of this being a winning assumption is $7/8$. They all need to memorize the 16 codewords of C_7. Suppose their assumption is correct. Then in order to win, they proceed as follows. Since the hat array $x_1 \dots x_7$ differs in exactly one place from a codeword $c_1 \dots c_7$, let us suppose that the discrepancy occurs at x_1. Then player #1 sees $c_2 \dots c_7$ and must make the choice $c_1 + 1$. Player #2 sees $x_1 c_3 \dots c_7$. But since $d(C_7) = 3$, she knows that whatever x_2 is, $x_1 x_2 c_3 \dots c_7 \notin C$. Therefore, in order to not make a mistake, she must abstain, as do the other five players. Assuming that the hat array $x_1 \dots x_7$ is not in C_7, they have won the game. The odds that they win the million bucks are a pretty good $7/8$.

Can you devise a strategy for how to proceed if there are four, five, or six players? Since there are no perfect codes in $V(n, 2)$ for $n = 4, 5, 6$, it isn't clear how to proceed. More information about this problem and related questions can be found in the article "The Hat Problem and Hamming Codes," by M. Bernstein, in *Focus Magazine*, November 2001.

Exercises

Exercise 6.8.1. List all binary linear codes C of length 3 with four codewords.

Exercise 6.8.2. Find a formula for the number of linear codes $C \subset V(n, p)$ of dimension k. (Suggestion: count the number of linearly independent subsets of $V(n, p)$ having k elements.)

Exercise 6.8.3. The international standard book number (ISBN) is a linear code $C \subset V(10, 11)$ that consists of all solutions $a_1 a_2 \cdots a_9 a_{10}$ of the equation

$$a_1 + 2a_2 + 3a_3 + \cdots + 10a_{10} = 0.$$

The digits are hyphenated to indicate the book's language and publisher. So, for example, the ISBN of *Fermat's Enigma*, by Simon Singh, published by Penguin Books (14) in 1997, is 1-14-026869-3. The actual ISBNs in use satisfy the condition $0 \leq a_i \leq 9$ for all $i \leq 9$, while a_{10} is also allowed to take the value 10, which is denoted by the Roman numeral X.

(i) How many ISBNs are in use?

(ii) Determine all x such that 0-13-832x4-4 is an ISBN.

(iii) Determine all x and y such that both 1-2-3832xy4-4 and 3-33-x2y377-6 are ISBNs.

Exercise 6.8.4. Show that if C is a linear code, then

$$d(C) = \min\{\omega(\mathbf{x}) \mid \mathbf{x} \in C,\ \mathbf{x} \neq \mathbf{0}\}.$$

Use the result to find $d(C)$ for the code C used to define ISBNs. Is this code error-correcting?

Exercise 6.8.5. Taking $\mathbb{F} = \mathbb{F}_{11}$, compute the generating matrix for the ISBN code.

Exercise 6.8.6. Consider the binary code $C \subset V(6, 2)$ that consists of 000000 and the following nonzero codewords:

$$100111,\ 010101,\ 001011,\ 110010,\ 101100,\ 011110,\ 111001.$$

(i) Determine whether C is linear.

(ii) Compute $d(C)$.

(iii) How many elements of C are nearest to 011111?

(iv) Determine whether 111111 is a codeword. If not, is there a codeword nearest 111111?

Exercise 6.8.7. Prove the first part of Proposition 6.44.

Exercise 6.8.8. Consider the binary code C_7 defined as the row space of the matrix

$$A = \begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 \end{pmatrix}.$$

in $V(7,2)$.

(i) Compute $d(C)$ and e.

(ii) Find the unique element of C that is nearest to 1010010. Do the same for 1110001.

Exercise 6.8.9. Let r be a positive integer and let $\mathbf{x} \in V(n,2)$. Consider the ball $B_r(\mathbf{x})$ of radius r about $\mathbf{x}$, i.e., $B_r(\mathbf{x}) = \{\mathbf{y} \in V(n,2) \mid d(\mathbf{x},\mathbf{y}) \leq r\}$. Show that

$$|B_r(\mathbf{x})| = \sum_{i=0}^{r} \binom{n}{i}.$$

Exercise 6.8.10. Generalize Exercise 6.8.9 from $V(n,2)$ to $V(n,p)$.

Exercise 6.8.11. * Show that if $C \subset V(n,2)$ is a linear code such that $\dim(C) = k$ and C is e-error-correcting, then

$$\sum_{i=0}^{e} \binom{n}{i} \leq 2^{(n-k)}.$$

Conclude that if $e = 1$, then $1 + n \leq 2^{n-k}$.

Exercise 6.8.12. Suppose $C \subset V(n,2)$ is a linear code with $\dim C = k$ and $d \geq 3$. Prove that C is perfect if and only if

$$\sum_{i=0}^{e} \binom{n}{i} = 2^{(n-k)}. \tag{6.33}$$

In particular, if $e = 1$, then C is perfect if and only if

$$(1+n)2^k = 2^n. \tag{6.34}$$

Exercise 6.8.13. Consider the code $C = \{00000, 11111\} \subset V(5,2)$.

(i) Determine e.

(ii) Show that C is perfect.

Exercise 6.8.14. Show that every binary $[2^k - 1, 2^k - 1 - k]$-code with $d = 3$ is perfect. Notice that C_7 is of this type.

Exercise 6.8.15. Can there exist a perfect code with $n = 5$ and $e = 2$?

Exercise 6.8.16. Suppose $n \equiv 2 \ mod(4)$. Show that there cannot be a perfect binary $[n,k]$-code with $e = 2$.

(iii) Does C present any possibilities for a five-player hat game?

Exercise 6.8.17. Show that every binary $[23, 12]$-code with $d = 7$ is perfect.

Chapter 7
Linear Mappings

The purpose of this chapter is to introduce linear mappings. Let V and W be vector spaces over a field $\mathbb{F}$. A linear mapping is a mapping $T : V \to W$ with domain V and target W that preserves linear combinations. The basic situation we will consider is that both V and W are finite-dimensional. Here we already know quite a bit, since an $m \times n$ matrix A over $\mathbb{F}$ defines a linear mapping $T_A : \mathbb{F}^n \to \mathbb{F}^m$ by putting $T_A(\mathbf{x}) = A\mathbf{x}$ (see Section 3.1.7). The basic rules of matrix algebra tell us that T_A preserves linear combinations. Thus, linear mappings are a generalization of matrix theory, and many ideas from matrix theory, such as the rank of a matrix A, its null space $\mathcal{N}(A)$, and the column space $\mathrm{col}(A)$, have natural interpretations in terms of linear mappings, as we shall see.

7.1 Definitions and Examples

In this section, we will define linear mappings and introduce several terms that we will use to explain the basic theory of linear mappings. We will also give several examples to illustrate some of the interesting types of linear mappings.

7.1.1 Mappings

Recall from Chap. 1 that if X and Y are sets, then a mapping $F : X \to Y$ is a rule that assigns to each element of the domain X a unique element $F(x)$ in the target Y. The image of F is the subset of the target $F(X) = \{y \in Y \mid y = F(x)\ \exists\ x \in X\}$. A mapping $F : \mathbb{F}^n \to \mathbb{F}^m$ is completely determined by

J.B. Carrell, *Groups, Matrices, and Vector Spaces*,
DOI 10.1007/978-0-387-79428-0_7

its component functions $f_1, f_2, \ldots, f_m$, which are obtained by writing

$$F(\mathbf{x}) = \sum_{i=1}^{m} f_i(\mathbf{x})\mathbf{e}_i = \begin{pmatrix} f_1(\mathbf{x}) \\ \vdots \\ f_m(\mathbf{x}) \end{pmatrix},$$

where $\mathbf{e}_1, \ldots, \mathbf{e}_m$ is the standard basis of $\mathbb{F}^m$. If V and W are arbitrary vector spaces over a field $\mathbb{F}$ and W is finite-dimensional, we can also define component functions $h_1, \ldots, h_m$ of F with respect to an arbitrary basis $\mathbf{w}_1, \ldots, \mathbf{w}_m$ of W in the same way by writing

$$F(\mathbf{v}) = \sum_{i=1}^{m} h_i(\mathbf{v})\mathbf{w}_i.$$

The component functions with respect to a basis are uniquely determined by F.

7.1.2 The definition of a linear mapping

In linear algebra, the most important mappings are those that preserve linear combinations. These are called *linear mappings*.

Definition 7.1. Suppose V and W are vector spaces over a field $\mathbb{F}$. Then a mapping $T : V \to W$ is said to be *linear* if

(i) for all $\mathbf{x}, \mathbf{y} \in V$, $T(\mathbf{x} + \mathbf{y}) = T(\mathbf{x}) + T(\mathbf{y})$, and

(ii) for all $r \in \mathbb{F}$ and all $\mathbf{x} \in V$, $T(r\mathbf{x}) = rT(\mathbf{x})$.

If $W = \mathbb{F}$, then T is called a *linear function*.

By definition, a linear mapping T preserves linear combinations: for all $r, s \in \mathbb{F}$ and all $\mathbf{x}, \mathbf{y} \in V$

$$T(r\mathbf{x} + s\mathbf{y}) = rT(\mathbf{x}) + sT(\mathbf{y}).$$

Thus a linear mapping also preserves linear combinations of an arbitrary number of vectors. Conversely, every mapping that preserves linear combinations is a linear mapping.

7.1.3 Examples

We now present some basic examples of linear mappings. The reader should note that some of the examples don't require a particular basis or choice of coordinates in their definition.

Example 7.1 (Identity mapping)**.** The mapping

$$I_V : V \to V$$

defined by $I_V(\mathbf{x}) = \mathbf{x}$ is called the *identity mapping*. This mapping is clearly linear. If $V = \mathbb{F}^n$, then $I_V = T_{I_n}$. ⊠

Example 7.2. If $\mathbf{a} \in \mathbb{R}^n$, the dot product with $\mathbf{a}$ defines a linear function $S_\mathbf{a} : \mathbb{R}^n \to \mathbb{R}$ by $S_\mathbf{a}(\mathbf{x}) = \mathbf{a} \cdot \mathbf{x} = \mathbf{a}^T\mathbf{x}$. It turns out that every linear function $S : \mathbb{R}^n \to \mathbb{R}$ has the form $S_\mathbf{a}$ for some $\mathbf{a} \in \mathbb{R}^n$. For example, let $\mathbf{a}^T = (1\ 2\ 0\ 1)$. Then the linear function $S_\mathbf{a} : \mathbb{R}^4 \to \mathbb{R}$ has the explicit form

$$S_\mathbf{a}(\mathbf{x}) = (1\ 2\ 0\ 1) \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = x_1 + 2x_2 + x_4.$$

We will consider the analogue of this function for an arbitrary inner product space in the next example. ⊠

Example 7.3. More generally, suppose V is an inner product space of dimension m. Then for every linear mapping $T : V \to \mathbb{R}$, there exists a unique $\mathbf{w} \in V$ such that $T(\mathbf{v}) = (\mathbf{v}, \mathbf{w})$. In fact, if $\mathbf{u}_1, \ldots, \mathbf{u}_m$ constitute an orthonormal basis of V, then I claim that

$$\mathbf{w} = \sum_{i=1}^{m} T(\mathbf{u}_i)\mathbf{u}_i.$$

The reader should check that $T(\mathbf{v}) = (\mathbf{v}, \mathbf{w})$ does in fact hold for all $\mathbf{v} \in V$ and that this choice of $\mathbf{w}$ is unique. ⊠

Example 7.4 (Diagonal mappings)**.** Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be the mapping

$$T \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} \mu_1 x_1 \\ \mu_2 x_2 \end{pmatrix}, \tag{7.1}$$

where μ_1 and μ_2 are real scalars. We leave it as an exercise to show that T is linear. In (7.1), $T(\mathbf{e}_1) = \mu_1\mathbf{e}_1$ and $T(\mathbf{e}_2) = \mu_2\mathbf{e}_2$. If both μ_1 and μ_2 are nonzero, the image under T of a rectangle with sides parallel to $\mathbf{e}_1$ and $\mathbf{e}_2$ is a parallel rectangle whose sides have been dilated by μ_1 and μ_2 and whose area has been changed by the factor $|\mu_1\mu_2|$. If $\mu_1\mu_2 \neq 0$, then T maps a circle about the origin to an ellipse about the origin. For example, the image of the unit circle $x^2 + y^2 = 1$ is the ellipse

$$(\frac{w_1}{\mu_1})^2 + (\frac{w_2}{\mu_2})^2 = 1,$$

as can be seen by putting $w_1 = \mu_1 x_1$ and $w_2 = \mu_2 x_2$.

⊠

In general, a linear mapping $T : V \to V$ is called *semisimple* if there exist a basis $\mathbf{v}_1, \ldots, \mathbf{v}_n$ of V and scalars $\mu_1, \ldots, \mu_n$ in $\mathbb{F}$ such that $T(\mathbf{v}_i) = \mu_i \mathbf{v}_i$ for all i. Note that we are not requiring that any or all μ_i be nonzero; but if all $\mu_i = 0$, then T is the zero linear mapping. The basis $\mathbf{v}_1, \ldots, \mathbf{v}_n$ of V for which $T(\mathbf{v}_i) = \mu_i \mathbf{v}_i$ is called an eigenbasis, and the scalars $\mu_1, \ldots, \mu_n$ are called the eigenvalues of T. The existence of an eigenbasis for T says a great deal about how T acts. The question of when a linear mapping T admits an eigenbasis is very basic. It will be solved when we study the classification of linear mappings. The solution is nontrivial, and not all linear mappings are semisimple.

Example 7.5. Recall from Example 6.17 that the *cross product* of two vectors $\mathbf{a}, \mathbf{b} \in \mathbb{R}^3$ is defined as

$$\mathbf{a} \times \mathbf{b} = (a_2 b_3 - a_3 b_2, -(a_1 b_3 - a_3 b_1), a_1 b_2 - a_2 b_1)^T. \tag{7.2}$$

The cross product defines a linear mapping $C_{\mathbf{a}} : \mathbb{R}^3 \to \mathbb{R}^3$ by

$$C_{\mathbf{a}}(\mathbf{v}) = \mathbf{a} \times \mathbf{v}.$$

Notice that $C_{\mathbf{a}}(\mathbf{a}) = \mathbf{0}$. A basic property of $C_{\mathbf{a}}$ is that $C_{\mathbf{a}}(\mathbf{x})$ is orthogonal to both $\mathbf{a}$ and $\mathbf{x}$. It follows that $C_{\mathbf{a}}$ cannot be semisimple unless $\mathbf{a} = \mathbf{0}$ (see Example 7.4). ⊠

7.1.4 Matrix linear mappings

Recall from Section 3.1.7 that if $A \in \mathbb{F}^{m \times n}$, then the mapping $T_A : \mathbb{F}^n \to \mathbb{F}^m$ defined by $T_A(\mathbf{x}) = A\mathbf{x}$ is called a *matrix linear mapping*. Every matrix linear mapping is certainly linear. We now show that every linear mapping $T : \mathbb{F}^n \to \mathbb{F}^m$ is a matrix linear mapping.

Proposition 7.1. *Every linear mapping $T : \mathbb{F}^n \to \mathbb{F}^m$ is of the form T_A for a unique $A \in \mathbb{F}^{m \times n}$. In fact,*

$$A = \big(T(\mathbf{e}_1)\ T(\mathbf{e}_2)\ \cdots\ T(\mathbf{e}_n)\big).$$

Conversely, if $A \in \mathbb{F}^{m \times n}$, then T_A is a linear mapping with domain $\mathbb{F}^n$ and target $\mathbb{F}^m$.

Proof. Since $\mathbf{x} = \sum_{i=1}^n x_i \mathbf{e}_i$,

$$T(\mathbf{x}) = T\big(\sum_{i=1}^n x_i \mathbf{e}_i\big) = \sum_{i=1}^n x_i T(\mathbf{e}_i) = \big(T(\mathbf{e}_1)\ T(\mathbf{e}_2)\ \cdots\ T(\mathbf{e}_n)\big)\mathbf{x}.$$

Thus, $T = T_A$, where $A = \big(T(\mathbf{e}_1)\ T(\mathbf{e}_2)\ \cdots\ T(\mathbf{e}_n)\big)$. Furthermore, A is uniquely determined by T, since if two linear mappings $S, T : \mathbb{F}^n \to \mathbb{F}^m$ have the property that $S(\mathbf{e}_i) = T(\mathbf{e}_i)$ for $i = 1, \ldots, n$, then $S(\mathbf{x}) = T(\mathbf{x})$ for all $\mathbf{x} \in \mathbb{F}^n$. The claim that every T_A is linear has already been shown. □

In particular, a linear function $T : \mathbb{F}^n \to \mathbb{F}$ is given by a $1 \times n$ matrix. Thus, there exists a unique $\mathbf{a} \in \mathbb{F}^n$ such that $T(\mathbf{x}) = \mathbf{a}^T\mathbf{x}$. Hence there exist unique scalars $a_1, a_2, \ldots, a_n \in \mathbb{F}$ such that $T(\mathbf{x}) = \sum_{i=1}^n a_i x_i$. When $\mathbb{F} = \mathbb{R}$, we may express this fact in terms of the dot product as $T(\mathbf{x}) = \mathbf{a} \cdot \mathbf{x}$ in Example 7.2.

Example 7.6. For example, the matrix of the identity mapping $I_{\mathbb{F}^n}$ on $\mathbb{F}^n$ is the identity matrix I_n. That is, $I_{\mathbb{F}^n} = T_{I_n}$. We will sometimes use I_n to denote $I_{\mathbb{F}^n}$, provided no confusion is possible. ⊠

Recall from Section 3.1.7 that if $S : \mathbb{F}^p \to \mathbb{F}^n$ and $T : \mathbb{F}^n \to \mathbb{F}^m$ are linear mappings with matrices $S = T_A$ and $T = T_B$ respectively, then the composition $T \circ S : \mathbb{R}^p \to \mathbb{R}^m$ is the matrix linear mapping associated to BA. That is,

$$T_B \circ T_A = T_{BA}.$$

Writing M_T for the matrix of T etc., we therefore have the identity

$$M_{T \circ S} = M_T M_S.$$

We will not repeat the proof, but it is short. Somewhat surprisingly, the key fact is that matrix multiplication is associative.

7.1.5 An Application: rotations of the plane

The relationship between composition and matrix multiplication can be applied to the group Rot(2) of rotations of $\mathbb{R}^2$ to give an extremely pretty and simple proof of the sum formulas for the trigonometric functions *sine* and *cosine*. Recall that $\mathcal{R}_\theta : \mathbb{R}^2 \to \mathbb{R}^2$ is the rotation of $\mathbb{R}^2$ about the origin through θ. Computing the images of $\mathcal{R}(\mathbf{e}_1)$ and $\mathcal{R}(\mathbf{e}_2)$, we have

$$\mathcal{R}_\theta(\mathbf{e}_1) = \cos\theta\mathbf{e}_1 + \sin\theta\mathbf{e}_2$$

and

$$\mathcal{R}_\theta(\mathbf{e}_2) = -\sin\theta\mathbf{e}_1 + \cos\theta\mathbf{e}_2.$$

I claim that rotations are linear. This can be seen as follows. Suppose $\mathbf{x}$ and $\mathbf{y}$ are any two noncollinear vectors in $\mathbb{R}^2$, and let P be the parallelogram they span. Then $\mathcal{R}_\theta$ rotates the whole parallelogram P about $\mathbf{0}$ to a new

parallelogram $\mathcal{R}_\theta(P)$ with edges $\mathcal{R}_\theta(\mathbf{x})$ and $\mathcal{R}_\theta(\mathbf{y})$ at $\mathbf{0}$. Since the diagonal $\mathbf{x}+\mathbf{y}$ of P is rotated to the diagonal of $\mathcal{R}_\theta(P)$, it follows that

$$\mathcal{R}_\theta(\mathbf{x}+\mathbf{y}) = \mathcal{R}_\theta(\mathbf{x}) + \mathcal{R}_\theta(\mathbf{y}).$$

Similarly, for every scalar r,

$$\mathcal{R}_\theta(r\mathbf{x}) = r\mathcal{R}_\theta(\mathbf{x}).$$

Therefore, $\mathcal{R}_\theta$ is linear, as claimed. The matrix R_θ of $\mathcal{R}_\theta$, calculated via the above formula, is $R_\theta = (\mathcal{R}_\theta(\mathbf{e}_1) \;\; \mathcal{R}_\theta(\mathbf{e}_2))$. Hence

$$R_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}. \tag{7.3}$$

(This verifies the formula of Example 4.5.) Thus

$$\mathcal{R}_\theta \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x\cos\theta - y\sin\theta \\ x\sin\theta + y\cos\theta \end{pmatrix}.$$

Let us now apply this result to trigonometry. If one first applies a rotation $\mathcal{R}_\psi$ and follows that by the rotation $\mathcal{R}_\theta$, the outcome is the rotation $\mathcal{R}_{\theta+\psi}$. Hence,

$$\mathcal{R}_{\theta+\psi} = \mathcal{R}_\theta \circ \mathcal{R}_\psi = \mathcal{R}_\psi \circ \mathcal{R}_\theta.$$

Therefore, since composition of linear mappings corresponds to multiplication of their matrices, $R_{\theta+\psi} = R_\theta R_\psi = R_\psi R_\theta$. In particular,

$$\begin{pmatrix} \cos(\theta+\psi) & -\sin(\theta+\psi) \\ \sin(\theta+\psi) & \cos(\theta+\psi) \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} \cos\psi & -\sin\psi \\ \sin\psi & \cos\psi \end{pmatrix}.$$

Expanding the product and comparing both sides gives us the trigonometric formulas for the sine and cosine of $\theta+\psi$:

$$\cos(\theta+\psi) = \cos\theta\cos\psi - \sin\theta\sin\psi$$

and

$$\sin(\theta+\psi) = \sin\theta\cos\psi + \cos\theta\sin\psi.$$

Exercises

Exercise 7.1.1. Let X and Y be sets and $\phi : X \to Y$ a mapping. Recall from Chap. 1 that ϕ is injective if and only if for $x \in X$, $\phi(x) = \phi(x')$ implies $x = x'$, ϕ is surjective if and only if $\phi(X) = Y$, and ϕ is a bijection if and only if it is both injective and surjective. Show the following:

(*i*) ϕ is injective if and only if there exists a mapping $\psi : F(X) \to X$ such that $\psi \circ \phi$ is the identity mapping $I_X : X \to X$.

(*ii*) ϕ is surjective if and only if there exists a mapping $\psi : Y \to X$ such that $\phi \circ \psi$ is the identity mapping I_Y, and

(*iii*) ϕ is a bijection if and only if there exists a mapping $\psi : Y \to X$ such that $\psi \circ \phi = I_X$ and $\phi \circ \psi = I_Y$.

Exercise 7.1.2. Show that every linear function $T : \mathbb{R} \to \mathbb{R}$ has the form $T(x) = ax$ for some $a \in \mathbb{R}$.

Exercise 7.1.3. Determine whether any of the following functions $f : \mathbb{R}^2 \to \mathbb{R}$ are linear:

(*i*) $f(x, y) = xy$,

(*ii*) $f(x, y) = x - y$,

(*iii*) $f(x, y) = e^{x+y}$.

Exercise 7.1.4. Suppose $T : \mathbb{F}^n \to \mathbb{F}^m$ is an arbitrary mapping and write

$$T(\mathbf{v}) = (f_1(\mathbf{v}), f_2(\mathbf{v}), \ldots, f_m(\mathbf{v}))^T.$$

Show that T is linear if and only if each component function f_i is a linear function.

Exercise 7.1.5. Find the matrix of the following linear mappings:

(*i*) $S(x_1, x_2, x_3) = (2x_1 - 3x_3, x_1 + x_2 - x_3, x_1, x_2 - x_3)^T$.

(*ii*) $T(x_1, x_2, x_3, x_4) = (x_1 - x_2 + x_3 + x_4, x_2 + 2x_3 - 3x_4)^T$.

(*iii*) $T \circ S$.

Exercise 7.1.6. Let V be a vector space over $\mathbb{F}$, and let W be a subspace of V. Let $\pi : V \to V/W$ be the quotient map defined by $\pi(\mathbf{v}) = \mathbf{v} + W$. Show that π is linear.

Exercise 7.1.7. Let $S : U \to V$ and $T : V \to W$ be linear mappings. Show that $T \circ S$ is also linear.

Exercise 7.1.8. Suppose A is a real $m \times n$ matrix. Show that when we view both $\text{row}(A)$ and $\mathcal{N}(A)$ as subspaces of $\mathbb{R}^n$,

$$\text{row}(A) \cap \mathcal{N}(A) = \{\mathbf{0}\}.$$

Is this true for matrices over other fields, for example $\mathbb{F}_p$ or $\mathbb{C}$?

Exercise 7.1.9. Let $V = \mathbb{C}$, and consider the mapping $S : V \to V$ defined by $S(z) = \alpha z$, where $\alpha \in \mathbb{C}$.

(*i*) Describe S as a mapping $S : \mathbb{R}^2 \to \mathbb{R}^2$.

(*ii*) Is S linear over $\mathbb{R}$? If so, find its matrix in $\mathbb{R}^{2\times 2}$.

Exercise 7.1.10. Show that every mapping $S : \mathbb{C} \to \mathbb{C}$ that is linear over $\mathbb{C}$ has the form $S(z) = \alpha z$ for a unique $A \in \mathbb{C}$.

Exercise 7.1.11. Let $T_A : \mathbb{R}^2 \to \mathbb{R}^2$ be the matrix linear mapping associated to the matrix $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. The purpose of this exercise is to determine when T is linear over $\mathbb{C}$. That is, since by definition, $\mathbb{C} = \mathbb{R}^2$ (with complex multiplication), we may ask when $T_A(\alpha\beta) = \alpha T_A(\beta)$ for all $\alpha, \beta \in \mathbb{C}$. Show that a necessary and sufficient condition for T_A to be $\mathbb{C}$-linear is that $a = d$ and $b = -c$.

Exercise 7.1.12. Show that every rotation $\mathcal{R}_\theta$ defines a $\mathbb{C}$-linear map $\mathcal{R}_\theta : \mathbb{C} \to \mathbb{C}$. Relate this map to the complex exponential $e^{i\theta}$.

Exercise 7.1.13. Let $C_\infty(\mathbb{R})$ be the space of infinitely differentiable functions on the real line $\mathbb{R}$. A function $f \in C_\infty(\mathbb{R})$ is said to be *even* if $f(-x) = f(x)$ for all $x \in \mathbb{R}$ and *odd* if $f(-x) = -f(x)$ for all $x \in \mathbb{R}$. Let $C_\infty(\mathbb{R})^{ev}$ and $C_\infty(\mathbb{R})^{odd}$ denote the set of even and odd functions in $C_\infty(\mathbb{R})$ respectively.

(*i*) Show that $C_\infty(\mathbb{R})^{ev}$ and $C_\infty(\mathbb{R})^{odd}$ are subspaces of $C_\infty(\mathbb{R})$.

(*ii*) Show that the mapping $D : C_\infty(\mathbb{R}) \to C_\infty(\mathbb{R})$ defined by $D(f) = f'$ sends $C_\infty(\mathbb{R})^{ev}$ to $C_\infty(\mathbb{R})^{odd}$.

Exercise 7.1.14. Let $\mathbb{F}$ be a Galois field, and let p be the characteristic of $\mathbb{F}$. Let $T : \mathbb{F} \to \mathbb{F}$ be the mapping defined by $T(x) = x^p$. Recall that the set of multiples $m1$ of 1 in $\mathbb{F}$, where $m = 0, 1, \ldots, p-1$, forms a subfield $\mathbb{F}' = \mathbb{F}_p$ of $\mathbb{F}$ and that $\mathbb{F}$ is a vector space over $\mathbb{F}'$. Show that T is a linear mapping of $\mathbb{F}$ with respect to this vector space structure. The linear mapping T is called the *Frobenius map*.

Exercise 7.1.15. As in Exercise 7.1.14, let $\mathbb{F}$ be a Galois field of characteristic p, and let $V = \mathbb{F}^{n\times n}$. Show that the mapping $F : V \to V$ defined by $F(A) = A^p$ is a linear mapping.

7.2 Theorems on Linear Mappings

In this section, we will define some more terms and prove several results about linear mappings, including a result that will generalize the rank–nullity identity $\dim \mathcal{N}(A) + \dim \operatorname{col}(A) = n$ derived in Example 6.28.

7.2.1 The kernel and image of a linear mapping

The kernel and image are two natural subspaces associated with a linear mapping. Let $T : V \to W$ be linear. The image of T has been defined in Chap. 1. It will be denoted by $\operatorname{im}(T)$. The other natural subspace associated with T is its kernel. First note that every linear mapping T maps $\mathbf{0}$ to $\mathbf{0}$, since

$$T(\mathbf{0}) = T(0\mathbf{0}) = 0T(\mathbf{0}) = \mathbf{0}.$$

Definition 7.2. The *kernel* of T is defined to be the set

$$\ker(T) = \{\mathbf{v} \in V \mid T(\mathbf{v}) = \mathbf{0}\}.$$

Since a linear mapping is a group homomorphism, the notion of the kernel of a linear mapping is a special case of the notion of the kernel of a homomorphism. Suppose $V = \mathbb{F}^n$, $W = \mathbb{F}^m$, and $T = T_A$, i.e., $T(\mathbf{x}) = A\mathbf{x}$. Then $\ker(T_A)$ and $\operatorname{im}(T_A)$ are related to linear systems. In fact, $\ker(T_A) = \mathcal{N}(A)$, while $\operatorname{im}(T_A)$ is the set of vectors $\mathbf{b} \in \mathbb{F}^m$ for which $A\mathbf{x} = \mathbf{b}$ has a solution. Thus, $\operatorname{im}(T_A) = \operatorname{col}(A)$, so both $\ker(T_A)$ and $\operatorname{im}(T_A)$ are subspaces. This is more generally true for arbitrary linear mappings.

Proposition 7.2. *The kernel and image of a linear mapping $T : V \to W$ are subspaces of V and W respectively.*

Proof. We leave this as an exercise. □

The following result gives a very useful characterization of injective (equivalently, one-to-one) linear mappings. The proof is almost word for word the proof given in Chap. 2 for the group-theoretic analogue.

Proposition 7.3. *A linear mapping T is injective if and only if* $\ker(T) = \{\mathbf{0}\}$.

Proof. Suppose $\ker(T) = \{\mathbf{0}\}$ and $T(\mathbf{x}) = T(\mathbf{y})$. Then $T(\mathbf{x} - \mathbf{y}) = \mathbf{0}$, so $\mathbf{x} - \mathbf{y} \in \ker(T)$. But this says that $\mathbf{x} - \mathbf{y} = \mathbf{0}$, so T is injective. Conversely, if T is injective and $\mathbf{x} \in \ker(T)$, then $T(\mathbf{x}) = \mathbf{0} = T(\mathbf{0})$, so $\mathbf{x} = \mathbf{0}$. Thus $\ker(T) = \{\mathbf{0}\}$. □

7.2.2 The Rank–Nullity Theorem

Let $A \in \mathbb{F}^{m\times n}$ and recall the rank–nullity identity $\dim \mathcal{N}(A)+\dim \operatorname{col}(A) = n$ (see (6.7)). The general rank–nullity theorem generalizes this identity to a linear mapping $T : V \to W$, where V is finite-dimensional.

Theorem 7.4 (Rank–Nullity Theorem)**.** *Let V and W be vector spaces over $\mathbb{F}$, and suppose $\dim V$ is finite. Then for every linear mapping $T : V \to W$,*

$$\dim \ker(T) + \dim \operatorname{im}(T) = \dim V. \tag{7.4}$$

Proof. If $\ker(T) = V$, then $\operatorname{im}(T) = \{\mathbf{0}\}$, so there is nothing to prove. On the other hand, if $\dim \ker(T) = 0$, then T is injective. Thus if $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ is a basis of V, it follows that $T(\mathbf{v}_1), \ldots, T(\mathbf{v}_n)$ is a basis of $\operatorname{im}(T)$. To verify this, it suffices to show that the $T(\mathbf{v}_i)$ are independent. But if $\sum a_i T(\mathbf{v}_i) = \mathbf{0}$, then $T(\sum a_i \mathbf{v}_i) = \mathbf{0}$, so $\sum a_i \mathbf{v}_i = \mathbf{0}$, since T is injective. Hence all a_i are equal to zero, since $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ are independent. Thus, $\dim \operatorname{im}(T) = \dim V$. Now suppose $\dim \ker(T) = k > 0$. By the dimension theorem (Theorem 6.11), we may choose a basis $\mathbf{v}_1, \mathbf{v}_2, \ldots \mathbf{v}_k$ of $\ker(T)$ and extend it to a basis $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ of V. Let $\mathbf{w}_i = T(\mathbf{v}_i)$. I claim that $\mathbf{w}_{k+1}, \ldots, \mathbf{w}_n$ are a basis of $\operatorname{im}(T)$. To see that they span, let $\mathbf{w} \in \operatorname{im}(T)$, say $\mathbf{w} = T(\mathbf{v})$. Write $\mathbf{v} = \sum a_i \mathbf{v}_i$. Then

$$\mathbf{w} = T(\mathbf{v}) = \sum_{i=1}^{n} a_i T(\mathbf{v}_i) = \sum_{i=k+1}^{n} a_i \mathbf{w}_i.$$

The proof that $\mathbf{w}_{k+1}, \ldots, \mathbf{w}_n$ are independent is identical to the proof in the case $\dim \ker(T) = 0$, so we will leave it to the reader. Hence $\dim \operatorname{im}(T) = n - k$, which proves the result. □

7.2.3 An existence theorem

The rank–nullity theorem tells us something about the behavior of a given linear mapping, but we do not yet know how to construct linear mappings. The following existence result will show that there is considerable flexibility in defining a linear mapping $T : V \to W$, provided we have a basis of V (for example, if V is finite-dimensional). We will show that the values of T on a basis of V can be arbitrarily described.

Proposition 7.5. *Let V and W be vector spaces over $\mathbb{F}$, and let $\mathbf{v}_1, \ldots, \mathbf{v}_n$ be a basis of V. Suppose $\mathbf{w}_1, \ldots, \mathbf{w}_n$ are arbitrary vectors in W. Then there exists a unique linear mapping $T : V \to W$ such that $T(\mathbf{v}_i) = \mathbf{w}_i$ for each i. In other words, a linear mapping is uniquely determined by giving its values on a basis.*

Proof. The proof is surprisingly simple. Since every $\mathbf{x} \in V$ has a unique expression

$$\mathbf{x} = \sum_{i=1}^{n} r_i \mathbf{v}_i,$$

we obtain a mapping $T : V \to W$ by setting

$$T(\mathbf{x}) = \sum_{i=1}^{n} r_i \mathbf{w}_i.$$

In fact, T is linear. Indeed, if $\mathbf{y} \in V$, say

$$\mathbf{y} = \sum_{i=1}^{n} s_i \mathbf{v}_i,$$

then $\mathbf{x} + \mathbf{y} = \sum (r_i + s_i)\mathbf{v}_i$, so

$$T(\mathbf{x} + \mathbf{y}) = \sum_{i=1}^{n} (r_i + s_i)\mathbf{w}_i = \sum_{i=1}^{n} r_i \mathbf{w}_i + \sum_{i=1}^{n} s_i \mathbf{w}_i = T(\mathbf{x}) + T(\mathbf{y}).$$

Similarly, $T(r\mathbf{v}) = rT(\mathbf{v})$. Moreover, T is unique, since every linear mapping is determined on a basis. □

If $V = \mathbb{F}^n$ and $W = \mathbb{F}^m$, there is an even simpler proof by appealing to matrix theory. Let $B = (\mathbf{v}_1\ \mathbf{v}_2\ \ldots \mathbf{v}_n)$ and $C = (\mathbf{w}_1\ \mathbf{w}_2\ \ldots \mathbf{w}_n)$. Then the matrix A of T satisfies $AB = C$. But B is invertible, since $\mathbf{v}_1, \ldots, \mathbf{v}_n$ is a basis of $\mathbb{F}^n$, so $A = CB^{-1}$.

7.2.4 Vector space isomorphisms

We will now answer the following question. When are two vector spaces indistinguishable as far as their algebraic properties are concerned? The answer is given by the notion of isomorphism.

Definition 7.3. We will say that two vector spaces V and W over the same field $\mathbb{F}$ are *isomorphic* if there exists a bijective linear mapping $T : V \to W$. Such a linear mapping T is called an *isomorphism*.

In other words, an isomorphism is a linear mapping that is both injective and surjective. A vector space isomorphism is also a group isomorphism. The point of the definition is that although two vector spaces V and W may look quite different, isomorphic vector spaces are indistinguishable if one considers only their internal algebraic properties (addition, scalar multiplication,

etc.). For example, one space V may be the solution space of a homogeneous linear differential equation with real coefficients, and the other, W, may be a far less exotic vector space such as $\mathbb{R}^n$. The question is, when are V and W isomorphic? The quite simple answer, provided by the next result, is an application of the dimension theorem and the fact that linear mappings are determined by assigning arbitrary values to a basis.

Proposition 7.6. *Two finite-dimensional vector spaces V and W over the same field are isomorphic if and only if they have the same dimension.*

Proof. Suppose $\dim V = \dim W$. To construct an isomorphism $T : V \to W$, choose bases $\mathbf{v}_1, \ldots, \mathbf{v}_n$ of V and $\mathbf{w}_1, \ldots, \mathbf{w}_n$ of W, and let $T : V \to W$ be the unique linear mapping (guaranteed by Proposition 7.5) such that $T(\mathbf{v}_i) = \mathbf{w}_i$ if $1 \leq i \leq n$. Since $\operatorname{im}(T)$ is a subspace of W containing a basis of W, it follows that T is surjective. By the rank–nullity theorem, $\dim \ker(T) = 0$; hence T is injective by Proposition 7.3. The converse follows from the rank–nullity theorem (Theorem 7.4) by similar reasoning. □

The set of isomorphisms $T : V \to V$ in fact forms a group, called the *general linear group* of V, which is denoted by $GL(V)$. If $V = \mathbb{F}^n$, then in fact, $GL(V) = GL(n, \mathbb{F})$.

Exercises

Exercise 7.2.1. Let the field be $\mathbb{F}_2$ and consider the matrix

$$A = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 & 0 \end{pmatrix}.$$

(*i*) Find a basis of $\operatorname{im}(A)$.

(*ii*) How many elements are in $\operatorname{im}(A)$?

(*iii*) Is $(01111)^T$ in $\operatorname{im}(A)$?

(*iv*) Without any further computation, find a basis of $\operatorname{im}(A^T)$.

Exercise 7.2.2. Let A be a real 3×3 matrix such that the first row of A is a linear combination of A's second and third rows.

(*i*) Show that $\mathcal{N}(A)$ is either a line through the origin or a plane containing the origin.

(*ii*) Show that if the second and third rows of A span a plane P, then $\mathcal{N}(A)$ is the line through the origin orthogonal to P.

Exercise 7.2.3. Prove Proposition 7.2 using only the basic definition of a linear mapping. That is, show that the kernel and image of a linear mapping T are subspaces of the domain and target of T respectively.

Exercise 7.2.4. Consider the mapping $C_{\mathbf{a}} : \mathbb{R}^3 \to \mathbb{R}^3$ defined by

$$C_{\mathbf{a}}(\mathbf{v}) = \mathbf{a} \times \mathbf{v},$$

where $\mathbf{a} \times \mathbf{v}$ is the cross product of $\mathbf{a}$ and $\mathbf{v}$.

(*i*) Show that $C_{\mathbf{a}}$ is linear and find its matrix.

(*ii*) Describe the kernel and image of $C_{\mathbf{a}}$.

Exercise 7.2.5. Suppose $\mathbf{a} \in \mathbb{R}^3$. Find the kernel of the linear mapping $C_{\mathbf{a}} + I_3$.

Exercise 7.2.6. Suppose $T : \mathbb{R}^2 \to \mathbb{R}^2$ is a linear mapping that sends pair of noncollinear vectors to noncollinear vectors. Suppose $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{R}^2$ are noncollinear. Show that T sends every parallelogram with sides parallel to $\mathbf{x}$ and $\mathbf{y}$ to another parallelogram with sides parallel to $T(\mathbf{x})$ and $T(\mathbf{y})$.

Exercise 7.2.7. Find the kernel and image of the linear mapping $S_{\mathbf{a}} : \mathbb{R}^n \to \mathbb{R}$ defined by $S_{\mathbf{a}}(\mathbf{x}) = \mathbf{a} \cdot \mathbf{x}$.

Exercise 7.2.8. Determine $\ker(C_{\mathbf{a}}) \cap \operatorname{im}(C_{\mathbf{a}})$ for the cross product linear mapping $C_{\mathbf{a}} : \mathbb{R}^3 \to \mathbb{R}^3$ for every nonzero $\mathbf{a}$.

Exercise 7.2.9. Suppose V is a finite-dimensional vector space and $T : V \to V$ is a linear mapping such that $T \circ T = O$. Show that $\operatorname{im}(T) \subset \ker(T)$. Is the converse true?

Exercise 7.2.10. Suppose V is a finite-dimensional vector space and $T : V \to V$ is a linear mapping such that $\operatorname{im}(T) \subset \ker(T)$. Show that $\dim V$ is an even integer.

Exercise 7.2.11. Suppose A is a symmetric real $n \times n$ matrix.

(*i*) Show that $\operatorname{col}(A) \cap \mathcal{N}(A) = \{\mathbf{0}\}$.

(*ii*) Conclude that $\mathbb{R}^n = \mathcal{N}(A) \oplus \operatorname{col}(A)$.

(*iii*) Suppose $A^2 = O$. Show that $A = O$.

Exercise 7.2.12. Find a nonzero 2×2 symmetric matrix A over $\mathbb{C}$ such that $A^2 = O$.

Exercise 7.2.13. Show that if V is a finite-dimensional vector space and T is a linear mapping with domain V, then $\dim T(V) \leq \dim V$.

Exercise 7.2.14. Suppose A is a square matrix over an arbitrary field. Show that if $A^k = O$ for some positive integer k, then $\dim \mathcal{N}(A) > 0$.

Exercise 7.2.15. Recall that if $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, then $\mathbf{x} \cdot \mathbf{y} = \mathbf{x}^T \mathbf{y}$. Use this to prove that for every $A \in \mathbb{R}^{n \times n}$, $A^T A$ and A have the same null space. Conclude that $A^T A$ and A have the same rank.

Exercise 7.2.16. This exercise deals with the *inverse* of a linear mapping T. Let V be a vector space over $\mathbb{F}$, and let $T : V \to V$ be a linear mapping such that $\ker(T) = \mathbf{0}$ and $\mathrm{im}(T) = V$. That is, T is an isomorphism. Prove the following statements.

(*i*) There exists a linear mapping $S : V \to V$ with the property that $S(\mathbf{y}) = \mathbf{x}$ if and only if $T(\mathbf{x}) = \mathbf{y}$. Note: S is called the *inverse* of T.

(*ii*) Show that S is an isomorphism, and $S \circ T = T \circ S = I_V$.

(*iii*) If $V = \mathbb{F}^n$, A is the matrix of T, and B is the matrix of S, then

$$BA = AB = I_n.$$

Exercise 7.2.17. Let $S : \mathbb{R}^n \to \mathbb{R}^m$ and $T : \mathbb{R}^m \to \mathbb{R}^p$ be two linear mappings both of which are injective. Show that the composition $T \circ S$ is also injective. Conclude that if A is of size $m \times n$ and has $\mathcal{N}(A) = \{\mathbf{0}\}$, and B is of size $n \times p$ and has $\mathcal{N}(B) = \{\mathbf{0}\}$, then $\mathcal{N}(BA) = \{\mathbf{0}\}$ too.

Exercise 7.2.18. Suppose V and W are vector spaces of the same dimension over a field $\mathbb{F}$, and let $T : V \to W$ be a linear mapping. Show that if T is either injective or surjective, then T is an isomorphism.

Exercise 7.2.19. Let W be a subspace of a finite-dimensional vector space V. Show there exists a linear mapping $T : V \to V$ such that $\ker(T) = W$.

Exercise 7.2.20. Suppose $T : V \to W$ is linear. Show that there exists a unique linear mapping $\overline{T} : V/\ker(T) \to W$ such that $\overline{T}(\mathbf{v} + \ker(T)) = T(\mathbf{v})$.

Exercise 7.2.21. Let U, V, and W be finite-dimensional vector spaces over the same field $\mathbb{F}$, and let $S : U \to V$ and $T : V \to W$ be linear. Show that:

(*i*) TS is injective if and only if S is injective and $\mathrm{im}(S) \cap \ker(T) = \{\mathbf{0}\}$.

(*ii*) TS is surjective if and only if T is surjective and $V = \mathrm{im}(S) + \ker(T)$.

(*iii*) Conclude that TS is an isomorphism if and only if S is injective, T is surjective, and $\dim U = \dim W$.

7.3 Isometries and Orthogonal Mappings

Throughout this section, V will denote a real finite-dimensional inner product space. Recall that the inner product on V defines a distance function $d(\mathbf{a}, \mathbf{b}) = |\mathbf{a} - \mathbf{b}|$. Linear mappings $T : V \to V$ that preserve distances are called *isometries*. (Actually, as we will see below, a distance-preserving map is automatically linear, so the definition can be simplified.) The isometries of V form an important group, called $O(V)$. When $V = \mathbb{R}^n$, $O(V)$ is the matrix group $O(n, \mathbb{R})$ of all orthogonal $n \times n$ matrices over $\mathbb{R}$. We will first consider isometries in general and then specialize to rotations and reflections. Finally, we will consider the isometries of $\mathbb{R}^2$. Here we prove the $O(2, \mathbb{R})$-dichotomy: every isometry of $\mathbb{R}^2$ is either a rotation or a reflection. We will also show that the dihedral group $D(m)$ can be realized as the group of all isometries of a regular m-gon in $\mathbb{R}^2$. This will give another proof that the order of $D(m)$ is $2m$.

7.3.1 Isometries and orthogonal linear mappings

A mapping $S : V \to V$ is said to be *orthogonal* if

$$(S(\mathbf{x}), S(\mathbf{y})) = (\mathbf{x}, \mathbf{y}) \tag{7.5}$$

for all $\mathbf{x}, \mathbf{y} \in V$. The term orthogonal comes from the fact that an orthogonal mapping preserves the orthogonal relationship between orthogonal pairs of vectors. Since orthogonal mappings preserve inner products, they preserve lengths of vectors, distances between vectors, and angles between vectors, since the angle θ between $\mathbf{v}$ and $\mathbf{w}$ is found by the identity

$$\mathbf{v} \cdot \mathbf{w} = |\mathbf{v}||\mathbf{w}| \cos\theta.$$

It turns out that all orthogonal mappings are linear.

Proposition 7.7. *Let $S : V \to V$ be orthogonal. Then S is linear. In fact, the orthogonal mappings are exactly the isometries. In particular, all distance-preserving mappings on V are linear. Conversely, every isometry is orthogonal.*

Proof. To show that S is linear, we first show that for all $\mathbf{a}, \mathbf{b} \in V$, $|S(\mathbf{a} + \mathbf{b}) - S(\mathbf{a}) - S(\mathbf{b})|^2 = 0$. But

$$|S(\mathbf{a}+\mathbf{b}) - S(\mathbf{a}) - S(\mathbf{b})|^2 = (S(\mathbf{a}+\mathbf{b}) - S(\mathbf{a}) - S(\mathbf{b}), S(\mathbf{a}+\mathbf{b}) - S(\mathbf{a}) - S(\mathbf{b}))$$

$$= (S(\mathbf{a}+\mathbf{b}), S(\mathbf{a}+\mathbf{b})) + (S(\mathbf{a}), S(\mathbf{a})) + (S(\mathbf{b}), S(\mathbf{b})) - 2(S(\mathbf{a}+\mathbf{b}), S(\mathbf{a}))$$

$$-2(S(\mathbf{a}+\mathbf{b}), S(\mathbf{b})) + 2(S(\mathbf{a}), S(\mathbf{b})).$$

Using the fact that S is orthogonal, it follows that

$$|S(\mathbf{a}+\mathbf{b}) - S(\mathbf{a}) - S(\mathbf{b})|^2 = (\mathbf{a}+\mathbf{b}, \mathbf{a}+\mathbf{b}) + (\mathbf{a}, \mathbf{a}) + (\mathbf{b}, \mathbf{b})$$

$$-2(\mathbf{a}+\mathbf{b}, \mathbf{a}) - 2(\mathbf{a}+\mathbf{b}, \mathbf{b}) + 2(\mathbf{a}, \mathbf{b}).$$

Expanding further, one sees that the right-hand side is zero. Thus, $S(\mathbf{a}+\mathbf{b}) = S(\mathbf{a}) + S(\mathbf{b})$. The proof that $S(r\mathbf{a}) = rS(\mathbf{a})$ for all $r \in \mathbb{F}$ is similar. Therefore, every orthogonal mapping is linear. To see that S is an isometry, note that

$$\begin{aligned} |S(\mathbf{b}) - S(\mathbf{a})|^2 &= |S(\mathbf{b}-\mathbf{a})|^2 \\ &= (S(\mathbf{a}-\mathbf{b}), S(\mathbf{a}-\mathbf{b})) \\ &= (\mathbf{b}-\mathbf{a}, \mathbf{b}-\mathbf{a}) \\ &= |\mathbf{b}-\mathbf{a}|^2. \end{aligned}$$

This completes the proof that orthogonal mappings are isometries. We will leave the rest of the proof as an exercise. □

More generally, every mapping $F : V \to V$ that preserves distances turns out to be orthogonal, and hence is an isometry (see Exercise 7.3.3).

7.3.2 Orthogonal linear mappings on $\mathbb{R}^n$

We are now going to show that the matrix of an isometry is orthogonal, and conversely, that every orthogonal matrix defines an isometry.

Proposition 7.8. *Every isometry $T : \mathbb{R}^n \to \mathbb{R}^n$ is the matrix linear mapping associated with a unique orthogonal matrix. Conversely, every orthogonal matrix defines a unique isometry.*

Proof. Suppose $T : \mathbb{R}^n \to \mathbb{R}^n$ is an isometry. Since T is linear, $T = T_Q$ for a unique $Q \in \mathbb{R}^{n\times n}$. But T is orthogonal, so $T_Q(\mathbf{x}) \cdot T_Q(\mathbf{y}) = Q\mathbf{x} \cdot Q\mathbf{y} = \mathbf{x} \cdot \mathbf{y}$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. Thus

$$(Q\mathbf{x})^T(Q\mathbf{y}) = (\mathbf{x}^T Q^T)(Q\mathbf{y}) = \mathbf{x}^T(Q^T Q)\mathbf{y} = \mathbf{x}^T\mathbf{y} \tag{7.6}$$

for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. Setting $Q\mathbf{e}_i = \mathbf{q}_i$, (7.6) implies $\mathbf{q}_i^T\mathbf{q}_j = \mathbf{e}_i^T\mathbf{e}_j$. Hence $Q^TQ = I_n$, so Q is orthogonal. We leave the converse to the reader. □

Recall that the set of orthogonal $n \times n$ matrices is the orthogonal group $O(n, \mathbb{R})$, so $O(n, \mathbb{R})$ is also the group of isometries of $\mathbb{R}^n$.

7.3.3 Projections

Let $\mathbf{a} \in \mathbb{R}^2$ be nonzero, and consider the line $\mathbb{R}\mathbf{a}$ spanned by $\mathbf{a}$. In Section 6.6.2, we called the mapping

$$P_{\mathbf{a}}(\mathbf{x}) = \left(\frac{\mathbf{a} \cdot \mathbf{x}}{\mathbf{a} \cdot \mathbf{a}}\right)\mathbf{a}$$

the *projection* onto $\mathbb{R}\mathbf{a}$. We leave it as an exercise to show that the projection $P_{\mathbf{a}}$ is linear. Note that if $\mathbf{u}$ is the unit vector determined by $\mathbf{a}$, then since $\mathbf{a} \cdot \mathbf{a} = |\mathbf{a}|^2$, it follows that

$$P_{\mathbf{u}}(\mathbf{x}) = P_{\mathbf{a}}(\mathbf{x}) = (\mathbf{u}^T\mathbf{x})\mathbf{u}.$$

Thus,

$$P_{\mathbf{u}}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = (u_1x_1 + u_2x_2)\begin{pmatrix} u_1 \\ u_2 \end{pmatrix},$$

so

$$P_{\mathbf{u}}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} u_1^2 & u_1u_2 \\ u_1u_2 & u_2^2 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}.$$

Hence the matrix of $P_{\mathbf{u}}$ is

$$\begin{pmatrix} u_1^2 & u_1u_2 \\ u_1u_2 & u_2^2 \end{pmatrix}. \tag{7.7}$$

Note that a projection matrix is symmetric. The image of the projection $P_{\mathbf{u}}$ is the line $\mathbb{R}\mathbf{u}$, while its kernel is the line orthogonal to $\mathbb{R}\mathbf{u}$. A reflection $P_{\mathbf{u}}$ has the property that $P_{\mathbf{u}} \circ P_{\mathbf{u}} = P_{\mathbf{u}}$, since

$$P_{\mathbf{u}} \circ P_{\mathbf{u}}(\mathbf{v}) = P_{\mathbf{u}}((\mathbf{v} \cdot \mathbf{u})\mathbf{u})) = (\mathbf{v} \cdot \mathbf{u})P_{\mathbf{u}}(\mathbf{u}) = (\mathbf{v} \cdot \mathbf{u})\mathbf{u} = P_{\mathbf{u}}(\mathbf{v}).$$

We will next apply projections to find a general formula for the reflection $\mathbb{R}^n$ through a hyperplane.

7.3.4 Reflections

We will first find an expression for a plane reflection and use that to suggest how to define a reflection of $\mathbb{R}^n$ through a hyperplane. Let ℓ be a line through the origin of $\mathbb{R}^2$. We want to consider how to reflect $\mathbb{R}^2$ through ℓ. The sought-after mapping $H : \mathbb{R}^2 \to \mathbb{R}^2$ fixes every vector on ℓ and sends every vector $\mathbf{v}$ orthogonal to ℓ to $-\mathbf{v}$. (If a reflection is linear, this is enough information to determine H.) Suppose $\mathbf{v} \in \mathbb{R}^2$ is on neither ℓ nor $\ell^\perp$. Then $\mathbf{v}$ and its reflection $H(\mathbf{v})$ form an isosceles triangle with equal sides $\mathbf{v}$ and $H(\mathbf{v})$, which H swaps.

Choosing a unit vector $\mathbf{u}$ on $\ell^\perp$, let us write

$$\mathbf{v} = P_{\mathbf{u}}(\mathbf{v}) + \mathbf{c},$$

where $\mathbf{c} \in \ell$. This is the orthogonal decomposition of $\mathbf{v}$ with respect to ℓ and $\ell^\perp$. Hence the component on ℓ is $\mathbf{c} = \mathbf{v} - P_{\mathbf{u}}(\mathbf{v})$. By Euclidean geometry, $H(\mathbf{v}) = -P_{\mathbf{u}}(\mathbf{v}) + \mathbf{c}$. Replacing $\mathbf{c}$ by $\mathbf{v} - P_{\mathbf{u}}(\mathbf{v})$, we get the formula

$$H(\mathbf{v}) = -P_{\mathbf{u}}(\mathbf{v}) + (\mathbf{v} - P_{\mathbf{u}}(\mathbf{v})) = \mathbf{v} - 2P_{\mathbf{u}}(\mathbf{v}).$$

Therefore,

$$H(\mathbf{v}) = \mathbf{v} - 2P_{\mathbf{u}}(\mathbf{v}) = \mathbf{v} - 2(\mathbf{u} \cdot \mathbf{v})\mathbf{u} = \mathbf{v} - 2(\mathbf{u}^T\mathbf{v})\mathbf{u}. \tag{7.8}$$

This expression immediately establishes the following result.

Proposition 7.9. *The reflection H of $\mathbb{R}^2$ through a line ℓ passing through $\mathbf{0}$ is a linear mapping.*

Example 7.7. Let us find the reflection H through the line ℓ given by $x = -y$. Now $\mathbf{u} = (\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})^T$ is a unit vector on $\ell^\perp$. Thus,

$$\begin{aligned} H\begin{pmatrix} a \\ b \end{pmatrix} &= \begin{pmatrix} a \\ b \end{pmatrix} - 2\left(\begin{pmatrix} a \\ b \end{pmatrix} \cdot \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}\right)\begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} \\ &= \begin{pmatrix} a - (a+b) \\ b - (a+b) \end{pmatrix} \\ &= \begin{pmatrix} -b \\ -a \end{pmatrix}. \end{aligned}$$

Thus the matrix of H is

$$\begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix}. \qquad \boxtimes$$

Now let Q denote the matrix of the reflection through the line orthogonal to a unit vector $\mathbf{u} \in \mathbb{R}^2$. Applying (7.7) and (7.8), one sees that

$$Q = \begin{pmatrix} 1 - 2u_1^2 & -2u_1u_2 \\ -2u_1u_2 & 1 - 2u_2^2 \end{pmatrix} = \begin{pmatrix} u_2^2 - u_1^2 & -2u_1u_2 \\ -2u_1u_2 & u_1^2 - u_2^2 \end{pmatrix}. \tag{7.9}$$

Thus, every 2×2 reflection matrix has the form

$$Q = \begin{pmatrix} a & b \\ b & -a \end{pmatrix}, \tag{7.10}$$

where $a^2 + b^2 = 1$. Conversely, as we will show below, a matrix of the form $Q = \begin{pmatrix} a & b \\ b & -a \end{pmatrix}$, where $a^2 + b^2 = 1$, is a reflection. Hence, the symmetric 2×2 orthogonal matrices are exactly the matrices of reflections.

Let us now ask what the linear mapping H defined by (7.8) does in the case of $\mathbb{R}^n$. The equation $\mathbf{v} \cdot \mathbf{u} = 0$ defines an $(n-1)$-dimensional subspace W of $\mathbb{R}^n$, namely the hyperplane through the origin orthogonal to $\mathbf{u}$. If $\mathbf{v} \in W$, then $H(\mathbf{v}) = \mathbf{v}$. On the other hand, if $\mathbf{v} = r\mathbf{u}$, then $H(\mathbf{v}) = rH(\mathbf{u}) = r(\mathbf{u} - 2\mathbf{u}) = -r\mathbf{u} = -\mathbf{v}$. In particular, H leaves the hyperplane W pointwise fixed and reverses vectors on the line $W^\perp$ orthogonal to this hyperplane.

Definition 7.4. Let $\mathbf{u} \in \mathbb{R}^n$ be a unit vector, and let W be the hyperplane in $\mathbb{R}^n$ orthogonal to the line $\mathbb{R}\mathbf{u}$. Then the linear mapping $H : \mathbb{R}^n \to \mathbb{R}^n$ defined by

$$H(\mathbf{v}) = \mathbf{v} - 2(\mathbf{v} \cdot \mathbf{u})\mathbf{u} = \mathbf{v} - 2(\mathbf{u}^T\mathbf{v})\mathbf{u} \tag{7.11}$$

is called the *reflection of* $\mathbb{R}^n$ *through* W.

Since reflecting $\mathbf{v} \in \mathbb{R}^n$ twice through a hyperplane W returns $\mathbf{v}$ to itself, a reflection H has the property that $H \circ H = I_{\mathbb{R}^n}$. This can be checked directly by matrix multiplication. Let P denote the matrix of the projection $P_{\mathbf{u}}$. Then the matrix Q of H is $Q = I_n - 2P$. Thus,

$$Q^2 = (I_n 2 - 2P)(I_n - 2P) = I_n - 4P + 4P^2 = I_n,$$

since $P^2 = P$.

Since Q is symmetric and $Q^2 = I_n$, Q is by definition orthogonal. Thus we get the following.

Proposition 7.10. *The matrix of a reflection is orthogonal. Hence, reflections are orthogonal linear mappings.*

7.3.5 Projections on a general subspace

As a final example of a linear mapping, let us work out the projection of finite-dimensional inner product space V onto an arbitrary subspace W. As we saw in Proposition 6.38, every $\mathbf{x} \in V$ admits a unique orthogonal decomposition

$$\mathbf{x} = \mathbf{w} + \mathbf{y},$$

where $\mathbf{w} \in W$ and $\mathbf{y} \in W^\perp$. The *projection of* V *onto* W is the mapping $P_W : V \to V$ defined by $P_W(\mathbf{x}) = \mathbf{w}$. The following proposition justifies calling P_W a projection.

Proposition 7.11. *The mapping $P_W : V \to V$ has the following properties:*

(i) P_W *is linear,*

(ii) $P_W(\mathbf{w}) = \mathbf{w}$ *if* $\mathbf{w} \in W$,

(iii) $P_W(W^\perp) = \{\mathbf{0}\}$, *and finally,*

(iv) $P_W + P_{W^\perp} = I$.

Proof. Since W is a finite-dimensional inner product space, we have shown that it has an orthonormal basis, say $\mathbf{u}_1, \ldots, \mathbf{u}_k$. Assume that $\mathbf{w} \in W$. By Proposition 6.36,

$$\begin{aligned}\mathbf{w} &= \sum_{i=1}^{k} (\mathbf{w}, \mathbf{u}_i)\mathbf{u}_i \\ &= \sum_{i=1}^{k} (\mathbf{x} - \mathbf{y}, \mathbf{u}_i)\mathbf{u}_i \\ &= \sum_{i=1}^{k} (\mathbf{x}, \mathbf{u}_i)\mathbf{u}_i - \sum_{i=1}^{k} (\mathbf{y}, \mathbf{u}_i)\mathbf{u}_i \\ &= \sum_{i=1}^{k} (\mathbf{x}, \mathbf{u}_i)\mathbf{u}_i.\end{aligned}$$

The last identity holds, since $\mathbf{y} \in W^\perp$ and all the $\mathbf{u}_i$ are in in W. Hence,

$$P_W(\mathbf{x}) = \sum_{i=1}^{k} (\mathbf{x}, \mathbf{u}_i)\mathbf{u}_i.$$

This shows that P_W is linear and $P_W(\mathbf{w}) = \mathbf{w}$ if $\mathbf{w} \in W$. Thus, *(i)* and *(ii)* hold. Statement *(iii)* follows from the fact that $(\mathbf{y}, \mathbf{u}_i) = 0$ for all $\mathbf{y} \in W^\perp$, and *(iv)* is a consequence of the decomposition $\mathbf{x} = \mathbf{w} + \mathbf{y}$, since $P_{W^\perp}(\mathbf{x}) = \mathbf{y}$. □

When $V = \mathbb{R}^n$, Exercise 7.3.9 below gives an interesting alternative expression for P_W that doesn't require knowing an orthonormal basis of W.

7.3.6 Dimension two and the $O(2, \mathbb{R})$-dichotomy

The orthogonal group $O(2, \mathbb{R})$ of all isometries of $\mathbb{R}^2$ contains the group $\mathrm{Rot}(2)$ of all rotations of $\mathbb{R}^2$ as a normal subgroup. Recall that every rotation matrix has the form $R_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$. We are now going to show that $O(2, \mathbb{R})$ has surprising decomposition.

Theorem 7.12 (The $O(2,\mathbb{R})$-dichotomy)**.** *Every 2×2 orthogonal matrix is either a rotation matrix $\begin{pmatrix} a & -b \\ b & a \end{pmatrix}$ or a reflection matrix $\begin{pmatrix} a & b \\ b & -a \end{pmatrix}$. Here $a^2+b^2=1$, so $a=\cos\theta$ and $b=\sin\theta$ for some θ. Consequently, every isometry of $\mathbb{R}^2$ is either a rotation or a reflection.*

Proof. We will first use the coset decomposition of $O(2,\mathbb{R})$. Note that by the product formula, $\det : O(2,\mathbb{R}) \to \mathbb{R}^*$ is a homomorphism. We have shown that $\det(Q) = \pm 1$ for all $Q \in O(2,\mathbb{R})$, since $Q^TQ = I_2$. But by Exercise 2.2.5, the value of a homomorphism on a coset of its kernel is constant and takes different values on different cosets. Let $Q \in O(2,\mathbb{R})$. Since the columns of Q are orthogonal unit vectors, we have

$$Q = \begin{pmatrix} a & c \\ b & d \end{pmatrix},$$

where $a^2+b^2=1$, $c^2+d^2=1$ and $ac+bd=0$. After some simplification, it follows that there are exactly two possibilities for Q:

$$Q_1 = \begin{pmatrix} a & -b \\ b & a \end{pmatrix} \quad \text{and} \quad Q_2 = \begin{pmatrix} a & b \\ b & -a \end{pmatrix}, \quad \text{where} \quad a^2+b^2=1.$$

Now, $\det(Q_1) = a^2+b^2 = 1$, while $\det(Q_2) = -a^2-b^2 = -1$. Thus Q_1 constitutes the kernel of det, so it follows that the kernel is $\mathrm{Rot}(2)$. Now every element of the form Q_2 can be written

$$Q_2 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}\begin{pmatrix} a & -b \\ b & a \end{pmatrix} = \begin{pmatrix} a & c \\ c & -a \end{pmatrix},$$

where $c=-b$. Thus we have to show that Q_2 is a reflection. We have

$$\begin{pmatrix} a & c \\ c & -a \end{pmatrix}\begin{pmatrix} -c \\ a+1 \end{pmatrix} = -\begin{pmatrix} -c \\ a+1 \end{pmatrix}$$

and

$$\begin{pmatrix} a & c \\ c & -a \end{pmatrix}\begin{pmatrix} a+1 \\ c \end{pmatrix} = \begin{pmatrix} a+1 \\ c \end{pmatrix}.$$

Put $\mathbf{v} = \begin{pmatrix} -c \\ a+1 \end{pmatrix}$ and $\mathbf{w} = \begin{pmatrix} a+1 \\ c \end{pmatrix}$, and note that $\mathbf{v}$ and $\mathbf{w}$ are orthogonal. Hence if $\mathbf{v}$ (and equivalently $\mathbf{w}$) is nonzero, there exists an orthonormal basis $\mathbf{u}_1 = \mathbf{v}/|\mathbf{v}|$ and $\mathbf{u}_2 = \mathbf{w}/|\mathbf{w}|$ of $\mathbb{R}^2$ such that $Q_2\mathbf{u}_1 = -\mathbf{u}_1$ and $Q_2\mathbf{u}_2 = \mathbf{u}_2$. If $\mathbf{v} = \mathbf{w} = \mathbf{0}$, then $a=-1$ and $c=0$. But in this case,

$$Q_2 = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}.$$

In either case, Q_2 is the matrix of a reflection. This shows that every element of $O(2,\mathbb{R})$ is either a rotation or a reflection. □

The $O(2,\mathbb{R})$-dichotomy rests on the fact that the kernel of the determinant homomorphism on $O(2,\mathbb{R})$ has two cosets: the identity coset Rot(2) and its complement, which is the coset consisting of reflection matrices. In general, the determinant determines a surjective homomorphism of $O(n,\mathbb{R}) \to \{\pm 1\}$ with kernel the group $O(n,\mathbb{R}) \cap SL(n,\mathbb{R})$. We will denote this group by $SO(n,\mathbb{R})$. In particular, $\mathrm{Rot}(2) = SO(2,\mathbb{R})$. (Also see Exercise 5.2.13.) Since $SO(n,\mathbb{R})$ is a normal subgroup of $O(n,\mathbb{R})$, we may apply the first isomorphism theorem (Theorem 2.20) to deduce that $O(n,\mathbb{R})/SO(n,\mathbb{R}) \cong \{\pm 1\}$. Thus, $SO(n,\mathbb{R})$ also has exactly two left cosets, and therefore

$$O(n,\mathbb{R}) = SO(n,\mathbb{R}) \cup QSO(n,\mathbb{R}), \tag{7.12}$$

where Q is any element of $O(n,\mathbb{R})$ such that $\det(Q) = -1$. However, it is not as easy to describe this coset as in the case $n = 2$.

Since the product of a rotation matrix R and a reflection matrix Q is a reflection matrix, one can ask how to determine the reflections RQ and QR. Similarly, if Q_1 and Q_2 are reflections, how does one describe the rotation Q_1Q_2? These questions are taken up in the exercises. There is another nice fact.

Proposition 7.13. *Every finite subgroup of $SO(2,\mathbb{R})$ is cyclic.*

Proof. In fact, $SO(2,\mathbb{R})$ is isomorphic to the circle group S^1 in $\mathbb{C}^*$ via the isomorphism $\varphi(R_\theta) = e^{i\theta}$ (see Section 4.2.2). The fact that φ is an isomorphism is due to the formula $R_\theta \begin{pmatrix} x \\ y \end{pmatrix} = e^{i\theta} z$, where $z = x + iy$. But we showed that every finite subgroup of $\mathbb{C}^*$ is cyclic in Proposition 2.26, so the same holds for $SO(2,\mathbb{R})$. □

7.3.7 The dihedral group as a subgroup of $O(2,\mathbb{R})$

Recall that the dihedral group $D(m)$ ($m \geq 1$) was originally defined by giving generators a and b satisfying the relations $a^m = b^2 = 1$ and $ab = ba^{-1}$. In this section we will show that $D(m)$ can be realized geometrically as a subgroup of $O(2,\mathbb{R})$. Consider the rotation matrix $a = R_{2\pi/m}$ and the reflection matrix $b = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$. Then $a^m = b^2 = I_2$. Moreover, it can be checked directly that $aba = b$, so $ab = ba^{-1}$. Thus the subgroup $\mathcal{D}(m)$ of $O(2,\mathbb{R})$ generated by a and b is a copy of $D(m)$.

Now suppose $m > 1$, and let $\{m\}$ denote an m-sided regular polygon in $\mathbb{R}^2$ centered at the origin having a vertex at $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$. For example, the vertices of

$\{m\}$ can be placed at the mth roots of unity $e^{2\pi i/m}$. Then $\{m\}$ is symmetric about the x-axis. If $m = 1$, we will assume $\{m\} = \{\begin{pmatrix}1\\0\end{pmatrix}\}$. Let $\mathcal{O}(\{m\}) \subset O(2,\mathbb{R})$ be the set of all orthogonal matrices that send $\{m\}$ onto itself.

Proposition 7.14. *Assume $m \geq 1$. Then $\mathcal{O}(\{m\}) = \mathcal{D}(m)$.*

Proof. We will leave it to the reader to verify that $\mathcal{O}(\{m\})$ is a subgroup of $O(2,\mathbb{R})$. By construction, it follows that $\{m\}$ is sent onto itself by both a and b. Thus $\mathcal{D}(m) \subset \mathcal{O}(\{m\})$. Since $|\mathcal{D}(m)| = 2m$, it suffices to show that $|\mathcal{O}(\{m\})| = 2m$ also. Since the vertices of $\{m\}$ are equidistant from the origin and elements of $\mathcal{O}(\{m\})$ preserve lengths, it follows that $\mathcal{O}(\{m\})$ has to permute the vertices of $\{m\}$. But an element of $\mathcal{O}(\{m\})$ is a linear mapping on $\mathbb{R}^2$, so it is determined by its values on two noncollinear vectors. Thus, $\mathcal{O}(\{m\})$ has to be finite. Let $\mathcal{O}_+ = \mathcal{O}(\{m\}) \cap SO(2,\mathbb{R})$. Being a finite subgroup of $SO(2,\mathbb{R})$, we know that $\mathcal{O}_+$ is cyclic. This implies that $|\mathcal{O}_+| = m$, since because $\{m\}$ has m vertices, no element of $\mathcal{O}_+$ can have order greater than m. But $a \in \mathcal{O}_+$, so $\mathcal{O}_+$ is the cyclic group generated by a. Since $b \in \mathcal{O}(\{m\})$, it follows that $\det : \mathcal{O}(\{m\}) \to \{\pm 1\}$ is surjective, so as above,

$$\mathcal{O}(\{m\}) = \mathcal{O}_+ \cup b\,\mathcal{O}_+.$$

Therefore, $|\mathcal{O}(\{m\})| = 2m$, as claimed. □

7.3.8 The finite subgroups of $O(2,\mathbb{R})$

We now classify the finite subgroups of $O(2,\mathbb{R})$. It turns out that there are no major surprises. Every finite subgroup of $O(2,\mathbb{R})$ is either cyclic or dihedral.

Theorem 7.15. *The only finite subgroups of $O(2,\mathbb{R})$ are:*

(i) the groups $\{I_2, b\}$, where b is a reflection;

(ii) the cyclic groups C_m consisting of rotations R_θ, where $\theta = 2k\pi/m$ with $0 \leq k \leq m-1$; and

(iii) the dihedral groups $\mathcal{D}(m)$ of symmetries of $\{m\}$, where $m > 1$.

Proof. We have already shown that every finite subgroup of $SO(2,\mathbb{R})$ is cyclic. Suppose G is a subgroup of $O(2,\mathbb{R})$ that is not contained in $SO(2,\mathbb{R})$. Then by the $O(2,\mathbb{R})$-dichotomy, G contains a reflection b. If $G \neq \{1, b\}$, then G also contains a rotation different from I_2. So let $G_+ = G \cap SO(2,\mathbb{R})$. Then $\det : G \to \{\pm 1\}$ is surjective, so as above, $G = G_+ \cup bG_+$. Thus, $G = \mathcal{D}(m)$, where $m = |G_+|$. □

This theorem (or rather its content) is attributed to Leonardo da Vinci. He was the first to explicitly list the symmetries of a regular polyhedron $\{m\}$. Leonardo's interest in symmetries involved architecture, in particular questions such as whether one can add structures at the corners of a building without destroying the rotational or reflective symmetry. Many interesting facts about symmetries can be found on the Internet as well as in a pair of classic books: *Symmetry*, by H. Weyl, and *Geometry*, by H.S.M. Coxeter. In the penultimate chapter, we will discuss the symmetry groups of the Platonic solids and classify all the finite subgroups of $SO(3,\mathbb{R})$. The classification is a considerably more difficult proof.

Exercises

Exercise 7.3.1. Suppose $\mathbf{a} \in \mathbb{R}^n$ is nonzero.

(*i*) Show that the projection

$$P_{\mathbf{a}}(\mathbf{x}) = \Big(\frac{\mathbf{a}\cdot\mathbf{x}}{\mathbf{a}\cdot\mathbf{a}}\Big)\mathbf{a}$$

onto the line $\mathbb{R}\mathbf{a}$ is linear.

(*ii*) Using the formula for $P_{\mathbf{a}}$, verify that $P_{\mathbf{a}}$ fixes every vector on $\mathbb{R}\mathbf{a}$ and sends every vector orthogonal to $\mathbf{a}$ to $\mathbf{0}$.

(*iii*) Verify that $P_{\mathbf{a}} \circ P_{\mathbf{a}} = P_{\mathbf{a}}$.

Exercise 7.3.2. Let $\mathbf{u}$ and $\mathbf{v}$ be an orthonormal basis of $\mathbb{R}^2$. Show directly that the following formulas hold for all $\mathbf{x} \in \mathbb{R}^2$:

(*i*) $P_{\mathbf{u}}(\mathbf{x}) + P_{\mathbf{v}}(\mathbf{x}) = \mathbf{x}$, and

(*ii*) $P_{\mathbf{u}}(P_{\mathbf{v}}(\mathbf{x})) = P_{\mathbf{v}}(P_{\mathbf{u}}(\mathbf{x})) = \mathbf{0}$.

Exercise 7.3.3. Let V be an inner product space, and suppose $S : V \to V$ preserves distances. Show that S is orthogonal, and conclude that $S : V \to V$ preserves distances if and only if S is an isometry.

Exercise 7.3.4. Find the matrix of each of the following linear mappings:

(*i*) the rotation $\mathcal{R}_{-\pi/4}$ of $\mathbb{R}^2$ through $-\pi/4$,

(*ii*) the reflection H of $\mathbb{R}^2$ through the line $x = y$,

(*iii*) the matrices of $H \circ \mathcal{R}_{-\pi/4}$ and $\mathcal{R}_{-\pi/4} \circ H$, where H is the reflection of part (*ii*),

(*iv*) the matrix of the rotation $H \circ \mathcal{R}_{-\pi/4} \circ H$ of $\mathbb{R}^2$.

Exercise 7.3.5. Let H be the reflection of $\mathbb{R}^2$ through the line ℓ through the origin, and let H' be the reflection through $\ell^\perp$. Describe the following linear mappings:

(*i*) HR_θ and $H'R_\theta H$,

(*ii*) $HR_\theta H$, and

(*iii*) $R_\theta H R_{-\theta}$.

Exercise 7.3.6. Let $V = \mathbb{C}$ and consider the mapping $T : V \to V$ defined by $T(z) = \overline{z}$. Describe T as a linear mapping with domain and target $\mathbb{R}^2$, and find its matrix. Also, give a geometric interpretation of T.

Exercise 7.3.7. Show directly that a reflection H of $\mathbb{R}^n$ is orthogonal by checking that for all $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{R}^n$,

$$H(\mathbf{x}) \cdot H(\mathbf{y}) = \mathbf{x} \cdot \mathbf{y}.$$

Exercise 7.3.8. Find the reflection of $\mathbb{R}^3$ through the plane P if:

(*i*) P is the plane $x + y + z = 0$;

(*ii*) P is the plane $ax + by + cz = 0$.

Exercise 7.3.9. This exercise gives a formula for the projection on a subspace W of $\mathbb{R}^m$ that does not require having an orthonormal basis of W. Let $A \in \mathbb{R}^{m\times n}$ be a matrix whose columns are a basis of W.

(*i*) Let $\mathbf{x} \in \mathbb{R}^n$. Under what condition is $A\mathbf{x}$ the projection of $\mathbf{x}$ on W?

(*ii*) Prove that A^TA is invertible. (Hint: consider $\mathbf{x}^TA^TA\mathbf{x}$.)

(*iii*) Prove that $A(A^TA)^{-1}A^T\mathbf{x} \in W$ if $\mathbf{x} \in \mathbb{R}^m$, and $A(A^TA)^{-1}A^T\mathbf{x} = \mathbf{x}$ if $\mathbf{x} \in W$.

(*iv*) Next show that for every $\mathbf{v} \in \mathbb{R}^m$,

$$\mathbf{v} - A(A^TA)^{-1}A^T\mathbf{v}$$

is orthogonal to W. Hint: show that $A^T(\mathbf{v} - A(A^TA)^{-1}A^T\mathbf{v}) = 0$.

(*v*) Conclude that $A(A^TA)^{-1}A^T\mathbf{x}$ is the projection of $\mathbf{x} \in \mathbb{R}^m$ onto W. Thus, P_W has matrix $A(A^TA)^{-1}A^T$.

Exercise 7.3.10. Let $Q = \begin{pmatrix} a & b \\ b & -a \end{pmatrix}$, where $a^2 + b^2 = 1$. Show that Q is a reflection by demonstrating that Q can be written in the form $\begin{pmatrix} 1-2u_1^2 & -2u_1u_2 \\ -2u_1u_2 & 1-2u_2^2 \end{pmatrix}$, where $u_1^2 + u_2^2 = 1$, and verifying that Q is the reflection through the line $\mathbb{R}\begin{pmatrix} -u_2 \\ u_1 \end{pmatrix}$.

7.4 Coordinates with Respect to a Basis and Matrices of Linear Mappings

We now come to a technical question. Suppose V and W are finite-dimensional vector spaces over the same field $\mathbb{F}$ of dimensions n and m respectively, and let $T : V \to W$ be a linear mapping. How can we represent T? For example, if V and W are $\mathbb{F}^n$ and $\mathbb{F}^m$ respectively, then T is a matrix linear mapping, hence is determined by a unique element $A \in \mathbb{F}^{m\times n}$. The assignment $T \to A = M_T \in \mathbb{F}^{m\times n}$ depends on the fact that there are natural coordinates on $\mathbb{F}^n$ and $\mathbb{F}^m$. Recall that for a general vector space, coordinates depend on choosing a basis. So, in order to represent a linear mapping $T : V \to W$ as a matrix, we must first choose bases for both V and W. The first job in this section is to show how to define the matrix of an arbitrary linear mapping T with respect to a choice of bases for V and W, and the second job is to investigate how a different choice of these bases affects the matrix of T. The result is called the change of basis formula.

7.4.1 Coordinates with respect to a basis

As usual, let V be a finite-dimensional vector space over $\mathbb{F}$, and let $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ be a basis of V. Recall that every $\mathbf{v} \in V$ has a unique expression

$$\mathbf{v} = r_1\mathbf{v}_1 + r_2\mathbf{v}_2 + \cdots + r_n\mathbf{v}_n.$$

The scalars $r_1, \ldots, r_n$ are called the *coordinates* of $\mathbf{v}$ with respect to $\mathcal{B}$. We will write $\mathbf{v} = (r_1, r_2, \ldots, r_n)_{\mathcal{B}}$. Notice that the notion of coordinates assumes that the basis $\mathcal{B}$ is ordered. The term coordinate deserves some explanation. A coordinate is actually a function on V with target $\mathbb{F}$. A basis $\mathcal{B}$ as above determines n coordinate functions $x_1, \ldots, x_n$, which are defined by putting $x_i(\mathbf{v}) = r_i$ when $\mathbf{v}$ has the (above) expansion in the basis $\mathcal{B}$. The coordinates of $\mathbf{v}$ are the values of the coordinate functions on $\mathbf{v}$. Note that the coordinate functions are linear: $x_i(a\mathbf{v} + b\mathbf{w}) = ax_i(\mathbf{v}) + bx_i(\mathbf{w})$.

Finding the coordinates of a vector in $\mathbb{F}^n$ with respect to a basis is a familiar problem in matrix inversion. Here is a preliminary example.

Example 7.8. Let us choose two different bases of $\mathbb{R}^2$, say

$$\mathcal{B} = \{\begin{pmatrix}1\\2\end{pmatrix}, \begin{pmatrix}0\\1\end{pmatrix}\} \quad \text{and} \quad \mathcal{B}' = \{\begin{pmatrix}1\\1\end{pmatrix}, \begin{pmatrix}1\\-1\end{pmatrix}\}.$$

Expanding $\mathbf{e}_1$ in terms of these two bases gives two different sets of coordinates. By inspection,

$$\mathbf{e}_1 = \begin{pmatrix}1\\0\end{pmatrix} = 1\begin{pmatrix}1\\2\end{pmatrix} - 2\begin{pmatrix}0\\1\end{pmatrix}$$

and

$$\mathbf{e}_1 = \begin{pmatrix}1\\0\end{pmatrix} = \frac{1}{2}\begin{pmatrix}1\\1\end{pmatrix} + \frac{1}{2}\begin{pmatrix}1\\-1\end{pmatrix}.$$

Thus, $\mathbf{e}_1 = (1,-2)_{\mathcal{B}}$, while $\mathbf{e}_1 = (\frac{1}{2},\frac{1}{2})_{\mathcal{B}'}$. ⊠

7.4.2 The change of basis matrix

The first question is how the coordinates of $\mathbf{v}$ with respect to $\mathcal{B}$ and $\mathcal{B}'$ are related. This is answered by setting up a linear system as follows: expanding the basis $\mathcal{B}'$ in terms of the basis $\mathcal{B}$ gives

$$\begin{pmatrix}1\\1\end{pmatrix} = a\begin{pmatrix}1\\2\end{pmatrix} + b\begin{pmatrix}0\\1\end{pmatrix} \quad \text{and} \quad \begin{pmatrix}1\\-1\end{pmatrix} = c\begin{pmatrix}1\\2\end{pmatrix} + d\begin{pmatrix}0\\1\end{pmatrix}.$$

Expressed in matrix form, these equations become

$$\begin{pmatrix}1 & 1\\1 & -1\end{pmatrix} = \begin{pmatrix}1 & 0\\2 & 1\end{pmatrix}\begin{pmatrix}a & c\\b & d\end{pmatrix}.$$

Thus,

$$\begin{pmatrix}a & c\\b & d\end{pmatrix} = \begin{pmatrix}1 & 0\\2 & 1\end{pmatrix}^{-1}\begin{pmatrix}1 & 1\\1 & -1\end{pmatrix} = \begin{pmatrix}1 & 1\\-1 & -3\end{pmatrix}.$$

Now suppose $\mathbf{v}$ has coordinates $(r,s)_{\mathcal{B}}$ and $(x,y)_{\mathcal{B}'}$ with respect to $\mathcal{B}$ and $\mathcal{B}'$. Then

$$\mathbf{v} = \begin{pmatrix}1 & 0\\2 & 1\end{pmatrix}\begin{pmatrix}r\\s\end{pmatrix} = \begin{pmatrix}1 & 1\\1 & -1\end{pmatrix}\begin{pmatrix}x\\y\end{pmatrix}.$$

Therefore,

$$\begin{pmatrix}r\\s\end{pmatrix} = \begin{pmatrix}1 & 1\\-1 & -3\end{pmatrix}\begin{pmatrix}x\\y\end{pmatrix}.$$

We can imitate this in the general case. Let

$$\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$$

and

$$\mathcal{B}' = \{\mathbf{v}'_1, \mathbf{v}'_2, \ldots, \mathbf{v}'_n\}$$

be two bases of V. Define the *change of basis matrix* $\mathcal{M}_{\mathcal{B}'}^{\mathcal{B}} \in \mathbb{F}^{n\times n}$ to be the matrix (a_{ij}) with entries determined by

$$\mathbf{v}'_j = \sum_{i=1}^{n} a_{ij}\mathbf{v}_i.$$

For example, suppose $n = 2$. Then

$$\mathbf{v}'_1 = a_{11}\mathbf{v}_1 + a_{21}\mathbf{v}_2,$$

$$\mathbf{v}'_2 = a_{12}\mathbf{v}_1 + a_{22}\mathbf{v}_2.$$

In matrix form as above, this looks like

$$(\mathbf{v}'_1\ \mathbf{v}'_2) = (\mathbf{v}_1\ \mathbf{v}_2)\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = (\mathbf{v}_1\ \mathbf{v}_2)\mathcal{M}_{\mathcal{B}'}^{\mathcal{B}},$$

where

$$\mathcal{M}_{\mathcal{B}'}^{\mathcal{B}} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}.$$

Notice that $(\mathbf{v}_1\ \mathbf{v}_2)$ is a generalized matrix in the sense that it is a 1×2 matrix with vector entries. A nice general property of this notation is that whenever $\mathbf{v}_1, \ldots, \mathbf{v}_n$ is a basis of V and $(\mathbf{v}_1 \cdots \mathbf{v}_n)A = (\mathbf{v}_1 \cdots \mathbf{v}_n)B$, then $A = B$.

Returning to the general case, let $\mathcal{B}$ and $\mathcal{B}'$ be the two bases of V defined above. Then,

$$(\mathbf{v}'_1\ \mathbf{v}'_2\ \cdots\ \mathbf{v}'_n) = (\mathbf{v}_1\ \mathbf{v}_2\ \cdots\ \mathbf{v}_n)\mathcal{M}_{\mathcal{B}'}^{\mathcal{B}}. \tag{7.13}$$

Example 7.9. For $\mathcal{B}$ and $\mathcal{B}'$ as in Example 7.8, we have

$$\mathcal{M}_{\mathcal{B}'}^{\mathcal{B}} = \begin{pmatrix} 1 & 1 \\ -1 & -3 \end{pmatrix}. \qquad \boxtimes$$

Proposition 7.16. *Let $\mathcal{B}$ and $\mathcal{B}'$ be bases of V. Then*

$$\mathcal{M}_{\mathcal{B}}^{\mathcal{B}'} = (\mathcal{M}_{\mathcal{B}'}^{\mathcal{B}})^{-1}.$$

Also

$$\mathcal{M}_{\mathcal{B}}^{\mathcal{B}} = I_n.$$

Proof. First of all, the identity $\mathcal{M}_{\mathcal{B}}^{\mathcal{B}} = I_n$ is clear. For the rest of the proof, let us assume $n = 2$ for simplicity. Now,

$$(\mathbf{v}_1\ \mathbf{v}_2) = (\mathbf{v}'_1\ \mathbf{v}'_2)\mathcal{M}_{\mathcal{B}}^{\mathcal{B}'} = (\mathbf{v}_1\ \mathbf{v}_2)\mathcal{M}_{\mathcal{B}'}^{\mathcal{B}}\mathcal{M}_{\mathcal{B}}^{\mathcal{B}'}.$$

Thus,

$$\mathcal{M}_{\mathcal{B}'}^{\mathcal{B}}\mathcal{M}_{\mathcal{B}}^{\mathcal{B}'} = \mathcal{M}_{\mathcal{B}}^{\mathcal{B}} = I_2.$$

Hence, $(\mathcal{M}_{\mathcal{B}'}^{\mathcal{B}})^{-1} = \mathcal{M}_{\mathcal{B}}^{\mathcal{B}'}$. □

Now let's see what happens when a third basis $\mathcal{B}'' = \{\mathbf{v}''_1, \ldots, \mathbf{v}''_n\}$ is thrown in. Iterating the expression in (7.13) gives

$$(\mathbf{v}''_1\ \cdots\ \mathbf{v}''_n) = (\mathbf{v}'_1\ \cdots\ \mathbf{v}'_n)\mathcal{M}_{\mathcal{B}''}^{\mathcal{B}'} = (\mathbf{v}_1\ \cdots\ \mathbf{v}_n)\mathcal{M}_{\mathcal{B}'}^{\mathcal{B}}\mathcal{M}_{\mathcal{B}''}^{\mathcal{B}'}.$$

This gives the following result.

Proposition 7.17. *Let $\mathcal{B}$, $\mathcal{B}'$, and $\mathcal{B}''$ be bases of V. Then*

$$\mathcal{M}_{\mathcal{B}''}^{\mathcal{B}} = \mathcal{M}_{\mathcal{B}'}^{\mathcal{B}}\mathcal{M}_{\mathcal{B}''}^{\mathcal{B}'}.$$

7.4.3 The matrix of a linear mapping

Now suppose $T : V \to W$ is a linear mapping. The purpose of this section is to associate a matrix to T with respect to a pair of chosen bases of V and W. Let these bases be

$$\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$$

for V and

$$\mathcal{B}' = \{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_m\}$$

for W. Then one can write

$$T(\mathbf{v}_j) = \sum_{i=1}^{m} c_{ij}\mathbf{w}_i. \tag{7.14}$$

Note: the ith component of $T(\mathbf{v}_j)$ with respect to $\mathcal{B}'$ is denoted by c_{ij}. This is exactly analogous to how we defined the change of basis matrix. Now we can define the matrix $\mathcal{M}_{\mathcal{B}'}^{\mathcal{B}}(T)$ of T with respect to the bases $\mathcal{B}$ and $\mathcal{B}'$.

Definition 7.5. The matrix of T with respect to the bases $\mathcal{B}$ and $\mathcal{B}'$ is defined to be the $m \times n$ matrix (c_{ij}). In other words, $\mathcal{M}_{\mathcal{B}'}^{\mathcal{B}}(T) = (c_{ij})$.

Let us put $T(\mathbf{v}_1\ \mathbf{v}_2 \cdots \mathbf{v}_n) = (T(\mathbf{v}_1)\ T(\mathbf{v}_2) \cdots T(\mathbf{v}_n))$. Expressing (7.14) in matrix form gives

$$T(\mathbf{v}_1\ \mathbf{v}_2 \cdots \mathbf{v}_n) = (T(\mathbf{v}_1)\ T(\mathbf{v}_2) \cdots T(\mathbf{v}_n)) = (\mathbf{w}_1\ \mathbf{w}_2 \cdots \mathbf{w}_m)\mathcal{M}_{\mathcal{B}'}^{\mathcal{B}}(T). \tag{7.15}$$

This notation is set up so that if $V = \mathbb{F}^n$, $W = \mathbb{F}^m$, and $T = T_A$, where $A \in \mathbb{F}^{m\times n}$, then $\mathcal{M}^{\mathcal{B}}_{\mathcal{B}'}(T) = A$ when $\mathcal{B}$ and $\mathcal{B}'$ are the standard bases, since $T_A(\mathbf{e}_j)$ is the jth column of A. For (7.15) says that

$$A = T_A I_n = I_m \mathcal{M}^{\mathcal{B}}_{\mathcal{B}'}(T_A) = \mathcal{M}^{\mathcal{B}}_{\mathcal{B}'}(T_A).$$

We remark that

$$\mathcal{M}^{\mathcal{B}}_{\mathcal{B}'}(I_V) = \mathcal{M}^{\mathcal{B}}_{\mathcal{B}'},$$

where $I_V : V \to V$ is the identity mapping.

7.4.4 The Case $V = W$

Now suppose $V = W$. In this case, we want to express the matrix of T in a single basis and then find its expression in another basis. So let $\mathcal{B}$ and $\mathcal{B}'$ be bases of V. As above, for simplicity, we assume $n = 2$ and put $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2\}$ and $\mathcal{B}' = \{\mathbf{v}'_1, \mathbf{v}'_2\}$. Hence $(\mathbf{v}'_1\ \mathbf{v}'_2) = (\mathbf{v}_1\ \mathbf{v}_2)\mathcal{M}^{\mathcal{B}}_{\mathcal{B}'}$. Since T is linear, we may write

$$\begin{aligned}(T(\mathbf{v}'_1)\ T(\mathbf{v}'_2)) &= (T(\mathbf{v}_1)\ T(\mathbf{v}_2))\mathcal{M}^{\mathcal{B}}_{\mathcal{B}'} \\ &= (\mathbf{v}_1\ \mathbf{v}_2)\mathcal{M}^{\mathcal{B}}_{\mathcal{B}}(T)\mathcal{M}^{\mathcal{B}}_{\mathcal{B}'} \\ &= (\mathbf{v}'_1\ \mathbf{v}'_2)\mathcal{M}^{\mathcal{B}'}_{\mathcal{B}}\mathcal{M}^{\mathcal{B}}_{\mathcal{B}}(T)\mathcal{M}^{\mathcal{B}}_{\mathcal{B}'}.\end{aligned}$$

Hence,

$$\mathcal{M}^{\mathcal{B}'}_{\mathcal{B}'}(T) = \mathcal{M}^{\mathcal{B}'}_{\mathcal{B}}\mathcal{M}^{\mathcal{B}}_{\mathcal{B}}(T)\mathcal{M}^{\mathcal{B}}_{\mathcal{B}'}.$$

We have therefore proved the following proposition.

Proposition 7.18. *Let $T : V \to V$ be linear and let $\mathcal{B}$ and $\mathcal{B}'$ be bases of V. Then*

$$\mathcal{M}^{\mathcal{B}'}_{\mathcal{B}'}(T) = \mathcal{M}^{\mathcal{B}'}_{\mathcal{B}}\mathcal{M}^{\mathcal{B}}_{\mathcal{B}}(T)\mathcal{M}^{\mathcal{B}}_{\mathcal{B}'}. \tag{7.16}$$

Thus, if $P = \mathcal{M}^{\mathcal{B}'}_{\mathcal{B}}$, we have

$$\mathcal{M}^{\mathcal{B}'}_{\mathcal{B}'}(T) = P\mathcal{M}^{\mathcal{B}}_{\mathcal{B}}(T)P^{-1}. \tag{7.17}$$

Example 7.10. Consider the linear mapping $T : \mathbb{R}^2 \to \mathbb{R}^2$ whose matrix with respect to the standard basis is

$$A = \begin{pmatrix} 1 & 0 \\ -4 & 3 \end{pmatrix}.$$

Let's find the matrix B of T with respect to the basis $(1, 1)^T$ and $(1, -1)^T$. Calling this basis $\mathcal{B}'$ and the standard basis $\mathcal{B}$, we have

$$\mathcal{M}_{\mathcal{B}'}^{\mathcal{B}} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}.$$

To compute B, we have to use the matrix equation

$$\begin{pmatrix} 1 & 0 \\ -4 & 3 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} B,$$

so

$$B = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}^{-1} \begin{pmatrix} 1 & 0 \\ -4 & 3 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}.$$

Computing the product gives

$$B = \begin{pmatrix} 0 & -3 \\ 1 & 4 \end{pmatrix}. \qquad \boxtimes$$

Example 7.11. Consider the reflection H of $\mathbb{R}^2$ through a line ℓ (containing $\mathbf{0}$). Let $\mathbf{v}_1$ be a nonzero element of ℓ and $\mathbf{v}_2$ a nonzero element of the line $\ell^{\perp}$. Then $H(\mathbf{v}_1) = \mathbf{v}_1$ and $H(\mathbf{v}_2) = -\mathbf{v}_2$. Since H is an isometry, it is natural to switch to an orthonormal basis. So let $\mathcal{B} = \{\mathbf{u}_1, \mathbf{u}_2\}$ be the orthonormal basis of $\mathbb{R}^2$ where $\mathbf{u}_i = \dfrac{1}{|\mathbf{v}_i|}\mathbf{v}_i$ for $i = 1, 2$. Then

$$\mathcal{M}_{\mathcal{B}}^{\mathcal{B}}(H) = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Now suppose that $\mathbf{u}_1 = \dfrac{1}{\sqrt{2}}(1,1)^T$ and $\mathbf{u}_2 = \dfrac{1}{\sqrt{2}}(1,-1)^T$. Then the matrix $(\mathbf{u}_1\ \mathbf{u}_2)$ is orthogonal. Let us use this to find the matrix A of H with respect to to the standard basis $\mathcal{B}' = \{\mathbf{e}_1, \mathbf{e}_2\}$ of $\mathbb{R}^2$. Since

$$\mathcal{M}_{\mathcal{B}'}^{\mathcal{B}'}(H) = \mathcal{M}_{\mathcal{B}}^{\mathcal{B}'} \mathcal{M}_{\mathcal{B}}^{\mathcal{B}}(H) \mathcal{M}_{\mathcal{B}'}^{\mathcal{B}},$$

and since $\mathcal{B}$ is an orthonormal basis, it follows that $\mathcal{M}_{\mathcal{B}'}^{\mathcal{B}} = (\mathbf{u}_1\ \mathbf{u}_2)^T$. Thus, by (7.16),

$$A = \mathcal{M}_{\mathcal{B}}^{\mathcal{B}'} \mathcal{M}_{\mathcal{B}}^{\mathcal{B}}(H) \mathcal{M}_{\mathcal{B}}^{\mathcal{B}'} =$$

$$\frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

This checks, since by definition, H sends $\mathbf{e}_1$ to $\mathbf{e}_2$ and sends $\mathbf{e}_2$ to $\mathbf{e}_1$. $\boxtimes$

7.4.5 Similar matrices

The relationship between the matrices of a linear mapping in different bases suggests that we recall a term introduced in Exercise 5.2.6.

Definition 7.6. Let A and B be $n \times n$ matrices over $\mathbb{F}$. Then we say that A *is similar to* B if there exists an invertible $P \in \mathbb{F}^{n\times n}$ such that $B = PAP^{-1}$.

It is not hard to see that similarity is an equivalence relation on $\mathbb{F}^{n\times n}$ (exercise: check this). An equivalence class for this equivalence relation is called a *conjugacy class* of $\mathbb{F}^{n\times n}$. The meaning of a conjugacy class is given in the next proposition.

Proposition 7.19. *Let V be a finite-dimensional vector space over $\mathbb{F}$ such that $\dim V = n$. Then the matrices that represent a given linear mapping $T : V \to V$ form a conjugacy class of $\mathbb{F}^{n\times n}$.*

Recall that a linear mapping $T : V \to V$ is semisimple if there exists a basis $\mathcal{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ of V for which $T(\mathbf{v}_i) = \mu_i \mathbf{v}_i$ for some scalars $\mu_1, \ldots, \mu_n$ in $\mathbb{F}$. Thus a linear mapping T is semisimple if and only if the conjugacy class of its matrix with respect to some basis of V contains a diagonal matrix. The semisimple linear mappings $T : V \to V$ are classified in the next chapter. Conjugacy classes are also important in group theory. We will say more about this in Chap. 11.

7.4.6 The matrix of a composition $T \circ S$

Suppose $S, T : V \to V$ are linear mappings. Recall that we saw in Proposition 3.6 that when $V = \mathbb{F}^n$, then $M_{T\circ S} = M_T M_S$. We will now prove that this fact also holds in general.

Proposition 7.20. *Assume that V is a finite-dimensional vector space with basis $\mathcal{B}$. Then*

$$\mathcal{M}_{\mathcal{B}}^{\mathcal{B}}(T \circ S) = \mathcal{M}_{\mathcal{B}}^{\mathcal{B}}(T)\mathcal{M}_{\mathcal{B}}^{\mathcal{B}}(S).$$

Proof. We leave this to the reader.

7.4.7 The determinant of a linear mapping

The product identity $\det(AB) = \det(A)\det(B)$ and the fact that $\det(A^{-1}) = \det(A)^{-1}$ imply that two similar matrices always have the same determinant, since

$$\det(PAP^{-1}) = \det(P)\det(A)\det(P^{-1}) = \det(A).$$

Using this, one can define the determinant of a linear mapping $T : V \to V$, provided V is finite-dimensional. The definition goes as follows.

Definition 7.7. Let V be a finite-dimensional vector space and suppose $T : V \to V$ is linear. Then the determinant $\det(T)$ of T is defined to be $\det(A)$, where $A \in \mathbb{F}^{n\times n}$ is any matrix representing T with respect to some basis of V.

In order to show that $\det(T)$ is well defined, we need to show that $\det(T)$ is independent of the choice of basis of V. But by Proposition 7.18, if A and B are matrices of T with respect to different bases, then A and B are similar, i.e., $B = PAP^{-1}$ for some invertible $P \in \mathbb{F}^{n\times n}$. Hence, $\det(B) = \det(A)$. Thus, $\det(T)$ is indeed well defined.

Example 7.12. Suppose $T : V \to V$ is semisimple. Then there exists a basis for which the matrix of T is a diagonal matrix $D = \operatorname{diag}(\mu_1 \ \cdots \ \mu_n)$. Thus $\det(T) = \mu_1 \dots \mu_n$, since the determinant of a diagonal matrix is the product of the diagonal entries. ⊠

Example 7.13. Let V be a finite-dimensional inner product space, and let $T : V \to V$ be an isometry. Since the determinant of an orthogonal matrix is ± 1 (since $Q^TQ = I_n$), it follows that $\det(T) = \pm 1$ also. ⊠

Proposition 7.21. *If $S, T : V \to V$ are linear mappings on a finite-dimensional vector space V, then* $\det(T \circ S) = \det(T)\det(S)$.

Proof. Apply Proposition 7.20 and the product formula. □

For example, we have the following.

Proposition 7.22. *Suppose $T : V \to V$ is a linear mapping such that $\det(T) \neq 0$. Then there exists a linear mapping $S : V \to V$ such that $T \circ S = S \circ T = I_V$. In particular, T is a bijection.*

Proof. Choose a basis $\mathcal{B}$ of V and let A be the matrix of T with respect to $\mathcal{B}$. Since $\det(T) \neq 0$, $\det(A) \neq 0$ too, so A has an inverse A^{-1}. Now let $S : V \to V$ be the linear mapping whose matrix with respect to $\mathcal{B}$ is A^{-1}. Then by Proposition 7.20,

$$\mathcal{M}_{\mathcal{B}}^{\mathcal{B}}(T \circ S) = \mathcal{M}_{\mathcal{B}}^{\mathcal{B}}(T)\mathcal{M}_{\mathcal{B}}^{\mathcal{B}}(S) = AA^{-1} = I_n.$$

Therefore, $T \circ S = I_V$. Similarly, $S \circ T = I_V$ too. The assertion that T is a bijection follows immediately from Proposition 1.1 or Exercise 7.1.1. □

Exercises

Exercise 7.4.1. Find the coordinates of the standard basis $\mathbf{e}_1,\ \mathbf{e}_2,\ \mathbf{e}_3$ of $\mathbb{R}^3$ in terms of the basis $(1,1,1)^T, (1,0,1)^T,\ (0,1,1)^T$, and find the matrix of the linear mapping $T((x_1,x_2,x_3)^T) = (4x_1 + x_2 - x_3, x_1 + 3x_3, x_2 + 2x_3)^T$ with respect to this basis.

Exercise 7.4.2. Consider the basis $(1,1,1)^T$, $(1,0,1)^T$, and $(0,1,1)^T$ of $\mathbb{R}^3$. Find the matrix of the linear mapping $T : \mathbb{R}^3 \to \mathbb{R}^3$ defined by $T(\mathbf{x}) = (1,1,1)^T \times \mathbf{x}$ with respect to this basis.

Exercise 7.4.3. Let $H : \mathbb{R}^2 \to \mathbb{R}^2$ be the reflection through the line $2x = y$. Find a basis of $\mathbb{R}^2$ such that the matrix of H is diagonal.

Exercise 7.4.4. Show that every projection $P_{\mathbf{a}} : \mathbb{R}^2 \to \mathbb{R}^2$ is semisimple by explicitly finding a basis for which the matrix of $P_{\mathbf{a}}$ is diagonal. Also, find this diagonal matrix.

Exercise 7.4.5. Let $\mathcal{R}_\theta$ be the usual rotation of $\mathbb{R}^2$. Does there exist a basis of $\mathbb{R}^2$ for which the matrix of $\mathcal{R}_\theta$ is diagonal?

Exercise 7.4.6. Show that matrix similarity is an equivalence relation on $\mathbb{F}^{n\times n}$.

Exercise 7.4.7. Find the matrix $\mathcal{M}_{\mathcal{B}}^{\mathcal{B}}(C_{\mathbf{a}})$ of the cross product $C_{\mathbf{a}} = \mathbf{a} \times \mathbf{x}$ in the following cases:

(*i*) $\mathbf{a} = \mathbf{e}_1$ and $\mathcal{B}$ is the standard basis;

(*ii*) $\mathbf{a} = \mathbf{e}_1$ and $\mathcal{B}$ is the basis $\{\mathbf{e}_1, \mathbf{e}_2 + \mathbf{e}_3, \mathbf{e}_2 - \mathbf{e}_3\}$.

Exercise 7.4.8. Let $V = (\mathbb{F}_2)^{2\times 2}$, and let $T : V \to V$ be defined by $T(B) = AB - BA$, where $A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ as an element of V.

(*i*) Show that T is a linear mapping.

(*ii*) Find the matrix of T with respect to a suitable basis (of your choice) of V.

(*iii*) Find $\det(T)$.

Exercise 7.4.9. Let V be a finite-dimensional vector space, and suppose $T : V \to V$ is a linear mapping. Find the relationship between between $\mathcal{N}\big(\mathcal{M}_{\mathcal{B}}^{\mathcal{B}}(T)\big)$ and $\mathcal{N}\big(\mathcal{M}_{\mathcal{B}'}^{\mathcal{B}'}(T)\big)$, where $\mathcal{B}$ and $\mathcal{B}'$ are any two bases of V.

Exercise 7.4.10. Let $T : V \to V$ be a linear mapping, where V has finite dimension. Show that the kernel of T is nontrivial if and only if $\det(T) = 0$.

Exercise 7.4.11. Suppose the characteristic of $\mathbb{F}$ is different from 2, and let $V = \mathbb{F}^{n\times n}$. Let $T : V \to V$ be the linear mapping given by $T(A) = A^T$. Find a basis of V for which the matrix of T is diagonal. (Hint: recall that A is the sum of a symmetric matrix and a skew-symmetric matrix.)

Exercise 7.4.12. Let $V = \mathbb{F}_4 = \{0, 1, \alpha, \beta\}$ be the four-element Galois field considered as a vector space over the prime field $\mathbb{F}_2$ (see Section 2.6.2).

(*i*) Show that 1 and α form a basis of V.

(*ii*) Show that the Frobenius map $F : V \to V$ given by $F(x) = x^2$ is a linear mapping.

(*iii*) Find the matrix of F with respect to the basis of (*i*).

(*iv*) Is F semi-simple?

7.5 Further Results on Mappings

As usual, all vector spaces will be over the field $\mathbb{F}$. If V and W are vector spaces, the space $L(V, W)$ denotes the space of all linear mappings $T : V \to W$. If $V = \mathbb{F}^n$ and $W = \mathbb{F}^m$, then $L(V, W) = \mathbb{F}^{m \times n}$. The purpose of this section is to study $L(V, W)$ for various choices of V and W.

7.5.1 The space $L(V, W)$

Mappings $F : V \to W$ can be added using pointwise addition and can be multiplied by scalars in a similar way. That is, if $F, G : V \to W$ are two mappings, their sum $F + G$ is the mapping formed by setting

$$(F + G)(\mathbf{v}) = F(\mathbf{v}) + G(\mathbf{v}).$$

Scalar multiplication is defined by putting

$$(aF)(\mathbf{v}) = aF(\mathbf{v})$$

for any scalar a. Hence, one can form linear combinations of mappings. It isn't hard to see that the set of all mappings with domain V and target W is a vector space over $\mathbb{F}$. Now let $L(V, W)$ denote the set of all linear mappings with domain V and target W. Then $L(V, W)$ is a vector space over $\mathbb{F}$ under the pointwise addition and scalar multiplication defined above. The following result gives the dimension of $L(V, W)$ in the finite-dimensional case.

Proposition 7.23. *Suppose V and W are finite-dimensional vector spaces, say $\dim V = n$ and $\dim W = m$. Then $\dim L(V, W) = mn$.*

Proof. Choose bases $\mathcal{B}$ of V and $\mathcal{B}'$ of W. Then T has matrix $M_T = \mathcal{M}^{\mathcal{B}}_{\mathcal{B}'}(T)$, and putting $\phi(T) = M_T$ defines a linear mapping $\Phi : L(V, W) \to \mathbb{F}^{m \times n}$. In fact, Φ is a bijection, and therefore Φ is an isomorphism. This implies $\dim L(V, W) = \dim \mathbb{F}^{m \times n} = mn$, since an isomorphism preserves dimension. □

7.5.2 The dual space

The space $V^* = L(V, \mathbb{F})$ of linear maps (or linear functions) from V to $\mathbb{F}$ is called the *dual space* of V. If V is a finite-dimensional vector space, then the previous result says that $\dim V^* = \dim V$. It turns out that even though V and V^* have the same dimension, there usually is no natural isomorphism

between them unless there is some additional structure. For example, if V is a finite-dimensional inner product space, then there is a natural isomorphism (see below). Given a basis of V, however, there is a natural basis of V^* known as the *dual basis*, which we now describe. If $\mathbf{v}_1, \ldots, \mathbf{v}_n$ is a basis of V, then the dual basis $\mathbf{v}_1^*, \ldots, \mathbf{v}_n^*$ of V^* is defined by specifying how each $\mathbf{v}_i^*$ acts on the basis $\mathbf{v}_1, \ldots, \mathbf{v}_n$. Since a linear mapping is uniquely defined by giving its values on a basis, this suffices to define a unique element of V^*. Thus put

$$\mathbf{v}_i^*(\mathbf{v}_j) = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j. \end{cases} \tag{7.18}$$

More succinctly, $\mathbf{v}_i^*(\mathbf{v}_j) = \delta_{ij}$. To justify the term dual basis, we prove the following proposition.

Proposition 7.24. *If $\mathbf{v}_1, \ldots, \mathbf{v}_n$ is a basis of V, then $\mathbf{v}_1^*, \ldots, \mathbf{v}_n^*$ is a basis of V^*.*

Proof. Since V and V^* have the same dimension, it suffices to show that the dual basis vectors are independent. Suppose $\sum_{i=1}^n a_i \mathbf{v}_i^* = \mathbf{0}$. By (7.18),

$$\Big(\sum_{i=1}^n a_i \mathbf{v}_i^*\Big)(\mathbf{v}_j) = \sum_{i=1}^n a_i \mathbf{v}_i^*(\mathbf{v}_j) = a_j.$$

Hence each a_j is equal to zero, so $\mathbf{v}_1^*, \ldots, \mathbf{v}_n^*$ are indeed independent. □

Example 7.14. In fact, the dual basis for $V = \mathbb{F}^n$ is already quite familiar. Recall that the ith component function $x_i : \mathbb{F}^n \to \mathbb{F}$ is defined by

$$x_i(a_1, \cdots, a_n) = a_i$$

for $i = 1, \ldots, n$. Then each x_i is in V^*. In fact, $x_1, \ldots, x_n$ is the basis of V^* dual to the standard basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$. ⊠

Example 7.15. Here is an infinite-dimensional example. Let $V = \mathbb{F}[x]$. Then the evaluation at an element $r \in \mathbb{F}$ is the map $e_r : V \to \mathbb{F}$ defined by $e_r(f) = f(r)$. Then $e_r \in V^*$ for every $r \in \mathbb{F}$. The kernel of e_r consists of all $f \in \mathbb{F}[x]$ such that $f(r) = 0$.

Let us now consider the case mentioned above in which V is a finite-dimensional inner product space.

Proposition 7.25. *Let V be a finite-dimensional inner product space over $\mathbb{R}$. If $\mathbf{v} \in V$, let $\varphi_{\mathbf{v}} : V \to \mathbb{R}$ be the element of V^* defined by*

$$\varphi_{\mathbf{v}}(\mathbf{x}) = (\mathbf{v}, \mathbf{x}),$$

where $(\ ,\ )$ is the inner product on V. Then the mapping $\Phi : V \to V^*$ defined by $\Phi(\mathbf{v}) = \varphi_{\mathbf{v}}$ is an isomorphism.

Proof. It is clear, by the properties of an inner product, that $\varphi_{\mathbf{v}} \in V^*$. Since $\dim V = \dim V^*$, we have only to show that Φ is injective. But if $\Phi(\mathbf{v}) = \mathbf{0}$, then $(\mathbf{v}, \mathbf{x}) = 0$ for all $\mathbf{x} \in V$. In particular, $(\mathbf{v}, \mathbf{v}) = 0$, so $\mathbf{v} = \mathbf{0}$ by the definition of an inner product, and hence by Proposition 7.3, Φ is injective. □

In a similar vein, a Hermitian inner product on a finite-dimensional vector space V over the complex numbers enables one to define an isomorphism from V to V^*, although the definition is slightly different, since $(\alpha\mathbf{x}, \mathbf{v}) = \overline{\alpha}(\mathbf{x}, \mathbf{v})$ for a Hermitian inner product (see Example 6.35).

It turns out that there is always a natural isomorphism between V and V^{**}. This is an interesting exercise.

7.5.3 Multilinear maps

Let V and W be vector spaces over $\mathbb{F}$. A mapping

$$T : V \times V \times \cdots \times V \to W \quad (k \text{ factors})$$

is called k-multilinear if for all $i = 1, \ldots, k$, the mapping $S_i : V \to W$ defined by

$$S_i(\mathbf{x}) = T(\mathbf{v}_1, \ldots, \mathbf{v}_{i-1}, \mathbf{x}, \mathbf{v}_{i+1}, \ldots, \mathbf{v}_k)$$

is a linear mapping for every $\mathbf{v}_1, \ldots, \mathbf{v}_{i-1}, \mathbf{v}_{i+1}, \ldots, \mathbf{v}_k \in V$.

Example 7.16. Let $V = \mathbb{F}$ and define $T(r_1, r_2, \ldots, r_k) = r_1 r_2 \cdots r_k$. Then T is k-multilinear on V. ⊠

More interestingly, let $V = \mathbb{F}^n$, where elements of $\mathbb{F}^n$ are viewed as columns. Define $D : \mathbb{F}^{n \times n} = \mathbb{F}^n \times \cdots \times \mathbb{F}^n \to \mathbb{F}$ by

$$D(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n) = \det(\mathbf{v}_1 \mathbf{v}_2 \ldots \mathbf{v}_n).$$

Proposition 7.26. *The mapping D is n-multilinear.*

Proof. Fix $\mathbf{a}_1, \ldots, \mathbf{a}_n \in \mathbb{F}^n$, and put $S_j(\mathbf{x}) = D(\mathbf{a}_1, \ldots, \mathbf{a}_{j-1}, \mathbf{x}, \mathbf{a}_{j+1}, \ldots, \mathbf{a}_n)$. We must show that $S_j(\mathbf{x} + \mathbf{y}) = S_j(\mathbf{x}) + S_j(\mathbf{y})$, and $S_j(r\mathbf{x}) = rS_j(\mathbf{x})$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{F}^n$ and $r \in \mathbb{F}$. Let $\mathbf{a}_j = (a_{1j} a_{2j} \ldots a_{nj})^T$. By (5.9),

$$D(\mathbf{a}_1, \ldots, \mathbf{a}_n) = \sum_{\sigma \in S(n)} \operatorname{sgn}(\sigma)\ a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{n\sigma(n)}.$$

Put $\mathbf{x} = (x_{1j}, x_{2j}, \ldots, x_{nj})^T$ and $\mathbf{y} = (y_{1j}, y_{2j}, \ldots, y_{nj})^T$. For each $\sigma \in S(n)$, there exists exactly one index i such that $\sigma(i) = j$. Thus, each term in the expansion of $S_j(\mathbf{x} + \mathbf{y})$ has the form

$$\operatorname{sgn}(\sigma)\ a_{1\sigma(1)} \cdots (x_{i\sigma(i)} + y_{i\sigma(i)}) \cdots a_{n\sigma(n)}.$$

This shows that $S_j(\mathbf{x} + \mathbf{y}) = S_j(\mathbf{x}) + S_j(\mathbf{y})$. Similarly, $S_j(r\mathbf{x}) = rS_j(\mathbf{x})$. Therefore, D is n-multilinear. □

Example 7.17. Let us try an example. Suppose $A = \begin{pmatrix} 1 & 3 \\ 2 & 1 \end{pmatrix}$. Then,

$$\det(A) = \det \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix} + \det \begin{pmatrix} 1 & 3 \\ 2 & 0 \end{pmatrix} = 1 - 6 = -5. \qquad \boxtimes$$

7.5.4 A characterization of the determinant

The determinant function on $\mathbb{F}^{n\times n}$ is multilinear, has the value 1 on I_n, and $\det(A) = 0$ if two columns of A are equal. In the next result, we will prove that the only function $\mathcal{D} : \mathbb{F}^{n\times n} \to \mathbb{F}$ having these three properties is the determinant.

Proposition 7.27. *Suppose $\mathcal{D} : \mathbb{F}^{n\times n} \to \mathbb{F}$ is a function satisfying the following properties:*

(i) $\mathcal{D}$ is n-multilinear with respect to columns,

(ii) $\mathcal{D}(A) = 0$ if two columns of A coincide; and

(iii) $\mathcal{D}(I_n) = 1$.

Then $\mathcal{D}(A) = \det(A)$ for all $A \in \mathbb{F}^{n\times n}$.

Proof. I claim that for every elementary matrix E, $\mathcal{D}(E) = \det(E)$. First, if E is the elementary matrix obtained by dilating the ith column of I_n by $r \in \mathbb{F}$, then by (i) and (iii), $\mathcal{D}(E) = r\mathcal{D}(I_n) = r = \det(E)$. Suppose E is the elementary matrix obtained by swapping the ith and jth columns of I_n, where $i < j$. Form the matrix $F \in \mathbb{F}^{n\times n}$ whose ith and jth columns are both $\mathbf{e}_i + \mathbf{e}_j$ and whose kth column, for each $k \neq i, j$, is $\mathbf{e}_k$. (Note: in the 2×2 case, F is the all ones matrix.) By (ii), $\mathcal{D}(F) = 0$. Expanding $\mathcal{D}(F)$ using (i) and applying (ii) twice, we get

$$\mathcal{D}(\mathbf{e}_1, \ldots, \mathbf{e}_j, \ldots, \mathbf{e}_i, \ldots, \mathbf{e}_n) + \mathcal{D}(\mathbf{e}_1, \ldots, \mathbf{e}_i, \ldots, \mathbf{e}_j, \ldots, \mathbf{e}_n) = 0.$$

The first term is $\mathcal{D}(E)$, while the second term is $\mathcal{D}(I_n) = 1$, so $\mathcal{D}(E) = -1 = \det(E)$. Finally, if E is the transvection that adds a multiple of the ith column of I_n to its jth column, then $\mathcal{D}(E) = 1 = \det(E)$ by (i), (ii), and (iii).

The next step is to show that for every elementary matrix E and $A \in \mathbb{F}^{n\times n}$, $\mathcal{D}(AE) = \mathcal{D}(E)\mathcal{D}(A)$. (Recall that column operations are done via right multiplication.) This follows from the previous proposition if E is a row dilation. If E is a column swap, an argument similar to showing that $\mathcal{D}(E) = -1$ shows that $\mathcal{D}(AE) + \mathcal{D}(A) = 0$. Therefore, $\mathcal{D}(AE) = -\mathcal{D}(A) = \mathcal{D}(E)\mathcal{D}(A)$. Likewise, $\mathcal{D}(AE) = \mathcal{D}(A) = \mathcal{D}(E)\mathcal{D}(A)$ when E is a transvection. This completes the second step.

Now suppose $A \in \mathbb{F}^{n\times n}$ has rank n. Then there exist elementary matrices $E_1, \ldots, E_k$ such that $A = E_1 \cdots E_k$. Applying $\mathcal{D}(AE) = \mathcal{D}(E)\mathcal{D}(A)$ and using the product formula (Theorem 5.1) for the determinant, one gets

$$\mathcal{D}(A) = \mathcal{D}(E_1)\cdots\mathcal{D}(E_k) = \det(E_1)\cdots\det(E_k) = \det(A).$$

On the other hand, if A has rank less than n, then the column reduced form of A has a column of zeros, so for suitable elementary matrices $E_1, \ldots, E_k$, $AE_1 \cdots E_k$ has a column of zeros. Thus, $\mathcal{D}(A) = 0 = \det(A)$. This completes the proof that $\mathcal{D}(A) = \det(A)$ for all $A \in \mathbb{F}^{n\times n}$. □

Notice that what the proposition shows is that two functions $\mathcal{D}_1$ and $\mathcal{D}_2$ on $\mathbb{F}^{n\times n}$ satisfying (i)–(iii) have to coincide. But it cannot actually be used as a definition of the determinant.

Exercises

Exercise 7.5.1. Let V be a finite-dimensional inner product space. Show how to define an inner product on V^* in two ways:

(i) using a basis and the dual basis;

(ii) without appealing to a basis.

Exercise* 7.5.1. Let V be any finite-dimensional vector space. Define the *double dual* V^{**} of V to be $(V^*)^*$. That is, V^{**} is the dual of the dual space of V. Show that the map $\Delta : V \to V^{**}$ defined by the condition

$$\Delta(\mathbf{v})(\varphi) = \varphi(\mathbf{v}) \tag{7.19}$$

for all $\mathbf{v} \in V$ and $\varphi \in V^*$ is an isomorphism. Thus Δ is a natural isomorphism from V onto V^{**}.

Exercise* 7.5.2. Let V and W be a pair of finite-dimensional vector spaces over $\mathbb{F}$ and let $T : V \to W$ be linear. Define the *adjoint* map $T^* : W^* \to V^*$ by

$$T^*(\omega)(\mathbf{v}) = \omega(T(\mathbf{v}))$$

for all $\omega \in W^*$ and $\mathbf{v} \in V$.

(*i*) Show that T^* is a well-defined linear map.

(*ii*) Suppose $\mathbf{v}_1, \ldots \mathbf{v}_n$ is a basis of V, and $\mathbf{w}_1, \ldots, \mathbf{w}_m$ is a basis of W, and let the matrix of T with respect to these bases be A. Find the matrix of T^* with respect to the dual bases of W^* and V^*.

(*iii*) Show that if T is injective, then T^* is surjective. Also show the reverse: if T is surjective, then T^* is injective.

(*iv*) Show that $\dim \mathrm{im}(T) = \dim \mathrm{im}(T^*)$.

(*v*) Show that if V is a subspace of W, then there exists a natural surjective linear map $S : W^* \to V^*$.

Chapter 8
Eigentheory

Suppose V is a finite-dimensional vector space over $\mathbb{F}$ and $T : V \to V$ is a linear mapping. An *eigenpair* for T consists of a pair $(\lambda, \mathbf{v})$, where $\lambda \in \mathbb{F}$ and $\mathbf{v} \in V$ is a nonzero vector such that $T(\mathbf{v}) = \lambda\mathbf{v}$. The scalar λ is called an *eigenvalue* of T, and $\mathbf{v}$ is called an *eigenvector* of T corresponding to λ. We will say that the linear mapping T is semisimple if there exist eigenpairs $(\lambda_1, \mathbf{v}_1), \ldots, (\lambda_n, \mathbf{v}_n)$ for T such that $\mathbf{v}_1, \ldots, \mathbf{v}_n$ form a basis of V. A basis consisting of eigenvectors is known as an *eigenbasis*, and the problem of finding an eigenbasis (or whether one exists) is an important step in understanding the structure of a linear mapping. The aim of this chapter is to develop the theory of eigenpairs and eigenbases and to eventually obtain a characterization the semisimple linear mappings. We will also introduce several geometric notions, such as the definition of a dynamical system, and we will give a number of group-theoretic applications of eigentheory. One of the nicest applications is the proof of another classical theorem of Euler, which in modern terms says that $SO(3, \mathbb{R})$ consists of all rotations of $\mathbb{R}^3$ about the origin. We will also discuss the symmetries of the Platonic solids, and finally, we will prove the celebrated Cayley–Hamilton theorem, which will be applied later in the proofs of Jordan canonical form and the Jordan–Chevalley decomposition theorem.

8.1 The Eigenvalue Problem and the Characteristic Polynomial

The purpose of this section is to introduce the basic terms and concepts connected with eigentheory: eigenvalues, eigenvectors, eigenpairs, eigenspaces, the characteristic polynomial, and the characteristic equation. As above, V

J.B. Carrell, *Groups, Matrices, and Vector Spaces*,
DOI 10.1007/978-0-387-79428-0_8

is a finite-dimensional vector space over $\mathbb{F}$, and $T : V \to V$ is a linear mapping. When $V = \mathbb{F}^n$, then $T = T_A$, where $A \in \mathbb{F}^{n\times n}$ is the matrix of T with respect to the standard basis.

8.1.1 First considerations: the eigenvalue problem for matrices

The problem of finding eigenpairs for a linear mapping T will be attacked via matrix theory by solving it for the matrix A of T with respect to a basis of V and then showing how an eigenpair for A determines an eigenpair for T. Suppose $\dim V = n$, so that $A \in \mathbb{F}^{n\times n}$. On the level of matrix theory, the eigenvalue problem is to find the values of $\lambda \in \mathbb{F}$ such that the linear system $A\mathbf{x} = \lambda\mathbf{x}$ has a nontrivial solution. The variables are λ and $\mathbf{x}$, so this is a nonlinear problem because of the $\lambda\mathbf{x}$ term. The way to circumvent this difficulty is by breaking the problem down into two separate problems. The first is to determine all $\lambda \in \mathbb{F}$ such that $\mathcal{N}(A - \lambda I_n) \neq \{\mathbf{0}\}$. Since the null space of a square matrix is nonzero if and only if its determinant is zero, the problem is to determine all $\lambda \in \mathbb{F}$ such that

$$\det(A - \lambda I_n) = 0. \tag{8.1}$$

One calls (8.1) the *characteristic equation* of A. The *eigenvalues* of A are the solutions λ in $\mathbb{F}$ of the characteristic equation. The second problem, which is straightforward, is to find the null space $\mathcal{N}(A - \lambda I_n)$ corresponding to an eigenvalue. The null space $\mathcal{N}(A - \lambda I_n)$ is called the *eigenspace* of A corresponding to λ and denoted by $E_\lambda(A)$. For every nonzero $\mathbf{v} \in \mathcal{N}(A - \lambda I_n)$, $(\lambda, \mathbf{v})$ is an eigenpair.

Remark. The properties of the determinant imply that $\det(A - \lambda I_n)$ is a polynomial in λ of degree n. Hence finding eigenvalues of A requires finding the roots of a polynomial. If $n > 2$, there is no easy way to do this, but there are a few remarks we can and will make later. The roots of polynomials of degree at most four can be found with great difficulty by radicals, but there is no general formula for roots of a polynomial of degree five or more. This is a famous result in Galois theory involving the Galois group of the equation. There are analytic techniques for approximating roots. On the other hand, the fundamental theorem of algebra guarantees that every $A \in \mathbb{C}^{n\times n}$ has n complex eigenvalues. This does not mean, however, that square matrices over $\mathbb{C}$ always admit an eigenbasis, as we will see.

Let us now consider an example.

Example 8.1. Consider the real matrix

$$A = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}.$$

The real eigenvalues of A are the real numbers λ such that

$$A - \lambda I_2 = \begin{pmatrix} 1-\lambda & 2 \\ 2 & 1-\lambda \end{pmatrix}$$

has rank 0 or 1. The characteristic equation of A is

$$\det(A - \lambda I_2) = (1-\lambda)^2 - 2 \cdot 2 = \lambda^2 - 2\lambda - 3 = (\lambda - 3)(\lambda + 1) = 0,$$

so $\lambda = 3, \ -1$ are the eigenvalues. Therefore, we seek $\mathcal{N}(A - 3I_2)$ and $\mathcal{N}(A + I_2)$. Clearly,

$$\mathcal{N}(A - 3I_2) = \mathcal{N}(\begin{pmatrix} -2 & 2 \\ 2 & -2 \end{pmatrix}) = \mathbb{R} \begin{pmatrix} 1 \\ 1 \end{pmatrix},$$

while

$$\mathcal{N}(A + I_2) = \mathcal{N}(\begin{pmatrix} 2 & 2 \\ 2 & 2 \end{pmatrix}) = \mathbb{R} \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

Thus, $(3, \begin{pmatrix} 1 \\ 1 \end{pmatrix})$ and $(-1, \begin{pmatrix} 1 \\ -1 \end{pmatrix})$ are eigenpairs for A. Note that they determine an eigenbasis. ⊠

8.1.2 The characteristic polynomial

Let $A \in \mathbb{F}^{n \times n}$. Instead of dealing with the characteristic equation of A, it is more useful to consider $\det(A - \lambda I_n)$ as a function of λ. Let x be a variable and put $p_A(x) = \det(A - xI_n)$. Then $p_A(x) \in \mathbb{F}[x]$. In calculating $p_A(x)$, one must expand a determinant containing a variable, so it turns out that row operations are not very helpful. This will turn out not to be a problem, because we will soon give a beautiful closed formula for $\det(A - xI_n)$. Let us now note some basic facts about $p_A(x)$.

Proposition 8.1. *If $A \in \mathbb{F}^{n \times n}$, then $p_A(x) = \det(A - xI_n)$ is a polynomial in x over $\mathbb{F}$. Its leading term is $(-1)^n x^n$, so the degree of $p_A(x)$ is n, and its constant term is $\det(A)$. The eigenvalues of A are the roots of $p_A(x) = 0$ in $\mathbb{F}$. In particular, an $n \times n$ matrix cannot have more than n eigenvalues.*

Proof. That $p_A(x)$ is a polynomial is a consequence of the definition of the determinant. The constant term is $p_A(0) = \det(A)$. The leading term comes

from the term

$$(a_{11} - x) \cdots (a_{nn} - x)$$

in the Leibniz expansion, which is clearly $(-1)^n x^n$. The last claim follows from the fact that a polynomial over $\mathbb{F}$ of degree n cannot have more than n roots in $\mathbb{F}$. □

One calls $p_A(x)$ the *characteristic polynomial* of A. The characteristic polynomial needn't have any roots in $\mathbb{F}$. However, one of the basic results in the theory of fields is that given $f(x) \in \mathbb{F}[x]$, there exists a field $\mathbb{F}'$ containing $\mathbb{F}$ such that $f(x)$ factors into linear terms in $\mathbb{F}'[x]$. That is, $\mathbb{F}'$ contains all roots of $f(x) = 0$. We will give a proof of this result in the appendix (see Section 8.7) at the end of this chapter. Although the fundamental theorem of algebra (Theorem 2.27) guarantees that the characteristic polynomial of an $n \times n$ matrix over $\mathbb{R}$ has n complex roots, none of these roots need be real, as the next example points out.

Example 8.2. The characteristic polynomial of the matrix

$$J = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

is $x^2 + 1 = 0$. Hence J is a real matrix that has no real eigenvalues. In fact, there cannot be real eigenvalues, since J is the rotation of $\mathbb{R}^2$ through $\pi/2$: no nonzero vector is rotated into a multiple of itself. On the other hand, if J is treated as a 2×2 complex matrix with eigenvalues $\pm i$, solving for corresponding eigenvectors gives eigenpairs $(i, (-1, i)^T)$ and $(-i, (1, i)^T)$. Thus J has two $\mathbb{C}$-eigenvalues and two independent eigenvectors in $\mathbb{C}^2$. In particular, J has an eigenbasis for $\mathbb{C}^2$. Thus,

$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ -i & i \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ -i & i \end{pmatrix} \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix},$$

so $J = MDM^{-1}$, where $M = \begin{pmatrix} 1 & 1 \\ -i & i \end{pmatrix}$. ⊠

The following example illustrates another possibility.

Example 8.3. Let $K = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}$. The characteristic polynomial of K is $x^2 - 1$, so the eigenvalues of K are ± 1. Thus K is a complex matrix with real eigenvalues. Notice that $K = iJ$, which explains why its eigenvalues are i times those of J. ⊠

The next Proposition gives an important property of the characteristic polynomial.

Proposition 8.2. *Similar matrices have the same characteristic polynomial.*

Proof. Suppose A and B are similar, say $B = MAM^{-1}$. Then

$$\begin{aligned}\det(B - xI_n) &= \det(MAM^{-1} - xI_n)\\ &= \det(M(A - xI_n)M^{-1})\\ &= \det(M)\det(A - xI_n)\det(M^{-1}).\end{aligned}$$

Since $\det(M^{-1}) = \det(M)^{-1}$, the proof is done. □

On the other hand, two matrices with the same characteristic polynomial are not necessarily similar. Because of this proposition, we can extend the definition of the characteristic polynomial of a matrix to the characteristic polynomial of an arbitrary linear mapping $T : V \to V$, provided V is finite-dimensional.

Definition 8.1. If V is a finite-dimensional vector space and $T : V \to V$ is linear, then we define the *characteristic polynomial* of T to be the polynomial $p_T(x)$ defined as $p_A(x)$ for each matrix A of T.

The eigenvalues of T are roots of its characteristic polynomial. The connection between the eigentheory for linear mappings and matrices will be made explicit in Proposition 8.4.

8.1.3 The characteristic polynomial of a 2 × 2 matrix

If A is of size 2×2, say

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix},$$

then the characteristic polynomial $p_A(x)$ of A is easy to find:

$$p_A(x) = (a - x)(d - x) - bc = x^2 - (a + d)x + (ad - bc).$$

As already noted, the constant term is $\det(A)$. The coefficient $a + d$ of x, that is, the sum of the diagonal entries, is the trace of A, that is, $\mathrm{Tr}(A)$. Hence,

$$p_A(x) = x^2 - \mathrm{Tr}(A)x + \det(A). \tag{8.2}$$

The quadratic formula gives an elegant formula for the eigenvalues λ:

$$\lambda = \frac{1}{2}\big(\mathrm{Tr}(A) \pm \sqrt{\mathrm{Tr}(A)^2 - 4\det(A)}\ \big). \tag{8.3}$$

Hence if $A \in \mathbb{R}^{2\times 2}$, it has real eigenvalues if and only if

$$\Delta(A) := \mathrm{Tr}(A)^2 - 4\det(A) = (a-d)^2 + 4bc \geq 0.$$

In particular, if $bc \geq 0$, then A has real eigenvalues. If $\Delta(A) > 0$, the roots are real and unequal, and if $\Delta(A) = 0$, they are real and identical. If $\Delta(A) < 0$, the roots are complex and unequal. In this case, the roots are conjugate complex numbers, since $p_A(x)$ has real coefficients.

Example 8.4. For example, if $A \in \mathbb{R}^{2\times 2}$ is symmetric, then since $b = c$, A has two real eigenvalues. If A is skew-symmetric (that is, $A^T = -A$) and $A \neq O$, then $p_A(x)$ has two unequal complex roots. ⊠

If $p_A(x) = 0$ has roots λ_1, λ_2, then $p_A(x)$ factors as

$$p_A(x) = (x-\lambda_1)(x-\lambda_2) = x^2 - (\lambda_1+\lambda_2)x + \lambda_1\lambda_2,$$

so a comparison of the coefficients gives the following:
(i) the trace of A is the sum of the eigenvalues of A:

$$\mathrm{T}r(A) = a + d = \lambda_1 + \lambda_2,$$

and
(ii) the determinant of A is the product of the eigenvalues of A:

$$\det(A) = ad - bc = \lambda_1\lambda_2.$$

Thus the characteristic polynomial of a 2×2 matrix can be calculated without pencil and paper. Our next task is to give a general formula extending the 2×2 case.

8.1.4 A general formula for the characteristic polynomial

As mentioned above, row operations are essentially of no use if one wants to find a characteristic polynomial by hand. The Laplace expansion is, in general, the only tool that obviates the need to resort to Leibniz's definition. It turns out, however, that there is a beautiful formula for the characteristic polynomial that reduces the computation to computing the principal minors of A.

Let $A \in \mathbb{F}^{n\times n}$. Since $p_A(x)$ is a polynomial in x of degree n with leading coefficient $(-1)^n x^n$ and constant term $\det(A)$, one can write

$$p_A(x) = (-1)^n x^n + (-1)^{n-1}\sigma_1(A)x^{n-1} + (-1)^{n-2}\sigma_2(A)x^{n-2} +$$

$$+ \cdots + (-1)\sigma_{n-1}(A)x + \det(A), \tag{8.4}$$

where the $\sigma_i(A)$, $1 \le i \le n-1$, are scalars given by the next result.

Theorem 8.3. *The coefficients $\sigma_i(A)$ for $1 \le i \le n$ are given by*

$$\sigma_i(A) := \sum \big(\textit{all principal } i \times i \textit{ minors of } A\big), \tag{8.5}$$

where the principal $i \times i$ minors of A are defined to be the determinants of the $i \times i$ submatrices of A obtained by deleting $n - i$ rows of A and then the same $n - i$ columns.

We will omit the proof, since it would require us to take a lengthy detour through exterior algebra. Note that by definition, the principal 1×1 minors are just the diagonal entries of A. Hence

$$\sigma_1(A) = a_{11} + a_{22} + \cdots + a_{nn},$$

so

$$\sigma_1(A) = \mathrm{Tr}(A).$$

Of course, $\sigma_n(A) = \det(A)$. In general, the number of $j \times j$ minors of A is the binomial coefficient $\binom{n}{n-j} = \frac{n!}{j!(n-j)!}$. Thus, the characteristic polynomial of a 4×4 matrix will involve four 1×1 principal minors, six 2×2 principal minors, four 3×3 principal minors, and a single 4×4 principal minor, the determinant. In all, there are 2^n terms involved, since by the Binomial theorem, $(1+1)^n = \sum_{j=0}^{n} \binom{n}{j}$.

Now suppose $\lambda_1, \ldots, \lambda_n$ are the roots of $p_A(x) = 0$. Then

$$\begin{aligned} p_A(x) &= (-1)^n\big((x-\lambda_1)(x-\lambda_2)\cdots(x-\lambda_n)\big) \\ &= (-1)^n x^n + (-1)^{n-1}(\lambda_1 + \lambda_2 + \cdots + \lambda_n)x^{n-1} + \cdots + \lambda_1\lambda_2\cdots\lambda_n. \end{aligned}$$

This generalizes the result of the 2×2 case. For example,

$$\sigma_1(A) = \sigma_1(\lambda_1, \ldots, \lambda_n) = \lambda_1 + \lambda_2 + \cdots + \lambda_n,$$

while

$$\sigma_n(A) = \sigma_n(\lambda_1, \ldots, \lambda_n) = \lambda_1\lambda_2\cdots\lambda_n.$$

Thus the trace of a matrix A is the sum of the roots of its characteristic polynomial, and its determinant is the product of its roots. The other functions $\sigma_i(\lambda_1, \ldots, \lambda_n) = \sigma_i(A)$ can be expressed in a similar manner. For example,

$$\sigma_2(A) = \sigma_2(\lambda_1, \ldots, \lambda_n) = \sum_{i<j} \lambda_i \lambda_j.$$

The functions $\sigma_i(\lambda_1, \ldots, \lambda_n)$ are called the elementary symmetric functions (symmetric because they remain unchanged after an arbitrary permutation of $\lambda_1, \ldots, \lambda_n$). It turns out that all the coefficients $\sigma_i(A)$ of $p_A(x)$ can be expressed in terms of the traces of powers of A. (This is a fact about symmetric functions due to Newton.) Consequently, there exist formulas for the characteristic polynomial that avoid determinants altogether. For example, if A is of size 3×3, then

$$\sigma_2(A) = \frac{1}{2}(\mathrm{Tr}(A)^2 - \mathrm{Tr}(A^2)),$$

while

$$\det(A) = \mathrm{Tr}(A)^3 + 2\mathrm{Tr}(A^3) - 3\mathrm{Tr}(A)\mathrm{Tr}(A^2).$$

Consequently,

$$p_A(x) = -x^3 + \mathrm{Tr}(A)x^2 - \frac{1}{2}(\mathrm{Tr}(A)^2 - \mathrm{Tr}(A^2))x$$

$$+\mathrm{Tr}(A)^3 + 2\mathrm{Tr}(A^3) - 3\mathrm{Tr}(A)\mathrm{Tr}(A^2).$$

One can find $\sigma_i(A)$ for all i from the determinantal formula

$$\sigma_i(A) = \frac{1}{i!} \det \begin{pmatrix} \mathrm{Tr}(A) & i-1 & 0 & \cdots \\ \mathrm{Tr}(A^2) & \mathrm{Tr}(A) & i-2 & \cdots \\ \vdots & \vdots & & \vdots \\ \mathrm{Tr}(A^{i-1}) & \mathrm{Tr}(A^{i-2}) & \cdots & 1 \\ \mathrm{Tr}(A^i) & \mathrm{Tr}(A^{i-1}) & \cdots & \mathrm{Tr}(A) \end{pmatrix}.$$

Exercises

Exercise 8.1.1. Prove the following: Suppose A is a square matrix over $\mathbb{F}$ and $(\lambda, \mathbf{v})$ is an eigenpair for A. Then for every scalar $r \in \mathbb{F}$, $(r\lambda, \mathbf{v})$ is an eigenpair for rA. Moreover, for every positive integer k, $(\lambda^k, \mathbf{v})$ is an eigenpair for A^k. Finally, A has an eigenpair of the form $(0, \mathbf{v})$ if and only if $\mathcal{N}(A)$ is nontrivial.

Exercise 8.1.2. This exercise describes the rational root test, which is used for finding rational roots of polynomials with integer coefficients. The rational root test says that the only rational roots of a polynomial

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

with integer coefficients can be expressed as p/q, where $(p, q) = 1$, q divides a_n, and p divides a_0.

(i) Give a proof of the rational root test.

(ii) Conclude that if A is a square matrix with integer entries, then the only possible rational eigenvalues are the integers that divide $\det(A)$.

(iii) Using the rational root test, find all integral eigenvalues of the matrix $A = \begin{pmatrix} 3 & -2 & -2 \\ 3 & -1 & -3 \\ 1 & -2 & 0 \end{pmatrix}$.

Exercise 8.1.3. Find the characteristic polynomial, eigenvalues, and if possible, a real eigenbasis for:

(i) the X-files matrix

$$X = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix},$$

(ii) the checkerboard matrix

$$C = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix},$$

(iii) the 4×4 X-files matrix

$$\begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix},$$

(iv) the 4×4 checkerboard matrix

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

Exercise 8.1.4. Find the characteristic polynomial and eigenvalues of

$$\begin{pmatrix} -3 & 0 & -4 & -4 \\ 0 & 2 & 1 & 1 \\ 4 & 0 & 5 & 4 \\ -4 & 0 & -4 & -3 \end{pmatrix}$$

in two ways, one using the Laplace expansion and the other using principal minors.

Exercise 8.1.5. The following matrix A was on a blackboard in the movie *Good Will Hunting*:

$$A = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 2 & 1 \\ 0 & 2 & 0 & 0 \\ 1 & 1 & 0 & 0 \end{pmatrix}.$$

Find the characteristic polynomial of A and show that there are four real eigenvalues.

Exercise 8.1.6. Using the determinantal formula for $\sigma_i(A)$, find a formula for $p_A(x)$ when A is of size 4×4.

Exercise 8.1.7. Find the characteristic polynomial of a 4×4 matrix A if you know that three eigenvalues of A are ± 1 and 2 and that $\det(A) = 6$.

Exercise 8.1.8. Suppose $A \in \mathbb{F}^{n \times n}$ has the property that $A = A^{-1}$. Show that if λ is an eigenvalue of A, then so is λ^{-1}. Use this to find the characteristic polynomial of A^{-1} in terms of the characteristic polynomial of A.

Exercise 8.1.9. Show that two similar matrices have the same trace and determinant.

Exercise 8.1.10. True or false: Two matrices with the same characteristic polynomial are similar. If false, supply a 2×2 counter example.

Exercise 8.1.11. If A is a square matrix, determine whether A and A^T have the same characteristic polynomial, hence the same eigenvalues.

Exercise 8.1.12. Show that 0 is an eigenvalue of A if and only if A is singular.

Exercise 8.1.13. True or false: If λ is an eigenvalue of A and μ is an eigenvalue of B, then $\lambda + \mu$ is an eigenvalue of $A + B$. If false, supply a 2×2 counter example.

Exercise 8.1.14. Suppose A and B are similar and $\mathbf{v}$ is an eigenvector of A. Find an eigenvector of B.

Exercise 8.1.15. Let A be a real 3×3 matrix such that A and $-A$ are similar. Show that:

(*i*) $\det(A) = \mathrm{Tr}(A) = 0$,

(*ii*) 0 is an eigenvalue of A, and

(*ii*) if some eigenvalue of A is nonzero, then A has an eigenbasis for $\mathbb{C}^3$.

Exercise 8.1.16. Let A be a matrix whose characteristic polynomial has the form $-x^3 + 7x^2 - bx + 8$. Suppose that the eigenvalues of A are integers.

(*i*) Find the eigenvalues of A.

(*ii*) Find the value of b.

Exercise 8.1.17. Let $A \in \mathbb{F}^{n\times n}$. What is the characteristic polynomial of A^3 in terms of that of A?

Exercise 8.1.18. An $n \times n$ matrix such that $A^k = O$ for some positive integer k is called *nilpotent*.

(*i*) Show all eigenvalues of a nilpotent matrix A are 0.

(*ii*) Conclude that the characteristic polynomial of A is $(-1)^n\lambda^n$. In particular, the trace of a nilpotent matrix is 0.

(*iii*) Find a 3×3 matrix A such that $A^2 \neq O$ but $A^3 = O$. (Hint: look for an upper triangular example.)

Exercise 8.1.19. Let the field be $\mathbb{F}_2$. Find the characteristic polynomials of the following X-matrices. Is either X-matrix similar to a diagonal matrix over $\mathbb{F}_2$?

$$X_1 = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}, \qquad X_2 = \begin{pmatrix} 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

Exercise 8.1.20. Show that the complex eigenvalues of a real $n \times n$ matrix occur in conjugate pairs λ and $\overline{\lambda}$. (Note: the proof of this that we gave for $n = 2$ does not extend. First show that if $p(x)$ is a polynomial with real coefficients, then $\overline{p(z)} = p(\overline{z})$ for every $z \in \mathbb{C}$.)

Exercise 8.1.21. Conclude from the previous exercise that a real $n \times n$ matrix, where n is odd, has at least one real eigenvalue. In particular, every 3×3 real matrix has a real eigenvalue.

Exercise 8.1.22. Show that the only possible real eigenvalues of an $n \times n$ real orthogonal matrix are ± 1.

Exercise 8.1.23. Find eigenpairs for the two complex eigenvalues of the rotation matrix R_θ if $\theta \neq 0, \pi$.

Exercise 8.1.24. Let $\mathbb{F}$ be a field.

(*i*) Show that $\mathbb{F}$ is a one-dimensional vector space over itself.

(*ii*) Show that every linear mapping $T : \mathbb{F} \to \mathbb{F}$ is semisimple. That is, show that there exists $\lambda \in \mathbb{F}$ such that $T(\mathbf{v}) = \lambda\mathbf{v}$ for all $\mathbf{v} \in \mathbb{F}$.

Exercise 8.1.25. Let

$$J = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

Show that J determines a linear mapping $T : \mathbb{C} \to \mathbb{C}$ for every θ and find the unique complex eigenvalue of T.

Exercise 8.1.26. Suppose $A, B \in \mathbb{F}^{n\times n}$ and assume that A is invertible. Show that B, AB, and BA all have the same characteristic polynomial.

Exercise 8.1.27. Find formulas for the elementary symmetric functions

$$\sigma_i(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$$

$(i = 1, 2, 3, 4)$ by expanding $(x - \lambda_1)(x - \lambda_2)(x - \lambda_3)(x - \lambda_4)$. Deduce an expression for all $\sigma_i(A)$ for an arbitrary 4×4 matrix A.

Exercise 8.1.28. * Assume that a, b, c are real and consider the matrix

$$A = \begin{pmatrix} a & b & c \\ b & c & a \\ c & a & b \end{pmatrix}.$$

(i) Find the characteristic polynomial of A.

(ii) Show that $\mathrm{Tr}(A)$ is an eigenvalue of A.

(iii) Use (ii) to show that $\det(A) \leq 0$ and conclude that $(a^3 + b^3 + c^3) \geq 3abc$.

8.2 Basic Results on Eigentheory

We now return to the general problem of understanding the eigentheory of a linear mapping $T : V \to V$. We have already shown how to obtain eigenpairs for a matrix, but it remains to show, first, how an eigenpair for a matrix determines an eigenpair for the linear mapping that the matrix represents, and second, how an eigenbasis for a matrix determines an eigenbasis for this linear mapping. We will also mention a sufficient condition for the existence of an eigenbasis that will be generalized in the next section. Finally, we will give a couple of nice applications of eigentheory. In particular, we will introduce the notion of a dynamical system and use it to study the Fibonacci sequence.

8.2.1 Eigenpairs for linear mappings

Throughout this section, V will denote a finite-dimensional vector space over a field $\mathbb{F}$, and $T : V \to V$ will be a linear mapping. Our first objective is to explain the correspondence between eigenpairs for matrices and eigenpairs for linear mappings. Suppose $\dim V = n$.

Claim. Assume that $A = \mathcal{M}_{\mathcal{B}}^{\mathcal{B}}(T)$ is the matrix of T with respect to a basis $\mathcal{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ of V. Let $(\mu, \mathbf{x})$ be an eigenpair for A, where $\mathbf{x} = (x_1, \ldots, x_n)^T \in \mathbb{F}^n$. If

$$\mathbf{v} = \sum_{i=1}^{n} x_i \mathbf{v}_i = (\mathbf{v}_1 \ \mathbf{v}_2 \cdots \mathbf{v}_n)\mathbf{x},$$

then $(\lambda, \mathbf{v})$ is an eigenpair for T.

Proof. Recall that by (7.15), A is determined from

$$(T(\mathbf{v}_1) \ T(\mathbf{v}_2) \cdots T(\mathbf{v}_n)) = (\mathbf{v}_1 \ \mathbf{v}_2 \cdots \mathbf{v}_n)A.$$

Since $A\mathbf{x} = \mu\mathbf{x}$,

$$\begin{aligned} T(\mathbf{v}) &= (T(\mathbf{v}_1) \ T(\mathbf{v}_2) \cdots T(\mathbf{v}_n))\mathbf{x} \\ &= (\mathbf{v}_1 \ \mathbf{v}_2 \cdots \mathbf{v}_n)A\mathbf{x} \\ &= \mu(\mathbf{v}_1 \ \mathbf{v}_2 \cdots \mathbf{v}_n)\mathbf{x} \\ &= \mu\mathbf{v}. \end{aligned}$$

Thus $(\mu, \mathbf{v})$ is indeed an eigenpair for T. Summarizing, we have the following result.

Proposition 8.4. *Let $T : V \to V$ be linear, and let $A = \mathcal{M}_{\mathcal{B}}^{\mathcal{B}}(T)$ be the matrix of T with respect to a basis $\mathcal{B}$ of V. Then every eigenpair $(\mu, \mathbf{x})$ for A gives an eigenpair $(\mu, \mathbf{v})$ for T, where $\mathbf{v}$ is the element of V whose coordinates with respect to $\mathcal{B}$ are the components of $\mathbf{x}$. Conversely, an eigenpair $(\mu, \mathbf{v})$ for T gives a corresponding eigenpair $(\mu, \mathbf{x})$ for A in the same manner.*

8.2.2 Diagonalizable matrices

Recall from Example 8.1 that the matrix $\begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$ has eigenvalues 3 and -1 and corresponding eigenspaces $\mathbb{R}\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ and $\mathbb{R}\begin{pmatrix} 1 \\ -1 \end{pmatrix}$. When this data is encoded in a matrix equation, we see that

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

This expression has the form $AP = PD$, where the columns of P are linearly independent eigenvectors and the entries of D are the corresponding eigenvalues. Since P is invertible, we get the factorization $A = PDP^{-1}$, where D is diagonal. Thus, A is similar to a diagonal matrix over $\mathbb{F}$. Let us now introduce the following term.

Definition 8.2. A matrix $A \in \mathbb{F}^{n\times n}$ is said to be *diagonalizable* over $\mathbb{F}$ if A can be written $A = PDP^{-1}$, where $D, P \in \mathbb{F}^{n\times n}$ and D is diagonal. In other words, A is diagonalizable over $\mathbb{F}$ if it is similar over $\mathbb{F}$ to a diagonal matrix $D \in \mathbb{F}^{n\times n}$.

We now describe what it means to say that a matrix is diagonalizable. Assume $A \in \mathbb{F}^{n\times n}$.

Proposition 8.5. *Suppose $\mathbf{w}_1, \ldots, \mathbf{w}_n$ is an eigenbasis of $\mathbb{F}^n$ for A with corresponding eigenvalues $\lambda_1, \ldots, \lambda_n$ in $\mathbb{F}$. Then A is diagonalizable over $\mathbb{F}$. In fact, $A = PDP^{-1}$, where $P = (\mathbf{w}_1 \ \cdots \ \mathbf{w}_n)$ and $D = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$. Conversely, if $A = PDP^{-1}$, where P and D are in $\mathbb{F}^{n\times n}$, P is invertible, and D is diagonal, then the columns of P are an eigenbasis of $\mathbb{F}^n$ for A, and the diagonal entries of D are the corresponding eigenvalues. That is, if the ith column of P is $\mathbf{w}_i$, then $(\lambda_i, \mathbf{w}_i)$ is an eigenpair for A.*

Proof. Let $\mathbf{w}_1, \ldots, \mathbf{w}_n$ be an eigenbasis, and put $P = (\mathbf{w}_1 \ \cdots \ \mathbf{w}_n)$. Then

$$AP = (A\mathbf{w}_1 \ \cdots \ A\mathbf{w}_n) = (\lambda_1\mathbf{w}_1 \ \cdots \ \lambda_n\mathbf{w}_n) = (\mathbf{w}_1 \ \cdots \ \mathbf{w}_n)D, \tag{8.6}$$

where $D = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$. Thus $AP = PD$, so A is diagonalizable over $\mathbb{F}$. The converse is proved in a similar manner. □

Recall, a linear mapping having an eigenbasis is called semisimple. Finding an eigenbasis for a linear mapping reduces to the problem of diagonalizing its matrix.

Proposition 8.6. *Suppose* $\dim V = n$, $\mathcal{B} = \{\mathbf{v}_1, \ \ldots \ , \mathbf{v}_n\}$ *is a basis of* V *and* $A = \mathcal{M}_{\mathcal{B}}^{\mathcal{B}}(T)$ *is the matrix of* T *with respect to* $\mathcal{B}$. *Then* T *is semisimple if and only if* A *is diagonalizable over* $\mathbb{F}$, *say* $A = PDP^{-1}$, *and an eigenbasis* $\mathbf{w}_1, \ \ldots, \ \mathbf{w}_n$ *for* T *is given by*

$$(\mathbf{w}_1 \ \mathbf{w}_2 \cdots \mathbf{w}_n) = (\mathbf{v}_1 \ \mathbf{v}_2 \cdots \mathbf{v}_n)P. \tag{8.7}$$

Proof. The proof is similar to the last, but we will give it anyway. Assume the notation already in use above, and suppose $A = PDP^{-1}$, where $P, D \in \mathbb{F}^{n \times n}$ and D is diagonal. Then

$$(T(\mathbf{v}_1) \ T(\mathbf{v}_2) \cdots T(\mathbf{v}_n)) = (\mathbf{v}_1 \ \mathbf{v}_2 \cdots \mathbf{v}_n)PDP^{-1},$$

so

$$(T(\mathbf{v}_1) \ T(\mathbf{v}_2) \cdots T(\mathbf{v}_n))P = (\mathbf{v}_1 \ \mathbf{v}_2 \cdots \mathbf{v}_n)PD.$$

Thus if $\mathbf{w}_1, \ldots, \mathbf{w}_n$ are defined by

$$(\mathbf{w}_1 \ \mathbf{w}_2 \cdots \mathbf{w}_n) = (\mathbf{v}_1 \ \mathbf{v}_2 \cdots \mathbf{v}_n)P,$$

then since P is nonsingular, $\mathbf{w}_1, \ldots, \mathbf{w}_n$ form a basis of V such that

$$\begin{aligned} T(\mathbf{w}_1 \ \mathbf{w}_2 \cdots \mathbf{w}_n) &= T(\mathbf{v}_1 \ \mathbf{v}_2 \cdots \mathbf{v}_n)P \\ &= (T(\mathbf{v}_1) \ T(\mathbf{v}_2) \cdots T(\mathbf{v}_n))P \\ &= (\mathbf{v}_1 \ \mathbf{v}_2 \cdots \mathbf{v}_n)(PDP^{-1})P \\ &= (\mathbf{v}_1 \ \mathbf{v}_2 \cdots \mathbf{v}_n)PD \\ &= (\mathbf{w}_1 \ \mathbf{w}_2 \cdots \mathbf{w}_n)D. \end{aligned}$$

Consequently, since D is diagonal, $\mathbf{w}_1, \ldots, \mathbf{w}_n$ form a basis of V such that $T(\mathbf{w}_i) = \lambda_i \mathbf{w}_i$ for each i. Hence T is semisimple. The converse is proved by reversing the argument. □

Consequently, we have the following corollary.

Corollary 8.7. *A linear mapping* $T : V \to V$ *is semisimple if and only if its matrix with respect to some basis of* V *is diagonal, and consequently its matrix with respect to any basis of* V *is diagonalizable.*

8.2.3 A criterion for diagonalizability

The following proposition states a well-known criterion for diagonalizability. Since we will prove a stronger result later, the proof will be omitted. We will also give what amounts to a one-line proof in Example 8.10.

Proposition 8.8. *An $n \times n$ matrix A over $\mathbb{F}$ with n distinct eigenvalues in $\mathbb{F}$ is diagonalizable. More generally, if V is a finite-dimensional vector space over $\mathbb{F}$ and $T : V \to V$ is a linear mapping with $\dim V$ distinct eigenvalues in $\mathbb{F}$, then T is semisimple.*

The criterion for $A \in \mathbb{F}^{n\times n}$ to have distinct eigenvalues, given that its eigenvalues lie in $\mathbb{F}$, is that its characteristic polynomial have simple roots. The multiple root test (Corollary 2.38) applied to $p_A(x)$ therefore gives the following.

Proposition 8.9. *A square matrix A over $\mathbb{F}$ has no repeated eigenvalues in $\mathbb{F}$ if $p_A(x)$ and its derivative $(p_A)'(x)$ have no common roots in $\mathbb{F}$.*

Example 8.5. The counting matrix

$$C = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}$$

has characteristic polynomial $p_C(x) = -x^3 + 15x^2 - 18x$; hence C has three distinct real eigenvalues. It follows that C is diagonalizable. ⊠

Example 8.6. The matrix

$$A = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 2 & 1 \\ 0 & 2 & 0 & 0 \\ 1 & 1 & 0 & 0 \end{pmatrix},$$

which is seen on a blackboard in the movie *Good Will Hunting*, has characteristic polynomial

$$p_A(x) = x^4 - 7x^2 - 2x + 4.$$

Now, -1 is a root, and hence

$$p_A(x) = (x + 1)(x^3 - x^2 - 6x + 4).$$

Since -1 is not a root of $q(x) = x^3 - x^2 - 6x + 4 = 0$, it follows that $p_A(x)$ has four distinct roots as long as $q(x)$ has three distinct roots. But $q'(x) = 3x^2 - 2x - 6$ has roots

$$r = \frac{2 \pm \sqrt{76}}{6},$$

and

$$q(\frac{2+\sqrt{76}}{6}) < 0, \quad \text{while} \quad q(\frac{2-\sqrt{76}}{6}) > 0.$$

Thus q and q' have no roots in common, so q has three distinct roots. Moreover, it follows from the two inequalities above that q has three real roots. (Reason: the coefficient of x^3 is positive.) Thus A has four distinct real eigenvalues; hence A is diagonalizable. ⊠

Note that the *Good Will Hunting* matrix A is real and symmetric. As we will see later, the principal axis theorem therefore guarantees that A is diagonalizable.

8.2.4 The powers of a diagonalizable matrix

In this section, we will consider the powers of a square matrix A. Let us begin with the following observation. Suppose $A \in \mathbb{F}^{n\times n}$ has an eigenpair $(\lambda, \mathbf{v})$ and let k be a positive integer. Since $A^k\mathbf{v} = A^{k-1}(A\mathbf{v}) = A^{k-1}(\lambda\mathbf{v}) = \lambda A^{k-1}\mathbf{v}$, it follows by iteration that

$$A^k\mathbf{v} = \lambda^k\mathbf{v}.$$

Now suppose A can be diagonalized as $A = PDP^{-1}$. Then for every positive integer k,

$$A^k = (PDP^{-1})(PDP^{-1})\cdots(PDP^{-1}) = PD^kP^{-1}.$$

If $D = \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$, then $D^k = \operatorname{diag}\big((\lambda_1)^k, (\lambda_2)^k, \ldots, (\lambda_n)^k\big)$. Let $P = (\mathbf{v}_1\ \mathbf{v}_2\ \cdots\ \mathbf{v}_n)$. Setting $\mathbf{v} = \sum a_i\mathbf{v}_i$, we have

$$A^k\mathbf{v} = \sum a_iA^k\mathbf{v}_i = \sum a_i(\lambda_i)^k\mathbf{v}_i. \tag{8.8}$$

If $A \in \mathbb{R}^{n\times n}$ and all its eigenvalues are real, one can say more:

(*i*) If A has only nonnegative eigenvalues, then A has a kth root for every positive integer k. In fact, $A^{\frac{1}{k}} = PD^{\frac{1}{k}}P^{-1}$.

(*ii*) If none of the eigenvalues of A are 0, then the negative powers of A are found from the formula $A^{-k} = (A^{-1})^k = PD^{-k}P^{-1}$.

(*iii*) If all the eigenvalues λ of A satisfy $|\lambda| < 1$, then $\lim_{m\to\infty} A^m = O$.

More generally, if $A \in \mathbb{C}^{n\times n}$, there are corresponding statements. The reader should attempt to formulate them. For nondiagonalizable matrices, kth roots need not exist.

8.2.5 The Fibonacci sequence as a dynamical system

The *Fibonacci numbers* a_k are defined by the Fibonacci sequence (a_k) as follows. Starting with arbitrary integers a_0 and a_1, put $a_2 = a_0 + a_1$, $a_3 = a_2 + a_1$, and in general, put $a_k = a_{k-1} + a_{k-2}$ if $k \geq 2$. The Fibonacci sequence can also be defined by the matrix identity

$$\begin{pmatrix} a_{k+1} \\ a_k \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} a_k \\ a_{k-1} \end{pmatrix},$$

provided $k \geq 1$. Hence putting

$$F = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix},$$

we obtain that

$$\begin{pmatrix} a_{k+1} \\ a_k \end{pmatrix} = F \begin{pmatrix} a_k \\ a_{k-1} \end{pmatrix} = F^2 \begin{pmatrix} a_{k-1} \\ a_{k-2} \end{pmatrix} = \cdots = F^k \begin{pmatrix} a_1 \\ a_0 \end{pmatrix}.$$

Thus putting

$$\mathbf{v}_k = \begin{pmatrix} a_{k+1} \\ a_k \end{pmatrix},$$

for $k = 0,\ 1,\ 2, \ldots$, we can therefore express the Fibonacci sequence in the form

$$\mathbf{v}_k = F^k \mathbf{v}_0. \tag{8.9}$$

Such a sequence is called the *dynamical system* defined by F. To analyze the Fibonacci sequence, we will diagonalize F. The characteristic equation of F is $x^2 - x - 1 = 0$, so the eigenvalues are

$$\phi = \frac{1+\sqrt{5}}{2}, \quad \mu = \frac{1-\sqrt{5}}{2}.$$

One checks that $\mathcal{N}(F - \phi I_2) = \mathbb{R}(\phi, 1)^T$ and $\mathcal{N}(F - \mu I_2) = \mathbb{R}(\mu, 1)^T$. Therefore, as in the previous example, F is diagonalized by

$$F = \begin{pmatrix} \phi & \mu \\ 1 & 1 \end{pmatrix} \begin{pmatrix} \phi & 0 \\ 0 & \mu \end{pmatrix} \begin{pmatrix} \phi & \mu \\ 1 & 1 \end{pmatrix}^{-1}.$$

Hence, by (8.9),

$$\begin{pmatrix} a_{m+1} \\ a_m \end{pmatrix} = F^m \begin{pmatrix} a_1 \\ a_0 \end{pmatrix} = \begin{pmatrix} \phi & \mu \\ 1 & 1 \end{pmatrix} \begin{pmatrix} \phi^m & 0 \\ 0 & \mu^m \end{pmatrix} \begin{pmatrix} \phi & \mu \\ 1 & 1 \end{pmatrix}^{-1} \begin{pmatrix} a_1 \\ a_0 \end{pmatrix}.$$

Since $\begin{pmatrix} \phi & \mu \\ 1 & 1 \end{pmatrix}^{-1} = \dfrac{1}{\phi - \mu} \begin{pmatrix} 1 & -\mu \\ -1 & \phi \end{pmatrix}$, we get

$$\begin{pmatrix} a_{m+1} \\ a_m \end{pmatrix} = \frac{1}{\phi - \mu} \begin{pmatrix} \phi^{m+1} - \mu^{m+1} & -\mu\phi^{m+1} + \mu^{m+1}\phi \\ \phi^m - \mu^m & \mu\phi^m + \mu^m\phi \end{pmatrix} \begin{pmatrix} a_1 \\ a_0 \end{pmatrix}.$$

For example, if $a_0 = 0$ and $a_1 = 1$, we see that

$$a_m = \frac{\phi^m - \mu^m}{\phi - \mu} = \frac{1}{\sqrt{5} \cdot 2^m}((1 + \sqrt{5})^m - (1 - \sqrt{5})^m). \tag{8.10}$$

Notice that

$$\lim_{m\to\infty} \frac{a_{m+1}}{a_m} = \lim_{m\to\infty} \frac{\phi^{m+1} - \mu^{m+1}}{\phi^m - \mu^m} = \phi,$$

since $\lim_{m\to\infty}(\mu/\phi)^m = 0$. Therefore, for large m, the ratio a_{m+1}/a_m is approximately ϕ. A little further computation gives the precise formulas

$$a_{2m} = \left[\frac{\phi^{2m}}{\sqrt{5}}\right] \quad \text{and} \quad a_{2m+1} = \left[\frac{\phi^{2m+1}}{\sqrt{5}}\right] + 1,$$

where $[r]$ denotes the integer part of the real number r.

The eigenvalue $\phi = \dfrac{1 + \sqrt{5}}{2}$ is the so-called golden ratio. It comes up in many unexpected and interesting ways. For more information, one can consult *The Story of ϕ, the World's Most Astonishing Number*, by Mario Livo.

Exercises

Exercise 8.2.1. Diagonalize the following matrices when possible:

$$A = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}, \quad B = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}, \quad C = \begin{pmatrix} -3 & 0 & -4 & -4 \\ 0 & 2 & 1 & 1 \\ 4 & 0 & 5 & 4 \\ -4 & 0 & -4 & -3 \end{pmatrix}.$$

Exercise 8.2.2. A 4×4 matrix over $\mathbb{R}$ has eigenvalues ± 1, trace 3, and determinant 0. Can A be diagonalized over $\mathbb{R}$? What about over $\mathbb{Q}$?

Exercise 8.2.3. Determine which of the following matrices are diagonalizable over the reals:

$$\begin{pmatrix} 1 & 0 & -1 \\ -1 & 1 & 1 \\ 2 & -1 & -2 \end{pmatrix}, \quad \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & -1 & 1 \end{pmatrix}, \quad \begin{pmatrix} 2 & -1 & 1 \\ 1 & 0 & 1 \\ 1 & -1 & -2 \end{pmatrix}, \quad \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}.$$

Exercise 8.2.4. Find all possible square roots, if any exist, of the following matrices:

$$\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}, \quad \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & -1 \\ 1 & -1 \end{pmatrix}.$$

Exercise 8.2.5. Do the same as in Problem 8.2.4 for the 4×4 all 1's matrix.

Exercise 8.2.6. Compute the nth power of each matrix of Exercise 8.2.4 and also that of the 3×3 all 1's matrix.

Exercise 8.2.7. Let F denote the Fibonacci matrix. Find F^4 in two ways, once directly and once using eigentheory.

Exercise 8.2.8. Assuming $a_0 = 0$ and $a_1 = 1$, find the thirteenth and fifteenth Fibonacci numbers using eigentheory.

Exercise 8.2.9. Show directly that

$$\frac{\phi^m - \mu^m}{\phi - \mu} = \frac{1}{\sqrt{5} \cdot 2^m}((1 + \sqrt{5})^m - (1 - \sqrt{5})^m)$$

is an integer, thus explaining the strange expression in Section 5.1.

Exercise 8.2.10. Let $A = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$. Find A^{10}.

Exercise 8.2.11. Give an example of a 2×2 matrix A over $\mathbb{C}$ that does not have a square root.

Exercise 8.2.12. If possible, find a square root for $A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$.

Exercise 8.2.13. Suppose G is a finite subgroup of $GL(n, \mathbb{C})$ of order m. Show that if $g \in G$, then every eigenvalue of g is an mth root of unity.

8.3 Two Characterizations of Diagonalizability

In this section, we first prove the strong form of the diagonalizability criterion stated in Proposition 8.10. Then we give two characterizations of the diagonalizable matrices (equivalently semisimple linear mappings). The first involves the eigenspace decomposition. The second is more interesting, since it gives a precise criterion for $A \in \mathbb{F}^{n\times n}$ to be diagonalizable in terms of whether A satisfies a certain polynomial equation. This brings in the notion of the minimal polynomial of a matrix.

8.3.1 Diagonalization via eigenspace decomposition

In the previous section, we stated the result that every $A \in \mathbb{F}^{n\times n}$ with n distinct eigenvalues in $\mathbb{F}$ is diagonalizable. We will now extend this result by dropping the assumption that the eigenvalues are distinct. Suppose $\lambda \in \mathbb{F}$ is an eigenvalue of A and $(x-\lambda)^k$ divides $p_A(x)$ for some $k > 1$. Then we say that λ is a repeated eigenvalue of A. We define the *algebraic multiplicity* of λ as the largest value of k such that $(x-\lambda)^k$ divides $p_A(x)$. The *geometric multiplicity* of λ is defined to be $\dim E_\lambda(A)$. It turns out that the algebraic multiplicity of an eigenvalue is always greater than or equal to its geometric multiplicity. (The proof of this will have to wait until Chap. 10.)

Proposition 8.10. *Suppose $\lambda_1, \ldots, \lambda_m \in \mathbb{F}$ are distinct eigenvalues of $A \in \mathbb{F}^{n\times n}$, and choose a set of linearly independent eigenvectors in the eigenspace $E_{\lambda_i}(A)$ for each λ_i, $1 \le i \le m$. Then the union of these linearly independent sets is linearly independent.*

Proof. First, notice that if $i \neq j$, then $E_{\lambda_i}(A) \cap E_{\lambda_j}(A) = \{\mathbf{0}\}$. For each i, let S_i denote a linearly independent subset of $E_{\lambda_i}(A)$, and put $S = \bigcup_{i=1}^m S_i$. Write $S = \{\mathbf{u}_1, \ldots, \mathbf{u}_s\}$ in some way, and suppose

$$\sum a_r \mathbf{u}_r = \mathbf{0}, \tag{8.11}$$

where all $a_r \in \mathbb{F}$. Let M_i, $i = 1, \ldots, m$, denote the set of indices j such that $\mathbf{u}_j \in S_i$, and put $\mathbf{v}_i = \sum_{j\in M_i} a_j \mathbf{u}_j$. Thus, $\mathbf{v}_i \in E_{\lambda_i}(A)$, and by assumption,

$$\sum_{i=1}^m \mathbf{v}_i = \mathbf{0}.$$

We will now show that all $\mathbf{v}_i$ are equal to $\mathbf{0}$. Suppose some $\mathbf{v}_i$ is not equal to $\mathbf{0}$. Without loss of generality, we suppose $i = 1$, so $\mathbf{v}_1 = -\sum_{i>1} \mathbf{v}_i \neq \mathbf{0}$. Now

let W denote the span of $Y = \{\mathbf{v}_2, \ldots, \mathbf{v}_m\}$. By the dimension theorem, we may select $\mathbf{w}_1, \ldots, \mathbf{w}_\ell \in Y$ that form a basis of W. Of course, each $\mathbf{w}_i$ is an eigenvector of A. Let μ_i denote the corresponding eigenvalue, and note that $\lambda_1 \neq \mu_i$ for all i. Write $\mathbf{v}_1 = \sum_i b_i \mathbf{w}_i$. By applying A to $\mathbf{v}_1$, we get $\lambda_1 \mathbf{v}_1 = \sum_i b_i \mu_i \mathbf{w}_i$. By multiplying $\mathbf{v}_1$ by λ_1, we also obtain that $\lambda_1 \mathbf{v}_1 = \sum_i \lambda_1 b_i \mathbf{w}_i$. Subtracting the two expressions for $\lambda_1 \mathbf{v}_1$ gives

$$\sum_i (\lambda_1 - \mu_i) b_i \mathbf{w}_i = \mathbf{0}.$$

But the $\mathbf{w}_i$ are independent, so $(\lambda_1 - \mu_i) b_i = 0$ for all i. Since $\lambda_1 - \mu_i \neq 0$ for all i, it follows that all b_i are zero. Hence $\mathbf{v}_1 = \mathbf{0}$, a contradiction. It follows that all $\mathbf{v}_i$ equal $\mathbf{0}$, so $\sum_{j \in M_i} a_j \mathbf{u}_j = \mathbf{0}$ for all i. Consequently, all a_r are equal to zero in the original expression (8.11). Therefore, S is linearly independent. □

We now state the first characterization of the diagonal matrices. Let $A \in \mathbb{F}^{n \times n}$.

Theorem 8.11. *Let $\lambda_1, \ldots, \lambda_m$ denote the distinct eigenvalues of A in $\mathbb{F}$. Then A is diagonalizable over $\mathbb{F}$ if and only if*

$$\sum_{i=1}^{m} \dim E_{\lambda_i}(A) = n. \tag{8.12}$$

In that case, if $\mathcal{B}_i$ is a basis of $E_{\lambda_i}(A)$, then

$$\mathcal{B} = \bigcup_{1 \leq i \leq m} \mathcal{B}_i$$

is an eigenbasis of $\mathbb{F}^n$, and we have the direct sum decomposition

$$\mathbb{F}^n = E_{\lambda_1}(A) \oplus \cdots \oplus E_{\lambda_m}(A). \tag{8.13}$$

Proof. Suppose (8.12) holds. Then Proposition 8.10 and the dimension theorem imply that A admits an eigenbasis. Therefore, A is diagonalizable. The converse statement is immediate. The rest of the proof amounts to applying results on direct sums, especially Proposition 6.26. □

The following example with repeated eigenvalues is rather fun to analyze.

Example 8.7. Let B denote the 4×4 all-ones matrix

$$B = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix}.$$

Now 0 is an eigenvalue of B. In fact, B has rank 1, so $\dim \mathcal{N}(B) = 3$. Thus the eigenspace $E_0(B)$ of 0 has dimension 3. Every eigenvector for the eigenvalue 0 satisfies the equation $x_1 + x_2 + x_3 + x_4 = 0$. The basic null vectors for this equation are $\mathbf{f}_1 = (-1, 1, 0, 0)^T$, $\mathbf{f}_2 = (-1, 0, 1, 0)^T$, $\mathbf{f}_3 = (-1, 0, 0, 1)^T$, and they give a basis of $\mathcal{N}(B)$. By Proposition 8.10, there is at most one other eigenvalue. The other eigenvalue can be found by inspection by noticing that every row of B adds up to 4. Thus, $\mathbf{f}_4 = (1, 1, 1, 1)^T$ is an eigenvector for $\lambda = 4$. By Proposition 8.10, we now have four linearly independent eigenvectors, hence an eigenbasis. Therefore, B is diagonalizable; in fact, B is similar to $D = \operatorname{diag}(0, 0, 0, 4)$. ⊠

In the above example, the eigenvalues of B were found by being a little clever and using some special properties of B. A more methodical way to find the eigenvalues would have been to compute the characteristic polynomial of B using principal minors. In fact, all principal minors of B are zero except for the 1×1 principal minors, namely the diagonal elements. Hence, $p_B(x) = x^4 - \operatorname{Tr}(A)x^3 = x^4 - 4x^3 = x^3(x - 4)$. Recall also that if all eigenvalues but one are known, the final eigenvalue can be found immediately from the trace.

The following result classifying the semisimple linear mappings follows immediately from Theorem 8.11. Let V be a finite-dimensional vector space over $\mathbb{F}$.

Corollary 8.12. *A linear mapping $T : V \to V$ is semisimple if and only if $\sum_i^m \dim E_{\lambda_i}(T) = \dim V$, where $\lambda_1, \ldots, \lambda_m$ denote the distinct eigenvalues of T.*

8.3.2 A test for diagonalizability

Is there a simple way of determining when a linear mapping is semisimple or when a matrix is diagonalizable? It turns out the answer is yes, as long as all its eigenvalues are known. Here, the algebraic and geometric multiplicities do not play a role. Let $T : V \to V$ be a linear mapping whose distinct eigenvalues (in $\mathbb{F}$) are $\lambda_1, \ldots, \lambda_m$.

Claim: if T is semisimple, then

$$(T - \lambda_1 I_V)(T - \lambda_2 I_V) \cdots (T - \lambda_m I_V) = O. \tag{8.14}$$

Proof. Indeed, if T is semisimple, then

$$V = E_{\lambda_1}(T) \oplus E_{\lambda_2}(T) \cdots \oplus E_{\lambda_m}(T).$$

Thus, to prove (8.14), it suffices to show that

$$(T - \lambda_1 I_V)(T - \lambda_2 I_V) \cdots (T - \lambda_m I_V)\mathbf{v}_i = \mathbf{0},$$

provided $\mathbf{v}_i \in E_{\lambda_i}(T)$ for some i. But this is clear, since

$$(T - \lambda_i I_V)(T - \lambda_j I_V) = (T - \lambda_j I_V)(T - \lambda_i I_V)$$

for all i and j. Thus, the left-hand side of (8.14) can be factored into the form $C(T - \lambda_i I_V)$, and since $C(T - \lambda_i I_V)\mathbf{v}_i = C\mathbf{0} = \mathbf{0}$, (8.14) follows. □

There is another way to prove (8.14). To do so, we need to evaluate a polynomial $f(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$ in $\mathbb{F}[x]$ on a matrix in $\mathbb{F}^{n\times n}$. The value of f at A is defined to be $f(A) = a_n A^n + a_{n-1}A^{n-1} + \cdots + a_1 A + a_0 I_n$. Thus, $f(A) \in \mathbb{F}^{n\times n}$. Similarly, if and $T : V \to V$ is a linear mapping, then by definition,

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

where T^i means T composed with itself i times. Thus, $f(T)$ is also a linear mapping with domain and target V. Note that if $A = PDP^{-1}$, where $A \in \mathbb{F}^{n\times n}$, then $f(A) = f(PDP^{-1}) = Pf(A)P^{-1}$. Thus, if $A = P\text{diag}(\lambda_1, \lambda_2 \ldots, \lambda_n)P^{-1}$ and $g(x) = (x - \lambda_1)(x - \lambda_2)\cdots(x - \lambda_m)$, then

$$g(A) = g(PDP^{-1}) = Pg(D)P^{-1} = P\text{diag}(g(\lambda_1), g(\lambda_2), \ldots, g(\lambda_m))P^{-1}.$$

But $\text{diag}(g(\lambda_1), g(\lambda_2), \ldots, g(\lambda_m)) = O$, so $g(A) = O$ also. □

The useful fact is that the converse is also true. We state and prove that next.

Theorem 8.13. *Suppose the distinct eigenvalues of a linear mapping $T : V \to V$ are $\lambda_1, \ldots, \lambda_m$. Then T is semisimple if and only if*

$$(T - \lambda_1 I_V)(T - \lambda_2 I_V) \cdots (T - \lambda_m I_V) = O. \tag{8.15}$$

Proof. The "only if" implication was just proved. The "if" assertion, which is harder to prove, is based on the following lemma.

Lemma 8.14. *Suppose $T_1, T_2, \ldots, T_N : V \to V$ are linear mappings that satisfy the following properties:*

(i) $T_1 \circ T_2 \circ \cdots \circ T_N = O$,

(ii) $T_i \circ T_j = T_j \circ T_i$ for all indices i, j,

(iii) for each $i = 1, \ldots, N-1$, $\ker(T_i) \cap \ker(T_{i+1} \circ \cdots \circ T_N) = \{\mathbf{0}\}$.

Then $V = \ker(T_1) \oplus \ker(T_2) \oplus \cdots \oplus \ker(T_N)$.

Proof. The proof is based on the following principle. If $P, Q : V \to V$ are linear mappings such that $P \circ Q = O$ and $\ker(P) \cap \ker(Q) = \{\mathbf{0}\}$, then $V = \ker(P) \oplus \ker(Q)$. This is proved by first noting that by the rank–nullity theorem, $\dim V = \dim \ker(Q) + \dim \operatorname{im}(Q)$. Since $P \circ Q = O$, we have $\operatorname{im}(Q) \subset \ker(P)$. Thus $\dim V \leq \dim \ker(Q) + \dim \ker(P)$. Now apply the Grassmann intersection formula, which says that

$$\dim(\ker(P) + \ker(Q)) = \dim \ker(P) + \dim \ker(Q) - \dim(\ker(P) \cap \ker(Q)).$$

Since $\dim(\ker(P) \cap \ker(Q)) = 0$, the previous inequality says that $\dim(\ker(P) + \ker(Q)) \geq \dim V$. But $\ker(P) + \ker(Q)$ is a subspace of V, so $V = \ker(P) + \ker(Q)$. Since $\ker(P) \cap \ker(Q) = \{\mathbf{0}\}$, it follows from Proposition 6.26 that $V = \ker(P) \oplus \ker(Q)$. Letting $P = T_1$ and $Q = T_2 \circ \cdots \circ T_N$, we have shown that $V = \ker(T_1) \oplus \ker(T_2 \circ \cdots \circ T_N)$. Now repeat the argument, replacing V by $\ker(T_2 \circ \cdots \circ T_N)$, P by T_2, and Q by $T_3 \circ \cdots \circ T_N$. This is allowed, since by (ii), both T_2 and $T_3 \circ \cdots \circ T_N$ map $\ker(T_2 \circ \cdots \circ T_N)$ into itself. Thus, $\ker(T_2 \circ \cdots \circ T_N) = \ker(T_2) \oplus \ker(T_3 \circ \cdots \circ T_N)$. Hence

$$V = \ker(T_1) \oplus \big(\ker(T_2) \oplus \ker(T_3 \circ \cdots \circ T_N)\big).$$

It follows that $V = \ker(T_1) \oplus \ker(T_2) \oplus \ker(T_3 \circ \cdots \circ T_N)$. The hypotheses allow us to iterate this argument until the conclusion is reached. □

Clearly, $(T - \lambda_i I_V)(T - \lambda_j I_V) = (T - \lambda_j I_V)(T - \lambda_i I_V)$, so to finish the proof of Theorem 8.13, we just have to check that

$$\ker(T - \lambda_i I_V) \cap \ker\big((T - \lambda_{i+1} I_V) \cdots (T - \lambda_m I_V)\big) = \{\mathbf{0}\}$$

for $i = 1, \ldots, m-1$. But if $\mathbf{x} \in \ker(T - \lambda_i I_V)$ and $\mathbf{x} \neq \mathbf{0}$, then

$$(T - \lambda_{i+1} I_V) \cdots (T - \lambda_m I_V)\mathbf{x} = (\lambda_i - \lambda_{i+1}) \cdots (\lambda_i - \lambda_m)\mathbf{x} \neq \mathbf{0},$$

since $\lambda_i \neq \lambda_j$ if $i \neq j$. Thus, by the lemma,

$$V = E_{\lambda_1}(T) \oplus \ \cdots \ \oplus E_{\lambda_m}(T),$$

which that proves T is semisimple. □

For matrices, we obtain the following corollary.

Corollary 8.15. *Suppose $A \in \mathbb{F}^{n \times n}$, and let $\lambda_1, \ldots, \lambda_m$ be the distinct eigenvalues of A in $\mathbb{F}$. Then A is diagonalizable over $\mathbb{F}$ if and only if*

$$(A - \lambda_1 I_n)(A - \lambda_2 I_n) \cdots (A - \lambda_m I_n) = O.$$

Remark. Of course, it is possible that there exist distinct λ_i, $1 \leq i \leq m$, such that $(A - \lambda_1 I_n)(A - \lambda_2 I_n) \cdots (A - \lambda_m I_n) = O$, where all the λ_i lie in a field $\mathbb{F}'$ containing $\mathbb{F}$. Some may lie in $\mathbb{F}$, of course. This simply means that A isn't diagonalizable over $\mathbb{F}$, but it is over $\mathbb{F}'$. An example of this is a 2×2 rotation matrix that is diagonalizable over $\mathbb{C}$ but not over $\mathbb{R}$. This illustrates an advantage that matrices have over linear mappings. A matrix over $\mathbb{F}$ may not be diagonalizable over $\mathbb{F}$, but it can be diagonalizable over a field containing $\mathbb{F}$. On the other hand, this concept does not make sense for a linear mapping with domain a vector space over $\mathbb{F}$, since we do not know how to enlarge the field over which V is defined. In fact, it is possible to do this, but it requires using tensor products.

Let us test the all-ones matrix with the second criterion.

Example 8.8. The 4×4 all-ones matrix B has eigenvalues 0 and 4. Therefore, to test it for diagonalizabilty (over $\mathbb{R}$), we have to show that $B(B - 4I_4) = O$. Calculating the product, we see that

$$B(B - 4I_4) = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} -3 & 1 & 1 & 1 \\ 1 & -3 & 1 & 1 \\ 1 & 1 & -3 & 1 \\ 1 & 1 & 1 & -3 \end{pmatrix} = O;$$

hence B is diagonalizable. ⊠

This test works well on upper and lower triangular matrices, since the eigenvalues are on the diagonal.

Example 8.9. Is the upper triangular matrix

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

diagonalizable? Since the eigenvalues are 1 and 2, we have to show that $(A - I_3)(A - 2I_3) = O$. Now,

$$(A - I_3)(A - 2I_3) = \begin{pmatrix} 0 & 2 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} -1 & 2 & 1 \\ 0 & -1 & 1 \\ 0 & 0 & 0 \end{pmatrix}.$$

The product is clearly nonzero, so A isn't diagonalizable. ⊠

This example shows that nondiagonalizable matrices exist. It also illustrates the general fact that the algebraic multiplicity of an eigenvalue is an upper bound on the dimension of its corresponding eigenspace, that is, the geometric multiplicity. Let us now reconsider a result we proved earlier.

Example 8.10 (Simple eigenvalues)**.** Recall from Proposition 8.10 that when $A \in \mathbb{F}^{n\times n}$ has n distinct eigenvalues in $\mathbb{F}$, then A is diagonalizable over $\mathbb{F}$. The Cayley–Hamilton theorem, which we will take up and prove in the next section, enables us to give an extremely short proof of this fact. Let $\lambda_1, \ldots, \lambda_n \in \mathbb{F}$ be the eigenvalues of A, which are assumed to all be different. Then

$$p_A(A) = (A - \lambda_1 I_n)(A - \lambda_2 I_n) \cdots (A - \lambda_n I_n). \tag{8.16}$$

But the Cayley–Hamilton theorem says that $p_A(A) = O$, so A is indeed diagonalizable by the diagonalizabilty test of Theorem 8.13. ⊠

Exercises

Exercise 8.3.1. Show from first principles that if λ and μ are distinct eigenvalues of A, then $E_\lambda(A) \cap E_\mu(A) = \{\mathbf{0}\}$.

Exercise 8.3.2. Diagonalize the following matrices if possible:

$$A = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}, \quad B = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}, \quad C = \begin{pmatrix} -3 & 0 & -4 & -4 \\ 0 & 2 & 1 & 1 \\ 4 & 0 & 5 & 4 \\ -4 & 0 & -4 & -3 \end{pmatrix}.$$

Exercise 8.3.3. Consider the real 2×2 matrix

$$A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}.$$

Determine whether A is diagonalizable.

Exercise 8.3.4. Determine which of the following matrices are diagonalizable over the reals:

$$\begin{pmatrix} 1 & 0 & -1 \\ -1 & 1 & 1 \\ 2 & -1 & -2 \end{pmatrix}, \quad \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & -1 & 1 \end{pmatrix}, \quad \begin{pmatrix} 2 & -1 & 1 \\ 1 & 0 & 1 \\ 1 & -1 & -2 \end{pmatrix}, \quad \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}.$$

Exercise 8.3.5. Does

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

have distinct eigenvalues? Is it diagonalizable?

Exercise 8.3.6. Let $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, where a, b, c, d are all positive real numbers. Show that A is diagonalizable.

Exercise 8.3.7. Suppose A is the matrix in Exercise 8.3.6. Show that A has an eigenvector in the first quadrant and another in the third quadrant.

Exercise 8.3.8. Show that if $A \in \mathbb{R}^{n\times n}$ admits an eigenbasis for $\mathbb{R}^n$, it also admits an eigenbasis for $\mathbb{C}^n$.

Exercise 8.3.9. Let $\mathbf{a} \in \mathbb{R}^3$, and let $C_{\mathbf{a}} : \mathbb{R}^3 \to \mathbb{R}^3$ be the cross product map $C_{\mathbf{a}}(\mathbf{v}) = \mathbf{a} \times \mathbf{v}$. Find the characteristic polynomial of $C_{\mathbf{a}}$ and determine whether $C_{\mathbf{a}}$ is semisimple.

Exercise 8.3.10. Let $V = \mathbb{F}^{n\times n}$ and let $\mathbb{T} : V \to V$ be the linear map defined by sending $A \in V$ to A^T. That is, $\mathbb{T}(A) = A^T$.

(*i*) Show that the only eigenvalues of $\mathbb{T}$ are ± 1.

(*ii*) Prove that if the characteristic of $\mathbb{F}$ is different from two, then $\mathbb{T}$ semisimple.

Exercise 8.3.11. Find an example of a nondiagonalizable 3×3 matrix A with real entries that is neither upper nor lower triangular such that every eigenvalue of A is 0.

Exercise 8.3.12. Let A be a 3×3 matrix with eigenvalues $0, 0, 1$. Show that $A^3 = A^2$.

Exercise 8.3.13. Let A be a 2×2 matrix such that $A^2 + 3A + 2I_2 = O$. Show that $-1, -2$ are eigenvalues of A.

Exercise 8.3.14. Suppose A is a 2×2 matrix such that $A^2 + A - 3I_2 = O$. Show that A is diagonalizable.

Exercise 8.3.15. Let U be an upper triangular matrix over $\mathbb{F}$ with distinct entries on its diagonal. Show that U is diagonalizable.

Exercise 8.3.16. Suppose that a 3×3 matrix A over $\mathbb{R}$ satisfies the equation $A^3 + A^2 - A + 2I_3 = O$.

(*i*) Find the eigenvalues of A.

(*ii*) Is A diagonalizable? Explain.

Exercise 8.3.17. We say that two $n \times n$ matrices A and B are simultaneously diagonalizable if they are diagonalized by the same matrix M, that is, if they share a common eigenbasis. Show that two simultaneously diagonalizable matrices A and B commute; that is, $AB = BA$.

Exercise 8.3.18. This is the converse to Exercise 8.3.17. Suppose $A, B \in \mathbb{F}^{n\times n}$ commute.

(*i*) Show that for every eigenvalue λ of A, $B(E_\lambda(A)) \subset E_\lambda(A)$.

(*ii*) Conclude that if S is the subset of $\mathbb{F}^{n\times n}$ consisting of diagonalizable matrices such that $AB = BA$ for all $A, B \in S$, then the elements of S are simultaneously diagonalizable.

Exercise 8.3.19. Let U be an arbitrary upper triangular matrix over $\mathbb{F}$ possibly having repeated diagonal entries. Show by example that U may not be diagonalizable, and give a condition to guarantee that it will be diagonalizable without any diagonal entries being changed.

Exercise 8.3.20. Let V be a finite dimensional vector space over $\mathbb{F}_p$, p a prime, and let μ be the mapping with domain and target $L(V, V)$ defined by $\mu(T) = T^p$. Show that μ is linear and find its characteristic polynomial when $V = (\mathbb{F}_p)^2$. Is μ diagonalizable?

8.4 The Cayley–Hamilton Theorem

The Cayley–Hamilton theorem gives an interesting and fundamental relationship between a matrix and its characteristic polynomial, which is related to our considerations about diagonalizabilty in the previous section. It is, in fact, one of the most famous and useful results in matrix theory.

8.4.1 Statement of the theorem

The version of the Cayley–Hamilton theorem we will prove below is stated as follows.

Theorem 8.16 (Cayley–Hamilton theorem)**.** *Let $\mathbb{F}$ be a field and suppose $A \in \mathbb{F}^{n\times n}$. Then A satisfies its characteristic polynomial, that is, $p_A(A) = O$. Consequently, if V is a finite-dimensional vector space over $\mathbb{F}$ and $T : V \to V$ is linear, then $p_T(T) = O_V$, where O_V is the zero mapping on V.*

The proof we give below in Section 8.4.4 was noticed by Jochen Kuttler and myself. It is more straightforward than the usual proof based on Cramer's rule. We first prove that $P_A(A) = O$ inductively just using matrix theory and the assumption that $\mathbb{F}$ contains all the eigenvalues of A. The second step in the proof, which doesn't involve matrix theory, is to show that there exists a field $\mathbb{F}'$ containing $\mathbb{F}$ such that all the roots of $p_A(x) = 0$ lie in $\mathbb{F}$. We say $\mathbb{F}'$ is a splitting field for p_A. Thus, $p_A(A) = O$ holds in $\mathbb{F}'^{n\times n}$, and hence $p_A(A) = O$ in $\mathbb{F}^{n\times n}$ too. And since it holds for A, it also holds for T.

8.4.2 The real and complex cases

If A is a real $n \times n$ matrix, its eigenvalues all lie in $\mathbb{C}$, and if A is diagonalizable over $\mathbb{C}$, say $A = MDM^{-1}$, then

$$p_A(A) = p_A(MDM^{-1}) = Mp_A(D)M^{-1} = M\mathrm{diag}(p_A(\lambda_1), p_A(\lambda_2), \ldots, p_A(\lambda_n))M^{-1} = O,$$

by the argument in Section 8.3.2. With some ingenuity, one can fashion a proof for arbitrary complex matrices (in particular, real matrices) by showing that every square matrix over $\mathbb{C}$ is the limit of a sequence of diagonalizable matrices.

Here is an important 2×2 example.

Example 8.11. For example, the characteristic polynomial of the matrix $J = \left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$ is $x^2 + 1$. Cayley–Hamilton asserts that $J^2 + I_2 = O$, which is easy to check directly. Notice that the eigenvalues of J are $\pm i$, so J is diagonalizable over $\mathbb{C}$, though not over $\mathbb{R}$. ⊠

8.4.3 Nilpotent matrices

A square matrix A is said to be *nilpotent* if $A^m = O$ for some integer $m > 0$. For example,

$$A = \begin{pmatrix} 1 & -1 \\ 1 & -1 \end{pmatrix}$$

is nilpotent, since $A^2 = O$. It follows directly from the definition of an eigenvalue that a nilpotent matrix cannot have a nonzero eigenvalue. (Convince yourself of this.) Thus if an $n \times n$ matrix A is nilpotent, its characteristic polynomial $p_A(x)$ is equal to $(-1)^n x^n$.

Proposition 8.17. *If an $n \times n$ matrix A is nilpotent, then $A^n = O$.*

Proof. Just apply Cayley–Hamilton. □

More generally, we also have the following definition.

Definition 8.3. A linear mapping $T : V \to V$ is said to be *nilpotent* if its only eigenvalue is 0.

So the previous proposition implies the following result.

Proposition 8.18. *If $T : V \to V$ is nilpotent, then $(T)^{\dim V} = O$.*

8.4.4 A proof of the Cayley–Hamilton theorem

We will first show that $p_A(A) = O$ for every $A \in \mathbb{F}^{n\times n}$, provided all the eigenvalues of A lie in $\mathbb{F}$. We will induct on n, the case $n = 1$ being trivial, since if $A = (a)$, then $p_A(x) = a - x$ and thus $p_A(A) = p_A(a) = 0$. Assume the result for $n - 1$, where $n > 1$. Let $(\lambda_1, \mathbf{v}_1)$ be an eigenpair for A, and extend $\mathbf{v}_1$ to a basis $\mathcal{B}$ of $\mathbb{F}^n$. Then A is similar to $B = \mathcal{M}_{\mathcal{B}}^{\mathcal{B}}(T_A)$, so $p_A(x) = p_B(x)$, and hence it suffices to show that $p_B(B) = O$. Now

$$B = \begin{pmatrix} \lambda_1 & * & \cdots & * \\ 0 & & & \\ \vdots & & B_1 & \\ 0 & & & \end{pmatrix},$$

where $B_1 \in \mathbb{F}^{(n-1)\times(n-1)}$. It follows from this block decomposition that $\det(B - xI_n) = (\lambda_1 - x)\det(B_1 - xI_{n-1})$. Since $p_A(x) = p_B(x)$ and the eigenvalues$\lambda_1, \lambda_2, \ldots, \lambda_n$ of A all lie in $\mathbb{F}$, it follows that the eigenvalues of B are in $\mathbb{F}$, so the eigenvalues $\lambda_2, \ldots, \lambda_n$ of B_1 are also in $\mathbb{F}$. These all being elements of $\mathbb{F}$, we may apply the induction hypothesis to B_1. First, notice

the following fact about block matrix multiplication: if C_1 and C_2 are of size $(n-1)\times(n-1)$, then

$$\begin{pmatrix} c_1 & * & \cdots & * \\ 0 & & & \\ \vdots & & C_1 & \\ 0 & & & \end{pmatrix}\begin{pmatrix} c_2 & * & \cdots & * \\ 0 & & & \\ \vdots & & C_2 & \\ 0 & & & \end{pmatrix} = \begin{pmatrix} c_1c_2 & * & \cdots & * \\ 0 & & & \\ \vdots & & C_1C_2 & \\ 0 & & & \end{pmatrix}.$$

Now calculate $p_B(B)$. By the previous comment,

$$\begin{aligned} p_B(B) &= (-1)^n(B-\lambda_1 I_n)(B-\lambda_2 I_n)\cdots(B-\lambda_n I_n) \\ &= (-1)^n \begin{pmatrix} 0 & * & \cdots & * \\ 0 & & & \\ \vdots & & B_1-\lambda_1 I_{n-1} & \\ 0 & & & \end{pmatrix}\begin{pmatrix} * & * & \cdots & * \\ 0 & & & \\ \vdots & & C & \\ 0 & & & \end{pmatrix}, \end{aligned}$$

where

$$C = (B_1-\lambda_2 I_{n-1})\cdots(B_1-\lambda_n I_{n-1}).$$

But clearly $C = p_{B_1}(B_1)$, so by the induction hypothesis, it follows that $C = O$. Thus

$$p_B(B) = (-1)^n \begin{pmatrix} 0 & * & \cdots & * \\ 0 & & & \\ \vdots & & B_1-\lambda_1 I_{n-1} & \\ 0 & & & \end{pmatrix}\begin{pmatrix} * & * & \cdots & * \\ 0 & & & \\ \vdots & & O & \\ 0 & & & \end{pmatrix}.$$

Carrying out the multiplication, one sees immediately that $p_B(B) = O$. □

The assumption that all the roots of the characteristic polynomial of A lie in $\mathbb{F}$ is actually unnecessary, due to the next result.

Lemma 8.19. *Let $\mathbb{F}$ be a field. Then for every polynomial $f(x)\in\mathbb{F}[x]$, there exists a field $\mathbb{F}'$ containing $\mathbb{F}$ and all the roots of $f(x)=0$.*

We will give a proof of this fact in Section 8.7. Thus the eigenvalues of A lie in an extension $\mathbb{F}'$ of $\mathbb{F}$, so $p_A(A) = O$. All that remains to be proved is that a linear mapping also satisfies its characteristic polynomial. Let V be a finite-dimensional vector space over $\mathbb{F}$ and $T: V\to V$ a linear mapping. Then for every matrix A of T with respect to a basis of V, we have $p_A(A) = O$. But as noted already, the matrix of $p_T(T)$ with respect to this basis is $p_A(A)$. Since $p_A(A) = O$, it follows that $p_T(T)$ is zero also. □

8.4.5 The minimal polynomial of a linear mapping

Let V be as usual and $T : V \to V$ a linear mapping. By the Cayley–Hamilton theorem, $p_T(T) = O_V$, and therefore there exists a polynomial $f(x) \in \mathbb{F}[x]$ of least degree and leading coefficient one such that $f(T) = O_V$. This follows by division with remainder. For if f_1 and f_2 satisfy the minimal polynomial criterion, then f_1 and f_2 must have the same degree k. Then $g = f_1 - f_2$ is a polynomial of degree at most $k-1$ such that $g(T) = O$. Thus $g = 0$. The *minimal polynomial* of T will be denoted by μ_T. The minimal polynomial μ_A of a matrix $A \in \mathbb{F}^{n\times n}$ is defined in exactly the same way, and if A is the matrix of T, then $\mu_A = \mu_T$.

Proposition 8.20. *Suppose $T : V \to V$ is a linear mapping whose eigenvalues lie in $\mathbb{F}$, and let $\lambda_1, \ldots, \lambda_m$ denote T's distinct eigenvalues in $\mathbb{F}$. Then the minimal polynomial $\mu_T(x)$ of T divides the characteristic polynomial $p_T(x)$, and $q(x) = (x - \lambda_1)\cdots(x - \lambda_m)$ divides $\mu_T(x)$.*

Proof. Since the degree of $\mu_T(x)$ is at most n, division with remainder in $\mathbb{F}[x]$ allows us to write $p_T(x) = h(x)\mu_T(x) + r(x)$, where either $r = 0$ or $\deg r < \deg \mu_T$. But $p_T(T) = p(T) = O$, $r(T) = O$ also. By the definition of μ_T, it follows that $r = 0$; hence polynomial $p(x)$ divides $p_T(x)$. To conclude that $q(x)$ divides $\mu_T(x)$, use division with remainder again. Let $(\lambda_j, \mathbf{v})$ be an eigenpair for T, and write $\mu_T(x) = f(x)(x - \lambda_j) + r$. Since $\mu_T(T)\mathbf{v} = (T - \lambda_j)\mathbf{v} = \mathbf{0}$, it follows that $rI_n = 0$, so $r = 0$. Thus q divides μ_T. □

The following corollary restates the diagonalization criterion of Theorem 8.13. As above, let $q(x) = (x - \lambda_1)\cdots(x - \lambda_m)$, where $\lambda_1, \ldots, \lambda_m$ denote T's distinct eigenvalues.

Corollary 8.21. *The linear mapping $T : V \to V$ is semisimple if and only if its minimal polynomial is $q(x)$.*

Proof. If T is semisimple, then all of its eigenvalues lie in $\mathbb{F}$, and by Theorem 8.13, $q(T) = O$. Therefore, $q(x)$ is the minimal polynomial of T. Conversely, if $q(x)$ is T's minimal polynomial, then the criterion of Theorem 8.13 implies that T is semisimple. □

Exercises

Exercise 8.4.1. Verify the Cayley–Hamilton theorem directly for

$$A = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}.$$

Exercise 8.4.2. Give a direct proof for the 2×2 case. That is, show that

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} = A^2 - \mathrm{Tr}(A)A + \det(A)I_2 = O.$$

Apparently, this was Cayley's contribution to the theorem.

Exercise 8.4.3. The following false proof of the Cayley–Hamilton theorem is well known and very persistent. In fact, it actually appears in an algebra book. Since setting $x = A$ in $p_A(x) = \det(A - xI_n)$ gives $\det(A - AI_n) = \det(A - A) = 0$, it follows that $p_A(A) = O$. What is incorrect about this "proof"?

Exercise 8.4.4. Show that a 2×2 matrix A is nilpotent if and only if $\mathrm{Tr}(A) = \det(A) = 0$. Use this to find a 2×2 nilpotent matrix with all entries different from 0.

Exercise 8.4.5. Use the Cayley–Hamilton theorem to deduce that $A \in \mathbb{F}^{n \times n}$ is nilpotent if and only if all eigenvalues of A are 0.

Exercise 8.4.6. Show that a nonzero nilpotent matrix is not diagonalizable.

Exercise 8.4.7. Without using the Cayley–Hamilton theorem, show that if an $n \times n$ matrix A is nilpotent, then in fact $A^n = O$. (Hint: use induction on n. Choose a basis of $\mathcal{N}(A)$ and extend to a basis of $\mathbb{F}^n$. Then apply the induction to $\mathrm{col}(A)$.)

Exercise 8.4.8. * Prove that every $A \in \mathbb{C}^{n \times n}$ is the limit of a sequence of diagonalizable matrices, and thus deduce the Cayley–Hamilton theorem over $\mathbb{C}$ in another way.

Exercise 8.4.9. Test the following matrices to determine which are diagonalizable.

(i) $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ and $\begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}$ $(\mathbb{F} = \mathbb{F}_2)$.

(ii) $\begin{pmatrix} 2 & -1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 2 \end{pmatrix}$ $(\mathbb{F} = \mathbb{Q})$.

(iii) $\begin{pmatrix} 1 & -1 & 2 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix}$ $(\mathbb{F} = \mathbb{Q})$.

(iv) $\begin{pmatrix} 5 & 4 & 2 & 1 \\ 0 & 1 & -1 & -1 \\ -1 & -1 & 3 & 0 \\ 1 & 1 & -1 & 2 \end{pmatrix}$ $(\mathbb{F} = \mathbb{Q})$. Note that the characteristic polynomial is $x^4 - 11x^3 + 42x^2 - 64x + 32 = (x-1)(x-2)(x-4)^2$.

(v) $\begin{pmatrix} 2 & 1 & 2 & 1 \\ 0 & 1 & 2 & 2 \\ 2 & 2 & 0 & 0 \\ 1 & 1 & 2 & 2 \end{pmatrix}$ $(\mathbb{F} = \mathbb{F}_3)$.

Exercise 8.4.10. Find the minimal polynomials of the matrices in parts (i)–(iii) of Exercise 8.4.9.

Exercise 8.4.11. Show that the characteristic polynomial of an $n \times n$ matrix divides a power of its minimal polynomial.

Exercise 8.4.12. This exercise shows the minimal polynomial exists without appealing to Cayley–Hamilton. Show that if V is a finite dimensional vector space and $T : V \to V$ is linear, then the powers $T^0, T, T^2, \ldots T^k$ are linearly dependent for some $k > 0$. Conclude that the minimal polynomial of T exists.

8.5 Self Adjoint Mappings and the Principal Axis Theorem

We now return to inner product spaces (both real and Hermitian) to treat one of the most famous results in linear algebra: every matrix that is symmetric over $\mathbb{R}$ or Hermitian over $\mathbb{C}$ is diagonalizable. Moreover, it has an orthonormal eigenbasis. This result is known as the principal axis theorem. It is also commonly referred to as the spectral theorem. Our goal is to present a geometric proof that explains some of the intuition underlying the ideas behind symmetric and Hermitian matrices. We will first treat self-adjoint linear mappings. These are the linear mappings whose matrices with respect to an orthonormal basis are symmetric in the real case and Hermitian in the complex case.

8.5.1 The notion of self-adjointness

Let V be either a real or Hermitian finite-dimensional inner product space. The inner product of $\mathbf{x}$ and $\mathbf{y}$ in V will be denoted by $(\mathbf{x}, \mathbf{y})$. Consider a linear mapping $T : V \to V$ such that

$$(T(\mathbf{x}), \mathbf{y}) = (\mathbf{x}, T(\mathbf{y}))$$

for all $\mathbf{x}, \mathbf{y} \in V$. In the case that V is a real vector space, we will say that T is a *self-adjoint linear mapping*, and in the complex case, we will say that T is a *Hermitian self-adjoint linear mapping*. The connection between self-adjointness and matrix theory comes from the following result.

Proposition 8.22. *Let V be a finite-dimensional Hermitian inner product space (respectively inner product space) over $\mathbb{C}$ (respectively $\mathbb{R}$). Then a linear mapping $T : V \to V$ is self-adjoint if and only if its matrix with respect to an arbitary Hermitian orthonormal basis (respectively orthonormal basis) of V is Hermitian (respectively symmetric).*

The proof is an exercise. A matrix $K \in \mathbb{C}^{n\times n}$ is said to be Hermitian if $K^H = K$. We remind the reader that $A^H = (\overline{A})^T$. Consequently, we get the following corollary.

Corollary 8.23. *A linear mapping $T : \mathbb{R}^n \to \mathbb{R}^n$ is self-adjoint if and only if $T = T_A$, where A is symmetric. Similarly, a linear mapping $T : \mathbb{C}^n \to \mathbb{C}^n$ is Hermitian self-adjoint if and only if $T = T_K$, where K is Hermitian.*

Proof. Let us give a proof in the real case. If T is self-adjoint, then $T(\mathbf{e}_i) \cdot \mathbf{e}_j = \mathbf{e}_i \cdot T(\mathbf{e}_j)$. This implies that $a_{ij} = a_{ji}$. Thus $A = A^T$. The converse is similar. □

It is more natural to consider the Hermitian case first. The reason for this will be clear later. The geometric consequences of the condition that a linear mapping is Hermitian self-adjoint are summed up in the following.

Proposition 8.24. *Suppose $T : V \to V$ is a Hermitian self-adjoint linear mapping. Then:*

(i) all eigenvalues of T are real;

(ii) eigenvectors of T corresponding to distinct eigenvalues are orthogonal;

(iii) if W is a subspace of V such that $T(W) \subset W$, then $T(W^\perp) \subset W^\perp$;

(iv) $\operatorname{im}(T) = \ker(T)^\perp$; and

(v) consequently, $V = \ker(T) \oplus \operatorname{im}(T)$.

Proof. We will first show that T has only real eigenvalues. Since $\mathbb{C}$ is algebraically closed, T has $\dim V$ complex eigenvalues. Let λ be any eigenvalue and suppose $(\lambda, \mathbf{w})$ is a corresponding eigenpair. Since $(T(\mathbf{w}), \mathbf{w}) = (\mathbf{w}, T(\mathbf{w}))$, we see that $(\lambda\mathbf{w}, \mathbf{w}) = (\mathbf{w}, \lambda\mathbf{w})$. This implies $\overline{\lambda}|\mathbf{w}|^2 = \lambda|\mathbf{w}|^2$. Since $|\mathbf{w}| \neq 0$, we have $\lambda = \overline{\lambda}$, so all eigenvalues of T are real. For (ii), assume that λ and μ are distinct eigenvalues of T with corresponding eigenvectors $\mathbf{v}$ and $\mathbf{w}$. Then $(T(\mathbf{v}), \mathbf{w}) = (\mathbf{v}, T(\mathbf{w}))$, so $(\lambda\mathbf{v}, \mathbf{w}) = (\mathbf{v}, \mu\mathbf{w})$. Hence, $(\lambda - \mu)(\mathbf{v}, \mathbf{w}) = 0$. (Recall that by (i), $\lambda, \mu \in \mathbb{R}$.) Since $\lambda \neq \mu$, we get (ii). For (iii), let $\mathbf{x} \in W$ and $\mathbf{y} \in W^\perp$. Since $T(\mathbf{x}) \in W$, $(T(\mathbf{x}), \mathbf{y}) = 0$. But $(\mathbf{x}, T(\mathbf{y})) = (T(\mathbf{x}), \mathbf{y})$, so $(\mathbf{x}, T(\mathbf{y})) = 0$. Since $\mathbf{x}$ is arbitrary, it follows that $T(\mathbf{y}) \in W^\perp$; hence (iii) follows. For (iv), first note that $\operatorname{im}(T) \subset \ker(T)^\perp$. Indeed, if $\mathbf{y} = T(\mathbf{x})$ and $\mathbf{w} \in \ker(T)$, then

$$(\mathbf{w}, \mathbf{y}) = (\mathbf{w}, T(\mathbf{x})) = (T(\mathbf{w}), \mathbf{x}) = 0.$$

By Proposition 6.38, $\dim V = \dim\ker(T) + \dim\ker(T)^\perp$, and by the rank–nullity theorem (Theorem 7.4), $\dim V = \dim\ker(T) + \dim\operatorname{im}(T)$. Hence, $\dim\operatorname{im}(T) = \dim\ker(T)^\perp$, so $\operatorname{im}(T) = \ker(T)^\perp$. Part (v) follows from (iv), since $V = W \oplus W^\perp$ for every subspace W of V. □

8.5.2 Principal Axis Theorem for self-adjoint linear mappings

We will now prove the principal axis theorem, starting with the Hermitian version.

Theorem 8.25. *(The Hermitian Principal Axis Theorem) Suppose V is a finite-dimensional Hermitian vector space, and let $T : V \to V$ be Hermitian self-adjoint. Then there exists a Hermitian orthonormal basis of V consisting of eigenvectors of T.*

Proof. We will induct on $\dim V$. The result is certainly true if $\dim V = 1$. Suppose it holds when $\dim V \leq n-1$, and let $\dim V = n$. Let $\lambda \in \mathbb{R}$ be any eigenvalue of T, and let us replace T by $S = T - \lambda I_V$. If $V = \ker(S)$, there is nothing to prove, since every Hermitian orthonormal basis of V is an eigenbasis. Thus we may assume that $V \neq \ker(S)$. Note that S is also self-adjoint, and $\ker(S)$ and $\mathrm{im}(S)$ are both stable under T, since S and T commute. Moreover, by the previous proposition, V decomposes into the orthogonal direct sum $V = \ker(S) \oplus \mathrm{im}(S)$. Since $0 < \dim \ker(S) \leq n-1$ and likewise, $0 < \dim \mathrm{im}(S) \leq n-1$, it follows by induction that $\ker(S)$ and $\mathrm{im}(S)$ each admits a Hermitian orthonormal eigenbasis for T. This finishes the proof. □

Corollary 8.26. *If $K \in \mathbb{C}^{n\times n}$ is Hermitian, then there exists an eigenbasis $\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_n$ for K that is Hermitian orthonormal. Hence, there exists a unitary matrix $U \in \mathbb{C}^{n\times n}$ such that $KU = UD$, where D is real diagonal. Thus, $K = UDU^{-1} = UDU^H$.*

Proof. Put $U = (\mathbf{w}_1\ \mathbf{w}_2\ \ldots\ \mathbf{w}_n)$. □

Similarly, we have the real principal axis theorem.

Theorem 8.27. *(The real Principal Axis Theorem) Suppose V is a finite-dimensional inner product space, and let $T : V \to V$ be self-adjoint. Then there exists an orthonormal basis of V consisting of eigenvectors of T.*

Proof. The proof is identical to that of the Hermitian principal axis theorem once we prove that T has only real eigenvalues. For this, we appeal to the matrix A of T with respect to an orthonormal basis of V. This matrix is symmetric, hence also Hermitian. Therefore, A has only real eigenvalues. But the eigenvalues of A are also the eigenvalues of T, so we have the desired result. □

Corollary 8.28. *If $A \in \mathbb{R}^{n\times n}$ is symmetric, then there exists an orthonormal eigenbasis $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n$ for A. Consequently, there exists an orthogonal matrix Q such that $A = QDQ^{-1} = ADQ^T$, where D is real diagonal.*

Proof. Take $Q = (\mathbf{u}_1\ \mathbf{u}_2\ \ldots\ \mathbf{u}_n)$. □

Remark. Symmetric and Hermitian matrices satisfy the condition that $AA^H = A^H A$. Thus they are normal matrices, and hence automatically admit a Hermitian orthonormal basis, by the normal matrix theorem, which we will prove in Chap. 9. The proof is conceptually much simpler, but it doesn't shed any light on the geometry. There are infinite-dimensional versions of the spectral theorem for bounded self-adjoint operators on Hilbert space that the reader can read about in a book on functional analysis. The finite-dimensional spectral theorem is due to Cauchy.

8.5.3 Examples of self-adjoint linear mappings

Let us now consider some examples.

Example 8.12 (Projections)**.** Let W be a subspace of $\mathbb{R}^n$. Recall from Section 7.3.5 that the projection $P_W : \mathbb{R}^n \to \mathbb{R}^n$ is defined by choosing an orthonormal basis $\mathbf{u}_1, \ldots, \mathbf{u}_m$ of W and putting

$$P_W(\mathbf{x}) = \sum_{i=1}^{m} (\mathbf{x} \cdot \mathbf{u}_i)\mathbf{u}_i.$$

Certainly P_W is linear. To see that it is self-adjoint, extend the orthonormal basis of W to an orthonormal basis $\mathbf{u}_1, \ldots, \mathbf{u}_n$ of $\mathbb{R}^n$. If $1 \leq i, j \leq m$, then $P_W(\mathbf{u}_i) \cdot \mathbf{u}_j = \mathbf{u}_i \cdot P_W(\mathbf{u}_j) = \mathbf{u}_i \cdot \mathbf{u}_j$. On the other hand, if one of the indices i, j exceeds m, then $P_W(\mathbf{u}_i) \cdot \mathbf{u}_j = 0$ and $\mathbf{u}_i \cdot P_W(\mathbf{u}_j) = 0$. Thus, P_W is self-adjoint. ⊠

Example 8.13 (Reflections)**.** Let H be a hyperplane in $\mathbb{R}^n$, say $H = (\mathbb{R}\mathbf{u})^\perp$, where $\mathbf{u}$ is a unit vector. Recall that the reflection of $\mathbb{R}^n$ through H is the linear mapping $Q(\mathbf{v}) = \mathbf{v} - 2(\mathbf{v}, \mathbf{u})\mathbf{u}$. Then $Q = I_n - 2P_W$, where $W = \mathbb{R}\mathbf{u}$. Since the sum of two self-adjoint maps is evidently self-adjoint, all reflections are also self-adjoint. ⊠

Example 8.14. Recall that in Example 6.33 we introduced the inner product on $\mathbb{R}^{n\times n}$ given by the Killing form

$$(A, B) = \mathrm{Tr}(AB^T),$$

where $\mathrm{Tr}(A)$ is the trace of A. In this example, we show that the linear mapping $T : \mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}$ defined by $T(A) = A^T$ is self-adjoint with respect to the Killing form. In other words, $(A^T, B) = (A, B^T)$ for all A, B. To see this, note that

$$(A^T, B) = \mathrm{Tr}(A^T B^T) = \mathrm{Tr}((BA)^T) = \mathrm{Tr}(BA),$$

while

$$(A, B^T) = \mathrm{Tr}(A(B^T)^T) = \mathrm{Tr}(AB).$$

But $\mathrm{Tr}(AB) = \mathrm{Tr}(BA)$ for all $A, B \in \mathbb{R}^{n\times n}$, so T is self-adjoint. ⊠

In fact, the identity $\mathrm{Tr}(AB) = \mathrm{Tr}(BA)$ holds for all $A, B \in \mathbb{F}^{n\times n}$, where $\mathbb{F}$ is any field.

Example 8.15. Here is another way to show that $T(A) = A^T$ is self-adjoint on $\mathbb{R}^{n\times n}$. We will show that T admits an orthonormal eigenbasis. Notice that since $T^2 = I_W$, where $W = \mathbb{R}^{n\times n}$, its only eigenvalues are ± 1. An eigenbasis

consists of the symmetric matrices E_{ii} and $(E_{ij} + E_{ji})/\sqrt{2}$, where $i \neq j$ with eigenvalue $\lambda = 1$, and the skew-symmetric matrices $(E_{ij} - E_{ji})/\sqrt{2}$ for $i \neq j$ with eigenvalue $\lambda = -1$. We claim that this eigenbasis is orthonormal (the verification is an exercise), so T is self-adjoint. A consequence of this orthonormal basis is that every real matrix can be orthogonally decomposed as the sum of a symmetric matrix and a skew-symmetric matrix. (Recall that we already proved in Chap. 6 that every square matrix over an arbitrary field of characteristic different from two can be uniquely expressed as the sum of a symmetric matrix and a skew-symmetric matrix.) ⊠

The next example is an opportunity to diagonalize a 4×4 symmetric matrix (one that was already diagonalized in Chap. 8) without any calculations.

Example 8.16. Let B be the 4×4 all-ones matrix. The rank of B is clearly one, so 0 is an eigenvalue and $\mathcal{N}(B) = E_0$ has dimension three. In fact, $E_0 = (\mathbb{R}(1,1,1,1)^T)^\perp$. Thus, $(1,1,1,1)^T$ is also an eigenvector. In fact, all the rows sum to 4, so the eigenvalue for $(1,1,1,1)^T$ is 4. Consequently, B is orthogonally similar to $D = \mathrm{diag}(4,0,0,0)$. To produce Q such that $B = QDQ^T$, we need to find an orthonormal basis of E_0. One can simply look for orthogonal vectors orthogonal to $(1,1,1,1)^T$. In fact, $\mathbf{v}_1 = 1/2(1,1,-1,-1)^T$, $\mathbf{v}_2 = 1/2(1,-1,1,-1)^T$, and $\mathbf{v}_3 = 1/2(1,-1,-1,1)^T$ will give such an orthonormal basis after they are normalized. We can thus write

$$B = \frac{1}{2}\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix} \begin{pmatrix} 4 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix} \frac{1}{2}.$$

Notice that the orthogonal matrix Q used here is symmetric, so $Q^{-1} = Q$. ⊠

8.5.4 A projection formula for symmetric matrices

One of the nice applications of the principal axis theorem is that it enables one to express any symmetric matrix as a linear combination of orthogonal projections. Suppose $A \in \mathbb{R}^{n \times n}$ is symmetric, and let $(\lambda_1, \mathbf{u}_1), \ldots, (\lambda_n, \mathbf{u}_n)$ be eigenpairs for A that give an orthonormal eigenbasis of $\mathbb{R}^n$. Then if $\mathbf{x} \in \mathbb{R}^n$, the projection formula (6.21) gives

$$\mathbf{x} = (\mathbf{u}_1^T \mathbf{x})\mathbf{u}_1 + \cdots + (\mathbf{u}_n^T \mathbf{x})\mathbf{u}_n.$$

Hence

$$A\mathbf{x} = \lambda_1(\mathbf{u}_1^T\mathbf{x})\mathbf{u}_1 + \cdots + \lambda_n(\mathbf{u}_n^T\mathbf{x})\mathbf{u}_n.$$

Thus

$$A = \lambda_1\mathbf{u}_1\mathbf{u}_1^T + \cdots + \lambda_n\mathbf{u}_n\mathbf{u}_n^T. \tag{8.17}$$

Since $\mathbf{u}_i\mathbf{u}_i^T$ is the matrix of the projection of $\mathbb{R}^n$ onto the line $\mathbb{R}\mathbf{u}_i$, the identity (8.17) indeed expresses A as a linear combination of orthogonal projections. This formula holds in the Hermitian case as well, provided one uses the Hermitian inner product.

The projection formula (8.17) can be put in a more elegant form. If $\mu_1, \ldots, \mu_k$ are the distinct eigenvalues of A and $E_1, \ldots, E_k$ are the corresponding eigenspaces, then

$$A = \mu_1 P_{E_1} + \mu_2 P_{E_2} + \cdots + \mu_k P_{E_k}. \tag{8.18}$$

For example, in the case of the all-ones matrix of Example 8.16,

$$A = 0P_{E_0} + 4P_{E_4} = 4P_{E_4}.$$

Exercises

Exercise 8.5.1. Let $C_{\mathbf{a}} : \mathbb{R}^3 \to \mathbb{R}^3$ be the cross product map $C_{\mathbf{a}}(\mathbf{v}) = \mathbf{c} \times \mathbf{v}$. True or false: $C_{\mathbf{a}}$ is self-adjoint.

Exercise 8.5.2. Prove Proposition 8.22.

Exercise 8.5.3. Are rotations of $\mathbb{R}^3$ self-adjoint?

Exercise 8.5.4. Orthogonally diagonalize the following matrices:

$$\begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 1 & 3 \\ 1 & 3 & 1 \\ 3 & 1 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix}.$$

(Try to diagonalize the first and third matrices without pencil and paper. You can also find an eigenvalue of the second by inspection.)

Exercise 8.5.5. Let $A = \begin{pmatrix} a & b \\ b & c \end{pmatrix} \in \mathbb{R}^{2\times 2}$.

(*i*) Show directly that both roots of the characteristic polynomial of A are real.

(*ii*) Prove that A is orthogonally diagonalizable without appealing to the principal axis theorem.

Exercise 8.5.6. Suppose B is a real symmetric 3×3 matrix such that $(1,\ 0,\ 1)^T \in \mathcal{N}(B - I_3)$, and $(1,\ 1,\ -1)^T \in \mathcal{N}(B - 2I_3)$. If $\det(B) = 12$, find B.

Exercise 8.5.7. Answer each question true or false. If true, give a brief reason. If false, give a counter example.

(*i*) The sum and product of two symmetric matrices are symmetric.

(*ii*) If two symmetric matrices A and B have the same eigenvalues, counting multiplicities, then A and B are orthogonally similar ($A = QBQ^{-1}$, where Q is orthogonal).

Exercise 8.5.8. Suppose A is a 3×3 symmetric matrix such that the trace of A is 4, the determinant of A is 0, and $\mathbf{v}_1 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$ and $\mathbf{v}_2 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$ are eigenvectors of A that lie in the image of T_A.

(*i*) Find the eigenvalues of A.

(*ii*) Find the eigenvalues corresponding to $\mathbf{v}_1$ and $\mathbf{v}_2$.

(*iii*) Finally, find A itself.

Exercise 8.5.9. Suppose A and B in $\mathbb{R}^{n\times n}$ are both symmetric and have a common eigenbasis. Show that AB is symmetric.

Exercise 8.5.10. Suppose A, B, and AB are symmetric. Show that A and B are simultaneously diagonalizable. Is BA symmetric?

Exercise 8.5.11. Let W be a subspace of $\mathbb{R}^n$. Simultaneously orthogonally diagonalize P_W and $P_{W^\perp}$.

Exercise 8.5.12. • Suppose $A \in \mathbb{R}^{3\times 3}$ is symmetric and the trace of A is an eigenvalue. Show that if A is invertible, then $\det(A)\mathrm{Tr}(A) < 0$.

Exercise 8.5.13. • Assume that a, b, c are all real. Let

$$A = \begin{pmatrix} a & b & c \\ b & c & a \\ c & a & b \end{pmatrix}.$$

(*i*) Show that if $\mathrm{Tr}(A) = 0$, then $\det(A) = 0$ too.

(*ii*) Diagonalize A if $\det(A) = 0$ but $\mathrm{Tr}(A) \neq 0$.

Exercise 8.5.14. Using Exercise 8.5.12 and the matrix A in Exercise 8.5.13, show that if $a + b + c > 0$, then

$$a^3 + b^3 + c^3 - 3abc > 0.$$

Exercise 8.5.15. • Diagonalize

$$A = \begin{pmatrix} aa & ab & ac & ad \\ ba & bb & bc & bd \\ ca & cb & cc & cd \\ da & db & dc & dd \end{pmatrix},$$

where a, b, c, d are arbitrary real numbers. (Note: it may help to factor A.)

Exercise 8.5.16. Prove that a real symmetric matrix A whose only eigenvalues are ± 1 is orthogonal.

Exercise 8.5.17. Suppose $A \in \mathbb{R}^n$ is symmetric. Show that if $A^k = O$ for some positive integer k, then $A = O$.

Exercise 8.5.18. Give a direct proof of the principal axis theorem in the 2×2 Hermitian case.

Exercise 8.5.19. Show that two real symmetric matrices A and B having the same characteristic polynomial are orthogonally similar. In other words, $A = QBQ^{-1}$ for some orthogonal matrix Q.

Exercise 8.5.20. • Let $A \in \mathbb{R}^n$ be symmetric, and let λ_m and λ_M be its minimum and maximum eigenvalues respectively.

(*i*) Use formula (8.17) to show that for every $\mathbf{x} \in \mathbb{R}^n$, we have

$$\lambda_m \mathbf{x}^T \mathbf{x} \leq \mathbf{x}^T A \mathbf{x} \leq \lambda_M \mathbf{x}^T \mathbf{x}.$$

(*ii*) Use this inequality to find the maximum and minimum values of $|A\mathbf{x}|$ on the ball $|\mathbf{x}| \leq 1$ in $\mathbb{R}^n$.

(*iii*) Show that the maximum and minimum values of $\mathbf{x}^T A \mathbf{x}$ for $|\mathbf{x}| = 1$ are eigenvalues of A.

Exercise 8.5.21. Show that if $Q \in \mathbb{R}^n$ is orthogonal and symmetric, then $Q^2 = I_n$. Moreover, if 1 is not an eigenvalue of Q, then $Q = -I_n$.

Exercise 8.5.22. Find the eigenvalues of $K = \begin{pmatrix} 2 & 3+4i \\ 3-4i & -2 \end{pmatrix}$ and diagonalize K.

Exercise 8.5.23. Unitarily diagonalize the rotation $R_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$.

Exercise 8.5.24. Using only the definition, show that the trace and determinant of a Hermitian matrix are real.

Exercise 8.5.25. Describe the relationship between $U(1, \mathbb{C})$ and $SO(2, \mathbb{R})$.

Exercise 8.5.26. Let $SU(2,\mathbb{C}) \subset U(2,\mathbb{C})$ denote the set of 2×2 unitary matrices of determinant one.

(*i*) Show that $SU(2,\mathbb{C})$ is a matrix group.

(*ii*) Describe the eigenvalues of the elements of $SU(2,\mathbb{C})$.

Exercise 8.5.27. Consider a 2×2 unitary matrix U such that one of U's columns is in $\mathbb{R}^2$. Is U orthogonal?

Exercise 8.5.28. Suppose W is a complex subspace of $\mathbb{C}^n$. Show that the projection P_W is Hermitian.

Exercise 8.5.29. Show how to alter the Killing form to define a Hermitian inner product on $\mathbb{C}^{n\times n}$.

Exercise 8.5.30. Verify that the basis in Example 8.15 is indeed an orthonormal basis of $\mathbb{R}^{n\times n}$.

Exercise 8.5.31. Find a one-to-one correspondence between the set of all isometries $\Phi : \mathbb{R}^2 \to \mathbb{R}^2$ and $O(2,\mathbb{R})$.

Exercise 8.5.32. Give a proof of the real principal axis theorem for a self-adjoint $T : V \to V$ by reducing it to the case of a symmetric $n \times n$ matrix by choosing an isometry $\Phi : V \to \mathbb{R}^n$ $(n = \dim V)$.

Exercise 8.5.33. Let V be a finite-dimensional inner product space, and suppose $T : V \to V$ is linear. Define the *adjoint* of T to be the map $T^* : V \to V$ determined by the condition that

$$(T^*(\mathbf{x}), \mathbf{y}) = (\mathbf{x}, T(\mathbf{y}))$$

for all $\mathbf{x}, \mathbf{y} \in V$.

(*i*) Show that the adjoint T^* is a well-defined linear mapping.

(*ii*) If $V = \mathbb{R}^n$, find the matrix of T^*.

8.6 The Group of Rotations of $\mathbb{R}^3$ and the Platonic Solids

The purpose of this section is give an application of eigentheory to group theory. We will show that the set of rotations of $\mathbb{R}^3$ is the matrix group $SO(3,\mathbb{R})$ of 3×3 orthogonal matrices of determinant one. After that, we will describe the Platonic solids and study their rotations.

8.6.1 Rotations of $\mathbb{R}^3$

The classical definition of a rotation of $\mathbb{R}^3$ originated with Euler. A mapping $\rho:\mathbb{R}^3\to\mathbb{R}^3$ that fixes every point on a line through the origin, called the axis of ρ, and rotates every plane orthogonal to the axis through the same fixed angle θ is called a *rotation* of $\mathbb{R}^3$. We will let $\mathrm{Rot}(\mathbb{R}^3)$ denote the set of all rotations of $\mathbb{R}^3$. Notice that it is not at all clear that $\mathrm{Rot}(\mathbb{R}^3)$ is a group. It must be shown that the composition of two rotations with different axes is also a rotation.

Our first objective is to show that rotations are linear. We will then show that the matrix of a rotation is orthogonal and has determinant one. Thus, $\mathrm{Rot}(\mathbb{R}^3)\subset SO(3,\mathbb{R})$. It is clear from the definition that a rotation preserves lengths and angles. Since the inner product on $\mathbb{R}^3$ has the property that

$$\mathbf{x}\cdot\mathbf{y}=|\mathbf{x}||\mathbf{y}|\cos\alpha$$

for all nonzero $\mathbf{x},\mathbf{y}\in\mathbb{R}^3$, α being the angle between $\mathbf{x}$ and $\mathbf{y}$, it follows that every transformation of $\mathbb{R}^3$ preserving both lengths and angles also preserves dot products. Thus if $\rho\in\mathrm{Rot}(\mathbb{R}^3)$, then

$$\rho(\mathbf{x})\cdot\rho(\mathbf{y})=\mathbf{x}\cdot\mathbf{y}. \tag{8.19}$$

Therefore every rotation is orthogonal. Hence, by Proposition 7.7, we have at once the following result.

Proposition 8.29. *Every rotation ρ of $\mathbb{R}^3$ is an orthogonal linear mapping. Consequently, the matrix of ρ with respect to the standard orthonormal basis of $\mathbb{R}^3$ is orthogonal.*

We will henceforth identify a rotation with its matrix with respect to the standard orthonormal basis, so that $\mathrm{Rot}(\mathbb{R}^3)\subset O(3,\mathbb{R})$. We will now classify the elements of $O(3,\mathbb{R})$ that are rotations.

Claim: every rotation ρ of $\mathbb{R}^3$ has determinant one. Indeed, a rotation ρ fixes a line L through the origin pointwise, so ρ has eigenvalue 1. Moreover, the plane

orthogonal to L is rotated through an angle θ, so there exists an orthonormal basis of $\mathbb{R}^3$ for which the matrix of ρ has the form

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{pmatrix}.$$

Hence $\det(\rho) = 1$, which gives the claim. Therefore, $\mathrm{Rot}(\mathbb{R}^3) \subset SO(3,\mathbb{R})$. We will now prove a theorem.

Theorem 8.30. $\mathrm{Rot}(\mathbb{R}^3) = SO(3,\mathbb{R})$.

Proof. It remains to show that $SO(3,\mathbb{R}) \subset \mathrm{Rot}(\mathbb{R}^3)$, i.e., that every element of $SO(3,\mathbb{R})$ is a rotation. Note that by our definition, the identity mapping I_3 is a rotation. Namely, I_3 is the rotation that fixes every line L through $\mathbf{0}$ and rotates every plane parallel to $L^\perp$ through zero degrees. Let $\sigma \in SO(3,\mathbb{R})$. I claim that 1 is an eigenvalue of σ, and moreover, if $\sigma \neq I_3$, the eigenspace E_1 of 1 is a line. To see this, we have to characterize σ's eigenvalues. Since complex eigenvalues occur in conjugate pairs, every 3×3 real matrix has a real eigenvalue; and since the real eigenvalues of an orthogonal matrix are either 1 or -1, the eigenvalues of σ are given by one of the following possibilities (recall that $\det(\sigma) = 1$):

(i) 1 of multiplicity three,

(ii) $1, -1$, where -1 has multiplicity two, and

(iii) 1, λ, $\overline{\lambda}$, where $|\lambda| = 1$ and $\lambda \neq \overline{\lambda}$ (since the complex roots of the characteristic polynomial of a real matrix occur in conjugate pairs).

In every case, 1 is an eigenvalue of σ, so $\dim E_1(\sigma) \geq 1$. Suppose $\sigma \neq I_3$ but $\dim E_1(\sigma) > 1$. Then $\dim E_1(\sigma) = 3$ is impossible, so $\dim E_1(\sigma) = 2$. Thus σ fixes the plane $E_1(\sigma)$ pointwise. Since σ preserves angles, it also has to send the line $L = E_1(\sigma)^\perp$ to itself. Thus L is an eigenspace. Since $\sigma \neq I_3$, the only possible eigenvalue for σ on L is -1. In this case, $\mathbb{R}^3$ has a basis, so that the matrix of σ is

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix},$$

contradicting the fact that $\det(\sigma) = 1$. Thus, if $\sigma \neq I_3$, $\dim E_1 = 1$. Therefore, σ fixes every point on a unique line L through the origin and maps the plane $L^\perp$ orthogonal to L into itself. For if $\mathbf{u} \in L$ and $\mathbf{u} \cdot \mathbf{v} = 0$, then

$$\mathbf{u} \cdot \mathbf{v} = \sigma(\mathbf{u}) \cdot \sigma(\mathbf{v}) = \mathbf{u} \cdot \sigma(\mathbf{v}) = 0.$$

Thus if $\mathbf{v} \in L^\perp$, then we must have $\sigma(\mathbf{v}) \in L^\perp$ as well. It remains to show that σ rotates $L^\perp$. Let $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3$ be an orthonormal basis in $\mathbb{R}^3$ such that

$\mathbf{u}_1, \mathbf{u}_2 \in L^\perp$ and $\mathbf{u}_3 \in L$, i.e., $\sigma(\mathbf{u}_3) = \mathbf{u}_3$. Since $\sigma\mathbf{u}_1$ and $\sigma\mathbf{u}_2$ are orthogonal unit vectors on $L^\perp$, we can choose an angle θ such that

$$\sigma\mathbf{u}_1 = \cos\theta\mathbf{u}_1 + \sin\theta\mathbf{u}_2$$

and

$$\sigma\mathbf{u}_2 = \pm(\sin\theta\mathbf{u}_1 - \cos\theta\mathbf{u}_2).$$

In matrix terms, this says that if $Q = (\mathbf{u}_1\ \mathbf{u}_2\ \mathbf{u}_3)$, then

$$\sigma Q = Q \begin{pmatrix} \cos\theta & \pm\sin\theta & 0 \\ \sin\theta & \pm(-\cos\theta) & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Since $\det(\sigma) = 1$ and $\det(Q) \neq 0$, it follows that

$$\det \begin{pmatrix} \cos\theta & \pm\sin\theta & 0 \\ \sin\theta & \pm(-\cos\theta) & 0 \\ 0 & 0 & 1 \end{pmatrix} = 1.$$

The only possibility is that

$$\sigma Q = Q \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{8.20}$$

Thus σ rotates the plane $L^\perp$ through θ, so it follows that $\sigma \in \mathrm{Rot}(\mathbb{R}^3)$. This proves that $SO(3,\mathbb{R}) = \mathrm{Rot}(\mathbb{R}^3)$. □

The fact that $\mathrm{Rot}(\mathbb{R}^3) = SO(3,\mathbb{R})$ gives an interesting corollary.

Corollary 8.31. $\mathrm{Rot}(\mathbb{R}^3)$ *is a matrix group. Hence, the composition of two rotations of* $\mathbb{R}^3$ *is another rotation.*

Remark. The fact that the composition of two rotations is a rotation is certainly not obvious from the definition of a rotation. This result is due to Euler. (Thus it may be said that Euler proved that $\mathrm{Rot}(\mathbb{R}^3)$ is a group before groups were defined. This is the second brush with groups associated with Euler, the first being the group U_m of multiplicative units in $\mathbb{Z}_m$.) The axis of the product of two rotations was found using what are called the Euler angles of the rotation. The most efficient way of finding the axis of the product of two rotations is to represent each rotation as a unit quaternion, say $\mathbf{q}_1$ and $\mathbf{q}_2$. Then the axis is read off from the quaternionic product $\mathbf{q} = \mathbf{q}_1\mathbf{q}_2$. The reader is referred to any article on unit quaternions for further details.

Notice that the orthogonal matrix Q defined above may be chosen to be an element of $SO(3,\mathbb{R})$. Therefore, the above argument gives another result.

Proposition 8.32. *Given $\sigma \in SO(3,\mathbb{R})$, there exists an element $Q \in SO(3,\mathbb{R})$ such that*

$$\sigma = Q \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix} Q^{-1}.$$

8.6.2 The Platonic solids

In order to apply our result about $SO(3,\mathbb{R})$, we will first define the Platonic solids and then determine their rotation groups. A half-space H in $\mathbb{R}^3$ consists of all points lying on or on one side of a plane P in $\mathbb{R}^3$. If P has the equation $ax + by + cz = d$, then there are two half-spaces, P_+ and P_-, which are defined respectively by the inequalities $ax + by + cz \geq d$ and $ax + by + cz \leq d$. A set R in $\mathbb{R}^3$ is called *bounded* if there exists $M > 0$ such that $|\mathbf{r}| < M$ for all $\mathbf{r} \in R$. It is *convex* if for every two points $\mathbf{x}, \mathbf{y}$ in R, the straight line segment between $\mathbf{x}$ and $\mathbf{y}$ is also in R. The intersection of a collection of half-spaces in $\mathbb{R}^3$ is always convex.

Definition 8.4. A *convex polyhedron* $\mathcal{P}$ in $\mathbb{R}^3$ is by definition a bounded region in $\mathbb{R}^3$ that is the intersection of a finite number of half-spaces in $\mathbb{R}^3$.

It turns out that we can require that the half-spaces $H_1, \ldots, H_k$ that define $\mathcal{P}$ have the property that each $F_i = \mathcal{P} \cap H_i$ is a polygon in $\mathbb{R}^3$. These polygons are the *faces* of $\mathcal{P}$. Two distinct faces F_i and F_j are either disjoint, meet in a common vertex, or meet along an edge common to both. The vertices and edges of all the F_i constitute the sets *vertices* and *edges* of $\mathcal{P}$. The *boundary* of $\mathcal{P}$ is the union of all its faces F_i. For example, a cube is a convex polyhedron whose boundary is made up of 6 faces, 12 edges, and 8 vertices. Notice that the faces of a cube are squares of side 1. Hence they are regular polygons of type $\{4\}$ in the notation of Section 7.3.7. To give an idea of the startlingly original ideas of Euler, yet another of his famous (and surprising) theorems states that for every convex polyhedron in $\mathbb{R}^3$ with V vertices, E edges, and F faces, $F - E + V = 2$. This number, which can be also be defined for the boundaries of arbitrary piecewise linear solids in $\mathbb{R}^3$, is called the Euler characteristic. For example, for the surface of a piecewise linear doughnut in $\mathbb{R}^3$, $F - E + V = 0$.

A convex polyhedron $\mathcal{P}$ with the property that all faces of $\mathcal{P}$ are congruent regular polygons and every vertex is on the same number of faces is called a *Platonic solid*. Up to position in $\mathbb{R}^3$ and volume, there are exactly five Platonic solids: the tetrahedron, the cube, the octahedron, the dodecahedron, and the icosahedron (or soccer ball). The Platonic solids were known to the classical Greek mathematician-philosophers. Plato famously attempted to associate four of them with the classical elements (earth, air, fire, and water), and later Kepler attempted to improve on Plato by associating them with the

known planets. Euclid proved in the *Elements* that the Platonic solids fall into the five classes of convex polyhedra mentioned just above.

Not surprisingly, all convex polyhedra are determined by their vertices. Hence the Platonic solids can be described by giving coordinates for their vertices. All Platonic solids have Euler characteristic 2. The most familiar one is the cube $\mathcal{C}$, which has 8 vertices $(\pm 1, \pm 1, \pm 1)$, 12 edges, and 6 faces. (Note that $F - E + V = 2$.) A tetrahedron has four vertices and four faces, which are equilateral triangles. Since $F + V = 2 + E$, it has six edges. A convenient tetrahedron that we will call $\mathcal{T}$ has cubical vertices $(1, 1, 1), (1, -1, -1), (-1, 1, -1), (-1, -1, 1)$. The octahedron $\mathcal{O}$ has 6 vertices, which we will take to be the midpoints of the faces of $\mathcal{C}$, namely $(\pm 1, 0, 0), (0, \pm 1, 0), (0, 0, \pm 1)$. It has 8 triangular faces, hence 12 edges. The icosahedron $\mathcal{I}$ has 12 vertices $(0, \pm 1, \pm\phi)$, $(\pm 1, \pm\phi, 0)$, $(\pm\phi, 0, \pm 1)$, 20 faces, and 30 edges. Finally, the dodecahedron $\mathcal{D}$ has 20 vertices: 8 of which are vertices of $\mathcal{C}$, $(0, \pm\phi^{-1}, \pm\phi)$, $(\pm\phi, 0, \pm\phi^{-1})$, $(\pm\phi, 0, \pm\phi^{-1})$. Here, $\phi = \frac{1+\sqrt{5}}{2}$ is the golden mean, which was encountered when we considered the Fibonacci sequence. Notice that all five convex polyhedra listed above have the property that their vertices are equidistant from the origin. Such a convex polyhedron is said to be *central*.

8.6.3 The rotation group of a Platonic solid

The set $\mathrm{Rot}(\mathcal{P})$ consisting of rotations of $\mathbb{R}^3$ that preserve a convex polyhedron $\mathcal{P}$ is called the *rotation group* of $\mathcal{P}$.

Proposition 8.33. *Suppose $\mathcal{P}$ is a convex polyhedron whose vertices span $\mathbb{R}^3$. Then $\mathrm{Rot}(\mathcal{P})$ is a finite subgroup of $SO(3, \mathbb{R})$.*

Proof. A convex polyhedron $\mathcal{P}$ is uniquely determined by its vertices, and every linear mapping of $\mathbb{R}^3$ sends each convex polyhedron $\mathcal{P}$ to another convex polyhedron. Thus if a linear mapping sends $\mathcal{P}$ into itself, it has to preserve the faces, edges and vertices of $\mathcal{P}$. Since a rotation of $\mathbb{R}^3$ is linear and since the vertices of $\mathcal{P}$ span $\mathbb{R}^3$, two rotations that coincide on the vertices are the same. Since the vertex set is finite, it follows that $\mathrm{Rot}(\mathcal{P})$ has to be finite too. □

The finite subgroups of $SO(3, \mathbb{R})$ are called the *polyhedral groups*. Thus, $\mathrm{Rot}(\mathcal{P})$ is always a polyhedral group. Conversely, every polyhedral group is the rotation group of a convex polyhedron. We will eventually classify all the polyhedral groups. When $\mathcal{P}$ is a central Platonic solid, there is a beautiful formula for the order $|\mathrm{Rot}(\mathcal{P})|$:

Proposition 8.34. *Suppose $\mathcal{P}$ is a central Platonic solid with f faces such that each face has e edges. Then $|\mathrm{Rot}(\mathcal{P})| = ef$.*

The proof will be given in Proposition 11.7 as an application of the orbit stabilizer theorem. Thus the rotation group of a central cube $\mathcal{C}$ or octahedron $\mathcal{O}$ has order 24. As we will prove next, both of these groups are isomorphic to the symmetric group $S(4)$. A central regular tetrahedron has four triangular faces, so its rotation group has order 12. A central regular dodecahedron $\mathcal{D}$ and icosahedron $\mathcal{I}$ both have rotation groups of order 60. In fact, $\mathrm{Rot}(\mathcal{I}) = \mathrm{Rot}(\mathcal{D}) \cong A(5)$.

8.6.4 The cube and the octahedron

Since the rotation group of a central cube $\mathcal{C}$ has order 24, one might suspect that it is isomorphic to $S(4)$. We will verify this, but first let us give an explicit description of $\mathrm{Rot}(\mathcal{C})$. For convenience, suppose the six vertices of $\mathcal{C}$ are $(\pm 1, \pm 1, \pm 1)$. Since all rotations of $\mathcal{C}$ send faces to faces, they also permute the six midpoints of the faces. The midpoints being $\pm\mathbf{e}_1, \pm\mathbf{e}_2$, and $\pm\mathbf{e}_3$, the orthogonal matrices that permute these vectors have the form

$$\sigma = (\pm\mathbf{e}_{\pi(1)} \ \pm\mathbf{e}_{\pi(2)} \ \pm\mathbf{e}_{\pi(3)}),$$

where $\pi \in S(3)$. The matrices of this form are called *signed permutation matrices*. They form the subgroup $SP(3)$ of $O(3, \mathbb{R})$. Note that $|SP(3)| = 48$. Observe that if $\sigma \in SP(3)$, then

$$\det \sigma = \det \ (\pm\mathbf{e}_{\pi(1)} \ \pm\mathbf{e}_{\pi(2)} \ \pm\mathbf{e}_{\pi(3)}) = (-1)^r \mathrm{sgn}(\pi),$$

where r is the number of -1's. This implies that $|SP(3) \cap SO(3, \mathbb{R})| = 24$. It is evident that $\mathrm{Rot}(\mathcal{C}) = SP(3) \cap SO(3, \mathbb{R})$, so we get another proof that $|\mathrm{Rot}(\mathcal{C})| = 24$. For simplicity, let G denote $\mathrm{Rot}(\mathcal{C})$. To see that $G \cong S(4)$, let D_1, D_2, D_3, D_4 denote the four diagonals of $\mathcal{C}$. Since G leaves $\mathcal{C}$ invariant, it follows that each element of G permutes the four diagonals. Thus if $\sigma \in G$, then $\sigma(D_i) = D_{\pi(i)}$ for a unique $\pi \in S(4)$. We will show that the mapping $\varphi : G \to S(4)$ given by $\varphi(\sigma) = \pi$ is an isomorphism. The proof that φ is a homomorphism is left to the reader. To show that φ is a bijection, it suffices, by the pigeonhole principle, to show that φ is injective, since G and $S(4)$ have the same order. Consider the four vectors $\mathbf{d}_1 = \mathbf{e}_1 + \mathbf{e}_2 + \mathbf{e}_3$, $\mathbf{d}_2 = -\mathbf{e}_1 + \mathbf{e}_2 + \mathbf{e}_3$, $\mathbf{d}_3 = \mathbf{e}_1 - \mathbf{e}_2 + \mathbf{e}_3$, and $\mathbf{d}_4 = -\mathbf{e}_1 - \mathbf{e}_2 + \mathbf{e}_3$, which lie on the four diagonals of $\mathcal{C}$ and point upward. Let D_i be the diagonal determined by $\mathbf{d}_i$. Then for each i, $\sigma(\mathbf{d}_i) = \epsilon_i \mathbf{d}_{\pi(i)}$, where each ϵ_i is either 1 or -1 depending on whether $\sigma(\mathbf{d}_i)$ points up or down. Since $\{\mathbf{d}_1, \mathbf{d}_2, \mathbf{d}_3\}$ and $\{\epsilon_1 \mathbf{d}_{\pi(1)}, \epsilon_2 \mathbf{d}_{\pi(2)}, \epsilon_3 \mathbf{d}_{\pi(3)}\}$ are both bases of $\mathbb{R}^3$ for every choice of signs and each $\sigma \in \mathrm{Rot}(\mathcal{C})$ is a linear mapping, it follows that σ is uniquely determined by the matrix identity

$$\sigma\left(\mathbf{d}_1\ \mathbf{d}_2\ \mathbf{d}_3\right) = \left(\sigma(\mathbf{d}_1)\ \sigma(\mathbf{d}_2)\ \sigma(\mathbf{d}_3)\right) = \left(\epsilon_1\mathbf{d}_{\pi(1)}\ \epsilon_2\mathbf{d}_{\pi(2)}\ \epsilon_3\mathbf{d}_{\pi(3)}\right).$$

Now assume $\varphi(\sigma) = (1)$. Then

$$\sigma\left(\mathbf{d}_1\ \mathbf{d}_2\ \mathbf{d}_3\right) = \left(\epsilon_1\mathbf{d}_1\ \epsilon_2\mathbf{d}_2\ \epsilon_3\mathbf{d}_3\right) = \left(\mathbf{d}_1\ \mathbf{d}_2\ \mathbf{d}_3\right)\operatorname{diag}(\epsilon_1, \epsilon_2, \epsilon_3).$$

Hence, $\sigma = D\mathcal{E}D^{-1}$, where $D = \left(\mathbf{d}_1\ \mathbf{d}_2\ \mathbf{d}_3\right)$ and $\mathcal{E} = \operatorname{diag}(\epsilon_1, \epsilon_2, \epsilon_3)$. In order to be in G, σ must be orthogonal, with $\det\sigma = 1$. In particular, $\sigma^{-1} = \sigma^T$ and $\det\mathcal{E} = 1$. Since $\mathcal{E}^{-1} = \mathcal{E}$, it follows that $\sigma = \sigma^{-1}$; hence to be orthogonal, σ must be symmetric. This implies that $D\mathcal{E}D^{-1} = (D^T)^{-1}\mathcal{E}D^T$, or equivalently, $D^TD\mathcal{E} = \mathcal{E}D^TD$. By direct calculation,

$$D^TD = \begin{pmatrix} 3 & 1 & 1 \\ 1 & 3 & -1 \\ 1 & -1 & 3 \end{pmatrix},$$

so the only way $D^TD\mathcal{E} = \mathcal{E}D^TD$ can occur with $\mathcal{E} = \operatorname{diag}(\pm1, \pm1, \pm1)$ and $\det\mathcal{E} = 1$ is if $\mathcal{E} = I_3$, for at least one entry of $\mathcal{E}$, say ϵ_i, must be 1. If the other diagonal entries are -1, then $\mathcal{E}D^TD \neq D^TD\mathcal{E}$. For example, when $\epsilon_1 = \epsilon_3 = -1$ and $\epsilon_2 = 1$, then the first column of $D^TD\mathcal{E}$ is $\begin{pmatrix} -3 \\ -1 \\ -1 \end{pmatrix}$, while that of $\mathcal{E}D^TD$ is $\begin{pmatrix} -3 \\ 1 \\ -1 \end{pmatrix}$. We conclude that $\sigma = I_3$, and consequently, φ is injective. This proves that φ is an isomorphism, so $\operatorname{Rot}(\mathcal{C}) \cong S(4)$. □

The octahedron $\mathcal{O}$ can be viewed as the unique convex polyhedron whose vertices are the midpoints of the faces of the cube $\mathcal{C}$, and every rotation of $\mathcal{C}$ is thus a rotation of the octahedron and conversely. Thus $\operatorname{Rot}(\mathcal{C}) = \operatorname{Rot}(\mathcal{O})$.

Perhaps a more enlightening way to realize $\operatorname{Rot}(\mathcal{C})$ is to describe the rotations directly by finding the axes about which $\mathcal{C}$ can be rotated. For example, every coordinate axis $\mathbb{R}\mathbf{e}_i$ is such a line; C can be rotated by $\pi/2$, π, and $3\pi/2$ about each coordinate axis. This gives a total of nine distinct rotations. One can also rotate C through π around the lines $x = y$ and $x = -y$ in the xy-plane. Repeating this for the other two coordinate planes gives six more rotations, so we have now accounted for 15 rotations, 16 including the identity. There are four more lines of symmetry about which one might be able to rotate C, namely the four diagonals joining opposite vertices. It seems to be a little harder to visualize whether there any rotations through these lines, so let us take a slightly different approach. Recall that the alternating group $A(3)$ can be realized as the 3×3 permutation matrices of determinant 1. One element $\sigma \in A(3)$ is the rotation

$$\sigma = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$$

sending $\mathbf{e}_1 \to \mathbf{e}_2 \to \mathbf{e}_3 \to \mathbf{e}_1$. Clearly, $\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$ is an eigenvector, so φ is in fact a rotation about the line $\mathbb{R}\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$. Since σ clearly has order 3, it is the rotation through $2\pi/3$. In this way, we account for another eight elements of $\mathrm{Rot}(C)$, since there are four diagonals, so we now have constructed all 24 rotations of C.

8.6.5 Symmetry groups

The geometric side of group theory is symmetry. To understand this aspect, suppose S is a subset of $\mathbb{R}^n$. A *symmetry of* S is defined to be an orthogonal linear mapping $\varphi : \mathbb{R}^n \to \mathbb{R}^n$ such that $\varphi(S) = S$. The set of all symmetries of S will be denoted by $\mathcal{O}(S)$. Recall that an orthogonal linear mapping is a linear map φ that preserves the inner product on $\mathbb{R}^n$. Since φ must also preserve lengths, distances, and angles, it preserves the geometry of S. We know that an orthogonal linear mapping $\varphi : \mathbb{R}^n \to \mathbb{R}^n$ is an orthogonal $n \times n$ matrix, so the set of symmetries of S is thus

$$\mathcal{O}(S) = \{\varphi \in O(n, \mathbb{R}) \mid \varphi(S) = S\}.$$

In particular, $\mathcal{O}(\mathbb{R}^n) = O(n, \mathbb{R})$. The first thing to note is the following.

Proposition 8.35. *For every $S \subset \mathbb{R}^n$, the set of symmetries $\mathcal{O}(S)$ is a subgroup of $O(n, \mathbb{R})$.*

Proof. If φ and ψ are elements of $\mathcal{O}(S)$, then $\psi\varphi$ and φ^{-1} are also symmetries of S. Hence $\psi\varphi^{-1} \in \mathcal{O}(S)$, so $\mathcal{O}(S)$ is indeed a subgroup. □

Of course, it is possible, in fact likely, that the only element of $\mathcal{O}(S)$ is I_n. For example, let $n = 2$ and let $S = \{(1, 0), (2, 0), (3, 0)\}$. On the other hand, if S is the unit circle $x^2 + y^2 = 1$ in $\mathbb{R}^2$, then $\mathcal{O}(S) = O(2, \mathbb{R})$. When S is a convex polyhedron in $\mathbb{R}^3$ (or even in $\mathbb{R}^n$), then $\mathcal{O}(S)$ has to permute the vertices of S. But since elements of $\mathcal{O}(S)$ preserve lengths, $\mathcal{O}(S)$ can move a vertex only into another vertex having the same distance from the origin. But if S is contained in $\mathbb{R}^n$ and has n linearly independent vertices of different lengths, then $\mathcal{O}(S)$ is the trivial group.

Example 8.17. Let us consider the n-cube $C(n)$ in $\mathbb{R}^n$ defined by

$$C(n) = \{(x_1, x_2, \ldots, x_n) \mid -1 \leq x_i \leq 1 \text{ for all } i = 1, 2, \ldots, n\}.$$

The matrix group $SP(n)$ consisting of all $n \times n$ matrices of the form

$$(\pm \mathbf{e}_{\pi(1)} \ \pm \mathbf{e}_{\pi(2)} \ \cdots \ \pm \mathbf{e}_{\pi(n)}),$$

where $\pi \in S(n)$, permutes the vectors $\pm\mathbf{e}_1, \pm\mathbf{e}_2, \ldots, \pm\mathbf{e}_n$ and hence sends $C(n)$ to $C(n)$. In fact, $SP(n) = \mathcal{O}(C(n))$. ⊠

Rotations in $\mathbb{R}^n$ for $n > 3$ are harder to define. We know that every rotation of $\mathbb{R}^3$ is given by an element of $SO(3, \mathbb{R})$. It can be shown by a similar analysis that every element of $SO(4, \mathbb{R})$ is given by a matrix that is a rotation in two orthogonal planes in $\mathbb{R}^4$. In general, elements of $O(n, \mathbb{R})$ satisfy the normal matrix criterion $A^T A = AA^T$. This implies, by the normal matrix theorem (see Theorem 9.2), that A is unitarily diagonalizable. That is, for every $A \in O(n, \mathbb{R})$, there is a Hermitian orthonormal basis of $\mathbb{C}^n$ consisting of eigenvectors of A. Then, associated to every pair of eigenvalues $\lambda, \overline{\lambda}$ of A, there exists a two-plane V in $\mathbb{R}^n$ such that A is a rotation of V through $e^{i\lambda}$. (Recall that since A is orthogonal, its eigenvalues satisfy $|\lambda| = 1$.)

Remark. The study of symmetry via group theory has been successful in several disciplines, e.g., chemistry, physics, and materials science. The symmetries of a class of pure carbon molecules called fullerenes offer a prime example. The most widely known of the fullerenes is a carbon molecule denoted by C_{60} (not to be confused with the cyclic group of order 60), named buckminsterfullerene (or buckyball for short) after Buckminster Fuller for its similarity to his famous geodesic domes, which is one of the most rigid molecules ever discovered. Buckminsterfullerene is a truncated icosahedron. To obtain a picture of the buckyball, we have to consider the truncated icosahedron, which is the solid obtained by cutting off each vertex of the icosahedron by a plane orthogonal to the line through the vertex and the center of the icosahedron at the same distance from the center for each vertex. To get a model for the truncated icosahedron, just take a close look at a soccer ball. Since every vertex of the icosahedron lies on five faces, the vertices are replaced by 12 regular pentagons. Hence, the truncated icosahedron has 60 vertices. It also has 32 faces. The symmetry group of the vertices of the truncated icosahedron is the same as for the icosahedron.

Exercises

Exercise 8.6.1. Prove the following: Suppose G is a subgroup of $O(3, \mathbb{R})$ and let

$$G_+ = G \cap SO(3, \mathbb{R}).$$

Then either $G_+ = G$ or G has exactly two cosets, and $|G| = 2|G_+|$.

Exercise 8.6.2. Let $\mathcal{P}$ be a central Platonic solid. Let $\mathcal{O}(\mathcal{P})$ be the set of all orthogonal matrices preserving $\mathcal{P}$.

(i) Verify that $\mathcal{O}(\mathcal{P})$ is a subgroup of $O(3, \mathbb{R})$.

(ii) Show that $|\mathcal{O}(\mathcal{P})| = 2|\mathrm{Rot}(\mathcal{P})|$.

Exercise 8.6.3. Consider the cube $\mathcal{C}$ with vertices at $(\pm 1, \pm 1, \pm 1)$. For the rotations of $\mathcal{C}$ through the diagonal along $(1, 1, 1)$, where do the vertices go? Even though this was worked out in the text, try to do it anyway without looking back.

Exercise 8.6.4. Let H denote the reflection through the xy-plane in $\mathbb{R}^3$. Show how to express the reflection through the yz-plane in the form $H\sigma$, where σ is a rotation.

Exercise 8.6.5. The symmetry group of the central 3-cube $\mathcal{C}$ of the previous exercise permutes the diagonals of $\mathcal{C}$, but it has order 48, which is twice the order of $S(4)$. Describe all the nonrotational symmetries.

Exercise 8.6.6. Show that

$$SP(n) = \{\sigma \in O(n, \mathbb{R}) \mid \sigma = (\pm \mathbf{e}_{\pi(1)} \ \pm \mathbf{e}_{\pi(2)} \ \cdots \ \pm \mathbf{e}_{\pi(n)}),\ \pi \in S(n)\}$$

is a subgroup of $O(n, \mathbb{R})$ of order $2^n n!$. Elements of $SP(n)$ are called *signed permutation matrices*.

Exercise 8.6.7. Consider the n-cube $C(n)$ with its 2^n vertices at the points $(\pm 1, \pm 1, \ldots, \pm 1)$. Show that the symmetry group of the set of vertices of $C(n)$ is the group $SP(n)$. Does this imply that the symmetry group of $C(n)$ is $SP(n)$?

Exercise 8.6.8. The 4-cube $C(4)$ has eight diagonals. They are represented by the semidiagonals $\pm\mathbf{e}_1 + \pm\mathbf{e}_2 + \pm\mathbf{e}_3 + \mathbf{e}_4$. Show that the symmetry group of $C(4)$ does not act transitively on the diagonals. Find a pair of diagonals D_1 and D_2 such that no $\sigma \in \mathrm{Sym}(C(4))$ satisfies $\sigma(D_1) = D_2$.

Exercise 8.6.9. Let $\mathcal{T}$ be the central regular tetrahedron in $\mathbb{R}^3$ described above. Show that the group $\mathcal{O}(\mathcal{T})$ is isomorphic to $S(4)$ and that $\mathrm{Rot}(\mathcal{T})$ is isomorphic to $A(4)$.

Exercise 8.6.10. Construct a convex polyhedron P by taking two copies of a tetrahedron, say T_1 and T_2, and gluing them together along two faces. The result is a convex polyhedron with six faces, nine edges, and five vertices. Assuming that P is central, compute the order of $\mathcal{O}(P)$.

Exercise 8.6.11. For a subset S of $\mathbb{R}^3$, let $\mathrm{Rot}(S)$ denote the group of all $\sigma \in SO(3,\mathbb{R})$ such that $\sigma(S) = S$. Find $\mathrm{Rot}(S)$ in the following cases:

(a) S is the half-ball $\{x^2 + y^2 + z^2 \leq 1, z \geq 0\}$,

(b) S is the solid rectangle $\{-1 \leq x \leq 1, -2 \leq y \leq 2, -1 \leq z \leq 1\}$.

Exercise 8.6.12. Suppose H is a reflection of $\mathbb{R}^2$. Show that there is a rotation ρ of $\mathbb{R}^3$ such that $\rho(\mathbf{x}) = H(\mathbf{x})$ for all $\mathbf{x} \in \mathbb{R}^2$. (Hint: consider the line through which H reflects $\mathbb{R}^2$.)

Exercise 8.6.13. Let G be a subgroup of $O(n,\mathbb{R})$. True or false: if G is not contained in $SO(n,\mathbb{R})$, then G is normal in $O(n,\mathbb{R})$.

Exercise 8.6.14. Describe how the alternating group $A(4)$ acts on the cube in $\mathbb{R}^3$ with vertices $(\pm 1, \pm 1, \pm 1)$.

8.7 An Appendix on Field Extensions

The purpose of this Appendix is to prove that given a field $\mathbb{F}$ and a polynomial $f \in \mathbb{F}[x]$, there exists a field $\mathbb{F}'$ containing both $\mathbb{F}$ and all the roots of $f(x) = 0$. The field $\mathbb{F}'$ is called a *splitting field* for the polynomial f.

To begin, we will construct a field $\mathbb{F}'$ containing $\mathbb{F}$ and at least one root of f. Let x be a variable, and put $V = \mathbb{F}[x]$. Then V is a vector space over $\mathbb{F}$ with an infinite basis $1,\ x,\ x^2, \ldots$. Given a nonconstant polynomial $f \in V$, let W be the subspace of V consisting of all polynomials of the form $h = gf$, for some $g \in \mathbb{F}[x]$. We leave it to the reader to check that W is indeed a subspace of V. Now form the quotient vector space V/W. Recall that the elements of V/W are cosets $g + W$, where $g \in V$, and that coset addition is given by $(g + W) + (h + W) = (g + h) + W$. Scalar multiplication by $a \in \mathbb{F}$ is given in an analogous way: $a(g + W) = ag + W$. Recall also that two cosets $g + W$ and $h + W$ are the same if and only if $h - g \in W$. That is, $h - g = qf$ for some $q \in \mathbb{F}[x]$. Since both V and W are infinite-dimensional, the following result may at first glance be surprising.

Proposition 8.36. *The quotient vector space V/W is a finite-dimensional vector space over $\mathbb{F}$. In fact,* $\dim V/W = \deg(f)$.

Proof. To prove that $\dim V/W = \deg(f)$, we exhibit a basis. Let $k = \deg(f) - 1$. Given $g \in \mathbb{F}[x]$, put $\overline{g} = g + W$. For convenience, let us set $\alpha = \overline{x}$ and $\alpha^i = \overline{x^i}$ for each nonnegative integer i. We claim that $\overline{1}, \alpha, \alpha^2, \ldots, \alpha^k$ is a basis of V/W. We first show independence. Suppose there exist $a_0, a_1, \ldots, a_k \in \mathbb{F}$ such that

$$\sum_{i=0}^{k} a_x \alpha^i = \overline{\mathbf{0}}.$$

By definition, this means that

$$h(x) = \sum_{i=0}^{k} a_i x^i \in W.$$

Thus $h = gf$ for some $g \in \mathbb{F}[x]$. This is impossible unless $g = 0$, so all a_i are equal to zero. To show that $\overline{1}, \alpha, \ldots, \alpha^k$ span, let $g \in \mathbb{F}[x]$ and apply division with remainder to write $g = qf + r$, where $q, r \in \mathbb{F}[x]$ and $\deg(r) < \deg(f)$. Then $\overline{g} = \overline{r}$. But $\overline{r}$ is in the span of $\overline{1}, \alpha, \ldots, \alpha^k$, so we have found a basis of V/W. Thus $\dim V/W = \deg(f)$, which finishes the proof. □

An element $f \in \mathbb{F}[x]$ is said to be *irreducible* if there is no factorization $f = gh$ in which both g and h lie in $\mathbb{F}[x]$ and both g and h have positive degree. The next theorem gives an important and fundamental result in field theory.

Theorem 8.37. *If $f \in \mathbb{F}[x]$ is irreducible, then V/W can be given the structure of a field $\mathbb{F}'$ such that $\mathbb{F}$ is a subfield of $\mathbb{F}'$. Moreover, $\alpha = x + W$ is a root of f in $\mathbb{F}'$.*

Proof. We must first define multiplication on V/W. Let $g, h \in \mathbb{F}[x]$, and put $\overline{g}\,\overline{h} = \overline{gh}$. To prove that this definition makes sense, it is necessary to show that if $\overline{g_1} = \overline{g_2}$ and $\overline{h_1} = \overline{h_2}$, then $\overline{g_1 h_1} = \overline{g_2 h_2}$. This is analogous to the proof that $\mathbb{Z}_m$ admits an associative and commutative multiplication given in Chap. 2, so we will omit the details. Note that $\overline{\mathbf{0}}$ is the additive identity, and $\overline{1}$ is the multiplicative identity. It remains to prove that if $\overline{g} \neq \overline{\mathbf{0}}$, then $\overline{g}^{-1}$ exists. That is, there exists $h \in \mathbb{F}[x]$ such that $\overline{h}\,\overline{g} = \overline{1}$. Since f is irreducible and $\overline{g} \neq \overline{\mathbf{0}}$, f by definition doesn't divide g. Therefore, f and g have no common factor of positive degree. This means there exist polynomials $a, b \in \mathbb{F}[x]$ such that $af + bg = 1$, by the algorithm for computing the greatest common divisor of two polynomials. Consequently, in $\mathbb{F}'$, $\overline{b}\,\overline{g} = \overline{1}$. Hence V/W with this multiplication is a field. We now have $\mathbb{F}'$. Note that $\mathbb{F}$ is contained in $\mathbb{F}'$ as the subfield $\{\overline{r} \mid r \in \mathbb{F} \subset \mathbb{F}[x]\}$. Finally, we must show that $f(\alpha) = \overline{\mathbf{0}}$. Let $f(x) = \sum c_i x^i$. Now,

$$f(\alpha) = \sum c_i \alpha^i = \sum c_i \overline{x^i} = \overline{f} = \overline{\mathbf{0}},$$

so α is indeed a root. Therefore, the proof is finished. □

Example 8.18. Let $\mathbb{F} = \mathbb{Q}$ and notice that $x^2 + x - 1$ is irreducible in $\mathbb{Q}[x]$. Its roots ϕ and μ were considered in Section 8.2.5. In particular, $\phi = \frac{1+\sqrt{5}}{2}$. The field $\mathbb{Q}'$ has dimension two over $\mathbb{Q}$. A vector space basis of $\mathbb{Q}'$ over $\mathbb{Q}$ is $1, \sqrt{5}$. ⊠

Finally, we need to modify the above construction to get a splitting field $\mathbb{F}'$ for f. Notice that $f \in \mathbb{F}[x] \subset \mathbb{F}'[x]$, but f is no longer irreducible in $\mathbb{F}'[x]$, since it has the root $\alpha = \overline{x}$ in $\mathbb{F}'$. Choose a new variable, say z, and consider $\mathbb{F}'[z]$. Dividing $f(z)$ by $(z - \alpha)$ gives $f(z) = g(z)(z - \alpha)$, for some $g \in \mathbb{F}'[z]$. Now $\deg(g) = \deg(f) - 1$, so if $\deg(f) = 2$, then Proposition 8.36 implies that $\mathbb{F}'$ necessarily contains all roots of f, as in the above example. If $\deg(f) > 2$, we check whether g is irreducible in $\mathbb{F}'[z]$. If so, we repeat the construction with g to obtain a field extension of $\mathbb{F}'$ (and hence $\mathbb{F}$) containing a root β of g different from α. Of course, β is also root of f. If g isn't irreducible, factor it until another irreducible factor is found and then perform another extension. By continuing in this manner, one eventually obtains a field extension of $\mathbb{F}$ containing all roots of the original polynomial f. In fact, this process will stop after at most $\deg(f)$ steps, since a polynomial of degree m has at most m distinct roots. □

Chapter 9
Unitary Diagonalization and Quadratic Forms

As we saw in Chap. 8, when V is a finite-dimensional vector space over $\mathbb{F}$, then a linear mapping $T : V \to V$ is semisimple if and only if its eigenvalues lie in $\mathbb{F}$ and its minimal polynomial has only simple roots. It would be useful to have a result that would allow one to predict that T is semisimple on the basis of a criterion that is simpler than finding the minimal polynomial, which, after all, requires knowing the roots of the characteristic polynomial. In fact, we also proved that when $\mathbb{F} = \mathbb{C}$ or $\mathbb{R}$, every self-adjoint operator is semisimple and even admits a Hermitian orthonormal basis. The matrices associated to self-adjoint operators, that is, Hermitian and symmetric matrices respectively, happen to be in a larger class of matrices said to be normal consisting of all $A \in \mathbb{C}^{n\times n}$ such that $AA^H = A^H A$. The normal matrix theorem asserts that the normal matrices are exactly those $A \in \mathbb{C}^{n\times n}$ that can be unitarily diagonalized, or equivalently, admit a Hermitian orthonormal basis. The first goal in this chapter is to prove this theorem. The second goal is to consider the topic of quadratic forms to which our diagonalization results may be applied. For example, we will classify the positive definite real quadratic forms or equivalently, the positive definite symmetric matrices. We will also introduce an equivalence relation on the set of all real quadratic forms called congruence and classify the equivalence classes. This result is known as Sylvester's law of inertia.

9.1 Schur Triangularization and the Normal Matrix Theorem

The key to understanding which matrices can be unitarily diagonalized is the Schur triangularization theorem, which says that an arbitrary $n \times n$ matrix over $\mathbb{C}$ is similar via a unitary matrix to an upper triangular matrix. Put

J.B. Carrell, *Groups, Matrices, and Vector Spaces*,
DOI 10.1007/978-0-387-79428-0_9

another way, if V is a finite-dimensional vector space over $\mathbb{C}$ with a Hermitian inner product, then every linear mapping $T : V \to V$ can be represented in a Hermitian orthonormal basis by an upper triangular matrix. One of the interesting aspects of the proof of the Schur triangularization theorem is that it uses the fact that the unitary group $U(n)$ is closed under multiplication.

9.1.1 Upper triangularization via the unitary group

We will now prove the Schur triangularization theorem.

Theorem 9.1 (Schur triangularization theorem)**.** *Suppose $A \in \mathbb{C}^{n\times n}$. Then there exist a unitary matrix U and an upper triangular matrix $T \in \mathbb{C}^{n\times n}$ such that $A = UTU^H$. Thus, every square matrix over $\mathbb{C}$ is unitarily similar to an upper triangular matrix over $\mathbb{C}$. If A is real and all its eigenvalues are also real, then A is similar to an upper triangular matrix over $\mathbb{R}$ via an orthogonal matrix.*

Proof. We will induct on n. Since the result is trivial if $n = 1$, suppose $n > 1$ and that the proposition is true for all $k \times k$ matrices over $\mathbb{C}$ whenever $k < n$. Since A's eigenvalues lie in $\mathbb{C}$, A has an eigenpair $(\lambda_1, \mathbf{u}_1)$ with $\lambda_1 \in \mathbb{C}$ and $\mathbf{u}_1 \in \mathbb{C}^n$. Let W be the subspace $(\mathbb{C}\mathbf{u}_1)^\perp$. By the Hermitian version of Proposition 6.38, $\dim W = n - 1$. Hence by the above discussion, there exists a Hermitian orthonormal basis $\{\mathbf{u}_2, \ldots, \mathbf{u}_n\}$ of W. Adjoining $\mathbf{u}_1$ gives the Hermitian orthonormal basis $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n$ of $\mathbb{C}^n$. Thus $U_1 = (\mathbf{u}_1\ \mathbf{u}_2\ \ldots \mathbf{u}_n)$ is unitary, and since $(\lambda_1, \mathbf{u}_1)$ is an eigenpair for A, we have

$$AU_1 = (A\mathbf{u}_1\ A\mathbf{u}_2\ \cdots A\mathbf{u}_n) = (\lambda_1\mathbf{u}_1\ A\mathbf{u}_2\ \cdots A\mathbf{u}_n).$$

Hence

$$U_1^H AU_1 = \begin{pmatrix} \mathbf{u}_1^H \\ \mathbf{u}_2^H \\ \vdots \\ \mathbf{u}_n^H \end{pmatrix} (\lambda_1\mathbf{u}_1\ A\mathbf{u}_2\ \cdots A\mathbf{u}_n) = \begin{pmatrix} \lambda_1\mathbf{u}_1^H\mathbf{u}_1 & * & \cdots & * \\ \lambda_1\mathbf{u}_2^H\mathbf{u}_1 & * & \cdots & * \\ \vdots & \vdots & & \vdots \\ \lambda_1\mathbf{u}_n^H\mathbf{u}_1 & * & \cdots & * \end{pmatrix}.$$

Therefore,

$$U_1^H AU_1 = \begin{pmatrix} \lambda_1 & * & \cdots & * \\ 0 & * & \cdots & * \\ \vdots & \vdots & & \vdots \\ 0 & * & \cdots & * \end{pmatrix}. \tag{9.1}$$

Now apply the induction hypothesis to the $(n-1)\times(n-1)$ matrix B in the lower right-hand corner of $U_1^H A U_1$ to get an $(n-1)\times(n-1)$ unitary matrix U' such that $(U')^H B U'$ is upper triangular. The matrix

$$U_2 = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & U' & \\ 0 & & & \end{pmatrix}$$

is clearly unitary, and

$$U_2^H (U_1^H A U_1) U_2 = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & (U')^H & \\ 0 & & & \end{pmatrix} \begin{pmatrix} \lambda_1 & * & \cdots & * \\ 0 & & & \\ \vdots & & B & \\ 0 & & & \end{pmatrix} \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & U' & \\ 0 & & & \end{pmatrix}$$

$$= \begin{pmatrix} \lambda_1 & * & \cdots & * \\ 0 & & & \\ \vdots & & (U')^H B U' & \\ 0 & & & \end{pmatrix},$$

so $U_2^H (U_1^H A U_1) U_2$ is upper triangular. We are therefore done, since $U = U_1 U_2$ is unitary due to the fact that $U(n)$ is a matrix group (see Proposition 6.40). If A and its eigenvalues are all real, then there exists a real eigenpair $(\lambda_1, \mathbf{u}_1)$; hence U_1 can be chosen to be orthogonal. The rest of the argument is the same with orthogonal matrices replacing unitary matrices. □

9.1.2 The normal matrix theorem

We now determine when the upper triangular matrix T in Proposition 9.1 is actually diagonal. In other words, we classify those $A \in \mathbb{C}^{n\times n}$ such that $A = UDU^H$ for some unitary matrix U, where $D = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$. To do so, we make the following definition.

Definition 9.1. A matrix $N \in \mathbb{C}^{n\times n}$ is said to be *normal* if

$$NN^H = N^H N. \tag{9.2}$$

After proving the normal matrix theorem, which is next, we will give a number of interesting examples of normal matrices. They fall into classes depending on what conditions their eigenvalues satisfy.

Theorem 9.2 (The normal matrix theorem)**.** *A matrix $A \in \mathbb{C}^{n\times n}$ is unitarily diagonalizable if and only if A is normal.*

Proof. The "only if" part is straightforward and left as an exercise. Suppose A is normal. Write $A = UTU^H$, where U is unitary and T is upper triangular. Since $A^HA = AA^H$ and U is unitary, it follows that $TT^H = T^HT$. Hence it suffices to show that an upper triangular normal matrix is diagonal. The key is to compare the diagonal entries of T^HT and TT^H. The point is that for every $n \times n$ complex matrix B, the kth diagonal entry of B^HB is the square of the length of the kth column of B, while the kth diagonal entry of BB^H is the square of the length of the kth row. But in an upper triangular matrix B, if the square of the length of the kth row equals the square of the length of the kth column for all k, then B is diagonal (the proof is left to the reader). Thus an upper triangular normal matrix is in fact diagonal. Therefore, A is indeed unitarily diagonalizable. □

Corollary 9.3. *If $A \in \mathbb{C}^{n\times n}$ is normal, then two eigenvectors of A with distinct eigenvalues are Hermitian orthogonal.*

Proof. We leave this as an exercise. □

9.1.3 The Principal axis theorem: the short proof

Recall that a matrix $A \in \mathbb{C}^{n\times n}$ is Hermitian if and only if $A^H = A$. A real Hermitian matrix is, of course, symmetric. It is clear that every Hermitian matrix is normal. Furthermore, using the normal matrix theorem, we may easily prove the following result.

Proposition 9.4. *The eigenvalues of a Hermitian matrix are real. In fact, the Hermitian matrices are exactly the normal matrices having real eigenvalues.*

Proof. Let A be Hermitian, and write $A = UDU^H$. Then $A = A^H$ if and only if $U^HAU = U^HA^HU$ if and only if $D = D^H$ if and only if D is real. □

Of course, the fact that Hermitian matrices have real eigenvalues was proved from first principles in Chap. 8. The normal matrix theorem immediately implies the following.

Theorem 9.5 (Principal axis theorem)**.** *A complex matrix A is Hermitian if and only if A can be unitarily diagonalized as $A = UDU^H$, where D is a real diagonal matrix. Similarly, a real matrix A is symmetric if and only if A can be orthogonally diagonalized $A = QDQ^T$, where D is also real.*

Proof. The Hermitian case is immediate, but if A is real symmetric, one needs to argue a little more to show that U can be taken to be orthogonal. By the Schur triangulation theorem, $A = QTQ^T$, where Q is orthogonal and T is real and upper triangular. But since A is symmetric, it follows that T is symmetric, and since T is also upper triangular, it has to be diagonal. Therefore, $A = QDQ^T$, as desired. □

This proof is surprisingly brief, but it does not lead to any of the insights of the first version of the principal axis theorem.

9.1.4 Other examples of normal matrices

To obtain other classes of normal matrices, one can impose other conditions on D in the expression $A = UDU^H$. Here is another example obtained in this way.

Example 9.1 (Skew-Hermitian matrices)**.** A matrix J is said to be *skew-Hermitian* if $J^H = -J$. It is easy to see that J is skew-Hermitian if and only if iJ is Hermitian. Since Hermitian matrices are normal, so are skew-Hermitian matrices. Also, the nonzero eigenvalues of a skew-Hermitian matrix are pure imaginary: they have the form $i\lambda$ for a nonzero $\lambda \in \mathbb{R}$ that is an eigenvalue of iJ. A real skew-Hermitian matrix S is skew-symmetric, i.e., $S^T = -S$. For example,

$$J = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \quad \text{and} \quad S = \begin{pmatrix} 0 & 1 & 2 \\ -1 & 0 & 2 \\ -2 & -2 & 0 \end{pmatrix}$$

are both skew-symmetric. The diagonal entries of a skew-symmetric matrix are zero, so the trace of a skew-symmetric is also zero. The determinant of a skew-symmetric matrix of odd order is also 0 (see Exercise 9.1.3 below), so a skew-symmetric matrix of odd order has 0 as an eigenvalue. The matrix J above shows that the determinant of a skew-symmetric matrix of even order can be nonzero. Note that the characteristic polynomial of S is $-\lambda^3 - 9\lambda$, so the eigenvalues of S are $0, \pm 3i$, confirming the observation that all nonzero eigenvalues of a skew-Hermitian matrix are purely imaginary. Moreover, the nonzero eigenvalues of S occur in conjugate pairs, since S is real. ⊠

The normal matrix theorem gives the following structure theorem, which describes the nonsingular skew-symmetric matrices.

Proposition 9.6. *Assume that n is even, say $n = 2m$, and $A \in \mathbb{R}^{n\times n}$ is an invertible skew-symmetric matrix. Then there exists an orthonormal basis $\mathbf{x}_1,\ \mathbf{y}_1, \mathbf{x}_2, \mathbf{y}_2, \ldots, \mathbf{x}_m,\ \mathbf{y}_m$ of $\mathbb{R}^n$ such that A sends each real two-plane $\mathbb{R}\mathbf{x}_k + \mathbb{R}\mathbf{y}_k$ onto itself, and the matrix of A on this two-plane has the form*

$$J_{\lambda_k} = \begin{pmatrix} 0 & \lambda_k \\ -\lambda_k & 0 \end{pmatrix},$$

where λ_k is a nonzero real number such that $i\lambda_k$ is an eigenvalue of A. (This basis is not necessarily unique, however.) Thus there exists an orthogonal matrix Q such that $A = Q\ diag(J_{\lambda_1}, \ldots, J_{\lambda_m})\ Q^T$. In particular, a nonsingular real skew-symmetric matrix is similar via the matrix group $O(n, \mathbb{R})$ to a matrix that is the direct sum of nonzero two-dimensional skew-symmetric blocks J_{λ_k}.

Proof. Since A is real, its eigenvalues occur in conjugate pairs. Moreover, since A is skew-symmetric, iA is Hermitian, so iA has only real eigenvalues. Thus the eigenvalues of A can be sorted into pairs $\pm\, i\lambda_k$, where λ_k is a nonzero real number, since $\det(A) \neq 0$ and k varies from 1 to m. Note, however, that we are not claiming that the λ_k are all distinct. We may choose a Hermitian orthonormal basis of $\mathbb{C}^n$ consisting of eigenvectors of A, and since eigenvectors for different eigenvalues are Hermitian orthogonal by Corollary 9.3, we may choose a Hermitian orthonormal basis of $\mathbb{C}^n$ consisting of pairs $\{\mathbf{u}_k, \overline{\mathbf{u}_k}\}$, where $A\mathbf{u}_k = i\lambda_k \mathbf{u}_k$. We now observe that this means that $\mathbf{x}_k = (\mathbf{u}_k + \overline{\mathbf{u}}_k)/\sqrt{2}$ and $\mathbf{y}_k = i(\overline{\mathbf{u}}_k - \mathbf{u}_k)/\sqrt{2}$ also are a basis over $\mathbb{C}$ of $\text{span}\{\mathbf{u}_k, \overline{\mathbf{u}_k}\}$. Moreover, $\mathbf{x}_k$ and $\mathbf{y}_k$ have the additional property that both lie in $\mathbb{R}^n$ and are Euclidean orthogonal: $(\mathbf{x}_k)^T \mathbf{y}_k = 0$. Furthermore, A leaves the real two-plane $P_k = \mathbb{R}\mathbf{x}_k + \mathbb{R}\mathbf{y}_k$ that they span invariant, and the matrix of A with respect to this basis of P_k is J_{λ_k}. Finally, note that the planes P_i and P_j are also orthogonal if $i \neq j$. Hence the matrix $Q = (\mathbf{x}_1\ \mathbf{y}_1\ \cdots\ \mathbf{x}_m\ \mathbf{y}_m)$ is orthogonal and $A = Q\ \text{diag}(J_{\lambda_1}, \ldots, J_{\lambda_m})\ Q^T$. □

One can extend the above result to odd skew-symmetric matrices. We will leave this to the reader.

Another natural condition one can put on the eigenvalues of a normal matrix is that they all have modulus one. This is investigated in the following example.

Example 9.2. Let $K = UDU^H$, where every diagonal entry of D is a unit complex number. Then D is unitary, hence so is K. Conversely, every unitary matrix is normal (see Exercise 9.1.10). Thus the unitary matrices are exactly the normal matrices such that every eigenvalue has modulus one. For example, the skew-symmetric matrix

$$J = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

is clearly orthogonal, hence unitary. The matrix J has eigenvalues $\pm i$, and we can easily compute that $E_i = \mathbb{C}(1, -i)^T$ and $E_{-i} = \mathbb{C}(1, i)^T$. Thus

$$J = U_1 D U_1^H = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ i & -i \end{pmatrix} \begin{pmatrix} -i & 0 \\ 0 & i \end{pmatrix} \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -i \\ 1 & i \end{pmatrix}.$$

The basis constructed in the above proposition is $\mathbf{u}_1 = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\ i\end{pmatrix}$ and $\mathbf{u}_1' = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\ -i\end{pmatrix}$. The way U acts as a complex linear mapping of $\mathbb{C}^2$ can be interpreted geometrically as follows: U rotates vectors on the complex line $\mathbb{C}(1,i)^T$ spanned by $(1,i)^T$ (thought of as a real two-plane) through $\frac{\pi}{2}$ and rotates vectors on the orthogonal axis by $-\frac{\pi}{2}$. Of course, as a mapping on $\mathbb{R}^2$, U is simply the rotation $R_{\pi/2}$. ⊠

Exercises

Exercise 9.1.1. Show that $A \in \mathbb{C}^{n\times n}$ is unitarily diagonalizable if and only if A^H is.

Exercise 9.1.2. True or false (discuss your reasoning):

(i) A complex symmetric matrix is normal.
(ii) The real part of a Hermitian matrix is symmetric and the imaginary part is skew-symmetric.
(iii) The real part of a normal matrix is normal.
(iv) If a normal matrix N has real eigenvalues, then N is Hermitian.

Exercise 9.1.3. Unitarily diagonalize the skew-symmetric matrices

$$J = \begin{pmatrix}0 & 1\\ -1 & 0\end{pmatrix} \quad \text{and} \quad S = \begin{pmatrix}0 & 1 & 2\\ -1 & 0 & 2\\ -2 & -2 & 0\end{pmatrix}.$$

Exercise 9.1.4. Let

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

Express A as $Q\mathrm{diag}(J_\lambda, J_\mu)Q^T$ as in Proposition 9.6.

Exercise 9.1.5. Formulate a result similar to Proposition 9.6 for skew-symmetric matrices $A \in \mathbb{R}^{n\times n}$, where n is odd, that have the property that $\dim \mathcal{N}(A) = 1$.

Exercise 9.1.6. Let S be a skew-Hermitian $n \times n$ matrix. Show the following:

(i) If n is odd, then $\det(S)$ is pure imaginary (but possibly zero), and if n is even, then $\det(S)$ is real.

(ii) If S is skew-symmetric, then $\det(S) = 0$ if n is odd, and $\det(S) \geq 0$ if n is even.

Exercise 9.1.7. Show from the definition of the determinant that the determinant of a Hermitian matrix is real.

Exercise 9.1.8. Prove Corollary 9.3. That is, show that two eigenvectors for different eigenvalues of a normal matrix are Hermitian orthogonal.

Exercise 9.1.9. If possible, find an example of a 3×3 real matrix N such that N is normal, but N is neither symmetric nor skew-symmetric.

Exercise 9.1.10. Let $U \in U(n)$. Show the following.

(i) Every eigenvalue of U also has modulus 1.

(ii) $\det(U)$ has modulus 1.

(iii) $|\mathrm{Tr}(U)| \leq n$.

Exercise 9.1.11. Suppose Q is an $n \times n$ orthogonal matrix with no real eigenvalues. True or false: n is even and $\det(Q) = 1$.

Exercise 9.1.12. Suppose all eigenvalues of a unitary matrix Q are 1. True or false: $Q = I_n$.

Exercise 9.1.13. Let $\mathcal{N}(n)$ denote the set of normal $n \times n$ complex matrices. Prove that $\mathcal{N}(n)$ is not a subspace of $\mathbb{C}^{n \times n}$.

Exercise 9.1.14. Suppose A and B are commuting normal matrices. Prove the following:

(i) $A + B$ and AB are also normal, and

(ii) A and B are simultaneously diagonalizable.

Exercise 9.1.15. Suppose, as in Exercise 9.1.14, that A and B are commuting normal matrices. What are the possible eigenvalues of $A + B$?

Exercise 9.1.16. Formulate the notion of a normal operator on a Hermitian inner product space from the definition of a normal matrix.

Exercise 9.1.17. True or false (discuss your reasoning):

(i) If A is normal and U is unitary, then UAU^H is normal.

(ii) If A is normal and invertible, then A^{-1} is normal.

(iii) If $A \in \mathbb{R}^{n \times n}$, then AA^T and A^TA have the same eigenvalues.

(iv) If A is normal and k is a positive integer, then A^k is normal.

9.2 Quadratic Forms

A quadratic form over a field $\mathbb{F}$ is a function $q : \mathbb{F}^n \to \mathbb{F}$ of the form $q(r_1, \ldots, r_n) = \sum_{i,j} a_{ij} r_i r_j$, where all the a_{ij} are in $\mathbb{F}$. Quadratic forms are used in many areas. For example, Lagrange's four squares theorem says that every positive integer n is the sum of the squares of four integers. That is, there exist $a, b, c, d \in \mathbb{Z}$ such that $n = a^2 + b^2 + c^2 + d^2$. Every real quadratic form can be expressed via a real symmetric matrix; hence every real quadratic form can be written as a sum of squares. In this section, we will develop some of the basic properties of quadratic forms over $\mathbb{F}$ such as diagonalization. We will also consider an interesting equivalence relation on the symmetric matrices over $\mathbb{F}$ called congruence, and we will prove Sylvester's law of inertia, which classifies the equivalence classes of congruent symmetric matrices over the reals $\mathbb{R}$. A consequence of this is the classification of positive definite matrices. We will also treat the corresponding equivalence relation for Hermitian matrices and Hermitian quadratic forms.

9.2.1 Quadratic forms and congruence

Throughout this section, we will assume that the field $\mathbb{F}$ has characteristic different from two. A *quadratic form* on $\mathbb{F}^n$ is a function $q : \mathbb{F}^n \to \mathbb{F}$ such that

$$q(r_1, \ldots, r_n) = \sum_{i,j=1}^{n} a_{ij} r_i r_j,$$

where all a_{ij} are in $\mathbb{F}$. Since the characteristic of $\mathbb{F}$ is not two, one can suppose that the coefficients a_{ij} of q are symmetric by replacing a_{ij} by $b_{ij} = \frac{1}{2}(a_{ij} + a_{ji})$. This leaves q unchanged, but by making its coefficients symmetric, we can associate a unique symmetric matrix to q in order to bring in our previous results on symmetric matrices. Indeed, q is expressed in terms of B by the identity

$$q(r_1, \ldots, r_n) = \mathbf{r}^T B \mathbf{r} \quad \text{where} \quad \mathbf{r} = \begin{pmatrix} r_1 \\ \vdots \\ r_n \end{pmatrix}.$$

Conversely, a symmetric matrix $A \in \mathbb{F}^{n \times n}$ defines the quadratic form

$$q_A(r_1, \ldots, r_n) = \mathbf{r}^T A \mathbf{r}.$$

Under a change of coordinates on $\mathbb{F}^n$ of the form $\mathbf{r} = C\mathbf{s}$, where $C \in \mathbb{F}^{n \times n}$ is invertible, the quadratic form q_A is transformed into a new quadratic form

as follows:

$$q_A(r_1, \ldots, r_n) = q_A((C\mathbf{s})^T) = \mathbf{s}^T C^T AC\mathbf{s}.$$

Thus, the quadratic form $q_B(s_1, s_2, \ldots, s_n)$ associated to $B = C^T AC$ satisfies

$$q_B(s_1, s_2, \ldots, s_n) = q_A(r_1, \ldots, r_n).$$

In other words, $q_B(\mathbf{s}^T)$ and $q_A(\mathbf{r}^T)$ are the same quadratic form expressed in different coordinate systems. This motivates the next definition.

Definition 9.2. Two symmetric $n \times n$ matrices A and B over $\mathbb{F}$ are said to be *congruent* if there exists a nonsingular $C \in \mathbb{F}^{n\times n}$ such that $A = C^T BC$. Quadratic forms associated to congruent matrices are said to be *equivalent*.

Proposition 9.7. *Congruence is an equivalence relation on the symmetric matrices in $\mathbb{F}^{n\times n}$. Moreover, equivalence of quadratic forms is also an equivalence relation on quadratic forms.*

Proof. It suffices to show that equivalence is an equivalence relation. First, $A = (I_n)^T AI_n$, which shows that A is equivalent to itself. If $A = C^T BC^T$, where C is nonsingular, then $B = (C^{-1})^T AC^{-1}$. Thus equivalence is symmetric. We leave the proof that equivalence is transitive as an exercise. □

9.2.2 Diagonalization of quadratic forms

The basic fact about quadratic forms is that every quadratic form q over a field $\mathbb{F}$ of characteristic different from two is equivalent to a quadratic form q' that is a sum of squares. That is,

$$q'(s_1, \ldots, s_n) = \sum_{i=1}^{n} a_i s_i^2,$$

where the a_i lie in $\mathbb{F}$. Equivalently, every symmetric matrix is congruent to a diagonal matrix. We will omit the proof, but note that we have already treated a special case. For example, by Proposition 4.19, a symmetric $n \times n$ matrix A over $\mathbb{F}$ such that each $k \times k$ submatrix A_k in the upper left-hand corner of A is invertible admits an LDL^T decomposition, where L is lower triangular unipotent. In that case, the kth diagonal entry of D is the kth pivot of A, namely, $d_k = \det A_k / \det A_{k-1}$, where by definition, $\det A_0 = 1$.

Example 9.3. Suppose $\mathbb{F} = \mathbb{Q}$, and let $A \in \mathbb{F}^{2\times 2}$ be symmetric. Put

$$A = \begin{pmatrix} a & b \\ b & c \end{pmatrix},$$

where $a \neq 0$. Then we have

$$\begin{pmatrix} 1 & 0 \\ -b/a & 1 \end{pmatrix} \begin{pmatrix} a & b \\ b & c \end{pmatrix} \begin{pmatrix} 1 & -b/a \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} a & 0 \\ 0 & c - b^2/a \end{pmatrix},$$

so A is congruent to a diagonal matrix (even if A is singular). ⊠

9.2.3 Diagonalization in the real case

The principal axis theorem says that a real quadratic form $q_A(r_1, \ldots, r_n) = \sum_{i,j=1}^{n} a_{ij} r_i r_j$, where the matrix $A = (a_{ij})$ is symmetric, can be diagonalized using orthogonal axes. To be specific, there exists an orthogonal matrix Q such that $A = QDQ^T$. The columns of Q are the principal axes, and setting $\mathbf{s} = Q^T\mathbf{r}$, the components of $\mathbf{s}$ are the coordinates with respect to the principal axes. Thus,

$$q_A(r_1, \ldots, r_n) = q_D(s_1, \ldots, s_n) = \sum \lambda_i s_i^2.$$

Proposition 9.8. *Every real quadratic form $q(r_1, \ldots, r_n)$ can be orthogonally diagonalized as*

$$q(r_1, \ldots, r_n) = \sum_{i=1}^{n} \lambda_i s_i^2,$$

where $\mathbf{r} = Q\mathbf{s}$ and Q is the orthogonal matrix of principal axes.

Example 9.4. Consider the quadratic form $q(x, y) = x^2 + 4xy + y^2$. Its associated symmetric matrix is

$$A = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}.$$

The eigenvalues of A are 3 and -1. (Reason: both rows sum to 3, and the trace of A is 2.) Noting that $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ and $\begin{pmatrix} 1 \\ -1 \end{pmatrix}$ are corresponding eigenvectors, we get $A = QDQ^T$, where $Q = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$ and $D = \operatorname{diag}(3, -1)$. Putting $\begin{pmatrix} u \\ v \end{pmatrix} = Q^T \begin{pmatrix} x \\ y \end{pmatrix}$ gives new coordinates (u, v) such that $q(x, y)$ is the difference of squares $q(x, y) = 3u^2 - v^2$. Thus the equation $q(x, y) = 1$ represents a hyperbola with axes $\frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ and $\frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix}$. ⊠

Remark. If q is a quadratic form on $\mathbb{R}^n$ and $c \in \mathbb{R}$ is a constant, the level set $V_c = \{(r_1, \ldots, r_n) \mid q(r_1, \ldots, r_n) = c\}$ is called a *quadratic variety*. It is an example of a real algebraic variety. When $n = 2$, a quadratic variety is called a conic section. If q has matrix A and both eigenvalues of A are positive, then V_c is an ellipse or a circle when $c > 0$. (It is a circle if A has equal eigenvalues.) If both eigenvalues are positive and $c < 0$, then V_c is actually a subset of $\mathbb{C}^2$ that does not meet $\mathbb{R}^2$. If A's eigenvalues have different signs, then V_c is a hyperbola in $\mathbb{R}^2$. In $\mathbb{R}^3$, the quadratic varieties are surfaces whose type is classified by the number of positive and negative eigenvalues of the associated symmetric matrix.

9.2.4 Hermitian forms

When $A \in \mathbb{C}^{n\times n}$ is Hermitian, a quadratic function $u_A : \mathbb{C}^n \to \mathbb{C}$ of the form

$$u_A(z_1, \ldots, z_n) = \sum_{i,j=1}^{n} a_{ij}\overline{z}_i z_j = \mathbf{z}^H A\mathbf{z}, \quad \text{where } \mathbf{z} = \begin{pmatrix} z_1 \\ \vdots \\ z_n \end{pmatrix}$$

is called a *Hermitian form*. Every Hermitian form can be unitarily diagonalized. All Hermitian forms are real-valued, since

$$(\mathbf{z}^H A\mathbf{z})^H = \mathbf{z}^H A^H \mathbf{z}^{HH} = \mathbf{z}^H A\mathbf{z}.$$

Similarly, a skew-Hermitian form takes only pure imaginary values. By an argument similar to that in Proposition 9.9, we get the following.

Proposition 9.9. *Let $A \in \mathbb{C}^{n\times n}$ be normal, so $A = UDU^H$, where U is unitary and $D = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$. Let $w_1, \ldots, w_n$ be the coordinates on $\mathbb{C}^n$ coming from a Hermitian orthonormal eigenbasis of $\mathbb{C}^n$ consisting of the columns of U. Then*

$$u_A(z_1, \ldots, z_n) = \sum_{i=1}^{n} \lambda_i \overline{w}_i w_i = \sum_{i=1}^{n} \lambda_i |w_i|^2.$$

9.2.5 Positive definite matrices

Given a real quadratic form q (or more generally, a Hermitian quadratic form h), when does q (or h) have a strict maximum or minimum at the origin? This corresponds to all the eigenvalues of its matrix being either positive or

negative. A minimum is attained if all are positive, and a maximum occurs if they are all negative. Thus one makes the following definition.

Definition 9.3. A real symmetric $n \times n$ matrix A is called *positive definite* if

$$\mathbf{x}^T A\mathbf{x} > 0 \quad \text{for all nonzero } \mathbf{x} \in \mathbb{R}^n.$$

An $n \times n$ Hermitian matrix A is *positive definite* if and only if

$$\mathbf{z}^H A\mathbf{z} > 0 \quad \text{for all nonzero } \mathbf{z} \in \mathbb{C}^n.$$

The following result describes the positive definite matrices in several equivalent ways.

Proposition 9.10. *For a real symmetric (respectively complex Hermitian) matrix A, the following conditions are equivalent:*

(i) all eigenvalues of A are positive;

(ii) A is positive definite;

(iii) the upper left $k \times k$ submatrix A_k has $\det(A_k) > 0$ for all k; and

(iv) A has an LDL^T decomposition in which D has positive diagonal entries.

Proof. For simplicity, we will give the proof for only the real symmetric case. The changes necessary to prove the Hermitian case are routine after replacing $\mathbf{x}^T A\mathbf{x}$ by $\mathbf{z}^H A\mathbf{z}$. We will show that each statement implies the following one and (iv) implies (i). Assume that $A \in \mathbb{R}^{n\times n}$ is symmetric, and (i) holds. Applying the principal axis theorem, we have $A = Q^T DQ$, where $Q = (\mathbf{q}_1 \cdots \mathbf{q}_n)$ is orthogonal and $D = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$ with all $\lambda_i > 0$. Thus, if $\mathbf{x} = \sum_{i=1}^n x_i \mathbf{q}_i$, then $\mathbf{x}^T A\mathbf{x} = \sum_{i=1}^n \lambda_i x_i^2$, which is positive unless $\mathbf{x} = \mathbf{0}$. Thus (i) implies (ii). Now suppose A is positive definite, and recall that if $k \le n$, then $\mathbb{R}^k \subset \mathbb{R}^n$ as $\mathbb{R}^k = \{\mathbf{x} \in \mathbb{R}^n \mid x_i = 0 \quad \text{for all} \quad i > k\}$. Being symmetric, it follows that A_k is positive definite on $\mathbb{R}^k$ for all k. Consequently, all eigenvalues of A_k are positive, so $\det(A_k) > 0$ for all k. We conclude that (ii) implies (iii). Moreover, by Proposition 4.17, (iii) immediately implies (iv). The last implication, (iv) implies (i), follows from the law of inertia, which will be proved below. □

We say that a symmetric matrix $A \in \mathbb{R}^{n\times n}$ is *negative definite* if $-A$ is positive definite, with a similar definition for the Hermitian case. Thus a negative definite matrix has only negative eigenvalues and has an LDL^T decomposition with negative pivots. Note that this means that $(-1)^k \det(A_k) > 0$. Finally, a symmetric matrix that has both positive and negative eigenvalues is called *indefinite*.

Example 9.5. Consider the matrix

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

By row operations using lower triangular elementary matrices of the third kind, we get

$$L^*A = \begin{pmatrix} 1 & 1 & 0 & 1 \\ 0 & 1 & -1 & -1 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & -1 \end{pmatrix}.$$

Hence A has an LDU decomposition, but three pivots are positive and one is negative. Therefore, A cannot be positive definite or negative definite.

9.2.6 The positive semidefinite case

A quadratic form q on $\mathbb{R}^n$ is said to have a *relative minimum* at $\mathbf{0}$ if $q(\mathbf{x}) \geq 0$ for all $\mathbf{x} \in \mathbb{R}^n$ and $q(\mathbf{x}) = 0$ for some nonzero $\mathbf{x}$. The meaning of a *relative maximum* is similar, and the same definitions apply to Hermitian forms on $\mathbb{C}^n$. Here, k satisfies $0 < k \leq n$. A real symmetric (respectively complex Hermitian) $n \times n$ matrix A is said to be *positive semidefinite* if $\mathbf{x}^T A\mathbf{x} \geq 0$ (respectively $\mathbf{z}^H A\mathbf{z} \geq 0$) for all $\mathbf{x} \in \mathbb{R}^n$ (respectively $\mathbf{z} \in \mathbb{C}^n$). A positive semidefinite matrix can have a nontrivial null space, but it can't have any negative eigenvalues, since if $(\lambda, \mathbf{v})$ is an eigenpair for A with $\lambda < 0$, then $\mathbf{v}^T A\mathbf{v} = \lambda \mathbf{v}^T\mathbf{v} = \lambda|\mathbf{v}|^2 < 0$.

Proposition 9.11. *Let A be positive semidefinite. Then all eigenvalues of A are nonnegative.*

This is immediate from the previous observation. Here is the main result about positive semidefinite matrices.

Proposition 9.12. *A real symmetric (respectively complex Hermitian) matrix A is positive semidefinite if and only if it admits an LDL^T (respectively LDL^H) decomposition in which D has only nonnegative entries.*

Proof. Suppose that A is real symmetric. (The proof in the Hermitian case is, as usual, similar.) Since it may happen that $\det(A_k) = 0$ for some k, we have to give a direct proof that A has an LDL^T decomposition. But we know that A has an $LPDU$ decomposition of the form $A = LPDL^T$, where L is lower triangular unipotent, P is a (unique) partial permutation matrix, D is diagonal, and PD is symmetric. It suffices to show that if PD is positive

semidefinite, then PD is a diagonal matrix with nonnegative entries. Suppose A is of size $n \times n$ and $P \neq I_n$ nor is P obtained by setting some rows of I_n equal to $\mathbf{0}$. To simplify the notation, suppose the first row of P is nonzero. Let's assume that the ith column is $d\mathbf{e}_1$, where $i > 1$ and $d \neq 0$. Since PD is symmetric, its first column is $d\mathbf{e}_i$. Thus PD interchanges $\mathbf{e}_1$ and $\mathbf{e}_i$. Putting $\mathbf{y} = y_1\mathbf{e}_1 + y_i\mathbf{e}_i$, we get

$$\mathbf{y}^T PD\mathbf{y} = (y_1\mathbf{e}_1 + y_i\mathbf{e}_i)^T(dy_1\mathbf{e}_i + dy_i\mathbf{e}_1) = 2dy_1y_i.$$

Hence, if PD is positive semidefinite, then $dy_1y_i \geq 0$ for all $y_1, y_i \in \mathbb{R}$. Since $i > 1$, this is clearly impossible, so if the first row of PD is nonzero, it has d in the $(1,1)$ entry. The argument is the same for all other rows, so PD is diagonal. Moreover, the entries of D must be nonnegative, so every positive semidefinite real symmetric matrix can be written LDL^T. The converse is left as an exercise. □

Example 9.6. Suppose $A \in \mathbb{R}^{m\times n}$. Then A^TA is symmetric, so we can ask whether A^TA is positive or positive semidefinite. In fact, for every $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{x}^TA^TA\mathbf{x} = (A\mathbf{x})^TA\mathbf{x} = |A\mathbf{x}|^2$, so $\mathbf{x}^TA^TA\mathbf{x} \geq 0$. Hence A^TA is positive semidefinite. In particular, all eigenvalues of A^TA are nonnegative. If $\mathcal{N}(A) = \{\mathbf{0}\}$, then $|A\mathbf{x}| > 0$, provided $\mathbf{x} \neq 0$, so A^TA is positive definite. The same is true for AA^T. ⊠

Exercises

Exercise 9.2.1. Let $A \in \mathbb{R}^{2\times 2}$ be symmetric.

(i) Show that if $\det(A) > 0$, then A is either positive definite or negative definite.

(ii) Also show that if $\mathrm{Tr}(A) = 0$, then A is indefinite.

Exercise 9.2.2. Suppose A is a symmetric matrix such that $\det(A) \neq 0$ and A has both positive and negative diagonal entries. Explain why A has to be indefinite.

Exercise 9.2.3. Show that if A is a positive definite 3×3 matrix, then the coefficients of its characteristic polynomial alternate in sign. Also show that if A is negative definite, the coefficients are all negative.

Exercise 9.2.4. Give an example of a 3×3 symmetric matrix A such that the coefficients of the characteristic polynomial of A are all negative, but A is not negative definite. (Could your answer be a diagonal matrix?)

Exercise 9.2.5. Let $A \in \mathbb{R}^{n\times n}$ be positive definite and suppose $S \in \mathbb{R}^{n\times n}$ is nonsingular.

(i) When is SAS^{-1} positive definite?

(ii) Is A^m positive definite for all integers m? (Note: $A^0 = I_n$.)

Exercise 9.2.6. Describe the quadratic surface $(x\ y\ z)A(x\ y\ z)^T = 1$ for the following choices of A:

$$\begin{pmatrix} 1 & 2 & -1 \\ 2 & 0 & 3 \\ 3 & -1 & 2 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 2 & 4 & 2 \\ 2 & 2 & 1 \\ 2 & 1 & 5 \end{pmatrix}.$$

Exercise 9.2.7. Decide whether $g(x, y, z) = x^2 + 6xy + 2xz + 3y^2 - xz + z^2$ has a maximum, minimum, or neither at $(0, 0, 0)$.

Exercise 9.2.8. Let $A \in \mathbb{R}^{n\times n}$, and suppose $A_i = 0$ for some $i < n$. Does this mean that A has a zero eigenvalue?

Exercise 9.2.9. Let $A \in \mathbb{R}^{m\times n}$ have the property that $\mathcal{N}(A) \neq \{\mathbf{0}\}$. Is $A^T A$ positive semidefinite? Can it be positive definite?

Exercise 9.2.10. For the following pairs A, B of symmetric matrices, determine whether A and B are congruent.

(i) A and B have the same characteristic polynomial.

(ii) $\det(A) < 0$, $\det(B) > 0$.

(iii) $A = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$ and $B = \begin{pmatrix} 1 & 2 \\ 2 & -1 \end{pmatrix}$.

(iv) A and B are similar.

(v) A and B are positive definite.

(vi) $AB = BA$.

Exercise 9.2.11. Show that if a symmetric (respectively Hermitian) matrix A is semipositive definite, then A has a symmetric (respectively Hermitian) kth root M for all $k > 0$. (That is, $A^k = M$.) Moreover, if A is positive definite, so is M.

Exercise 9.2.12. Show that the product AB of a positive definite matrix A and a symmetric matrix B has real eigenvalues, even though AB is not necessarily symmetric. (Hint: show that AB is similar to a symmetric matrix. Exercise 9.2.11 may help.)

9.3 Sylvester's Law of Inertia and Polar Decomposition

In the final section of this chapter, we will prove two interesting results about congruence classes of Hermitian matrices. The first is a famous result of Sylvester that classifies the congruence class of a Hermitian matrix in terms of its signature. The signature of a Hermitian matrix is a triple that tabulates the number of eigenvalues that are positive, negative, or zero. The law of inertia says that two Hermitian matrices are congruent if and only if their signatures coincide. This result tells us, for example, that the signs of the pivots of a real symmetric matrix determine its signature. The second result says that every element of $GL(n, \mathbb{C})$ has a unique factorization as KU, where K is positive definite Hermitian and U is unitary. This gives the structure of the congruence class with signature $(n, 0, 0)$. Namely, the class consisting of positive definite Hermitian matrices is in one-to-one correspondence with the coset space $GL(n, \mathbb{C})/U(n)$. Similarly, the congruence class of positive definite real symmetric matrices is in one-to-one correspondence with the coset space $GL(n, \mathbb{R})/O(n, \mathbb{R})$.

9.3.1 The law of inertia

Let $A \in \mathbb{C}^{n\times n}$ be Hermitian, and let $n_+(A)$, $n_-(A)$, and $n_0(A)$ denote, respectively, the number of positive, negative, and zero eigenvalues of A. For example, $n_0(A) = \dim \mathcal{N}(A)$. We will call the triple $(n_+(A), n_-(A), n_0(A))$ the *signature* of A. Sylvester's law of inertia, to be proved next, gives an elegant answer to the question of determining the signature of A. Since by Corollary 4.22, congruent matrices have the same rank, it follows from the rank–nullity theorem that if A and B are congruent, then $n_0(A) = n_0(B)$. For simplicity, we will state and prove the law of inertia in the real symmetric case only. The Hermitian version may be formulated without any surprises and proved in essentially the same way.

Theorem 9.13 (Sylvester's law of inertia)**.** *Let A and B be congruent real symmetric matrices. Then A and B have the same signature. That is,*

$$n_+(A) = n_+(B), \quad n_0(A) = n_0(B) \quad \text{and} \quad n_-(A) = n_-(B).$$

Conversely, two real symmetric matrices having the same signature are congruent. In particular, if a real symmetric matrix A has a symmetric LDU decomposition $A = LDL^T$, then the signs of its eigenvalues are the same as the signs on the diagonal of D (which are the pivots of A).

Proof. First choose orthogonal matrices P and Q such that $A = PDP^T$ and $B = QEQ^T$, where D and E are diagonal. Notice that without affecting these expressions, we may assume that the positive diagonal entries of D are $d_1, \ldots, d_s$ and that those of E are $e_1, \ldots, e_t$. This follows from the fact that for every permutation matrix P, $PDP^T = PDP^{-1}$ and D have the same diagonal up to the permutation corresponding to P. Notice that A and D, as well as B and E, have the same eigenvalues, hence the same signature. Thus, to show that if A and B are congruent, then they have the same signature, it will suffice to show that D and E have the same signature. Write $B = CAC^T$, where $C \in \mathbb{R}^{n\times n}$ is invertible. Thus

$$B = CAC^T = CPDP^TC^T = QEQ^T.$$

Hence $E = MDM^T$, where $M = Q^TCP$. To show that D and E have the same signature, it suffices to show that $n_+(D) = n_+(E)$, since D and E, being congruent, have $n_0(D) = n_0(E)$, as remarked above. Let us assume $n_+(D) = s < n_+(E) = t$, and let $f_i(\mathbf{x}) \in (\mathbb{R}^n)^*$ denote the ith component function of $M^T\mathbf{x}$. Since $\mathbf{x}^TE\mathbf{x} = \mathbf{x}^TMDM^T\mathbf{x}$, when $\mathbf{x}^T = (x_1, \ldots, x_t, 0, \ldots, 0)$, we have

$$\sum_{i=1}^{t} e_i x_i^2 = \sum_{j=1}^{n} d_j f_j(x_1, \ldots, x_t, 0, \ldots, 0)^2. \tag{9.3}$$

Since $t > s$, there exist $a_1, \ldots, a_t \in \mathbb{R}$ not all zero such that

$$f_j(a_1, \ldots, a_t, 0, \ldots, 0) = 0 \quad \text{for} \quad j = 1, \ldots, s.$$

Indeed, fewer than t homogeneous equations in t variables have a nontrivial solution. Thus,

$$0 < \sum_{i=1}^{t} e_i a_i^2 = \sum_{j=s+1}^{n} d_j f_j(a_1, \ldots, a_t, 0, \ldots, 0)^2 \leq 0,$$

since $d_j \leq 0$ if $j > s$. This is a contradiction, so we must have $n_+(D) = n_+(E)$. This shows that D and E have the same signature, and hence so do A and B. We will leave the proof of the converse as an exercise. □

Example 9.7. Let

$$A = \begin{pmatrix} 1 & 1 & 2 \\ 1 & 1 & 3 \\ 2 & 3 & 2 \end{pmatrix}.$$

Then the quadratic form associated to A is

$$Q(x, y, z) = x^2 + 2xy + 4xz + 2y^2 + 6yz + 2z^2.$$

A routine calculation gives

$$LA = DU = \begin{pmatrix} 1 & 1 & 2 \\ 0 & 1 & 1 \\ 0 & 0 & -3 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -3 \end{pmatrix} \begin{pmatrix} 1 & 1 & 2 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}.$$

Since A is nonsingular and symmetric, the expression $LA = DU$ implies that $A = U^T DU$. Hence A is congruent to $\mathrm{diag}(1, 1, -3)$, so its signature is $(2, 1, 0)$. The quadratic surface

$$x^2 + 2xy + 4xz + 2y^2 + 6yz + 2z^2 = 1$$

is a hyperboloid of one sheet. ⊠

Remark. The law of inertia also holds in the Hermitian case. In fact, the proof is essentially the same, so we will leave the details to the reader.

9.3.2 The polar decomposition of a complex linear mapping

Recall that every nonzero complex number has a unique polar representation $z = |z|e^{i\theta}$ with $0 \leq \theta < 2\pi$. The purpose of this section is to generalize this fact to linear mappings. Recall that $e^{i\theta}$ is an element of the group of unit complex numbers. By definition, this group is $U(1) = \{z \in \mathbb{C} \mid zz^H = 1\}$. The polar representation says that group-theoretically, $\mathbb{C}^* = GL(1, \mathbb{C}) = (\mathbb{R}_{>0})U(1)$. This representation generalizes to the $n \times n$ case when we replace $U(1)$ by the $n \times n$ unitary group and $\mathbb{R}_{>0}$ by the set of positive definite $n \times n$ Hermitian matrices (which is not a group).

Proposition 9.14. *Every $A \in GL(n, \mathbb{C})$ can be uniquely expressed in either of two ways as $A = KU = UK'$, where $U \in U(n)$, K is the unique Hermitian positive definite matrix such that $K^2 = AA^H$, and $K' = U^H KU$.*

Proof. Let $H = AA^H$. Then H is Hermitian. Since A is nonsingular, H is positive definite. (Proof: let $(\lambda, \mathbf{x})$ be an eigenpair for H. Since A is nonsingular, $A^H\mathbf{x} \neq \mathbf{0}$. Thus,

$$0 < |A^H\mathbf{x}|^2 = (A^H\mathbf{x})^H A^H\mathbf{x} = \mathbf{x}^H(AA^H)\mathbf{x} = \mathbf{x}^H(\lambda\mathbf{x}) = \lambda|\mathbf{x}|^2,$$

so $\lambda > 0$.) By the result of Exercise 9.2.11, we can write $H = K^2$, where K is also positive definite Hermitian. Now put $U = K^{-1}A$. Then $A = KU$, so it suffices to show that U is unitary. But

$$UU^H = K^{-1}A(A^H(K^{-1})^H) = K^{-1}H(K^{-1})^H = K^{-1}K^2K^{-1} = I_n,$$

so U is indeed unitary. The $U'K'$ factorization follows because

$$KU = (UU^H)KU = U(U^HKU) = UK',$$

and $K' = U^HKU$ is also positive definite Hermitian by the law of inertia. Finally, to prove uniqueness, notice that if K is Hermitian positive definite and $A = KU$, then necessarily $K^2 = AA^H$, since $K = AU^H$, so $K^H = UA^H$, and hence $K^2 = KK^H = (AU^H)(UA^H) = AA^H$. But the square root of a positive definite matrix is unique (verify this), so the polar representation must be unique. □

The real version of the polar representation takes an expected similar form. The proof is identical to the complex case.

Proposition 9.15. *Every $A \in GL(n,\mathbb{R})$ can be uniquely expressed in either of two ways as $A = SQ = QS'$, where $Q \in O(n,\mathbb{R})$, S is the unique positive definite real matrix such that $S^2 = AA^T$, and $S' = Q^TSQ$.*

Polar decomposition gives rise to a nice interpretation of the coset spaces $GL(n,\mathbb{C})/U(n)$ and $GL(n,\mathbb{R})/O(n,\mathbb{R})$. Let $PD(n,\mathbb{C})$ (respectively $PD(n,\mathbb{R})$) denote the $n \times n$ positive definite Hermitian (respectively positive definite real symmetric) matrices. Neither $PD(n,\mathbb{C})$ nor $PD(n,\mathbb{R})$ is a group, because neither is closed under multiplication. However, $PD(n,\mathbb{C})$ and $PD(n,\mathbb{R})$ are both closed under inverses. The upshot of polar decomposition is that the quotient maps $\pi : GL(n,\mathbb{C}) \to GL(n,\mathbb{C})/U(n)$ and $\pi : GL(n,\mathbb{R}) \to GL(n,\mathbb{R})/O(n,\mathbb{R})$ have the property that $\pi(PD(n,\mathbb{C})) = GL(n,\mathbb{C})/U(n)$ and $\pi(PD(n,\mathbb{R})) = GL(n,\mathbb{R})/O(n,\mathbb{R})$. This gives a surjective mapping $\overline{\pi} : PD(n,\mathbb{C}) \to GL(n,\mathbb{C})/U(n)$ with a corresponding map in the real case defined in exactly the same way. I claim that $\overline{\pi}$ is injective in each case. Here is the proof in the real case. (The complex case is essentially the same.) Suppose $R, T \in PD(n,\mathbb{R})$ and $\overline{\pi}(R) = \overline{\pi}(T)$. That is, $RO(n,\mathbb{R}) = TO(n,\mathbb{R})$. By the criterion for equality of left cosets, it follows that $R^{-1}T \in O(n,\mathbb{R})$. Since R^{-1} is also positive definite, Exercise 9.2.12 implies that $R^{-1}T$ has real eigenvalues. But if $R^{-1}T \in O(n,\mathbb{R})$, this implies that the eigenvalues of $R^{-1}T$ are either 1 or -1. But -1 is impossible, for if -1 is an eigenvalue, there exists a nonzero $\mathbf{x} \in \mathbb{R}^n$ such that $R^{-1}T(\mathbf{x}) = -\mathbf{x}$. Thus $T(\mathbf{x}) = R(-\mathbf{x}) = -R(\mathbf{x})$, so

$$\mathbf{x}^TT(\mathbf{x}) = -\mathbf{x}^TR(\mathbf{x}).$$

Since R and T are both positive definite, this is impossible. Hence 1 is the only eigenvalue of $R^{-1}T$. But since $R^{-1}T \in O(n, \mathbb{R})$, $R^{-1}T$ is also normal, hence is unitarily diagonalizable by the normal matrix theorem (Theorem 9.2). Hence $R^{-1}T = I_n$, so $R = T$. Therefore, we have proved the following result.

Proposition 9.16. *The mappings* $\overline{\pi} : PD(n, \mathbb{C}) \to GL(n, \mathbb{C})/U(n)$ *and* $\overline{\pi} : PD(n, \mathbb{R}) \to GL(n, \mathbb{R})/O(n, \mathbb{R})$ *are bijective.*

Chapter 10
The Structure Theory of Linear Mappings

Throughout this chapter, V will be a finite-dimensional vector space over $\mathbb{F}$. Our goal is to prove two theorems that describe the structure of an arbitrary linear mapping $T : V \to V$ having the property that all the roots of its characteristic polynomial lie in $\mathbb{F}$. To describe this situation, let us say that $\mathbb{F}$ contains the eigenvalues of T. A linear mapping $T : V \to V$ is also called an endomorphism of V, and in this chapter, we will usually use that term. The structure theory for the endomorphisms of V is one of the nicest chapters in the theory of finite-dimensional vector spaces. The first result we will prove, known as the Jordan–Chevalley decomposition, asserts that every endomorphism T as above can be uniquely expressed as the sum $T = S + N$ of a semisimple endomorphism S and nilpotent endomorphism N such that $TS = ST$ and $SN = NS$ (hence $TN = NT$). The endomorphisms S and N are called, respectively, the *semisimple part of* T and the *nilpotent part of* T, and the expression $T = S + N$ is known as the *Jordan–Chevalley decomposition of* T.

The second structure theorem, known as the Jordan canonical form, is a refinement of the Jordan–Chevalley decomposition. It says that there exists a basis of V such that the matrix of T is a direct sum of Jordan blocks. A Jordan block is a matrix of the form $J = \mu I_n + N$, where N is an upper triangular $n \times n$ matrix with ones on the superdiagonal and zeros everywhere else. In particular, N is nilpotent. Notice that μ is the unique eigenvalue of J. A matrix J in this form is said to be in Jordan canonical form.

These structure theorems require that all the eigenvalues of the endomorphism T lie in $\mathbb{F}$. If $\mathbb{F}$ is algebraically closed, for example if $\mathbb{F} = \mathbb{C}$, then they apply to all endomorphisms. We can satisfy this eigenvalue assumption by letting $V = \mathbb{F}^n$, since then we are dealing with $n \times n$ matrices over $\mathbb{F}$, and by the appendix to Chap. 8, there exists an extension field $\mathbb{F}'$ of $\mathbb{F}$ containing every eigenvalue of A. Thus A can be decomposed as $A = S' + N'$, where S'

J.B. Carrell, *Groups, Matrices, and Vector Spaces*,
DOI 10.1007/978-0-387-79428-0_10

and N' are matrices over $\mathbb{F}'$ that have the same roles as S and N; that is, S' is semisimple, N' is nilpotent, $S'N' = N'S'$, and $AS' = S'A$.

10.1 The Jordan–Chevalley Theorem

From now on, assume that V is a vector space over $\mathbb{F}$ such that $\dim V = n$, and suppose $T : V \to V$ is an arbitrary linear mapping whose characteristic polynomial is $p_T(x) = (x - \lambda_1)^{\mu_1} \cdots (x - \lambda_m)^{\mu_m}$, where $\lambda_1, \ldots, \lambda_m$ are the distinct eigenvalues of T and all λ_i lie in $\mathbb{F}$. By the Cayley–Hamilton theorem,

$$p_T(T) = (T - \lambda_1 I_V)^{\mu_1} \cdots (T - \lambda_m I_V)^{\mu_m} = O. \tag{10.1}$$

By Theorem 8.13, a necessary and sufficient condition that T be semisimple is that $(T - \lambda_1 I_V) \cdots (T - \lambda_m I_V) = O$. The Jordan–Chevalley theorem, which we will presently prove, will answer the question of how far T varies from being semisimple.

10.1.1 The statement of the theorem

To state the theorem, we need to define the invariant subspaces of T.

Definition 10.1. The subspaces $C_i = \ker(T - \lambda_i I_V)^{\mu_i} \subset V$ are called the *invariant subspaces* of T. If A is the matrix of T with respect to a basis of V, then the subspaces $C_i = \mathcal{N}((A - \lambda_i I_n)^{\mu_i})$ are the invariant subspaces for A.

Note that $T(C_i) \subset C_i$, so each C_i is invariant under T and similarly for A. The invariant subspaces are also called the *cyclic subspaces* of T or A.

Lemma 10.1. *Let* $T_i = (T - \lambda_i I_V)^{\mu_i}$ *so that* $C_i = \ker(T_i)$. *Then*

$$V = C_1 \oplus C_2 \oplus \cdots \oplus C_m.$$

Proof. We will apply Lemma 8.14. Thus we have to check that $T_i \circ T_j = T_j \circ T_i$ for all i and j, $T_1 \circ \cdots \circ T_m = O$, and $\ker(T_i) \cap \ker(T_{i+1} \circ \cdots \circ T_m) = \{\mathbf{0}\}$ for all i. In fact, the first statement is clear, and the second is the Cayley–Hamilton theorem. Thus, we need to check only that $\ker(T_i) \cap \ker(T_{i+1} \circ \cdots \circ T_m) = \{\mathbf{0}\}$ for all i. Note first that $E_{\lambda_i}(T) \cap \ker(T_{i+1} \circ \cdots \circ T_m) = \{\mathbf{0}\}$. For if $\mathbf{v} \in E_{\lambda_i}(T)$, then

$$T_{i+1} \circ \cdots \circ T_m(\mathbf{v}) = (\lambda_i - \lambda_{i+1})^{\mu_{i+1}} \cdots (\lambda_i - \lambda_m)^{\mu_m} \mathbf{v}.$$

Thus if $\mathbf{v} \in E_{\lambda_i}(T) \cap \ker(T_{i+1} \circ \cdots \circ T_m)$, then $\mathbf{v} = \mathbf{0}$. Now suppose $\mu_i > 1$ and take $\mathbf{v} \in \ker(T_i) \cap \ker(T_{i+1} \circ \cdots \circ T_m)$ such that $\mathbf{v} \neq \mathbf{0}$. Let $a \leq \mu_i$ be the least positive integer such that $(T - \lambda_i I_V)^a \mathbf{v} = \mathbf{0}$. By the previous case, we may assume $a > 1$, so let $\mathbf{w} = (T - \lambda_i I_V)^{a-1}\mathbf{v}$. Then by definition, $\mathbf{w} \in E_{\lambda_i}(T)$ and $\mathbf{w} \neq \mathbf{0}$. But $\mathbf{w} \in \ker(T_{i+1} \circ \cdots \circ T_m)$ also, which implies $\mathbf{w} = \mathbf{0}$, contradicting the choice of $\mathbf{w}$. Hence $\mathbf{v} = \mathbf{0}$, so the lemma is proved. □

Let us now state the theorem.

Theorem 10.2 (Jordan–Chevalley decomposition theorem)**.** *Let $T : V \to V$ be an endomorphism all of whose eigenvalues lie in $\mathbb{F}$ and whose characteristic polynomial is given by (10.1), where $\lambda_1, \ldots, \lambda_m$ are the distinct eigenvalues of T. Suppose $C_1, C_2, \ldots, C_m$ are the invariant subspaces of T. Then*

$$V = C_1 \oplus C_2 \oplus \cdots \oplus C_m. \tag{10.2}$$

Let $S : V \to V$ be the unique endomorphism such that $S(\mathbf{v}) = \lambda_i \mathbf{v}$ if $\mathbf{v} \in C_i$. Then:

(i) S is semisimple, and the linear mapping $N = T - S$ is nilpotent;

(ii) S and N commute, and both commute with T: $SN = NS$, $NT = TN$, and $ST = TS$;

(iii) the decomposition $T = S + N$ of T into the sum of a semisimple linear mapping and a nilpotent linear mapping that commute is unique; and finally,

(iv) $\dim C_i = \mu_i$ for all i.

The proof is given in the next section. The reader who wishes to see an example in which the decomposition is explicitly computed can go directly to Example 10.1.

Definition 10.2. The decomposition $T = S + N$ is called the *Jordan–Chevalley decomposition* of T. The mapping S is called the *semisimple part* of T, and N is called its *nilpotent part*.

Corollary 10.3. *An $n \times n$ matrix A over $\mathbb{F}$ whose eigenvalues all lie in $\mathbb{F}$ can be expressed in exactly one way as a sum $A = L + M$ of two commuting matrices L and M over $\mathbb{F}$ such that L is diagonalizable and M is nilpotent.*

Proof. First, let $T_A = T_L + T_M$ be the Jordan–Chevalley decomposition of T_A. This gives the decomposition $A = L + M$, as asserted. □

The Jordan–Chevalley decomposition enables us finally to clear up an intriguing question: is the algebraic multiplicity of an eigenvalue always an upper bound for its geometric multiplicity?

Corollary 10.4. *Let λ be an eigenvalue of an endomorphism $T : V \to V$ having algebraic multiplicity μ. Then* $\dim E_\lambda(T) \leq \mu$.

Proof. Let λ_i be an eigenvalue. By part (iv) of the Jordan–Chevalley theorem, the algebraic multiplicity of λ_i, namely μ_i, is the dimension of the invariant subspace C_i. But $E_{\lambda_i}(T) \subset C_i$, so the geometric multiplicity of λ_i, namely $\dim E_{\lambda_i}(T)$, is at most μ_i. □

10.1.2 The multiplicative Jordan–Chevalley decomposition

There is also a multiplicative version of the Jordan–Chevalley decomposition for nonsingular matrices. Let us first mention that the notion of a unipotent matrix used in the $LPDU$ decomposition is a special case of what is generally known as a unipotent matrix: a matrix $A \in \mathbb{F}^{n\times n}$ is called *unipotent* if $A - I_n$ is nilpotent.

Proposition 10.5. *Suppose $A \in GL(n, \mathbb{F})$ and all the eigenvalues of A lie in $\mathbb{F}$. Then one can write A uniquely in the form $A = A_s A_u$, where A_s is semisimple, A_u is unipotent, and $A_s A_u = A_u A_s$. Moreover, this decomposition is unique.*

Proof. Let $A = S + N$ be the Jordan–Chevalley decomposition of A. Since the eigenvalues of A and S are the same, it follows that S is nonsingular. Put $A_s = S$, so that A_s is semisimple. Next put $A_u = I_n + S^{-1}N$. Since S and N commute, $S^{-1}N$ is nilpotent; hence A_u is unipotent. This shows that $A = A_s A_u$ and $A_s A_u = A_u A_s$. To prove uniqueness, note that $A = A_s(I_n + N) = A_s + A_s N$. Since A_s and N commute, $A = A_s + A_s N$ is the Jordan–Chevalley decomposition of A. Hence, A_s and A_u are unique. □

This product decomposition applies to all $A \in GL(n, \mathbb{C})$, for example, or to every $GL(n, \mathbb{F})$ with $\mathbb{F}$ algebraically closed. The matrices A_s and A_u are known as the *semisimple* and *unipotent* parts of A. The multiplicative version of the Jordan–Chevalley decomposition implies an interesting fact about finite subgroups of $GL(n, \mathbb{C})$, or indeed $GL(n, \mathbb{F})$ if $\mathbb{F}$ is algebraically closed of characteristic zero.

Proposition 10.6. *Let $\mathbb{F}$ be algebraically closed of characteristic zero. Then every finite subgroup $G < GL(n, \mathbb{F})$ consists of diagonalizable matrices.*

Proof. Assume $A \in G$, and write its Jordan–Chevalley decomposition as $A = A_s A_u$. Since G is finite, A has finite order, say m. Then $A^m = (A_s A_u)^m = (A_s)^m (A_u)^m = I_n$. Now $(A_s)^m$ is semisimple, and $(A_u)^m$ is unipotent, so by the uniqueness of the multiplicative Jordan–Chevalley decomposition,

$(A_s)^m = (A_u)^m = I_n$. To finish the proof, we have to show that $(A_u)^m = I_n$ implies $A_u = I_n$. Write $A_u = I_n + N$, where N is nilpotent. Then

$$(A_u)^m = I_n + mN + \binom{m}{2}N^2 + \cdots + N^m.$$

Hence, $mN(I_n + \binom{m}{2}N + \cdots + N^{m-1}) = O$. Since $I_n + \binom{m}{2}N + \cdots + N^{m-1}$ is unipotent, it is invertible, so $N = O$. Therefore $A_u = I_n$, so $A = A_s$. □

In fact, the above proof shows that there is a stronger conclusion.

Proposition 10.7. *Suppose $\mathbb{F}$ is algebraically closed of characteristic p and the order of $G < GL(n, \mathbb{F})$ is prime to p. Then every element of G is diagonalizable.*

Proof. By Lagrange's theorem, the order m of every $A \in G$ is prime to p, since $|G|$ is. Therefore, by the above argument, $mN = O$, so $N = O$. Hence $A_u = I_n$, so $A = A_s$. □

If $\mathbb{F}$ has characteristic p, then there exist finite subgroups of $GL(n, \mathbb{F})$ of order p all of whose elements are unipotent. For example, a transvection matrix $I_n + E_{ij}$, where $i \neq j$, generates a subgroup of $GL(n, \mathbb{F})$ of order p isomorphic to the additive group of $\mathbb{F}_p$.

Remark. When $\mathbb{F} = \mathbb{C}$, the proof of Proposition 10.6 can also be done using the standard Hermitian inner product on $\mathbb{C}^n$. If one averages the inner products $(A\mathbf{v}, A\mathbf{w})$ over all $A \in G$, then one obtains a Hermitian inner product on $\mathbb{C}^n$ for which the matrices $A \in G$ are normal. Thus, every element of G is unitarily diagonalizable, by the normal matrix theorem.

10.1.3 The proof of the Jordan–Chevalley theorem

As proved in Lemma 10.1,

$$V = C_1 \oplus \cdots \oplus C_m.$$

Thus each $\mathbf{v} \in V$ has a unique expression $\mathbf{v} = \sum_{i=1}^m \mathbf{c}_i$ with $\mathbf{c}_i \in C_i$. Applying Proposition 7.5, there exists a unique linear mapping $S : V \to V$ defined by setting

$$S(\mathbf{v}) = \sum_{i=1}^m \lambda_i \mathbf{c}_i. \tag{10.3}$$

Moreover, by definition, S is semisimple. Now put $N = T - S$. We claim that N is nilpotent. For this, it suffices to show that there exists an integer $r > 0$ such that $N^r = O$ on V. By definition, if $\mathbf{v} \in C_i$, then

$$N^{\mu_i}(\mathbf{v}) = (T - S)^{\mu_i}(\mathbf{v}) = (T - \lambda_i I_V)^{\mu_i}(\mathbf{v}) = \mathbf{0}.$$

But V is the sum of the C_i, so $N^r = O$ on V if $r > \mu_i$ for all i. This shows that N is nilpotent and completes the verification of part (i). We will leave part (ii), that T, S, and N all commute with each other, to the reader. We next prove part (iii): the decomposition $T = S + N$ satisfying (i) and (ii) is unique. Let $T = S' + N'$ be another decomposition of T, where S' is semisimple, N' is nilpotent, and $S'N' = N'S'$. Since S' is semisimple, we can write $V = W_1 \oplus \cdots \oplus W_k$, where the W_i are the eigenspaces for S', and since $TS' = S'T$, it follows that $T(W_i) \subset W_i$ for each i. Thus, $N'(W_i) \subset W_i$ also. But since N' is nilpotent, $\ker(N') \cap W_i \neq \{\mathbf{0}\}$. Thus, there exists $\mathbf{v} \in W_i$ such that $(T - S')\mathbf{v} = \mathbf{0}$. This means that $T(\mathbf{v}) = S'(\mathbf{v}) = \nu_i \mathbf{v}$, where ν_i is the eigenvalue of S on W_i. Thus, every eigenvalue of S' is an eigenvalue of T. Now it follows from the argument in Lemma 10.1 that $C_j = \ker(T - \lambda_j I_n)^{\mu_j + k}$ for all $k \geq 0$. Thus if $\nu_i = \mu_j$, then $W_i \subset C_j$. But this implies $W_i = C_j$, since $\sum \dim W_j = \sum \dim C_i = \dim V$. Hence $S' = S$ and $N' = N$.

It remains to prove (iv), that is, that $\mu_i = \dim C_i$ for all i. Note that the matrix A of T has the block form

$$A = \begin{pmatrix} A_1 & O & \cdots & O \\ O & A_2 & \cdots & O \\ \vdots & \vdots & \ddots & \vdots \\ O & \cdots & O & A_m \end{pmatrix},$$

where A_i is the matrix of T on C_i. It follows from the product rule for determinants that

$$p_A(x) = p_{A_1}(x) \cdots p_{A_m}(x) = (x - \lambda_1)^{\ell_1} \cdots (x - \lambda_m)^{\ell_m}, \tag{10.4}$$

since the only eigenvalue of A_i is λ_i. Therefore, $\dim C_i = \ell_i = \mu_i$. This proves (iv). □

10.1.4 An example

The following example shows that even the 3×3 case can be a little complicated.

Example 10.1. Let $V = \mathbb{C}^3$, and consider the endomorphism T_A, where

$$A = \begin{pmatrix} 1 & 2 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 2 \end{pmatrix}.$$

The characteristic polynomial of A is $-(x-1)^2(x-2)$, so the distinct eigenvalues are 2 and 1, which is repeated. Note that $(A - I_3)(A - 2I_3) \neq O$, so A is not semisimple. The matrices $(A - I_3)^2$ and $A - 2I_3$ row reduce respectively to

$$\begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 1 & 0 & -3 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{pmatrix}.$$

Thus $C_1 = \operatorname{span}\{\mathbf{e}_1, \mathbf{e}_2\}$, and $C_2 = \mathbb{C} \begin{pmatrix} 3 \\ 1 \\ 1 \end{pmatrix}$. Hence the semisimple part of A is the matrix S determined by

$$S \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad S \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad S \begin{pmatrix} 3 \\ 1 \\ 1 \end{pmatrix} = 2 \begin{pmatrix} 3 \\ 1 \\ 1 \end{pmatrix}.$$

As usual, S is found by $SP = PD$. Therefore,

$$S = \begin{pmatrix} 1 & 0 & 3 \\ 0 & 1 & 1 \\ 0 & 0 & 2 \end{pmatrix},$$

and we get N by subtraction:

$$N = \begin{pmatrix} 0 & 2 & -2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

(As a check, make sure that $SN = NS$.) Thus $A = S + N$ is the Jordan–Chevalley decomposition. ⊠

Notice that if P is the matrix that diagonalizes S, i.e.,

$$P = \begin{pmatrix} 1 & 0 & 3 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix},$$

then a change of basis using P puts the matrix of T_A into block diagonal form. Namely,

$$P^{-1}AP = \begin{pmatrix} 1 & 2 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix}.$$

By choosing P more carefully, we can guarantee that $P^{-1}TP$ is in Jordan canonical form. For this, see Example 10.4.

10.1.5 The Lie bracket

Given two endomorphisms $S, T : V \to V$, their *Lie bracket* $[S, T]$ is defined by $[S, T] = S \circ T - T \circ S$. Thus $[S, T]$ is an endomorphism of V that measures how much S and T fail to commute. Now let $g\ell(V)$ denote the vector space of all endomorphisms of V. Recall from Section 7.5.1 that $g\ell(V)$ has dimension $(\dim V)^2$. By fixing S, one obtains an endomorphism of $g\ell(V)$, called the *adjoint of* S, which is denoted by $\mathrm{ad}(S)$ and defined by the rule $\mathrm{ad}(S)(T) = [S, T]$. For example, if $V = \mathbb{F}^n$, then $g\ell(V) = \mathbb{F}^{n\times n}$, and we know that a basis of $g\ell(V)$ is given by the matrices E_{ij}. Then if $i \neq j$, we have $\mathrm{ad}(E_{ij})(E_{ji}) = E_{ii} - E_{jj}$. We leave it as an exercise to compute $\mathrm{ad}(E_{ij})(E_{rs})$ for all r, s (see Exercise 10.1.6) below). Now suppose $\mathbb{F}$ is algebraically closed. Then a basic theorem about the adjoint mapping says that if we denote the semisimple and nilpotent parts of S by S_{ss} and S_{nilp}, then the Jordan–Chevalley decomposition of $ad(S)$ is $\mathrm{ad}(S) = \mathrm{ad}(S_{ss}) + \mathrm{ad}(S_{nilp})$. In other words, $\mathrm{ad}(S)_{ss} = \mathrm{ad}(S_{ss})$ and $\mathrm{ad}(S)_{nilp} = \mathrm{ad}(S_{nilp})$. We ask the reader to verify this when $g\ell(V) = \mathbb{F}^{2\times 2}$ in the exercises below.

Exercises

Exercise 10.1.1. Show that invariant subspaces of a linear mapping T are actually invariant under T. That is, $T(C_i) \subset C_i$ for each i.

Exercise 10.1.2. Discuss the Jordan–Chevalley decomposition of the following types of endomorphisms or matrices:

(*i*) rotations of $\mathbb{R}^2$ (i.e., elements of $SO(2, \mathbb{R})$),

(*ii*) rotations of $\mathbb{R}^3$,

(*iii*) the cross product mapping $C_{\mathbf{a}}(\mathbf{x}) = \mathbf{a} \times \mathbf{x}$ on $\mathbb{R}^3$, and

(*iv*) 2×2 complex matrices of determinant and trace zero.

Exercise 10.1.3. Let $A \in \mathbb{F}^{n\times n}$. Show how to express the semisimple and nilpotent parts of A^2 and A^3 in terms of those of A.

Exercise 10.1.4. Describe all real 2×2 matrices that are both symmetric and nilpotent.

Exercise 10.1.5. Find the Jordan–Chevalley decomposition of the matrices in parts (i)–(iii) of Exercise 8.5.9.

Exercise 10.1.6. Consider the standard basis E_{ij}, $1 \leq i, j \leq 2$, of $\mathbb{F}^{2\times 2}$.

(i) Compute the matrices of $\mathrm{ad}(E_{11})$ and $\mathrm{ad}(E_{12})$ with respect to the standard basis.

(ii) Calculate the Jordan–Chevalley decomposition of each of $\mathrm{ad}(E_{11})$ and $\mathrm{ad}(E_{12})$.

Exercise 10.1.7. Show that an $n \times n$ matrix A is unipotent if and only if its characteristic polynomial is $(1 - x)^n$.

Exercise 10.1.8. Prove that the Lie bracket on $g\ell(V)$ satisfies the Jacobi identity

$$[R, [S, T]] + [T, [R, S]] + [S, [T, R]] = 0.$$

Exercise 10.1.9. Show that the Jacobi identity on $g\ell(V)$ can be restated as the identity

$$ad([S.T]) = [ad(S), ad(T)].$$

This says that if $g\ell(V)$ is made into a ring where the multiplication is $R \circ S = [R, S]$, then ad is a ring homomorphism.

Exercise 10.1.10. Let $T : \mathbb{C}^9 \to \mathbb{C}^9$ be an endomorphism with characteristic polynomial

$$(t + 1)^2(t - 1)^3(t^2 + 1)^2 = (t + 1)(t - 1)^3(t - i)^2(t + i)^2,$$

and suppose the minimal polynomial of T is $(t + 1)(t - 1)^2(t^2 + 1)^2$. What is the rank of the nilpotent part of T on each invariant subspace of T?

10.2 The Jordan Canonical Form

The Jordan canonical form is one of the central results in linear algebra. It asserts that if all the eigenvalues of $A \in \mathbb{F}^{n\times n}$ lie in $\mathbb{F}$, then A is similar to a matrix over $\mathbb{F}$ that is the direct sum of Jordan blocks. In particular, for an endomorphism $T : V \to V$ having all its eigenvalues in $\mathbb{F}$, there exists a basis of V that is an eigenbasis for the semisimple part of T and simultaneously a string basis, as defined below, for the nilpotent part.

10.2.1 Jordan blocks and string bases

A *Jordan block* is an $n \times n$ matrix J over $\mathbb{F}$ of the form

$$J = \lambda I_n + N,$$

where $\lambda \in \mathbb{F}$ and $N = (n_{ij})$, where

$$n_{ij} = \begin{cases} 1 & \text{if j = i+1,} \\ 0 & \text{otherwise.} \end{cases}$$

In other words, N is an upper triangular matrix with ones on its superdiagonal and zeros elsewhere. The matrix N is called a *nilpotent Jordan block*. Notice that if $n = 1$, then $J = (\lambda)$. An $n \times n$ matrix A is said to be in *Jordan canonical form* if it has the form

$$A = \begin{pmatrix} J_1 & O & \cdots & O \\ O & J_2 & \cdots & O \\ \vdots & \vdots & \ddots & \vdots \\ O & \cdots & O & J_m \end{pmatrix}, \tag{10.5}$$

where $J_1, \ldots, J_m$ are Jordan blocks.

Example 10.2. There are four 3×3 matrices in Jordan canonical form having eigenvalue λ:

$$J_1 = \begin{pmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{pmatrix},\ J_2 = \begin{pmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda \end{pmatrix},\ J_3 = \begin{pmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{pmatrix} \text{ and } J_4 = \lambda I_3.$$

The first matrix, J_1, is itself a Jordan block; J_2 and J_3 have two Jordan blocks; and J_4 has three Jordan blocks. The reader can check that J_2 and J_3 are similar. ⊠

Example 10.3. The 5×5 nilpotent matrix

$$N = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

is a sum of two nilpotent Jordan blocks, one of size 3×3 and the other of size 2×2. As a linear mapping, N sends

$$\mathbf{e}_3 \to \mathbf{e}_2 \to \mathbf{e}_1 \to \mathbf{0} \quad \text{and} \quad \mathbf{e}_5 \to \mathbf{e}_4 \to \mathbf{0}. \tag{10.6}$$

⊠

Definition 10.3. Suppose $N : V \to V$ is a linear mapping. A basis of V that can be partitioned into disjoint subsets $\{\mathbf{w}_1, \ldots, \mathbf{w}_\ell\}$ such that $N(\mathbf{w}_1) = \mathbf{0}$ and $N(\mathbf{w}_i) = \mathbf{w}_{i-1}$ for $i = 2, \ldots, \ell$ is called an *N-string basis.* The subsets $\{\mathbf{w}_1, \ldots, \mathbf{w}_\ell\}$ will be called *N-strings.*

In Example 10.6, the sequences in (10.6) give the N-strings, and $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3, \mathbf{e}_4, \mathbf{e}_5\}$ is the N-string basis. Note that if $N = O$, then the N-strings consisting of the singletons $\{\mathbf{e}_i\}$ form a string basis. But in fact, every basis determines a string basis for the zero matrix.

Put schematically, N acts on the N-strings as follows:

$$\mathbf{w}_\ell \to \mathbf{w}_{\ell-1} \to \cdots \to \mathbf{w}_1 \to \mathbf{0}.$$

The matrix of N with respect to the above N-string is the nilpotent Jordan block of size $(\ell + 1) \times (\ell + 1)$, and the matrix of N with respect to a string basis is a direct sum of nilpotent Jordan blocks.

10.2.2 Jordan canonical form

Here is the main result.

Theorem 10.8 (The Jordan canonical form)**.** *Let V be a finite-dimensional vector space over $\mathbb{F}$, and suppose $T : V \to V$ is an endomorphism whose eigenvalues all lie in $\mathbb{F}$. Then there exists an eigenbasis for the semisimple*

part of T that is also a string basis for the nilpotent part of T. Thus, there exists a basis of V for which the matrix A of T has the form

$$A = \begin{pmatrix} J_1 & O & \cdots & O \\ O & J_2 & \cdots & O \\ \vdots & \vdots & \ddots & \vdots \\ O & \cdots & O & J_m \end{pmatrix}, \tag{10.7}$$

where each J_i is a Jordan block. Note: the Jordan blocks do not necessarily have different eigenvalues.

Let us review the situation. By the Jordan–Chevalley decomposition theorem, V is the direct sum of the invariant subspaces $C_1, \ldots, C_m$ of T. Hence, choosing a basis of each C_i and taking the union of these bases gives a basis of V for which the semisimple part S of T is the diagonal matrix $\operatorname{diag}(\lambda_1 I_{\mu_1}, \ldots, \lambda_m I_{\mu_m})$, where $\mu_i = \dim C_i$. Thus, if we show that each C_i has a string basis for the nilpotent part N of T, then we obtain a basis of V for which the matrix of T is in Jordan canonical form. This will be done in the next section.

10.2.3 String bases and nilpotent endomorphisms

We will now prove that a nilpotent endomorphism $N : V \to V$ of an arbitrary finite-dimensional vector space has a string basis. We begin by noticing that string bases have the following property.

Proposition 10.9. *In a string basis of V to N, the number of strings is* $\dim \ker(N)$.

Proof. Let the strings be $\mathbf{v}_{k_i}^{(i)} \to \cdots \to \mathbf{v}_1^{(i)} \to \mathbf{0}$ for $i = 1, \ldots, \ell$. Certainly the terminal vectors $\mathbf{v}_1^{(i)}$, $1 \le i \le \ell$, span $\ker(N)$. Let $\mathbf{w} \in \ker(N)$, and use the string basis to write $\mathbf{w} = \sum_{i,j} a_{ij} \mathbf{v}_j^{(i)}$. Then $N(\mathbf{w}) = \mathbf{0}$, so $\sum_{i,j} a_{ij} \mathbf{v}_{j-1}^{(i)} = \mathbf{0}$, where $\mathbf{v}_{j-1}^{(i)} = \mathbf{0}$ if $j = 1$. Thus, $a_{i,j} = 0$ if $j > 1$. Consequently, $\mathbf{w} = \sum_i a_{i1} \mathbf{v}_1^{(i)}$. This proves that the $\mathbf{v}_1^{(i)}$, $1 \le i \le \ell$, span $\ker(N)$. Since the $\mathbf{v}_1^{(i)}$ are linearly independent, $\dim \ker(N) = \ell$. □

We now prove the main result.

Proposition 10.10. *Suppose $N : V \to V$ is a nilpotent linear mapping. Then V admits a string basis for N.*

Proof. We will induct on $\dim V$. The case $\dim V = 1$ is obvious, so assume that the proposition is true if $\dim V < r$, and suppose $\dim V = r$. We may as well assume $N \neq O$, so $N(V) \neq \{\mathbf{0}\}$. Let $W = \ker(N)$ and $U = \operatorname{im}(N)$. By the rank–nullity theorem (cf. Theorem 7.4) we have $\dim U + \dim W = \dim V$. Since N is nilpotent, $\dim W > 0$, so $\dim U < \dim V$. By definition, $N(U) \subset U$, so N is a nilpotent linear mapping on U. Hence, $\dim(U \cap W) > 0$. Now apply the induction hypothesis to U to get a basis of N-strings. By Proposition 10.9, there are exactly $\dim(U \cap W)$ N-strings in every string basis of U. Let

$$\mathbf{v}_{k_i}^{(i)} \to \cdots \to \mathbf{v}_2^{(i)} \to \mathbf{v}_1^{(i)}$$

$(i = 1, \ldots, \ell = \dim(U \cap W))$ be these strings. Since $\mathbf{v}_{k_i}^{(i)} \in \operatorname{im}(N)$, there exists $\mathbf{v}_{k_i+1}^{(i)} \in V$ such that $N(\mathbf{v}_{k_i+1}^{(i)}) = \mathbf{v}_{k_i}^{(i)}$. Thus we can form a new N-string

$$\mathbf{v}_{k_i+1}^{(i)} \to \mathbf{v}_{k_i}^{(i)} \to \cdots \to \mathbf{v}_2^{(i)} \to \mathbf{v}_1^{(i)} \tag{10.8}$$

in V. Now adjoin to the strings in (10.8) additional $\mathbf{w}_j \in W$ so that the $\mathbf{v}_1^{(i)}$ and the $\mathbf{w}_k$ form a basis of W. The number of $\mathbf{w}_k$ is exactly $\dim W - \dim(U \cap W)$. Now count the number of vectors. We have $\dim U$ vectors from the strings (10.8), and there are $\dim(U \cap W)$ of these strings, each contributing a basis vector of W. Hence because of how the string basis of U was obtained, we have a total of

$$\dim U + \dim(U \cap W) + (\dim W - \dim(U \cap W)) = \dim U + \dim W$$

vectors. But as noted above, $\dim U + \dim W = \dim V$, so it suffices to show that these vectors are independent. Thus suppose

$$\sum_{i,j} a_{ij} \mathbf{v}_j^{(i)} + \sum b_k \mathbf{w}_k = \mathbf{0}. \tag{10.9}$$

Applying N to (10.9) gives an expression

$$\sum_{i,j} a_{ij} \mathbf{v}_{j-1}^{(i)} = \mathbf{0},$$

where for each i, $1 < j \leq k_i + 1$. But then all these a_{ij} are equal to zero. Hence the original expression (10.9) becomes

$$\sum a_{i1} \mathbf{v}_1^{(i)} + \sum b_k \mathbf{w}_k = \mathbf{0}.$$

Thus the remaining coefficients are also zero, since the $\mathbf{v}_1^{(i)}$ and the $\mathbf{w}_k$ form a basis of W. Hence, we have constructed an N-string basis of V. □

There is another way to prove the existence of a string basis that the reader is invited to consider in the exercises. Since $N(\ker(N)) \subset \ker(N)$, it follows that N induces a linear mapping $\overline{N} : V/\ker(N) \to V/\ker(N)$. One sees that $\overline{N}$ is also nilpotent, so one can again use induction.

Remark. Each Jordan block of a matrix A corresponds to a subspace of one of A's invariant subspaces, and the sum of the sizes of all the Jordan blocks for a fixed eigenvalue μ of A is the dimension of the invariant subspace corresponding to μ.

Corollary 10.11. *Every $A \in \mathbb{F}^{n\times n}$ whose characteristic polynomial decomposes into linear factors over $\mathbb{F}$ is similar over $\mathbb{F}$ to a matrix in Jordan canonical form (10.7). In particular, every square matrix over $\mathbb{C}$ is similar over $\mathbb{C}$ to a matrix in Jordan canonical form.*

Example 10.4. Let us find the Jordan canonical form of

$$A = \begin{pmatrix} 1 & 2 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 2 \end{pmatrix}.$$

Recall from Example 10.1 that if

$$P = \begin{pmatrix} 1 & 0 & 3 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix},$$

then A is similar using P to a matrix in block diagonal form. Namely,

$$A = P \begin{pmatrix} 1 & 2 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix} P^{-1}.$$

Thus, we want to find a 2×2 matrix R such that

$$R \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix} R^{-1} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}.$$

Let $R = \operatorname{diag}(1/2, 2)$. Conjugation by R multiplies the first row by $1/2$ then multiplies the first column by 2. Thus, increasing R to a 3×3 matrix in an obvious way, we put

$$Q = PR' = \begin{pmatrix} 1 & 0 & 3 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 2 & 0 & 3 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}.$$

Then

$$A = Q \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix} Q^{-1},$$

giving the Jordan canonical form of A.

10.2.4 Jordan canonical form and the minimal polynomial

Let $T : V \to V$ be a linear mapping, and suppose $\mu_T(x) \in \mathbb{F}[x]$ denotes the minimal polynomial of T. Let $\lambda_1, \ldots, \lambda_m$ be the distinct eigenvalues of T. As usual, we assume $\lambda_1, \ldots, \lambda_m \in \mathbb{F}$. Let $C_1, \cdots, C_m$ be the invariant subspaces of V, and let $\mu_i = \dim C_i$. Now

$$\mu_T(x) = (x - \lambda_1)^{a_1} \cdots (x - \lambda_m)^{a_m},$$

where each a_i is greater than zero. If we choose a basis of each C_i and consider the Jordan–Chevalley decomposition $T = S + N$, where S is the semisimple part and N is the nilpotent part of T, then $\mu_T(T) = O$ means that each $N_i^{a_i}$ equals O on C_i. But this means that no nilpotent Jordan block of N_i can be larger than $a_i \times a_i$.

Here is an example.

Example 10.5. Let $T : \mathbb{C}^9 \to \mathbb{C}^9$ be an endomorphism with characteristic polynomial

$$(t+1)^2(t-1)^3(t^2+1)^2 = (t+1)(t-1)^3(t-i)^2(t+i)^2,$$

and suppose its minimal polynomial is $(t+1)(t-1)^2(t^2+1)^2$. Let us try to find the Jordan canonical form of T. Let C_1, C_2, C_3, and C_4 be the invariant subspaces for the eigenvalues -1, 1, i, $-i$ respectively of T. Then $\dim C_1 = 2$, $\dim C_2 = 3$, $\dim C_3 = 2$, and $\dim C_4 = 2$. By the above comments, the nilpotent part N of T is zero on C_1, so T has two 1×1 Jordan blocks with eigenvalue -1 on C_1. Now, T has a 3×3 Jordan block with eigenvalue 1 on C_2. Finally, T has a 2×2 Jordan block for i on C_3 and a 2×2 Jordan block for $-i$ on C_4. ⊠

10.2.5 The conjugacy class of a nilpotent matrix

Since the eigenvalues of a nilpotent matrix are zero, the Jordan canonical form is similar to a matrix of the form

$$N = \begin{pmatrix} J_{n_1} & O & \cdots & O \\ O & J_{n_2} & \cdots & O \\ \vdots & \vdots & \ddots & \vdots \\ O & \cdots & O & J_{n_s} \end{pmatrix}, \tag{10.10}$$

where J_{n_i} is the $n_i \times n_i$ nilpotent Jordan block and $n_1, \ldots, n_s$ are the lengths of the strings in an N-string basis of V. By conjugating N with a permutation matrix, we may also assume that the blocks are arranged so that $n_1 \geq n_2 \geq \cdots \geq n_s$. Such a sequence $n_1 \geq n_2 \geq \cdots \geq n_s$ of positive integers such that $n_1 + n_2 + \cdots + n_s = n$ is called a *partition* of n. Thus every nilpotent $n \times n$ matrix over $\mathbb{F}$ determines a unique partition of n. But Theorem 10.8 says that if two nilpotent $n \times n$ matrices over $\mathbb{F}$ determine the same partition of n, then they are conjugate or similar by an element of $GL(n, \mathbb{F})$. Thus the conjugacy classes of nilpotent $n \times n$ matrices over $\mathbb{F}$ are in one-to-one correspondence with the partitions of n. The partition function $\pi(n)$ is the function that counts the number of partitions of n. It starts out slowly and grows rapidly with n. For example, $\pi(1) = 1$, $\pi(2) = 2$, $\pi(3) = 3$, and $\pi(4) = 5$, while $\pi(100) = 190,569,292$. This unexpected connection between partitions, which lie in the domain of number theory, and the conjugacy classes of nilpotent matrices, which lie in the domain of matrix theory, has led to some interesting questions.

Exercises

Exercise 10.2.1. Suppose V is a finite-dimensional vector space and $T : V \to V$ is an endomorphism. Show that if $\ker(T) = \ker(T^2)$, then $\ker(T^m) = \ker(T)$ for all $m > 0$. What does this say about the minimal polynomial of T?

Exercise 10.2.2. Suppose $T : V \to V$ is an endomorphism, and suppose there exists a T-string basis of V. Show that T is nilpotent.

Exercise 10.2.3. Find the Jordan canonical form of the matrices in parts (i)–(iii) of Exercise 8.5.9.

Exercise 10.2.4. Let A be the 4×4 all-ones matrix over $\mathbb{F}_2$. Find the Jordan canonical form of A and resolve the paradox.

Exercise 10.2.5. Let $\mathbf{a} \in \mathbb{R}^3$. Find the Jordan canonical form over $\mathbb{C}$ of the cross product $C_{\mathbf{a}} : \mathbb{R}^3 \to \mathbb{R}^3$.

Exercise 10.2.6. Compute $\pi(5)$ and write down all $\pi(5)$ 5×5 nilpotent matrices in Jordan canonical form with decreasing blocks along the diagonal.

Exercise 10.2.7. Show that the nilpotent matrix N in (10.10) is similar via a permutation matrix to a nilpotent matrix N' in Jordan canonical form such that the block sizes form a decreasing sequence.

Exercise 10.2.8. True or false. State your reasoning.

(i) Two matrices over $\mathbb{F}$ with the same characteristic polynomial must be similar.

(ii) Two matrices with the same minimal polynomial must be similar.

Exercise 10.2.9. Show without appealing to Jordan canonical form that if $A \in \mathbb{F}^{n\times n}$ and all the roots of $p_A(x)$ lie in $\mathbb{F}$, then A is similar over $\mathbb{F}$ to an upper triangular matrix. (Hint: Assume first that A is nilpotent. Let k be the least positive integer for which $A^k = O$. Then

$$\mathcal{N}(A) \subset \mathcal{N}(A^2) \subset \cdots \subset \mathcal{N}(A^k) = \mathbb{F}^n.$$

Now construct a basis $\mathcal{B}$ of $\mathbb{F}^n$ such that $B = \mathcal{M}_{\mathcal{B}}^{\mathcal{B}}(A)$ is upper triangular. Then show that B can't have any nonzero entries on its diagonal. Finally, consider the Jordan–Chevalley decomposition of A.)

Exercise 10.2.10. Let $N : V \to V$ be a nilpotent linear mapping, where V is a finite-dimensional vector space.

(i) Show that $N(\ker(N)) \subset \ker(N)$, and conclude that N induces a nilpotent linear mapping $\overline{N} : V/\ker(N) \to V/\ker(N)$.

(ii) Show by induction that V admits an N-string basis subordinate to N using induction on $\dim V$ with part (i).

Exercise 10.2.11. This exercise requires the use of limits. Let $A \in \mathbb{C}^{n\times n}$. Show that A is nilpotent if and only if there exists a homomorphism $\varphi : \mathbb{C}^* \to GL(n, \mathbb{C})$ such that $\lim_{t\to 0} \varphi(t)A\varphi(t)^{-1} = O$. This fact is a special case of the Hilbert–Mumford criterion.

Chapter 11
Theorems on Group Theory

Our treatment of matrix theory and the theory of finite-dimensional vector spaces and their linear mappings is finished, and we now return to the theory of groups. The most important results in group theory are of two types: deep results about finite groups, and results about linear algebraic groups. Linear algebraic groups are matrix subgroups of some $GL(n, \mathbb{F})$ that are solutions of polynomial equations. We give an introduction to this subject in the last chapter. One of the deepest results in finite group theory is the theorem that all finite groups of odd order are solvable. Another result, which is truly remarkable, is that all the finite simple groups have now been described. This was accomplished by a joint effort of many mathematicians who filled in the details of an ambitious program that eventually consumed more than 10,000 pages in mathematics research journals. The final step in this project was the discovery of a so-called sporadic simple group known as the monster, whose order is

$$808017424794512875886459904617107570057543 68 \times 10^9,$$

a number that is said to exceed the number of elementary particles in the universe. But the discovery of this monstrous group did not close the subject. In fact, it opened up several fascinating questions in areas such as number theory that are usually not related to group theory. Moreover, there are even conjectures that the monster has deep yet to be discovered connections to physics. Linear algebraic groups, which we will briefly introduce in the last chapter, are closely related to two of the major areas of pure mathematics, algebraic geometry and representation theory. It is harder to state results in this area succinctly, but we will make an attempt in the next chapter. In the last fifty years, there has been a massive amount of important research in these three areas.

J.B. Carrell, *Groups, Matrices, and Vector Spaces*,
DOI 10.1007/978-0-387-79428-0_11

Our starting point is the theory of group actions and orbits, which is the collection of ideas necessary for the orbit stabilizer theorem. The main consequence of the theory of orbits is Cauchy's theorem and a collection of theorems known as the Sylow theorems, originally proved in 1872 by a Norwegian high-school teacher, Ludwig Sylow, which still form one of the most basic sets of tools in the theory of finite groups. If G has order mp^n, where p is prime, $n \geq 1$, and m is prime to p, then a subgroup H of order p^n is called a Sylow p-subgroup. The Sylow theorems describe several properties of the Sylow subgroups of a finite group G. For example, according to the first Sylow theorem, G has a Sylow p-subgroup for every prime p dividing the order of G, and by the second Sylow theorem, any two Sylow p-subgroups of G are conjugate. The finite abelian groups are completely classified by the Sylow theorems. The orbit method will also be used to show that the only finite subgroups of $SO(3, \mathbb{R})$ are the dihedral groups, cyclic rotation groups, and the symmetry groups of the Platonic solids. This extends our classification of the finite subgroups of $SO(2)$.

11.1 Group Actions and the Orbit Stabilizer Theorem

We will now lay the foundation for a number of results on the structure of finite groups. The basic notion is the idea of a group action, which leads directly to a simple yet extremely powerful observation known as the orbit stabilizer theorem, which is the basis for several counting arguments employed below. Following the usual practice in group theory, from now on we will write $H < G$ (or equivalently, $G > H$) whenever H is a proper subgroup of G and write $H \leq G$ if H is a subgroup that may coincide with G.

11.1.1 Group actions and G-sets

Let X be a set (finite or infinite), and recall that $\mathrm{Sym}(X)$ denotes the group of all bijections of X.

Definition 11.1. A group G is said to *act* on X if there exists a homomorphism $\varphi : G \to \mathrm{Sym}(X)$. If such a homomorphism φ exists, then $g \in G$ acts on X by sending each element $x \in X$ to the element $g \cdot x$ defined by

$$g \cdot x = \varphi(g)(x).$$

A mapping $\varphi : G \to \mathrm{Sym}(X)$ is a homomorphism if and only if for every $g, h \in G$ and $x \in X$, we have

$$(gh) \cdot x = g \cdot (h \cdot x). \tag{11.1}$$

The corresponding mapping $G \times X \to X$ sending $(g, x) \to g \cdot x$ will be called an *action* of G on X, and we will say that X is a *G-set*. Note that if G acts on X, then the identity of G acts as the identity bijection on X; that is, $1 \cdot x = x$ for all $x \in X$. In general, there may be other elements of G that act as the identity on X, namely the elements of $\ker(\varphi)$.

Definition 11.2. Let X be a G-set, and suppose $x \in X$. The set

$$G \cdot x = \{y \in X \mid y = g \cdot x \ \exists \ g \in G\}$$

is called the G-orbit of x. The *stabilizer* G_x of x is defined by

$$G_x = \{g \in G \mid g \cdot x = x\}.$$

For each $x \in X$, the mapping $G \to X$ defined by $g \to g \cdot x$ of G onto $G \cdot x$ is called the *orbit map associated to* x. Finally, the action of G on X is said to be *transitive* if given any pair of elements $x, y \in X$, there exists $g \in G$ such that $g \cdot x = y$.

For example, every orbit $G \cdot x$ is also a G-set and G acts transitively (in the obvious way) on $G \cdot x$. In particular, the orbit map $g \to g \cdot x$ of G to $G \cdot x$ is surjective. The symmetric group $S(n)$ acts transitively on $X_n = \{1, 2, \ldots, n\}$. Another transitive action is given in the following example.

Example 11.1. A group G acts on itself by left multiplication as follows: if $g \in G$, let $L_g : G \to G$ be the map given by $L_g(h) = gh$. This map is called *left translation* by g. Left translation is a bijection of G, and the mapping $\varphi : G \to \mathrm{Sym}(G)$ defined by $\varphi(g) = L_g$ is a homomorphism, by associativity. Recall that left translation was already used in the proof of Cayley's theorem (Theorem 2.7). In the proof of Proposition 2.5 it was shown that left translation by g is a bijection of G. ⊠

Proposition 11.1. *Let X be a G-set, and suppose $x \in X$. Then the stabilizer G_x is a subgroup of G. Moreover, the stabilizer of $g \cdot x$ is gG_xg^{-1}. Consequently, if G_x is finite and x and y lie in the same G-orbit, then $|G_x| = |G_y|$.*

The proof is left to the reader. We next show that a group action on X defines an equivalence relation on X whose equivalence classes are exactly the orbits of G.

Definition 11.3. Let X be a G-set, and let $x, y \in X$. Then we say that x is *congruent to y modulo G* if $x \in G \cdot y$.

Proposition 11.2. *Let X be a G-set. Then congruence modulo G is an equivalence relation on X whose equivalence classes are the G-orbits. Therefore, the G-orbits give a decomposition of X into disjoint subsets.*

Proof. Since $1 \cdot x = x$, $x \in G \cdot x$, so every element of X is congruent to itself. If x is congruent to y, then $x \in G \cdot y$. Thus $x = g \cdot y$, so $y = g^{-1} \cdot x$. Hence y is congruent to x. Finally, if x is congruent to y, and y is congruent to z, then $x = g \cdot y$ and $y = h \cdot z$, so $x = g \cdot (h \cdot z) = (gh) \cdot z$. Hence $x \in G \cdot z$, so x is congruent to z. It follows that congruence modulo G is an equivalence relation. □

Corollary 11.3 (The orbit identity)**.** *Suppose a group G acts on a finite set X. Then G has finitely many orbits, say $\mathcal{O}_1, \ldots, \mathcal{O}_r$, and $|X| = |\mathcal{O}_1| + \cdots + |\mathcal{O}_r|$.*

Example 11.2 (Left and right cosets)**.** Suppose H is a subgroup of G. Then H acts on G on the left by $h \cdot a = ha$ $(h \in H,\ a \in G)$. The orbits of this action are the right cosets Ha of H. Similarly, H acts on G on the right by $h \cdot a = ah^{-1}$. The orbits of the right action are the left cosets aH of H. ⊠

Example 11.3 (Double cosets)**.** Let H and K be subgroups of G. Then the product group $H \times K$ acts on G by $(h, k) \cdot a = hak^{-1}$. The orbits of this action have the form HaK. An orbit of the $H \times K$-action on G is called a *double coset* of the pair (H, K), or sometimes an (H, K)-*double coset*. Thus such double cosets are either disjoint or equal. ⊠

Recall that in a finite group G, all cosets aH of a subgroup H have exactly $|H|$ elements. This isn't true for double cosets, however. For example, the number of elements in HaH is $|H|$ if $a \in H$, while $|HaH| > |H|$ if $a \notin H$. The reader may wish to find a formula for $|HaH|$. (See Exercise 11.1.16.)

Example 11.4. We have already seen that $GL(n, \mathbb{F})$ has a decomposition

$$GL(n, \mathbb{F}) = \mathcal{L}(n, \mathbb{F}) \cdot P(n) \cdot \mathcal{D}(n, \mathbb{F}) \cdot \mathcal{U}(n, \mathbb{F}), \tag{11.2}$$

where $\mathcal{L}(n, \mathbb{F})$ and $\mathcal{U}(n, \mathbb{F})$ are respectively the lower triangular and upper triangular unipotent $n \times n$ matrices over $\mathbb{F}$. Let $\mathcal{T}(n, \mathbb{F}) = \mathcal{D}(n, \mathbb{F}) \cdot \mathcal{U}(n, \mathbb{F})$. Then $\mathcal{T}(n, \mathbb{F})$ is the subgroup of $GL(n, \mathbb{F})$ consisting of all invertible upper triangular $n \times n$ matrices over $\mathbb{F}$. Thus,

$$GL(n, \mathbb{F}) = \mathcal{L}(n, \mathbb{C}) \cdot P(n) \cdot \mathcal{T}(n, \mathbb{F})$$

is a double coset decomposition of $GL(n, \mathbb{F})$ consisting of the double cosets $\mathcal{L}(n, \mathbb{F})\sigma\mathcal{T}(n, \mathbb{F})$, where σ varies over the group $P(n)$ of $n \times n$ permutation matrices. This particular decomposition is called the *Birkhoff decomposition*. There is a similar decomposition

$$GL(n, \mathbb{F}) = \mathcal{T}(n, \mathbb{F}) \cdot P(n) \cdot \mathcal{T}(n, \mathbb{F}),$$

called the *Bruhat decomposition*. We will discuss the Bruhat decomposition in the final chapter. ⊠

11.1.2 The orbit stabilizer theorem

We now come to the orbit stabilizer theorem, which gives a powerful method for counting the number of elements in an orbit of a finite group.

Theorem 11.4 (The orbit stabilizer theorem)**.** *Let G be a finite group and let X be a G-set. Then for every $x \in X$, the mapping $g \to g \cdot x$ induces a bijection $\pi_x : G/G_x \to G \cdot x$ from the coset space G/G_x to the orbit $G \cdot x$. Hence*

$$|G \cdot x| = \frac{|G|}{|G_x|}.$$

In particular, $|G \cdot x|$ divides $|G|$.

Proof. The proof is similar to the proof of Lagrange's Theorem. Fix $x \in X$ and let H denote the stabilizer G_x. For each $a \in G$ and $h \in H$,

$$(ah) \cdot x = a \cdot (h \cdot x) = a \cdot x.$$

Hence, we can define a map $\pi_x : G/H \to G \cdot x$ by setting $\pi_x(aH) = a \cdot x$ for every $aH \in G/H$. This map is certainly surjective, since the orbit map $G \to G \cdot x$ is surjective. To finish the proof, we need to show that π_x is injective. Suppose $\pi_x(aH) = \pi_x(bH)$. Then $a \cdot x = b \cdot x$, so $a^{-1}b \cdot x = x$. Thus $a^{-1}b \in H$, and therefore $aH = bH$. Hence π_x is injective. Since $|G/H| = \dfrac{|G|}{|H|}$ by Lagrange, the proof is complete. □

11.1.3 Cauchy's theorem

Lagrange's theorem, which says that the order of a subgroup of a finite group G divides the order of G, does not have a converse. For example, the alternating group $A(4)$ has order 12, but it has no subgroup of order 6. Our first application of the orbit stabilizer theorem is Cauchy's theorem, which guarantees, however, that a group whose order is divisible by a prime p has an element of order p. Cauchy's proof of this theorem, which appeared in 1845, was unsatisfactory, being very long, hard to understand, and, as was noticed around 1980, logically incorrect, although the flaw was fixable. The theorem itself is historically important as a precursor of the Sylow theorems. As we will now see, it also is an example of a nontrivial result whose proof is made both elegant and brief by a clever application of the orbit stabilizer theorem.

Proposition 11.5 (Cauchy's theorem)**.** *A finite group G whose order is divisible by a prime p has an element of order p.*

Proof. Define

$$\Sigma = \{(a_1, a_2, \ldots, a_p) \mid \text{all } a_i \in G \text{ and } a_1 a_2 \cdots a_p = 1\}.$$

Since $(a_1, a_2, \ldots, a_p) \in \Sigma$ if and only if $a_p = (a_1 a_2 \cdots a_{p-1})^{-1}$, it follows that $|\Sigma| = |G|^{p-1}$. Thus p divides $|\Sigma|$. Note that the cyclic group $C_p =< \zeta >$ of order p acts on $G^p = G \times G \times \cdots \times G$ by rightward cyclic shifts. That is, for each $r = 1, \ldots, p$,

$$\zeta^r \cdot (g_1, g_2, \ldots, g_p) = (g_{p-r+1}, \ldots, g_p, g_1, \ldots, g_{p-r}).$$

This action is well defined on Σ, since if $a_1 a_2 \cdots a_p = 1$, then $(a_j a_{j+1} \cdots a_p)$ $(a_1 \cdots a_{j-1}) = 1$ also, due to the fact that in every group, if $xy = 1$, then $yx = 1$ too. Since $|C_p| = p$, the stabilizer of an arbitrary element of Σ has order one or p. Thus, the orbit stabilizer theorem implies that every orbit of C_p has either one or p elements. Since Σ is the disjoint union of orbits, it follows that the number of elements of Σ whose orbits consist of a single element is divisible by p. But the orbit of $\sigma = (1, 1, \ldots, 1)$ has one element, so there are at least $p-1$ more elements $\tau \in \Sigma$ such that $C_p \cdot \tau = \{\tau\}$. Every such τ has the form $(t, t, \ldots, t)$, where $t \neq 1$, so it follows that $t^p = 1$. □

From Cauchy's theorem one immediately concludes the following.

Corollary 11.6. *If G is finite and the order of every element of G is divisible by p, then the order of G is p^n for some n.*

In the second application of the orbit stabilizer theorem, we will prove the formula for the order of the rotation group of a Platonic solid stated in Section 8.7.

Proposition 11.7. *If $\mathcal{P}$ is a Platonic solid with f faces and if each face has e edges, then $|Rot(\mathcal{P})| = ef$.*

Proof. Let X denote the set consisting of the f faces of $\mathcal{P}$. Now, $Rot(\mathcal{P})$ acts on X, and it can be shown that for every two faces, there is an element of $Rot(\mathcal{P})$ that takes one face into the other. To see this, one has only to verify that every two adjacent faces can rotated one into the other in $Rot(\mathcal{P})$, which can be seen by inspection. The orbit stabilizer theorem says that $|Rot(\mathcal{P})| = f|H|$, where H is the subgroup of $Rot(\mathcal{P})$ consisting of all rotations that fix a face. But the faces are regular polygons with e edges, so H has order at most e. In fact, the order of H is exactly e. This is evident for the cube, tetrahedron, and octahedron. It also holds for the dodecahedron and icosahedron, but we will skip the details. Therefore, $|Rot(S)| = ef$. □

11.1.4 Conjugacy classes

The action of G on itself defined by $g \cdot a = gag^{-1}$ is called the *conjugation action*. We have already seen this action in a number of contexts, for example in the question of when an invertible matrix is diagonalizable. Recall that an orbit of this action is called a *conjugacy class*. Conjugation defines an action, since the mapping $\varphi : G \to \mathrm{Sym}(G)$ defined by $\varphi(g)(a) = gag^{-1}$ is a homomorphism. Recall that $\varphi(g)$ is in fact the inner automorphism σ_g. We will denote the conjugacy class of $a \in G$ by G^a.

Definition 11.4. If $a \in G$, the *centralizer* $Z_G(a)$ of a is defined to be

$$Z_G(a) = \{g \in G \mid gag^{-1} = a\}.$$

In other words, the centralizer of a consists of all $g \in G$ commuting with a. Note that the cyclic group $< a >$ generated by a satisfies $< a > \leq Z_G(a)$. Clearly, $Z_G(a)$ is the stabilizer G_a of $a \in G$ for the conjugation action, so $Z_G(a)$ is a subgroup of G. If $a \in Z(G)$, the center of G, then $Z_G(a) = G$. The orbit stabilizer theorem applied to the conjugation action of G on itself has the following consequence.

Proposition 11.8. *Let G be finite. Then for every $a \in G$, the number of elements of the G-conjugacy class G^a is $|G|/|Z_G(a)|$. Consequently, $|G^a|$ divides $|G|$.*

To test the power of our results so far, let us consider what what can be said about groups of order 15.

Example 11.5 (Groups of order 15)**.** By Cauchy's theorem, G contains elements a and b of orders 3 and 5 respectively. A natural question is whether the center of G is trivial. If not, then $|Z(G)|$ =3, 5, or 15. Any one of these possibilities implies that $G/Z(G)$ is cyclic, so G is abelian by the result in Exercise 11.1.7. But if G is abelian, then ab has order 15, so $G = C_{15}$. So suppose $Z(G) = (1)$. Since G is the union of the distinct conjugacy classes and the order of a conjugacy class divides $|G|$, we have $|G| = 15 = 1 + 3m + 5n$, where m is the number of conjugacy classes of order 3, n the number of order 5, and 1 is the order of the conjugacy class of the identity. The only solution to this equation in nonnegative integers is $m = 3$ and $n = 1$. Thus, three conjugacy classes have order 3, and one has order 5. Consider a conjugacy class G^x such that $|G^x| = 3$. By the orbit stabilizer theorem, $|Z_G(x)| = 5$. Since $x \in Z_G(x)$, it follows that x has order 5, so all elements of G^x have order 5, since the elements in a conjugacy class all have the same order. This gives nine elements of order 5. But by Proposition 2.13, the number of elements of order p in G is divisible by $p - 1$. Since 4 does not divide 9, we obtain a contradiction. Thus, $|Z(G)| > 1$, so by the remarks above, G is the cyclic group C_{15}. ⊠

To summarize, we have

Proposition 11.9. *If G has order 15, then G is cyclic.*

11.1.5 Remarks on the center

Recall from Section 2.1 that the *center* of a group G is defined to be the subgroup

$$Z(G) = \bigcap_{g \in G} Z_G(g) = \{g \in G \mid ag = ga \ \forall a \in G\}.$$

The center of G is a normal subgroup. Note that if G is finite, then the order of its center is the number of elements of G whose conjugacy class consists of a single element. We now consider a nontrivial example.

Proposition 11.10 (The center of $GL(n, \mathbb{F})$)**.** *The center of $GL(n, \mathbb{F})$ consists of the subgroup of scalar matrices $\{cI_n \mid c \in \mathbb{F}^*\}$.*

Proof. Let C belong to the center of $GL(n, \mathbb{F})$. We will first show that C is diagonal. If the characteristic of $\mathbb{F}$ is greater then two, then any matrix C commuting with all elementary matrices of type I is diagonal. If the characteristic is two, then row operations of type three must also be used to see that C is diagonal. But a nonsingular diagonal matrix that commutes with any row swap matrix has the form $C = cI_n$ for some $c \in \mathbb{F}^*$, so we are done. □

Let Z denote the center of $GL(n, \mathbb{F})$. The quotient group $GL(n, \mathbb{F})/Z$ is denoted by $PGL(n, \mathbb{F})$ and is referred to as the *projective general linear group.* By a similar argument, the center of $SL(n, \mathbb{F})$ consists of all cI_n such that $c^n = 1$. For example, if $\mathbb{F} = \mathbb{C}$, then Z is the group of nth roots of unity.

11.1.6 A fixed-point theorem for p-groups

We have already shown that G is a p-group if and only if every subgroup of G is also a p-group. We will now prove a well known result about p-groups which is a consequence of a fixed-point formula that we will also use in the proof of the Sylow theorems.

Suppose G is a group acting on a set X. We say that $x \in X$ is a *G-fixed point* if $G_x = G$, that is, the stabilizer G_x of x is all of G. Let X^G denote the set of G-fixed points in X. The fixed-point formula goes as follows.

Proposition 11.11. *Suppose G is a p-group acting on a finite set X such that p divides $|X|$. Then $|X^G|$ is also divisible by p.*

Proof. Let $\mathcal{O}_1, \ldots, \mathcal{O}_r$ be the orbits of G. Since $|\mathcal{O}_i| = |G|/|G_x|$ for every $x \in \mathcal{O}_i$, it follows that either $|\mathcal{O}_i| = 1$ or $|\mathcal{O}_i|$ is divisible by p. By the orbit identity, $|X| = \sum |\mathcal{O}_i|$, and the fact that p divides $|X|$, it follows that the number of orbits such that $|\mathcal{O}_i| = 1$ is also divisible by p. This is $|X^G|$, which gives the result. □

Now G acts on itself by conjugation, and the set of fixed points for this action is the center $Z(G)$ of G. Thus we have the following corollary.

Corollary 11.12. *The center of a p-group G is nontrivial.*

Proof. Since the fixed-point set of the conjugation action is $Z(G)$ and $1 \in Z(G)$, it follows that $|Z(G)| = mp$ for some $m \geq 1$. □

Here is another corollary.

Proposition 11.13. *If $|G| = p^2$, then G is abelian.*

The proof is left to the reader. □

11.1.7 Conjugacy classes in the symmetric group

We will now derive a well-known result describing the conjugacy classes in $S(n)$. To this end, we need to introduce cycles and disjoint cycle notation.

Definition 11.5. An element σ of $S(n)$ is called a *k-cycle* if the following two conditions hold:

(i) There exists an integer i in $[1, n]$ such that $i, \sigma(i), \sigma^2(i), \ldots, \sigma^{k-1}(i)$ are all distinct and $\sigma^k(i) = i$.

(ii) $\sigma(j) = j$ for all j in $[1, n]$ distinct from $\sigma^m(i)$ for all $m = 1, \ldots, k$.

The k-cycle σ as defined above will be denoted by $(i\, \sigma(i)\, \sigma^2(i)\, \cdots\, \sigma^{k-1}(i))$. There are other possible representations of this k-cycle, as the following example illustrates.

Example 11.6. For example, a transposition (ij) is a 2-cycle. The permutation $\sigma = [2, 3, 4, 1]$ in $S(4)$ that sends 1 to 2, 2 to 3, 3 to 4, and 4 to 1 is the 4-cycle (1234). It can also be represented as (2341), (3412), or (4123). In other words, a representation of a k-cycle in $S(n)$ is not unique. ⊠

The ambiguity in cycle notation pointed out in the above example can be avoided by letting the leading entry of a cycle $(a_1 a_2 \dots a_k)$ be the least integer among the a_i. The identity is represented by the cycle (1). Cycles are multiplied, like transpositions, by composing their permutations. For example, $(123)(13) = [1,3,2,4] = (23)$. Two cycles that don't share a common letter such as (13) and (24) are said to be *disjoint*. Disjoint cycles have the property that they commute, since they act on disjoint sets of letters. Hence the product of two or more disjoint cycles can be written in any order. For example, in $S(6)$, we have $(13)(24)(56) = (56)(24)(13)$. Nondisjoint cycles, however, do not in general commute: for example, nondisjoint transpositions never commute, since $(ab)(bc) = (abc)$, while $(bc)(ab) = (acb)$. Two more examples are $(123)(13) = (23)$ (as above), while $(13)(123) = (12)$; also $(123)(34) = (1234)$, while $(34)(123) = (1243)$. The important point is that every element of $S(n)$ can expressed as a product of *disjoint cycles*, as we now prove.

Proposition 11.14. *Every element σ of $S(n)$ different from (1) can be written as a product of one or more disjoint cycles of length greater than one. The disjoint cycles are unique up to their order.*

Proof. Given σ, we can construct a cycle decomposition as follows. Let i be the first integer in $[1,n]$ such that $\sigma(i) \neq i$. Now consider the sequence $i, \sigma(i), \sigma^2(i), \dots$. Since all $\sigma^j(i) \in [1,n]$, there has to be a least $k > 1$ such that $\sigma^k(i) \in \{i, \sigma(i), \dots, \sigma^{k-1}(i)\}$. I claim that $\sigma^k(i) = i$. For if $\sigma^k(i) = \sigma^\ell(i)$, where ℓ satisfies $1 \leq \ell < k$, then $\sigma(\sigma^{k-1}(i)) = \sigma(\sigma^{\ell-1}(i))$. Thus $\sigma^{k-1}(i) = \sigma^{\ell-1}(i)$, since σ is a bijection. Since $\ell < k$, this contradicts the definition of k. Hence $\sigma^k(i) = i$. It follows that $(i\sigma(i) \cdots \sigma^{k-1}(i))$ is a k-cycle. Now repeat this construction starting with the least $j \in [1,n]$ such that $j > i$ and j does not occur in the cycle we have just constructed. Proceeding in this way, we eventually construct a family of disjoint cycles whose product is σ such that each $j \in [1,n]$ such that $\sigma(j) \neq j$ belongs to exactly one cycle (and no j such that $\sigma(j) = j$ appears in any cycle). For the uniqueness, suppose σ has two disjoint cycle representations, say $\sigma = c_1 \cdots c_k = c'_1 \dots c'_\ell$. Assuming that i is the least integer such that $\sigma(i) \neq i$, it follows that i occurs in some c_j and some c'_m. We can assume that $j = m = 1$, so by construction, $c_1 = c'_1$. By repeating the argument, we may conclude that each c_i is a c'_j for a unique j and conversely, so the disjoint cycle representations of σ coincide up to order of the factors. □

Example 11.7. Let's express the element $\sigma = [2,3,1,5,4] \in S(5)$ as a product of disjoint cycles. Now, $\sigma(1) = 2, \sigma(2) = 3, \sigma(3) = 1$. Thus, (123) will be a cycle. The other will be (45), so $\sigma = (123)(45)$. Similarly, $\tau = [5,3,2,1,4]$ will have (154) and (23) as its disjoint cycles, so $\tau = (154)(23)$ ⊠

We can now describe the conjugacy classes in $S(n)$.

Proposition 11.15. *Two elements of $S(n)$ are conjugate if and only if their disjoint cycle representations have the same number of cycles of each length.*

Proof. Suppose σ is a k-cycle, say $\sigma = (i\ \sigma(i) \cdots \sigma^{k-1}(i))$. Then $\psi = \tau\sigma\tau^{-1}$ is also a k-cycle. To see this, let $\tau^{-1}(j) = i$. Since $\psi^k = (\tau\sigma\tau^{-1})^k = \tau\sigma^k\tau^{-1}$, it follows that $\psi^k(j) = \tau\sigma^k(i) = \tau(i) = j$. If $m < k$, then $\psi^m(j) \neq j$, so $(j\ \psi(j) \cdots \psi^{k-1}(j))$ is a k-cycle. But if ψ contains another cycle, then reversing this argument shows that σ also contains another cycle, which cannot happen by assumption. Thus ψ is a k-cycle. It follows that two elements in the same conjugacy class have the same cycle structure. For if $\sigma = c_1 \cdots c_k$, where the c_i are mutually disjoint cycles, then

$$\tau\sigma\tau^{-1} = (\tau c_1\tau^{-1})(\tau c_2\tau^{-1}) \cdots (\tau c_k\tau^{-1}),$$

and the cycles $(\tau c_i\tau^{-1})$ are also mutually disjoint. Conversely, if two permutations ψ and σ have the same disjoint cycle structure, then they are conjugate. To see this, let $\psi = c_1 \cdots c_s$ and $\sigma = d_1 \cdots d_s$ be disjoint cycle representations of ψ and σ, where c_r and d_r have the same length for each r. Write $c_r = (i\ \psi(i) \cdots \psi^m(i))$ and $d_r = (j\ \sigma(j) \cdots \sigma^m(j))$, and let $\tau_r \in S(n)$ be the product of the transpositions $(\sigma^k(i)\ \psi^k(j))$ for $k = 0, \ldots, m-1$. These transpositions commute pairwise, so the order in the product is irrelevant. We leave it to the reader to check that $\tau_r c_r \tau_r^{-1} = d_r$, while $\tau_j c_r \tau_j^{-1} = c_r$ for $j \neq r$. Consequently, if $\tau = \tau_1 \cdots \tau_s$, then $\tau\psi\tau^{-1} = \sigma$. Therefore, ψ and σ are conjugate. □

For example, (12)(34) and (13)(24) must be conjugate, since they are products of two disjoint 2-cycles. Clearly (23)(12)(34)(23) = (13)(24).

Remark. Therefore, each conjugacy class of $S(n)$ corresponds to a unique partition $n = a_1 + a_2 + \cdots + a_m$, where $a_1 \geq a_2 \geq \cdots \geq a_m > 0$, and conversely, every such partition of n determines a unique conjugacy class. Recall also that by Jordan canonical form, the conjugacy classes of nilpotent $n \times n$ matrices over $\mathbb{C}$ are also in one-to-one correspondence with the partitions of n. It follows that conjugacy classes of nilpotent matrices over $\mathbb{C}$ are in one-to-one correspondence with the conjugacy classes of $P(n)$, the group of $n \times n$ permutation matrices.

Exercises

Exercise 11.1.1. Let G be an arbitrary group. The conjugation action is the action of G on itself defined by $g \cdot h = ghg^{-1}$. Show that the conjugation action is indeed a group action.

Exercise 11.1.2. Show that if X is a G-set and $y = g \cdot x$, then $G_y = gG_xg^{-1}$.

Exercise 11.1.3. Does $g \cdot a = ag$ define an action of a group G on itself? If not, adjust the definition so as to get an action.

Exercise 11.1.4. Prove Proposition 11.1.

Exercise 11.1.5. Show that a subgroup H of a group G is normal in G if and only if H is a union of G-conjugacy classes.

Exercise 11.1.6. Show directly that $Z_G(a)$ is a subgroup of G.

Exercise 11.1.7. • Show that if $G/Z(G)$ is cyclic, then G is abelian (and hence $G = Z(G)$).

Exercise 11.1.8. Let $G = S(3)$. Write out all three G-conjugacy classes. Also, find all six centralizers.

Exercise 11.1.9. Recall the signature homomorphism $\text{sgn} : S(n) \to \{\pm 1\}$. Show that if $\sigma \in S(n)$ is a k-cycle, then $\text{sgn}(\sigma) = (-1)^{k-1}$.

Exercise 11.1.10. Write out the conjugacy classes for the dihedral groups $D(3)$ and $D(4)$. (Think geometrically.)

Exercise 11.1.11. Describe the centers of the dihedral groups $D(n)$ for $n =$ 3 and 4.

Exercise 11.1.12. Let p and q be distinct primes. Show that a group of order pq is abelian and hence cyclic if and only if its center is nontrivial. Show by example that there exist groups of order pq whose center is trivial.

Exercise 11.1.13. Generalize the fact that all groups of order 15 are cyclic by showing that if p and $p + 2$ are both prime, then every group of order $p(p + 2)$ is cyclic.

Exercise 11.1.14. Show that if $|G| = 65$, then G is cyclic.

Exercise 11.1.15. Suppose $|G| = 30$. Show that there are two elements of order 3 and four of order 5. How many elements of order 2 are there?

Exercise 11.1.16. • Let G be a finite group and H a subgroup. Find a formula for the order of the double coset $|HaH|$ if $a \notin H$.

Exercise 11.1.17. Let $\sigma = (1632457) \in S(7)$.
(*i*) Find the disjoint cycle representation of σ.
(*ii*) What is the partition of 7 corresponding to σ?
(*iii*) Determine the sign $\text{sgn}(\sigma)$ of σ.
(*iv*) Finally, express σ as the product of simple transpositions.

11.2 The Finite Subgroups of $SO(3,\mathbb{R})$

Before continuing our treatment of group theory, we will pause to give a geometric application of the orbit method. Our plan is to determine the orders of the polyhedral groups, that is, the finite subgroups of $SO(3,\mathbb{R})$, and give a partial classification of these groups that will extend the result that every finite subgroup of $O(2,\mathbb{R})$ is either a cyclic group C_m consisting of m rotations or a dihedral group $\mathcal{D}(m)$ consisting of m rotations and m reflections. In both cases, C_m and $\mathcal{D}(m)$ act on the m-sided regular polygon $\{m\}$ centered at the origin in $\mathbb{R}^2$. The complete classification of the polyhedral groups is a natural extension to the Platonic solids of the two-dimensional classification, but the proof is much more complicated, and we will skip some of the details. It turns out that the list of finite subgroups of $SO(3,\mathbb{R})$ is brief and easy to state. Here is the result.

Theorem 11.16. *Every polyhedral group is isomorphic to one of C_m, $D(m)$, $A(4)$, $S(4)$, $A(5)$, and each of these groups is the rotation group of a convex polyhedron in $\mathbb{R}^3$.*

Polyhedral groups of orders 12, 24, and 60 have already been discussed in Section 8.7.2. They occur as the rotation groups of the Platonic solids (see Section 8.7).

11.2.1 The order of a finite subgroup of $SO(3,\mathbb{R})$

Let G be polyhedral. Our plan is to apply the orbit stabilizer theorem to count the number of elements in G in terms of the set of what are known as its poles. We first introduce some notation. Let $G^\times = G \setminus \{I_3\}$, and let $S^2 = \{(x,y,z) \in \mathbb{R}^3 \mid x^2+y^2+z^2 = 1\}$ be the unit sphere in $\mathbb{R}^3$. A point $\mathbf{p}$ of S^2 is called a *pole of* G if the stabilizer $G_\mathbf{p}$ satisfies $G_\mathbf{p} \neq \{I_3\}$. Hence each pole is on the axis of rotation of some $\sigma \in G^\times$. Since G consists of linear mappings, $G_\mathbf{p} = G_{-\mathbf{p}}$ for every $\mathbf{p} \in S^2$. Let $\mathcal{P}$ denote the set of all poles of G, and note that $\mathcal{P} = -\mathcal{P}$. Hence we can write $\mathcal{P} = \mathcal{P}^+ \cup \mathcal{P}^-$, where $\mathcal{P}^- = -\mathcal{P}^+$, and the union is disjoint. Since each $\sigma \in G^\times$ fixes a unique doubleton, and if $\mathbf{p},\mathbf{q} \in S^2$ satisfy $\mathbf{q} \neq \pm\mathbf{p}$, then $G_\mathbf{q} \cap G_\mathbf{p}$ is empty, and it follows that $G^\times$ can be expressed as a disjoint union

$$G^\times = \bigcup_{\mathbf{p}\in\mathcal{P}^+} (G_\mathbf{p})^\times.$$

This leads to the observation that

$$|G^\times| = \sum_{\mathbf{p}\in\mathcal{P}^+} |(G_\mathbf{p})^\times|.$$

Consequently,

$$|G| - 1 = \sum_{\mathbf{p} \in \mathcal{P}^+} (|G_{\mathbf{p}}| - 1).$$

It is now convenient to reformulate this identity. First of all,

$$|G| - 1 = \frac{1}{2} \sum_{\mathbf{p} \in \mathcal{P}} (|G_{\mathbf{p}}| - 1). \tag{11.3}$$

We leave it to the reader to check that $\mathcal{P}$ is stable under G. Let us call the G-orbit $G \cdot \mathbf{p}$ of $\mathbf{p} \in \mathcal{P}$ a *polar orbit*. Since G is finite, we can choose distinct polar orbits $G \cdot \mathbf{p}_1, G \cdot \mathbf{p}_2, \ldots, G \cdot \mathbf{p}_k$ such that

$$\mathcal{P} = G \cdot \mathbf{p}_1 \cup G \cdot \mathbf{p}_2 \cup \cdots \cup G \cdot \mathbf{p}_k.$$

We showed that $|G_{\mathbf{p}}| = |G_{\mathbf{q}}|$ if $\mathbf{p}$ and $\mathbf{q}$ are in the same orbit. Moreover, $|G| = |G \cdot \mathbf{p}||G_{\mathbf{p}}|$. Thus,

$$\begin{aligned} 2(|G| - 1) &= \sum_{\mathbf{p} \in \mathcal{P}} (|G_{\mathbf{p}}| - 1) \\ &= \sum_{i=1}^{k} \Big(\sum_{\mathbf{p} \in G \cdot \mathbf{p}_i} (|G_{\mathbf{p}}| - 1) \Big) \\ &= \sum_{i=1}^{k} |G \cdot \mathbf{p}_i| (|G_{\mathbf{p}_i}| - 1) \\ &= \sum_{i=1}^{k} (|G| - |G \cdot \mathbf{p}_i|) \\ &= |G| \sum_{i=1}^{k} (1 - \frac{1}{|G_{\mathbf{p}_i}|}). \end{aligned}$$

Therefore, we have proved the following proposition.

Proposition 11.17. *Suppose G is a polyhedral group having polar orbits $G \cdot \mathbf{p}_1, G \cdot \mathbf{p}_2, \ldots, G \cdot \mathbf{p}_k$. Then*

$$2(1 - \frac{1}{|G|}) = \sum_{i=1}^{k} (1 - \frac{1}{|G_{\mathbf{p}_i}|}). \tag{11.4}$$

The upshot of this elegant identity is that $|G|$ depends only on the orders $|G_{\mathbf{p}_i}|$ of the stabilizers of its poles.

11.2.2 The order of a stabilizer $G_{\mathbf{p}}$

Let us now apply (11.4) to investigate the possible values of $|G|$. First of all, if $\mathbf{p}$ is a pole, then $|G_{\mathbf{p}}| \geq 2$, so

$$\frac{1}{2} \leq 1 - \frac{1}{|G_{\mathbf{p}}|} \leq 1.$$

But

$$2(1 - \frac{1}{|G|}) < 2,$$

and hence there cannot be more than three poles. Thus, $k = 1$, 2, or 3. So we need to consider these three cases.

(1) Suppose $k = 1$. Then G acts transitively on the poles. Let $G \cdot \mathbf{p}$ be the unique orbit. Thus $|G| = |G \cdot \mathbf{p}||G_{\mathbf{p}}|$, so using the formula for $|G|$ gives

$$2(|G \cdot \mathbf{p}||G_{\mathbf{p}}| - 1) = |G \cdot \mathbf{p}|(|G_{\mathbf{p}}| - 1).$$

Simplifying gives $|G \cdot \mathbf{p}|(|G_{\mathbf{p}}| + 1) = 2$. This is impossible, since $|G_{\mathbf{p}}| + 1 \geq 3$, so G cannot act transitively on the poles.

(2) Next, let $k = 2$ and let $G \cdot \mathbf{p}$ and $G \cdot \mathbf{q}$ be the two orbits. Then

$$2(1 - \frac{1}{|G|}) = 2 - \frac{1}{|G_{\mathbf{p}}|} - \frac{1}{|G_{\mathbf{q}}|}.$$

Canceling and cross multiplying by $-|G|$ gives us that

$$2 = \frac{|G|}{|G_{\mathbf{p}}|} + \frac{|G|}{|G_{\mathbf{q}}|} = m + n,$$

where m and n are positive integers by Lagrange. Thus $m = n = 1$ and $G = G_{\mathbf{p}} = G_{\mathbf{q}}$. In other words, every element of G fixes both $\mathbf{p}$ and $\mathbf{q}$. Since every rotation has an axis, this means that G has just one axis, and $\mathbf{q} = -\mathbf{p}$. Thus G consists of rotations about the line $\ell = \mathbb{R}\mathbf{p}$. In other words, G is determined by a finite subgroup of rotations of the plane orthogonal to ℓ. In this case, G is cyclic.

(3) Now suppose G has three polar orbits $G \cdot \mathbf{p}$, $G \cdot \mathbf{q}$, and $G \cdot \mathbf{r}$. Then the order formula gives

$$2(1 - \frac{1}{|G|}) = 3 - \frac{1}{|G_{\mathbf{p}}|} - \frac{1}{|G_{\mathbf{q}}|} - \frac{1}{|G_{\mathbf{r}}|}.$$

Consequently,

$$1 + \frac{2}{|G|} = \frac{1}{|G_{\mathbf{p}}|} + \frac{1}{|G_{\mathbf{q}}|} + \frac{1}{|G_{\mathbf{r}}|} > 1.$$

This implies that the triple $(|G_{\mathbf{p}}|^{-1}, |G_{\mathbf{q}}|^{-1}, |G_{\mathbf{r}}|^{-1})$ has to be one of the following:

$$(\frac{1}{2}, \frac{1}{2}, \frac{1}{m}), \quad (\frac{1}{2}, \frac{1}{3}, \frac{1}{3}), \quad (\frac{1}{2}, \frac{1}{3}, \frac{1}{4}), \quad (\frac{1}{2}, \frac{1}{3}, \frac{1}{5}).$$

The order formula (11.4) says that $|G|$ is respectively $2m$, 12, 24, or 60.

Let us analyze each case. Assume first that $(|G_{\mathbf{p}}|^{-1}, |G_{\mathbf{q}}|^{-1}, |G_{\mathbf{r}}|^{-1}) = (\frac{1}{2}, \frac{1}{2}, \frac{1}{m})$. If $m = 1$, then $|G| = 2$, so there can only be two poles. Hence, this case cannot occur. Suppose $m = 2$. Then G has order four. Thus G is abelian, and since G has three polar orbits, it cannot be cyclic. Hence every element σ of G except 1 has order two. Suppose $\sigma \neq 1$. Then σ is a rotation through π about its axis. Thus, σ is diagonalizable with orthonormal eigenvectors, say $\mathbf{u}, \mathbf{v}, \mathbf{w}$ and corresponding eigenvalues 1, -1, -1. It follows that every element of G is diagonalizable, and since G is abelian, all elements of G are simultaneously diagonalizable. Hence the only possibility is that G is conjugate in $SO(3, \mathbb{R})$ to the subgroup $H = \{I_3, \mathrm{diag}(1, -1, -1), \mathrm{diag}(-1, 1, -1), \mathrm{diag}(-1, -1, 1)\}$. Here, the polar orbits for G are $\{\pm\mathbf{u}\}$, $\{\pm\mathbf{v}\}$, and $\{\pm\mathbf{w}\}$. Notice that if α, β, γ denote the three rotations in H, then $\alpha^2 = \beta^2 = \gamma^2$ and $\alpha\beta = \gamma$. Thus H is isomorphic to the dihedral group $D(2)$.

Now suppose $m > 2$. Since $|G| = 2m$, $|G \cdot \mathbf{p}| = |G \cdot \mathbf{q}| = m$, while $|G \cdot \mathbf{r}| = 2$. But $G \cdot (-\mathbf{r}) = -G \cdot \mathbf{r}$, so $|G \cdot (-\mathbf{r})| = 2$ also. Since $m > 2$, the only possibility is that $G \cdot \mathbf{r} = G \cdot (-\mathbf{r}) = \{\pm\ \mathbf{r}\}$, so $\sigma(\mathbf{r}) = \pm\ \mathbf{r}$ for every $\sigma \in G$. Now let P denote the subspace $(\mathbb{R}\mathbf{r})^{\perp}$ of $\mathbb{R}^3$ orthogonal to $\mathbb{R}\mathbf{r}$. Then for every $\sigma \in G$ and $\mathbf{p} \in P$, we have

$$0 = \mathbf{r} \cdot \mathbf{p} = \sigma(\mathbf{r}) \cdot \sigma(\mathbf{p}) = \pm\ \mathbf{r} \cdot \sigma(\mathbf{p}).$$

Thus the whole group G acts on P. Now choose an orthonormal basis $\mathbf{u}_1, \mathbf{u}_2$ of P, and define a linear mapping $\varphi : P \to \mathbb{R}^2$ by requiring $\varphi(\mathbf{u}_i) = \mathbf{e}_i$ for $i = 1, 2$. Then φ is a isometry; that is, $\varphi(\mathbf{x} \cdot \mathbf{y}) = \mathbf{x} \cdot \mathbf{y}$ for all $\mathbf{x}, \mathbf{y} \in P$. Next, put $\tau(\sigma) = \varphi\sigma\varphi^{-1}$ for all $\sigma \in G$. Then $\tau(\sigma) \in O(2, \mathbb{R})$. To see this, put $\mu = \tau(\sigma)$. Then, as the reader should verify, $\mu(\mathbf{a}) \cdot \mu(\mathbf{b}) = \mathbf{a} \cdot \mathbf{b}$ for all $\mathbf{a}, \mathbf{b} \in \mathbb{R}^2$, so μ is an orthogonal linear mapping by Proposition 7.7. Consequently, we have defined a mapping $\tau : G \to O(2, \mathbb{R})$. The reader can check that τ is a homomorphism, so the image $\tau(G)$ is a finite subgroup of $O(2, \mathbb{R})$. By the definition of τ, it follows that the eigenvalues of σ on P are the eigenvalues of $\tau(\sigma)$ on $\mathbb{R}^2$. Thus $\tau(G)$ contains both rotations and reflections, so $\tau(G)$ is a dihedral subgroup of $O(2, \mathbb{R})$ consisting of all orthogonal linear mappings fixing a regular polygon $\{m\}$ in $\mathbb{R}^2$. But observe that in fact, τ is injective. For if $\tau(\sigma) = I_2$, then the eigenvalues of σ on P are both 1. But this means

that $\sigma(\mathbf{r}) = \mathbf{r}$ too; for otherwise, $\sigma(\mathbf{r}) = -\mathbf{r}$, so $\det(\sigma) = -1$, contradicting the assumption that $\sigma \in SO(3,\mathbb{R})$. Hence, if $\tau(\sigma) = I_2$, then $\sigma = I_3$, and consequently τ is injective, since its kernel is trivial. We conclude that G is isomorphic to the dihedral group $D(m)$ of order $2m$.

Next assume that G has the triple $(\frac{1}{2}, \frac{1}{3}, \frac{1}{3})$. Then the order of G is 12. If one studies a list of all groups of order 12, it turns out that the only possibility is that G is the rotation group of a central regular tetrahedron. The poles are the four vertices $\mathbf{v}$, the four midpoints $\mathbf{e}$ of the edges, and the four centers $\mathbf{c}$ of the faces, and G acts transitively on each class of pole. Thus the polar orbits have $|G \cdot \mathbf{v}| = 4$, $|G \cdot \mathbf{e}| = 6$, and $|G \cdot \mathbf{c}| = 4|$, which verifies that the orders of the stabilizers are $3, 2, 3$.

For $(\frac{1}{2}, \frac{1}{3}, \frac{1}{4})$, the order of G is 24. We have already shown that the rotation group of a central cube is isomorphic to $S(4)$. As in the previous case, the poles consist of the eight vertices $\mathbf{v}$, the four midpoints $\mathbf{e}$ of the twelve edges, and the centers $\mathbf{c}$ of the six faces, and G acts transitively on the set of poles of each type. It can be shown that this realization of $S(4)$ is the only polyhedral group of order 24, but the proof is complicated.

The final case $(\frac{1}{2}, \frac{1}{3}, \frac{1}{5})$ gives a group G of order 60. Recall from Remark 8.7.5 that an icosahedron is a regular solid with 12 vertices, 30 edges, and 20 triangular faces. Thus, the rotation group of a central icosahedron has order 60, which proves the existence of a polyhedral group of order 60. One may take the vertices of $\mathcal{I}$ to be the points

$$(\pm 1, \pm\phi, 0), (0, \pm 1, \pm\phi), (\pm\phi, 0, \pm 1),$$

where $\phi = \frac{(1+\sqrt{5})}{2}$ is the golden ratio encountered in Chap. 8. At each of the vertices, there are five adjoining triangular faces. Therefore, for each vertex $\mathbf{v}$, $|G_{\mathbf{v}}| = 5$. The centers of the faces $\mathbf{m}$ are also poles, and clearly $|G_{\mathbf{m}}| = 3$. The other poles consist of the midpoints $\mathbf{e}$ of the edges with $|G_{\mathbf{e}}| = 2$.

There is an amusing way of showing that G is isomorphic to $A(5)$. It turns out that one can color the 20 faces of $\mathcal{I}$ with five colors so that no two adjoining faces have the same color and each color is used four times. If the colors are labeled 1,2,3,4,5, then G, which permutes $\mathcal{I}$'s 20 faces, becomes the set of all permutations of $\{1, 2, 3, 4, 5\}$ with signature $+1$. In other words, $G \cong A(5)$.

11.3 The Sylow Theorems

Given a finite group G, what can one say about its subgroups? Both Lagrange's theorem and Cauchy's theorem make assertions about the subgroups of G in terms of the order of G, which is the most basic invariant of a finite group. The Sylow theorems, which were discovered by Sylow in 1872 during his research on algebraic equations, give much more precise information. Suppose $|G| = p^n m$, where p is a prime that is prime to m and $n \geq 1$. A subgroup of G of order p^n is called a *Sylow p-subgroup*. Sylow's theorems assert the following: if $1 \leq k \leq n$, there exists a subgroup H of order p^k, and every such H is contained in a Sylow p-subgroup; all Sylow p-subgroups for a given prime p are conjugate in G; and finally, the number of Sylow p-subgroups of G is congruent to one modulo p and divides $|G|$. The proofs given below are elegant and brief, and above all, they explain why the theorems are true. The tools used are Cauchy's theorem, the orbit stabilizer theorem, and the fixed-point result for p-groups in Proposition 11.11.

11.3.1 The first Sylow theorem

As above, let G be a finite group and p a prime p dividing $|G|$, say $|G| = p^n m$, where $(p, m) = 1$. We now prove the first Sylow theorem, which ensures that Sylow p-subgroups always exist and a bit more.

Theorem 11.18. *For each k such that $1 \leq k \leq n$, G contains a subgroup of order p^k, and every subgroup of order p^k is contained in a Sylow p-subgroup of G.*

Proof. Suppose H is a subgroup of G of order p^k, where $1 \leq k < n$. We will show that H is contained in a subgroup K of G of order p^{k+1}. Iterating this argument will yield a subgroup of G of order p^n that contains H, giving the required Sylow p-subgroup. The existence of a subgroup of order p^k, for each $k \leq n$, then follows by applying this result to the subgroup generated by an element of order p, whose existence is guaranteed by Cauchy's theorem. Now, H acts on $X = G/H$ by left translation. Moreover, H is a p-group, and $|X|$ is divisible by p. Therefore, Proposition 11.11 tells us that either the fixed-point set X^H is empty or $|X^H| = jp$ for some $j > 0$. But $1H \in X^H$, so the latter holds. Note that $gH \in (G/H)^H$ if and only if $gH \in N_G(H)/H$, since $HgH = gH$ if and only if $gHg^{-1} = H$. Thus, $N_G(H)/H$ is a group whose order is divisible by p. By Cauchy's theorem, $N_G(H)/H$ contains a subgroup K of order p. Consider the inverse image $H' = \pi^{-1}(K)$ of the quotient map $\pi : N_G(H) \to N_G(H)/H$. Then H' is a subgroup of $N_G(H)$ containing H, which π maps to K. Thus, by the isomorphism theorem, Theorem 2.20, $H'/H \cong K$, so $[H' : H] = |K| = p$. But Lagrange's theorem implies that $|H'| = p^{k+1}$, so the proof is complete. □

11.3.2 The second Sylow theorem

The second Sylow theorem asserts that for a given prime p, every pair of Sylow p-subgroups of G are conjugate. This says that a finite group G is rigid in the sense that all subgroups of G of certain orders are isomorphic by an inner automorphism of G.

Theorem 11.19. *Every pair of Sylow p-subgroups of a finite group G are conjugate, and if p divides $|G|$, then the number of Sylow p-subgroups of G also divides $|G|$.*

Proof. Let H and K be Sylow p-subgroups of G, and let H act on G/K. It suffices to show that the fixed-point set $(G/K)^H$ is nonempty, say $HgK = gK$. For if $HgK = gK$, it follows that $gKg^{-1} = H$. Suppose $(G/K)^H$ is empty. Then the orbit stabilizer theorem implies that for every H-orbit $\mathcal{O}$, $|\mathcal{O}|$ is a multiple of p, since H is a p-group. This implies that p divides $|G/K|$, which contradicts the assumption that $|K| = p^n$. Consequently, every two Sylow p-subgroups are conjugate. Thus G acts transitively via conjugation on the set of its Sylow p-subgroups, so by another application of the orbit stabilizer theorem, the number of Sylow p-subgroups of G divides $|G|$. □

The fact that Sylow p-subgroups are conjugate gives the following result.

Corollary 11.20. *If G is a finite abelian group, then G contains a unique Sylow p-subgroup for every prime p dividing $|G|$.*

11.3.3 The third Sylow theorem

The third Sylow theorem is analogous to the result that in a finite group, the number of elements of order p is divisible by $p - 1$.

Theorem 11.21. *If G is a finite group and p is a prime dividing $|G|$, then the number of Sylow p-subgroups of G is congruent to 1 modulo p and divides $|G|$.*

Proof. Assume that the prime p divides $|G|$, and let $\mathcal{S}_p$ denote the set of Sylow p-subgroups of G. Fix $S \in \mathcal{S}_p$, and let N_S denote the normalizer $N_G(S)$. I claim that the only Sylow p-subgroup of N_S is S. For if $R \in \mathcal{S}_p$ satisfies $R < N_S$, then there exists $g \in N_S$ such that $gSg^{-1} = R$. But since $gSg^{-1} = S$ by the definition, it follows that $R = S$. Thus, if $R, S \in \mathcal{S}_p$ and $R \neq S$, then $|N_R \cap N_S|$ cannot be divisible by $|S|$. Now let N_S act on $\mathcal{S}_p$. Certainly, S is a fixed point; i.e. $N_S \cdot S = S$. On the other hand, if $R \neq S$, then $|N_S \cdot R|$ is divisible by p. Indeed, by the orbit stabilizer theorem, $|N_S \cdot R| = |N_S|/|N_R \cap N_S|$, so this follows from the fact that the denominator is not divisible by $|S|$. Therefore, the orbit identity implies that $|\mathcal{S}_p| = 1 + mp$. □

11.3.4 Groups of order 12, 15, and 24

We will now look at some examples.

Example 11.8. First consider groups of order 12. In such a group, the number of Sylow 2-subgroups is either 1 or 3, and the number of Sylow 3-subgroups is either 1 or 4. The direct products $C_3 \times C_4 = C_{12}$ and $C_3 \times U_8$, where U_8 is the group of units modulo 8, are both abelian groups of order 12; hence each has exactly one Sylow subgroup of order 3 and one of order 4. Note: the direct product of G and H is the group $G \times H$ consisting of all the pairs (g, h) with $g \in G$ and $H \in H$ with the almost obvious group structure. The general definition is given in Definition 11.6 below. The alternating group $A(4)$ has four cyclic subgroups of order 3 generated respectively by (123), (134), (124), and (234). Thus there are eight elements of order 3. This leaves exactly three elements of order 2 or 4. Thus the Sylow 2-subgroup, which has order 4, has to be normal. This group turns out to be the well-known Klein 4-group

$$V_4 = \{(1), (12)(34), (14)(23), (13)(24)\},$$

consisting of all elements of $S(4)$ that are the product of two disjoint transpositions. In fact, V_4 is normal in $S(4)$. The dihedral group $D(6)$ has six rotations, which form a cyclic normal subgroup. The two rotations of order 3 determine a normal subgroup, which is therefore the unique Sylow 3-subgroup of $D(6)$. By the $O(2, \mathbb{R})$ dichotomy, the remaining six elements are reflections. Since reflections about orthogonal lines commute, and since a six-sided regular polygon P in $\mathbb{R}^2$ has three pairs of orthogonal lines through which P may be reflected, there are three Sylow 2-subgroups. Thus $D(6)$ has one Sylow 3-subgroup and three Sylow 2-subgroups. ⊠

Example 11.9. Returning to the group of order 15, we see that the number of Sylow 3-subgroups has the form $1+3m$ and divides 15. The only possibility is that there is exactly one Sylow 3-subgroup. Similarly, the number of Sylow 5-subgroups has the form $1 + 5m$ and divides 15. Hence there is exactly one Sylow 5-subgroup. The order of an element has to divide 15, so the possible orders are 3, 5, and 15. Since we just showed that there are two elements of order 3 and four elements of order 5, there has to be an element of order 15 (in fact, eight of them). Thus we see once again that a group of order 15 is cyclic. ⊠

Example 11.10. The symmetric group $S(4)$ and the matrix group $SL(2, \mathbb{F}_3)$ both have order 24. Are they isomorphic? A necessary condition for two finite groups to be isomorphic is that they have the same Sylow subgroups. By the third Sylow theorem, the number of Sylow 3-subgroups in a group of order 24 is one or four, and the number of Sylow 2-subgroups is one or three. We will now show that $S(4)$ has three Sylow 2-subgroups. Recall that $S(4)$ is

isomorphic to the group of all rotations of the unit cube $\mathcal{C}$ in $\mathbb{R}^3$ centered at the origin. Let S_1 denote the intersection of the cube with the xy-plane, and label the vertices of S as 1, 2, 3, 4 in a clockwise manner, with 1 denoting the vertex in the first octant. The subgroup

$$H_1 = \{(1), (1234), (1432), (13)(24), (12)(34), (14)(23), (13), (24)\}$$

is a copy of $D(4)$. The first four elements are rotations of S_1, which rotate $\mathcal{C}$ about the z-axis. The other four elements are reflections of S_1, which are rotations of the cube through $\pi/2$. The other two Sylow 2-subgroups are copies of $D(4)$, which act the same way on the squares S_2 and S_3 obtained by intersecting the cube with the other two coordinate hyperplanes. It turns out that $SL(2, \mathbb{F}_3)$ has only one Sylow subgroup H of order 8, so H is normal. The details are tedious, so we will omit them. For example, the elements of order 4 are

$$\begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}, \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 2 \\ 2 & 2 \end{pmatrix}, \begin{pmatrix} 2 & 2 \\ 1 & 2 \end{pmatrix}.$$

Thus, $S(4)$ and $SL(2, \mathbb{F}_3)$ have different Sylow 2-subgroups, so they cannot be isomorphic. ⊠

Exercises

Exercise 11.3.1. Show that $A(4)$ does not have a subgroup of order 6; hence the converse of Lagrange's theorem is false.

Exercise 11.3.2. Determine the Sylow subgroups for the dihedral group $D(6)$.

Exercise 11.3.3. Let p and q be distinct primes, and let G be a group of order pq. Show that the following statements are equivalent:

(*i*) G is abelian.

(*ii*) The center $Z(G)$ is nontrivial.

(*iii*) G is cyclic.

Give an example of a group of order pq that is not cyclic.

Exercise 11.3.4. True or false with reasoning: every group of order 10 is abelian.

Exercise 11.3.5. Show that if p and $p+2$ are both prime, then every group of order $p(p+2)$ is cyclic.

Exercise 11.3.6. Let p be a prime. Show that $\mathcal{U}(n, \mathbb{F}_p)$ and $\mathcal{L}(n, \mathbb{F}_p)$ are Sylow p-subgroups of both $GL(n, \mathbb{F}_p)$ and $SL(n, \mathbb{F}_p)$. How is $\mathcal{U}(n, \mathbb{F}_p)$ conjugated onto $\mathcal{L}(n, \mathbb{F}_p)$?

Exercise 11.3.7. Suppose $p = 5$ and $n = 3$.

(i) How many Sylow p-subgroups does $GL(n, \mathbb{F}_p)$ have?

(ii) Answer part (i) for $SL(n, \mathbb{F}_p)$.

Exercise 11.3.8. Find the orders of all the Sylow p-subgroups of $GL(n, \mathbb{F}_p)$ for primes $p \leq 7$.

Exercise 11.3.9. Consider the group U_n of units modulo n for $n = 50$.

(i) Discuss the Sylow subgroups of U_n.

(ii) Determine whether U_n is cyclic. (Suggestion: consider part (i).)

Exercise 11.3.10. The order of U_n is 100 if $n = 202$. Is either of the Sylow subgroups cyclic?

Exercise 11.3.11. • Let G be a finite group. Show that if $|G| = mp$, where p is prime and $p > m$, then there is only one Sylow p-subgroup H, and H is normal in G.

Exercise 11.3.12. • Show that a group of order pq, p and q distinct primes, is cyclic whenever p does not divide $q - 1$ and q does not divide $p - 1$.

11.4 The Structure of Finite Abelian Groups

So far, we have encountered a few examples of finite abelian groups: the cyclic groups C_m, the multiplicative group $\mathbb{F}^*$ of a Galois field $\mathbb{F}$, the center $Z(G)$ of an arbitrary finite group G, and the subgroups of all these groups. More interestingly, as we saw in Section 2.1, the group of multiplicative units (i.e., invertible elements) in $\mathbb{Z}/n\mathbb{Z}$ is an abelian group U_n of order $\phi(n)$, where ϕ is Euler's phi-function. We will now describe the structure of all finite abelian groups. The Sylow theorems imply a finite abelian group is the direct product of its Sylow subgroups. After that, the job is to describe the abelian p-groups.

11.4.1 Direct products

The direct product of a finite number of groups is a generalization of the direct sum of a finite number of vector spaces. And as for vector spaces, direct products in the group setting come in two flavors, internal and external. The simpler case is the external direct product, so we will introduce it first.

Definition 11.6. Let G_1 and G_2 be arbitrary groups. Their *external direct product* is the usual Cartesian product $G_1 \times G_2$ with the group operation

$$(g_1, h_1)(g_2, h_2) = (g_1 g_2, h_1 h_2).$$

It is easy to check that the external direct product $G_1 \times G_2$ is a group with identity $(1, 1)$, where 1 denotes both the identity of G_1 and that of G_2. The notion of external direct product can be extended without difficulty from two groups to any finite number of groups. We will leave the details to the reader. Note that if V_1 and V_2 are vector spaces over different fields, then $V_1 \times V_2$ is an abelian group but not a vector space. If V_1 and V_2 have the same scalar field $\mathbb{F}$, then the external direct sum of vector spaces applies, and $V_1 \times V_2$ is a vector space over $\mathbb{F}$.

Proposition 11.22. *Suppose $G = G_1 \times G_2 \times \cdots \times G_m$ is the external direct product of the groups $G_1, G_2, \ldots, G_m$. Then:*

(i) G is abelian if each G_i is, and

(ii) for every $\sigma \in S(m)$, $G \cong G_{\sigma(1)} \times G_{\sigma(2)} \times \cdots \times G_{\sigma(m)}$.

Proof. We leave the proof to the reader. □

We will now consider the internal direct product, which for groups is a generalization of the internal direct sum of subspaces (keeping in mind that subspaces of a vector space are normal subgroups). Let $A_1, A_2, \ldots, A_k$ be subsets of the group G. Define $A_1 A_2 \cdots A_k$ to be the totality of products

$a_1a_2\cdots a_k$, where each a_i is in A_i. Also, define $< A_1, A_2, \ldots, A_k >$ to be the smallest subgroup of G containing each A_i. Suppose each A_i is normalized by G; that is, $gA_ig^{-1} = A_i$ for all $g \in G$. Then $A_iA_j = A_jA_i$ for all i, j. For if we choose any $a_i \in A_i$ for each i, it follows that $a_ia_j = a_j(a_j^{-1}a_ia_j)$, so $A_iA_j \subset A_jA_i$. Hence by symmetry, $A_iA_j = A_jA_i$. Consequently, the product $A_1A_2\cdots A_k$ is independent of how the sets A_i are ordered.

We now need the following lemma.

Lemma 11.23. *If $H_1, H_2, \ldots, H_k$ are normal subgroups of G, then the product $H_1H_2\cdots H_k$ is a normal subgroup of G, and*

$$< H_1, H_2, \ldots, H_k >= H_1H_2\cdots H_k.$$

Proof. We will use induction to prove that $H_1H_2\cdots H_k$ is a subgroup of G, which will show that $< H_1, H_2, \ldots, H_k >= H_1H_2\cdots H_k$, since $H_1H_2\cdots H_k \subset< H_1, H_2, \ldots, H_k >$. Notice that if $H_1H_2\cdots H_k$ is a subgroup, then it has to be normal, since for all $g \in G$,

$$g(H_1H_2\cdots H_k)g^{-1} = (gH_1g^{-1})(gH_2g^{-1})\cdots(gH_kg^{-1}),$$

and every H_i is a normal subgroup. The lemma holds for $k = 1$, so assume $k > 1$ and that it holds for $k - 1$. Thus $H = H_1H_2\cdots H_{k-1}$ is a subgroup of G. We need to show that HH_k is a subgroup, so we must show that if $a, b \in H$ and $g, h \in H_k$, then $(ag)(bh)^{-1} \in HH_k$. Now, $(ag)(bh)^{-1} = a(gh^{-1})b^{-1} \in HH_kH = HH_k$, since $HH_k = H_kH$ by the remark preceding the lemma. Hence $H_1H_2\cdots H_k$ is a subgroup, so the proof is complete. □

We now come to the internal direct product of an arbitrary number of normal subgroups.

Definition 11.7. Let G be a group and let $H_1, H_2, \ldots, H_k$ be normal subgroups of G. We say that G is the *internal direct product* of $H_1, H_2, \ldots, H_k$ if

$$G =< H_1, H_2, \ldots, H_k >,$$

and for each i with $1 \leq i < k$,

$$< H_1, H_2, \ldots, H_i > \cap H_{i+1} = (1).$$

If G is the internal direct product of $H_1, \ldots, H_k$, we will write $G = ((H_1, \ldots, H_k))$.

Proposition 11.24. *Let $G = ((H_1, H_2, \ldots, H_k))$. Then for each $g \in G$, there exists exactly one expression $g = h_1h_2\cdots h_m$, where $h_i \in H_i$ for each i. In particular, if G is finite, then*

$$|G| = \prod_{i=1}^{m} |H_i|. \tag{11.5}$$

Proof. An expression $g = h_1 h_2 \cdots h_m$ exists by Lemma 11.23. Assume $g = h_1 \cdots h_m = k_1 \cdots k_m$ with $h_i, k_i \in H_i$ for all i. Then $h_m k_m^{-1} \in < H_1, \ldots, H_{m-1} >$. Therefore, $h_m k_m^{-1} = 1$, so $h_1 \cdots h_{m-1} = k_1 \cdots k_{m-1}$. Continuing in this manner, it follows that $h_i = k_i$ for each i. This shows that the above expression for g is unique.

□

Proposition 11.25. *Suppose* $G = ((H_1, H_2, \ldots, H_k))$. *Then* $G \cong H_1 \times H_2 \times \cdots \times H_k$.

Proof. Define a map $\varphi : H_1 \times \cdots \times H_k \to G$ by $\varphi(h_1, \ldots, h_k) = h_1 \cdots h_k$. By Proposition 11.24, φ is a bijection. Since H_i and H_j commute elementwise for all i and j, it also follows that φ is also a homomorphism, so φ is an isomorphism. □

Example 11.11. Let's reconsider U_8 (see Example 2.18). For simplicity, we denote the coset $m + 8\mathbb{Z}$ by $\overline{m}$. Thus $U_8 = \{\overline{1}, \overline{3}, \overline{5}, \overline{7}\}$. Since U_8 is abelian, every subgroup is normal. Using coset multiplication, we have $\overline{3} \cdot \overline{5} = \overline{15} = \overline{7}$. Hence if $H_1 = \{\overline{1}, \overline{3}\}$ and $H_2 = \{\overline{1}, \overline{5}\}$, then $U_8 = ((H_1, H_2)) \cong H_1 \times H_2$. ⊠

Example 11.12. Let's next consider $U_{15} = \{\overline{1}, \overline{2}, \overline{4}, \overline{7}, \overline{8}, \overline{11}, \overline{13}, \overline{14}\}$. Since $|U_{15}| = 8$, the possible orders of its elements are 1, 2, 4, and 8. It is easy to check that $\overline{2}, \overline{7}, \overline{8}$, and $\overline{13}$ have order 4, and $\overline{4}$, $\overline{11}$, and $\overline{14}$ have order 2. Moreover, $\overline{2}^2 = \overline{7}^2 = \overline{8}^2 = \overline{13}^2 = \overline{4}$. Thus $< \overline{a} > \cap < \overline{11} > = < \overline{a} > \cap < \overline{14} > = (1)$ for $a = 2, 7, 8, 13$. Hence if $H = < \overline{a} >$, $K = < \overline{11} >$, and $L = < \overline{14} >$, we have

$$U_{15} = ((H, K)) \cong H \times K \cong ((H, L)) \cong H \times L = C_4 \times C_2.$$

⊠

11.4.2 The structure theorem for finite abelian groups

Suppose G is a finite abelian group of order $p_1^{n_1} \cdots p_m^{n_m}$, where $p_1, p_2, \ldots, p_m$ are the distinct prime factors of $|G|$, and let G_i denote the unique Sylow p_i-subgroup G. The first of two results about the structure of G is stated as follows.

Proposition 11.26. *Let* G *be as above. Then* G *is the direct product (both internal and external) of its Sylow subgroups* $G_1, \ldots, G_m$.

Proof. Since G is abelian, its Sylow subgroups are normal and commute elementwise. Let us induct on m. Assume that the product $G_1 \cdots G_i$ is direct for each $i < m$. By Proposition 11.24, $|G_1 \cdots G_i| = p_1^{n_1} \cdots p_i^{n_i}$. Thus $G_1 \cdots G_{m-1} \cap G_m = (1)$, since every element of G_m has order p_m^k for some k, and p_m does not divide $|G_1 \cdots G_{m-1}|$. Therefore, the product $G_1 \cdots G_m$ is direct, and consequently, $G = ((G_1, \ldots, G_m))$. It follows that $G \cong G_1 \times \cdots \times G_m$. □

It remains to give a complete description of the abelian p-groups in more detail. This turns out to be surprisingly complicated, so we will simply outline the proof. Here is the main result:

Theorem 11.27. *A finite abelian group of order p^n, p a prime, is isomorphic to a product of cyclic p-groups. In particular, every finite abelian group G is the external direct sum of cyclic groups of prime power order.*

Proof. The proof is by induction on n. Choose an element $y \in G$ of maximal order, and let $H =< y >$. By induction, G/H is a product of cyclic groups, say $G/H =< a_1 H > \times \cdots \times < a_k H >$. Now G/H is also a p-group, so for each i, $a_i^{p^{r_i}} = y^{s_i}$ for some $r_i, s_i > 0$. We assert that p^{r_i} divides s_i. Put $j_i = s_i/p^{r_i}$ and $u_i = a_i y^{-j_i}$. Then it turns out that $G =< u_1 > \times \cdots \times < u_k > \times H$, which gives the induction step and finishes the proof (modulo omitted details). □

Example 11.13. According to the theorem, if G is abelian of order 16, then the possibilities for G are

$$C_2 \times C_2 \times C_2 \times C_2, \quad C_4 \times C_2 \times C_2, \quad C_4 \times C_4, \quad C_8 \times C_2, \quad C_{16}.$$

11.4.3 The Chinese Remainder Theorem

Applying the above results gives the following corollary.

Corollary 11.28. *Suppose G is a finite cyclic group. Then its Sylow subgroups $G_1, \ldots, G_m$ are cyclic of prime power order, and $G = G_1 \times \cdots \times G_m$.*

Proof. Since G is cyclic, division with remainder tells us that every subgroup of G is also cyclic. Hence the Sylow subgroups of G are cyclic, so the result follows from Proposition 11.26. □

In particular, let $n = p_1^{n_1} \cdots p_m^{n_m}$, as usual. Since the additive group $\mathbb{Z}/n\mathbb{Z}$ is cyclic, its Sylow subgroups are $(\mathbb{Z}/p_1^{n_1}\mathbb{Z}), \ldots, (\mathbb{Z}/p_m^{n_m}\mathbb{Z})$. Therefore,

$$\mathbb{Z}/n\mathbb{Z} \cong (\mathbb{Z}/p_1^{n_1}\mathbb{Z}) \times \cdots \times (\mathbb{Z}/p_m^{n_m}\mathbb{Z}).$$

□

As a corollary, one gets the Chinese remainder theorem: given congruence equations

$$x \equiv a_1 \bmod(p_1^{n_1}), \ \ldots \ , x \equiv a_m \bmod(p_m^{n_m}),$$

there exists a unique class $w + n\mathbb{Z}$ in $\mathbb{Z}/n\mathbb{Z}$ such that setting $x = w$ solves all the above congruences.

Exercises

Exercise 11.4.1. Find at least three values of n such that U_n is cyclic. Is there a general rule?

Exercise 11.4.2. • Show that in every cyclic group, the number of elements of order m is divisible by $\phi(m)$. Give an example of a finite abelian group in which the number of elements of order m is not divisible by $\phi(m)$.

Exercise 11.4.3. Suppose each G_i is cyclic. When is $G = G_1 \times G_2 \times \cdots \times G_m$ cyclic?

Exercise 11.4.4. Suppose G can be expressed as a direct product HK. Show that

(i) $K \cong G/H$, and

(ii) $G \cong H \times G/H$.

Exercise 11.4.5. Find all the ways of decomposing U_{21} as a product of cyclic subgroups.

Exercise 11.4.6. Let $\mathbb{F}$ be a Galois field. It is a theorem that the multiplicative group $\mathbb{F}^*$ of $\mathbb{F}$ is always cyclic. Prove directly that $\mathbb{F}^*$ is cyclic in the following cases:

(*i*) $|\mathbb{F}| = 16$,

(*ii*) $|\mathbb{F}| = 32$, and

(*ii*) $|\mathbb{F}| = 25$.

Exercise 11.4.7. Let G be a group, H a subgroup, and N a normal subgroup of G. Then G is said to be the *semidirect product of H and N* if at least one of the following statements holds.

(*i*) $G = NH$ and $N \cap H = 1$.

(*ii*) Every element of G can be written in a unique way as a product nh, where $n \in N$ and $h \in H$.

(*iii*) The natural inclusion homomorphism $H \to G$ composed with the natural projection $G \to G/N$ induces an isomorphism $H \cong G/N$.

(*iv*) There exists a homomorphism $G \to H$ that is the identity on H whose kernel is N.

Show that the all four statements are equivalent.

11.5 Solvable Groups and Simple Groups

One of the most celebrated results in classical algebra is the unsolvability of the quintic. Roughly, this means that there exist polynomials $f(x)$ of degree five with rational coefficients whose roots cannot be expressed in terms of radicals involving those coefficients. More precisely, there is no formula that expresses the roots of an arbitrary fifth-degree polynomial $f(x) = x^5 + a_1x^4 + a_2x^3 + a_3x^2 + a_4x + a_5$ starting from an algebraic combination α of the coefficients a_i of f, takes an mth root $\alpha^{1/m}$ for some m, then repeats the process on the coefficients of $\alpha^{1/m}$ and so on until a closed formula that gives all the roots of $f(x)$ is obtained. Unsolvable quintics can be quite ordinary: $2x^5 + 10x + 5$ is an example. The definitive criterion for solvability by radicals first appeared in Galois's posthumous 1846 paper: a polynomial is solvable by radicals if and only if its Galois group is solvable. (For a definition of the Galois group of a polynomial, see Section 11.5.5.) The Galois group of $2x^5 + 10x + 5$ turns out to be $S(5)$, which, as we will see below, is not a solvable group. The simple groups lie at the opposite end of the group spectrum from the solvable groups, and in fact, they serve as the main source of examples of unsolvable groups. Recall that a nontrivial group G is said to be *simple* if its only normal subgroups are G itself and the trivial subgroup (1). A famous classical result is that the alternating groups $A(n)$ for $n > 4$ are simple. It follows that the symmetric groups $S(n)$ for $n > 4$ are not solvable. Simple groups play a role in the structure theory because of composition series and the Jordan–Hölder theorem, which we will state without proof below. As mentioned in the introduction to this chapter, the finite simple groups have now been classified, the final step not having come until 1982 with the construction of the Monster.

11.5.1 The definition of a solvable group

The process described above for solving by radicals hints at what the definition of a solvable group might be. To define solvability, we must first introduce the notion of a subnormal series in a group.

Definition 11.8. A *subnormal series for a group* G is a finite sequence of subgroups

$$G = H_1 > H_2 > \cdots > H_{r-1} > H_r = (1) \tag{11.6}$$

of G such that H_{i+1} is normal in H_i for $i = 1, \ldots, r-1$. The quotient groups H_i/H_{i+1} are called the *factors* of the subnormal series (11.6).

For example, if $H \neq (1)$ is a proper normal subgroup of G, then $G > H > (1)$ is an example of a subnormal series. A simple group G has no nontrivial

normal subgroups, so $G > (1)$ is the only subnormal series for G. A subnormal series (11.6) is called a *composition series* if each factor is a simple group. If H is normal in G, then G/H is simple if and only if there exists no normal subgroup N of G such that $H < N < G$. (The proof of this assertion is an exercise.) Thus a composition series is a subnormal series of maximal length. It follows that every finite group has a composition series.

Now we can define what a solvable group is.

Definition 11.9. A group G is called *solvable* if it has a subnormal series (11.6) such that each factor H_i/H_{i+1} is abelian.

Of course, abelian groups are solvable. The next result points out an important class of solvable groups.

Proposition 11.29. *Every p-group is solvable.*

Proof. Suppose $|G| = p^n$. If $n = 1$, then surely G is solvable, so let us argue by induction on n. Assume that all p-groups of order p^{n-1} are solvable. By the first Sylow theorem, G has a subgroup H of order p^{n-1}, and by assumption, H is solvable. To show that G is solvable, it suffices to show that H is normal in G, since in that case, G/H has order p and hence is abelian. Let H act on G/H by left multiplication. The orbit stabilizer theorem then says that for every orbit $\mathcal{O}$, either $|\mathcal{O}| = 1$ or $|\mathcal{O}| = p$. But the latter case cannot occur, since the orbit of the coset H is itself. It follows that every orbit is a single point. Thus $HgH = gH$ for all $g \in G$. This implies that H is normal in G. □

A famous result in finite group theory known as Burnside's theorem (1904) states that a finite group whose order is divisible by at most two primes is solvable. Burnside's proof used techniques from outside group theory, so it remained an open problem to find a purely group-theoretic proof. Such a proof was not discovered until 1970. In 1963, Walter Feit and John Thompson published a 255-page paper proving that every finite group of odd order is solvable, hence ensuring that there are nonabelian solvable groups that are not p-groups. Note that the Feit–Thompson theorem implies Burnside's result for two odd primes.

The symmetric groups $S(3)$ and $S(4)$ are both solvable. (For example, apply Burnside's theorem.) We will leave this claim as an exercise for $S(3)$ and give a proof for $S(4)$ in the next example. It turns out, however, that $S(n)$ is not solvable for $n > 4$.

Example 11.14 ($S(4)$ is solvable)**.** We need to display a subnormal series for $S(4)$ whose derived factors are abelian. The alternating group $A(4)$ (see Section 5.2) is normal in $S(4)$ and $S(4)/A(4) \cong C_2$, so it is abelian. Thus, $S(4) > A(4)$ is a first step. As we have seen, the Klein 4-group V_4 (see Example 11.8) is a normal subgroup of $A(4)$ of order 4. Since the order of $A(4)$ is 12, $A(4)/V_4$ has order three, and hence it is also abelian. But V_4 is abelian, since it has order p^2. It follows that $S(4) > A(4) > V_4 > (1)$ is a subnormal series for $S(4)$ whose factors are abelian. Hence $S(4)$ is solvable. ⊠

11.5.2 The commutator subgroup

In this section, we will find an explicit test for solvability. To do so, we need to introduce the commutator subgroup of a group and its derived series. The *commutator subgroup of* G is the smallest subgroup of G that contains all products of the form $ghg^{-1}h^{-1}$, where $g, h \in G$. This subgroup is denoted by $[G, G]$.

Proposition 11.30. *The commutator subgroup* $[G, G]$ *of* G *is normal in* G*, and* $G/[G, G]$ *is abelian. In particular, if* G *is simple, then* $G = [G, G]$*. Moreover, if* N *is a normal subgroup of* G *such that* G/N *is abelian, then* N *contains* $[G, G]$*.*

Proof. We will leave the proof as an exercise. □

The commutator groups for $S(n)$ provide a nice example. Recall that $A(n) = \ker(\text{sgn})$, where $\text{sgn} : S(n) \to \{\pm 1\}$ is the signature homomorphism defined by the expression

$$\text{sgn}(\sigma) = \prod_{i<j} \frac{\sigma(i) - \sigma(j)}{i - j}.$$

It follows from the fact that *sgn* is a homomorphism that $[S(n), S(n)] \leq A(n)$.

Proposition 11.31. *The commutator subgroup of* $S(n)$ *is* $A(n)$ *for all* n*.*

Proof. We just showed that $[S(n), S(n)] \leq A(n)$. Notice that all 3-cycles in $S(n)$ are commutators, since

$$(abc) = (acb)(ab)(abc)(ab).$$

To show that $A(n) = [S(n), S(n)]$, it thus suffices to show that $A(n)$ is generated by 3-cycles. To see this, recall that $S(n)$ is generated by transpositions; hence elements of $A(n)$ are products of an even number of transpositions. These products can be either disjoint, such as $(ab)(cd)$, or nondisjoint, such as $(ab)(bc)$. Thus the claim follows from the two identities

$$(ab)(cd) = (acb)(acd), \quad \text{and} \quad (ab)(bc) = (abc).$$

□

The commutator subgroup of G has its own commutator subgroup. To avoid making the notation too clumsy, let $G^{(1)}$ denote $[G, G]$, and for each $i > 1$ define $G^{(i)} = [G^{(i-1)}, G^{(i-1)}]$. The *derived series of* G is the sequence

$$G \geq G^{(1)} \geq G^{(2)} \geq \cdots \geq G^{(i)} \geq \cdots . \tag{11.7}$$

Proposition 11.32. *A group G is solvable if and only if its derived series has the property that $G^{(k)}$ is the identity subgroup for some k.*

Proof. The if assertion follows immediately from Proposition 11.30 and the definition of a solvable group. The only if assertion follows from the second assertion of Proposition 11.30, since if H/N is abelian, then N contains $[H, H]$. For if $G = N_1 > N_2 > \cdots > N_{k-1} > N_k = 1$ is a subnormal series such that N_i/N_{i+1} is abelian for all i, then $G^{(i)} < N_{i+1}$. Thus $G^{(k-1)} = 1$. □

Corollary 11.33. *Subgroups and quotients of a solvable group are solvable.*

Proof. Let G be solvable. If H is a subgroup of G, then its derived series satisfies $H^{(i)} < G^{(i)}$, so the derived series of H has to terminate at the identity, since the derived series of G does. Likewise, if N is normal in G, then the derived series of G/N is the image of the derived series of G, so it likewise terminates. □

We now give a nontrivial example of a solvable matrix group.

Proposition 11.34. *The upper triangular subgroup $\mathcal{T}(n, \mathbb{F}) < GL(n, \mathbb{F})$ is solvable.*

Proof. First notice that every element $T \in \mathcal{T}(n, \mathbb{F})$ can be factored as $T = DU = VD$, where $D \in \mathcal{D}(n, \mathbb{F})$ and both U and V are in $\mathcal{U}(n, \mathbb{F})$. Moreover, $\mathcal{D}(n, \mathbb{F}) < N_{GL(n,\mathbb{F})}(\mathcal{U}(n, \mathbb{F}))$. This implies that the commutator of $\mathcal{T}(n, \mathbb{F})$ is contained in $\mathcal{U}(n, \mathbb{F})$, for if $A = DU$ and $B = EV$ are in $\mathcal{T}(n, \mathbb{F})$, then there exist $W, Y \in \mathcal{U}(n, \mathbb{F})$ such that $UE = EW$ and $D^{-1}V^{-1} = YD^{-1}$. Thus,

$$\begin{aligned}[A, B] &= ABA^{-1}B^{-1} \\ &= (DU)(EV)(U^{-1}D^{-1})(V^{-1}E^{-1}) \\ &= (DE)(WVU^{-1}Y)(D^{-1}E^{-1}).\end{aligned}$$

Hence $[A, B] \in \mathcal{U}(n, \mathbb{F})$. Thus it suffices to show that $\mathcal{U}(n, \mathbb{F})$ is solvable. For each k with $1 \leq k < n$, let $U_k = \{U \in \mathcal{U}(n, \mathbb{F}) \mid u_{ij} = 0 \text{ if } 0 < j - i \leq k\}$. In other words, U_k consists of all $U \in \mathcal{U}(n, \mathbb{F})$ that have zero on the first through kth superdiagonals. We leave it to the reader to check that $[U_k, U_k] < U_{k+1}$. But $[U_n, U_n] = \{I_n\}$, so it follows from Proposition 11.32 that $\mathcal{T}(n, \mathbb{F})$ is solvable. □

11.5.3 An example: $A(5)$ is simple

In this section, we will prove that the alternating group $A(5)$ is simple. The proof uses the following characterization of a normal subgroup, whose proof is left to the reader.

Proposition 11.35. *A subgroup H of a group G is normal in G if and only if H is a union of G-conjugacy classes.*

Proposition 11.36. *$A(5)$ is simple.*

Proof. It suffices to show that no subgroup of $A(5)$ is a union of $A(5)$-conjugacy classes. To prove this, we have to describe these conjugacy classes. Since $A(5)$ is normal in $S(5)$, $A(5)$ is the union of $S(5)$-conjugacy classes. Note that the $S(5)$-conjugacy classes in $A(5)$ are unions of $A(5)$-conjugacy classes. By Proposition 11.15, the conjugacy classes in $S(5)$ are represented by

$$(1),\ (12),\ (123),\ (1234),\ (12345),\ (12)(34),\ \text{and}\ \ (123)(45).$$

Since the signature of a k-cycle is 1 if k is odd and -1 is k is even, it follows that $A(5)$ is the union of the $S(5)$-conjugacy classes of the elements

$$(1),\ (123),\ (12345),\ \text{and}\ \ (12)(34). \tag{11.8}$$

Let us first compute the orders of the $A(5)$-conjugacy classes of these elements. By Proposition 11.8, we need to compute their centralizers in $A(5)$. But $Z_{A(5)}(1) = A(5)$, $Z_{A(5)}(123) = \langle(123)\rangle$, $Z_{A(5)}(12345) = \langle(12345)\rangle$, and $Z_{A(5)}((12)(34)) = \{1, (12), (34), (12)(34)\}$. Since $|A(5)| = 60$, Proposition 11.8 implies that the conjugacy classes in $A(5)$ of the elements in (11.8) have respectively 1, 20, 12, and 15 elements. Since these classes give a total of 48 elements, we have to account for another 12 elements. The reader can check that $(12)(12345)(12) = (13452)$ is not conjugate in $A(5)$ to (12345). Thus its $A(5)$-conjugacy class accounts for the remaining 12 elements. Therefore, the order of a normal subgroup of $A(5)$ must be a sum of 1 and some subset of 12, 12, 15, and 20. Hence the order of a nontrivial normal subgroup of $A(5)$ can be only one of 13, 16, 21, 25, 28, 33, 36, 40, 45, and 48. But none of these numbers divides 60, so $A(5)$ must be simple. □

In fact, $A(n)$ is simple for all $n \geq 5$. The standard proof of this is to show that every normal subgroup of $A(n)$ for $n \geq 5$ contains a 3-cycle. But the only normal subgroup of $A(n)$ that contains a 3-cycle is all of $A(n)$, so $A(n)$ is simple. A complete proof can be found in *Abstract Algebra*, by Dummit and Foote (pp. 128–130). The alternating groups also appear in another class of finite simple groups. To describe this class, we first consider the matrix groups $SL(2, \mathbb{F}_p)$. Since the center $Z(G)$ of a group G is always a normal subgroup, the question is whether $G/Z(G)$ is simple. Now, $Z(SL(2, \mathbb{F}_p)) = \{cI_2 \mid c^2 = 1\}$ is nontrivial for $p > 2$. The quotient group $PSL(2, p) = SL(2, \mathbb{F}_p)/Z(SL(2, \mathbb{F}_p))$ is called the *projective linear group of degree two.*

Let us next compute $|PSL(2,p)|$. Since $SL(2,\mathbb{F}_p)$ is the kernel of the homomorphism $\det : GL(2,\mathbb{F}_p) \to (\mathbb{F}_p)^*$, and det is surjective, $|SL(2,\mathbb{F}_p)| = |GL(2,\mathbb{F}_p)|/(p-1)$. But from a computation from Chap. 6, $|GL(2,\mathbb{F}_p)| = (p^2-1)(p^2-p)$ (since $(p^2-1)(p^2-p)$ is the number of pairs of linearly independent vectors in $(\mathbb{F}_2)^2$). Thus if $p > 2$, then

$$|PSL(2,p)| = \frac{|SL(2,\mathbb{F}_p)|}{2} = \frac{(p^2-1)(p^2-p)}{2(p-1)} = \frac{p(p^2-1)}{2}.$$

It follows that $|PSL(2,2)| = 6$, while $|PSL(2,3)| = 12$ and $|PSL(2,5)| = 60$. In fact, using fractional linear transformations and projective geometry, one can write down explicit isomorphisms $PSL(2,2) \cong S(3)$, $PSL(2,3) \cong A(4)$, and $PSL(2,5) \cong A(5)$. The general result is that if $p > 3$, then $PSL(2,p)$ is simple. The proof takes several pages to write down, and we will skip it.

11.5.4 Simple groups and the Jordan–Hölder theorem

The Jordan–Hölder theorem is stated as follows:

Theorem 11.37. *Any two composition series for a group G have the same number of composition factors, and their composition factors are isomorphic up to order.*

The proof is somewhat long and complicated, and we will omit it. As we noted earlier, a finite group always admits at least one composition series, while infinite groups need not admit any. The integers, for example, do not have a composition series.

Example 11.15. For example, the cyclic group C_{12} gives a nice illustration of the Jordan–Hölder Theorem. The group C_{12} has three composition series:

$$C_{12} > C_6 > C_3 > (1), \quad C_{12} > C_6 > C_2 > (1), \text{and} \quad C_{12} > C_4 > C_2 > (1).$$

The corresponding composition factors taken in order are

$$\{C_2, C_2, C_3\}, \quad \{C_2, C_3, C_2\}, \quad \text{and} \quad \{C_3, C_2, C_2\}.$$

Two groups with isomorphic composition factors need not be isomorphic, however, as the following example shows. ⊠

Example 11.16. Let G be a cyclic group of order eight, and let $H < GL(3, \mathbb{F}_2)$ be the group of upper triangular unipotent matrices. Then G and H have order eight, but G and H are not isomorphic, since G is abelian and H isn't. Now a composition series for G is $G = C_8 > C_4 > C_2 > (1)$ with composition factors C_2, C_2, C_2. To get a composition series for H, put

$$H_1 = \left\{ \begin{pmatrix} 1 & 0 & c \\ 0 & 1 & b \\ 0 & 0 & 1 \end{pmatrix} \right\}, \quad \text{and} \quad H_2 = \left\{ \begin{pmatrix} 1 & 0 & c \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \right\}.$$

Then the composition factors for the series $H > H_1 > H_2 > \{I_3\}$ are also C_2, C_2, C_2. Note that the composition factors for G are multiplicative groups, while those for H are additive groups. Nevertheless, the composition factors are isomorphic. ⊠

Remark. The previous example shows that a group isn't determined by its composition series. However, it does suggest the question of how groups can be recovered from their composition series. This is known as the extension problem.

For the final result of this section, we classify the simple solvable finite groups.

Proposition 11.38. *The only finite groups that are both simple and solvable are the cyclic groups of prime order.*

Proof. Suppose G is simple and solvable. Since G has no nontrivial normal subgroups, its only subnormal series is $G > (1)$, which forces G to be abelian, since it is assumed to be solvable. But the only simple abelian groups are the cyclic groups of prime order. □

11.5.5 A few brief remarks on Galois theory

To conclude our introduction to solvable groups, it seems necessary to give a brief overview of Galois theory and the notion of the Galois group of a field extension, which is at the heart of Galois theory. Let $\mathbb{F}$ be a subfield of a field $\mathbb{E}$. Put another way, $\mathbb{E}$ is an extension field of $\mathbb{F}$. As we noted in Chap. 6, $\mathbb{E}$ is a vector space over $\mathbb{F}$; when the vector space dimension of $\mathbb{E}$ over $\mathbb{F}$ is finite, it is customary to denote it by $[\mathbb{E} : \mathbb{F}]$. In that case, one calls $\mathbb{E}$ a *finite extension* of $\mathbb{F}$. For example, $\mathbb{C}$ is a finite extension of $\mathbb{R}$ with $[\mathbb{C} : \mathbb{R}] = 2$.

In the appendix to Chap. 8, we considered a method for extending a field $\mathbb{F}$ to a field containing all the roots of a polynomial $f(x) \in \mathbb{F}[x]$. We will not need to refer to that technique here, but the reader may wish to review it. Let us consider an example. Assume $\mathbb{F} = \mathbb{Q}$ and suppose $f(x) = x^4 - 3$. Letting

α denote the positive real root $\sqrt[4]{3}$, it follows that the four roots of $f(x) = 0$ are $\pm\,\alpha, \pm\, i\alpha$. Now let $\mathbb{E} = \mathbb{Q}(\alpha, i)$ denote the smallest field containing $\mathbb{Q}$, α, and i. Then $\mathbb{E}$ contains all four roots of $f(x) = 0$ as well as $\alpha^2, \alpha^3, i, i\alpha^2, i\alpha^3$. It turns out that $1, \alpha, \alpha^2, \alpha^3, i, i\alpha, i\alpha^2, i\alpha^3$ are linearly independent over $\mathbb{Q}$, and the set of all linear combinations

$$a_1 + a_2\alpha + a_3\alpha^2 + a_4\alpha^3 + a_5 i + a_6 i\alpha + a_7 i\alpha^2 + a_8 i\alpha^3,$$

where $a_1, \ldots, a_8 \in \mathbb{Q}$, forms a field that is a subfield of $\mathbb{C}$. We leave the verification of this assertion to the reader. This field must evidently be $\mathbb{E}$, so $[\mathbb{E} : \mathbb{Q}] = 8$. Thus the smallest field containing $\mathbb{Q}$ and all the roots of $x^4 - 3 = 0$ has dimension eight over $\mathbb{Q}$.

We now make the key definition.

Definition 11.10. Let $\mathbb{E}$ be an extension field of $\mathbb{F}$ that is the smallest field containing all roots of a polynomial $f(x) \in \mathbb{F}[x]$ without repeated roots. Then the *Galois group* $\mathrm{Gal}(\mathbb{E}/\mathbb{F})$ of $\mathbb{E}$ relative to $\mathbb{F}$ is the set of all field isomorphisms $\phi : \mathbb{E} \to \mathbb{E}$ such that $\phi(a) = a$ for all $a \in \mathbb{F}$.

By definition, $\mathrm{Gal}(\mathbb{E}/\mathbb{F})$ is a group under composition. Since $\phi(a) = a$ for all $a \in \mathbb{F}$, it follows that each $\phi \in \mathrm{Gal}(\mathbb{E}/\mathbb{F})$ is also an endomorphism of the vector space $\mathbb{E}$ over $\mathbb{F}$. Note that for every $g(x) \in \mathbb{F}[x]$ and $\beta \in \mathbb{E}$ such that $g(\beta) = 0$, the definition of $\mathrm{Gal}(\mathbb{E}/\mathbb{F})$ implies $\phi(g(\beta)) = g(\phi(\beta)) = 0$ for all $\phi \in \mathrm{Gal}(\mathbb{E}/\mathbb{F})$. In other words, elements of $\mathrm{Gal}(\mathbb{E}/\mathbb{F})$ have to permute the roots in $\mathbb{E}$ of arbitrary polynomials in $\mathbb{F}[x]$.

Let us return to the polynomial $f(x) = x^4 - 3$. Now, every $\phi \in \mathrm{Gal}(\mathbb{E}/\mathbb{Q})$ permutes the roots $\pm\alpha, \pm i\alpha^2$ of $f(x)$. Furthermore, if two elements $\phi, \psi \in \mathrm{Gal}(\mathbb{E}/\mathbb{Q})$ satisfy $\phi(\mu) = \psi(\mu)$ for each root μ, then $\phi = \psi$. Since $f(x) = 0$ has four roots in $\mathbb{E}$, $\mathrm{Gal}(\mathbb{E}/\mathbb{Q}) < S(4)$. Let us now determine $\mathrm{Gal}(\mathbb{E}/\mathbb{Q})$. The polynomial f has four distinct roots in $\mathbb{E}$ that are not in $\mathbb{Q}$. Notice that since $\pm i$ satisfy $x^2 + 1 = 0$, the above remark implies that every $\phi \in \mathrm{Gal}(\mathbb{E}/\mathbb{Q})$ must have $\phi(i) = \pm i$. However, ϕ can send α to any other root. Thus $\mathrm{Gal}(\mathbb{E}/\mathbb{Q})$ has an element of order four and an element of order two, and it can be shown by a tedious calculation that they generate $\mathrm{Gal}(\mathbb{E}/\mathbb{Q})$. To identify $\mathrm{Gal}(\mathbb{E}/\mathbb{Q})$, let $\varphi \in \mathrm{Gal}(\mathbb{E}/\mathbb{Q})$ satisfy $\varphi(\alpha) = \alpha$ and $\varphi(i\alpha) = -i\alpha$. Then $\varphi(i) = -i$. Next define $\tau \in \mathrm{Gal}(\mathbb{E}/\mathbb{Q})$ by $\tau(\alpha) = -i\alpha$ and $\tau(i) = i$. Then $\varphi^2 = \tau^4 = 1$ and $\varphi\tau\varphi = \tau^{-1}$ in $\mathrm{Gal}(\mathbb{E}/\mathbb{Q})$. We claim (again omitting the details) that φ and τ also generate $\mathrm{Gal}(\mathbb{E}/\mathbb{Q})$. Therefore, $\mathrm{Gal}(\mathbb{E}/\mathbb{Q})$ is isomorphic to the dihedral group $D(4)$, of order eight.

In general, we have the following assertion.

Proposition 11.39. *If $f(x) \in \mathbb{F}[x]$ is a polynomial with only simple roots and $\mathbb{E}$ is the smallest extension field of $\mathbb{F}$ containing all the roots of $f(x)$, then $|\mathrm{Gal}(\mathbb{E}/\mathbb{F})| \leq n!$, where $n = [\mathbb{E} : \mathbb{F}]$.*

The fundamental theorem of Galois theory stated below explains the correspondence between the subfields of $\mathbb{E}$ containing $\mathbb{F}$ and the subgroups of $\mathrm{Gal}(\mathbb{E}/\mathbb{F})$.

Theorem 11.40. *Let $\mathbb{E}$ be an extension field of the field $\mathbb{F}$ that is obtained by adjoining all the roots of an irreducible polynomial $f(x)$ over $\mathbb{F}$ with simple roots. Then there is a one-to-one correspondence between subgroups of* $\mathrm{Gal}(\mathbb{E}/\mathbb{F})$ *and fields $\mathbb{K}$ such that $\mathbb{F} \subset \mathbb{K} \subset \mathbb{E}$ in which a subgroup H corresponds to the subfield $\mathbb{K} = \mathbb{E}^H$ of elements of $\mathbb{E}$ fixed by all elements of H, and a subfield $\mathbb{K}$ such that $\mathbb{F} \subset \mathbb{K} \subset \mathbb{E}$ corresponds to* $\mathrm{Gal}(\mathbb{E}/\mathbb{K})$. *Furthermore, if H is a normal subgroup, then* $\mathrm{Gal}(\mathbb{K}/\mathbb{F}) \cong \mathrm{Gal}(\mathbb{E}/\mathbb{F})/\mathrm{Gal}(\mathbb{E}/\mathbb{K})$, *where $\mathbb{K}$ is the fixed subfield $\mathbb{E}^H$.*

Returning to $\mathbb{Q}(\alpha, i)$, let us determine what the theorem says. The group $\mathrm{Gal}(\mathbb{E}/\mathbb{Q}(i))$ is the subgroup of $\mathrm{Gal}(\mathbb{E}/\mathbb{Q})$ containing all elements of $\mathrm{Gal}(\mathbb{E}/\mathbb{Q})$ fixing i. The element τ fixes i and generates a cyclic group $< \tau >$ of order four. Therefore, $\mathrm{Gal}(\mathbb{E}/\mathbb{Q}(i))$ is evidently $< \tau >$. Notice that $< \tau >$ is a normal subgroup of $\mathrm{Gal}(\mathbb{E}/\mathbb{Q})$. Note also that $\mathrm{Gal}(\mathbb{Q}(i)/\mathbb{Q}) \cong C_2$. The conclusion from this is that $|\mathrm{Gal}(\mathbb{E}/\mathbb{Q})| = 8$, which bypasses the tedious calculations mentioned above and shows that $geq \cong D(4)$.

Exercises

Exercise 11.5.1. Show directly that $S(3)$ is solvable.

Exercise 11.5.2. True or false: a nonabelian simple group is its own commutator.

Exercise 11.5.3. True or false: all dihedral groups are solvable.

Exercise 11.5.4. Show, without using the Sylow theorems, that the Klein 4-group is normal in $A(4)$. (Suggestion: make a list of the conjugacy classes of $A(4)$ and show that V_4 is a union of conjugacy classes.)

Exercise 11.5.5. Show that if H is normal in G, then G/H is simple if and only if there does not exist a normal subgroup N of G such that $H < N < G$.

Exercise 11.5.6. Show that a solvable group has an abelian normal subgroup.

Exercise 11.5.7. Prove that a finite group G is solvable if and only if G has a subnormal series whose factors are cyclic.

Exercise 11.5.8. Suppose a finite group G contains a subgroup H that is not solvable. Can G itself be solvable? If so, give an example.

Exercise 11.5.9. Suppose $G = H_0 > H_1 > \cdots > H_k$ is a subnormal series. Show that if G is finite, then $|G| = h_1 \cdots h_k$, where $h_i = |H_{i-1}/H_i|$ is the order of the ith factor.

Exercise 11.5.10. Show that the elementary matrix matrix $E = \begin{pmatrix} 1 & \lambda \\ 0 & 1 \end{pmatrix}$ over a field $\mathbb{F} \neq \mathbb{F}_2$ is a commutator in $GL(2, \mathbb{F})$. Use this to show that $GL(2, \mathbb{F})$ is its own commutator.

Exercise 11.5.11. Explain why $\mathbb{Z}$ does not have a composition series.

Exercise 11.5.12. Show that every finite group has a composition series.

Exercise 11.5.13. Does $S(5)$ possess two distinct composition series?

11.6 Appendix: $S(n)$, Cryptography, and the Enigma

One of the most interesting chapters in the history of World War 2 is how the British cryptographers at Bletchley Park were able to solve the German cipher machine known as the Enigma enabling the Allied military to read virtually all the top secret military transmissions of the German military. However, until just before the beginning of World War 2, the British had no idea what sort of cipher the German military was using and were shocked to learn that the mathematicians of the Polish cipher bureau had been able to decipher the Enigma since 1932. Just before Germany invaded Poland in 1939, the Poles were able to give the British an actual German Enigma they had reverse engineered along with their knowledge of how to operate it.

The main tool the Polish cryptographers used was group theory. Their accomplishment still stands as both the first and undoubtedly most important use of abstract algebra in cryptography or any other endeavor outside of pure mathematics. Since we are primarily interested in the role of group theory, we will not mention many of the fascinating aspects of this story, such as, for example, how Enigma led to the development of the computer. There is an article by the principal character, Marian Rejewski, in the *Annals of the History of Computing*, Vol. 3, Number 3, July, 1981, which gives a fascinating first-hand account. There are now many books and articles on the Enigma. *Enigma*, by W. Kozachuk (published in 1984), is an excellent, though not easy to find, account. It is the only book that contains appendices written by Rejewski himself explaining his breakthroughs. Kozachuk was himself a Polish army officer and a military historian. Another excellent account is given in *Intercept*, by Jozef Garlinski. Both books are fascinating.

11.6.1 Substitution ciphers via $S(26)$

A cipher is an algorithm for disguising a message so that only the sender and the intended recipient for can read it. Cryptology, the mathematical discipline of ciphers, consists of two areas, cryptography and cryptanalysis. A cryptographer makes ciphers and a cryptanalyst tries to break them. A *substitution cipher* is created by permuting the alphabet using an element of $S(26)$. A substitution cipher is one of the oldest ciphers in existence. Let us consider an example. The following ciphertext can be deciphered in a few minutes by analyzing the frequencies of the letters in the message. Guessing the letters used in the one- and two-letter words is useful.

F KVZSDVS XNZVN
NSKZOFOSK OL ULWESO
ZOK ULHDRSWK ZK CLKO
FCUWSR DLWON XNZOSNSFR

Let us make a few observations. The frequencies of letters in the ciphertext should roughly correspond to the frequencies of letters in plaintext, that is, the message in English that has been enciphered. Notice, for example, that there are eight S's, so there is a strong probability that S represents E or I. There are six Z's, so Z is another candidate for E or I. But there are no commonly used two-letter words that begin with E, and several that begin with I, so Z very likely represents I. Since I has now been used, and the only two one-letter words are A and I, we may infer that F represents A. This is a start, but there is still work to do. One of the obvious ways of making this cipher stronger would be to remove the spaces between words, since that would conceal the one and two letter words.

11.6.2 The Enigma

A more sophisticated substitution cipher than the simple substitution described above could employ several substitutions. For example, one might encipher the first letter by a permutation σ_1, the second by another permutation σ_2, the third by σ_3, and so on. Since there are 26! possible substitutions, if the sequence of permutations was sufficiently random and didn't repeat often, the cipher would be very hard to break, and statistics would be of little help. This sort of variation of the substitution cipher has long been incorporated in commercial cipher machines. The most famous of these machines is the Enigma, which was manufactured in Germany and adapted by the German military in 1929 for its military transmissions. In 1928, the Polish Cipher Bureau was tipped off that the German military was interested in the Enigma when an Enigma machine was inadvertently shipped to Poland marked as radio parts. The cipher bureau learned of the misplaced package through customs because of the anxiety of the German officials who had mistakenly sent the wrong package, which they demanded to have returned immediately. The mistake was discovered on a Saturday, so Polish customs had time to allow the cipher bureau experts to inspect the contents of the package, which they realized was an Enigma cipher machine. It was carefully repackaged and returned, and apparently the Germans never suspected that the Poles had learned about their error.

Around then, the Poles noticed that the German military began to use an entirely new system of encipherment. They correctly surmised the source, and the cipher bureau purchased a commercial Enigma for further study. The chief of the cipher bureau made the astute observation that traditional cryptological methods (linguistics and statistics) would not be of any use against such a machine, and he organized a course in the mathematics department at the University of Poznan to train cryptologists, hoping that he would find some brilliant students. There were indeed three outstanding students, who

were recruited, and by 1932, they had actually succeeded in recreating a German military Enigma, the so-called the Enigma double. This breakthrough was based on a brilliant observation about permutation groups by Marian Rejewski.

To explain the role group theory played in helping Rejewski and his colleagues duplicate the Enigma, we need to describe how it worked. It somewhat resembled a portable typewriter. It had a keyboard with a key for each letter but no space bar, no shift, and no keys for punctuation. Mounted above and behind the keyboard where the keys would strike the paper was a lamp board displaying 26 lights labeled a through z. Pressing a key had the affect of causing one of the lamps to light up. An Enigma required an operator, who typed in the plaintext, and an assistant, who recorded the ciphertext as the lamps were lit up in sequence. If, say, a was pressed five times in succession, a sequence of five lamps lit up. For example, pressing $aaaaa$ might produce $bsfgt$. The sequence would eventually reappear, but not for a long time, in fact, not until a had been pressed $(26)^3$ times. Curiously, due to the way the keyboard was wired to the lamp board, if a was pressed, the lamp corresponding to a could not light up. Eventually, the British cryptographers figured out how to exploit this feature to their great benefit. Once a message was enciphered, it was sent in Morse code. A recieved enciphered message was deciphered in exactly the same way. After it was decoded from the Morse code to reveal the ciphertext, the operator typed the ciphertext on the keyboard, and the assistant read off the plaintext as the lamps lit up one after another.

Now let us turn to how the machine functioned, which will explain why enciphering and deciphering were the same process. The original version of the Enigma contained three adjacent rotors on a horizontal axle. Each of the rotors had 26 terminals equally spaced around both its left-hand and right-hand circumferences. The terminals around each circumference represented the alphabet arranged in the usual order, and each terminal on the left circumference of a rotor was wired internally to a single terminal on the right. When a key was pressed, a current passed through the left rotor from the left-hand terminal to one on the right, then through the middle rotor, and finally through the right-hand rotor, thereby undergoing three permutations σ_L, σ_M, σ_R in that order. The current then passed through a fixed disk at the end of the axle with 26 terminals around its inner circumference, each wired to another terminal. This disk was called the reflector. The current then returned through the right-hand rotor, middle rotor, and left-hand rotor to the key that had caused its lamp to illuminate.

Let ρ denote the permutation of the reflector. Suppose the a key is struck. Then the lamp that is illuminated by striking a is

$$\sigma_L^{-1}\sigma_M^{-1}\sigma_R^{-1}\rho\sigma_R\sigma_M\sigma_L(a).$$

Suppose this is w. Notice that ρ has the property that $\rho = \rho^{-1}$. In a group, such an element is called an *idempotent*. The permutation $\sigma_L^{-1}\sigma_M^{-1}\sigma_R^{-1}\rho\sigma_R\sigma_M\sigma_L$, being conjugate to the idempotent ρ, is also idempotent. Thus if pressing a lights up w, then pressing w lights up a. This was a most convenient feature of Enigma and explains why encipherment and decipherment were performed in the same way.

What complicated the encipherment is that each rotor could be independently rotated through all 26 positions. Every time a key was pressed, the first rotor moved forward one terminal. This shift corresponded to the cyclic permutation $\pi = (abc \ldots xyz)$, of order 26. Hence the second letter of plaintext would be enciphered by

$$\pi^{-1}\sigma_L^{-1}\pi\sigma_M^{-1}\sigma_R^{-1}\rho\sigma_R\sigma_M\pi^{-1}\sigma_L\pi.$$

Notice that we have inserted $\pi^{-1}\sigma_L\pi$ for σ_L, since the middle and right rotors were stationary. Without conjugating by π, all three rotors would advance $1/26$ revolution together. As soon as the first 26 letters had been enciphered and the left-hand rotor had made a complete revolution, the middle rotor advanced $1/26$th of a revolution. The 27th letter was thus enciphered by

$$\sigma_L^{-1}\pi^{-1}\sigma_M^{-1}\pi\sigma_R^{-1}\rho\sigma_R\pi^{-1}\sigma_M\pi\sigma_L$$

since $\pi^{26} = (1)$. As soon as $26^2 = 626$ letters were enciphered, the right-hand rotor moved forward $1/26$th of a revolution, and so on. The rotors thus kept cycling through different permutations until $26^3 = 17576$ keys had been pressed, after which the cycle repeated.

11.6.3 Rejewski's theorem on idempotents in $S(n)$

We will now pause to analyze some properties of idempotents in $S(n)$. To take a quick example, note that $(ab)(cd)(ef)$ is an idempotent, but $(ab)(bd)$ is not. Since an idempotent ρ has the property that $\rho^2 = 1$, it follows that ρ's disjoint cycles must be transpositions, since a k-cycle has order k. The reflector ρ on the Enigma was the product of 13 disjoint transpositions, since every terminal had to be paired with a different terminal, since pairing a terminal with itself would mean that sometimes pressing a key would fail to illuminate a lamp, since a key could not light up its own lamp. Thus, by the binomial theorem, the total number of possible Enigma reflectors is

$$\frac{26!}{2^{13}} = 49229914688306352 \times 10^6.$$

Although it was certainly not obvious at the time, it turned out that the key to unlocking how the Enigma rotors were wired is what happens when two idempotents are multiplied. This question was answered by Marian Rejewski in 1932 in the following theorem, which has also been referred to as "the theorem that won World War Two."

Theorem 11.41 (Rejewski's theorem)**.** *Let σ and τ be idempotents in $S(n)$ with the same fixed points in $\{1, \ldots, n\}$. Then the number of disjoint cycles in $\sigma\tau$ of each length is even (including the possibility of length 0). Thus if $\sigma\tau$ has a disjoint cycle of length $k > 0$, then it has an even number of them. Conversely, an element of $S(n)$ that has the property that there is an even number of disjoint cycles of each possible length in its disjoint cycle representation is a product of two idempotents (though possibly in several ways).*

Considering an example will give a good idea why the first assertion is true, but its converse is harder to justify.

Example 11.17. The permutations

$$\sigma = (a\ e)(b\ f)(c\ g)(h\ d) \quad \text{and} \quad \tau = (b\ e)(f\ c)(h\ g)(a\ d)$$

are idempotents in $S(8)$ with the same fixed letters, namely i through z. To see how to construct the disjoint cycles in $\sigma\tau$, consider the following arrangement:

$$\begin{array}{ccccccccc} a & & h & & c & & b & & a \\ & d & & g & & f & & e & \end{array}.$$

Note that τ acts by reading diagonally down from left to right, while σ acts by reading diagonally up from left to right. Thus the disjoint cycle decomposition of $\sigma\tau$ is revealed by reading the top row from left to right to get one cycle, and the bottom row from right to left to get the other. Thus $\sigma\tau = (a\ h\ c\ b)(e\ f\ g\ d)$. Similarly, $\tau\sigma = (a\ b\ c\ h)(d\ g\ f\ e)$. The cycles of each length occur in pairs. In each case, there are only two cycles, and both are of length four. This construction works for the product of any two idempotents with the same fixed letters. The converse statement is harder but more important. It is this fact that led to Rejewski's breakthrough. ⊠

11.7 Breaking the Enigma

As we noted above, for two Enigmas with the same initial rotor settings, enciphering and deciphering were the same operation. The operator who received a message had only to type in the ciphertext as the assistant read off the plaintext on the lamp board. To ensure that the starting positions were always the same, a *daily key* schedule was compiled in a codebook issued to all the Enigma operators. If on September 5, 1940, the daily key was xsf, then on that day all Enigmas would begin sending and deciphering with the left rotor set at x, the middle at s, and the right at f. To increase security, each operator also selected another three-letter key, a so-called *telegram key*, e.g., arf. Then before enciphering took place, the operator, with the Enigma set to the daily key xsf, enciphered the telegram key arf. As an error-detecting device, the operator actually typed $arfarf$, producing a six-letter string, let us say $wkuygh$. This six-letter string was then sent by Morse code as the first six letters of the enciphered message. After sending his doubly enciphered telegram key, the operator set his rotors to arf and proceeded to encipher the plaintext. The operator on the receiving end, with his Enigma set to the daily key xsf, typed in $wkuygh$. The deciphered string $arfarf$ told him to reset his rotors to arf before typing in the ciphertext. Of course, if something like $arfark$ was received, this signaled a transmission error, and the message wasn't deciphered until the doubly enciphered telegram key was resent. The double enciphering of the telegram key was necessary. Radio transmissions could be disrupted by static, and there was always the possibility of human error under the difficulties experienced in wartime. But it turned out to be the weak link. To see why, see whether you can detect a pattern in the first six-letter groups from fifteen messages all intercepted on September 5. Note that all groupings are double encipherings of different telegram keys.

wkuygh wctyuo qvtnno kophau evprmu
qmlnxz wvqymk dgybhj bxcdla mijwce
dbobth yoeiaw ntplbu yugicf lhmqzp

The interesting feature of these six-letter groups is that whenever two have the same first letter, they have the same fourth letter, and conversely. This also holds for the second and fifth letters and the third and sixth letters. A good cryptographer would notice this feature immediately, but an untrained eye (such as the author's) might not see it for quite a while, or ever. This clearly suggested a double encryption hypothesis. Working on this assumption, Rejewski had the wonderful insight to string together the first and fourth letters of all the first six-letter groups for all the intercepted messages on a particular day, since they were all enciphered with the same daily key: xsf in the case of September 5. Here is what the above fifteen intercepts give:

$$dbd\ldots mwyi\ldots qnlq\ldots odb\ldots fv\ldots er\ldots kh\ldots.$$

However, working with 60–80 daily intercepts, he was sometimes able to string together the whole alphabet, getting an element $\pi \in S(26)$. Let us see what the construction of π tells us. Looking at the first grouping, we know that the operator typed $arfarf$, which produced $wkuygh$. So typing a on the keyboard produced w on the lamp board via an idempotent σ_1. Now let σ_2 be the idempotent that sends a to y. Then since $\pi(w) = y$, we have $\pi\sigma_1(a) = \sigma_2(a)$. But since all the Enigmas were set up on September 5 with the same daily key xsf, the pairings σ_1 and σ_2 would be the same for all Enigmas. Hence $\pi\sigma_1 = \sigma_2$. Consequently, $\pi = \sigma_2\sigma_1$ in $S(26)$! Thus we see the surprising way in which the product of two idempotents figured. It would be possible, though not easy, to find the disjoint cycle representation of π, and from this, one might be able to deduce σ_1 and σ_2. Recall that neither σ_1 nor σ_2 could have any fixed letters. Let us call such idempotents *pairings*. Thus, it might be possible to find the pairings σ_1 and σ_2 from π. Repeating this process for the second and fifth and the third and sixth letters gave two more elements of $S(26)$ that were also products of two pairings. There was an unavoidable problem, however: a factorization into pairings is not unique. Let us take a couple of simple examples to illustrate how the pairings might be found.

Example 11.18. Let us shorten the alphabet to a through h, and let $\pi = (ahc)(dgb)$. We want to write this as $\sigma\tau$, where σ and τ are pairings. We can clearly see that $e \to e$ and $f \to f$. Imitating the procedure illustrated after Rejewski's theorem, consider the three possibilities taking the cyclic permutations of b, d, g into account:

$$\begin{array}{cccc} a & h & c & a \\ & b & g & d \end{array},$$

$$\begin{array}{cccc} a & h & c & a \\ & g & d & b \end{array},$$

and

$$\begin{array}{cccc} a & h & c & a \\ & d & b & g \end{array}.$$

Thus there are three possible solutions:

$$\sigma = (a\ b)(h\ g)(c\ d)(e\ f), \qquad \tau = (a\ d)(c\ g)(b\ h)(e\ f),$$

$$\sigma = (a\ g)(d\ h)(b\ c)(e\ f), \qquad \tau = (a\ b)(c\ d)(g\ h)(e\ f),$$

$$\sigma = (a\ d)(b\ h)(c\ g)(e\ f), \qquad \tau = (a\ g)(c\ d)(d\ h)(e\ f).$$

Notice that we included $(e\ f)$ in each solution in order to ensure that σ and τ are pairings, not just idempotents. The transposition $(e\ f)$ disappears in the product.

Example 11.19. Consider the permutation

$$\pi = (d\ e\ p\ z\ v\ l\ y\ q)(a\ r\ o\ n\ j\ f\ m\ x)(b\ g\ k\ tu)(w\ s\ c\ i\ h).$$

Thus we consider pairs of arrays such as

$$\begin{array}{cccccccc} d & e & p & z & v & l & y & q & d \\ x & m & f & j & n & o & r & a & \end{array}$$

and

$$\begin{array}{cccccc} b & g & k & t & u & b \\ h & i & c & s & w & \end{array}.$$

One possible solution is therefore

$$\sigma = (d\ x)(e\ m)(f\ p)(j\ z)(n\ v)(l\ o)(r\ y)(a\ q)(b\ h)(g\ i)(k\ c)(s\ t)(u\ w)$$

and

$$\tau = (a\ d)(q\ r)(o\ y)(ln\ )(j\ v)(f\ z)(m\ p)(e\ x)(b\ w)(u\ s)(t\ c)(k\ i)(g\ h).$$

Since we obtain all solutions by cyclicly permuting the second rows of the two arrays, there are 128 solutions in all.

Rejewski's solution helped reveal the some of the pairings. If he had known the daily keys, then he would have gotten some real insight into the wiring of the rotors. But it wasn't always necessary to know the daily keys, because the Enigma operators often chose easy to guess keys such as *aaa* or *abc*, or they might always choose the same key. But what turned out to be a huge break for the Poles was that French intelligence uncovered a disgruntled German code clerk who sold them the daily keys for a two-month period in 1932. They were given to the Poles, who used this windfall along with Rejewski's factorizations. Using several other clever and imaginative devices, the three Polish cryptographers were able to decipher their first Enigma message at the end of 1932. By 1934, they had completely solved the puzzle of the wiring of the rotors and were able to build an exact replica of the Enigma (an Enigma double). What made this even more amazing was that Poland was economically depressed, having become an independent country only at the end of the First World War, and the financial outlay for this project was a serious strain on its national treasury. Yet because of the wisdom of the head of its cipher bureau and the ingenuity of its cryptographers, the Poles

were years ahead of the British and French, who despite their great economic advantage, had been unable to unravel the Enigma's mystery.

Rejewski had even constructed a primitive computer, which he called a bomb, to test for the daily keys. Fortunately, a couple of months before Germany's shocking invasion of Poland in September 1939, two of the duplicate Enigmas were handed over to the French, who gave one to the British. With this windfall, the British cryptographers at Bletchley Park were immediately able to decipher a certain amount of the intercepted radio traffic, somewhere on the order of 150 intercepts per day. But after the war started, the Germans upgraded their system, so the cryptanalysts were stymied until they figured out what modifications the Germans had made. In 1943, the British, under the leadership of Alan Turing, built the first true electronic computer to test for the daily keys. They also called it the bomb, apparently in honor of the Polish original. With the bomb, the Bletchley Park cryptanalysts were eventually able to read virtually all of the top secret communications of the German High Command, often before the officers for whom the communiqués were intended.

The breaking of the Enigma surely shortened the war. In fact, after learning in the 1970s the extent to which the British had penetrated the Enigma ciphers, the head of German Enigma security, a mathematical logician, stated that it was a good thing, since it must have shortened the war. He was happy to learn that a fellow logician, Alan Turing, had played such an important role. A well-known German mathematician, who was responsible for the security of the Enigmas for a branch of the German military, was quoted in an obituary as saying that it was a job he hadn't been very good at.

Acknowledgments The above account of how the Enigma was broken was originally explained to me by my colleague Professor Hugh Thurston, who was a British cryptanalyst at Bletchley Park during the war. He characterized the code-breaking activity at Bletchley Park as "one of the few just war efforts human history can boast of."

Chapter 12
Linear Algebraic Groups: an Introduction

The purpose of this chapter is to give a brief informal introduction, with very few proofs, to the subject of linear algebraic groups, a far-reaching generalization of matrix theory and linear algebra. A very readable treatment with much more information is contained in the book *Linear Algebraic Groups and Finite Groups of Lie Type*, by Gunter Malle and Donna Testerman. A *linear algebraic group* is a matrix group G contained in a general linear group $GL(n, \mathbb{F})$, for some field $\mathbb{F}$ and positive integer n, whose elements are precisely the roots, or zeros, of a finite set of polynomial equations in n^2 variables. Linear algebraic groups have proved to be indispensable in many areas of mathematics, e.g., number theory, invariant theory, algebraic geometry, and algebraic combinatorics, to name some. The results we will describe concern, for the most part, the case $\mathbb{F} = \mathbb{C}$, but results that are valid for $\mathbb{C}$ are usually true whenever $\mathbb{F}$ is algebraically closed and of characteristic zero. There is also a great deal of interest in linear algebraic groups over a field of characteristic $p > 0$, since many such groups give examples of finite simple groups.

12.1 Linear Algebraic Groups

In order to define what a linear algebraic group is, we must first make some remarks about polynomials on $\mathbb{F}^{n\times n}$. Let V be a finite-dimensional vector space over $\mathbb{F}$ with basis $\mathbf{v}_1, \ldots, \mathbf{v}_n$, and let $x_1, \ldots, x_n$ be the dual basis of V^*. Thus $x_i : V \to \mathbb{F}$ is the linear function defined by $x_i(\mathbf{v}_j) = \delta_{ij}$, where δ_{ij} is the Kronecker delta function. A *monomial* in $x_1, \ldots, x_n$ is a function of the form $x_{i_1}^{a_1} x_{i_2}^{a_2} \cdots x_{i_k}^{a_k} : V \to \mathbb{F}$, where $a_1, \ldots, a_k$ are positive integers, $1 \leq i_j \leq n$ for all indices and $i_1 < i_2 < \cdots < i_k$. A polynomial function in $x_1, \ldots, x_n$ over $\mathbb{F}$ is a linear combination over $\mathbb{F}$ of a finite set of monomials. Thus a typical

J.B. Carrell, *Groups, Matrices, and Vector Spaces*,
DOI 10.1007/978-0-387-79428-0_12

polynomial has the form

$$f(x_1, \ldots, x_n) = \sum c_{i_1 i_2 \ldots i_k} x_{i_1}^{a_1} x_{i_2}^{a_2} \cdots x_{i_k}^{a_k},$$

where the coefficients $c_{i_1 i_2 \ldots i_k}$ are in $\mathbb{F}$, and only finitely many are nonzero. Let $\mathbb{F}[x_1, \ldots, x_n]$ denote the set of all such polynomials. If $V = \mathbb{F}^{n \times n}$ with basis E_{ij}, $1 \leq i, j \leq n$, then the dual basis is denoted by x_{ij}, where $1 \leq i, j \leq n$. A basic example of a polynomial in $\mathbb{F}[x_{ij}]$ is given by the determinant

$$\det(x_{ij}) = \sum_{\pi \in S(n)} \operatorname{sgn}(\pi)\ x_{\pi(1)1} x_{\pi(2)2} \cdots x_{\pi(n)n}.$$

We now define what it means for a matrix group to be closed.

Definition 12.1. A subgroup G of $GL(n, \mathbb{F})$ is said to be *closed* if G consists of the common zeros of a finite set of polynomials $f_1, \ldots, f_k \in \mathbb{F}[x_{ij}]$. A closed subgroup G of $GL(n, \mathbb{F})$ is called a *linear algebraic group*.

The group $GL(n, \mathbb{F})$ is by default a linear algebraic group, since $GL(n, \mathbb{F})$ is the set of common zeros of the empty set of polynomials on $\mathbb{F}^{n \times n}$. The special linear group $SL(n, \mathbb{F})$ is a closed subgroup of $GL(n, \mathbb{F})$, since $SL(n, \mathbb{F})$ is defined by the setting $\det(x_{ij}) - 1 = 0$. That is, $SL(n, \mathbb{F})$ consists of the matrices A such that $\det(A) - 1 = 0$. The groups $P(n), \mathcal{U}(n, \mathbb{F}), \mathcal{L}(n, \mathbb{F})$, and $\mathcal{D}(n, \mathbb{F})$ involved in the $LPDU$ are also closed subgroups of $GL(n, \mathbb{F})$, as is $\mathcal{T}(n, \mathbb{F})$. For example, $\mathcal{T}(n, \mathbb{F})$, the subgroup consisting of all upper triangular elements of $GL(n, \mathbb{F})$, is the set of zeros of the polynomials x_{ij}, where $i > j$. The elements of the group $P(n)$ of $n \times n$ permutation matrices satisfy the equations $x_{ij}(x_{ij} - 1) = 0$ for all i, j, since every entry of a permutation matrix is zero or one. But these equations don't capture the fact that the columns of P are orthogonal; for that, we use the identity $P^T P = I_n$. Thus, $P \in P(n)$ if and only if P satisfies the set of polynomial equations

$$\sum_k x_{ki} x_{kj} = \delta_{ij} \quad \text{and} \quad x_{ij}(x_{ij} - 1) = 0 \quad (i, j, k = 1, \ldots, n).$$

Hence $P(n)$ is a linear algebraic group; in fact, this example shows that linear algebraic groups can be finite. (The reader may wish to show that in fact, every finite subgroup of $GL(n, \mathbb{F})$ is closed.)

From the definition of a linear algebraic group, it is clear that the intersection of two linear algebraic groups is a linear algebraic group, and the product of two linear algebraic groups is also a linear algebraic group. In order to understand the importance of the condition that the elements of a linear algebraic group are cut out by polynomial equations, one must introduce some basic concepts from algebraic geometry, in particular, the notion of an affine variety. What is very useful is that the important subgroups of a linear algebraic group themselves also turn out to be linear algebraic groups.

We know make a general definition that applies to arbitrary groups. The reason for the terminology should become clear later.

Definition 12.2. A group G, not necessarily a linear algebraic group, is said to be *connected* if G has no normal subgroup of finite index greater than one.

By definition, a finite group different from $\{1\}$ cannot be connected. To see another example of a nonconnected group, consider $O(n, \mathbb{F})$ for which the characteristic of $\mathbb{F}$ is different from two. In that case, $P(n)$ is naturally embedded as a subgroup of $O(n, \mathbb{F})$. Thus, the determinant homomorphism $\det : O(n, \mathbb{F}) \to \{\pm 1\}$ is surjective. Consequently, by the first isomorphism theorem, its kernel is a proper normal subgroup of index two. If the characteristic of $\mathbb{F}$ is zero, then $GL(n, \mathbb{F})$, $SL(n, \mathbb{F})$, and $SO(n, \mathbb{F})$ are all connected.

Exercise 12.1.1. Show that if $\mathbb{F}$ is a Galois field, then $GL(n, \mathbb{F})$ is not connected.

12.1.1 Reductive and semisimple groups

We will now get a bit technical for a while in order to focus on two of the most important classes of linear algebraic groups: reductive and semisimple groups. We will concentrate on linear algebraic groups $G < GL(n, \mathbb{C})$. Each such G has a unique maximal, closed, normal, and connected solvable subgroup $Rad(G)$, called the *radical* of G. For example, if $G = GL(n, \mathbb{C})$, then $Rad(G) = \mathbb{C}^* I_n$. (Note: $\mathbb{C}^* I_n$ is closed, since it is the set of all zeros in $GL(n, \mathbb{C})$ of the polynomials x_{ij} and $x_{ii} - x_{jj}$, where $i \neq j$). If $G = \mathcal{T}(n, \mathbb{C})$ or $\mathcal{D}(n, \mathbb{C})$, then $Rad(G) = G$, since G is solvable. We now state the first main definition.

Definition 12.3. Let $G < GL(n, \mathbb{C})$ denote a connected linear algebraic group. Then G is called *reductive* if the only unipotent element of $Rad(G)$ is the identity. It is called *semisimple* if the only closed, normal, connected abelian subgroup of G is the identity. It is called *simple* if G is nonabelian and G has no nontrivial closed normal subgroup. Finally, G is *almost simple* if every closed normal subgroup of G is finite.

It is not hard to see that a semisimple linear algebraic group is reductive, but the converse isn't true. For example, by the above remarks, the general linear group $GL(n, \mathbb{C})$ is reductive. However, $GL(n, \mathbb{C})$ is not semisimple, since its radical $\mathbb{C}^* I_n$ is a nontrivial closed normal connected abelian subgroup. Another example of a reductive group that is not semisimple is $\mathcal{D}(n, \mathbb{C})$. The special linear group $SL(n, \mathbb{C})$ is semisimple. This is hard to prove from first principles, so we will omit the proof. Note that $SL(n, \mathbb{C})$ contains a normal abelian subgroup, namely its center $Z = \{\zeta I_n \mid \zeta^n = 1\}$. But Z is not connected, since $\{I_n\}$ is a normal subgroup of finite index.

12.1.2 The classical groups

A list of all semisimple linear algebraic groups over $\mathbb{C}$ has been known since the early part of the twentieth century. There are four infinite families known as the classical groups, and five exceptional groups, which we will not describe here. When $\mathbb{C}$ is replaced by a Galois field, all the classical and exceptional groups become what are called groups of Lie type. With the exception of 26 so-called sporadic groups, all the finite simple groups are of this type. We have already encountered three of the families of classical groups: $SL(n,\mathbb{C})$ (the special linear groups), $SO(2n,\mathbb{C})$ (the even orthogonal groups), and $SO(2n+1,\mathbb{C})$ (the odd orthogonal groups). It may seem artificial to distinguish the even and odd special orthogonal groups, but the mechanism used to classify the semisimple groups forces us to put these groups into different categories. The fourth family consists of the *symplectic groups* $Sp(2n,\mathbb{C})$. By definition,

$$Sp(2n,\mathbb{C}) = \{A \in GL(2n,\mathbb{C}) \mid A\Omega A^T = \Omega\},$$

where

$$\Omega = \begin{pmatrix} O & I_n \\ -I_n & O \end{pmatrix}.$$

If $A \in Sp(2n,\mathbb{C})$, then $A^{-1} = \Omega A^T \Omega^{-1}$, and since $\Omega^2 = -I_{2n}$, we get $A^{-1} = -\Omega A^T \Omega$. Also, by definition, $\det(AA^T) = 1$, so $\det(A) = \pm 1$. In fact, it can be shown that $\det(A) = 1$, so $Sp(2n,\mathbb{C}) < SL(n,\mathbb{C})$. This follows by showing that $Sp(2n,\mathbb{C})$ is connected, since the kernel of det is a normal subgroup. The symplectic groups are used in symplectic geometry, physics, and the theory of alternating quadratic forms.

12.1.3 Algebraic tori

A linear algebraic group $T < GL(n,\mathbb{C})$ is called an *algebraic torus*, or simply a *torus*, if T is connected and abelian and every element of T is semisimple. Algebraic tori are basic examples of reductive groups. This follows from the multiplicative Jordan–Chevalley decomposition, since the only unipotent element of a torus is the identity. If T is a subgroup of a linear algebraic group $G < GL(n,\mathbb{C})$, we call T a *subtorus* of G if T is an algebraic torus in $GL(n,\mathbb{C})$. A subtorus T of G that is not a proper subgroup of any other subtorus in G is called a *maximal torus* in G. For example, $\mathcal{D}(n,\mathbb{C})$ is a maximal torus in $GL(n,\mathbb{C})$. More generally, we have the following.

Proposition 12.1. *Every maximal torus $T < GL(n,\mathbb{C})$ is conjugate to $\mathcal{D}(n,\mathbb{C})$.*

Exercise 12.1.2. Show that $T = \left\{ \begin{pmatrix} a & b \\ -b & a \end{pmatrix} \mid a, b \in \mathbb{C},\ a^2 + b^2 \neq 0 \right\}$ and $\mathcal{D}(2, \mathbb{C})$ are both maximal tori in $GL(2, \mathbb{C})$, and prove that they are conjugate in $GL(2, \mathbb{C})$.

Exercise 12.1.3. Let T be the algebraic torus defined in the previous exercise. Show that $\mathcal{D}(2, \mathbb{R})$ and $T \cap GL(2, \mathbb{R})$ are not conjugate in $GL(2, \mathbb{R})$.

We next state an important fact.

Theorem 12.2. *Every reductive linear algebraic group G contains a maximal torus, and any two maximal tori in G are conjugate by an element of G. Moreover, if $T < G$ is an algebraic torus, then T is contained in a maximal torus.*

Let us now describe some maximal tori for classical groups.

Case 1: $SL(n, \mathbb{C})$. This is the easiest case to describe. In fact, the subgroup of all diagonal matrices in $SL(n, \mathbb{C})$, that is, $SL(n, \mathbb{C}) \cap \mathcal{D}(n, \mathbb{C})$, is a maximal torus.

Case 2: $SO(2n, \mathbb{C})$. Starting with the maximal torus $T = \{\mathrm{diag}(t, t^{-1}) \mid t \in \mathbb{C}^*\}$ in $SL(2, \mathbb{C})$, one can conjugate T into $SO(2, \mathbb{C})$, since

$$R = P\mathrm{diag}(t, t^{-1})P^{-1} = \begin{pmatrix} a & b \\ -b & a \end{pmatrix},$$

where $a = (t^2 + 1)/t$, $b = (-it^2 + 1)/t$, and $P = \dfrac{1}{\sqrt{2}} \begin{pmatrix} 1 & i \\ i & 1 \end{pmatrix}$. We will call R a rotation matrix. It turns out that the set of all matrices of the form

$$\mathcal{R} = \begin{pmatrix} R_1 & O & \cdots & O \\ O & R_2 & \cdots & O \\ \vdots & \vdots & \ddots & \vdots \\ O & O & \cdots & R_n \end{pmatrix},$$

where $R_1, \ldots, R_n$ are arbitrary rotation matrices, is a maximal torus in $SO(2n, \mathbb{C})$.

Case 3: $SO(2n + 1, \mathbb{C})$. To obtain a maximal torus in $SO(2n + 1, \mathbb{C})$, consider the natural inclusion homomorphism $i : SO(2n, \mathbb{C}) \to SO(2n + 1, \mathbb{C})$ defined by

$$R \to \begin{pmatrix} R & \mathbf{0} \\ \mathbf{0} & 1 \end{pmatrix}.$$

Then the image of $\mathcal{R}$ is a maximal torus in $SO(2n + 1, \mathbb{C})$.

Case 4: $Sp(2n, \mathbb{C})$. Finally, the set of all $T = \mathrm{diag}(x_1, \ldots, x_n, x_1^{-1}, \ldots, x_n^{-1})$, where each x_i is in $\mathbb{C}^*$, is a maximal torus for $Sp(2n, \mathbb{C})$.

A maximal torus T in a linear algebraic group G defines a group action of T on G by conjugation, namely $(t, g) \to t \cdot g = tgt^{-1}$. This action gives rise to the so-called roots and weights of the pair (G, T), which completely determine G up to isomorphism (in the sense of linear algebraic groups) when G is semisimple. Space does not permit us to elaborate on this theme, since it requires introducing the Lie algebra of G. Instead, we will concentrate on another group that we have already seen in a special case. This is the Weyl group of (G, T). We will then discuss the role of the Weyl group in the theory of linear algebraic groups.

12.1.4 The Weyl group

The *Weyl group* of a pair (G, T) consisting of a linear algebraic group G and a maximal torus $T < G$ is the group $W(G, T) = N_G(T)/T$. We will see below that if $G = GL(n, \mathbb{C})$ and $T = \mathcal{D}(n, \mathbb{C})$, then $W(G, T)$ is isomorphic to the group $P(n)$ of $n \times n$ permutation matrices. Recall that $P(n)$ is a basic component of the $LPDU$ decomposition in $GL(n, \mathbb{C})$. The matrices P and D are always unique, though L and U need not be. We will give some examples of Weyl groups and then explain how they play a role similar to that of $P(n)$ for $GL(n, \mathbb{C})$ for arbitrary reductive linear algebraic groups. Let us begin with a fundamental result.

Theorem 12.3. *If G is reductive, then the normalizer $N_G(T)$ is also a linear algebraic group, and the Weyl group $W(G, T)$ is finite.*

For simplicity, we will denote $W(G, T)$ by W as long as it is clear what G and T are. Since any two maximal tori in G are conjugate, the Weyl group of (G, T) is independent of the choice of T up to an isomorphism induced by an inner automorphism of G. The following result computes W when G is either $GL(n, \mathbb{C})$ or $SL(n, \mathbb{C})$.

Proposition 12.4. *If $G = GL(n, \mathbb{C})$ or $SL(n, \mathbb{C})$, then $W \cong S(n)$.*

Proof. Suppose first that $G = GL(n, \mathbb{C})$ and $T = \mathcal{D}(n, \mathbb{C})$. We will show that $N_G(T)$ is the semidirect product $P(n)T$. Note that if $P \in P(n)$ and $D = \mathrm{diag}(d_1, \ldots, d_n)$, then $PDP^{-1} = \mathrm{diag}(d_{\sigma(1)}, \ldots, d_{\sigma(n)})$, where $\sigma \in S(n)$ is the unique permutation such that $P = P_\sigma$. To see this, it suffices to check that $PDP^{-1} = \mathrm{diag}(d_{\sigma(1)}, \ldots, d_{\sigma(n)})$ whenever $\sigma = (i\ i+1)$ is a simple transposition. Since $(i\ i+1)$ is conjugate to (12), it suffices to let $\sigma = (12)$ and assume that D is of size 2×2. Hence we get the result, since

$$PDP^{-1} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} b & 0 \\ 0 & a \end{pmatrix}.$$

Since $P(n) \cap T = (1)$, the subgroup H generated by $P(n)$ and T is the semi-direct product of $P(n)$ and T, and $H \leq N_G(T)$. Moreover, transvections, or elementary matrices of type III, do not normalize T. For example, in the 2×2 case,

$$\begin{pmatrix} 1 & 0 \\ u & 1 \end{pmatrix} \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -u & 1 \end{pmatrix} = \begin{pmatrix} a & 0 \\ u(a-b) & b \end{pmatrix}.$$

It follows that $N_G(T) = H = P(n)T$, as claimed. It follows from Exercise 11.4.7 that $W \cong P(n)$ via the natural homomorphism $P(n) \to W$ defined by $P \to PT$. Hence $W \cong S(n)$ also. When $G = SL(n, \mathbb{C})$, the above argument doesn't work, because if $P = P_\sigma$ and $\operatorname{sgn}(\sigma) = -1$, then $P \notin SL(n, \mathbb{C})$, since $\det(P_\sigma) = \operatorname{sgn}(\sigma)$. This is fixed by replacing a P_σ with $\operatorname{sgn}(\sigma) = -1$ by the matrix RP_σ, where $R = \operatorname{diag}(-1, 1, \ldots, 1)$. It is clear that conjugation by RP_σ has the same effect as conjugation by P_σ, so $W \cong S(n)$ in the $SL(n, \mathbb{C})$ case also. □

We will skip the details for the computation of the Weyl groups of the remaining classical groups. The complete result is contained in the following theorem.

Theorem 12.5. *The Weyl groups of the classical groups are as follows:*

(i) for $G = GL(n, \mathbb{C})$ or $SL(n, \mathbb{C})$, $W \cong S(n)$;

(ii) for $G = SO(2n+1, \mathbb{C})$ or $SP(2n, \mathbb{C})$, $W \cong SP(n)$, where $SP(n)$ denotes the group of signed permutation matrices; and

(iii) for $G = SO(2n, \mathbb{C})$, $W \cong SP(n)^+$, the subgroup of $SP(n)$ consisting of all signed permutation matrices having an even number of minus signs.

The group $SP(n)$ consists of all elements of $O(n, \mathbb{C})$ having integer entries. In other words, every element of $SP(n)$ is obtained from a permutation matrix P by allowing ± 1 wherever a 1 occurs in P. Thus the order of $SP(n)$ is $2^n n!$. Since $\det(P_\sigma) = \operatorname{sgn}(\sigma)$ for the permutation matrix P_σ associated to $\sigma \in S(n)$, it follows that $SP(n)^+ = SP(n) \cap SL(n, \mathbb{C})$. Thus $|SP(n)^+| = 2^{n-1} n!$.

Each of the Weyl groups above is generated by reflections of $\mathbb{R}^n$. We will give a minimal set of reflections that generate in each case. Such reflections are called *simple*.

Case 1: $W = P(n)$. For each $i = 1, \ldots, n-1$, let H_i denote the reflection through the hyperplane orthogonal to $\mathbf{e}_i - \mathbf{e}_{i+1}$. The matrix of H_i with respect to the standard basis is the row swap matrix P_i obtained from I_n by swapping the ith and $(i+1)$st rows. Then $P_1, \ldots, P_{n-1}$ are the simple reflections generating W. Note that P_i is the permutation matrix corresponding to the simple transposition $(i\ i+1)$.

Case 2: $W = SP(n)$. Here the simple reflections are $P_1, \ldots, P_{n-1}$ together with the reflection $P_n = \operatorname{diag}(1, \ldots, 1, -1)$, which reflects $\mathbb{R}^n$ through the hyperplane $x_n = 0$.

Case 3: $W = SP(n)^+$. In the final case, the simple reflections consist of $P_1, \ldots, P_{n-1}$ together with the reflection H_n through the hyperplane orthogonal to $\mathbf{e}_{n-1} + \mathbf{e}_n$. The matrix of H_n is $-P_{n-1}$. Thus the simple reflections are $P_1, \ldots, P_{n-2}, \pm P_{n-1}$.

The simple reflections listed above arise from the root systems of types A, B, C, D in the theory of reflection groups (that is, groups generated by reflections). The simple reflections allow one to define the notion of the *length* of an element w of W as the minimal k such that there exists an expression $w = P_{i_1} \cdots P_{i_k}$ with each P_j a simple reflection. The length of w is denoted by $\ell(w)$. We will mention below an interesting role that the length function plays. An excellent source for information and further study on reflection groups, including their root systems, is Reflection Groups and Coxeter Groups (Cambridge Studies in Advanced Mathematics), by James E. Humphreys.

12.1.5 Borel subgroups

If G is a connected linear algebraic group, a closed, connected subgroup $\mathcal{B}$ of G that is solvable and not properly contained in any other closed, connected solvable subgroup of G is called a *Borel subgroup* of G. Borel subgroups play a very important role in the structure theory of linear algebraic groups, as we will presently see. We have already seen that the upper triangular subgroups $\mathcal{T}(n, \mathbb{C})$ and $\mathcal{T}(n, \mathbb{C}) \cap SL(n, \mathbb{C})$ of $GL(n, \mathbb{C})$ and $SL(n, \mathbb{C})$ are solvable. It turns out that they are Borel subgroups of $GL(n, \mathbb{C})$ and $SL(n, \mathbb{C})$ respectively. This is easy to see for $n = 2$. The commutator $ABA^{-1}B^{-1}$ of $A, B \in GL(2, \mathbb{C})$ is in general not in $\mathcal{T}(2, \mathbb{C})$. In fact, let G be a closed subgroup of $GL(2, \mathbb{C})$ properly containing $\mathcal{T}(2, \mathbb{C})$. Then there exists $A \in G$ such that the entry a_{21} is nonzero. Then one can find $B \in \mathcal{T}(2, \mathbb{C})$ such that $C = ABA^{-1}B^{-1}$ has a nonvanishing $(2, 1)$ entry. For example, assuming $ad - bc = 1$ to avoid denominators, then

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 - ac & a^2 + ac - 1 \\ -c^2 & 1 + ac + c^2 \end{pmatrix}.$$

It follows that every commutator subgroup $G^{(i)}$ is nontrivial, so G cannot be solvable. Hence, $\mathcal{T}(2, \mathbb{C})$ is a Borel subgroup of $GL(2, \mathbb{C})$. The reader can then extend this reasoning to conclude that $\mathcal{T}(n, \mathbb{C})$ is a Borel subgroup of $GL(n, \mathbb{C})$. Similarly, $\mathcal{T}(n, \mathbb{C}) \cap SL(n, \mathbb{C})$ is a Borel subgroup of $SL(n, \mathbb{C})$.

The existence of a Borel subgroup $\mathcal{B}$ in a linear algebraic group G follows from the fact that an algebraic torus is a connected solvable (in fact, abelian) linear algebraic group. Now, if H is any closed connected solvable linear algebraic group such that $T \leq H \leq G$, then either H is maximal or there

exists a closed connected solvable linear algebraic group H' of G such that $H < H'$. We need to know that iterating this remark will produce a maximal closed connected solvable linear algebraic group $\mathcal{B} \leq G$ such that $T \leq \mathcal{B}$. In fact, if $G < GL(n, \mathbb{C})$, then there cannot be a strictly increasing sequence of linear algebraic groups

$$H_1 < H_2 < \cdots < H_k < H_{k+1} < \cdots < G.$$

The tool needed to guarantee this is a fundamental theorem in abstract algebra known as the Hilbert basis theorem, which is a result about ideals in $\mathbb{F}[x_1, \ldots, x_m]$. The existence of a Borel subgroup $\mathcal{B}$ such that $T \leq \mathcal{B}$ thus follows. It can also be shown that every Borel subgroup in a linear algebraic group $G < GL(n, \mathbb{C})$ can be obtained as $G \cap \mathcal{B}$ for some Borel $\mathcal{B}$ in $GL(n, \mathbb{C})$.

Remark. The Borel subgroups of the orthogonal groups and the symplectic groups are harder to describe, but a nice description is given in the abovementioned book by Malle and Testerman on page 38.

12.1.6 The conjugacy of Borel subgroups

Every subgroup $H < GL(n, \mathbb{C})$ conjugate to a linear algebraic group $G < GL(n, \mathbb{C})$ is also a linear algebraic group. For if $H = gGg^{-1}$, and $p(X)$ is a polynomial on $\mathbb{C}^{n \times n}$ such that $p(X) = 0$ if $X \in G$, then the function $q(X) = p(g^{-1}Xg)$ is also a polynomial on $\mathbb{C}^{n \times n}$ such that $q(Y) = 0$ if $Y \in H$, since $q(Y) = p(g^{-1}Yg) = p(g^{-1}gXg^{-1}g) = p(X) = 0$. It follows that every subgroup of G that is conjugate to a Borel subgroup is itself a Borel subgroup. Thus the lower triangular matrices also give Borel subgroups in $GL(n, \mathbb{C})$, and those in $SL(n, \mathbb{C})$ give a Borel subgroup in $SL(n, \mathbb{C})$. Moreover, there are also Borel subgroups that are quite hard to describe.

The Borel subgroups of a reductive linear algebraic group G have two fundamental and deep properties: any two Borel subgroups of G are conjugate by an element of G, and the normalizer of every Borel subgroup of $\mathcal{B} < G$ is $\mathcal{B}$ itself (cf. Malle and Testerman). These two facts give rise to a nice parameterization of the set of Borel subgroups of G.

Proposition 12.6. *Let $\mathcal{B}_G$ denote the set of all Borel subgroups of G, and let $\mathcal{B}$ denote a fixed Borel subgroup. Then the mapping $g\mathcal{B} \to g\mathcal{B}g^{-1}$ defines a bijection between the space of left cosets $G/\mathcal{B}$ and $\mathcal{B}_G$.*

Recall that we showed that every maximal torus T in G is contained in a Borel subgroup $\mathcal{B}$ of G. We just remarked that $N_G(\mathcal{B}) = \mathcal{B}$ for every Borel subgroup, and in fact, $N_G(T) \cap \mathcal{B} = T$. Since $T < \mathcal{B}$, it makes sense to define $w\mathcal{B}w^{-1}$ by $n_w \mathcal{B} n_w^{-1}$ for every representative n_w of w in the Weyl group W of (G, T). The interesting point is that $w\mathcal{B}w^{-1}$ is also a Borel subgroup of G

containing T. Moreover, if $w_1, w_2 \in W$, then $w_1 \mathcal{B} w_1^{-1} = w_2 \mathcal{B} w_2^{-1}$ if and only if $w_1 = w_2$.

Theorem 12.7. *The correspondence sending $w \in W$ to wBw^{-1} is a bijection from the Weyl group of (G,T) onto the set of Borel subgroups of G containing T. In particular, the number of Borel subgroups of G containing T is $|W|$.*

12.1.7 The flag variety of a linear algebraic group

If $\mathcal{B}$ is a Borel subgroup of G, then the coset space $G/\mathcal{B}$ turns out to have the structure of an algebraic variety, which is the fundamental concept in the field of algebraic geometry. This allows one to use results from algebraic geometry to study the abstract set $\mathcal{B}_G$ of all Borel subgroups in G. When $G = GL(n, \mathbb{C})$, we can explicitly describe $G/\mathcal{B}$ in terms from linear algebra. Let us suppose $\mathcal{B} = \mathcal{T}(n, \mathbb{C})$ and fix $A \in GL(n, \mathbb{C})$. Let U be an arbitrary element of $\mathcal{B}$. Since multiplication on the right by each U performs rightward column operations, the spans of the first k columns of A and AU are the same for all k. Let V_k denote this subspace. Since A is invertible, $\dim V_k = k$ for all k. Thus the coset $A\mathcal{B}$ uniquely determines a strictly increasing sequence of subspaces of $\mathbb{C}^n$, namely

$$V_1 \subset V_2 \subset \cdots \subset V_{n-1} \subset \mathbb{C}^n. \tag{12.1}$$

Conversely, every such sequence uniquely determines a coset of $G/\mathcal{B}$. This gives us a bijection from $G/\mathcal{B}$ onto the set of all sequences of the form (12.1). These sequences are called *complete flags* in $\mathbb{C}^n$. The set of complete flags in $\mathbb{C}^n$ is often denoted by $\mathrm{Flag}(\mathbb{C}^n)$. It has many applications in geometry and related areas. If G is an arbitrary reductive group and $\mathcal{B}$ a Borel subgroup, then the coset space $G/\mathcal{B}$ is known as the *flag variety* of G.

Now let T be a maximal torus in a reductive group G and let $\mathcal{B}$ be a Borel subgroup of G such that $T < \mathcal{B}$. Consider the action of T on the flag variety $G/\mathcal{B}$ by left translation given explicitly by $(t, g\mathcal{B}) \to tg\mathcal{B}$. As we have already seen in the proofs of the Sylow theorems, it is often very useful to know the fixed-point set of a group action.

Proposition 12.8. *The fixed-point set $(G/\mathcal{B})^T$ of the left multiplication action of T on $G/\mathcal{B}$ is precisely the set of cosets $w\mathcal{B}$ as w varies through the Weyl group. In particular, T has exactly $|W|$ fixed points on $G/\mathcal{B}$.*

Proof. Suppose $Tg\mathcal{B} = g\mathcal{B}$. Then $g^{-1}Tg < \mathcal{B}$, so $T < g\mathcal{B}g^{-1}$. Thus $g\mathcal{B}g^{-1}$ is a Borel subgroup of G containing T. Therefore, by Cartan's theorem, $g\mathcal{B}g^{-1} = w\mathcal{B}w^{-1}$ for some $w \in W$. Consequently, $g\mathcal{B} = w\mathcal{B}$ by Proposition 12.6. □

The fact that the number of fixed points of T on $G/\mathcal{B}$ is $|W|$ translates into a statement about the topology of the flag variety $G/\mathcal{B}$ of G, namely, the Euler characteristic of $G/\mathcal{B}$ is $|W|$, the order of the Weyl group. This is a famous theorem of André Weil that was discovered around 1930. The Euler characteristic is a topological invariant of a space that generalizes the number $F - E + V$, which measures the number of holes in a two-dimensional surface without boundary.

Example 12.1. It is interesting to compute $(G/\mathcal{B})^T$ when $G = GL(n, \mathbb{C})$ in two ways. First, let us view $GL(n, \mathbb{C})/\mathcal{B}$ as the complete flags in $\mathbb{C}^n$. Note that if $T = \mathcal{T}(n, \mathbb{C})$, then every flag of the form

$$\mathbb{C}\mathbf{e}_{i_1} \subset \mathrm{span}_{\mathbb{C}}\{\mathbf{e}_{i_1}, \mathbf{e}_{i_2}\} \subset \cdots \subset \mathrm{span}_{\mathbb{C}}\{\mathbf{e}_{i_1}, \mathbf{e}_{i_2}, \ldots, \mathbf{e}_{i_{n-1}}\} \subset \mathbb{C}^n$$

is fixed by T. Here $i_1, i_2, \ldots, i_{n-1}$ are $n-1$ distinct integers in $[1, n]$. Since there are exactly $n!$ such flags, and $|S(n)| = n!$, these flags comprise the T-fixed points. Now let us compute $(G/\mathcal{B})^T$ another way, this time using the $LPDU$ decomposition. The points (or cosets) $P\mathcal{B}$, where $P \in P(n)$, are fixed under T, because $TP\mathcal{B} = PT\mathcal{B}$, since the normalizer in $GL(n, \mathbb{C})$ of T is $P(n)$. Thus, $P^{-1}TP = T$, so $TP = PT$, whence $TP\mathcal{B} = P\mathcal{B}$. Hence, the points $P\mathcal{B}$ as P ranges over $P(n)$ are fixed under T. But in the $LPDU$ decomposition of an element of $GL(n, \mathbb{C})$, we know that P and D are unique. Thus if $P, P' \in P(n)$, then $P\mathcal{B} = P'\mathcal{B}$ if and only if $P = P'$. Therefore, since $|P(n)| = n!$, we have found $(GL(n, \mathbb{C})/\mathcal{B})^T$ in two (equivalent) ways. ⊠

12.1.8 The Bruhat decomposition of $GL(n, \mathbb{F})$

We are now going to discuss how the $LPDU$ decomposition generalizes to an arbitrary reductive linear algebraic group $G < GL(n, \mathbb{F})$, where $\mathbb{F}$ is an arbitrary field. We will begin by finding a slightly different version of $LPDU$ for $G = GL(n, \mathbb{F})$. According to $LPDU$, we can decompose $GL(n, \mathbb{F})$ as the product of four subgroups, namely

$$GL(n, \mathbb{F}) = \mathcal{L}(n, \mathbb{F})P(n)\mathcal{D}(n, \mathbb{F})\mathcal{U}(n, \mathbb{F}).$$

Of course, $\mathcal{D}(n, \mathbb{F})\mathcal{U}(n, \mathbb{F})$ is the upper triangular Borel subgroup $\mathcal{T}(n, \mathbb{F})$. Another Borel subgroup of $GL(n, \mathbb{F})$ is, in fact, the set $\mathcal{T}(n, \mathbb{F})^-$ of all lower triangular elements of $GL(n, \mathbb{F})$. Since $\mathcal{T}(n, \mathbb{F})$ and $\mathcal{T}(n, \mathbb{F})^-$ are Borel subgroups containing the maximal torus $\mathcal{D}(n, \mathbb{F})$, they are conjugate by an element of the Weyl group $P(n)$. In fact, this element is

$$P_0 = (\mathbf{e}_n \ \mathbf{e}_{n-1} \ \cdots \ \mathbf{e}_2 \ \mathbf{e}_1).$$

Put another way, if P_0 is the permutation matrix with ones on the antidiagonal, then $\mathcal{L}(n,\mathbb{F}) = P_0\mathcal{T}(n,\mathbb{F})P_0^{-1}$. Now $GL(n,\mathbb{F}) = P_0GL(n,\mathbb{F})$ and $P_0^{-1} = P_0$. Thus,

$$GL(n,\mathbb{F}) = P_0GL(n,\mathbb{F}) = P_0\mathcal{L}(n,\mathbb{F})P_0P_0P(n)\mathcal{T}(n,\mathbb{F}) = \mathcal{T}(n,\mathbb{F})P(n)\mathcal{T}(n,\mathbb{F}).$$

Recall that $P_\sigma = (\mathbf{e}_{\sigma(1)} \ \cdots \ \mathbf{e}_{\sigma(n)})$. Hence

$$GL(n,\mathbb{C}) = \bigcup_{\sigma\in S(n)} \mathcal{T}(n,\mathbb{F})P_\sigma\mathcal{T}(n,\mathbb{F}). \tag{12.2}$$

This is a double coset decomposition of $GL(n,\mathbb{C})$ in which the coset representatives come from the subgroup $P(n)$, which is isomorphic to $S(n)$. The double coset decomposition (12.2) is called the *Bruhat decomposition* of $GL(n,\mathbb{F})$. Applied to the flag variety of $GL(n,\mathbb{F})$, the Bruhat decomposition implies

$$\mathrm{Flag}(\mathbb{F}^n) = \bigcup_{\sigma\in S(n)} \mathcal{T}(n,\mathbb{F})\cdot P_\sigma\mathcal{T}(n,\mathbb{F}).$$

Put another way, $\mathrm{Flag}(\mathbb{F}^n)$ is the union of the $\mathcal{T}(n,\mathbb{F})$ orbits of the cosets $P_\sigma\mathcal{T}(n,\mathbb{F}) \in \mathrm{Flag}(\mathbb{F}^n)$. These points are also the fixed points of the action of $\mathcal{D}(n,\mathbb{F})$ on $\mathrm{Flag}(\mathbb{F}^n)$. The double cosets $\mathcal{T}(n,\mathbb{F})P_\sigma\mathcal{T}(n,\mathbb{F})$ in $GL(n,\mathbb{F})$ are called *Bruhat cells*. The $\mathcal{T}(n,\mathbb{F})$-orbits of the cosets $P_\sigma\mathcal{T}(n,\mathbb{F})$ in $\mathrm{Flag}(\mathbb{F}^n)$ are called *Schubert cells*. There are exactly $n!$ Bruhat cells and the same number of Schubert cells.

Let us now suppose $\mathbb{F} = \mathbb{C}$. Then the Bruhat cells have an interesting connection with the length function on the Weyl group $P(n)$. Suppose first that $w = P_\sigma$, and let the length be given by $\ell(w) = r$. This means that w has a minimal expression as $w = s_1s_2\cdots s_r$, where each s_i is one of the reflections $P_1,\ldots,P_{n-1}$ through the hyperplanes in $\mathbb{R}^n$ orthogonal to $\mathbf{e}_1-\mathbf{e}_2,\ldots,\mathbf{e}_{n-1}-\mathbf{e}_n$ respectively. The length $\ell(w)$ is also the minimal number of simple transpositions $(i\ i+1)$ needed for an expression of σ as a product of transpositions. (Note: if $w = 1$, then we agree that $\ell(w) = 0$.) It is a nice exercise to prove that if $w = P_0$, then $\ell(w) = n(n-1)/2$. Now the Borel $\mathcal{T}(n,\mathbb{C})$ is defined by setting the coordinate functions $x_{ij} = 0$ for $i > j$ in $GL(n,\mathbb{C})$. Intuitively, therefore, the dimension of $\mathcal{T}(n,\mathbb{C})$ is $n(n+1)/2$. In fact, we can parameterize $\mathcal{T}(n,\mathbb{C})$ using n free variables a_{1j}, $n-1$ free variables a_{2j}, and so on. Thus, the dimension of $\mathcal{T}(n,\mathbb{C})$ should be $n+(n-1)+(n-2)+\cdots+2+1 = n(n+1)/2$. Now, the dimension of $\mathcal{T}(n,\mathbb{C})P_0\mathcal{T}(n,\mathbb{C})$ is the same as the dimension of $P_0^{-1}\mathcal{T}(n,\mathbb{C})P_0\mathcal{T}(n,\mathbb{C}) = \mathcal{L}(n,\mathbb{C})\mathcal{T}(n,\mathbb{C})$. But recall that $\mathcal{L}(n,\mathbb{C})\mathcal{T}(n,\mathbb{C})$ consists of all $A \in GL(n,\mathbb{C})$ with $LPDU$ decomposition having $P = I_n$. In fact, we showed that

$$\mathcal{L}(n,\mathbb{C})\mathcal{T}(n,\mathbb{C}) = \{A \in GL(n,\mathbb{C}) \mid \det(A_i) \neq 0,\ i = 1,\ldots,n\},$$

where A_i is the $i \times i$ matrix in the upper left-hand corner of A. Furthermore, if $A \in \mathcal{L}(n,\mathbb{C})\mathcal{T}(n,\mathbb{C})$, then in the factorization $A = LDU$, L, D, and U are all unique. Now $\mathcal{L}(n,\mathbb{C})$ is described by $n(n-1)/2$ independent variables a_{ij}, where $i > j$, $\mathcal{D}(n,\mathbb{C})$ is described by n independent variables a_{ii}, and $\mathcal{U}(n,\mathbb{C})$ by another $n(n-1)/2$ independent variables a_{ij}, where $j > i$. Thus $\mathcal{L}(n,\mathbb{C})\mathcal{T}(n,\mathbb{C})$ is described by $n(n-1)/2 + n(n-1)/2 + n = n(n-1) + n = n^2$ independent variables. Similarly, if $P = I_n$, which corresponds to $P = P_w$ with $w = 1$, then $\dim \mathcal{T}(n,\mathbb{C})P\mathcal{T}(n,\mathbb{C}) = \dim \mathcal{T}(n,\mathbb{C})$. The general formula for the dimension of a Bruhat cell is as follows.

Proposition 12.9. *If $w = P_\sigma$, then the Bruhat cell $\mathcal{T}(n,\mathbb{C})w\mathcal{T}(n,\mathbb{C})$ has dimension $\ell(w) + \dim \mathcal{T}(n,\mathbb{C}) = \ell(w) + n(n+1)/2$. The dimension of the corresponding Schubert cell is $\ell(w)$.*

The very interesting proof is beyond the scope of this introduction. The reader may well want to verify it in some special cases. For example, when $n = 3$ and σ is the transposition (12), then $\mathcal{B}P_\sigma\mathcal{B}'$ has the form

$$\mathcal{B}P_\sigma\mathcal{B}' = \begin{pmatrix} a & b & * \\ 0 & c & * \\ 0 & 0 & d \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} r & s & * \\ 0 & t & * \\ 0 & 0 & u \end{pmatrix} = \begin{pmatrix} br & at+bs & * \\ cr & cs & * \\ 0 & 0 & du \end{pmatrix}, \tag{12.3}$$

where the asterisks stand for some entries in $\mathbb{C}$. The $(3,1)$ and $(3,2)$ entries of this matrix are zero, but the other entries are (essentially) not restricted except for the condition that the determinant of $\mathcal{B}P_\sigma\mathcal{B}'$ is nonzero. This (admittedly not rigorous) argument gives that the dimension of $\mathcal{B}P_\sigma\mathcal{B}'$ is 7, which has the form $\ell(12) + \dim \mathcal{B}$.

12.1.9 The Bruhat decomposition of a reductive group

We now have the necessary ingredients to generalize the Bruhat decomposition to an arbitrary reductive linear algebraic group G, namely a maximal torus T, the Weyl group $W = N_G(T)/T$, and a Borel subgroup $\mathcal{B}$ of G containing T. Recall that W is a finite group. It will play the role that $P(n)$ plays when $G = GL(n,\mathbb{C})$. However, W isn't in general a subgroup of G. But if $T \subset \mathcal{B}$, we can still multiply the identity coset $\mathcal{B}$ by a coset of T. So let $w = n_wT$, where $n_w \in N_G(T)$, and define $w\mathcal{B}$ to be the coset $n_w\mathcal{B}$. This is well defined, since two representatives of w differ by an element of T. Thus, the Bruhat cell $\mathcal{B}w\mathcal{B}$ in G is well defined: it is the union of the cosets $bn_w\mathcal{B}$, where b varies through $\mathcal{B}$. The double coset $\mathcal{B}w\mathcal{B}$ is called a *Bruhat cell* in G.

In the flag variety $G/\mathcal{B}$, the $\mathcal{B}$-orbit $\mathcal{B} \cdot w\mathcal{B}$ of the coset $w\mathcal{B}$ is called a *Schubert cell*. The $LPDU$ decomposition of $GL(n, \mathbb{C})$ generalizes as follows.

Theorem 12.10. *Let G be a reductive linear algebraic group over $\mathbb{C}$, $\mathcal{B}$ a Borel subgroup of G, and T a maximal torus in G such that $T < \mathcal{B}$. Then G is the union of the Bruhat cells $\mathcal{B}w\mathcal{B}$ as w varies through W, so $G = \mathcal{B}W\mathcal{B}$. Moreover, $N_G(T) \cap \mathcal{B} = T$; hence if $w \neq w'$ in W, then $w\mathcal{B} \neq w'\mathcal{B}$. Thus, the number of distinct Bruhat cells in G is the order of W. Finally, the dimension of the Bruhat cell $\mathcal{B}w\mathcal{B}$ is $\ell(w) + \dim \mathcal{B}$.*

12.1.10 Parabolic subgroups

A closed subgroup $\mathcal{P}$ of a linear algebraic group G is called *parabolic* if $\mathcal{P}$ contains a Borel subgroup. We will classify the parabolic subgroups of G after we give an example.

Example 12.2. Consider the standard case $G = GL(n, \mathbb{C})$, $\mathcal{B} = \mathcal{T}(n, \mathbb{C})$, and $T = \mathcal{D}(n, \mathbb{C})$. As we have seen, $W = S(n) \cong P(n)$. Now suppose $j + k = n$. Let $P(j, k) < P(n)$ denote the subgroup consisting of permutation matrices with block decomposition

$$P = \begin{pmatrix} P_1 & O \\ O & P_2 \end{pmatrix},$$

where $P_1 \in P(j)$ and $P_2 \in P(k)$. But $P = P_\tau$ for a unique $\tau \in S(n)$. Since $P_1 = P_\mu$ and $P_2 = P_\nu$ for unique $\mu \in S(j)$ and $\nu \in S(k)$, we can define an injective homomorphism $\varphi : S(j) \times S(k) \to S(n)$ by $\varphi(\mu \times \nu) = \tau$. Let $S(j, k) < S(n)$ denote the image of φ. Then $S(j, k) \cong P(j, k)$. Then $\mathcal{B}P(j, k)\mathcal{B}$ is the set of all matrices of the form

$$\begin{pmatrix} A & * \\ O & B \end{pmatrix},$$

where $A \in GL(j, \mathbb{C})$ and $B \in GL(k, \mathbb{C})$. Hence, $\mathcal{B}P(j, k)\mathcal{B}$ is a closed subgroup of $GL(n, \mathbb{C})$.

More generally, let $\mathcal{P}(j_1, \ldots, j_k)$ denote all matrices of the form

$$\begin{pmatrix} A_1 & * & \cdots & * \\ O & A_2 & \cdots & * \\ \vdots & \vdots & \ddots & \vdots \\ O & O & \cdots & A_k \end{pmatrix},$$

where $A_i \in GL(j_i, \mathbb{C})$ and $j_1 + \cdots + j_k = n$. Then the $\mathcal{P}(j_1, \ldots, j_k)$ are the parabolic subgroups of $GL(n, \mathbb{C})$ containing $\mathcal{T}(n, \mathbb{C})$. ⊠

Now suppose G is reductive and fix a maximal torus T and a Borel subgroup $\mathcal{B}$ such that $T < \mathcal{B} < G$. Let $\mathcal{P}$ be a parabolic subgroup in G containing $\mathcal{B}$, and define the Weyl group $W_{\mathcal{P}}$ to be $N_{\mathcal{P}}(T)/T$. Then we have the following theorem.

Theorem 12.11. *The parabolic subgroup $\mathcal{P}$ is a union of certain Bruhat cells. In particular, $\mathcal{P} = \mathcal{B}W_{\mathcal{P}}\mathcal{B}$. Moreover, the number of Bruhat cells in $\mathcal{P}$ is $|W_{\mathcal{P}}|$.*

Not every subgroup $Z < W$ gives a parabolic subgroup in this manner. For example, let $\sigma = (i\ j)$ be a transposition that is not simple; that is, $|i - j| > 1$. Then if $Z = \{I_n, P_\sigma\}$, $\mathcal{B}Z\mathcal{B}$ is not a subgroup of $GL(n, \mathbb{C})$. A natural question then is which subgroups of Weyl groups have the form $W_{\mathcal{P}}$ for some parabolic $\mathcal{P}$. In fact, these subgroups can be described in a simple way. An element $P \neq 1$ of W is called *simple* if $\mathcal{B}P\mathcal{B} \cup \mathcal{B}$ is a subgroup of G. Note that if P is simple, then $P^2 = 1$. (We leave the proof to the reader.) Hence the simple elements have order two. Let $S \subset W$ denote the set of all simple elements. In the standard case, the simple elements are the reflections P_σ, where σ is a simple transposition. The next result is a complete description of the parabolic subgroups of the reductive group G.

Theorem 12.12. *A subgroup $Z < W$ has the property that $\mathcal{B}Z\mathcal{B}$ is a parabolic subgroup of G such that $\mathcal{B} < \mathcal{B}Z\mathcal{B}$ if and only if there exists a subset J of S such that $Z =< \sigma \mid \sigma \in J >$. Moreover, every parabolic subgroup of G is conjugate in G to a parabolic $\mathcal{P}$ containing $\mathcal{B}$.*

Corollary 12.13. *The set of simple reflections in W generates W.*

Proof. Let Z be the subgroup of W such that $G = \mathcal{B}Z\mathcal{B}$. Since $G = \mathcal{B}W\mathcal{B}$, it follows that $W = Z$. □

Corollary 12.14. *Every parabolic subgroup of $GL(n, \mathbb{C})$ containing $\mathcal{T}(n, \mathbb{C})$ has the form $\mathcal{P}(j_1, \ldots, j_k)$ for some choice of the j_i.*

12.2 Linearly reductive groups

Let us now treat a new idea. Suppose $G < GL(n, \mathbb{F})$ is a linear algebraic group. How does one describe the G-sets in $V = \mathbb{F}^n$? This is one of the questions treated in an area known as representation theory. The main question is what one can say if we restrict the notion of G-sets to G-invariant subspaces. A subspace W of V that is also a G-set is called a *G-invariant subspace*. Two of the basic questions are which G-invariant subspaces have complementary G-invariant subspaces and which G-invariant subspaces do not have any nontrivial proper G-invariant subspaces.

12.2.1 Invariant subspaces

Suppose $G < GL(n, \mathbb{F})$ is a linear algebraic group, and let V denote $\mathbb{F}^n$. A G-invariant subspace W of V is called *G-irreducible* if there is no nontrivial G-invariant subspace U of W. We say that G is *linearly reductive* or *completely reducible* if whenever W is a nontrivial G-invariant subspace of V, there exists a G-invariant subspace U of V such that $V = U \oplus W$. If V has no proper G-invariant subspace except $\{\mathbf{0}\}$, then we say that G acts irreducibly on V. For example, $GL(n, \mathbb{F})$ acts irreducibly on $\mathbb{F}^n$.

Here are two basic examples. First of all, we have the following result.

Proposition 12.15. *If $T < GL(n, \mathbb{F})$ is an algebraic torus, then T is linearly reductive.*

Proof. (sketch) By definition, every element of T is semisimple, and since T is abelian, it follows that all elements of T are simultaneously diagonalizable. Thus, T is conjugate to a subgroup of $\mathcal{D}(n, \mathbb{F})$. It follows from this that T is linearly reductive. □

The second basic example is the group $P(n)$ of $n \times n$ permutation matrices acting on $\mathbb{R}^n$. Observe that the line $\ell = \mathbb{R}(\mathbf{e}_1 + \mathbf{e}_2 + \cdots + \mathbf{e}_n)$ is stable under $P(n)$, so the hyperplane H orthogonal to ℓ is also. This hyperplane has equation $x_1 + x_2 + \cdots + x_n = 0$ and is clearly invariant under $P(n)$.

Proposition 12.16. *The only nontrivial $P(n)$-invariant subspaces of $\mathbb{R}^n$ are H and ℓ. In particular, the action of $P(n)$ on H is irreducible.*

Exercise 12.2.1. Prove this proposition.

12.2.2 Maschke's theorem

Maschke's theorem says the following.

Theorem 12.17. *A finite group $G < GL(n, \mathbb{F})$ is linearly reductive if either $\mathbb{F}$ is of characteristic zero or $|G|$ is prime to the characteristic of $\mathbb{F}$.*

Proof. Let $V = \mathbb{F}^n$ and suppose W is a G-invariant subspace of V. Let $T : V \to W$ be any linear mapping such that $T(\mathbf{w}) = \mathbf{w}$ if $\mathbf{w} \in W$. We now alter T using an averaging trick. In order to do this, we will use the fact that $|G|$ is invertible in $\mathbb{F}$. This is guaranteed, since the characteristic of $\mathbb{F}$ either is zero or is prime to $|G|$. Let $\varphi : V \to W$ be defined by

$$\varphi(\mathbf{v}) = \frac{1}{|G|} \sum_{g \in G} g \circ T(g^{-1}(\mathbf{v})). \tag{12.4}$$

Then $\varphi(\mathbf{w}) = \mathbf{w}$ if $\mathbf{w} \in W$, and for every $h \in G$ and $\mathbf{v} \in V$, $\varphi(h(\mathbf{v})) = h(\varphi(\mathbf{v}))$. (The proof of this is left for the reader.) Now, $\ker \varphi \cap W = \{\mathbf{0}\}$, so by the rank–nullity theorem, $V = W \oplus \ker \varphi$. But $\ker \varphi$ is G-invariant, since if $\mathbf{v} \in \ker \varphi$, then $\varphi(g(\mathbf{v})) = g(\varphi(\mathbf{v})) = g\mathbf{0} = \mathbf{0}$. □

12.2.3 Reductive groups

When $G < GL(n, \mathbb{F})$ is linearly reductive, it follows that $V = \mathbb{F}^n$ admits a direct sum decomposition

$$V = W_1 \oplus W_2 \oplus \cdots \oplus W_k,$$

where each W_i is a G-irreducible subspace. We just saw that if the characteristic of $\mathbb{F}$ is zero, then every finite subgroup $G < GL(n, \mathbb{F})$ is linearly reductive. Reductive groups are fundamental partly because of the following classical result.

Theorem 12.18. *If $\mathbb{F}$ is algebraically closed of characteristic zero, then every reductive subgroup $G < GL(n, \mathbb{F})$ is linearly reductive.*

The mapping φ used in the proof of Maschke's theorem is known as a *Reynolds operator*. When G is not finite but $\mathbb{F} = \mathbb{C}$, the proof uses a Reynold's-type operator defined by the Haar integral over G. It was an open question until the 1980s whether reductive subgroups of $GL(n, \mathbb{F})$, $\mathbb{F}$ algebraically closed of positive characteristic, are linearly reductive. The answer is yes if the notion of linearly reductive is replaced by a slightly weaker notion.

12.2.4 Invariant theory

Suppose $G < GL(n, \mathbb{F})$, where $\mathbb{F}$ has characteristic zero. A polynomial $f \in \mathbb{F}[x_1, \ldots, x_n]$ is said to be G-invariant if f is constant on every G-orbit. In other words, for all $(a_1, \ldots, a_n) \in \mathbb{F}^n$, we have $f(g(a_1, \ldots, a_n)) = f(a_1, \ldots, a_n)$ for all $g \in G$. For example, if G acts on $\mathbb{F}^{n \times n}$ by conjugation, then the determinant $\det \in \mathbb{F}[x_{ij}]$, $1 \leq i, j \leq n$, is $GL(n, \mathbb{F})$-invariant. We remark that G-invariants need to be defined differently when $\mathbb{F}$ has positive characteristic.

In the nineteenth century, mathematicians who worked in the field of invariant theory concentrated on the problem of constructing invariants, especially fundamental invariants, namely G-invariant polynomials $f_1, \ldots, f_k$ such that every G-invariant f can be written uniquely as

$$f = \sum c_{\alpha_1, \ldots, \alpha_k} f_1^{\alpha_1} \cdots f_k^{\alpha_k}, \tag{12.5}$$

where all $c_{\alpha_1, \ldots, \alpha_k}$ are in $\mathbb{F}$. David Hilbert, one of the greatest mathematicians of the nineteenth and twentieth centuries, was the first to realize that a new approach had to be taken in order to further the field, and he proved in 1888 that such invariants must exist in a certain general setting without explicitly constructing them. This approach was initially condemned, but eventually it was redeemed by his famous 1893 paper that established bounds on the degrees of the (unknown) generators. Hilbert's paper not only revolutionized invariant theory, it established a new field called commutative algebra, which is still a very active area.

An important example illustrating invariant theory is the fundamental theorem on symmetric polynomials. A polynomial $f \in \mathbb{F}[x_1, \ldots, x_n]$ is said to be *symmetric* if

$$f(x_1, \ldots, x_n) = f(x_{\sigma(1)}, \ldots, x_{\sigma(n)})$$

for all $\sigma \in S(n)$. Since $S(n)$ and $P(n)$ are isomorphic via the isomorphism $\sigma \to P_\sigma$, symmetric polynomials are exactly the $P(n)$-invariant polynomials for the natural action of $P(n)$ on $\mathbb{F}^n$. Recall the elementary symmetric functions $\sigma_1, \sigma_2, \ldots, \sigma_n$, which were defined when we discussed the characteristic polynomial. Namely,

$$\sigma_1 = x_1 + \cdots + x_n, \quad \sigma_2 = \sum_{i<j} x_i x_j, \quad \ldots \quad , \sigma_n = x_1 \cdots x_n.$$

The fundamental theorem on symmetric polynomials is the following.

Theorem 12.19. *Let $\mathbb{F}$ be a field of characteristic zero. Then every symmetric polynomial $f \in \mathbb{F}[x_1, \ldots, x_n]$ can be expressed in exactly one way in the form $f = g(\sigma_1, \sigma_2, \ldots, \sigma_n)$ for some $g \in F[x_1, \ldots, x_n]$.*

This basic fact has a beautiful generalization known as the Chevalley–Shephard–Todd theorem. I will state a special case originally proved by Chevalley in 1955. By a *reflection* of $\mathbb{F}^n$, we mean an element of $GL(n,\mathbb{F})$ having order two that fixes pointwise a hyperplane in $\mathbb{F}^n$.

Theorem 12.20. *Assume that the field $\mathbb{F}$ has characteristic zero, and let $G < GL(n,\mathbb{F})$ be a finite group generated by reflections. Then there exist G-invariants $\tau_1,\dots,\tau_n \in \mathbb{F}[x_1,\dots,x_n]$ such that every G-invariant $f \in \mathbb{F}[x_1,\dots,x_n]$ can be expressed uniquely as $f = g(\tau_1,\dots,\tau_n)$ for some $g \in \mathbb{F}[x_1,\dots,x_m]$.*

Recall that Weyl groups are examples of finite groups generated by reflections. A very simple case of Chevalley's theorem is illustrated by the following example.

Example 12.3. Consider the group $SP(n)$ of $n \times n$ signed permutation matrices. Recall that a signed permutation matrix is an orthogonal matrix whose only entries are 0 and ± 1. The fundamental invariants of $SP(n)$ are easy to guess: they are $\tau_1 = x_1^2 + \cdots + x_n^2$, $\tau_2 = \sum_{i<j} x_i^2 x_j^2$, and in general, $\tau_k(x_1,\dots,x_n) = \sigma_k(x_1^2,\dots,x_n^2)$ for all $k = 1,\dots,n$. ⊠

Finally, let us mention a result that connects invariant theory and reductive groups.

Theorem 12.21. *Suppose $\mathbb{F}$ is algebraically closed and $G < GL(n,\mathbb{F})$ is reductive. Assume that $\mathbf{v} \neq \mathbf{0}$ is a vector in $\mathbb{F}^n$ such that $g(\mathbf{v}) = \mathbf{v}$ for all $g \in G$. Then there exists a G-invariant $f \in \mathbb{F}[x_1,\dots,x_n]$ such that $f(\mathbf{v}) \neq 0$ but $f(\mathbf{0}) = 0$.*

This implies, for example, that there exist invariant polynomials $f_1,\dots,f_k$ such that every G-invariant f can be represented as in (12.5). Invariant theory for reductive groups is now called geometric invariant theory, or GIT for short. GIT was founded in a famous 1965 paper by David Mumford in which he conjectured the above theorem.

Bibliography

Groups and Fields

Alperin, J. L.; Bell, Rowen B. *Groups and representations*. Graduate Texts in Mathematics, 162. Springer-Verlag, New York, 1995.
Artin, Michael *Algebra*. Prentice Hall, Inc., Englewood Cliffs, NJ, 1991.
Dummit, David S.; Foote, Richard M. *Abstract algebra*. Third edition. John Wiley and Sons, Inc., Hoboken, NJ, 2004.
Herstein, I. N. *Abstract algebra*. Third edition. With a preface by Barbara Cortzen and David J. Winter. Prentice Hall, Inc., Upper Saddle River, NJ, 1996.
Humphreys, John F. *A course in group theory.* Oxford Science Publications. The Clarendon Press, Oxford University Press, New York, 1996.
Rotman, Joseph J. *An introduction to the theory of groups*. Fourth edition. Graduate Texts in Mathematics, 148. Springer-Verlag, New York, 1995.
van der Waerden, B. L. *Modern Algebra. Vol. I.* Translated from the second revised German edition by Fred Blum. With revisions and additions by the author. Frederick Ungar Publishing Co., New York, N. Y., 1949.

Matrix Theory

Gantmacher, F. R. *The theory of matrices. Vol. 1*. Translated from the Russian by K. A. Hirsch. Reprint of the 1959 translation. AMS Chelsea Publishing, Providence, RI, 1998.
Herstein, I. N.; Winter, David J. *Matrix theory and linear algebra*. Macmillan Publishing Company, New York; Collier Macmillan Publishers, London, 1988.
Strang, Gilbert *Linear algebra and its applications*. Second edition. Academic Press [Harcourt Brace Jovanovich, Publishers], New York-London, 1980.

Determinants

Lang, Serge *Linear algebra*. Reprint of the third edition. Undergraduate Texts in Mathematics. Springer-Verlag, New York, 1989.

J.B. Carrell, *Groups, Matrices, and Vector Spaces*,
DOI 10.1007/978-0-387-79428-0

Muir, Thomas *A treatise on the theory of determinants*. Revised and enlarged by William H. Metzler Dover Publications, Inc., New York 1960.
Shilov, Georgi E. *Linear algebra*. Revised English edition. Translated from the Russian and edited by Richard A. Silverman. Dover Publications, Inc., New York, 1977.
Turnbull, H. W. *The theory of determinants, matrices, and invariants*. 3rd ed. Dover Publications, Inc., New York 1960.

Vector Spaces

Artin, Michael *Algebra*. Prentice Hall, Inc., Englewood Cliffs, NJ, 1991.
Birkhoff, Garrett; MacLane, Saunders *A Survey of Modern Algebra*. Macmillan Company, New York, 1941.
Halmos, Paul R. *Finite-dimensional vector spaces*. Reprinting of the 1958 second edition. Undergraduate Texts in Mathematics. Springer-Verlag, New York-Heidelberg, 1974.
Hoffman, Kenneth; Kunze, Ray *Linear algebra*. Second edition Prentice-Hall, Inc., Englewood Cliffs, N.J. 1971.
Lang, Serge *Linear algebra*. Reprint of the third edition. Undergraduate Texts in Mathematics. Springer-Verlag, New York, 1989.
Samelson, Hans *An introduction to linear algebra*. Pure and Applied Mathematics. Wiley-Interscience [John Wiley and Sons], New York-London-Sydney, 1974.
Herstein, I. N.; Winter, David J. *Matrix theory and linear algebra*. Macmillan Publishing Company, New York; Collier Macmillan Publishers, London, 1988.

Linear Transformations

Coxeter, H. S. M. *Introduction to geometry.* Reprint of the 1969 edition. Wiley Classics Library. John Wiley & Sons, Inc., New York, 1989.
Gelfand, I. M. *Lectures on linear algebra*. With the collaboration of Z. Ya. Shapiro. Translated from the second Russian edition by A. Shenitzer. Reprint of the 1961 translation. Dover Books on Advanced Mathematics. Dover Publications, Inc., New York, 1989.
Herstein, I. N.; Winter, David J. *Matrix theory and linear algebra*. Macmillan Publishing Company, New York; Collier Macmillan Publishers, London, 1988.
Weyl, H. *Symmetry.* Reprint of the 1952 original. Princeton Science Library. Princeton University Press, Princeton, NJ, 1989.

Eigentheory

Lax, Peter D. *Linear algebra*. Pure and Applied Mathematics (New York). A Wiley-Interscience Publication. John Wiley and Sons, Inc., New York, 1997.
Strang, Gilbert *Linear algebra and its applications*. Second edition. Academic Press [Harcourt Brace Jovanovich, Publishers], New York-London, 1980.

Unitary Diagonalization and Quadratic Forms

Gelfand, I. M. *Lectures on linear algebra*. With the collaboration of Z. Ya. Shapiro. Translated from the second Russian edition by A. Shenitzer. Reprint of the 1961 translation. Dover Books on Advanced Mathematics. Dover Publications, Inc., New York, 1989.
Samelson, Hans *An introduction to linear algebra*. Pure and Applied Mathematics. Wiley-Interscience [John Wiley and Sons], New York-London-Sydney, 1974.
Strang, Gilbert *Linear algebra and its applications*. Second edition. Academic Press [Harcourt Brace Jovanovich, Publishers], New York-London, 1980.

Theory of Linear Mappings

Birkhoff, Garrett; MacLane, Saunders *A Survey of Modern Algebra*. Macmillan Company, New York, 1941.
Gelfand, I. M. *Lectures on linear algebra*. With the collaboration of Z. Ya. Shapiro. Translated from the second Russian edition by A. Shenitzer. Reprint of the 1961 translation. Dover Books on Advanced Mathematics. Dover Publications, Inc., New York, 1989.
Hoffman, Kenneth; Kunze, Ray *Linear algebra*. Second edition Prentice-Hall, Inc., Englewood Cliffs, N.J. 1971.
Strang, Gilbert *Linear algebra and its applications*. Second edition. Academic Press [Harcourt Brace Jovanovich, Publishers], New York-London, 1980.

Linear Algebraic Groups

Alperin, J. L.; Bell, Rowen B. *Groups and representations*. Graduate Texts in Mathematics, 162. Springer-Verlag, New York, 1995.
Dieudonné, Jean: Carrell, James B. *Invariant Theory, Old and New*. Academic Press, New York-London, 1971.
Humphreys, James E. *Linear algebraic groups*. Graduate Texts in Mathematics, No. 21. Springer-Verlag, New York-Heidelberg, 1975.
Humphreys, James E. *Reflection groups and Coxeter groups*. Cambridge Studies in Advanced Mathematics, 29. Cambridge University Press, Cambridge, 1990.
Malle, Gunter; Testerman, Donna *Linear algebraic groups and finite groups of Lie type*. Cambridge Studies in Advanced Mathematics, 133. Cambridge University Press, Cambridge, 2011.

Index

J.B. Carrell, *Groups, Matrices, and Vector Spaces*,
DOI 10.1007/978-0-387-79428-0