FUNDAMENTALS OF LINEAR ALGEBRA

James B. Carrell
carrell@math.ubc.ca

(July, 2005)

Contents

Chapter 1

Introduction

This textbook is meant to be a mathematically complete and rigorous introduction to abstract linear algebra for undergraduates, possibly even first year students, specializing in mathematics.

Linear algebra is one of the most applicable areas of mathematics. It is used by the pure mathematician and by the mathematically trained scientists of all disciplines. This book is directed more at the former audience than the latter, but it is hoped that the writing is sufficiently clear with enough detail so that the anyone reading the text can understand it. While the book is written in an informal style and has many elementary examples, the propositions and theorems are generally carefully proved, and the interested student will certainly be able to experience the theorem-proof style of text.

We have throughout tried very hard to emphasize the fascinating and important interplay between algebra and geometry. The exercises are also intended to emphasize this aspect. Some of them are very easy, some are medium hard and a few are quite challenging . The hope is that the student will find them to be stimulating and a reason to think deeply about the material.

The first two Chapters of the text cover standard beginning topics in linear algebra: matrices, linear systems, Gaussian elimination, inverses of matrices and the LDU decomposition. In this material, we manage to define the notion of a matrix group and give several examples, such as the general linear group, the orthogonal group and the group of $n \times n$ permutation matrices. In Chapter 3, we define the notion of a field and construct the prime fields $\mathbb{F}_p$ as examples that will be used later on. We then introduce the notion of an abstract vector space over an arbitrary field and discuss

some important speial cases, including $\mathbb{R}^n$ and $\mathbb{C}^n$ as inner product spaces.

The fourth Chapter is devoted to proving the fundamental theorems on finite dimensional vector spaces. We also treat the properties of dimension. We also prove the Hausdorff Intersection Theorem and use it to treat direct sums. We next construct the quotient of a vector space by a subspace. Direct sums and quotients are used later when we begin to study eigentheory in earnest.

Chapter 5 is an introduction to linear coding theory. The basic results about error correcting codes are proven, and we treat perfect codes in some detail. As well as being a timely subject, the topic of linear coding theory illustrates as well as anything I know of how powerful and useful the results of elementary linear algebra really are.

In Chapter 6, we discuss linear transformations. We show how to associate a matrix to a linear transformation (depending on a choice of bases) and prove that two matrices representing a linear transformation from a space to itself are similar. We also define the notion of an eigenpair and what is meant by a semi-simple linear transformation.

The next topic we cover is the theory of determinants. We give a rigorous definition and derivation of the basic properties of det, starting from the classical definition of the determinant of a matrix as an alternating sum (thereby avoiding the use of the Laplace expansion). A number of geometric applications are also given.

Chapter 8 covers the essential results about eigen-theory. We study the usual notions: the characteristic polynomial, eigenspaces and so on. Having introduced direct sums, we are able to show that a linear transformation with the same domain and target is semi-simple if and only if the dimensions of its eigenspaces add up to the dimension of the domain. Furthermore, having introduced quotients, we are able to show that every $n \times n$ matrix over the complex numbers is similar to an upper triangular matrix; or, equivalently, every linear transformation admits a flag basis. As a corollary, we show that the geometric multiplicity of an eigenvalue is at most the algebraic multiplicity, a result that is not easy to show from scratch. To concluse, we state the Cayley-Hamilton Theorem and give a partial proof. The full proof is given in Chapter 13.

Chapter 9 treats inner product spaces. For the most part, we concentrate on real and complex n-space, $\mathbb{R}^n$ and $\mathbb{C}^n$ and treat some of the standard topics such as least squares, projections, pseudo-inverses and orthonormal bases. We discuss Graham-Schmidt orthogonalization and derive the basic QR-factorization. In the last section, we show that the matrix group $SO(3)$ is exactly the set of all rotations of $\mathbb{R}^3$ (in the sense of Euler). We also

compute the rotations of a cube and regular octahedron.

In Chapter 10, we classify the complex matrices that admit orthogonal diagonalization: i.e. the normal matrices (those which commute with their Hermitian transpose). The fundamental results, such as the Principle Axis Theorem, concerning self adjoint operators on a finite dimensional vector space, follow easily, but we also give a geometric treatment of the Principle Axis Theorem because of the fundamental nature of this result.

Chapter 11 is devoted to applications, particularly of symmetric matrices. We introduce quadratic forms, prove Sylvester's Theorem and derive the relationship between the signs of pivots and the number of positive eigenvalues. We then consider some graph theory, and study the adjacency matrix of a graph. Finally, we consider the QR-algorithm for approximating eigenvalues and the power method, which is the QR-algorithm on the space of complete flags.

In Chapter 13, we polish off the theory of linear transformations by proving the most fundamental result about linear transformations, namely the Jordan Decomposition Theorem. We then obtain the Cayley-Hamilton Theorem, which is an easy consequence. Finally, we give a short discussion of the Jordan Normal Form of a linear transformation or matrix, tying it in with the notion of a flag basis introduced in Chapter 8.

I would like to thank Kai Behrend for several helpful observations. He would also like to thank Jingyi Chen and Kevin Purbhoo for their valuable suggestions and Peter Kiernan for telling him about the power method. Finally, I would like to thank Ann Kostant for her generous advice and encouragement.

Chapter 2

Linear Equations and Matrices

The purpose of this chapter is to learn about linear systems. We will restrict our discussion for now to equations whose coefficients are real numbers. In order to develop the algorithmic approach to linear systems known as Gaussian reduction, we will introduce the notion of a matrix so that we can approach any system via its coefficient matrix. This allows us to state a set of rules called row operations to bring our equations into a normal form called the reduced row echelon form of the system. The set of solutions may then be expressed in terms of fundamental and particular solutions. Along the way, we will develope the criterion for a system to have a unique solution. After we have developed some further algebraic tools, which will come in the next chapter, we'll be able to considerably strengthen the techniques we developed in this chapter.

2.1 Linear equations: the beginning of algebra

The subject of algebra arose from studying equations. For example, one might want to find all the real numbers x such that $x = x^2 - 1$. To solve, we could rewrite our equation as $x^2 - x - 6 = 0$ and then factor its left hand side. This would tell us that $(x - 3)(x + 2) = 0$, so we would conclude that either $x = 3$ or $x = -2$ since either $x - 3$ or $x + 2$ has to be zero. Finding the roots of a polynomial is a nonlinear problem, whereas the topic to be studied here is the theory of linear equations.

The simplest linear equation is the equation $ax = b$. The letter x is the variable, and a and b are fixed numbers. For example, consider $4x = 3$. The

solution is $x = 3/4$. In general, if $a \neq 0$, then $x = b/a$, and this solution is unique. If $a = 0$ and $b \neq 0$, there is no solution, since the equation says $0 = b$. And in the case where a and b are both 0, every real number x is a solution. This points out a general property of linear equations. Either there is a unique solution (i.e. exactly one), no solution or infinitely many solutions.

More generally, if $x_1, x_2, \ldots x_n$ are variables and $a_1, a_2, \ldots a_n$ and c are fixed real numbers, then the equation

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = c$$

is said to be a *linear equation*. The a_i are the *coefficients*, the x_i the *variables* and c is the *constant*. While in familiar situations, the coefficients are real numbers, it will turn out that in other important settings, such as coding theory, the coefficients might be elements of some general field. We will study fields in the next chapter. For now, let us just say that in a field it is possible to carry out division. The real numbers are a field, but the integers are not (3/4 isn't an integer).

Let's take another example. Suppose you are planning to make a cake using 10 ingredients, and you want the cake to have 2000 calories. Let a_i be the number of calories per gram of the ith ingredient. Presumably, each a_i is nonnegative, although in the future, foods with negative calories may actually be available. Similarly, let x_i be the number of grams of the ith ingredient. Then $a_1x_1 + a_2x_2 + \cdots + a_{10}x_{10}$ is the total number of calories in the recipe. Since you want the total number of calories in your cake to be exactly 2000, you consider the equation $a_1x_1 + a_2x_2 + \cdots + a_{10}x_{10} = 2000$. The totality of possible solutions $x_1, x_2, \ldots, x_{10}$ for this equation is the set of all possible recipes you can concoct.

The following more complicated example illustrates how linear equations can be used in nonlinear problems. Let $\mathbb{R}$ denote the real numbers, and suppose we want to know something about the set of common solutions of the equations $z = x^2 + xy^5$ and $z^2 = x + y^4$. These equations represent two surfaces in real three space $\mathbb{R}^3$, so we'd expect the set of common solutions to lie on a curve. Here it's impossible to express the solutions in a closed form, but we can study them locally using linear methods. For example, both surfaces meet at $(1, 1, 1)$, and they both have a tangent plane at $(1, 1, 1)$. The tangent line to the curve of intersection at $(1, 1, 1)$ is the intersection of these two tangent planes. This will give us a linear approximation to the curve near $(1, 1, 1)$.

Nonlinear systems such as in the above example are usually difficult to solve; their theory involves highly sophisticated mathematics. On the other

hand, it turns out that systems of linear equations are handled quite simply by elementary methods, and modern computers make it possible to solve gigantic linear systems with fantastic speed.

A general linear system consisting of m equations in n unknowns will look like:

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{23}x_2 + \cdots + a_{2n}x_n &= b_2 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m. \end{aligned} \tag{2.1}$$

Notice how the coefficients a_{ij} are labelled. The first index gives its row and the second index its column. The case where all the constants b_i are zero is called the *homogeneous* case. Otherwise, the system is said to be *nonhomogeneous*

The main problem, of course, is to find a procedure or algorithm for describing the *solution set* of a linear system. The principal procedure for solving a linear system is called *Gaussian reduction*. We will take this up below.

2.2 Matrices

To simplify the cumbersome notation for a system used above, we will now introduce the notion of a matrix.

Definition 2.1. A *matrix* is simply a rectangular array of real numbers. An $m \times n$ matrix is an array having m rows and n columns, such as

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{23} & \dots & a_{2n} \\ \vdots & \vdots & \dots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}. \tag{2.2}$$

If $m = n$, we say A is *square of degree* n. The set of all $m \times n$ matrices with real entries will be denoted by $\mathbb{R}^{m\times n}$.

2.2.1 Matrix Addition and Vectors

It turns out to be very useful to introduce addition and multiplication for matrices. We will begin with sums.

Definition 2.2. The *matrix sum* (or simply the *sum*) $A + B$ of two $m \times n$ matrices A and B is defined to be the $m\times n$ matrix C such that $c_{ij} = a_{ij}+b_{ij}$ for all pairs of indices (i, j). The *scalar multiple* αA of A by a real number α is the matrix obtained by multiplying each entry of A by α.

Example 2.1. Here are a couple of examples. Let

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

Then

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

Doubling A gives

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

The $m \times n$ matrix all of whose entries are zero is called the *zero matrix*. If O is the $m \times n$ zero matrix and A is any $m \times n$ matrix, then $A + O = A$. Thus O is the *additive identity for matrix addition*. Now that the additive identity for matrix addition is defined, we can observe that the matrix $-A$ is the *additive inverse* of A, in the sense that $A + (-A) = (-A) + A = O$.

A column matrix is usually simply called a *vector*. The set of all $n \times 1$ column matrices (or vectors) is denoted by $\mathbb{R}^n$. Vectors with the same number of components are combined via the component-wise addition and scalar multiplication defined above. We will use the notation $(u_1, u_2, \ldots, u_n)^T$ to express the column matrix

$$\begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix}$$

in a more compact form. What the superscript ^{T} stands for will be clarified later. Vectors will usually be written as bold faced letters. For example, $\mathbf{x}$ will stand for

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}.$$

If $\mathbf{u}_1, \mathbf{u}_2, \ldots \mathbf{u}_m$ are vectors in $\mathbb{R}^n$ and if $a_1, a_2, \ldots, a_m$ are scalars, that is elements of $\mathbb{R}$, then the vector

$$a_1\mathbf{u}_1 + a_2\mathbf{u}_2 + \cdots + a_m\mathbf{u}_m$$

is called a *linear combination* of $\mathbf{u}_1, \mathbf{u}_2, \ldots \mathbf{u}_m$.

2.2.2 Some Examples

So far we have only considered matrices over the real numbers. After we define fields in the next Chapter, we will be able to study matrices over arbitrary fields, which will give us a much wider range of applications. Briefly, a field is a set with the operations of addition and multiplication which satisfies the basic algebraic properties of the integers. However, fields also have division in the sense that every element of a field has a multiplicative inverse. we will leave the precise meaning of this statement for the next chapter.

The field with the smallest number of elements is the integers mod 2, which is denoted by $\mathbb{F}_2$. This field consists of two elements 0 and 1 with

addition being defined by $0+0=0$, $0+1=1+0=1$ and $1+1=0$. Multiplication is defined so that 1 is its usual self: $0\times 1=0$ and $1\times 1=1$. except that $1+1$ is defined to be 0: $1+1=0$. $\mathbb{F}_2$ is very useful in computer science since adding 1 represents a change of state (off to on, on to off), while adding 0 represents status quo.

Matrices over $\mathbb{F}^2$ are themselves quite interesting. For example, since $\mathbb{F}_2$ has only two elements, there are precisely 2^{mn} such matrices. Addition of such matrices has an interesting property, as the following example shows.

Example 2.2. For example,

$$\begin{pmatrix}1&0&1\\0&1&1\end{pmatrix}+\begin{pmatrix}1&1&1\\1&1&1\end{pmatrix}=\begin{pmatrix}0&1&0\\1&0&0\end{pmatrix},$$

and

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

In the first sum, the parity of every element in the first matrix is reversed. In the second, we see every matrix over $\mathbb{F}_2$ is its own additive inverse.

Example 2.3. Random Key Crypts. 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 does. First he assigns the number 1 to a, 2 to b and so forth up to 26 to z. He also assigns 0 to the space between two words. He then computes the binary expansion of each integer between 1 and 26. Thus 1=1, 2=10, 3=11, 4=100, ..., 26=11010. He now converts his message into a sequence of five digit strings. Note that 00000 represents a space. The result is his encoded message, which is normally referred to as the *plaintext*. To make things more compact, he arranges the plaintext into a matrix. For example, if there are 5 words, hence 4 spaces, he could make a 3×3 matrix of 5 digit strings of zeros and ones.

Let's denote the matrix containing the plaintext by P, and suppose P is $m\times n$. Now the fun starts. Rocky and Bullwinkle have a list of $m\times n$ matrices of zeros and ones that only they know. The flying squirrel selects one of these matrices, say number 47, and tells Bullwinkle. Let E be matrix number 47. Cryptographers call E the *key*. Now he sends the *ciphertext* $enc_E(P)=P+E$ to Bullwinkle. If only Rocky and Bullwinkle know E, then the matrix P containing the plaintext is secure. Even if Boris and Natasha succeed in learning the ciphertext $P+E$, they will still have to know E to

find out what P is. The trick is that the key E has to be sufficiently random so that neither Boris nor Natasha can guess it. For example, if E is the all ones matrix, then P isn't very secure since Boris and Natasha will surely try it. Notice that once Bullwinkle receives the ciphertext, all he has to do is add the key E to recover the plaintext P since

$$enc_E(P) + E = (P + E) + E = P + (E + E) = P + O = P.$$

This is something even a mathematically challenged moose can do.

The hero's encryption scheme is extremely secure if the key E is sufficiently random and only used once. (Such a crypt is called a *one time pad*.) However, if he uses E to encrypt another plaintext message Q, and Boris and Natasha pick up both $enc_E(P) = P + E$ and $enc_E(Q) = Q + E$, then they can likely find out what both P and Q say. The reason for this is that

$$(P + E) + (Q + E) = (P + Q) + (E + E) = P + Q + O = P + Q.$$

The point is that knowing $P + Q$ may be enough for a good cryptographer to deduce both P and Q. But, as a one time pad, the random key is quite secure (in fact, apparently secure enough for communications on the hot line between Washington and Moscow).

Example 2.4. (**Scanners**) We can also interpret matrices over $\mathbb{F}_2$ in another natural way. Consider a black and white photograph as being 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, the black and white photo is therefore transformed into a matrix over $\mathbb{F}_2$. Now suppose we want to compare two black and white photographs whose matrices A and B are both $m \times n$. It's inefficient for a computer to scan the two matrices to see in how many positions they agree. However, when A and BB are added, the sum $A + B$ has a 1 in in any component where A and B differ, and a 0 wherever they coincide. For example, the sum two identical photographs is the zero matrix, and the sum of two complementary photographs is the all ones matrix. An obvious measure of how similar the two matrices A and B are is the number of non zero entries of $A + B$, i.e. $\Sigma(a_{ij} + b_{ij})$. This easily tabulated number is known as the *Hamming distance* between A and B.

2.2.3 Matrix Product

We will introduce matrix multiplication in the next Chapter. To treat linear systems, however, we need to define the product $A\mathbf{x}$ of a $m \times n$ matrix A

and a (column) vector $\mathbf{x}$ in $\mathbb{R}^n$. Put

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{23} & \dots & a_{2n} \\ \vdots & \vdots & \dots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \tag{2.3}$$

$$= \begin{pmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ a_{21}x_1 + a_{23}x_2 + \cdots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \end{pmatrix}.$$

¿From now on, one should think of the left hand side of the linear system $A\mathbf{x} = \mathbf{b}$ as a product.

Let us point out a basic property of multiplication.

Proposition 2.1. *The matrix product Ax is distributive. That is, for any $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{R}^n$ and any $A \in \mathbb{R}^{m\times n}$, $A(\mathbf{x}+\mathbf{y}) = A\mathbf{x} + A\mathbf{y}$.*

Proof. This is obvious from the distributive property of real numbers. □

2.3 Reduced Row Echelon Form and Row Operations

The purpose of this section is to define the two fundamental concepts in the title, which will turn out to be the main tools for solving an arbitrary linear system. Let us begin with the observation that a linear system $A\mathbf{x} = \mathbf{b}$ has an associated pair of matrices as described in the next definition.

Definition 2.3. The *coefficient matrix* of the linear system $A\mathbf{x} = \mathbf{b}$ in (2.1) is the $m \times n$ matrix A in (2.2). The *augmented coefficient matrix* of (2.1) is the $m \times (n+1)$ matrix

$$(A|\mathbf{b}) = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} & b_1 \\ a_{21} & a_{23} & \dots & a_{2n} & b_2 \\ \vdots & \vdots & \cdots & \vdots & \\ a_{m1} & a_{m2} & \dots & a_{mn} & b_m \end{pmatrix}. \tag{2.4}$$

We will start by defining the notion of row echelon form.

Definition 2.4. A matrix A is said to be in *row echelon form* if

(i) the first non zero entry in every row is to the right of the first non zero entry in all the rows above, and

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

The first non zero entry in a row is called its *pivot* or *corner entry*. A matrix A in row echelon form is said to be in *reduced row echelon form*, or simply *reduced*, if each corner entry is 1.

For a reason connected with matrix multiplication, the reduced $n \times n$ matrix in reduced row echelon form with n corners is called the $n \times n$ *identity matrix*. The identity matrix is denoted by I_n. For example,

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

Here are some more examples of reduced matrices:

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

Notice that the last matrix in this example would be the coefficient matrix of a system with variables $x_1, x_2, \ldots, x_5$ in which the variable x_1 doesn't actually appear. The only variables that one needs to solve for are $x_2, \ldots, x_5$. As mentioned above, the $n \times n$ identity matrix I_n is reduced, as is the $m \times n$ matrix of zeros.

We next define the three elementary row operations.

Definition 2.5. Let A be an arbitrary matrix. The three elementary row operations on A are as follows:

(I) interchange two rows of A;

(II) multiply a row of A by a non zero scalar; and

(III) replace a row of A by itself plus a multiple of a different row.

We will call operations of type I *row swaps*. Type II operations are called *row dilations*, and type III operations are called *transvections*. (We will generally not use this term.)

Proposition 2.2. *Every matrix A can be put into reduced row echelon form by a (not unique) sequence of elementary row operations.*

Before giving a proof, let us work an example.

Example 2.5. Consider the counting matrix

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

We can row reduce C as follows:

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

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

Notice that we have indicated the row operations.

Proof of Proposition 2.2. If $a_{11} \neq 0$, we can make it 1 by the dilation which divides the first row by a_{11}. We can then use row operations of type III to make all other entries in the first column zero. If $a_{11} = 0$, but the first column has a non zero entry somewhere, swapping the first row with the row containing this non zero entry puts a nonzero entry in the $(1, 1)$ position. Next, divide the new first row by the inverse of the $(1, 1)$ entry, getting a one in the $(1, 1)$ position. We now have a corner entry in the $(1, 1)$ position, and so we can use operations of type III to make all the elements in the first column below the $(1, 1)$ entry 0. If the first column consists entirely of zeros, we can proceed directly to the second column. To continue, we repeat the above steps on the second column, except that the question is whether the $(2, 2)$ entry a'_{22} is zero or not. If $a'_{22} \neq 0$, we can make the $(2, 2)$ entry 1 by dividing the second row by a'_{22}. After that, the $(1, 2)$ entry can be made 0 by a type III operation. If $a'_{22} = 0$, we look for a nonzero entry below it, and if there is one, a row swap will put it into the $(2, 2)$ position. If every entry below the $(1, 2)$ position is zero, we can proceed directly to the third clumn. Continuing in this manner, we will eventually obtain a reduced matrix. □

REMARK: Of course, the steps leading to a reduced form are not unique. Nevertheless, the reduced form of A itself turns out to be unique, This isn't obvious, but it can be proven (see Propositions 3.18 and 5.15). We now make an important definition assuming this.

Definition 2.6. The reduced form of an $m \times n$ matrix A is denoted by A_{red}. The *row rank*, or simply, the *rank* of an $m \times n$ matrix A is the number of non-zero rows in A_{red}.

2.4 Solving Linear Systems via Gaussian Reduction

Gaussian reduction is an algorithmic procedure for finding the solution set of a linear system. We will say that *two linear systems are equivalent* if their solution sets are equal. The strategy in Gaussian reduction is to replace the original system with a sequence of equivalent systems until the final system is in *reduced row echelon form*. That is, its coefficient matrix is in reduced row echelon form. The sequence of equivalent systems is produced by applying row operations.

2.4.1 Row Operations and Equivalent Systems

Let A be an $m \times n$ matrix and consider the linear system $A\mathbf{x} = \mathbf{b}$. The augmented coefficient matrix of this system is $(A|\mathbf{b})$. The first thing is to point out the role of row operations. What happens when one performs an elementary row operation on $(A \mid \mathbf{b})$? In fact, I claim that the new system is equivalent to the original system.

For example, row swaps simply interchange two equations, so they clearly leave the solution set unchanged. Similarly, multiplying the ith equation by a non-zero constant a does likewise, since the original system can be recaptured by multiplying the ith equation by a^{-1}. The only question is whether a row operation of type III changes the solutions. Suppose the ith equation is replaced by itself plus a multiple k of the jth equation, where $i \neq j$. Then any solution of the original system is still a solution of the new system. But any solution of the new system is also a solution of the original system since subtracting k times the jth equation from the ith equation of the new system gives us back the original system. Therefore the systems are equivalent.

To summarize this, we state

Proposition 2.3. *Performing a sequence of row operations on the augmented coefficient matrix of a linear system gives a new system which is equivalent to the original system.*

To reiterate, to solve a linear system by Gaussian reduction, the first step is to put the augmented coefficient matrix in reduced row echelon form via a sequence of row operations. The next step will be to find the solution set.

2.4.2 The Homogeneous Case

Solving a linear system $A\mathbf{x} = \mathbf{b}$ involves several steps. The first step is to solve the associated homogeneous system.

Definition 2.7. A linear system $A\mathbf{x} = \mathbf{b}$ is said to be *homogeneous* if $\mathbf{b} = \mathbf{0}$. The solution set of a homogeneous linear system $A\mathbf{x} = \mathbf{0}$ is called the *null space* of A. The null space of A is denoted throughout by $\mathcal{N}(A)$.

The efficient way to describe the solution set of a homogeneous linear system to use vectors. Note first that since performing row operations on $(A|\mathbf{0})$ doesn't alter the last column, we only need to use the coefficient matrix A.

The following example shows how to write down the null space.

Example 2.6. Consider the homogeneous linear system

$$\begin{aligned} 0x_1 + x_2 + 2x_3 + 0x_4 + 3x + 5 - x_6 &= 0 \\ 0x_1 + 0x_2 + 0x_3 + x_4 + 2x_5 + 0x_6 &= 0. \end{aligned}$$

Notice that the coefficient matrix A is already reduced. Indeed,

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

The procedure is to solve for the variables corresponding to the columns with corners, which we call the *corner variables*. Since the corner variables have nonzero coefficients, they can be expressed in terms of the remaining variables, which are called the *free variables*. For A, the corner columns are the second and fourth, so x_2 and x_4 are the corner variables, and the variables x_1, x_3, x_5 and x_6 are the free variables. Solving for x_2 and x_4 gives

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

In this expression, the corner variables are dependent variables which are functions of the free variables. Now let $\mathbf{x} = (x_1, x_2, x_3, x_4, x_5, x_6)^T$ denote an arbitrary vector in $\mathbb{R}^6$ in the solution set of the system, and let us call $\mathbf{x}$ the *general solution vector*. Notice that we have expressed it as a column vector. Replacing the corner variables by their expressions in terms of the free variables gives a new expression for the general solution vector involving only the free variables. Namely

$$\mathbf{x} = (x_1, -2x_3 - 3x_5 + x_6,\ x_3,\ -2x_5,\ x_5,\ x_6)^T.$$

The general solution vector now depends only on the free variables, and there is a solution for any choice of these variables.

Using a little algebra, we can compute the vector coefficients of each one of the free variables in $\mathbf{x}$. These vectors are called the *fundamental solutions*. In this example, the general solution vector $\mathbf{x}$ has the form

$$\mathbf{x} = x_1\mathbf{f}_1 + x_3\mathbf{f}_2 + x_4\mathbf{f}_3 + x_5\mathbf{f}_4, \tag{2.5}$$

where

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

The equation (2.5) tells us that every solution of $A\mathbf{x} = \mathbf{0}$ is a linear combination of the fundamental solutions $\mathbf{f}_1, \ldots, \mathbf{f}_4$.

This example illustrates a trivial but useful fact.

Proposition 2.4. *In an arbitrary homogeneous linear system with coefficient matrix A, every solution is a linear combination of the fundamental solutions, and the number of fundamental solutions is the number of free variables. Thus,*

$$\#\text{corner variables} + \#\text{free variables} = \#\text{variables}. \tag{2.6}$$

Proof. The proof that every solution is a linear combination of the fundamental solutions goes exactly like the above example, so we will omit it. Equation (2.6) is an application of the fact that every variable is either a free variable or a corner variable, but not both. □

We will eventually prove several refinements of this property which will say considerably more about the structure of the solution set.

Let us point out something a bit unusual in Example 2.6. The variable x_1 never actually appears in the system, but it does give a free variable and a corresponding fundamental solution $(1, 0, 0, 0, 0, 0)^T$. Suppose instead of A the coefficient matrix is

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

Now $(1,0,0,0,0,0)^T$ is no longer a fundamental solution. In fact the solution set is now a subset of $\mathbb{R}^5$. The corner variables are x_1 and x_3, and there are now only three fundamental solutions corresponding to the free variables x_2, x_4, and x_5.

Even though identity (2.6) is completely obvious, it gives some very useful information. Here is a typical application.

Example 2.7. Consider a linear system with 25 variables and assume there are 10 free variables. Then there are 15 corner variables, so the system has to have at least 15 equations. That is, there have to be at least 15 linear constraints on the 25 variables.

We can also use (2.6) to say when the homogeneous system $A\mathbf{x} = \mathbf{0}$ has a unique solution (that is, exactly one solution). Note that $\mathbf{0}$ is always a solution: the *trivial solution*. Hence if the solution is to be unique, then the only possibility is that $\mathcal{N}(A) = \{\mathbf{0}\}$. But this happens exactly when there are no free variables, since if there is a free variable there have to be non trivial solutions. Thus a homogeneous system has a unique solution if and only if every variable is a corner variable, which is the case exactly when the number of corner variables is the number of columns of A. By the same reasoning, if a homogeneous system has more variables than equations, there have to be non trivial solutions, since there has to be at least one free variable.

2.4.3 The Non-homogeneous Case

A system $A\mathbf{x} - \mathbf{b}$ with $\mathbf{b} \neq \mathbf{0}$ is said to be *non-homogeneous*. A non-homogeneous system requires that we use an augmented coefficient matrix $(A \mid \mathbf{b})$.

To resolve the non-homogeneous case, we need to observe a result sometimes called the *Super-Position Principle*.

Proposition 2.5. *If a system with augmented coefficient matrix $(A|\mathbf{b})$ has a particular solution $\mathbf{p}$, then any other solution has the form $\mathbf{p} + \mathbf{x}$, where $\mathbf{x}$ is an arbitrary element of $\mathcal{N}(A)$.*

Proof. The proof is quite easy. Suppose $\mathbf{p} = (p_1, \ldots, p_n)^T$, and let $\mathbf{x} = (x_1, \ldots, x_n)^T$ be an element of $\mathcal{N}(A)$. By the distributivity of matrix multiplication (Proposition 2.1),

$$A(\mathbf{p} + \mathbf{x}) = A\mathbf{p} + A\mathbf{x} = \mathbf{b} + \mathbf{0} = \mathbf{b}.$$

Conversely, if $\mathbf{q}$ is also particular solution, then $\mathbf{p}-\mathbf{q}$ is a solution to the homogeneous system, since

$$A(\mathbf{p}-\mathbf{q}) = A\mathbf{p} - A\mathbf{q} = \mathbf{b} - \mathbf{b} = \mathbf{0}.$$

Thus, $\mathbf{p}-\mathbf{q}$ is an element of $\mathcal{N}(A)$. Therefore $\mathbf{q} = \mathbf{p} + \mathbf{x}$, where $\mathbf{x} = \mathbf{q} - \mathbf{p}$, as asserted. This completes the proof. □

In the above proof, we made the statement that $A(\mathbf{p}+\mathbf{x}) = A\mathbf{p} + A\mathbf{x}$. This follows from a general algebraic identity called the distributive law which we haven't yet discussed. However, our particular use of the distributive law is easy to verify from first principles.

Example 2.8. Consider the system involving the counting matrix C of Example 2.5:

$$\begin{aligned} 1x_1 + 2x_2 + 3x_3 &= a \\ 4x_1 + 5x_2 + 6x_3 &= b \\ 7x_1 + 8x_2 + 9x_3 &= c, \end{aligned}$$

where a, b and c are fixed arbitrary constants. This system has augmented coefficient matrix

$$(C|\mathbf{b}) = \begin{pmatrix} 1 & 2 & 3 & a \\ 4 & 5 & 6 & b \\ 7 & 8 & 9 & c \end{pmatrix}.$$

We can use the same sequence of row operations as in Example 2.5 to put $(C|\mathbf{b})$ into reduced form $(C_{red}|\mathbf{c})$ but to minimize the arithmetic with denominators, we will actually use a different sequence.

$$(C|\mathbf{b}) \overset{R_2-R_1}{\rightarrow} \begin{pmatrix} 1 & 2 & 3 & a \\ 3 & 3 & 3 & b-a \\ 7 & 8 & 9 & c \end{pmatrix} \overset{R_3-2R_2}{\rightarrow} \begin{pmatrix} 1 & 2 & 3 & a \\ 3 & 3 & 3 & b-a \\ 1 & 2 & 3 & c-2b+2a \end{pmatrix} \overset{R_3-R_1}{\rightarrow}$$

$$\begin{pmatrix} 1 & 2 & 3 & a \\ 3 & 3 & 3 & b-a \\ 0 & 0 & 0 & c-2b+a \end{pmatrix} \overset{(-1/3)R_3}{\rightarrow} \begin{pmatrix} 1 & 2 & 3 & a \\ -1 & -1 & -1 & (1/3)a-(1/3)b \\ 0 & 0 & 0 & c-2b+a \end{pmatrix} \overset{R_2+R_1}{\rightarrow}$$

$$\begin{pmatrix} 1 & 2 & 3 & a \\ 0 & 1 & 2 & (4/3)a-(1/3)b \\ 0 & 0 & 0 & c-2b+a \end{pmatrix} \overset{R_1-2R_2}{\rightarrow} \begin{pmatrix} 1 & 0 & -1 & (-5/3)a+(2/3)b \\ 0 & 1 & 2 & (4/3)a-(1/3)b \\ 0 & 0 & 0 & c-2b+a \end{pmatrix}.$$

The reduced system turns out to be the same one we obtained by using the sequence in Example 11.2. We get

$$\begin{aligned} 1x_1 + 0x_2 - 1x_3 &= (-5/3)a + (2/3)b \\ 0x_1 + 1x_2 + 2x_3 &= (4/3)a - (1/3)b \\ 0x_1 + 0x_2 + 0x_3 &= a - 2b + c \end{aligned}$$

Clearly the above system may in fact have no solutions. Indeed, from the last equation, we see that whenever $a-2b+c \neq 0$, there cannot be a solution, since the left side of the third equation is always zero. Such a system is called *inconsistent*. For a simpler example of an inconsistent system, think of three lines in $\mathbb{R}^2$ which don't pass through a common point. This is an example where the system has three equations but only two variables.

Example 2.9. Let's solve the system of Example 2.8 for $a = 1$, $b = 1$ and $c = 1$. In that case, the original system is equivalent to

$$\begin{aligned} 1x_1 + 0x_2 - 1x_3 &= -1 \\ 0x_1 + 1x_2 + 2x_3 &= 1 \\ 0x_1 + 0x_2 + 0x_3 &= 0 \end{aligned}$$

It follows that $x_1 = -1 + x_3$ and $x_2 = 1 - 2x_3$. This represents a line in $\mathbb{R}^3$.

The line of the previous example is parallel to the line of intersection of the three planes

$$\begin{aligned} 1x_1 + 2x_2 + 3x_3 &= 0 \\ 4x_1 + 5x_2 + 6x_3 &= 0 \\ 7x_1 + 8x_2 + 9x_3 &= 0. \end{aligned}$$

In fact, one can see directly that these three planes meet in a line since our computation with row operations shows that the vectors normal to the three planes are contained in a single plane through the origin. On the other hand, when $a - 2b + c \neq 0$, what happens is that the line of intersection of any two of the planes is parallel to the third plane but doesn't meet it.

2.4.4 Criteria for Consistency and Uniqueness

To finish our treatment of systems (for now), we derive the criteria for consistency and uniqueness. The key concept is the notion of the rank of a matrix, which we saw earlier. Recall that the *rank* of an $m \times n$ matrix A is the number of corners in its reduced row echelon form.

Clearly the rank of an $m \times n$ matrix A is at most the minimum of m and n. However, we don't yet know whether the number of corners in a matrix is well defined, since two different sequences of row operations might produce two different reduced matrices. Hence we'll need to assume for now that the rank is a well defined concept. In fact, we'll prove in Proposition 3.18 that the reduced row echelon form of an arbitrary matrix is unique.

Proposition 2.6. *Let A be an $m \times n$ matrix. Then:*

(*i*) $\mathcal{N}(A) = \{\mathbf{0}\}$ *if and only if the rank of A is n;*

(*ii*) *if $A\mathbf{x} = \mathbf{b}$ is consistent and the rank of A is n, then the solution is unique;*

(*iii*) *the linear system $A\mathbf{x} = \mathbf{b}$ is consistent if and only if the ranks of A and $(A \mid \mathbf{b})$ are the same; and*

(*iv*) *if A is $n \times n$ and has rank n, the system $A\mathbf{x} = \mathbf{b}$ has a unique solution for all $\mathbf{b} \in \mathbb{R}^n$.*

The converse of statement (iv) is also true.

Proof. The first statement was already proved at the end of Section 2.4.2 using (2.6). The only way to have $\mathcal{N}(A) = \{\mathbf{0}\}$ is if every variable in the system $A\mathbf{x} = \mathbf{0}$ is a corner variable, which is the same as saying A has rank n. For the second statement, let $\mathbf{u}$ and $\mathbf{v}$ be two solutions of $A\mathbf{x} = \mathbf{b}$. Then $A(\mathbf{u} - \mathbf{v}) = \mathbf{b} - \mathbf{b} = \mathbf{0}$. Thus $\mathbf{u} - \mathbf{v} \in \mathcal{N}(A)$, so $\mathbf{u} - \mathbf{v} = \mathbf{0}$ by (i). The third statement follows as in the previous example, because if the rank of $(A \mid \mathbf{b})$ is greater than the rank of A, then the last equation is equivalent to the inconsistent equation $0 = 1$. For (iv), let A have rank n. Then $(A \mid \mathbf{b})$ also has rank n, since A is $n \times n$ and hence the rank of $(A \mid \mathbf{b})$ can't exceed n. Thus $A\mathbf{x} = \mathbf{b}$ has a unique solution for all $\mathbf{b} \in \mathbb{R}^n$ by (ii) and (iii). It remains to show the converse of (iv) that if A and $(A \mid \mathbf{b})$ have the same rank for all $\mathbf{b}$, then A has rank n. But if the rank of A is less than n, one can (exactly as in Example 2.8) produce a $\mathbf{b}$ for which $(A \mid \mathbf{b})$ has rank greater than the rank of A. We will leave filling in all the details as an exercise. □

Systems where $m = n$ are an important special case as they are neither under determined (fewer equations than unknowns) nor over determined (more equations than unknowns). When A is $n \times n$ of rank n, the system $A\mathbf{x} = \mathbf{b}$ is said to be *nonsingular*. Thus the nonsingular systems are the square systems which are always consistent and always have unique solutions. We will also say that an $n \times n$ matrix A is *nonsingular* if it has maximal rank n. If the rank of A is less than n, we will call A *singular*.

Exercises

Exercise 2.1. Consider the linear system

$$\begin{aligned} x_1 + 2x_2 + 4x_3 + 4x_4 &= 7 \\ x_2 + x_3 + 2x_4 &= 3 \\ x_1 + 0x_2 + 2x_3 + 0x_4 &= 1 \end{aligned}$$

(a) Let A be the coefficient matrix of the associated homogeneous system. Find the reduced form of A.

(b) Determine whether the system is consistent and, if so, find the general solution.

(c) Find the fundamental solutions of $A\mathbf{x} = \mathbf{0}$ and show that the every solution of $A\mathbf{x} = \mathbf{0}$ is a linear combination of the fundamental solutions.

(d) Is the system $A\mathbf{x} = \mathbf{b}$ consistent for all $\mathbf{b} \in \mathbb{R}^3$? If not, find an equation which the components of $\mathbf{b}$ must satisfy.

Exercise 2.2. Show that a real 2×2 matrix $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is nonsingular if and only if $ad - bc \neq 0$.

Exercise 2.3. Consider the system

$$ax + by = e$$

$$cx + dy = f.$$

Show that if $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is nonsingular, then the unique solution is given by $x = (de - bf)/(ad - bc)$ and $y = (af - ce)/(ad - bc)$.

Exercise 2.4. If A is 9×27, explain why the system $A\mathbf{x} = \mathbf{0}$ must have at least 18 fundamental solutions.

Exercise 2.5. Consider the system $A\mathbf{x} = \mathbf{0}$ where $A = \left(\begin{smallmatrix} 1 & -1 & 2 & -1 & 1 \\ -2 & 2 & 1 & -2 & 0 \end{smallmatrix}\right)$. Find the fundamental solutions and show that every element in $\mathcal{N}(A)$ is a linear combination of the fundamental solutions.

Exercise 2.6. Let A be the 2×5 matrix of Problem 2.5. Solve the compounded linear system

$$\left(A \middle| \begin{smallmatrix} 1 & -1 \\ -2 & 0 \end{smallmatrix}\right).$$

Exercise 2.7. Set up a linear system to determine whether $(1, 0, -1, 1)^T$ is a linear combination of $(-1, 1, 2, 0)^T$, $(2, 1, 0, 1)^T$ and $(0, 1, 0, -1)^T$ with real coefficients.

Exercise 2.8. Set up a linear system to determine whether $(1, 0, 1, 1)^T$ is a linear combination of $(1, 1, 0, 0)^T$, $(0, 1, 0, 1)^T$ and $(0, 1, 1, 1)^T$ with coefficients in $\mathbb{F}_2$.

Exercise 2.9. A baseball team has won 3 mores games at home than on the road, and lost 5 more at home than on the road. If the team has played a total of 42 games, and if the number of home wins plus the number of road losses is 20, determine the number of home wins, road wins, home losses and road losses.

Exercise 2.10. For what real values of a and b does the system

$$\begin{aligned} x + ay + a^2 z &= 1 \\ x + ay + abz &= a \\ bx + a^2 y + a^2 bz &= a^2 b \end{aligned}$$

have a unique solution?

Exercise 2.11. True or False: If the normals of three planes in $\mathbb{R}^3$ through the origin lie in a plane through the origin, then the planes meet in a line.

Exercise 2.12. Suppose A is a 12×15 matrix of rank 12. How many fundamental solutions are there in $\mathcal{N}(A)$?

Exercise 2.13. Find the number of nonsingular real 2×2 matrices having the property that all entries are either 0 or 1?

Exercise 2.14. Determine the number of nonsingular real 3×3 matrices with only 0 or 1 entries under the following conditions:

(i) exactly 3 entries are nonzero;

(ii) exactly 4 entries nonzero; and

(iii) exactly 8 entries are nonzero.
Bonus: find the total number of nonsingular real 3×3 matrices with all entries either 0 or 1.

Exercise 2.15. * Determine the number of nonsingular 3×3 matrices over $\mathbb{F}_2$.

Exercise 2.16. Find the ranks of each of the following matrices:

$$\begin{pmatrix} 1 & 2 & 3 \\ 1 & 4 & 9 \\ 1 & 8 & 27 \end{pmatrix}, \quad \begin{pmatrix} 1 & 2 & 2 \\ 1 & 4 & 4 \\ 1 & 8 & 8 \end{pmatrix}.$$

Exercise 2.17. * Find the rank of

$$A = \begin{pmatrix} 1 & a & a^2 \\ 1 & b & b^2 \\ 1 & c & c^2 \end{pmatrix},$$

where a, b, c are arbitrary real numbers.

Exercise 2.18. Find an equation which describes all possible values of a, b, c such that $(a, b, c)^T$ is a linear combination of $(-2, 1, 0)^T$, $(2, 1, 2)^T$ and $(0, 1, 1)^T$.

Exercise 2.19. If a chicken and a half lay and egg and a half in a day and a half, how many eggs does a single chicken lay in one day? Can you express this to linear equations.

2.5 Summary

This Chapter is an introduction to linear systems and matrices. We began by introducing the general linear system of m equations in n unknowns with real coefficients. There are two types of systems called homogeneous and non-homogeneous according to whether the constants on the right hand sides of the equations are all zeros or not. The solutions make up the solution set. If the system is homogeneous, every solution is a linear combination of the fundamental solutions. (We will see in the next Chapter that this means that the solution set is a subspace of $\mathbb{R}^n$.) In order to write the system in a convenient form, we introduced the coefficient matrix for homogeneous systems and the augmented coefficient matrix for non-homogeneous systems. We then wrote down the three row operations of Gaussian reduction. The row operations give a specific set of rules for bringing the coefficient matrix and augmented coefficient matrix into a normal form known as reduced form or reduced row echelon form. The point is that performing a row operation on the coefficient matrix (or augmented coefficient matrix) gives a new coefficient matrix (or augmented coefficient matrix) whose associated linear system has exactly the same solution space (or set in the non-homogeneous case).

After a matrix has been put into reducued form, one can read off its rank (the number of non-zero rows). We then obtained criteria which are necessary and sufficient for the existence and uniqueness of solutions. A non-homogeneous system has a solution if and only if its augmented coefficient matrix and coefficient matrix have the same rank. A unique solution exists if and only if the augmented coefficient matrix and coefficient matrix have the same rank and the rank is the number of unknowns.

Chapter 3

More Matrix Theory

The purpose of this chapter is to expand our knowledge of matrix theory. In particular we will study matrix inverses and use inverses to and to sharpen what we can say about linear systems. We will also see how row operations and matrix multiplication are related. Finally, will derive the $LPDU$ decomposition of a square matrix and apply it to show that the reduced row echelon form of an arbitrary matrix is unique.

3.1 Matrix Multiplication

In this section, we will define the product of two matrices and state the basic properties of the resulting matrix algebra. Let $\mathbb{R}^{m\times n}$ denote the set of all $m \times n$ matrices with real entries, and let $(\mathbb{F}_2)^{m\times n}$ denote the set of all $m \times n$ matrices over $\mathbb{F}_2$.

We have already defined matrix addition and the multiplication of a matrix by a scalar, and we've seen how to multiply an $m \times n$ matrix and a column vector with n components. We wil now define matrix multiplication. In general, the product AB of two matrices A and B is defined only when the number of columns of A equals the number of rows of B. Suppose $A = (\mathbf{a}_1\ \mathbf{a}_2\ \cdots\ \mathbf{a}_n)$ is $m \times n$ and $B = \begin{pmatrix}\mathbf{b}_1 & \mathbf{b}_2 & \cdots & \mathbf{b}_p\end{pmatrix}$ is $n \times p$. Since we already know the definition of each $A\mathbf{b}_j$, let us simply put

$$AB = \begin{pmatrix}A\mathbf{b}_1 & A\mathbf{b}_2 & \cdots & A\mathbf{b}_p\end{pmatrix}. \tag{3.1}$$

To write this out more precisely, let $C = AB$, and suppose the entry of C in the i-th row and k-th column is denoted c_{ik}. Then, using summation

notation, we have

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

so

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

Thus, for real matrices, we have

$$\mathbb{R}^{m \times n} \cdot \mathbb{R}^{n \times p} \subset \mathbb{R}^{m \times p},$$

where $\cdot$ denotes matrix multiplication.

Another way of putting the definition is to say that if the columns of A are $\mathbf{a}_1, \ldots, \mathbf{a}_n$, then the r-th column of AB is

$$b_{1r}\mathbf{a}_1 + b_{2r}\mathbf{a}_2 + \ldots b_{nr}\mathbf{a}_n. \tag{3.2}$$

Hence the r-th column of AB is the linear combination of all n columns of A using the n entries in the r-th column of B as the scalars. One can also express AB as a linear combination of the rows of B. The reader is invited to work this out explicitly. We will in fact use it to express row operations in terms of matrix multiplication. below

Example 3.1. Here are two examples.

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

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 for there exist 2×2 matrices A and B such that $AB \neq BA$, even though both products AB and BA are defined. *In general, matrix multiplication is not commutative.* In fact, almost any pair of 2×2 matrices you choose will not commute. In general, the multiplication of $n \times n$ matrices is not commutative. The only exception is that all 1×1 commute (why?).

3.1.1 The Transpose of a Matrix

Another operation on matrices is *transposition*, or taking the transpose. If A is $m \times n$, the *transpose* A^T of A is the $n \times m$ matrix $A^T := (c_{rs})$, where $c_{rs} = a_{sr}$. This is easy to remember: the ith row of A^T is just the ith column of A.

Example 3.2. If

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

then

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

An example of a 2×2 symmetric matrix is

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

Note that A and A^T have the same entries on the diagonal. Another simple fact is that

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

Definition 3.1. A matrix A which is equal to its transpose (that is, $A = A^T$) is called *symmetric*.

Clearly, every symmetric matrix is square. The symmetric matrices over $\mathbb{R}$ turn out to be especially fundamental, as we will see later.

The *dot product* $\mathbf{v} \cdot \mathbf{w}$ of two vectors $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ is defined to be the matrix product

$$\mathbf{v} \cdot \mathbf{w} = \mathbf{v}^T \mathbf{w} = \sum_{i=1}^{n} v_i w_i.$$

Proposition 3.1. *Let A and B be $m \times n$ matrices. Then*

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

Furthermore,

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

Proof. The first identity is left as an exercise. The product transpose identity can be seen as follows. The (i, j)-entry of $B^T A^T$ is the dot product of the i-th row of B^T and the j-th column of A^T. Since this is the same thing as the dot product of the j-th row of A and the i-th column of B, which is the (j, i)-entry of AB, and hence the (i, j)-entry of $(AB)^T$, we see that $(AB)^T = B^T A^T$. Suggestion: try this out on an example.

3.1.2 The Algebraic Laws

Except for the commutativity of multiplication, the usual algebraic properties of addition and multiplication in the reals also hold for matrices.

Proposition 3.2. *Assuming all the sums and products below are defined, then matrix addition and multiplication satisfy:*

(1) **the associative law***: Matrix addition and multiplication are associative:*

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

(2) **the distributive law***: Matrix addition and multiplication are distributive:*

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

(3) **the scalar multiplication law***: For any scalar r,*

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

(4) **the commutative law for addition***: Matrix addition is commutative: $A+B=B+A$.*

Verifying these properties is a routine exercise, so we will omit the details. I suggest working a couple of examples to convince yourself, if necessary. Though the associative law for multiplication doesn't seem to be exciting, it often turns to be extremely useful. We will soon see some examples of why.

Recall that the $n\times n$ *identity matrix* I_n is the matrix having one in each diagonal entry and zero in each entry off the diagonal. For example,

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

Note the inteersting fact that we can also construct the identity matrix over $\mathbb{F}_2$. The off diagonal entries are 0, of course, and the diagonal entries consist of the nonzero element 1, which is is the multiplicative identity of $\mathbb{F}_2$.

We have

Proposition 3.3. *If A is an $m\times n$ matrix (over $\mathbb{R}$ or $\mathbb{F}_2$), then $AI_n=A$ and $I_mA=A$.*

Proof. This is an exercise in using the definition of multiplication. □

Exercises

Exercise 3.1. Make up three matrices A, B, C so that AB and BC are defined. Then compute AB and $(AB)C$. Next compute BC and $A(BC)$. Compare your results.

Exercise 3.2. Prove the assertion $(A+B)^T = A^T + B^T$ in Proposition 3.1.

Exercise 3.3. Suppose A and B are symmetric $n \times n$ matrices. (You can even assume $n = 2$.)

(a) Decide whether or not AB is always symmetric. That is, whether $(AB)^T = AB$ for all symmetric A and B?

(b) If the answer to (a) is no, what condition ensures AB is symmetric?

Exercise 3.4. Suppose B has a column of zeros. How does this affect a product of the form AB? What if A has a row or a column of zeros?

Exercise 3.5. Show how to express the rows of the matrix product AB as linear combinations of the rows of B.

Exercise 3.6. Verify Proposition 3.3 for all A in either $\mathbb{R}^{m\times n}$ or $\mathbb{F}_2^{m\times n}$.

Exercise 3.7. Find all 2×2 matrices $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ such that $AB = BA$, where $B = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$.

Exercise 3.8. 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 3.7.

Exercise 3.9. Prove that if $S \in \mathbb{R}^{2\times 2}$ commutes with every matrix $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathbb{R}^{2\times 2}$, then $S = sI_2$ for some $s \in \mathbb{R}$. Matrices of the form aI_n are called *scalar matrices.*

Exercise 3.10. Let A be the 2×2 matrix over $\mathbb{F}_2$ such that $a_{ij} = 1$ for each i, j. Compute A^m for any integer $m > 0$. Does this question make sense if $m < 0$? (Note A^j is the product $AA\cdots A$ of A with itself j times.)

Exercise 3.11. Let A be the $n \times n$ matrix over $\mathbb{R}$ such that $a_{ij} = 2$ for all i, j. Find a formula for A^j for any positive integer j.

Exercise 3.12. 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 the reals.

3.2 Elementary Matrices and Row Operations

The purpose of this section is make a connection between matrix multiplication and row operations. What we will see is that row operations can be done by matrix multiplication. This may be somewhat unexpected, so the reader might want to recall the result of Exercise 3.5.

Let us first consider the $2 \times n$ case. Here, the following three types of 2×2 matrices are used:

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

These matrices enable us to do row operations of types I, II and III respectively via left or pre-multiplication. Hence they are called *elementary matrices*. For example,

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} c & d \\ a & b \end{pmatrix},$$

$$\begin{pmatrix} r & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} ra & rb \\ c & d \end{pmatrix},$$

and

$$\begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} a + sc & b + sd \\ c & d \end{pmatrix}.$$

The same idea works in general.

Definition 3.2. An $n \times n$ matrix obtained from I_n by performing a single row operation is called an *elementary $n \times n$ matrix*.

Here is the main point.

Proposition 3.4. *Let A be $n \times p$, and assume E is an elementary $n \times n$ matrix. Then EA is the matrix obtained by performing the row operation corresponding to E on A.*

Proof. Recall that for any $E \in \mathbb{R}^{n \times n}$, the rows of EA are linear combinations of the rows of A using the entries of $E = (e_{ij})$ as scalars. In fact, if $\mathbf{a}_i$ is the i-th row of A, then the i-th row of EA is

$$e_{i1}\mathbf{a}_1 + e_{i2}\mathbf{a}_2 + \cdots + e_{in}\mathbf{a}_n.$$

Thus, if E is obtained by interchanging the j-th and k-th rows of I_n, then in EA, the j-th and k-th rows of A have been interchanged, and the other rows

are unchanged. For example, suppose we wish to interchange the first and second rows of A. If we put $EA = B$, then we want to have $\mathbf{b}_1 = \mathbf{a}_2$, $\mathbf{b}_2 = \mathbf{a}_1$ and $\mathbf{b}_j = \mathbf{a}_j$ if $j \neq 1, 2$. Thus we let $e_{12} = e_{21} = 1$, $e_{11} = e_{22} = 0$ and let e_{ij} be the corresponding element of I_n for the other pairs (i, j). Similarly, if E is the matrix obtained by multiplying the i-th row of I_n by r, then in EA the i-th row of A has been multiplied by r, and all other rows are unchanged. The argument for the third type of row operation is similar. □

In fact, since $EI_n = E$, the matrix E performing the desired row operation is unique. We can now state an easy but important result.

Proposition 3.5. *An arbitrary $n \times p$ matrix A can be put into reduced form by a performing sequence of left multiplications on A using $n \times n$ elementary matrices. In other words, we can write $A_{red} = BA$, where B is a product of elementary $n \times n$ matrices.*

Proof. We know from Proposition 2.2 that any matrix can be put into reduced form by a sequence of row operations. But row operations are performed by left multiplication by elementary matrices.

□

This procedure can be represented as follows. First, replace A with $A_1 = E_1A$, then A_1 with $A_2 = E_2(E_1A)$ and so forth. Hence we get the sequence

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

the last matrix being A_{red}. This gives us a matrix

$$B = (E_k(E_{k-1}\cdots(E_1A)\cdots))$$

with the property that $BA = A_{red}$.

It needs to be emphasized that there is no reason B should be unique, since one can easily devise other sequences of row operations that reduce A, Nevertheless, it does turn out B is unique in certain cases. One of these is the case where A is a nonsingular.

By the associative law, we can express B without using parentheses, writing it simply as $B = E_kE_{k-1}\cdots E_1$.

Example 3.3. Let's compute the matrix B produced by the sequence of row operations in Example 2.5 which puts the counting matrix C in reduced form. Examining the sequence of row operations, we see that B is the

product

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

Thus

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

Be careful to express the product in the correct order. The first row operation is done by the matrix on the far right and the last by the matrix on the far left. Thus

$$BC = \begin{pmatrix} -5/3 & 2/3 & 0 \\ 4/3 & -1/3 & 0 \\ 1 & -2 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} = \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{pmatrix}.$$

That is, $BC = C_{red}$.

In the above computation, you do not need to explicitly multiply the elementary matrices out. Just 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 $(A|I_3)$ and carry out the sequence of row operations. The final result will be $(A_{red}|B)$. Thus if we start with

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

we end with

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

Example 3.4. To do another example, consider the matrix

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

with the catch that we consider A as a 3×3 matrix over $\mathbb{F}_2$. To row reduce A, we can use the following steps: $R_3 \to R_3+R_1$, $R_3 \to R_3+R_2$, $R_2 \to R_2+R_3$, and $R_1 \to R_1+R_3$. Hence the matrix B such that $BA = A_{red}$ is

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

Computing the product, we obtain

$$B = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \end{pmatrix}.$$

The reader should check that $BA = I_3$.

3.2.1 Application to Linear Systems

How does this method apply to solving a linear system $A\mathbf{x} = \mathbf{b}$? Starting with $A\mathbf{x} = \mathbf{b}$ and multiplying by an elementary matrix E gives a new linear system $EA\mathbf{x} = E\mathbf{b}$ equivalent to the original system (by Proposition 2.3). Continuing in this way, we obtain

Proposition 3.6. *Given a linear system $A\mathbf{x} = \mathbf{b}$, there exists a square matrix B which is a product of elementary matrices, such that the original system is equivalent to $A_{red}\mathbf{x} = B\mathbf{b}$.*

Proof. Just apply Propositions 3.5 and 2.3. □

The advantage of knowing the matrix B which brings A into reduced form is that one can handle an arbitrary number of systems as easily as one. That is, one can just as easily solve a matrix linear equation $A\mathbf{X} = \mathbf{D}$, where $\mathbf{X} = (x_{ij})$ is a matrix of variables and $\mathbf{D} = (D_{jk})$ is a matrix of constants. If A is $m \times n$ and D has p columns, then X is $n \times p$ and D is $m \times p$. This matrix equation is equivalent to $A_{red}X = BD$.

Exercises

Exercise 3.13. Find the reduced row echelon form for each of the following matrices, which are assumed to be over $\mathbb{R}$:

$$A_1 = \begin{pmatrix} 1 & 1 & 0 \\ 2 & 3 & 1 \\ 1 & 2 & 1 \end{pmatrix}, \quad A_2 = \begin{pmatrix} 1 & 2 & -1 & 1 \\ 2 & 3 & 1 & 0 \\ 0 & 1 & 2 & 1 \end{pmatrix}, \quad A_3 = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}.$$

Exercise 3.14. Find matrices B_1, B_2 and B_3 which are products of elementary matrices such that B_iA_i is reduced, where A_1, A_2, A_3 are the matrices of Exercise 3.13.

Exercise 3.15. Find the reduced row echelon form for each of the following matrices, assuming each matrix is defined over $\mathbb{F}_2$:

$$C_1 = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix}, \quad C_2 = \begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix}, \quad C_3 = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}.$$

Exercise 3.16. Find matrices D_1, D_2 and D_3 defined over $\mathbb{F}_2$ which are products of elementary matrices such that D_iC_i is reduced, where C_1, C_2, C_3 are the matrices of Exercise 3.15.

Exercise 3.17. Prove carefully that if E is an elementary $n \times n$ matrix and F is the elementary matrix that performs the reverse operation, then $FE = EF = I_n$.

Exercise 3.18. Write down all the 3×3 elementary matrices E over $\mathbb{F}_2$. For each E, find the matrix F defined in the previous exercise such that $FE = EF = I_3$.

Exercise 3.19. List all the row reduced 2×3 matrices over $\mathbb{F}_2$.

Exercise 3.20. Let E be an arbitrary elementary matrix.

(i) Show that E^T is also an elementary matrix.

(ii) Explain how to compute AE.

Exercise 3.21. In this exercise, we will introduce column operations. The reader should base their answers on the case of row operations.

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

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

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

3.3 Matrix Inverses

Given an elementary $n \times n$ matrix E, one can easily see that there exists an elementary matrix F such that $FE = I_n$. A little thought will convince you that $EF = I_n$ as well. Doing a row operation then undoing it produces the same result as first undoing it and then doing it. Either way you are back to where you started. This essential property is generalized in the next

Definition 3.3. Suppose two $n \times n$ matrices A and B have the property that $AB = BA = I_n$. Then we say A is an *inverse* of B (and B is an inverse of A). A matrix with an inverse is said to be *invertible*.

Let us first show

Proposition 3.7. *Suppose $A \in \mathbb{R}^{n\times n}$ or $A \in (\mathbb{F}_2)^{n\times n}$ and A has an inverse B. Then B is unique.*

Proof. Suppose that A has two inverses B and C. Then

$$B = BI_n = B(AC) = (BA)C = I_nC = C.$$

Thus $B = C$, so the inverse is unique. □

Note that the associative law was used in the above proof in an essential way. From now on, we will use A^{-1} to denote the (unique) inverse of A, if it exists.

The reader should check that any elementary matrix E has an inverse, which is also elementary.

Example 3.5. For example, the inverses of the 2×2 elementary matrices 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

$$E_3 = \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix} \Rightarrow E_3^{-1} = \begin{pmatrix} 1 & -s \\ 0 & 1 \end{pmatrix}.$$

3.3.1 A Necessary and Sufficient Condition for Existence

Recall that a linear system $A\mathbf{x} = \mathbf{b}$ is called nonsingular if the coefficient matrix A is square of maximal rank. In general, we say that an $n \times n$ matrix A is *nonsingular* if A has rank n or, equivalently, $A_{red} = I_n$. We therefore have the

Proposition 3.8. *If an $n \times n$ matrix A is nonsingular, there exists an $n \times n$ matrix B such that $BA = I_n$.*

Proof. This follows from Proposition 3.4 since $A_{red} = I_n$. □

Thus a nonsingular matrix has a *left inverse*. We will now prove the main result on matrix inverses. It will tell us that in particular a left inverse of a square matrix is a two sided inverse.

Theorem 3.9. *Suppose $A \in \mathbb{R}^{n\times n}$ or $A \in (\mathbb{F}_2)^{n\times n}$. Then we have the following:*

(i) A is nonsingular if and only if A has a left inverse;

(ii) if A has a left inverse B, then A is invertible and B is A's unique inverse (that is, if $BA = I_n$, then $AB = I_n$, and so $B = A^{-1}$); and

(iii) in particular, A is nonsingular if and only if it is invertible.

Proof. We'll suppose $A \in \mathbb{R}^{n\times n}$. The other case is handled just the same. We already know that if A is nonsingular, it has a left inverse. Hence suppose A has a left inverse B. To show that the rank of A is n, it suffices, by Proposition 2.6, to show that if $A\mathbf{x} = \mathbf{0}$, then $\mathbf{x} = \mathbf{0}$. Suppose $A\mathbf{x} = \mathbf{0}$, then

$$\mathbf{0} = B\mathbf{0} = B(A\mathbf{x}) = (BA)\mathbf{x} = I_n\mathbf{x} = \mathbf{x}. \tag{3.3}$$

Thus indeed, $\mathbf{x} = \mathbf{0}$, so A has rank n. Hence, if A has a left inverse, A is nonsingular, and (i) is finished. Next suppose A has a left inverse. Then, by (i), A has rank n, so we know the system $A\mathbf{x} = \mathbf{b}$ is consistent for every $\mathbf{b} \in \mathbb{R}^n$. Hence, the system $A\mathbf{x} = \mathbf{e}_i$ has a solution for each i, where $\mathbf{e}_i$ is the ith column of I_n. It follows that there exists an $n \times n$ matrix X so that $AX = I_n$. We now show that $B = X$. In fact, repeating the argument in Proposition 3.7, we see that

$$B = BI_n = B(AX) = (BA)X = I_nX = X.$$

Thus, if A has a left inverse, it has an inverse, proving (ii). Finally, suppose $BA = I_n$. Then A has rank n, so A has an inverse, say C. Repeating the

argument just given replacing X with C, it follows that $B = C$. Thus a left inverse of A is necessarily A^{-1}. □

One of the appealing applications of Theorem 3.9 is the following formula for the unique solution of a nonsingular system.

Corollary 3.10. *If A nonsingular, then the system*

$$A\mathbf{x} = \mathbf{b}$$

has the unique solution $\mathbf{x} = A^{-1}\mathbf{b}$.

Proof. This is an exercise. □

Notice that this solution is analogous to the solution of $ax = b$ which one may express as $x = b/a$ when $a \neq 0$. The difference is that $\mathbf{b}/A$ is not defined. We have to express the solution in the only sensical way, namely $\mathbf{x} = A^{-1}\mathbf{b}$.

The product of any two invertible $n \times n$ matrices A and B is also invertible. Indeed, $(AB)^{-1} = B^{-1}A^{-1}$. For

$$(B^{-1}A^{-1})AB = B^{-1}(A^{-1}A)B = B^{-1}I_nB = B^{-1}B = I_n.$$

This is used in the proof of the following useful Proposition.

Proposition 3.11. *Any product of elementary matrices is invertible, and, conversely, any invertible matrix is a product of elementary matrices.*

The proof is left as an exercise.

3.3.2 Methods for Finding Inverses

We have two ways of finding the matrix B so that $BA = A_{red}$. The first is simply to multiply out the sequence of elementary matrices which row reduces A. This is not as bad as it sounds since multiplying elementary matrices is elementary. The second method is to form the augmented matrix $(A \mid I_n)$ and row reduce. The final result will be in the form $(I_n \mid B)$. This is the method used in most textbooks. Let's begin with an example.

Example 3.6. Suppose we want to find an inverse for

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

Since we only need to solve the matrix equation $XA = I_3$, we can use our previous strategy of row reducing $(A \mid I_3)$.

$$(A \mid I_3) = \begin{pmatrix} 1 & 2 & 0 & 1 & 0 & 0 \\ 1 & 3 & 1 & 0 & 1 & 0 \\ 0 & 1 & 2 & 0 & 0 & 1 \end{pmatrix} \to \begin{pmatrix} 1 & 2 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & -1 & 1 & 0 \\ 0 & 1 & 2 & 0 & 0 & 1 \end{pmatrix} \to$$

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

$$\begin{pmatrix} 1 & 0 & 0 & 5 & -4 & 2 \\ 0 & 1 & 0 & -2 & 2 & -1 \\ 0 & 0 & 1 & 1 & -1 & 1 \end{pmatrix}.$$

Hence

$$A^{-1} = B = \begin{pmatrix} 5 & -4 & 2 \\ -2 & 2 & -1 \\ 1 & -1 & 1 \end{pmatrix},$$

since, by construction, $BA = I_3$.

Example 3.7. To take a slightly more interesting example, let

$$A = \begin{pmatrix} 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix},$$

where the entries of A are elements of $\mathbb{F}_2$. Using the above procedure, we see 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 may be (and generally 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 somewhat less obvious third technique which is sometimes also useful. If we form the *augmented coefficient matrix* $(A \mid \mathbf{b})$, where $\mathbf{b}$ represents the column vector with components $b_1, b_2, \ldots b_m$ and perform the row reduction of this augmented matrix, the result will be in the form $(I_n \mid \mathbf{c})$, where the components of $\mathbf{c}$ are certain linear combinations of the components of $\mathbf{b}$. The coefficients in these linear combinations give us the entries of A^{-1}. Here is an example.

Example 3.8. Again let

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

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,

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

3.3.3 Matrix Groups

In this section, we will introduce the concept of a matrix group and give a number of examples. Matrix groups are sometimes also called linear groups. The matrix group structure will be useful later in stating and proving some of the main results on matrices. The basic example of a matrix group is the set $GL(n, \mathbb{R})$ of all invertible elements of $\mathbb{R}^{n \times n}$. That is,

$$GL(n, \mathbb{R}) = \{A \in \mathbb{R}^{n \times n} \mid A^{-1} \text{ exists}\}. \tag{3.4}$$

Notice that, by definition, every element in $GL(n, \mathbb{R})$ has an inverse. Moreover, I_n is an element of $GL(n, \mathbb{R})$, and if A and B are elements of $GL(n, \mathbb{R})$, then so is their product AB. These three properties define what we mean by a matrix group.

Definition 3.4. A subset G of $\mathbb{R}^{n\times n}$ is called a *matrix group* if the following three conditions hold:

(i) if $A, B \in G$, then $AB \in G$,

(ii) $I_n \in G$, and

(iii) if $A \in G$, then $A^{-1} \in G$.

It turns out that these three axioms are enough to give the class of matrix groups an extremely rich structure. Of course, as already noted above, $GL(n, \mathbb{R})$ is a matrix group (commonly called the *general linear group*). In fact, if $G \subset \mathbb{R}^{n\times n}$ is a matrix group, then, by definition, $G \subset GL(n, \mathbb{R})$. A subset of $GL(n, \mathbb{R})$ which is also a matrix group is called a *subgroup* of $GL(n, \mathbb{R})$. Thus matrix groups are subgroups of some $GL(n, \mathbb{R})$.

The simplest example of a subgroup of $GL(n, \mathbb{R})$ is $\{I_n\}$: this is the so called *trivial subgroup*. To get some more interesting interesting examples, let us consider permutation matrices.

Example 3.9 (Permutation Matrices)**.** A matrix P obtained from I_n by a finite (possibly vacuous) sequence of row swaps is called a *permutation matrix*. In other words, a permutation matrix is a matrix $P \in \mathbb{R}^{n\times n}$ such that there are row swap matrices $S_1, \ldots, S_k \in \mathbb{R}^{n\times n}$ for which $P = S_1 \cdots S_k$. (Recall that a row swap matrix is by definition an elementary matrix obtained by interchanging two rows of I_n.) Clearly, I_n is a permutation matrix, and any product of permutation matrices is also a permutation matrix. It remains to see that the inverse of a permutation matrix is also a permutation matrix. Let $P = S_1 \cdots S_k$ be a permutation matrix. Then $P^{-1} = S_k^{-1} \cdots S_1^{-1}$. But every row swap S has the property that $S = S^{-1}$, so P^{-1} is indeed a permutation matrix, namely $S_k \cdots S_1$.

Let $P(n)$ denote the set of $n \times n$ permutation matrices. One can also describe $P(n)$ as the set of all matrices obtained from I_n by permuting the rows of I_n. Thus $P(n)$ is the set of all $n \times n$ matrices whose only entries are 0 or 1 such that every row and every column has exactly one non-zero entry. It follows from elementary combinatorics that $P(n)$ has exactly $n!$ elements.

The inverse of a permutation matrix has a beautiful expression.

Proposition 3.12. *If P is a permutation matrix, then $P^{-1} = P^T$.*

Proof. We leave this as an exercise. □

Example 3.10. $P(3)$ consists of the following five 3×3 permutation matrices and I_3:

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

The first, second and fifth matrices in the above list are row swaps, and the other two are products of row swaps.

Definition 3.5 (The orthogonal group)**.** Let $Q \in \mathbb{R}^{n \times n}$. Then we say that Q is *orthogonal* if and only if $Q^T Q = I_n$. The set of all $n \times n$ orthogonal matrices is denoted by $O(n, \mathbb{R})$. We call $O(n, \mathbb{R})$ the *orthogonal group*.

Proposition 3.13. *$O(n, \mathbb{R})$ is a subgroup of $GL(n, \mathbb{R})$.*

Proof. It follows immediately from the definition and Theorem 3.9 that if Q is orthogonal, then $Q^T = Q^{-1}$. Thus $Q = (Q^{-1})^T$, and we see that $(Q^{-1})^T Q^{-1} = QQ^{-1} = I_n$. Consequently, if Q is orthogonal, then Q^{-1} is orthogonal. The identity I_n is clearly orthogonal, so it remains to show that the product of two orthogonal matrices is orthogonal. Let Q and R be orthogonal. Then

$$(QR)^T(QR) = (R^T Q^T)(QR) = R^T(Q^T Q)R = R^T I_n R = I_n.$$

Hence $O(n, \mathbb{R})$ is a subgroup of $GL(n, \mathbb{R})$. □

By Proposition 3.12, we have $P(n) \subset O(n, \mathbb{R})$. That is, every permutation matrix is orthogonal. Hence $P(n)$ is also a subgroup of $O(n, \mathbb{R})$.

The orthogonal group for $O(2, \mathbb{R})$ is especially interesting. It has an important subgroup which is denoted as $SO(2)$ and called the *rotation group* because rotation matrices act on $\mathbb{R}^2$ as rotations. This subgroup consists of the matrices

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

The fact that $SO(2)$ is a subgroup of $O(2, \mathbb{R})$ follows from trigonometry. For example, the sum formulas for $\cos(\theta + \mu)$ and $\sin(\theta + \mu)$, are equivalent to the formula

$$R_\theta R_\mu = R_\mu R_\theta = R_{\theta+\mu}. \tag{3.5}$$

We will investigate other aspects of $O(2, \mathbb{R})$ in the Exercises.

Exercises

Exercise 3.22. Find the inverse of each of the following real matrices or show that the inverse does not exist.

(a) $\begin{pmatrix} 1 & 2 \\ 4 & 1 \end{pmatrix}$ (b) $\begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & -1 \\ 1 & 1 & 0 \end{pmatrix}$ (c) $\begin{pmatrix} 1 & 0 & -2 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \end{pmatrix}$ (d) $\begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & -1 \\ 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix}$.

Exercise 3.23. 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 3.24. If possible, find the inverse of $A = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \end{pmatrix}$ over $\mathbb{F}_2$.

Exercise 3.25. Let $A = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}$. Show that A has an inverse over $\mathbb{R}$ but A does not have an inverse over $\mathbb{F}_2$.

Exercise 3.26. Determine whether the following 5×5 matrix A over $\mathbb{F}_2$ has an inverse, and if it does find it:

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

Exercise 3.27. 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 3.28. Suppose A has an inverse. Find a formula for the inverse of A^T?

Exercise 3.29. Prove Proposition 3.11. That is, show that A is invertible if and only if A is a product of elementary matrices.

Exercise 3.30. Suppose A is $n \times n$ and there exists a right inverse B, i.e. $AB = I_n$. Show A invertible.

Exercise 3.31. 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 3.32. Show that if A and B are $n \times n$ and have inverses, then $(AB)^{-1} = B^{-1}A^{-1}$. What is $(ABCD)^{-1}$ if all four matrices are invertible?

Exercise 3.33. Suppose A is invertible $m \times m$ and B is $m \times n$. Solve the equation $AX = B$.

Exercise 3.34. Suppose A and B are both $n \times n$ and AB is invertible. Show that both A and B are invertible. (See what happens if $B\mathbf{x} = \mathbf{0}$.)

Exercise 3.35. 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 3.36. Let A and B be $n \times n$ matrices over $\mathbb{R}$.

(i) If the inverse of A^2 is B, show that the inverse of A is AB.

(ii) If 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 3.37. Is it TRUE or FALSE that if an $n \times n$ matrix with integer entries has an inverse, then the inverse also has integer entries?

Exercise 3.38. Show that a symmetric orthogonal matrix is its own inverse.

Exercise 3.39. 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: compute Q^TQ.)

Exercise 3.40. Using the sum formulas for $\cos(\theta + \mu)$ and $\sin(\theta + \mu)$, prove that $R_\theta R_\mu = R_{\theta+\mu}$ for all real numbers θ and μ.

Exercise 3.41. Two vectors $\mathbf{x},\ \mathbf{y} \in \mathbb{R}^2$ are said to be *orthogonal* if $\mathbf{x}^T\mathbf{y} = 0$: that is, $x_1y_1 + x_2y_2 = 0$. Using the definition of an orthogonal matrix, prove that the columns of an orthogonal matrix are orthogonal as are the rows, and, furthermore, each column and row has length one. (Note: the square of the length of $\mathbf{x}$ is $\mathbf{x}^T\mathbf{x}$.)

Exercise 3.42. * Show that every element H of $O(2,\mathbb{R})$ that isn't a rotation matrix satisfies $H^T = H$, $H^2 = I_2$ and $H \neq I_2$.

Exercise 3.43. 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 of S_1 and S_2.

Exercise 3.44. Let S_1 and S_2 be the permutation matrices defined in Exercise 3.43. Show that $(S_1S_2)^3 = I_3$.

Exercise 3.45. Show that every permutation matrix is orthogonal. Deduce that if P is a permutation matrix, then $P^{-1} = P^T$. This proves Proposition 3.12.

Exercise 3.46. 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 3.47. You are a code-breaker (more accurately, a cryptographer) assigned to crack a secret cipher constructed as follows. The sequence 01 represents A, 02 represents B and so forth up to 26, which represents Z. A space between words is indicated by inserting 00. A text can thus be encoded as a sequence. For example, 19080400020907000415 07 stands for "the big dog". We can think of this as a vector in $\mathbb{R}^{22}$. Suppose a certain text has been encoded as a sequence of length 14,212=44×323, and the sequence has been broken into 323 consecutive intervals of length 44. Next, suppose each sub-interval is multiplied by a single 44×44 matrix C. The new sequence obtained by laying the products end to end is called the cipher text, because it has now been enciphered, 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 sub-intervals will you need to decipher to break the whole cipher by deducing the matrix C?

3.4 The $LPDU$ Factorization

Recall that an invertible matrix A in $\mathbb{R}^{n\times n}$ can be expressed as a product of elementary $n \times n$ matrices. In this section, we will prove a much more explicit and general result: every $n \times n$ matrix A 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 for understanding the properties of matrices. For example, we will use it below to show that the reduced row echelon form of a matrix is unique. It will also be important when we determine the signs of the eigenvalues of an arbitrary symmetric matrix in Chapter 12. The $LPDU$ decomposition is an important theoretical tool for research in matrix theory, and it is widely used for solving large systems of linear equations.

3.4.1 The Basic Ingredients: $L,\ P,\ D$ and U

Let us now introduce the cast of characters in the $LPDU$ decomposition. First of all, a matrix A is called *lower triangular* if all the entries of A strictly above the diagonal are zero. Put another way, $a_{ij} = 0$ if $i < j$. SIMILARLY, A is *upper triangular* if its entries strictly below the diagonal are zero. Clearly, the transpose of a lower triangular matrix is upper triangular and vice versa. A square matrix A which is either upper or lower triangular and such that each diagonal entry $a_{ii} = 1$ is called *unipotent*. In our cast, the L's will be lower triangular unipotent, and the U's will be upper triangular unipotent.

Example 3.11. 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 $U = L^T$ is

$$U = \begin{pmatrix} 1 & a & b \\ 0 & 1 & c \\ 0 & 0 & 1 \end{pmatrix}.$$

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.

Notice that a type III row operations where a row is replaced by itself plus a multiple of a higher row is performed via left multiplication by a lower triangular unipotent matrix. We will usually call these downward row operations. Here is a basic fact.

Proposition 3.14. *The class $\mathcal{L}_n$ of all lower triangular unipotent $n \times n$ matrices is a subgroup of $GL(n, \mathbb{R})$. Similarly, the class $\mathcal{U}_n$ of all upper triangular unipotent matrices is also a subgroup of $GL(n, \mathbb{R})$.*

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 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 $\mathcal{L}_n$ is a subgroup of $GL(n, \mathbb{R})$, it remains to show that the inverse of a lower triangular unipotent matrix A is also in $\mathcal{L}_n$. But this follows since inverting a lower triangular unipotent matrix only requires downward row operations. (Row swaps aren't needed since A is lower triangular, and dilations aren't needed either since lower triangular unipotent matrices only have 1's on the diagonal.) Thus, A^{-1} is a product of lower triangular elementary matrices of type III. But these are elements of $\mathcal{L}_n$, so A^{-1} is also in $\mathcal{L}_n$. The proof for $\mathcal{U}_n$ is similar. In fact, one can simply transpose the proof just given. □

As just noted, every lower triangular unipotent matrix is the product of downward row operations. Indeed, there exist $E_1, E_2, \ldots, E_k$ of this type such that $E_k \cdots E_2 E_1 L = I_n$. Therefore, $L = E_1^{-1} E_2^{-1} \cdots E_k^{-1}$. But the inverse of each E_i is also a lower triangular elementary matrix of type III. Hence, $L = E_1^{-1} E_2^{-1} \cdots E_k^{-1}$. The analogous fact holds for upper triangular unipotent matrices.

Continuing the introduction of the cast of characters, recall from Example 3.9, that an $n \times n$ matrix which can be expressed as a product of elementary matrices of type II (i.e. row swaps) is called a *permutation matrix*. We've already seen that the set $P(n)$ of $n \times n$ permutation matrices is a matrix group, and, moreover, the inverse of a permutation matrix P is P^T. The $n \times n$ permutation matrices are exactly those matrices which

can be obtained by rearranging the rows of I_n. We now make the following definition:

Definition 3.6. An $n \times n$ matrix which is obtained from a permutation matrix by replacing some of the 1s by 0s is called a *partial permutation matrix*. The set of $n \times n$ partial permutation matrices is denoted by Π_n.

Note that by definition, an invertible partial permutation matrix is a permutation matrix. Note also that the product of two $n \times n$ partial permutation matrices is also a partial permutation matrix. However, the partial permutation matrices don't form a matrix group (why?).

The last character D stands for a diagonal matrix. Recall a matrix $D = (d_{ij})$ is called *diagonal* if and only if $d_{ij} = 0$ whenever $i \neq j$. Since a diagonal matrix 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, the set of invertible diagonal matrices is a matrix subgroup of $GL(n, \mathbb{R})$. We will denote this matrix group by $\mathcal{D}_n$.

3.4.2 The Main Result

We now arrive at the main theorem, which gives a normal form for an arbitrary square matrix.

Theorem 3.15. *Every $n \times n$ matrix A over $\mathbb{R}$ or $\mathbb{F}_2$ can be expressed in the form $A = LPDU$, where L is lower triangular unipotent, P is a partial permutation matrix, D is an invertible diagonal matrix, and U is upper triangular unipotent. If A is invertible, then P is a full permutation matrix, and P and D are unique. In fact, the partial permutation matrix P is unique in all cases.*

This result can be expressed in the product form

$$\mathbb{R}^{n \times n} = \mathcal{L}_n \cdot \Pi_n \cdot \mathcal{D}_n \cdot \mathcal{U}_n.$$

That is, every $n \times n$ matrix is the product of the four types of matrices: $\mathcal{L}_n$, Π_n, $\mathcal{D}_n$ and $\mathcal{U}_n$. Specializing to the invertible matrices, we have

$$GL(n, \mathbb{R}) = \mathcal{L}_n \cdot P(n) \cdot \mathcal{D}_n \cdot \mathcal{U}_n.$$

In particular, $GL(n, \mathbb{R})$ is the product of its four subgroups: $\mathcal{L}_n$, $P(n)$, $\mathcal{D}_n$ and $\mathcal{U}_n$. The $\mathbb{F}_2$ case is actually easier since there are only three types of matrices involved.

The idea behind the proof is very simple. Starting from A, one do downwaresd row operations on the left and rightward column operations on

the right until all that is left of A is PD. Before reading the proof, the reader may wish to look at the examples after the proof to see explicitly what to do.

Proof. Let an arbitrary $n \times n$ A be given. If the first column of A consists entirely of zeros, go immediately to the second column. Otherwise, let a_{i1}be the first nonzero entry in A's first column. Perform a sequence of row operations to make the entries below a_{i1} equal to zero. This transforms the first column of A into

$$(0, \ldots, 0, d_1, 0, \ldots, 0)^T, \tag{3.6}$$

where $d_1 = a_{i1}$. This reduction is performed by downward row operations: in other words, by pre-multiplying A by a sequence of lower triangular elementary matrices of type III. By Proposition 3.14, we therefore obtain a lower triangular unipotent matrix L_1 so that the first column of L_1A has the form (3.6). The next step is to use the first non zero entry d_1 in the first column to annihilate all the entries in the i-th row of A to the right of the first column. Since post multiplying by elementary matrices performs column operations, this amounts to multiplying L_1A on the right by a sequence of upper triangular unipotent matrices. This produces an upper triangular unipotent matrix U_1 such that the first column of $(L_1A)U_1$ has the form (3.6) and the i-th row is

$$(d_1, 0, \ldots, 0). \tag{3.7}$$

We now have the first column and i-th row of A in the desired form and from now on, they will be unchanged. If the first column of A is zero, we will for convenience put $L_1 = U_1 = I_n$.

To continue, scan to the right until we find a nonzero column in L_1AU_1, say this is the j-th. Let b_{kj} be its first nonzero entry, and put $d_j = b_{kj}$. Of course, $k \neq i$. Then we can find lower and upper triangular unipotent matrices L_2 and U_2 such that the first and j-th columns of $L_2L_1AU_1U_2$ have the single nonzero entry d_1 and d_j, and the same holds for i-th and k-th rows. Furthermore, the columns between the first and j-th columns, if any, are all zero. Continuing, 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.

Since multiplying a matrix B on the right by a diagonal matrix D multiplies the i-th column of B by d_{ii}, any matrix C with the property that every row and column has at most one nonzero entry can be written $C = PD$, where P is a partial permutation matrix and D is an invertible diagonal

matrix. If C is invertible, then P is a full permutation matrix, and D and P are unique. In fact, d_{ii} is the nonzero entry in C's i-th column. If C is singular, D isn't unique. It follows that $L'A'U = PD$, where P is a partial permutation matrix and D is a diagonal matrix.

We now show that the partial permutation matrix P is always unique. Let A have two decompositions

$$A = L_1P_1D_1U_1 = L_2P_2D_2U_2 \tag{3.8}$$

according to the Theorem. Then we can write

$$LP_1D_1 = P_2D_2U, \tag{3.9}$$

where $L = (L_2)^{-1}L_1$ and $U = U_2(U_1)^{-1}$. As L is lower triangular unipotent and U is upper triangular unipotent, LP_1D_1 can have nonzero elements below a nonzero entry of P_1, but no nonero entries to the right of a nonzero entry of P_1. The reverse is the case for P_2D_2U, so the only conclusion is that $LP_1D_1 = P_1D_1$ and $P_2D_2 = P_2D_2U$. But this implies $P_1D_1 = P_2D_2$. Since P_1 and P_2 are partial permutation matrices and D_1 and D_2 are invertible, P_1 and P_2 have to coincide.

Finally, if A is invertible and $A = LPDU$, then P is invertible, so P is a permutation matrix. We still have to show that if A is invertible, then D is unique, Suppose (3.9) holds. Repeating the previous argument, we see that $P_1D_1 = P_2D_2$. But $P_1 = P_2 = P$, and P is invertible, so it follows immediately that $D_1 = D_2$. □

Note that the above proof even gives an algorithm for finding the $LPDU$ factorization.

Example 3.12. To illustrate, let

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

Since the first non zero entry in the first column of A is $a_{13} = -1$, we can put $L_1 = I_3$. Then next two steps are to subtract the first column twice from the second and to subtract it once from the third. The result is

$$AU_1 = L_1AU_1 = \begin{pmatrix} 0 & 2 & -2 \\ 0 & 4 & -5 \\ -1 & 0 & 0 \end{pmatrix},$$

where

$$U_1 = \begin{pmatrix} 1 & -2 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Next we subtract twice the first row from the second, which gives

$$L_2 L_1 A U_1 = \begin{pmatrix} 0 & 2 & -2 \\ 0 & 0 & -1 \\ -1 & 0 & 0 \end{pmatrix},$$

where

$$L_2 = \begin{pmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Finally, we add the second column to the third, getting

$$L_2 L_1 A U_1 U_2 = \begin{pmatrix} 0 & 2 & 0 \\ 0 & 0 & -1 \\ -1 & 0 & 0 \end{pmatrix},$$

with

$$U_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}.$$

Now

$$U = U_1 U_2 = \begin{pmatrix} 1 & -2 & -3 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}.$$

Also,

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

After computing $L = (L_2 L_1)^{-1} = L_2^{-1}$ and $U = (U_1 U_2)^{-1}$, we obtain the $LPDU$ factorization

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

Here is another example.

Example 3.13. Let A be the matrix

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

Then

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

Thus

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

Hence

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

This has the form $LAU = P$, so $A = L^{-1}PU^{-1}$, so the Example is complete once L^{-1} and U^{-1} have been calculated. We leave this to the reader.

3.4.3 Further Uniqueness in $LPDU$

If A is invertible, the Theorem says that P and D in $A = LPDU$ is unique. The i-th diagonal entry d_{ii} of D is called the i-th *pivot* of A. It turns out that the pivots of A have quite a bit of significance. (See Chapter 12.)

Example 3.14. 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

$$A = \begin{pmatrix} 1 & 0 \\ -c/a & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a & 0 \\ 0 & (ad-bc)/a \end{pmatrix} \begin{pmatrix} 1 & -b/a \\ 0 & 1 \end{pmatrix}.$$

However, 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

$$PDU = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} c & 0 \\ 0 & b \end{pmatrix} \begin{pmatrix} 1 & d/c \\ 0 & 1 \end{pmatrix}.$$

This tells us that, in general, L and U aren't necessarily unique.

3.4.4 Further Uniqueness in $LPDU$

If A is invertible, the Theorem says that P and D in $A = LPDU$ is unique. The i-th diagonal entry d_{ii} of D is called the i-th *pivot* of A. It turns out that the pivots of A have quite a bit of significance. (See Chapter 12.)

The general two by two case occurs when $a_{11} \neq 0$. Here, the permutation matrix P turns out to be I_2. In this case, A can be row reduced without row swaps. For a general $n \times n$ A, $a_{11} \neq 0$, since having $a_{11} = 0$ puts a condition on A. Similarly, after putting the first column of A in the form $(1, 0, \ldots, 0)^T$, the new $(2, 2)$ entry will in general still be nonzero. Continuing in this way, we see that a sufficiently general matrix A will have permutation matrix $P = I_n$. This is exactly the situation where A can be row reduced without row swaps.

The next Proposition points out an extremely nice property of the general case.

Proposition 3.16. *If an invertible matrix A admits an LDU decomposition (i.e. $P = I_n$), then the matrices L, D and U are unique.*

Proof. We already know D is unique. So if A has two LDU decompositions, we have

$$A = L_1 D U_1 = L_2 D U_2.$$

Thus

$$L_1^{-1} L_2 = D U_1 U_2^{-1} D^{-1}. \tag{3.10}$$

But in (3.10), the left hand side is lower triangular unipotent and the right hand side is upper triangular, since D is diagonal. This tells us immediately that $L_1^{-1} L_2 = I_n$. Hence $L_1 = L_2$, so it also follows that $U_1 U_2^{-1} = D^{-1} D = I_n$. Hence $U_1 = U_2$ too, and the proof is finished. □

Going back to the 2×2 case considered in Example 3.14, the LDU decomposition for A is therefore unique when $a \neq 0$. We also pointed out in the same example that if $a = 0$, then L and U are not unique, although P and D are.

Now consider an $n \times n$ system $A\mathbf{x} = \mathbf{b}$. If A is invertible, solving consists of finding A^{-1}. If we write $A = LPDU$, 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}, both of which are expressed by simple formulas. In the non invertible case, we put $\mathbf{x} = (DU)^{-1}\mathbf{y}$. The system then becomes $LP\mathbf{y} = \mathbf{b}$, which is equivalent to $P\mathbf{y} = L^{-1}\mathbf{b}$ and easily solved.

With a linear system $A\mathbf{x} = \mathbf{b}$, where $A = LPDU$ is invertible, one can avoid having to deal with the permutation matrix P. In fact, it is always possible to post multiply A by another permutation matrix Q, suitably concocted to move zero pivots out of the way by switching columns, so as to get a matrix AQ which which has a factorization $AQ = L'D'U'$. The only affect on the system is renumbering the variables. As above, let $\mathbf{y} = Q^{-1}\mathbf{x} = Q^T\mathbf{x}$. Then

$$A\mathbf{x} = A(QQ^{-1})\mathbf{x} = A(QQ^T)\mathbf{x} = (AQ)\mathbf{y} = (L'D'U')\mathbf{y},$$

so we only have to solve $(L'D'U')\mathbf{y} = \mathbf{b}$.

3.4.5 The symmetric LDU decomposition

Suppose A is an invertible symmetric matrix which 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 3.15. 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. 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 $D = \text{diag}(1, 2, -1)$.

Summarizing, we state

Proposition 3.17. *If A is an (invertible) $n \times n$ symmetric matrix whose $LPDU$ decomposition has $P = I_n$, then A can be uniquely factored in the form $A = LDL^T$.*

The interested reader may wish to consider what happens when an invertible symmetric matrix A has zero pivots (see Exercise 3.60).

3.4.6 $LPDU$ and Reduced Row Echelon Form

The purpose of this Section is to relate the $LPDU$ decomposition of A to its reduced row echelon form. We will ultimately show that the reduced row echelon form of A of an arbitrary $m \times n$ matrix is unique. Let us make a series of observations.

Observation 1. It is enough to assume $m = n$. Indeed, if $m < n$, one can make A $n \times n$ by adjoining $n - m$ rows of zeroes at the bottom of A, and if $n < m$, one can adjoin $m - n$ columns of zeroes on the right of A.

Observation 2. If $A = LPDU$, then the nonzero columns of P are exactly the columns where A has a corner. This follows from comparing the algorithm for row reducing A with the method for finding the $LPDU$ factorization.

Observation 3. Suppose the rows of the partial permutation matrix P are permuted to get a partial permutation matrix Q whose first k rows are nonzero, where k is the number of ones in P. Then QDU is reduced (but is not necessarily in reduced row echelon form), where D and U are the same as in $LPDU$.

Observation 4. Any reduced row echelon form of A has the form QU' for some upper triangular unipotent matrix U'. For we can write $QD = D'Q$ for some (invertible) diagonal matrix D'. Then replacing D' by I_n is the same thing as making the corner entries all ones.

Now we show

Proposition 3.18. *The reduced row echelon form of an $m \times n$ matrix is unique. In particular, the rank of a matrix is well defined.*

Proof. As noted in Observation 1, we can restrict ourselves to $n \times n$ matrices A. By post multiplying A by a suitable permutation matrix, say R, we may

assume that the matrix Q in Observation 3 has the form $Q = \begin{pmatrix} I_k & 0 \\ 0 & 0 \end{pmatrix}$. Thus any matrix in reduced row echelon form obtained from AR has the form

$$\begin{pmatrix} I_k & a \\ O_1 & O_2 \end{pmatrix},$$

where a is $k \times (n-k)$, O_1 is the $(n-k) \times k$ zero matrix and O_2 is the $(n-k) \times (n-k)$ zero matrix. Suppose

$$\begin{pmatrix} I_k & b \\ O_1 & O_2 \end{pmatrix}$$

is another such matrix obtained from A by row operations. Then there exists a nonsingular $n \times n$ matrix G such that

$$G \begin{pmatrix} I_k & a \\ O_1 & O_2 \end{pmatrix} = \begin{pmatrix} I_k & b \\ O_1 & O_2 \end{pmatrix}.$$

Now write

$$G = \begin{pmatrix} J & K \\ L & M \end{pmatrix},$$

where J is $k \times k$, K is $k \times (n-k)$, L is $k \times (n-k)$ and M is $(n-k) \times (n-k)$. Carrying out the multiplication, we see that $JI_k = I_k$ and $Ja = b$. But this means $a = b$, so the reduced row echelon form of AR is indeed unique. It follows that the reduced row echelon form of A is also unique, since the reduced row echelon form of A is just BR^{-1}, where B is the reduced row echelon form of AR. It also follows immediately that the rank of A is uniquely defined. □

Thus if $A = LPDU$, then the rank of A is the rank of its partial permutation matrix P. This leads to an interesting, even surprising, Corollary.

Corollary 3.19. *For any matrix A, the rank of A equals the rank of A^T.*

We'll leave the proof as an Exercise.

The reason this Corollary is surprising is that there isn't any obvious connection between the reduced row echelon form of A and that of A^T. Nevertheless, we've done a lot of work to get the $LPDU$ factorization, so we should expect some payoffs. Another quite different proof of Proposition 3.18 is given in Chapter 5.

Exercises

Exercise 3.48. Find the $LPDU$ decompositions of the following matrices:

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

Exercise 3.49. Find the $LPDU$ factorization of the $\mathbb{F}_2$ matrix

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

Exercise 3.50. Find the $LPDU$ factorization of the $\mathbb{F}_2$ matrix

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

Exercise 3.51. 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 elementary matrices of type III.

Exercise 3.52. Find the 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 3.53. Show directly that an invertible upper triangular matrix B can be expressed $B = DU$, where D is a diagonal matrix with non zero diagonal entries and U is upper an triangular matrix all of whose diagonal entries are ones. Is this still true if B is singular?

Exercise 3.54. 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 3.55. Find a 3×3 matrix A such that the matrix L in the $A = LPDU$ decomposition isn't unique.

Exercise 3.56. Let A be $n \times n$, say $A = LPDU$. Show how to express the $LPDU$ decomposition of A^T.

Exercise 3.57. Assume A is symmetric and has an LDU decomposition. Show that if all the diagonal entries of D are non-negative, then A can be written $A = CC^T$, where C is lower triangular. This expression is called the Cholesky decomposition of A. The Cholesky decomposition is frequently used in biostatistics, where A may typically be at least 5000×5000.

Exercise 3.58. Find the LDU decomposition of the matrix

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

Exercise 3.59. 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 3.60. Prove the following:

Proposition 3.20. *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 3.61. Let P be a partial permutation matrix. Show that the rank of P equals the rank of P^T. (Hint: how is the rank related to the number of 1s?)

Exercise 3.62. Let A be $n \times n$ and write $A = LPDU$. Show that the rank of A is the same as the rank of P. Deduce from this that the rank of A is the rank of A^T. (Hint: the rank of $LPDU$ is the same as the rank of PDU. But DU is upper triangular of rank n. If the ith row of P is zero, then so is the ith row of PDU. This implies P and PDU have the same rank.)

Exercise 3.63. If A is nonsingular, what is the rank of AB? How about BA?

Exercise 3.64. True or False: If A and B are $n \times n$, then AB and BA have the same rank. Explain your reasoning. For example, if F give a counter-example.

3.5 Summary

We began this Chapter with the notion of matrix multiplication. To form the product AB, the number of rows of B must be the same as the number of columns of A. We saw that matrix multiplication is associative. We introduced elementaary matrices and showed that left multiplication by an elementary matrix performs a row operation. Thus matrices can be row reduced by pre-multiplication by elementary matrices. This leads naturally to the notion of the inverse of an $n \times n$ matrix A, which is a matrix B such that $AB = BA = I_n$. We saw $BA = I_n$ is enough to guarantee $AB = I_n$ also, and we also saw that the invertible $n \times n$ matrices are exactly those of rank n. A key fact is that a square linear system $A\mathbf{x} = \mathbf{b}$ with invertible coefficient matrix has the unique solution $\mathbf{x} = A^{-1}\mathbf{b}$.

After discussing inverses, we introduced matrix groups, or as they are also known, linear groups, and gave several examples. We then applied the notion of a matrix group to find way of factoring an arbitrary matrix into the form $LPDU$. This is an often used method both in applied mathematics, for solving large systems, and in pure mathematics, in the study of matrix groups and other branches of algebra.

Chapter 4

Fields and Vector Spaces

In the first two chapters, we considered linear equations and matrices over the reals. We also introduced the off-on field $\mathbb{F}_2 = \{0, 1\}$, where $1 + 1 = 0$, to give an example where we can solve linear systems and invert matrices without having to rely on the real numbers. We will begin this chapter with the introduction of the general notion of a *field*, which generalizes both the reals $\mathbb{R}$ and $\mathbb{F}_2$. This will immediately give us a whole new way considering of matrices, matrix algebra and, of course, linear systems.

Our second goal here is to introduce the notion of an abstract vector space, which generalizes both the set of all real n-tuples $\mathbb{R}^n$ and the set of all n-bit strings $(\mathbb{F}_2)^n$. We will make a few other general definitions, and, finally, conclude by considering a special class of vector spaces known as inner product spaces.

4.1 What is a Field?

The goal of this section is to define the notion of a field and to give some of the basic examples: the rational numbers, the smallest field containing the integers, the real numbers and the complex numbers, unquestionably the most important of all the fields we will consider. We will also introduce the prime fields $\mathbb{F}_p$. These are the finite fields defined by arithmetic modulo a prime p.

4.1.1 The Definition of a Field

Since algebra is the business of solving equations, let's consider the most trivial equation $ax = b$. First consider what happens if a and b are elements

of the integers $\mathbb{Z}$, say $a = 4$ and $b = 3$. Thus we want to solve the equation $4x = 3$. The algebraic operation we use for solving this problem is, of course, known as division, and it expresses x as $3/4$. This is the only solution, so we have to go outside of the of the integers to find a solution. Thus we introduce fractions, or quotients of integers.

The quotient r/s, where r, s are integers and $s \neq 0$ is called a *rational number*. The set of all rational numbers will be denoted by $\mathbb{Q}$, which reminds us of the term quotient. Two rationals r/s and u/v are equal if there is an integer k such that $r = ku$ and $s = kv$. Addition and multiplication in $\mathbb{Q}$ are defined by:

$$\frac{a}{b} + \frac{c}{d} = \frac{ad + bc}{bd}, \tag{4.1}$$

and

$$\frac{a}{b} \cdot \frac{c}{d} = \frac{ac}{bd}. \tag{4.2}$$

Clearly the sum and product of two rational numbers is a well defined rational number (since $bd \neq 0$ if $b, d \neq 0$).

The same problem would occur if we were given the less trivial job of solving a system of linear equations over the integers such as

$$\begin{aligned} ax + by &= m \\ cx + dy &= n, \end{aligned}$$

where a, b, c, d, m, n are all integers. Using Gaussian reduction to arrive at a solution, we will find that if $ad - bc \neq 0$, then

$$\begin{aligned} x &= \frac{dm - bn}{ad - bc} \\ y &= \frac{-cm + an}{ad - bc}. \end{aligned}$$

is a unique solution. Once again the rationals are needed.

More generally, solving any linear system addition, subtraction, multiplication and division. A field will be a set which has these four algebraic operations provided they satisfy certain other properties we haven't mentioned yet.

We will first define the notion of a *binary operation*. Addition and multiplication on the integers are two basic examples of binary operations. Let S be any set, finite or infinite. The *Cartesian product* of S with itself is the set $S \times S$ of all ordered pairs (x, y) of elements $x, y \in S$. Note, we call (x, y) an ordered pair to emphasize that $(x, y) \neq (y, x)$ unless $x = y$. Thus,

$$S \times S = \{(x, y) \mid x, y \in S\}.$$

Definition 4.1. A *binary operation* on S is a function F with domain $S \times S$ which takes its values $F(x, y)$ in S.

The notation for a function, or equivalently a mapping, F whose domain is a set A whose values are in a set B is $F : A \to B$. Thus an operation is a function $F : S \times S \to S$. We will often express a binary operation by writing something like $x \cdot y$ or $x * y$ for $F(x, y)$. So, for example, the operation of addition on $\mathbb{Z}$ is a function $S : \mathbb{Z} \times \mathbb{Z} \to \mathbb{Z}$ defined by $S(m, n) = m + n$. We also need the notion of a subset being closed with respect to a binary operation.

Definition 4.2. Let F be a binary operation on a set S. A subset T of S such that $F(x, y) \in T$ for allwhenever $x, y \in T$ is said to be *closed under the binary operation*.

For example, the positive integers are closed under both addition and multiplication. The odd integers are closed under multiplication, but not closed under addition since, for instance, $1 + 1 = 2$.

We now state the definition of a field.

Definition 4.3. Assume $\mathbb{F}$ is a set with two binary operations which are called addition and multiplication. The sum and product of two elements $a, b \in \mathbb{F}$ will be denoted by $a + b$ and ab respectively. Suppose addition and multiplication satisfy the following properties are satisfied for all $a, b, c \in \mathbb{F}$:

(*i*) $a + b = b + a$ (addition is commutative);

(*ii*) $(a + b) + c = a + (b + c)$ (addition is associative);

(*iii*) $ab = ba$ (multiplication is commutative);

(*iv*) $a(bc) = (ab)c$ (multiplication is associative);

(*v*) $a(b + c) = ab + ac$ (multiplication is distributive);

Suppose further that

(*vi*) $\mathbb{F}$ contains an element 0 called the *additive identity* such that $a+0 = a$ and an element 1 called the *multiplicative identity* such that $1a = a$ for all $a \in \mathbb{F}$;

(*vii*) $0 \neq 1$;

(*viii*) for every $a \in \mathbb{F}$, there is an element $-a$ called the *additive inverse* of a such that $a + (-a) = 0$; and

(ix) for every $a \neq 0$ in $\mathbb{F}$, there is an element a^{-1}, called the *multiplicative inverse* of a such that $aa^{-1} = 1$.

Then $\mathbb{F}$ is called a *field*.

We will write $a - b$ for $a + (-b)$. In particular, $a - a = 0$. In any field $\mathbb{F}$,

$$a0 = a(0 + 0) = a0 + a0,$$

For adding $-a0$ to both sides and using the associativity of addition, we get

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

Hence

$$a0 = 0$$

for all $a \in \mathbb{F}$. The converse of this fact is one of the most important properties of a field.

Proposition 4.1. *Let $\mathbb{F}$ be a field. Then $a0 = 0$ for all $a \in \mathbb{F}$. Moreover, whenever $ab = 0$, either $a = 0$ or $b = 0$. Put another way, if neither a nor b is zero, then $ab \neq 0$.*

Proof. The first claim was just proven. For the second claim, suppose $ab = 0$ but $a \neq 0$. Then

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

Hence $b = 0$, which completes the proof. □

The property that $ab = 0$ implies either $a = 0$ or $b = 0$ will be used repeatedly. Another basic fact is

Proposition 4.2. *In any field $\mathbb{F}$, the additive and multiplicative identities are unique. Moreover, the additive and multiplicative inverses are also unique.*

Proof. To show 0 is unique, suppose 0 and $0'$ are two additive identities. By definition,

$$0' = 0' + 0 = 0$$

so 0 is indeed unique. The proof that 1 is unique is similar. To see additive inverses are unique, let $a \in \mathbb{F}$ have two additive inverses b and c. By associativity,

$$b = b + 0 = b + (a + c) = (b + a) + c = 0 + c = c.$$

Thus $b = c$. The proof for multiplicative inverses is similar. □

It follows from the uniqueness that for all $a \in \mathbb{F}$,

$$-a = (-1)a.$$

For

$$0 = 0a = (1-1)a = 1a + (-1)a.$$

4.1.2 Arbitrary Sums and Products

In a field, we can take the sum and product of any finite collection of elements. However, we have to say how to define and interpret expressions such as

$$\sum_{i=1}^{k} x_i \quad \text{and} \quad \prod_{i=1}^{k} x_i.$$

Suppose we want to define the sum $x_1 + x_2 + \cdots + x_n$ of n arbitrary elements of a field $\mathbb{F}$. We accomplish this with mathematical induction. The sum or product of one element is unambiguously defined. So suppose the sum $x_1 + x_2 + \cdots + x_{n-1}$ has been unambiguously defined. Then put

$$x_1 + x_2 + \cdots + x_{n-1} + x_n = (x_1 + x_2 + \cdots + x_{n-1}) + x_n.$$

Likewise, put

$$x_1 x_2 \cdots x_n = (x_1 x_2 \cdots x_{n-1}) x_n.$$

Then it follows from induction that the sum and product of any number of elements is well defined. In fact, in the above sum and product, the parens can be put anywhere, as we now show.

Proposition 4.3. *In any field* $\mathbb{F}$,

$$x_1 + x_2 + \cdots + x_{n-1} + x_n = \Big(\sum_{i=1}^{r} x_i\Big) + \Big(\sum_{i=r+1}^{n} x_i\Big), \tag{4.3}$$

for any r *with* $1 \le r < n$. *Similarly,*

$$x_1 x_2 \cdots x_n = \Big(\prod_{i=1}^{r} x_i\Big)\Big(\prod_{j=r+1}^{n} x_j\Big), \tag{4.4}$$

for all r *with* $1 \le r \le n-1$. *Moreover, the sum and product on the left hand side of* (4.3) *and* (4.4) *respectively can be reordered in any manner.*

Proof. We will give the proof for sums and leave products to the reader, as the details in both cases are the same. We use induction on n. There is nothing to show for $n = 1$, so suppose $n > 1$ and the result is true for $n-1$. If $r = n-1$, there is also nothing to show. Thus assume $r < n-1$. Then

$$\begin{aligned} x_1 + x_2 + \cdots + x_{n-1} + x_n &= (x_1 + x_2 + \cdots + x_{n-1}) + x_n \\ &= \big(\sum_{i=1}^{r} x_i + \sum_{j=r+1}^{n-1} x_j\big) + x_n \\ &= \sum_{i=1}^{r} x_i + \big(\sum_{j=r+1}^{n-1} x_j + x_n\big) \\ &= \big(\sum_{i=1}^{r} x_i\big) + \big(\sum_{j=r+1}^{n} x_j\big) \end{aligned}$$

Hence the result is true for n, which completes the proof.

To see that the left hand side of (4.3) can we written in any order, let $y_1, y_2, \ldots, y_n$ be any reordering. Here $n > 1$ since otherwise. there's nothing to prove. Assume the result holds for $n-1$. Now $y_n = x_k$ for some index k, and we can assume $k < n$. By the first argument,

$$\begin{aligned} x_1 + x_2 + \cdots + x_n &= (x_1 + \cdots + x_k) + (x_{k+1} + \cdots + x_n) \\ &= (x_{k+1} + \cdots + x_n) + (x_1 + \cdots + x_k) \\ &= (x_{k+1} + \cdots + x_n + x_1 + \cdots + x_{k-1}) + x_k \\ &= (x_{k+1} + \cdots + x_n + x_1 + \cdots + x_{k-1}) + y_n \\ &= (y_1 + \cdots + y_{n-1}) + y_n \\ &= y_1 + \cdots + y_{n-1} + y_n. \end{aligned}$$

The next to last step uses the induction hypothesis since $y_1, y_2, \ldots, y_{n-1}$ forms a rearrangement of $x_1, \ldots, x_{k-1}, x_{k+1}, \ldots, x_n$. □

4.1.3 Examples

We now give some examples.

Example 4.1 ($\mathbb{Q}$)**.** The fact that the rationals satisfy all the field axioms is a consequence of the basic arithmetic properties of the integers: associativity, commutativity and distributivity and the existence of 0 and 1. Indeed, all one needs to do is to use (4.1) and (4.2) to prove the field axioms for $\mathbb{Q}$ from

these properties of the integers. Note that the integers $\mathbb{Z}$ are not a field, since field axiom (*viii*) isn't satisfied by $\mathbb{Z}$. The only integers which have multiplicative inverses are ± 1.

Example 4.2 ($\mathbb{R}$)**.** The second example of a field is the set of real numbers $\mathbb{R}$. The construction of the real numbers is actually somewhat technical, so we won't try to explain it. For most purposes, it suffices to think of $\mathbb{R}$ as being 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 $a_1 \neq 0$. Note that there can be infinitely many b_j to the right of the decimal point. We also have to make appropriate identifications for repeating decimals such as $1 = .999999\ldots$. A very useful property of the reals is the fact that they have an ordering $>$ such that any 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 $a_1 x_1 + a_2 x_2 + \cdots + a_n x_n > c$. The reals also have the *Archimedean property*: if $a, b > 0$, then there exists an $x > 0$ so that $ax > b$. In other words, linear inequalities have solutions.

Example 4.3 ($\mathbb{F}_2$)**.** The field $\mathbb{F}_2$ consisting of 0 and 1 was introduced in the previous chapter. The condition $1 + 1 = 0$ is forced on us, for $1 + 1 = 1$ would give $1 = 0$, violating the definition of a field. However, we haven't completely verified the field axioms for $\mathbb{F}_2$ since associativity hasn't been fully verified. We wll leave this as an exercise.

Another of the basic fields is $\mathbb{C}$, the complex numbers, but we will postpone discussing it until the next section.

4.1.4 An Algebraic Number Field

Many examples of fields arise by extending a given field. We will now give an example of a field called an algebraic number field which is obtained by adjoining the square root of an integer to the rationals $\mathbb{Q}$. In order to explain this example, let us first recall the

Theorem 4.4 (Fundamental Theorem of Arithmetic)**.** *Let m be an integer greater than 1. Then m can be factored $m = p_1 p_2 \cdots p_k$, where $p_1, p_2, \ldots, p_k$ are primes. Moreover, this factorization is unique up to the order of the factors.*

Recall that a positive integer p is called *prime* if $p > 1$ and its only positive divisors are 1 and itself. For a proof of the Fundamental Theorem

of Arithmetic, the reader is referred to a text on elementary number theory. We say that a positive integer m is *square free* if its prime factorization has no repeated factors. For example, $10 = 2 \cdot 5$ is square free while $12 = 4 \cdot 3$ isn't.

Let $m \in \mathbb{Z}$ be positive and square free, and let $\mathbb{Q}(\sqrt{m})$ denote the set of all real numbers of the form $a + b\sqrt{m}$, where a and b are arbitrary rational numbers. It is easy to see that sums and products of elements of $\mathbb{Q}(\sqrt{m})$ are also elements of $\mathbb{Q}(\sqrt{m})$. Clearly 0 and 1 are elements of $\mathbb{Q}(\sqrt{m})$. Hence, assuming the field axioms for $\mathbb{Q}$ allows us to conclude without any effort that all but one of the field axioms are satisfied in $\mathbb{Q}(\sqrt{m})$. We still have to prove that any non zero element of $\mathbb{Q}(\sqrt{m})$ has a multiplicative inverse.

So assume $a + b\sqrt{m} \neq 0$. Thus at least one of a or b is non zero. It suffices to assume the a and b are integers without common prime factors; that is, a and b are *relatively prime* . The trick is to notice that

$$(a + b\sqrt{m})(a - b\sqrt{m}) = a^2 - mb^2.$$

Hence, if $a^2 - mb^2 \neq 0$, then

$$\frac{1}{a + b\sqrt{m}} = \frac{a - b\sqrt{m}}{a^2 - mb^2},$$

so$(a + b\sqrt{m})^{-1}$ exists in $\mathbb{R}$ and by definition is an element of $\mathbb{Q}(\sqrt{m})$.

To see that indeed $a^2 - mb^2 \neq 0$, suppose not. Then

$$a^2 = mb^2.$$

This implies that m divides a^2, hence any prime factor p_i of m has to divide a itself. In other words, $a = cm$ for some $m \in \mathbb{Z}$. Hence $c^2m^2 = mb^2$, so $b^2 = mc^2$. Repeating the argument, we see that m divides b. This implies that the original assumption that a and b are relatively prime is violated, so $a^2 - mb^2 \neq 0$. Therefore we have proven

Proposition 4.5. *If m is a square free positive integer, then $\mathbb{Q}(\sqrt{m})$ is a field.*

The field $\mathbb{Q}(\sqrt{m})$ is in fact the smallest field containing both the rationals $\mathbb{Q}$ and $\sqrt{m}$.

Exercises

Exercise 4.1. Prove that in any field $(-1)a = -a$.

Exercise 4.2. Show directly that $\mathbb{F} = \{a + b\sqrt{2} \mid a, b \in \mathbb{Q}\}$ is a field under the usual operations of addition and multiplication in $\mathbb{R}$. Also, find $(1 - \sqrt{2})^{-1}$ and $(3 - 4\sqrt{2})^{-1}$.

Exercise 4.3. Let $\mathbb{Z}$ denote the integers. 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:

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

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

4.2 The Integers Modulo a Prime p

Let p denote an arbitrary prime. The purpose of this section is to define a field $\mathbb{F}_p$ with p elements for every prime p.

Let us first make a definition.

Definition 4.4. A field with only a finite number of elements is called a *Galois field*. A field with a prime number of elements is known as a *prime field*.

Since every field has to contain at least two elements, the simplest example of a Galois field is $\mathbb{F}_2$. Of course, $\mathbb{F}_2$ is also a prime field. To define a field $\mathbb{F}_p$ with p elements for any $p > 1$, we will use modular arithmetic, or arithmetic modulo p.

Everyone knows how to tell time, which is an example of using addition modulo 12. The thirteenth hour is 1pm, the fourteenth is 2pm and so forth. Arithmetic modulo p is addition and multiplication in which all that is used are remainders after division by p.

To begin, put

$$\mathbb{F}_p = \{0, 1, 2, \ldots, p-1\}. \tag{4.5}$$

The elements of $\mathbb{F}_p$ are integers, but they should be thought of as remainders. The way we will add them is as follows. If a and b in $\mathbb{F}_p$, first take their sum in the usual way to get the integer $a + b$. If $a + b < p$, then we define the sum $a + b$ in $\mathbb{F}_p$ to be $a + b$. But, if $a + b \geq p$, we use *division with remainder*. This is the principle explained in the next Proposition.

Proposition 4.6. *Suppose a and b are non-negative integers with $b \neq 0$. Then one can uniquely express a as $a = qb + r$, where q is a non-negative integer and $0 \leq r < b$.*

Proof. If $a < b$, put $r = a$, and if $a = b$, put $r = 0$. If $a > b$, some positive multiple sb of b satisfies $sb > a$. Let s be the least positive integer such that this happens. Here we are using the fact that every non-empty set of positive integers has a least element. Put $q = s - 1$. Thus $a = qb + (a - qb)$, so we are done if we show that $r = a - qb$ satisfies $0 \leq r < b$. Now $a - qb = a - (s-1)b \geq 0$ by definition. Also $a - qb = (a - sb) + b < b$ since $a < sb$. Thus $0 \leq r < b$, so the proof is finished. □

Thus, if $a + b \geq p$, write

$$a + b = qp + r,$$

where q is a nonnegative integer and r is an integer such that $0 \leq r < p$. Then the *sum of* a and b in $\mathbb{F}_p$ is defined to be r. This operation is called *addition modulo* p. It is a special case of *modular addition*. To define the *product of* a and b in $\mathbb{F}_p$, we use the remainder upon dividing ab by p in exactly the same way. Note that 0 and 1 are clearly additive and multiplicative identities.

Example 4.4. Let's carry out the definitions of addition and multiplication in $\mathbb{F}_3 = \{0, 1, 2\}$. Of course, 0 and 1 are always the identities, so all sums and products involving them are determined. To completely determine addition in $\mathbb{F}_3$, we only have to define $1 + 1$, $1 + 2$ and $2 + 2$. First of all, $1 + 1 < 3$, so by definition, $1 + 1 = 2$. To find $2 + 2$, first take the usual sum 4, then express $4 = 3 + 1$ as in Proposition 4.6. The remainder is 1, so $2 + 2 = 1$ in $\mathbb{F}_3$. Similarly, $1 + 2 = 0$ in $\mathbb{F}_3$. Thus $-2 = 1$ and $-1 = 2$. To find all products, it remains to find $2 \cdot 2$. But $2 \cdot 2 = 4$ in usual arithmetic, so $2 \cdot 2 = 1$ in $\mathbb{F}_3$. Thus $2^{-1} = 2$. A good way to summarize addition and multiplication is to construct addition and multiplication tables. The addition table for $\mathbb{F}_3$ is

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

We suggest that the reader construct the multiplication table for $\mathbb{F}_3$.

With the above definitions of addition and multiplication, we can prove

Theorem 4.7. *If p is a prime, then $\mathbb{F}_p$, as defined above, is a field.*

We will skip the proofs that addition and multiplication are commutative, associative and distributive. The existence of additive inverses is easy to see; the inverse of a is $p - a$. We have already noted that 0 and 1 are the identities, and clearly $0 \neq 1$. The rest of the proof involves showing that multiplicative inverses exist. This will require the basic facts about the integers stated above and an interesting diversion into some of the combinatorics of finite sets, namely the *Pigeon Hole Principle*.

Let us first make a general definition.

Definition 4.5. Let X and Y be sets, and let $\phi : X \to Y$ be a mapping. The set X is called the *domain* of ϕ and Y is called the *target* of ϕ. The mapping ϕ is called *one to one* or *injective* if $\phi(x) = \phi(x')$ implies $x = x'$. If every $y \in Y$ has the form $y = \phi(x)$ for some $x \in X$. then ϕ is said to be

onto or *surjective*. In other words, ϕ is surjective if the *image of* ϕ

$$\phi(X) = \{\phi(x) \mid x \in X\} \subset Y$$

of ϕ is Y. If ϕ is both injective and surjective, then ϕ is said to be a *bijection*. A bijection is also called a *one to one correspondence*.

If X is finite, the number of elements of X is denoted by $|X|$.

Proposition 4.8 (The Pigeon Hole Principle)**.** *Let X and Y be finite sets with $|X| = |Y|$, and suppose $\phi : X \to Y$ is a map. If ϕ is either injective or surjective, then ϕ is a bijection.*

Proof. If ϕ is injective, then X and its image $\phi(X)$ have the same number of elements. But this implies $\phi(X) = Y$, so ϕ is surjective, hence a bijection. On the other hand, suppose ϕ is surjective, i.e. $\phi(X) = Y$. Then $|X| \geq |Y|$. But if $\phi(x) = \phi(x')$ where $x \neq x'$, then in fact $|X| > |Y|$. This contradicts the assumption that $|X| = |Y|$, hence ϕ is a bijection. □

We now return to the proof that every nonzero element a of $\mathbb{F}_p$ has an inverse a^{-1}. First, we show multiplication in $\mathbb{F}_p$ satisfies the conclusion of Proposition 4.1:

Proposition 4.9. *Let p be a prime number. If $ab = 0$ in $\mathbb{F}_p$, then either $a = 0$ or $b = 0$ (or both).*

Proof. Since $ab = 0$ in $\mathbb{F}_p$ is the same thing as saying that p divides the usual product ab in $\mathbb{Z}$, the Proposition follows from the fact that if the prime number p divides ab, then p divides a or p divides b. This is an immediate consequence of the Fundamental Theorem of Arithmetic (Theorem 4.4). □

This Proposition implies that multiplication by a fixed non-zero element $a \in \mathbb{F}_p$ induces an injective map

$$\begin{aligned} \phi_a : \mathbb{F}_p \setminus \{0\} &\longrightarrow \mathbb{F}_p \setminus \{0\} \\ x &\longmapsto ax \end{aligned}$$

Here $\mathbb{F}_p \setminus \{0\}$ is the set $\mathbb{F}_p$ without 0. To see that ϕ_a is injective, let $\phi_a(x) = \phi_a(y)$, that is $ax = ay$. Thus $a(x - y) = 0$, so $x - y = 0$ since $a \neq 0$ (Proposition 4.9). Therefore ϕ_a is indeed injective. Since $F_p \setminus \{0\}$ is a *finite* set, the Pigeon Hole Principle says that ϕ_a is a bijection. In particular, there exists an $x \in \mathbb{F}_p \setminus \{0\}$, such that $ax = 1$. Hence we have shown the existence of an inverse of a.

If we relax the definition of a field by not assuming multiplicative inverses exist, the resulting system is called a *ring*. Every field is a ring, but there

exist rings such as the integers which aren't fields. For another example, consider $\mathbb{Z}_4$ to be $\{0, 1, 2, 3\}$ with addition and multiplication modulo 4. Then $\mathbb{Z}_4$ is a ring, but not a field since $2 \cdot 2 = 0$ in $\mathbb{Z}_4$ (hence 2 is not invertible). In fact, if q is a composite number, then the ring $\mathbb{Z}_q$ (defined in an analogous way) is a not a field. Note that the integers $\mathbb{Z}$ also form a ring which is not a field.

4.2.1 A Field with Four Elements

To show there exist fields that aren't prime fields, we will now construct a field $\mathbb{F} = \mathbb{F}_4$ with 4 elements. The method is to simply explicitly write down $\mathbb{F}$'s addition and multiplication tables. Let $0, 1, \alpha$, and β denote the elements of $\mathbb{F}$. The addition table is defined as follows. (Note that we have ignored addition by 0.)

$+$	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

Then we have

Proposition 4.10. *The set $\mathbb{F}_4 = \{0, 1, \alpha, \beta\}$ having 0 and 1 as identities and addition and multiplication defined as above is a field.*

The verification that $\mathbb{F}_4$ satifies the field axioms can be done by hand, and so we will omit it. A general way of constructing the Galois fields will be given in a later chapter.

Since $\alpha^2 = \beta$ and $\beta^2 = \alpha$, it follows that $\alpha^3 = \beta^3 = 1$. Hence $\alpha^4 = \alpha$ and $\beta^4 = \beta$, so all elements of $\mathbb{F}_4$ satisfy the equation $x^4 - x = 0$ since 0 and 1 trivially do. Now by Section 4.12 below, we can view $x^4 - x$ as a polynomial in a variable x over the field $\mathbb{F}_2$, where we have the identity $x^4 - x = x^4 + x$. Thus, we can factor $x^4 - x = x(x + 1)(x^2 + x + 1)$ (remember $1 + 1 = 0$ so $2x = 2x^2 = 0$). The elements α and β are the roots of $x^2 + x + 1 = 0$. We will give appropriate generalizations of these statements for all Galois fields in a later chapter.

We will prove in the next chapter that the number of elements in a Galois field $\mathbb{F}$ is always a power of a prime, i.e. is p^n for some prime p. This prime

is called the characteristic of the field and is the topic of the next section. In fact, it will be shown later that for every prime p and integer $n > 0$, there eists a Galois field with p^n elements, and any two Galois fields with the same number of elements are essentially the same.

4.2.2 The Characteristic of a Field

If $\mathbb{F}$ is a finite field, then some positive multiple r of the identity $1 \in \mathbb{F}$ has to be 0. Indeed, the positive multiples $r1$ of 1 can't all be different, so there have to be an $m > 0$ and $n > 0$ with say $m > n$ such that $m1 = n1$ in $\mathbb{F}$. But this implies $r1 = 0$ with $r = m - n$. I claim that the least positive integer r such that $r1 = 0$ is a prime. For if r can be expressed as a product $r = st$, where s, t are integers greater than 1, then $r1 = (st)1 = (s1)(t1) = 0$. As $\mathbb{F}$ is a field, it follows that either $s1 = 0$ or $t1 = 0$, contradicting the choice of r. Therefore r is a prime. We now make the following definition:

Definition 4.6. Let $\mathbb{F}$ be an arbitrary field. If $q1 = 0$ for some positive integer q, then, we say that $\mathbb{F}$ has *positive characteristic*. In that case, the least such q, which we showed above to be a prime, is called the *characteristic of* $\mathbb{F}$. If $q1 \neq 0$ for all $q > 0$, we say $\mathbb{F}$ has *characteristic* 0.

To summarize, we state

Proposition 4.11. *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 nonzero, then it's a prime. If $p1 = 0$, then $pa = p(1a) = (p1)a = 0a = 0$ for all $a \in \mathbb{F}$. □

Example 4.5. The characteristic of $\mathbb{F}_p$ is p. The characteristic of the field $\mathbb{F}_4$ of Example 4.2.1 is 2.

Proposition 4.12. *The characteristics of $\mathbb{Q}$, $\mathbb{R}$ and $\mathbb{C}$ are all 0. Moreover, the characteristic of any subfield of a field of characteristic 0 is also 0.*

The notion of the characteristic has a nice application.

Proposition 4.13. *If $\mathbb{F}$ is a field of characteristic $p > 0$, then for any $a_1, \ldots, a_n \in \mathbb{F}$,*

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

This can be proved by induction using the Binomial Formula, which says that if x and y are commuting variables and n is a positive integer,

$$(x + y)^n = \sum_{i=0}^{n} \binom{n}{i} x^{n-i} y^i, \tag{4.6}$$

where

$$\binom{n}{i} = \frac{n!}{(n-i)!i!}.$$

Another application of the characteristic is:

Proposition 4.14. *For every non zero $a \in \mathbb{F}_p$, $a^{p-1} = 1$. In particular, $a^{-1} = a^{p-2}$.*

Proof. This is an exercise. □

Example 4.6. For example, suppose we want to compute the inverse of 5 in $\mathbb{F}_{23}$. If you have a calculator handy, then you will see that $5^{21} = 476837158203125$, which is congruent to 14 modulo 23. Thus, $5^{-1} = 14$.

4.2.3 Connections With Number Theory

Proposition 4.14 implies a basic result in number theory (and conversely). First, we state the definition of a congruence.

Definition 4.7. Let a, b, c be integers. Then we say a is *congruent to b modulo c* if $a - b$ is divisible by c.

The congruence is expressed by writing $a \equiv b \bmod c$. Stating Proposition 4.14 in terms of congruences gives a classical result due to Fermat.

Theorem 4.15 (Fermat's Little Theorem)**.** *Suppose $p > 1$. Then for any integer $a \not\equiv 0 \bmod p$, $a^{(p-1)} \equiv 1 \bmod p$.*

There are further connections between properties of prime fields and elementary number theory. For any integers a and b which are not both 0, let $d > 0$ be the largest integer which divides both a and b. We call d the *greatest common divisor* of a and b. The greatest common divisor, or simply, gcd of a and b is traditionally denoted (a, b). For example, $(4, 10) = 2$. A basic fact about the gcd proved in any book on number theory is

Proposition 4.16. *Let a and b be integers which are not both 0, and let d be their gcd. Then there exist integers u and v such that $au + bv = d$. Conversely, if there exist integers u and v such that $au + bv = d$, then $d = (a, b)$.*

Definition 4.8. Let a, b, c be integers. Then we say a is *congruent to b modulo c* if $a - b$ is divisible by c. If a is congruent to b modulo c, we write $a \equiv b \bmod c$.

The following result gives another proof that non-zero elements of $\mathbb{F}_p$ have multiplicative inverses.

Proposition 4.17. *Let a, b, q be positive integers. Then the congruence equation $ax \equiv 1$ mod q has a solution if and only if $(a, q) = 1$.*

Fermat's Little Theorem suggests that one way to test whether m is prime is to see if $a^{(m-1)} \equiv 1 \bmod m$ for a few well chosen integers a. This doesn't give a foolproof test, but it is very good, and in fact it serves as the basis of the of some of the recent research in number theory on the topic of testing for primeness.

Exercises

Exercise 4.4. Prove that in any field $(-1)a = -a$.

Exercise 4.5. Show directly that $\mathbb{F} = \{a + b\sqrt{2} \mid a, b \in \mathbb{Q}\}$ is a field under the usual operations of addition and multiplication in $\mathbb{R}$. Also, find $(1 - \sqrt{2})^{-1}$ and $(3 - 4\sqrt{2})^{-1}$.

Exercise 4.6. Describe addition and multiplication for the field $\mathbb{F}_p$ having p elements for $p = 5$. That is, construct addition and multiplication tables for $\mathbb{F}_5$. Check that every element $a \neq 0$ has a multiplicative inverse.

Exercise 4.7. Construct addition and multiplication tables for $\mathbb{F}_7$, and use them to find both $-(6 + 6)$ and $(6 + 6)^{-1}$ in $\mathbb{F}_7$.

Exercise 4.8. Let $\mathbb{F}$ be a field and suppose that $\mathbb{F}' \subset \mathbb{F}$ is a subfield, that is, $\mathbb{F}'$ is a field for the operations of $\mathbb{F}$. Show that $\mathbb{F}$ and $\mathbb{F}'$ have the same characteristic.

Exercise 4.9. Show that the characteristic of $\mathbb{F}_p$ is p.

Exercise 4.10. Strengthen Proposition 4.11 by showing that if $\mathbb{F}$ is a field of characteristic $p > 0$, and $m1 =$ for some integer m, then m is a multiple of p.

Exercise 4.11. Suppose the field $\mathbb{F}$ contains $\mathbb{F}_p$ as a subfield. Show that the characteristic of $\mathbb{F}$ is p.

Exercise 4.12. Suppose that $\mathbb{F}$ has characteristic $p > 0$. Show that the multiples of 1 (including 0) form a subfield of $\mathbb{F}$ with p elements.

Exercise 4.13. Show that if $\mathbb{F}$ is a finite field of characteristic p, then for any $a, b \in \mathbb{F}$, we have $(a + b)^p = a^p + b^p$.

Exercise 4.14. Prove Proposition 4.13. That is, show that if $\mathbb{F}$ is a finite field of characteristic p, then for any $a_1, \ldots, a_n \in \mathbb{F}$,

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

Hint: Use induction with the aid of the result of Exercise 4.13 and the binomial theorem.

Exercise 4.15. Use Proposition 4.13 to show that $a^p = a$ for all $a \in \mathbb{F}_p$. Use this to deduce Proposition 4.14.

Exercise 4.16. In the definition of the field $\mathbb{F}_4$, could we have altered the definition of multiplication by requiring $\alpha^2 = \beta^2 = 1$, but leaving all the other rules as is, and still get a field?

Exercise 4.17. 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 4.18. Show that $\mathbb{F}$ is a finite field of characteristic p, then $\mathbb{F}$ is *perfect*. That is, every element in $\mathbb{F}_p$ is a pth power. (Hint: use the Pigeon Hole Principle.)

Exercise 4.19. Use Fermat's Theorem to find 9^{-1} in $\mathbb{F}_{13}$. Use this to solve the equation $9x \equiv 15 \bmod 13$.

Exercise 4.20. Find at least one primitive element β for $\mathbb{F}_{13}$? (Calculators should be used here.) Also, express 9^{-1} using this primitive element instead of Fermat's Theorem.

Exercise 4.21. Write out the addition and multiplication tables for $\mathbb{F}_6$. Is $\mathbb{F}_6$ is a field? If not, why not?

4.3 The Field of Complex Numbers

We will now introduce the field $\mathbb{C}$ of complex numbers. The complex numbers are astonishingly rich, and an incredible amount of mathematics depends on them. From our standpoint, the most notable fact about the complex numbers is that they form an *algebraically closed field* in the sense that every polynomial function

$$f(x) = x^n + a_1 x^{n-1} + \cdots + a_{n-1}x + a_n$$

with complex coefficients has a root in $\mathbb{C}$. (See Example 4.12 for a complete definition of the notion of a polynomial.) That is, there exists an $\alpha \in \mathbb{C}$ such that $f(\alpha) = 0$. This statement, which is due to C. F. Gauss, is called the Fundamental Theorem of Algebra.

4.3.1 The Construction

The complex numbers arise from the problem that if a is a positive real number, then $x^2 + a = 0$ apparently doesn't have any roots. In order to give it roots, we have to make sense of an expression such as $\sqrt{-a}$. The solution turns turns out to be extremely natural. The real xy-plane $\mathbb{R}^2$ with its usual component-wise addition also has a multiplication such that certain points (namely points on the y-axis), when squared, give points on the negative x-axis. If we interpret the points on the x-axis as real numbers, this solves our problem. It also turns out that under this multiplication on $\mathbb{R}^2$, every nonzero pair $(a, b)^T$ has a multiplicative inverse. The upshot is that we obtain the field $\mathbb{C}$ of complex numbers. The marvelous and deep consequence of this definition is that $\mathbb{C}$ contains not only numbers such as $\sqrt{-a}$, it contains the roots of all polynomial equations with real coefficients.

Let us now give the details. The definition of multiplication on $\mathbb{R}^2$ is easy to state and has a natural geometric meaning discussed below. First of all, we will call the x-axis the *real axis*, and identify a point of the form $(a, 0)^T$ with the real number a. That is, $(a, 0)^T = a$. Hence multiplication on $\mathbb{R}$ can be reformulated as $ab = (a, 0)^T \cdot (b, 0)^T = (ab, 0)^T$. We extend this multiplication to all of $\mathbb{R}^2$ by putting

$$(a, b)^T \cdot (c, d)^T = (ac - bd, ad + bc)^T. \tag{4.7}$$

(Note: do not confuse this with the inner product on $\mathbb{R}^2$.)

We now make the following definition.

Definition 4.9. Define $\mathbb{C}$ to be $\mathbb{R}^2$ with the usual component-wise addition (vector addition) and with the multiplication defined by (4.7).

Addition and multiplication are clearly binary operations. Notice that $(0,a)^T \cdot (0,a)^T = (-a^2,0)^T$, so that $(0,a)^T$ is a square root of $-a^2$. It is customary to denote $(0,1)^T$ by i so

$$i = \sqrt{-1}.$$

Since any point of $\mathbb{R}^2$ can be uniquely represented

$$(a,b)^T = a(1,0)^T + b(0,1)^T, \tag{4.8}$$

we can therefore write

$$(a,b)^T = a + ib.$$

In other words, by identifying the real number a with the vector $a(1,0)^T$ on the real axis, we can express any element of $\mathbb{C}$ as a sum of a real number, its *real part*, and a multiple of i, its *imaginary part*. Thus multiplication takes the form

$$(a+ib)(c+id) = (ac-bd) + i(ad+bc).$$

The Fundamental Theorem of Algebra is stated as follows:

Theorem 4.18. *A polynomial equation*

$$f(z) = z^n + a_{n-1}z^{n-1} + \cdots + a_1 z + a_0 = 0$$

with complex (but possibly real) coefficients has n complex roots.

There are many proofs of this theorem, but none of them are elementary enough to repeat here. Every known proof draws on some deep result from another field, such as complex analysis or topology.

An easy but important consequence is that given any polynomial $p(z)$ with complex coefficients, there exist $r_1, \ldots, r_n \in \mathbb{C}$ which are not necessarily all distinct such that

$$p(z) = (z-r_1)(z-r_2)\ldots(z-r_n).$$

We now prove

Theorem 4.19. *$\mathbb{C}$ is a field containing $\mathbb{R}$ as a subfield.*

Proof. The verification of this theorem is simply a computation. The real number 1 is the identity for multiplication in $\mathbb{C}$, and $0 = (0,0)^T$ is the identity for addition. If $a + ib \neq 0$, then $a + ib$ has a multiplicative inverse, namely

$$(a+ib)^{-1} = \frac{a-ib}{a^2+b^2}. \tag{4.9}$$

The other properties of a field follow easily from the fact that $\mathbb{R}$ is a field. $\square$

4.3.2 The Geometry of $\mathbb{C}$

We now make some more definitions which lead to some beautiful geometric properties of $\mathbb{C}$. First of all, the *conjugate* $\overline{z}$ of $z = a + ib$ is defined by $\overline{z} = a - ib$. It is easy to check the following identities:

$$\overline{w + z} = \overline{w} + \overline{z} \quad \text{and} \tag{4.10}$$

$$\overline{wz} = \overline{w}\ \overline{z}. \tag{4.11}$$

The real numbers are obviously the numbers which are equal to their conjugates. Complex conjugation is the transformation from $\mathbb{R}^2$ to itself which sends a point to its reflection through the real axis.

Formula (4.9) for $(a + ib)^{-1}$ above can now be expressed in a new way. Let $z = a + ib \neq 0$. Since $z\overline{z} = a^2 + b^2$, we get

$$z^{-1} = \frac{\overline{z}}{a^2 + b^2}.$$

Notice that the denominator of the above formula is the square of the length of z. The length of a complex number $z = a + ib$ is called its *modulus* and is denoted by $|z|$. Thus

$$|z| = (z\overline{z})^{1/2} = (a^2 + b^2)^{1/2},$$

and

$$z^{-1} = \frac{\overline{z}}{|z|^2}.$$

Since $\overline{wz} = \overline{w}\ \overline{z}$, the modulus of a product is the product of the moduli:

$$|wz| = |w||z|. \tag{4.12}$$

In particular, the product of two unit length complex numbers also has length one. Now the complex numbers of unit length are just those on the unit circle C=$\{x^2 + y^2 = 1\}$. Every point of C can be represented in the form $(\cos\theta, \sin\theta)$ for a unique angle θ such that $0 \leq \theta < 2\pi$. It is convenient to use a complex valued function of $\theta \in \mathbb{R}$ to express this. We define the *complex exponential* to be the function

$$e^{i\theta} := \cos\theta + i\sin\theta. \tag{4.13}$$

The following proposition is geometrically clear.

Proposition 4.20. *Any $z \in \mathbb{C}$ can be represented as $z = |z|e^{i\theta}$ for some $\theta \in \mathbb{R}$. θ is unique up to a multiple of 2π.*

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

$$e^{i(\theta+\mu)} = e^{i\theta}e^{i\mu}, \tag{4.14}$$

which follows from the standard trigonometric formulas for the sine and cosine of the sum of two angles. (We will give a simple geometric proof of this when we study rotations in the plane.) This gives complex multiplication a geometric interpretation. Writing $w = |w|e^{i\mu}$, we see that

$$wz = (|w|e^{i\mu})(|z|e^{i\theta}) = (|w||z|)e^{i(\mu+\theta)}.$$

In other words, the product wz is obtained by multiplying the lengths of w and z and adding their arguments. (This gives another verification that $|wz| = |w||z|$.)

Exercises

Exercise 4.22. Express all solutions of the equation $z^3 + 1 = 0$ in the form $e^{i\theta}$ and interpret them as complex numbers.

Exercise 4.23. Factor $z^4 - 1$ into the form $(z - \alpha_1)(z - \alpha_2)(z - \alpha_3)(z - \alpha_4)$.

Exercise 4.24. Find all solutions of the linear system

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

Exercise 4.25. Suppose $p(x) \in \mathbb{R}[x]$. Show that the roots of $p(x) = 0$ occur in conjugate pairs, that is $\lambda, \mu \in \mathbb{C}$ where $\overline{\lambda} = \mu$.

Exercise 4.26. Let a, b, c, d be arbitrary integers. Show that there exist integers m and n such that $(a^2 + b^2)(c^2 + d^2) = m^2 + n^2$.

4.4 Vector Spaces

Whenever term "space" is used in a mathematical context, it refers to a vector space, viz. real or complex n-space, the space of continuous functions on the line, the space of self adjoiont linear operators and so on. The purpose of this section is to define the notion of a vector space and to give some examples.

4.4.1 The Notion of a Vector Space

There many situations in which one deals with sets whose elements can be added and multiplied by scalars, in a way that is analogous with vector addition and scalar multiplication in $\mathbb{R}^n$. For example, consider the set of all real valued functions whose domain is a closed interval $[a, b]$ in $\mathbb{R}$, which we will denote by $\mathbb{R}^{[a,b]}$. Addition and scalar multiplication of functions is usually defined pointwise. That is, if f and g are elements of $\mathbb{R}^{[a,b]}$, then $f + g$ is defined at $x \in [a, b]$ by putting

$$(f + g)(x) = f(x) + g(x).$$

Likewise, if r is any real number, then $rf \in \mathbb{R}^{[a,b]}$ takes the value

$$(rf)(x) = rf(x)$$

at $x \in [a, b]$. The key point is that we have defined sums and scalar multiples such that $\mathbb{R}^{[a,b]}$ is closed under these two operations in the sense introduced in Section 4.1. and all scalar multiples of a single $f \in \mathbb{R}^{[a,b]}$ are also elements of $\mathbb{R}^{[a,b]}$. When a set S admits an addition (resp. scalar multiplication) with this property, we will say that S is *closed* under addition (resp. scalar multiplication).

A more refined example is the set $C[a, b]$ of all continuous real valued functions on $[a, b]$. Since $C[a, b] \subset \mathbb{R}^{[a,b]}$, and the definitions of addition and scalar multiplication already been given for $\mathbb{R}^{[a,b]}$, we can just adopt the addition and scalar multiplication we already have in $\mathbb{R}^{[a,b]}$. There is something to worry about, however. We need to chack that $C[a, b]$ is closed under the addition and scalar multiplication from $\mathbb{R}^{[a,b]}$. But this is guaranteed by a basic theorem from calculus: the pointwise 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 any real scalar r.

We now give the definition of a vector space over a field $\mathbb{F}$. It will be clear that, under the definitions of addition and scalar multiplication given above, $\mathbb{R}^{[a,b]}$ is a vector space over $\mathbb{R}$.

Definition 4.10. Let $\mathbb{F}$ be a field and V a set. Assume that there is a binary operation on V called addition which assigns to each pair of elements $\mathbf{a}$ and $\mathbf{b}$ of V a unique sum $\mathbf{a}+\mathbf{b} \in V$. Assume also that there is a second operation, called scalar multiplication, which assigns to any $r \in \mathbb{F}$ and any $\mathbf{a} \in V$ a unique scalar multiple $r\mathbf{a} \in V$. Suppose that addition and scalar multiplication together satisfy the following axioms.

(1) Vector addition is commutative. That is, $\mathbf{a}+\mathbf{b} = \mathbf{b}+\mathbf{a}$ for all $\mathbf{a}, \mathbf{b} \in V$.

(2) Vector addition is also associative. That is, $(\mathbf{a}+\mathbf{b})+\mathbf{c} = \mathbf{a}+(\mathbf{b}+\mathbf{c})$ for all $\mathbf{a}, \mathbf{b}, \mathbf{c} \in V$.

(3) There is an additive identity $\mathbf{0} \in V$ so that $\mathbf{0}+\mathbf{a} = \mathbf{a}$ for all $\mathbf{a} \in V$.

(4) For all $\mathbf{a} \in V$, $1\mathbf{a} = \mathbf{a}$, where 1 is the multiplicative identity of $\mathbb{F}$.

(5) 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 $\mathbf{v}$.

(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 $r(\mathbf{a}+\mathbf{b}) = r\mathbf{a}+r\mathbf{b}$, and $(r+s)\mathbf{a} = r\mathbf{a}+s\mathbf{a}$.

Then V is called a *vector space over* $\mathbb{F}$.

You will eventually come to realize that all of the above conditions are needed. Just as for fields, the additive identity $\mathbf{0}$ is unique, and additive inverses are unique: each vector has exactly one negative. We will call $\mathbf{0}$ the *zero vector*.

Proposition 4.21. *In a vector space, there can only be 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 additive an 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. □

The astute reader will have noticed that this proof is a carbon copy of the proof of Proposition 4.2.

Proposition 4.22. *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 $\mathbf{v}$. Hence, by the uniqueness of the additive inverse (see Proposition 4.21), $(-1)\mathbf{v} = -\mathbf{v}$ for all $\mathbf{v} \in V$. □

If $\mathbf{v}_1, \ldots, \mathbf{v}_k \in V$, then we 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 we verified for sums in a field, the terms in this sum can be associated in any convenient way, since addition is associative. Similarly, the terms $\mathbf{v}_i$ can be taken in any order without changing the sum, since addition is commutative. 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 $\mathbf{v}_1, \ldots, \mathbf{v}_k \in V$.

4.4.2 Examples

Example 4.7. The basic example of a vector space is $\mathbb{R}^n$, the set of all (ordered) n-tuples of real numbers. By an n-tuple, we mean a sequence consisting of n real numbers. Our n-tuples, will usually be denoted by bold faced lower case letters and written as columns such as

$$\mathbf{r} = \begin{pmatrix} r_1 \\ r_2 \\ \vdots \\ r_n \end{pmatrix}.$$

The entries $r_1, \ldots, r_n$ of an n-tuple $\mathbf{r}$ are called its *components*, r_i being the ith *component*. It's important to recall that the order of the components matters: e.g.

$$\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \neq \begin{pmatrix} 2 \\ 3 \\ 1 \end{pmatrix}.$$

Addition and scalar multiplication are carried out component-wise:

$$\mathbf{a} + \mathbf{b} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} + \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} = \begin{pmatrix} a_1 + b_1 \\ a_2 + b_2 \\ \vdots \\ a_n + b_n \end{pmatrix}.$$

and

$$r\mathbf{a} = r \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} ra_1 \\ ra_2 \\ \vdots \\ ra_n \end{pmatrix}.$$

Example 4.8. Recall that to save space, we will frequently use the transpose operator to express a column vector in row vector form. We can imitate the construction of $\mathbb{R}^n$ for any field $\mathbb{F}$ and $n \geq 1$. Let $\mathbb{F}^n$ denote the set of all n-tuples $(a_1, a_2, \ldots, a_n)^T$ of elements of $\mathbb{F}$. Defining addition and scalar multiplication component-wise, we have

$$r(a_1, a_2, \ldots, a_n)^T = (ra_1, ra_2, \ldots, ra_n)^T$$

for all $r \in \mathbb{F}$, and

$$\mathbf{a} + \mathbf{b} = (a_1, a_2, \ldots, a_n)^T + (b_1, b_2, \ldots, b_n)^T = (a_1 + b_1, a_2 + b_2, \ldots, a_n + b_n)^T,$$

for all $\mathbf{a}, \mathbf{b} \in \mathbb{F}^n$. Then $\mathbb{F}^n$ is a vector space over $\mathbb{F}$.

Example 4.9. Let $\mathbb{F}^{m \times n}$ denote the $m \times n$ matrices over $\mathbb{F}$. Then $\mathbb{F}^{m \times n}$ is a vector space over $\mathbb{F}$ with component-wise addition and scalar multiplication. Note that we can easily identify $\mathbb{F}^{m \times n}$ vector space with $\mathbb{F}^{mn}$.

Example 4.10. (See Example 2.2.) If $\mathbb{F} = \mathbb{F}_2$, the the elements of $\mathbb{F}^n$ are called *n-bit strings* and written as binary words. For example, if $n = 4$, we have 4-bit strings such as 0000, 1000, 0100, 1100 and so forth. Since there are 4 places to put either a 0 or a 1, there are exactly $2^4 = 16$ 4-bit strings. Binary strings have an interesting property: each string is its own

additive inverse. Also, the string 1111 consisting of 1's changes the parity of each component. For example, $0101 + 1111 = 1010$. n-bit strings are the fundamental objects of coding theory.

Example 4.11. This example generalizes $\mathbb{R}^{[a,b]}$. Let S be any set and define $\mathbb{R}^S$ to be the set of all real valued functions whose domain is S. We define addition and scalar multiplication pointwise, exactly as for $\mathbb{R}^{[a,b]}$. Then R^S is a vector space over $\mathbb{R}$. Notice that $\mathbb{R}^n$ is nothing but $\mathbb{R}^S$, where $S = \{1, 2, \ldots, n\}$. Indeed, specifying the n-tuple $\mathbf{a} = (a_1, a_2, \ldots a_n)^T \in \mathbb{R}^n$ is the same as defining the function $f_{\mathbf{a}} : S \to \mathbb{R}$ where $f_{\mathbf{a}}(i) = a_i$.

Example 4.12 (Polynomials over a field)**.** Let $\mathbb{F}$ be a field and suppose x denotes a variable. We will assume it makes sense to talk about the powers x^i, where i is any positive integer, and that if $i \neq j$, then $x^i \neq x^j$. Note that $x^1 = x$, and $x^0 = 1$. Define a *polynomial* over $\mathbb{F}$ to be an expression

$$p(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0,$$

where n is an arbitrary non-negative integer, and the coeffficents $a_i \in \mathbb{F}$. Let $\mathbb{F}[x]$ to be the set of all polynomials over $\mathbb{F}$. If $a_n \neq 0$, we say that f has *degree* n. Here we impose that $1x^i = x^i$ for every $i \geq 0$ and $a(bx^i) = (ab)x^i$ for al $a, b \in \mathbb{F}$. We also agree that two polynomials $p(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0$ and $q(x) = b_k x^k + b_{k-1} x^{k-1} + \cdots + b_1 x + b_0$ are equal if and only if $a_i = b_i$ for each index i.

The addition of polynomials is defined by adding the coefficients of the corresponding x^i. In particular, $ax^i + bx^i = (a + b)x^i$ for all $a, b \in \mathbb{F}$. Scalar multiplication is defined from the rule $a(bx^i) = (ab)x^i$. Then, with these operations, $\mathbb{F}[x]$ is a vector space over $\mathbb{F}$.

We may also multiply two polynomials in the natural way by putting $(ax^i)(bx^j) = (ab)x^{i+j}$ and assuming distributivity. Along with addition, this makes $\mathbb{F}[x]$ into a ring.

Example 4.13. When the field $\mathbb{F}$ is $\mathbb{Q}$, $\mathbb{R}$ or $\mathbb{C}$, we can interpret polynomials as functions. For example, the set $\mathcal{P}(\mathbb{R})$ of all real valued polynomial functions with domain $\mathbb{R}$ consists of the functions

$$p(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0,$$

where $a_0, a_1, \ldots, a_n \in \mathbb{R}$ and $n \geq 0$. Here, x denotes the function on $\mathbb{R}$ defined by $x(t) = t$ for all $t \in \mathbb{R}$. $\mathcal{P}(\mathbb{R})$ is a real vector space under pointwise addition and scalar multiplication. Pointwise addition of two polynomials $p(x)$ and $q(x)$ amounts to adding the coefficients of x^i in each polynomial

for every i. Notice that the space $\mathcal{P}_n(\mathbb{R})$ consisting of all polynomials of degree at most n is indistinguishable from $\mathbb{R}^{n+1}$ as a vector space over $\mathbb{R}$.

Example 4.14. Consider the differential equation

$$y'' + ay' + by = 0, \tag{4.15}$$

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 $\mathbb{R}$ which satisfy (4.15) is a real vector space.

Exercises

Exercise 4.27. 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 4.28. Let S be a set, and define $\mathbb{F}^S$ to be the set of all functions $f : S \to F$.

(i) Show how to make $\mathbb{F}^S$ into a vector space over $\mathbb{F}$.

(ii) Let n be a positive integer. If $S = \{1, 2, \ldots, n\}$, show how to identify $\mathbb{F}^S$ and $\mathbb{F}^n$.

4.5 Inner Product Spaces

The purpose of this section will be to introduce an fundamental class of real and complex vector spaces known as inner product spaces . There will be many applications of this concept in later chapters.

4.5.1 The Real Case

The notion of a real inner product space is modeled on Euclidean n-space $\mathbb{R}^n$ (Example 4.15) with its usual dot product $\mathbf{x} \cdot \mathbf{y} = \mathbf{x}^T\mathbf{y}$.

Definition 4.11. A real vector space V is called an *inner product space* if to every pair of elements $\mathbf{a}, \mathbf{b} \in V$, there is a scalar $(\mathbf{a}, \mathbf{b}) \in \mathbb{R}$ satisfying the following properties for all $\mathbf{a}, \mathbf{b}, \mathbf{c} \in V$ and $r \in \mathbb{R}$:

(1) $(\mathbf{a}, \mathbf{b}) = (\mathbf{b}, \mathbf{a})$,

(2) $(\mathbf{a} + \mathbf{b}, \mathbf{c}) = (\mathbf{a}, \mathbf{c}) + (\mathbf{b}, \mathbf{c})$ for all $\mathbf{c} \in V$,

(3) $(r\mathbf{a}, \mathbf{b}) = r(\mathbf{a}, \mathbf{b})$ for all $r \in \mathbb{R}$, and

(4) if $\mathbf{a} \neq \mathbf{0}$, then $(\mathbf{a}, \mathbf{a}) > 0$.

The *length* $|\mathbf{a}|$ of $\mathbf{a} \in V$ is defined by

$$|\mathbf{a}| = \sqrt{(\mathbf{a}, \mathbf{a})}, \tag{4.16}$$

and the *distance* between $\mathbf{a}$ and $\mathbf{b}$ is defined by

$$d(\mathbf{a}, \mathbf{b}) = |\mathbf{a} - \mathbf{b}|. \tag{4.17}$$

Property (4) says that an inner product is *positive definite*. Notice the identity $|r\mathbf{a}| = |r||\mathbf{a}|$. This in fact tells us that the length of the zero vector $\mathbf{0}$ is 0: $|\mathbf{0}| = 0$.

Example 4.15 (Euclidean n-space)**.** By Euclidean n-space, we mean $\mathbb{R}^n$ with the Euclidean inner product

$$\mathbf{a} \cdot \mathbf{b} = \sum_{i=1}^{n} a_i b_i. \tag{4.18}$$

Note that for $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$, we will use the notation $\mathbf{a} \cdot \mathbf{b}$ instead of $(\mathbf{a}, \mathbf{b})$. The inner product $\mathbf{a} \cdot \mathbf{b}$ can also be written as the matrix product $\mathbf{a}^T\mathbf{b}$ since

$$\mathbf{a}^T\mathbf{b} = \begin{pmatrix} a_1 & a_2 & \cdots & a_n \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} = \sum_{i=1}^{n} a_i b_i.$$

4.5.2 Orthogonality

One of the most important properties of an inner product space is that there is a well defined notion of orthogonality.

Definition 4.12. Two vectors $\mathbf{a}$ and $\mathbf{b}$ in an inner product space V are said to be *orthogonal* if $(\mathbf{a}, \mathbf{b}) = 0$.

Two orthogonal vectors are also said to be perpendicular. The zero vector is orthogonal to every vector and, by property (4) of the inner product, $\mathbf{0}$ is in fact the only vector orthogonal to itself. Since two vectors $\mathbf{a}, \mathbf{b} \in \mathbb{R}^2$ are orthogonal if and only if and only if $a_1b_1 + a_2b_2 = 0$, it follows that if $a_1b_2 \neq 0$, then $\mathbf{a}$ and $\mathbf{b}$ are orthogonal if and only if $a_2/a_1 = -b_1/b_2$. Thus, the slopes of such orthogonal vectors in $\mathbb{R}^2$ are negative reciprocals.

Now let V be an inner product space with inner product $(\ ,\ )$.

Proposition 4.23 (Pythagoras's Theorem)**.** *Two vectors* $\mathbf{a}$ *and* $\mathbf{b}$ *in an inner product space* V *are orthogonal if and only if*

$$|\mathbf{a} + \mathbf{b}|^2 = |\mathbf{a} - \mathbf{b}|^2 = |\mathbf{a}|^2 + |\mathbf{b}|^2.$$

Proof. Expanding $|\mathbf{a} + \mathbf{b}|^2 = (\mathbf{a} + \mathbf{b}, \mathbf{a} + \mathbf{b})$, we get

$$|\mathbf{a} + \mathbf{b}|^2 = (\mathbf{a}, \mathbf{a}) + 2(\mathbf{a}, \mathbf{b}) + (\mathbf{b}, \mathbf{b}) = |\mathbf{a}|^2 + 2(\mathbf{a}, \mathbf{b}) + |\mathbf{b}|^2.$$

Hence $(\mathbf{a}, \mathbf{b}) = 0$ if and only if $|\mathbf{a} + \mathbf{b}|^2 = |\mathbf{a}|^2 + |\mathbf{b}|^2$. The equation for $\mathbf{a} - \mathbf{b}$ also follows immediately from this. □

If $\mathbf{a}$ and $\mathbf{b}$ are elements of V and $\mathbf{b} \neq \mathbf{0}$, then there exists a unique $\lambda \in \mathbb{R}$ such that $\mathbf{a} = \lambda\mathbf{b} + \mathbf{c}$, and $(\mathbf{b}, \mathbf{c}) = 0$. This is checked directly by computing $(\mathbf{b}, \mathbf{c})$, where $\mathbf{c} = \mathbf{a} - \lambda\mathbf{b}$. In fact, $(\mathbf{b}, \mathbf{c}) = 0$ if and only if $\lambda = (\mathbf{a}, \mathbf{b})/(\mathbf{b}, \mathbf{b})$. Applying Pythagoras, we see that

$$|\mathbf{a}|^2 = \lambda^2|\mathbf{b}|^2 + |\mathbf{c}|^2.$$

What results from this is the famous

Proposition 4.24 (Cauchy-Schwartz Inequality)**.** *For any* $\mathbf{a}, \mathbf{b} \in V$,

$$|(\mathbf{a}, \mathbf{b})| \leq |\mathbf{a}||\mathbf{b}| \tag{4.19}$$

with equality if and only if $\mathbf{a}$ *is a multiple of* $\mathbf{b}$.

Proof. The result is certainly true if $\mathbf{b} = \mathbf{0}$. Hence we can assume $\mathbf{b} \neq \mathbf{0}$. Since $|\mathbf{c}| \geq 0$, $|\mathbf{a}|^2 \geq \lambda^2 |\mathbf{b}|^2$. Using the definition of λ, we see that $(\mathbf{a}, \mathbf{b})^2 \leq |\mathbf{a}|^2 |\mathbf{b}|^2$, so taking square roots gives us the inequality. Equality holds if and only if $\mathbf{c} = \mathbf{0}$, which holds if and only if $\mathbf{a}$ and $\mathbf{b}$ are collinear. □

The Cauchy-Schwartz 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 second important conclusion from the above discussion and the proof of the Cauchy-Schwartz Inequality is

Proposition 4.25. *Let* $\mathbf{a}$ *and* $\mathbf{b}$ *be elements of an inner product space* V*, and suppose* $\mathbf{b} \neq \mathbf{0}$*. Then we can uniquely decompose* $\mathbf{a}$ *as the sum of two orthogonal vectors*

$$\mathbf{a} = \lambda \mathbf{b} + \mathbf{c}, \tag{4.20}$$

where $(\mathbf{b}, \mathbf{c}) = 0$*. The unique scalar* λ *is* $(\mathbf{a}, \mathbf{b})/(\mathbf{b}, \mathbf{b})$.

The vector $\big((\mathbf{a}, \mathbf{b})/(\mathbf{b}, \mathbf{b})\big)\mathbf{b}$ is called the *projection of* $\mathbf{a}$ *on* $\mathbf{b}$. Finally, we define the angle between nonzero vectors.

Definition 4.13. The *angle* between two nonzero vectors $\mathbf{a}$ and $\mathbf{b}$ in an inner product space V is defined to be unique angle $\theta \in [0, \pi]$ such that

$$\cos\theta = \frac{(\mathbf{a}, \mathbf{b})}{|\mathbf{a}||\mathbf{b}|}. \tag{4.21}$$

Thus we obtain an expression for the inner product which is often taken as a definition. Namely,

$$(\mathbf{a}, \mathbf{b}) = |\mathbf{a}||\mathbf{b}| \cos\theta.$$

We now consider a more esoteric example.

Example 4.16. The vector space $C[a, b]$ of continuous real valued functions on the interval $[a, b]$ which was defined above also carries an inner product,

which enables us (at least partially) to extend our intuition about $\mathbb{R}^n$ to $C[a,b]$. The inner product (f,g) of $f,g \in C[a,b]$ is defined by putting

$$(f,g) = \int_a^b f(t)g(t)\mathrm{d}t.$$

The first three axioms for the Euclidean inner product on $\mathbb{R}^n$ are verified by applying standard facts about integration proved (or at least stated) in any calculus book. Recall that the last axiom requires that $(f,f) \geq 0$ and $(f,f) = 0$ only if $f = 0$. The verification of this requires recalling how the Riemann integral is defined. We leave it as an exercise in elementary real analysis. Thus the length $||f||$ of an $f \in C[a,b]$ is defined to be

$$||f|| := (f,f)^{1/2} = \big(\int_a^b f(t)^2 \mathrm{d}t \big)^{1/2},$$

and the distance between $f,g \in C[a,b]$ is defined to be

$$d(f,g) = ||f-g|| = \big(\int_a^b (f(t)-g(t))^2 \mathrm{d}t \big)^{1/2}.$$

The Cauchy-Schwartz Inequality for $C[a,b]$ says that for any $f,g \in C[a,b]$, we have

$$|\int_a^b f(t)g(t)\mathrm{d}t| \leq \big(\int_a^b f(t)^2\mathrm{d}t\big)^{1/2}\big(\int_a^b g(t)^2\mathrm{d}t\big)^{1/2},$$

with equality if and only if one of the functions is a constant multiple of the other.

Two functions $f,g \in C[a,b]$ are *orthogonal* if and only if $\int_a^b f(t)g(t)\mathrm{d}t = 0$. For example, since $\int_0^{2\pi} \cos t \sin t \mathrm{d}t = 0$, $\cos t$ and $\sin t$ are orthogonal on $[0,2\pi]$. Although the notion of orthogonality for $C[a,b]$ doesn't have any obvious geometric meaning, it nevertheless enables us to extend our intuitive concept of orthogonality into a new situation. In fact, this extension turns out to be extremely important since it leads to the idea of expanding a function in terms of possibly infinitely many mutually orthogonal functions. These infinite series expansions are called Fourier series.

Example 4.17. Suppose $[a,b] = [-1,1]$. Then the functions 1 and x are orthogonal. In fact, x^k and x^m are orthogonal if k is even and m is odd, or vice versa. Indeed,

$$(x^k, x^m) = \int_{-1}^1 x^k \cdot x^m \mathrm{d}x = \int_{-1}^1 x^{k+m}\mathrm{d}x = 0,$$

since $k+m$ is odd. On the other hand, the projection of x^2 on the constant function 1 is $r1$, where $r = \frac{1}{2}\int_{-1}^{1} 1 \cdot x^2 \mathrm{d}x = \frac{1}{3}$. Thus, $x^2 - 1/3$ is orthogonal to the constant function 1 on $[-1,1]$, and $x^2 = (x^2 - 1/3) + 1/3$ is the orthogonal decomposition of x^2 on $[-1,1]$.

4.5.3 Hermitian Inner Products

When V is a complex vector space, the notion of a inner product needs to be changed somewhat. The main example is what is known as the standard Hermitianinner product on $\mathbb{C}^n$. After considering this case, we will give the general definition.

Example 4.18 (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

$$\mathbf{w} \bullet \mathbf{z} = \overline{\mathbf{w}}^T \mathbf{z} = \begin{pmatrix} \overline{w_1} & \overline{w_2} & \cdots & \overline{w_n} \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{pmatrix} = \sum_{i=1}^{n} \overline{w_i} z_i. \tag{4.22}$$

The vector $\overline{\mathbf{w}}^T$ is the frequently called *Hermitian transpose* of $\mathbf{w}$. We'll denote it by $\mathbf{w}^H$ for short. Thus,

$$\mathbf{w} \bullet \mathbf{z} = \mathbf{w}^H \mathbf{z}.$$

Notice that $\mathbf{w} \bullet \mathbf{z}$ is not necessarily real, although $\mathbf{w} \bullet \mathbf{w}$ is in fact a nonnegative real number. In fact, $\mathbf{w}^H \mathbf{w} > 0$ provided $\mathbf{w} \neq \mathbf{0}$. We define the length $|\mathbf{w}|$ of $\mathbf{w}$ 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 the general definition.

Definition 4.14. 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 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*) $(\alpha \mathbf{w}, \mathbf{z}) = \alpha(\mathbf{w}, \mathbf{z})$ and $(\mathbf{w}, \beta \mathbf{z}) = \overline{\beta}(\mathbf{w}, \mathbf{z})$, and

(*v*) 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.*

Exercises

Exercise 4.29. A nice application of Cauchy-Schwartz is that 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 4.30. Prove the law of cosines: If a triangle has sides with lengths $a,\ b,\ c$ and θ is the angle between the sides of lengths a and b, then $c^2 = a^2 + b^2 - 2ab\cos\theta$. (Hint: Consider $\mathbf{c} = \mathbf{b} - \mathbf{a}$.)

Exercise 4.31. Orthogonally decompose the vector $(1, 2, 2)^T$ in $\mathbb{R}^3$ as $\mathbf{p}+\mathbf{q}$ where $\mathbf{p}$ is a multiple of $(3, 1, 2)^T$.

Exercise 4.32. Consider the real vector space $V = C[0, 2\pi]$ with the inner product defined in Section 4.5.

(i) Find the length of $\sin^2 t$ in V.

(ii) Compute the inner product $(\cos t, \sin^2 t)$.

(iii) Find the projection of $\sin^2 t$ on each of the functions $1, \cos t$, and $\sin t$ in V.

(iv) Are 1, $\cos t$ and $\sin t$ mutually orthogonal as elements of V?

(v) How would you define the orthogonal projection of $\sin^2 t$ onto the subspace W of V spanned by 1, $\cos t$, and $\sin t$?

(vi) Describe the subspace W of part (v) as an $\mathbb{R}^n$.

Exercise 4.33. 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 4.34. Let $f, g \in C[a, b]$. Give a formula for the scalar t which minimizes

$$||f - tg||^2 = \int_a^b (f(x) - tg(x))^2 dx.$$

Exercise 4.35. Show that the Hermitian inner product on $\mathbb{C}^n$ satisfies all the conditions listed in Definition 4.14.

4.6 Subspaces and Spanning Sets

The purpose of this section is to define and study the fundamental notions of a subspaces and spanning sets.

4.6.1 The Definition of a Subspace

Definition 4.15. Let V be an arbitrary vector space over a field $\mathbb{F}$. A nonempty subset W of V is called a *linear subspace of* V , or simply a *subspace*, provided the following two conditions hold:

(i) $\mathbf{a} + \mathbf{b} \in W$ whenever $\mathbf{a}, \mathbf{b} \in W$, and

(ii) $r\mathbf{a} \in W$ whenever $r \in \mathbb{F}$.

Clearly, every subspace of a vector space contains the zero vector $\mathbf{0}$. In fact, $\{\mathbf{0}\}$ is itself a subspace, which is called the *trivial subspace*. Consider, for example, a linear equation $ax + by + cz = d$, where $a, b, c, d \in \mathbb{F}$. If $d \neq 0$, then its solution set can't be a subspace of $\mathbb{F}^3$ since $x = y = z = 0$ will not be a solution, and hence the solution set won't contain $\mathbf{0} = (0, 0, 0)^T$. On the other hand, as we note below, the solution set of an arbitrary homogeneous system $ax + by + cz = 0$ is a subspace of $\mathbb{F}^3$. (See Example 4.19 below.)

The following Proposition is an immediate consequence of the definition.

Proposition 4.26. *A subspace W of V is a vector space over $\mathbb{F}$ in its own right.*

Proof. It appears we have to check all of the vector space axioms for W. Fortunately this isn't the case, however. We know by assumption that W is a nonempty subset of V which is closed under addition and scalar multiplication. But then W contains $\mathbf{0}$, since $0\mathbf{w} = \mathbf{0}$ for any $\mathbf{w} \in W$, and every element $\mathbf{w}$ of W has its additive inverse $-\mathbf{w}$ in W, since $-\mathbf{w} = (-1)\mathbf{w}$. But the rest of the vector space axioms hold in W since they already hold in V. Therefore, all the vector space axioms hold for W. □

Example 4.19. The solutions $(x, y, z)^T \in \mathbb{R}^3$ of a homogeneous linear equation $ax+by+cz = 0$, with $a, b, c \in \mathbb{R}$ make up the subspace P of $\mathbb{R}^3$ consisting of all vectors orthogonal to $(a, b, c)^T$. The fact that P is a subspace follows from the properties of the inner product: the sum of any two solutions is also a solution, and any scalar multiple of a solution is a solution. More generally, the solution set of any homogeneous linear equation in n variables with coefficients in a field $\mathbb{F}$ is a subspace of $\mathbb{F}^n$.

The subspaces of $\mathbb{R}^2$ are easily described. They are $\{\mathbf{0}\}$, any line through $\mathbf{0}$ and $\mathbb{R}^2$ itself. The subspaces of $\mathbb{R}^3$ are considered in an exercise.

A basic method for constructing subspaces of a given vector space V is to take all linear combinations of a fixed collection of vectors in V.

Proposition 4.27. *Let $\mathbf{v}_1, \ldots, \mathbf{v}_k$ be vectors in V, and let W be the set of all linear combinations of $\mathbf{v}_1, \ldots, \mathbf{v}_k$. Then W is a subspace of V.*

Proof. This follows readily from Proposition 4.26. The sum of two linear combinations of $\mathbf{v}_1, \ldots, \mathbf{v}_k$ is also a linear combination of $\mathbf{v}_1, \ldots, \mathbf{v}_k$, and any scalar multiple of a linear combination of $\mathbf{v}_1, \ldots, \mathbf{v}_k$ is again such a linear combination. □

Definition 4.16. The subspace of V consisting of all linear combinations of $\mathbf{v}_1, \ldots, \mathbf{v}_k$ is called the *span* of $\mathbf{v}_1, \ldots, \mathbf{v}_k$. We denote this subspace by $\operatorname{span}\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$.

We will denote the subspace in V spanned a vector $\mathbf{v} \neq \mathbf{0}$ by $\mathbb{F}\mathbf{v}$. A subspace spanned by two noncollinear vectors $\mathbf{u}$ and $\mathbf{v}$ is called a plane.

Suppose $\mathbb{F}$ is a prime field, say $\mathbb{F}_p$. Then if $\mathbf{v} \in \mathbb{F}^n$, we can ask how many elements does the line $\mathbb{F}\mathbf{v}$ have? If $\mathbf{v} = \mathbf{0}$, the answer is certainly one. Otherwise, recall from Exercise 4.27 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. Therefore $|\mathbb{F}\mathbf{v}| = |\mathbb{F}| = p$.

Proposition 4.27 says that subspaces are closed under taking linear combinations. It also asserts the converse. The set of all linear combinations of a collection of vectors in V is a subspace of V. We will denote the subspace spanned by $\mathbf{v}_1, \ldots, \mathbf{v}_k$ by $\operatorname{span}\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$.

As previously noted, lines L and planes P in $\mathbb{R}^3$ containing $\mathbf{0}$ are subspaces of $\mathbb{R}^3$. A line L is by definition $\operatorname{span}\{\mathbf{a}\}$ for some (in fact, any) nonzero $\mathbf{a} \in L$. Is every plane P the span of a set of vectors? Clearly, yes, since all we need to do to find two vectors that span P is to choose the two fundamental solutions defined in Section 2.4.2.

On the other hand, suppose $\mathbf{a}$ and $\mathbf{b}$ are two non-collinear vectors in $\mathbb{R}^3$. Their *cross product* is a vector $\mathbf{n} = \mathbf{a} \times \mathbf{b}$ which is orthogonal to both $\mathbf{a}$ and $\mathbf{b}$. The cross product is defined as

$$\mathbf{a} \times \mathbf{b} = (a_2b_3 - a_3b_2, -(a_1b_3 - a_3b_1), a_1b_2 - a_2b_1)^T. \tag{4.23}$$

The cross product $\mathbf{a} \times \mathbf{b}$ is orthogonal to any linear combination of $\mathbf{a}$ and $\mathbf{b}$. Thus we obtain a homogeneous equation satisfied by exactly those vectors in $P = \operatorname{span}\{\mathbf{a}, \mathbf{b}\}$. If $\mathbf{n} = (r, s, t)^T$, then an equation is $rx + sy + tz = 0$.

Example 4.20. Let P be the plane spanned by $(1,1,2)^T$ and $(-1,0,1)^T$. Then $(1,1,2)^T \times (-1,0,1)^T = (1,-3,1)^T$ is orthogonal to P, so an equation for P is $x - 3y + z = 0$.

The cross product $\mathbf{a} \times \mathbf{b}$ is defined for any two vectors in $\mathbb{F}^3$ for any field $\mathbb{F}$, and, in fact, this gives a method for obtaining an equation for the plane in $\mathbb{F}^3$ spanned by any two noncollinear vectors $\mathbf{a}$ and $\mathbf{b}$ for any field $\mathbb{F}$. In general, however, one cannot make any statements about orthogonality since the notion of an inner product doesn't exist for vector spaces over an arbitrary field.

Exercises

In the following exercises, $V(n,p)$ denotes $(\mathbb{F}_p)^n$.

Exercise 4.36. Which of the following subsets of $\mathbb{R}^2$ is not a subspace?

(a) The line $x = y$;

(b) The unit circle;

(c) The line $2x + y = 1$;

s(d) The first octant $x, y \geq 0$.

Exercise 4.37. Prove that every line through the origin and plane through the origin in $\mathbb{R}^3$ are subspaces. Use this to list all subspaces of $\mathbb{R}^3$.

Exercise 4.38. Describe all subspaces of $\mathbb{R}^4$.

Exercise 4.39. Show that a subset of $V(n,p)$ which contains the zero vector and is closed under addition is a subspace.

Exercise 4.40. Find all the 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 4.41. Suppose $A \in \mathbb{F}^{m\times n}$. Show that the null space $\mathcal{N}(A)$ is a subspace of $\mathbb{F}^n$.

Exercise 4.42. How many points lie on a line in $V(n,p)$?

Exercise 4.43. How many points lie on a plane in $V(3,2)$? Generalize this to $V(n,p)$?

Exercise 4.44. Let $\mathbb{F} = \mathbb{F}_2$. Find all solutions in $\mathbb{F}^4$ of the equation $w+x+y+z = 0$. Compare the number of solutions with the number of elements $\mathbb{F}^4$ itself has?

Exercise 4.45. Find a spanning set for the plane $3x - y + 2z = 0$ in $\mathbb{R}^3$.

Exercise 4.46. Find a spanning set for the plane $x + y + z = 0$ in $(\mathbb{F}_2)^3 = V(3,2)$.

Exercise 4.47. 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 4.48. Find an equation for the plane in $(\mathbb{F}_2)^3$ through the origin containing both $(1,1,1)^T$ and $(0,1,1)^T$.

Exercise 4.49. Let $\mathbf{a}, \mathbf{b} \in \mathbb{R}^3$. Show that the cross product $\mathbf{a} \times \mathbf{b}$ is orthogonal to any linear combination of $\mathbf{a}$ and $\mathbf{b}$.

Exercise 4.50. Let L be the line obtained by intersecting the two planes through the origin in $\mathbb{R}^3$. Express L as $\text{span}\{\mathbf{a}\} = \mathbb{R}\mathbf{a}$ where $\mathbf{a}$ is a cross product.

Exercise 4.51. Let $\mathbb{F}$ be any 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 $V + W = \{u \in \mathbb{F}^n \mid u = v + w \, \exists v \in V, w \in W\}$. Show that $V + W$ is a subspace of $\mathbb{F}^n$.

Exercise 4.52. Find the total number of subspaces of $(\mathbb{F}_2)^2$.

Exercise 4.53. Repeat Exercise 4.52 for $(\mathbb{F}_2)^3$.

4.7 Summary

The purpose of this chapter was to introduce two fundamental notions: fields and vector spaces. Fields are the number systems where we can add, subtract, multiply and divide in the usual sense. The basic examples were the rationals $\mathbb{Q}$, which form the smallest field containing the integers, the reals (which are hard to define, so we didn't), the prime fields $\mathbb{F}_p$, which are based on modular arithmetic, and the complex numbers $\mathbb{C}$, which has the crucial property that it is algebraically closed. The key property of a vector space is that we can add elements (i.e. vectors) and operate on them by scalars. General vector spaces do not have a multiplication, although some specific examples do.

We also considered a special class of vector spaces over $\mathbb{R}$, namely inner product spaces, and pointed out that the analogue of an inner product for a vector space over $\mathbb{C}$ is a Hermitian inner product space. The basic example of an inner product space is $\mathbb{R}^n$ with its dot product. Inner product spaces enable us to formulate the notion of orthogonality and use projections, which will be studied in much more detail in a later chapter. An important example of an inner product space is $C[a, b]$, the space of continuous real valued functions on $[a, b]$, the inner product being given by integration over $[a, b]$. Many of the results about the most familiar inner product space, namely $\mathbb{R}^n$, extend easily to a general inner product space such as $C[a, b]$. This is one of the most attractive features of the notion.

Finally, we considered the notion of a subspace of a vector space V and introduced the idea of the subspace spanned by a finite subset of V. This is the subspace consisting of all linear combinations of elements of the subset. Another important example of a subspace of $\mathbb{F}^n$ is the set of all solutions of a homogeneous linear system $A\mathbf{x} = \mathbf{0}$, where A is an $m \times n$ matrix over $\mathbb{F}$.

Chapter 5

Finite Dimensional Vector Spaces

Vector spaces which are spanned by a finite number of vectors are said to be *finite dimensional*. The purpose of this chapter is explain the basic theory of finite dimensional vector spaces, including the notions of linear independence, bases and dimension. Indeed, the development of a workable definition for the notion of dimension is one of the most important contributions of linear algebra. We will also describe some ways of constructing vector spaces such as direct sums and quotients.

5.1 The Notion of Dimension

Roughly speaking, the dimension of a vector space should be the largest number of degrees of freedom available in the space. The dimension should also be the minimal number of parameters required to describe the space. The meaning of this is clear for subsets of $\mathbb{R}^n$ if $n = 1, 2$ or 3. For example, the path traced out by a point moving smoothly through $\mathbb{R}^3$ is intuitively one dimensional because it depends on a single parameter. Similarly, a smooth surface is two dimensional. (On the other hand, a definition of the dimension of a non-smooth path or surface can be very tricky to formulate.) In particular, a plane through the origin in $\mathbb{R}^2$ can be described as the set of all linear combinations $x\mathbf{v} + y\mathbf{w}$, where $\mathbf{v}$ and $\mathbf{w}$ are any two non-collinear vectors on the plane and x, y vary over $\mathbb{R}$. The objects we will be treating here are all linear, and, as we will see, their dimensions are defined in a natural and computable way.

If $\mathbb{F}$ is an arbitrary field, say $\mathbb{F}_p$, one doesn't have any physical intuition

to fall back on. Thus, a definition of the dimension of a vector space over $\mathbb{F}$, such as $\mathbb{F}^n$ has to be made in a less intuitive fashion, in such a way that the answer for $\mathbb{F}^n$ should still be n. This is where the efficacy of abstract algebra becomes apparent.

5.1.1 Linear Independence

Let V denote a not necessarily finite dimensional vector space over an arbitrary field $\mathbb{F}$. Before defining the notion of the dimension of V, we must first introduce some preliminary notions, starting with linear independence. Put informally, a set of vectors is *linearly independent* if no one of them can be expressed as a linear combination of the others. In other words. two vectors are linearly independent when they don't lie on the same line through the origin, and three vectors are independent when they don't lie on the same plane through the origin. The formal definition, which we now give, is stated in a slightly different way.

Definition 5.1. Let $\mathbf{w}_1, \ldots, \mathbf{w}_k$ be in V. We say that $\mathbf{w}_1, \ldots, \mathbf{w}_k$ are *linearly independent* (or, simply, *independent*) if and only if the only linear combination

$$a_1\mathbf{w}_1 + a_2\mathbf{w}_2 + \cdots + a_k\mathbf{w}_k = \mathbf{0}, \tag{5.1}$$

with $a_1, a_2, \ldots, a_k \in \mathbb{F}$ is the trivial combination $a_1 = a_2 = \cdots = a_k = 0$. If (5.1) has a solution where some $a_i \neq 0$, we say that $\mathbf{w}_1, \ldots, \mathbf{w}_k$ are *linearly dependent* (or, simply, *dependent*). We will also say that a finite subset S of V is independent if the vectors contained in S are independent.

Notice that we're only defining linear indepence for a finite number of vectors. The reader might want to contemplate how to do this for infinite sets. Notice that any finite set of vectors in V containing $\mathbf{0}$ is dependent (why?).

Let us begin by relating the formal definition to the above discussion.

Proposition 5.1. *A finite 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 $\mathbf{w}_1, \ldots, \mathbf{w}_k$ be the vectors, and suppose one of the vectors, say $\mathbf{w}_1$, is a linear combination of the others. Then

$$\mathbf{w}_1 = a_2\mathbf{w}_2 + \cdots + a_k\mathbf{w}_k.$$

Thus

$$\mathbf{w}_1 - a_2\mathbf{w}_2 - \cdots - a_k\mathbf{w}_k = \mathbf{0},$$

so (5.1) has a solution with $a_1 = 1$. Therefore $\mathbf{w}_1, \ldots, \mathbf{w}_k$ are dependent. Conversely, suppose $\mathbf{w}_1, \ldots, \mathbf{w}_k$ are dependent. This means that there is a solution $a_1, a_2, \ldots, a_k$ of (5.1), where some $a_i \neq 0$. We can assume (by reordering the vectors) that the nonzero coefficient is a_1. We can thus write

$$\mathbf{w}_1 = b_2\mathbf{w}_2 + \cdots + b_k\mathbf{w}_k,$$

where $b_i = -a_i/a_1$, so the proof is done. □

The following Proposition gives one of the most important properties of linearly independent sets.

Proposition 5.2. *Assume that $\mathbf{w}_1, \ldots, \mathbf{w}_k$ are linearly independent vectors in V and suppose $\mathbf{v}$ is in their span. Then $\mathbf{v} = \sum_{i=1}^{k} r_i\mathbf{w}_i$ for exactly one linear combination of $\mathbf{w}_1, \ldots, \mathbf{w}_k$.*

Proof. By assumption, there exists an expression

$$\mathbf{v} = r_1\mathbf{w}_1 + r_2\mathbf{w}_2 + \cdots + r_k\mathbf{w}_k,$$

where $r_1, \ldots, r_k \in \mathbb{F}$. Suppose there is another expression, say

$$\mathbf{v} = s_1\mathbf{w}_1 + s_2\mathbf{w}_2 + \cdots + s_k\mathbf{w}_k$$

where the s_i are also elements of $\mathbb{F}$. By subtracting and doing a bit of algebraic manipulation, we get that

$$\mathbf{0} = \mathbf{v} - \mathbf{v} = (r_1 - s_1)\mathbf{w}_1 + (r_2 - s_2)\mathbf{w}_2 + \cdots + (r_k - s_k)\mathbf{w}_k.$$

Since the $\mathbf{w}_i$ are independent, every coefficient $r_i - s_i = 0$, and this proves the Proposition. □

When $V = \mathbb{F}^n$, the definition of linear independence involves a linear system. Recalling that vectors in $\mathbb{F}^n$ are viewed as column vectors, consider the $n \times m$ matrix

$$A = (\mathbf{w}_1 \ \cdots \ \mathbf{w}_m).$$

By the theory of linear systems, we have

Proposition 5.3. *The vectors $\mathbf{w}_1, \ldots, \mathbf{w}_m$ in $\mathbb{F}^n$ are linearly independent exactly when the system $A\mathbf{x} = \mathbf{0}$ has no nontrivial solution. This is the case exactly when the rank of A is m. In particular, more than n vectors in $\mathbb{F}^n$ are linearly dependent.*

Proof. The first statement follows from the definition of $A\mathbf{x}$, and the second and third follow from Proposition 2.6. □

5.1.2 The Definition of a Basis

As above, let V be a vector space over a field $\mathbb{F}$.

Definition 5.2. A collection of vectors in V which is linearly independent and spans V is called a *basis* of V.

Notice that by our convention, a basis of V is necessarily finite. The definition of linear independence can easily be reformulated so that the notion of an infinite basis makes sense. Our main concern, however, is the case where V is finite dimensional; that is, V is spanned by finitely many vectors.

Let us now consider some examples.

Example 5.1 (The standard basis of $\mathbb{F}^n$)**.** The *standard basis* of $\mathbb{F}^n$ is the set of the columns of the identity matrix I_n. The i-th column of I_n will always be denoted by $\mathbf{e}_i$. There is a slight ambiguity in this notation, because the definition of standard basis depends on n, so it would be better to denote the standard basis vectors in $\mathbb{F}^n$ by $\mathbf{e}_i(n)$ instead of $\mathbf{e}_i$. We'll ignore this point because the context will usually make it clear which $\mathbb{F}^n$ we are considering. 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 is clear that $\mathbf{e}_1, \ldots, \mathbf{e}_n$ do indeed give a basis of $\mathbb{F}^n$.

Example 5.2 (Lines and planes)**.** A nonzero vector in $\mathbb{R}^n$ spans a line, and clearly a nonzero vector is linearly independent. Hence a line through $\mathbf{0}$ has a basis consisting of any nonzero vector on the line. (Thus there isn't a unique basis.) A plane P containing the origin is spanned by any pair of non collinear vectors on P, and any two non collinear vectors on P are linearly independent. In fact, every basis of P consists of two independent vectors on P.

It should noted that the trivial vector space $\{\mathbf{0}\}$ does not have a basis. Indeed, in order to have a basis, $\{\mathbf{0}\}$ has to be independent, which it isn't.

Proposition 5.2 allows us to deduce an elementary but important characterization of a basis.

Proposition 5.4. *The vectors $\mathbf{v}_1, \ldots, \mathbf{v}_r$ in V form a basis of V if and only if every vector $\mathbf{v}$ in V admits a unique expression*

$$\mathbf{v} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_r\mathbf{v}_r,$$

where $a_1, a_2, \ldots a_r$ are elements of $\mathbb{F}$.

Proof. We leave this as an exercise. □

Here is another characterization of a basis.

Proposition 5.5. *A subset of V which is a maximal linearly independent set is a basis.*

Proof. By a maximal linearly independent set, we mean an independent subset $\{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$ of V such that if $\mathbf{w}$ is any element of V, then $\{\mathbf{v}_1, \ldots, \mathbf{v}_r, \mathbf{w}\}$ is dependent. The point is that a maximal linearly independent subset spans V. For if $\mathbf{w} \in V$, then from the fact that $\mathbf{v}_1, \ldots, \mathbf{v}_r, \mathbf{w}$ are dependent, there are scalars $a_1, \ldots, a_r$, and b not all zero such that

$$a_1\mathbf{v}_1 + \cdots + a_r\mathbf{v}_r + b\mathbf{w} = \mathbf{0}.$$

But if $b = 0$, then all the $a_i = 0$ as well (why?), so $b \neq 0$. Hence $\mathbf{w}$ is a linear combination of $\mathbf{v}_1, \ldots, \mathbf{v}_r$, so the proof is done. □

Example 5.3 (Column spaces)**.** Suppose A is an $m \times n$ matrix over $\mathbb{F}$. The *column space* $\mathrm{col}(A)$ of A is the subspace of $\mathbb{F}^m$ spanned by the columns of A. The column space has an important interpretation in terms of linear systems. Let $\mathbf{a}_1, \ldots, \mathbf{a}_n$ be the columns of A. By definition, $\mathbf{b} \in \mathrm{col}(A)$ if and only if there are scalars $r_1, \ldots, r_n$ such that $\mathbf{b} = r_1\mathbf{a}_1 + \cdots + r_n\mathbf{a}_n$. In matrix terms, this means $A\mathbf{r} = \mathbf{b}$. Thus,

Proposition 5.6. *The column space of the $m \times n$ matrix A consists of all $\mathbf{b} \in \mathbb{F}^n$ such that the linear system $A\mathbf{x} = \mathbf{b}$ is consistent. If A has rank n, then its columns are independent and hence, by definition, form a basis of* $\mathrm{col}(A)$.

Proof. The first statement is obvious from the above remarks. If A has rank n, then Proposition 2.6 tells us that the system $A\mathbf{x} = \mathbf{0}$ has a unique solution, which is equivalent to its columns being independent. Hence, by definition form a basis. □

If the rank of A is less than n, the columns are dependent, so we don't have a basis. The natural way to proceed is to try to find a set of columns which are independent and still span $\mathrm{col}(A)$. This is the general problem of extracting a basis from a spanning set, which we will treat below.

Example 5.4 (Basis of the null space)**.** Let A be an $m \times n$ matrix over $\mathbb{F}$. As pointed out in Chapter 2 (for $\mathbb{R}$, but equally valid for any $\mathbb{F}$), the fundamental solutions of $A\mathbf{x} = \mathbf{0}$ span the null space $\mathcal{N}(A)$, which is a subspace of $\mathbb{F}^n$. In fact, they are also independent, so the fundamental solutions comprise a basis of $\mathcal{N}(A)$.

Exercises

Exercise 5.1. Are the vectors $(0,2,1,0)^T$, $(1,0,0,1)^T$ and $(1,0,1,1)^T$ in $\mathbb{R}^4$ are independent? Can they form a basis of $\mathbb{R}^4$?

Exercise 5.2. 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 5.3. Find the fundamental solutions of the system

$$\begin{aligned} 0x_1 + x_2 + 2x_3 + 0x_4 + 3x + 5 - x_6 &= 0 \\ 0x_1 + 0x_2 + 0x_3 + x_4 + 2x_5 + 0x_6 &= 0. \end{aligned}$$

and show they are linearly independent.

Exercise 5.4. 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 fundamental solutions of $A\mathbf{x} = \mathbf{0}$ are a basis of $\mathcal{N}(A)$.

(ii) Repeat (i) when A is considered to be a matrix over $\mathbb{F}_3$.

Exercise 5.5. Prove the assertion made in Example 5.4 that the fundamental solutions are a basis of $\mathcal{N}(A)$.

Exercise 5.6. Show that any subset of a linearly independent set is linearly independent.

Exercise 5.7. Suppose $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k$ are mutually orthogonal unit vectors in $\mathbb{R}^m$. Show $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k$ are independent.

Exercise 5.8. Show that m independent vectors in $\mathbb{F}^m$ are a basis.

Exercise 5.9. Find a basis for the space $\mathbb{R}[x]_n$ of all polynomials with real coefficients having degree at most n.

Exercise 5.10. True or False: Any four vectors in $\mathbb{R}^3$ are dependent. (Supply reasoning.)

Exercise 5.11. 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 5.12. Let $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{w}$ be a basis of $\mathbb{R}^3$.

(a) Determine whether or not $3\mathbf{u}+2\mathbf{v}+\mathbf{w}$, $\mathbf{u}+\mathbf{v}+0\mathbf{w}$, and $-\mathbf{u}+2\mathbf{v}-3\mathbf{w}$ are independent.

(b) Do the vectors in part (a) span $\mathbb{R}^3$? Supply reasoning.

(c) 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 $a_1, a_2, \ldots, c_3$ are arbitrary scalars.

Exercise 5.13. Find a basis for the set of invertible 3×3 matrices over an arbitrary field $\mathbb{F}$. (Hint: thimk.)

5.2 Bases and Dimension

We now at last get down to defining the notion of dimension. Lines and planes have dimension one and two respectively. But we've already noticed that they also have bases with one and two elements respectively. We've also remarked that the dimension of a vector spaceshould be the maximal number of degrees of freedom, which is cleary the maximal number of independent vectors. Thus if the number of vectors in a basis is the maximal number of independent vectors, then the definition of dimension should be the number of vectors in a basis. We will now see that this is the case.

5.2.1 The Definition of Dimension

Let V denote a finite dimensional vector space over a field $\mathbb{F}$. We now state one of the definition of dimension, one of the most fundamental notions in linear algebra.

Definition 5.3. The *dimension* of the finite dimensional vector space V is defined to be the number of elements in a basis of V. For convenience, we will define the dimension of the trivial vector space $\{\mathbf{0}\}$ to be 0, even though $\{\mathbf{0}\}$ doesn't have a basis. 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.

Of course, by Proposition 5.5, the definition amounts to saying that the dimension of V is the maximal number of independent vectors in V. But it actually says more. The definition first asserts that every nontrivial finite dimensional vector space V has a basis. It also asserts that any two bases of V have the same number of elements. These assertions aren't obvious, so we have to prove them before we can use the definition. This will be done in the Dimension Theorem, which is stated and proved below.

Let us point out a subtlety in the definition when $\mathbb{F} = \mathbb{C}$ and $V = \mathbb{C}^n$. Since $\mathbb{C} = \mathbb{R}^2$, $\mathbb{C}^n$ is the same as $\mathbb{R}^{2n}$. Thus, if we ask for the dimension of $\mathbb{C}^n$, the answer could be either n or $2n$ and still be consistent with having the dimension of $\mathbb{F}^n$ be n. What we need to clarify is that $\mathbb{C}^n$ is both a vector space over $\mathbb{C}$ and a vector space over $\mathbb{R}$, and $\dim_{\mathbb{C}} \mathbb{C}^n = n$, while $\dim_{\mathbb{R}} \mathbb{C}^n = \dim_{\mathbb{R}} \mathbb{R}^{2n} = 2n$. Hence when we speak of the dimension of $\mathbb{C}^n$, we need to differentiate between whether we are speaking of the complex dimension (which is n) or the real dimension (which is $2n$).

5.2.2 Examples

Let us consider some examples.

Example 5.5. The dimension of a line is 1 and that of a plane is 2. The dimension of the hyperplane $a_1x_1 + \cdots + a_nx_n = 0$ in $\mathbb{R}^n$ is $n-1$, provided some $a_i \neq 0$, since the $n-1$ fundamental solutions form a basis of the hyperplane. These remarks are also true in $\mathbb{F}^n$ for any $\mathbb{F}$.

Example 5.6. Let $A = (\mathbf{w}_1\ \mathbf{w}_2\ \ldots\ \mathbf{w}_n) \in \mathbb{F}^{n\times n}$ have rank n. Then, by Proposition 2.6, the columns of A are a basis of $\mathbb{F}^n$.

Example 5.7. Recall from Example 4.13 that if n is a positive integer, then $\mathcal{P}_n(\mathbb{R})$ denotes the space of polynomials

$$f(x) = a_nx^n + a_{n-1}x^{n-1} + \cdots + a_1x + a_0$$

on $\mathbb{R}$ of degree at most n. Let us show, for example, that $1, x, x^2, x^3$ form a basis of $\mathcal{P}_3(\mathbb{R})$. To see independence, we have to show that if

$$f(x) = \sum_{i=0}^{3} a_ix^i = 0,$$

then each $a_i = 0$. That is, we have to show that if $f(r) = 0$ for all $r \in \mathbb{R}$, then each $a_i = 0$. Now if $f(r) = 0$ for all $r \in \mathbb{R}$, then

$$f(0) = a_0 = 0,\ f'(0) = a_1 = 0,\ f''(0) = 2a_2 = 0,\ f'''(0) = 3a_3 = 0.$$

Hence we have the asserted linear independence. It is obvious that $1, x, x^2, x^3$ span $\mathcal{P}_3(\mathbb{R})$, so the claim is established.

One can use the same argument to show that $1, x, x^2, \ldots, x^n$ are a basis of $\mathcal{P}_n(\mathbb{R})$ for all $n \geq 0$.

Example 5.8. 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 we are dealing with the vector space $\mathcal{P}_3$ of the last example. The solution space of the equation $y'' + y = 0$ consists of all linear combinations of the functions $\sin x$ and $\cos x$.

5.2.3 The Dimension Theorem

We will now prove a theorem that shows that the definition of dimension makes sense and more. This is one of the main results in linear algebra.

Theorem 5.7 (The Dimension Theorem)**.** *Assume V is a non-trivial finite dimensional vector space over a field $\mathbb{F}$. Then V has a basis. In fact, any spanning set for V contains a basis. Furthermore, any linearly independent subset of V is contained in a basis. Finally, any two bases of V have the same number of elements.*

Proof. We first show every spanning set contains a basis. Let $\mathbf{w}_1, \ldots, \mathbf{w}_k$ span V. Consider the set of all subsets of $\{\mathbf{w}_1, \ldots, \mathbf{w}_k\}$ which also span V, and let $\{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$ be any such subset where r is minimal. There is no problem showing such a subset exists, since $\{\mathbf{w}_1, \ldots, \mathbf{w}_k\}$ has only 2^k subsets. It suffices to show $\mathbf{v}_1, \ldots, \mathbf{v}_r$ are independent, so suppose

$$a_1\mathbf{v}_1 + \cdots + a_r\mathbf{v}_r = \mathbf{0}.$$

If $a_i \neq 0$, 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. This contradicts the minimality of r, and hence $\mathbf{v}_1, \ldots, \mathbf{v}_r$ are independent. Thus, $\mathbf{v}_1, \ldots, \mathbf{v}_r$ form a basis, so every spanning set contains a basis. Since, by assumption, V has a finite spanning set, it has a basis.

We next show that any 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 $a_k = 0$ also since the $\mathbf{w}_k$ are independent. This proves the claim.

Now suppose $W \neq V$. We will use the basis $\mathbf{v}_1, \ldots, \mathbf{v}_r$ obtained above. If each $\mathbf{v}_i \in W$, then $W = V$ and we are done. Otherwise, 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 $W_1 = \text{span}\{\mathbf{w}_1, \ldots, \mathbf{w}_m, \mathbf{v}_i\}$. Clearly we may continue, at each step adding one of the $\mathbf{v}_j$, if necessary, always maintaining an independent subset of V. Eventually we have to

obtain a subspace V' of V containing $\mathbf{v}_1, \ldots, \mathbf{v}_r$. But $\mathbf{v}_1, \ldots, \mathbf{v}_r$ span V, so $V' = V$. Thus $\mathbf{w}_1, \ldots, \mathbf{w}_m$ are contained in a basis of V.

It remains to show that any two bases of V have the same number of elements. This is proved by the following replacement technique. Suppose $\mathbf{u}_1, \ldots, \mathbf{u}_m$ and $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are two bases of V. Without any loss of generality, suppose $m \leq n$. We can certainly write

$$\mathbf{v}_1 = r_1\mathbf{u}_1 + r_2\mathbf{u}_2 \cdots + r_m\mathbf{u}_m.$$

Since $\mathbf{v}_1 \neq \mathbf{0}$, some $r_i \neq 0$, so we may suppose, if necessary by renumbering indices, 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 $\mathbf{v}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m$ 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 $r_1 \neq 0$, this gives two distinct ways of expanding $\mathbf{v}_1$ in terms of the first basis, which contradicts the uniqueness statement in Proposition 5.4. Hence $x_1 = 0$. It follows immediately that all $x_i = 0$ (why?), so $\mathbf{v}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m$ are independent. The proof that $\mathbf{v}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m$ span V is left as an exercise. Hence we have produced a new basis of V where $\mathbf{v}_1$ replaces $\mathbf{u}_1$. Now repeat the argument with the new basis by writing

$$\mathbf{v}_2 = x_1\mathbf{v}_1 + x_2\mathbf{u}_2 + \cdots + \mathbf{u}_m.$$

Clearly, one of the $x_i\mathbf{u}_i$ in this expression is nonzero. Thus, after renumbering, we can assume the coefficient of $\mathbf{u}_2$ is nonzero. Repeating the above argument, it follows 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 $\mathbf{u}_i$'s, 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 $\mathbf{v}_1, \ldots, \mathbf{v}_m$, which contradicts the linear independence of $\mathbf{v}_1, \ldots, \mathbf{v}_n$. We thus conclude $m = n$, and the Dimension Theorem is therefore proven. □

There is a useful consequence of the Dimension Theorem.

Corollary 5.8. *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, any subset of V containing more that $\dim V$ elements is dependent.*

Proof. This is an exercise. □

5.2.4 An Application

Let's begin with a nice application of the Dimension Theorem. Let p be a prime and consider a finite dimensional vector space V over $\mathbb{F} = \mathbb{F}_p$. Then the dimension of V determines the number of elements of V as follows.

Proposition 5.9. *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, which we know is possible. Then 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 Proposition 5.4 tells us that different choices of the a_i give different elements of V, it follows that there are exactly $p \cdot p \cdots p = p^k$ such linear combinations. Thus V contains p^k elements. □

For example, a line in $\mathbb{F}^n$ has p elements, a plane has p^2 and so forth. We can apply the last result to a finite field $\mathbb{F}$. It is clear that the characteristic of $\mathbb{F}$ is positive (since $\mathbb{F}$ is finite), so suppose it is the prime p. By Exercise 4.12, 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 $\mathbb{F}'$. Moreover, since $\mathbb{F}$ itself is finite, it follows that $\mathbb{F}$ is finite dimensional over $\mathbb{F}'$. For every $a \in \mathbb{F}$ has the expression $a = 1a$ which means $\mathbb{F}$ spans itself over $\mathbb{F}'$ since $1 \in \mathbb{F}'$. Applying Proposition 5.9, we get

Proposition 5.10. *Let $\mathbb{F}$ be a finite field of characteristic p. Then $|\mathbb{F}| = p^n$ where n is the dimension of $\mathbb{F}$ over the subfield $\mathbb{F}'$ of $\mathbb{F}$ consisting of all multiples of 1.*

The only example of a field we have seen having p^n elements, where $n > 1$, is the field with 4 elements constructed in Section 4.2.1. It can be shown that there is a field of order p^n for every prime p and integer $n > 0$ and that this field is essentially unique.

5.2.5 Examples

Let's next consider some more examples.

Example 5.9 (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 $\mathbb{F}^{mn}$. The matrix

analogue of the standard basis of $\mathbb{F}^{m\times n}$ is the set of $m \times n$ matrices E_{ij} which have a 1 in the i-th row and j-th column and a zero everywhere else. We leave the proof that they form a basis as an easy exercise. Therefore, $\dim \mathbb{F}^{m\times n} = mn$.

Example 5.10 (Linear Systems)**.** The fundamental identity (2.6) for an $m \times n$ homogeneous linear system $A\mathbf{x} = \mathbf{0}$ can now be expressed in terms of dimension as follows:

$$\dim \mathcal{N}(A) + \operatorname{rank}(A) = n. \tag{5.2}$$

We will interpret the rank of A as a dimension below.

Example 5.11 (The row space)**.** The *row space* $\operatorname{row}(A)$ of a matrix $A \in \mathbb{F}^{m\times n}$ is the span of its rows. The $\operatorname{row}(A)$ is a subspace of $\mathbb{F}^n$, so clearly $\dim \operatorname{row}(A) \leq n$, by Corollary 5.8. We will relate the row space to row operations and find its dimension in the next section.

Example 5.12 (The 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 of $\mathbb{F}^{n\times n}$ we found in Example 5.9 doesn't work for $\mathbb{F}_s^{n\times n}$ since E_{ij} isn't symmetric if $i \neq j$. To repair this problem, put $S_{ij} = E_{ij}+E_{ji}$ when $i \neq j$. Then $S_{ij} \in \mathbb{F}_s^{n\times n}$. 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

$$\dim \mathbb{F}_s^{n\times n} = (n-1) + (n-2) + \cdots + 2 + 1 + n = \sum_{i=1}^{n} i = n(n+1)/2.$$

Example 5.13 (The skew symmetric matrices)**.** The set $\mathbb{F}_{ss}^{n\times n}$ of skew symmetric $n \times n$ matrices over $\mathbb{F}$ are another interesting subspace of $\mathbb{F}^{n\times n}$. A square matrix $A \in \mathbb{F}^{n\times n}$ is called *skew symmetric* if $A^T = -A$. 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 $\operatorname{char}(\mathbb{F}) \neq 2$. 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

$$\begin{pmatrix} 0 & b \\ -b & 0 \end{pmatrix}.$$

Thus $E_{12} - E_{21}$ is a basis. We leave it as an exercise to show $\dim \mathbb{F}_{ss}^{n \times n} = n(n-1)/2$ for all n. We will return to these examples below.

Exercises

Exercise 5.14. Find a basis for the subspace of $\mathbb{R}^4$ spanned by

$$(1,0,-2,1),\ (2,-1,2,1),\ (1,1,1,1),\ (0,1,0,1),\ (0,1,1,0)$$

containing the first and fifth vectors.

Exercise 5.15. Let W be a subspace of V. Show that if $w_1, w_2, \ldots, w_k$ are independent vectors in W and $v \in V$ but $v \notin W$, then $w_1, w_2, \ldots, w_k, v$ are independent.

Exercise 5.16. Suppose V is a finite dimensional vector space over a field $\mathbb{F}$, say $\dim V = n$, and let W be a subspace of V. Prove the following:

(*i*) W is finite dimensional. (Hint: Show that if W is not finite dimensional, then W contains more than n independent vectors.)

(*ii*) In fact, $\dim W \leq \dim V$.

(*iii*) If $\dim W = \dim V$, then $W = V$.

This proves Corollary 5.8.

Exercise 5.17. Consider the subspace W of $(\mathbb{F}_2)^4$ 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 5.18. Let W and X be subspaces of a finite dimensional vector space V of dimension n. What are the minimum and maximum dimensions that $W \cap X$ can have? Discuss the case where W is a hyperplane (i.e. $\dim W = n-1$) and X is a plane (i.e. $\dim X = 2$).

Exercise 5.19. Let $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n$ be mutually orthogonal unit vectors in $\mathbb{R}^n$. Are $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n$ a basis of $\mathbb{R}^n$?

Exercise 5.20. Suppose W is a subspace of $\mathbb{F}^n$ of dimension k. Show the following:

(i) Any k linearly independent vectors in W span W, hence are a basis of W.

(ii) Any k vectors in W that span W are linearly independent, hence are a basis of W.

Exercise 5.21. Show that the functions

$$1, x, x^2, x^3, \ldots, x^n, \ldots$$

are linearly independent on any open interval (a, b).

Exercise 5.22. Is $\mathbb{R}$ a vector space over $\mathbb{Q}$? If so, is $\dim_{\mathbb{Q}} \mathbb{R}$ finite or infinite?

Exercise 5.23. Let $\mathbf{x} \in \mathbb{R}^n$ be any nonzero vector. Let $W \subset \mathbb{R}^{n\times n}$ consist of all matrices A such that $A\mathbf{x} = \mathbf{0}$. Show that W is a subspace and find its dimension.

Exercise 5.24. Consider the set $\mathbb{F}_s^{n\times n}$ of symmetric $n \times n$ matrices over $\mathbb{F}$.

(a) Show that $\mathbb{F}_s^{n\times n}$ is a subspace of $\mathbb{F}^{n\times n}$.

(a) Show that the set of matrices S_{ij} with $i < j$ defined in Example 5.12 together with the E_{ii} make up a basis of $\mathbb{F}_s^{n\times n}$.

Exercise 5.25. Let $\mathbb{F}_{ss}^{n\times n}$ be the $n \times n$ skew symmetric matrices over $\mathbb{F}$.

(a) Show that $\mathbb{F}_{ss}^{n\times n}$ is a subspace of $\mathbb{F}^{n\times n}$.

(b) Find a basis of $\mathbb{F}_{ss}^{n\times n}$ and compute its dimension.

(c) Find a basis of $\mathbb{F}^{n\times n}$ which uses only symmetric and skew symmetric matrices.

Exercise 5.26. Show that the set of $n \times n$ upper triangular real matrices is a subspace of $\mathbb{R}^{n\times n}$. Find a basis and its dimension.

Exercise 5.27. Let V be a vector space over $\mathbb{F}_p$ of dimensionv$n$. 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 with m elements is exactly $(p^n - 1)(p^n - p) \cdots (p^n - p^{m-2})(p^n - p^{m-1})$. (Use Proposition 5.9 and use part of the proof of the Dimension Theorem.)

Exercise 5.28. Use Exercise 5.27 to show that the number of subspaces of dimension m in an n-dimensional vector space V over $\mathbb{F}_p$ is

$$\frac{(p^n - 1)(p^n - p) \cdots (p^n - p^{m-2})(p^n - p^{m-1})}{(p^m - 1)(p^m - p) \cdots (p^n - p^{m-2})(p^m - p^{m-1})}.$$

Note: the set of m-dimensional subspaces of a finite dimensional vector space is an important object called a Grassmannian.

5.3 Some Results on Matrices

We now apply the ideas invoved in the Dimension Theorem to obtain some results about matrices.

5.3.1 A Basis of the Column Space

The Dimension Theorem (Theorem 5.7) guarantees that any spanning set of a finite dimensional vector space contains a basis. In fact, the subsets which give bases are exactly the minimal spanning subsets. However, there is still the question of whether there is an explicit method for actually extracting one of these subsets. We will now answer this for subspaces of $\mathbb{F}^n$. The method is based on row reduction.

Suppose $\mathbf{w}_1, \ldots, \mathbf{w}_k \in \mathbb{F}^n$, and let W be the subspace they span. To construct a subset of these vectors spanning W, consider the $n \times k$ matrix $A = (\mathbf{w}_1 \ \ldots \ \mathbf{w}_k)$. We must find columns of A which are a basis of the column space $W = \text{col}(A)$. An answer is given by

Proposition 5.11. *The columns of A that correspond to a corner entry in A_{red} are a basis of the column space $\text{col}(A)$ of A. In particular, the dimension of $\text{col}(A)$ is the number of corner entries in A_{red}.*

Proof. The key observation is that $A\mathbf{x} = \mathbf{0}$ if and only if $A_{red}\mathbf{x} = \mathbf{0}$ (why?). Any nontrivial solution $\mathbf{x}$ gives an expression of linear dependence among the columns of both A and A_{red}. (For example, if the fifth column of A_{red} is the sum of the first four columns of A_{red}, this also holds for A.) But the columns of A_{red} containing a corner entry are standard basis vectors of $\mathbb{F}^n$, hence independent. Hence the corner columns of A are also linearly independent. On the other hand, the corner columns of A_{red} span column space of A_{red} (but, of course, not of $\text{col}(A)$), so they are a basis of the column space of A_{red}. Since every non corner column in A_{red} is a linear combination of the corner columns of A_{red}, the same is true for A from what we said above, and therefore, the corner columns in A also span $\text{col}(A)$. The last claim is an immediate consequence. □

This result may seem a little surprising since it involves row reducing A which of course changes $\text{col}(A)$. We immediately get the following corollary.

Corollary 5.12. *For any $A \in \mathbb{F}^{m\times n}$, the dimension $\text{col}(A)$ of A is the rank of A.*

Example 5.14. To consider a simple example, let

$$A = \begin{pmatrix} 1 & 2 & 2 \\ 4 & 5 & 8 \\ 7 & 8 & 14 \end{pmatrix}.$$

Then

$$A_{red} = \begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Proposition 5.11 implies the first two columns are a basis of $\mathrm{col}(A)$. Notice that the first and third columns are dependent in both A and A_{red} as we remarked above. The vector $\mathbf{x} = (2, 0, -1)^T$ expresses this. The Proposition says that the first two columns are a basis of the column space, but makes no assertion about the second and third columns, which in fact are also a basis.

Example 5.15 (Linear Systems and Dimension)**.** The identity originally stated in Proposition 2.4, which says that the number of corner variables plus the number of free variables is the total number of variables can now be put into final form. For any $A \in \mathbb{F}^{m\times n}$,

$$\dim \mathrm{col}(A) + \dim \mathcal{N}(A) = \dim \mathbb{F}^n. \tag{5.3}$$

5.3.2 The Row Space of A and the Ranks of A and A^T

Recall that the row space of an $m \times n$ matrix A over $\mathbb{F}$ is the subspace $\mathrm{row}(A) \subset \mathbb{F}^n$ spanned by the rows of A. The goal of this subsection is to relate the row space to row operations and then to derive a surprising connection between A and A^T.

We first look at how row operations affect the row space (not!).

Proposition 5.13. *Elementary row operations leave the row space of A unchanged. That is, for any elementary matrix E,* $\mathrm{row}(EA) = \mathrm{row}(A)$. *Consequently A and A_{red} always have the same row space. Moreover, the nonzero rows of A_{red} are a basis of* $\mathrm{row}(A)$. *Hence*

$$\dim \mathrm{row}(A) = \mathrm{rank}(A).$$

Proof. Let E be any $m \times m$ elementary matrix over $\mathbb{F}$. We first show that $\mathrm{row}(EA) = \mathrm{row}(A)$. If E is a row swap or a row dilation, this is clear. So we only have to look at what happens if E is an elementary row operation of the type III. Suppose E replaces the ith row $\mathbf{r}_i$ by $\mathbf{r'}_i = \mathbf{r}_i + k\mathbf{r}_j$, where

$k \neq 0$ and $j \neq i$. Since all other rows of EA and A are the same and since $\mathbf{r}'_i$ is itself a linear combination of two rows of A, every row of EA is a linear combination of rows of A. Hence $\text{row}(EA) \subset \text{row}(A)$. But since E^{-1} is also of type III,

$$\text{row}(A) = \text{row}((E^{-1}E)A) = \text{row}(E^{-1}(EA)) \subset \text{row}(EA),$$

so $\text{row}(EA) = \text{row}(A)$. Therefore row operations do not change the row space, and the first claim of the proposition is proved.

It follows that the nonzero rows of A_{red} span $\text{row}(A)$. We will be done if the nonzero rows of A_{red} are independent. But this holds for the same reason the rows of I_n are independent. Every nonzero row of A_{red} has a 1 in the component corresponding to its corner entry, and in this column, all the other rows have a zero. Therefore the only linear combination of the nonzero rows which can give the zero vector is the one where every coefficient is zero. Hence the nonzero rows of A_{red} are independent. (This is the same argument that shows the fundamental solutions of a homogeneous linear system are independent.) Thus they form a basis of $\text{row}(A)$, so $\dim \text{row}(A)$ is the number of nonzero rows of A_{red}. Since this is also $\text{rank}(A)$, we get the final conclusion $\dim \text{row}(A) = \text{rank}(A)$. □

The surprising connection, which you have already noticed, is

Corollary 5.14. *For any $m \times n$ matrix A over a field $\mathbb{F}$,*

$$\dim \text{row}(A) = \dim \text{col}(A).$$

Therefore, the ranks of A and A^T are the same.

Proof. We just saw that $\dim \text{row}(A)$ equals $\text{rank}(A)$. But in Proposition 5.11, we also saw that $\dim \text{col}(A)$ also equals $\text{rank}(A)$. Finally, $\text{rank}(A^T) = \dim \text{col}(A^T) = \dim \text{row}(A) = \text{rank}(A)$, so we are done. □

So far, there is no obvious connection between $\text{row}(A)$ and $\text{col}(A)$, except that they have the same dimension. We will see later that there is another connection given in terms of orthogonal complements.

Let us cap off the discussion with some more examples.

Example 5.16. The 3×3 counting matrix C of Example 2.5 has reduced form

$$C_{red} = \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{pmatrix}.$$

Thus $\dim \text{row}(C) = 2$. Clearly, the first two rows of C are a basis of $\text{row}(C)$ since they span and are clearly independent.

Example 5.17. Suppose $\mathbb{F} = \mathbb{F}_2$ and

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

A is already reduced so its rows are a basis of $\text{row}(A)$, which is thus a three dimensional subspace of $\mathbb{F}^6$. The number of elements in $\text{row}(A)$ is therefore 8, by Proposition 5.9. The 7 nonzero vectors are

$$(100111),\ (010101),\ (001011),\ (110010),\ (101100),\ (011110),\ (1111001).$$

Note that all combinations of 0's and 1's occur in the first three components, since the corners are in these columns. In fact, the first three components tell you which linear combination is involved.

5.3.3 The Uniqueness of Row Echelon Form

Finally, let us return to the fact that the reduced row echelon form of an arbitrary $A \in \mathbb{F}^{m\times n}$ is unique . Recall that we gave a proof in Section 3.4.6, which left some of the details to the reader. Note that the original purpose of showing this was so that we would know that the notion of the rank of a matrix A (as the number of nonzero rows in A_{red}) is well makes sense. We now know another way of defining the rank of A, namely as $\dim \text{row}(A)$, so knowing the uniqueness of the reduced row echelon form isn't as crucial as before. Let us, however, give another proof. First, we restate the result.

Proposition 5.15. *The reduced row echelon form of an $m \times n$ matrix is unique.*

Proof. Suppose $A \in \mathbb{F}^{m\times n}$ has two reduced row echelon forms R and S. It doesn't hurt to suppose A has rank m, so let us do so. We know, by Proposition 5.13, that the nonzero rows of each of these matrices give bases of $\text{row}(A)$ as subspaces of $\mathbb{F}^n$. Let $\mathbf{r}_i$ be the i-th row of R and $\mathbf{s}_i$ the i-th row of S. We'll first show that the corner entries in each matrix occur in the same place. For example, if the first corner entry in R is to the left of the first corner entry in S, then it's impossible to express $\mathbf{r}_1$ as a linear combination of the rows of S, contradicting Proposition 5.13. Hence the first corner entry in R cannot be to the left of that of S. By symmetry, it can't be to the right either, so the first corner entries are in the same column. Now assume the first $(k-1)$ corner entries in each matrix are in the same columns. If the corner entry in the k-th row of S is to the left off that in

the k-th row of R, then, we see that writing $\mathbf{s}_k = \sum a_i \mathbf{r}_i$ is impossible. The same holds if the corner entry in the k-th row of S is to the right off that in the k-th row of R. (Just expand $\mathbf{r}_k = \sum a_i \mathbf{s}_i$.) Hence, by induction, the corners in both R and S occur in the same columns, and hence in the same entries.

Thus if we write $\mathbf{s}_1 = \sum a_i \mathbf{r}_i$, we must necessarily have $a_1 = 1$ and $a_i = 0$ if $i > 1$. For all the components of $\mathbf{s}_1$ corresponding to corner columns of S are zero. But if $a_i \neq 0$ for some $i > 1$, this couldn't happen since the $\mathbf{r}_i$ and the $\mathbf{s}_i$ have their first nonzero entries in the same components. Now argue again by induction. Suppose $\mathbf{r}_i = \mathbf{s}_i$ if $i < k$. We claim $\mathbf{r}_k = \mathbf{s}_k$. Again writing $\mathbf{s}_k = \sum a_i \mathbf{r}_i$, we have $a_i = 0$ if $i < k$, and $a_k = 1$, and, by the above reasoning, $a_i = 0$ if $i > k$. This completes the induction, and therefore $R = S$. $\square$

Exercises

Exercise 5.29. In this problem, the field is $\mathbb{F}_2$. Consider the matrix

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

(a) Find a basis of row(A).

(b) How many elements are in row(A)?

(c) Is (01111) in row(A)?

(d) Find a basis of col(A).

Exercise 5.30. If A and B are $n \times n$ matrices so that B is invertible (but not necessarily A), show that the ranks of A, AB and BA are all the same.

Exercise 5.31. True or False: $\text{rank}(A) \geq \text{rank}(A^2)$. Explain your answer.

Exercise 5.32. Suppose A and B lie in $\mathbb{F}^{n\times n}$ and $AB = O$. Show that $\text{col}(B) \subset \mathcal{N}(A)$, and conclude that $rank(A) + rank(B) \leq n$.

Exercise 5.33. Let A be an $n \times n$ matrix over the reals $\mathbb{R}$. Which of the following matrices have the same rank as A? Briefly explain the reason.

(a) A^T;

(b) $A + A^T$;

(c) $A^T A$ (hint: if $A^T A\mathbf{x} = \mathbf{0}$, note that $\mathbf{x}^T A^T A\mathbf{x} = |A\mathbf{x}|^2$);

(d) AA^T; and

(e) A^{-1}.

5.4 Intersections and Sums of Subspaces

The purpose of this chapter is to consider intersections and sums of subspaces of a finite dimensional vector space. The intersection of two distinct planes in $\mathbb{R}^3$ through the origin is a line through the origin. It is easy to see that the intersection of two subspaces W and Y of a finite dimensional vector space V is a subspace, but not so easy to see what the dimension of this subspace is. For example, instead of the two planes in $\mathbb{R}^3$, consider the possibilities for intersecting two three dimensional subspaces of $\mathbb{R}^4$ or $\mathbb{R}^5$. The dimension can't be more than three, but can it be zero, or is there a lower bound as in the example of two planes?

The answer to these questions is given by the Hausdorff Intersection Formula, which we will prove below. The Hausdorff Intersection Formula gives the relationship between the dimension of the intersection of two subspaces and the dimesion of their sum (see Definition 5.4). Studying the sum of subspaces leads directly to the notion of a direct sum, which we will also introduce here. Direct sums will be used in the study of eigenspaces in a later chapter.

5.4.1 Intersections and Sums

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 5.16. *The intersection $W \cap Y$ of the subspaces W and Y of V is also a subspace of V. More generally, the intersection of any collection of subspaces of V is also a subspace.*

Proof. This is an exercise. □

Proposition 5.16 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 is the intersection of a finite number of hyperplanes in $\mathbb{F}^n$, and each hyperplane is given by a homogeneous linear equation.

Another simple way of forming a new subspace is to take the subspace spanned by W and Y. This is defined as follows.

Definition 5.4. The *sum* of W and Y is defined to be the set

$$W + Y = \{\mathbf{w} + \mathbf{y} \mid \mathbf{w} \in W, \mathbf{y} \in Y\}.$$

More generally, we can form the sum $V_1 + \cdots + V_k$ of an arbitrary (finite) number of subspaces $V_1, V_2, \cdots, V_k$ of V. The sum $V_1 + \cdots + V_k$ is sometimes written as $\sum_{i=1}^{k} V_i$ or more simply $\sum V_i$.

Proposition 5.17. *If $V_1, V_2, \cdots, V_k$ are subspaces of V, then $\sum_{i=1}^{k} V_i$ 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 an exercise also. □

5.4.2 The Hausdorff Intersection Formula

We now come to the problem of what can one say about the intersection of two three dimensional subspaces of $\mathbb{R}^4$ or $\mathbb{R}^5$. It turns out that the answer is a consequence of an elegant formula, which the reader may already have guessed, that gives the relationship between the dimension of the sum $W+Y$ and the dimension of the intersection $W \cap Y$, assuming both W and Y are subspaces of the same finite dimensional vector space.

Theorem 5.18. *If W and Y are subspaces of a finite dimensional vector space V, then*

$$\dim(W + Y) = \dim W + \dim Y - \dim(W \cap Y). \tag{5.4}$$

Proof. As $W \cap Y$ is a subspace of V and V is finite dimensional, the Dimension Theorem implies $W \cap Y$ has a basis, say $\mathbf{x}_1, \ldots, \mathbf{x}_k$, and this basis extends to a basis of W, say $\mathbf{x}_1, \ldots, \mathbf{x}_k, \mathbf{w}_{k+1}, \ldots, \mathbf{w}_{k+r}$. Likewise, $\mathbf{x}_1, \ldots, \mathbf{x}_k$ extends to a basis of Y, say $\mathbf{x}_1, \ldots, \mathbf{x}_k, \mathbf{y}_{k+1}, \ldots, \mathbf{y}_{k+s}$. I claim

$$\mathcal{B} = \{\mathbf{x}_1, \ldots, \mathbf{x}_k, \mathbf{w}_{k+1}, \ldots, \mathbf{w}_{k+r}, \mathbf{y}_{k+1}, \ldots, \mathbf{y}_{k+s}\}$$

is a basis of $W + Y$. It is not hard to see that $\mathcal{B}$ spans. We leave this to the reader. To see $\mathcal{B}$ is independent, suppose

$$\sum_{i=1}^{k} \alpha_i \mathbf{x}_i + \sum_{j=k+1}^{k+r} \beta_j \mathbf{w}_j + \sum_{m=k+1}^{k+s} \gamma_m \mathbf{y}_m = \mathbf{0}. \tag{5.5}$$

Thus

$$\sum \gamma_m \mathbf{y}_m = -\big(\sum \alpha_i \mathbf{x}_i + \sum \beta_j \mathbf{w}_j\big).$$

But the left hand side of this expression lies in Y while the right hand side is in W. It follows that

$$\sum \gamma_m \mathbf{y}_m \in Y \cap W.$$

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, (5.5) becomes the expression

$$\sum \alpha_i \mathbf{x}_i + \sum \beta_j \mathbf{w}_j = \mathbf{0}.$$

Thus 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

$$\dim(W + Y) = k + r + s = (k + r) + (k + s) - k,$$

which is exactly $\dim W + \dim Y - \dim(W \cap Y)$. □

A question such as how do two three dimensional subspaces of a five dimensional space intersect can now be settled. First, note

Corollary 5.19. *If W and Y are subspaces of a finite dimensional vector space V, then*

$$\dim(W \cap Y) \geq \dim W + \dim Y - \dim V. \tag{5.6}$$

In particular, if $\dim W + \dim Y > \dim V$, *then* $\dim(Y \cap W) > 0$.

Proof. Since W and Y are both subspaces of V, $\dim V \geq \dim(W + Y)$. Now apply the Hausdorff Formula. □

Hence two three dimensional subspaces of a five dimensional space intersect in at least a line, though the intersection can also have dimension two or three. (Since $\dim V = 5$, then $\dim(W \cap Y) \geq 3 + 3 - 5 = 1$) However, if $\dim V = 6$, the above inequality only says $\dim(Y \cap W) \geq 0$, hence doesn't say anything new.

Example 5.18 (Intersection of hyperplanes)**.** Recall that a subspace W of V is called a *hyperplane* if $\dim V = \dim W + 1$. 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 5.19. Recall that in §5.2.5, we found bases of the subspaces $\mathbb{F}_s^{n\times n}$ and $\mathbb{F}_{ss}^{n\times n}$ of $n \times n$ symmetric and skew symmetric matrices, provided the characteristic of $\mathbb{F} \neq 2$. In particular, we found that $\dim \mathbb{F}_s^{n\times n} = n(n+1)/2$ and $\dim \mathbb{F}_{ss}^{n\times n} = n(n-1)/2$.

Since $\mathrm{char}(\mathbb{F}) \neq 2$, any $A \in \mathbb{F}^{n\times n}$ can be expressed as the sum of a symmetric matrix and a skew symmetric matrix. Namely,

$$A = 2^{-1}(A + A^T) + 2^{-1}(A - A^T). \tag{5.7}$$

Thus $\mathbb{F}^{n\times n} = \mathbb{F}_s^{n\times n} + \mathbb{F}_{ss}^{n\times n}$. Moreover, since $\mathrm{char}(\mathbb{F}) \neq 2$, the only matrix that is both symmetric and skew symmetric is the zero matrix. Hence $\mathbb{F}_s^{n\times n} \cap \mathbb{F}_{ss}^{n\times n} = \{\mathbf{0}\}$. Hence, by the Hausdorff Intersection Formula,

$$\dim \mathbb{F}^{n\times n} = \dim(\mathbb{F}_s^{n\times n}) + \dim(\mathbb{F}_{ss}^{n\times n}).$$

This agrees with the result in §5.2.5.

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 $\mathbb{F}_{ss}^{n\times n}$. In particular, the decomposition of A in (5.7) as the sum of a symmetric matrix and a skew symmetric matrix is unique (provided $\mathrm{char}(\mathbb{F}) \neq 2$).

5.4.3 Direct Sums of Several Subspaces

By the Hausdorff 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. This observation is related to the following definition.

Definition 5.5. We say that V is the *direct sum* of two subspaces W and Y if $V = W + Y$ and for any $\mathbf{v} \in V$, 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 V is the direct sum of a collection of subspaces $V_1, \ldots, V_k$ if $V = \sum V_i$ and for any $\mathbf{v} \in V$, 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 5.20. *Suppose V is finite dimensional. Then a necessary and sufficient condition that $V = W \oplus Y$ is that $V = W + Y$ and $W \cap Y = \{\mathbf{0}\}$. Moreover, $V = W \oplus Y$ if and only if $V = W + Y$ and $\dim V = \dim W + \dim Y$.*

Proof. First, assume $V = W + Y$ and $W \cap Y = \{\mathbf{0}\}$. To see $V = W \oplus Y$, let $\mathbf{v}$ have two expressions $\mathbf{v} = \mathbf{w} + \mathbf{y} = \mathbf{w}' + \mathbf{y}'$. Then $\mathbf{w} - \mathbf{w}' = \mathbf{y}' - \mathbf{y}$ is an element of $W \cap Y = \{\mathbf{0}\}$, so $\mathbf{w} = \mathbf{w}'$ 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 any non-zero $\mathbf{w} \in W \cap Y$ has two expressions $\mathbf{w} = \mathbf{w} + \mathbf{0} = \mathbf{0} + \mathbf{w}$. This violates the definition of a direct sum, so $W \cap Y = \{\mathbf{0}\}$. For the second claim, suppose $V = W \oplus Y$. Then, $\dim(W \cap Y) = 0$, so the Hausdorff Intersection Formula says $\dim V = \dim(W + Y) = \dim W + \dim Y$. Conversely, if $V = W + Y$ and $\dim V = \dim W + \dim Y$, then $\dim(W \cap Y) = 0$, so $V = W \oplus Y$. □

As a consequence, we get the following general result about $\mathbb{F}^{n\times n}$, mentioned in the previous section.

Proposition 5.21. *Assume char*$(\mathbb{F}) \neq 2$. *Then every* $A \in \mathbb{F}^{n\times n}$ *can be uniquely expressed as in (5.7 as the sum of a symmetric matrix and a skew symmetric matrix.*

We can also extend Proposition 5.20 to any number of subspaces as follows.

Proposition 5.22. *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. We will prove the if statement and leave the only if statement as an exercise. Let us induct on k. If $k = 1$, there's nothing to prove, so assume the result for a $k \geq 1$. Let $V = \sum_{i=1}^{k+1} V_i$, where $V_1, \ldots, V_k, V_{k+1}$ are subspaces of V, and suppose $\dim V = \sum_{i=1}^{k+1} \dim V_i$. Put $W = \sum_{i=1}^{k} V_i$. It is straightforward that $\dim W \leq \sum_{i=1}^{k} \dim V_i$. Since $V = W + V_{k+1}$,

$$\dim V \leq \dim W + \dim V_{k+1} \leq \sum_{i=1}^{k} \dim V_i + \dim V_{k+1} = \dim V.$$

It follows that $\dim W = \sum_{i=1}^{k} \dim V_i$. Proposition 5.20 thus says that $V = W \oplus V_{k+1}$. Moreover, the induction hypothesis says $W = \bigoplus_{i=1}^{k} V_i$. Now suppose $\sum_{i=1}^{k+1} \mathbf{v}_i = \mathbf{0}$, where each $\mathbf{v}_i \in V_i$. Then we know that $\sum_{i=1}^{k} \mathbf{v}_i = \mathbf{0}$ and $\mathbf{v}_{k+1} = \mathbf{0}$. But as $W = \bigoplus_{i=1}^{k} V_i$, it follows that $\mathbf{v}_i = \mathbf{0}$ if $1 \leq i \leq k$. Hence $V = \bigoplus_{i=1}^{k+1} V_i$, as was to be shown. □

Here is another basic example where direct sums occur.

Example 5.20 (Orthogonal Complements)**.** For example, let V be a subspace of $\mathbb{R}^n$. The *orthogonal complement* $V^\perp$ of V is defined to be

$$V^\perp = \{\mathbf{w} \in \mathbb{R}^n \mid \mathbf{w} \cdot \mathbf{v} = 0 \text{ for all } \mathbf{v} \in V\}.$$

Orthogonal complements in $\mathbb{R}^n$ provide examples of direct sums, since by Exercise 5.34, $\dim V + \dim V^\perp = n$ and $V \cap V^\perp = \{\mathbf{0}\}$ (why?). Thus, for any subspace V,

$$\mathbb{R}^n = V \oplus V^\perp. \tag{5.8}$$

5.4.4 External Direct Sums

Suppose 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.

Definition 5.6. The *external direct sum* of V and W is the vector space denoted by $V \times W$ consisting of all pairs $(\mathbf{v}, \mathbf{w})$, where $\mathbf{v} \in V$ and $\mathbf{w} \in W$. Addition is defined component-wise 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 is defined by

$$r(\mathbf{v}, \mathbf{w}) = (r\mathbf{v}, r\mathbf{w}).$$

The alert reader will have noted that $\mathbb{F} \times \mathbb{F}$ is nothing else than $\mathbb{F}^2$, and, more generally, $\mathbb{F}^k \times \mathbb{F}^m = \mathbb{F}^{k+m}$. Thus the external direct sum is a generalization of the construction of $\mathbb{F}^n$. The direct sum operation can be extended (inductively) to any number of vector spaces over $\mathbb{F}$. In fact, $\mathbb{F}^n$ is just the n-fold external direct sum of $\mathbb{F}$.

Note also that V and W can both be considered (in a natural way) as subspaces of $V \times W$ (why?).

Proposition 5.23. *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 5.34. Given a subspace W of $\mathbb{R}^n$, show that $W^\perp$ is a subspace of $\mathbb{R}^n$ and describe a method for constructing a basis of $W^\perp$.

Exercise 5.35. If W is a subspace of $\mathbb{R}^n$, show that $\dim W + \dim W^\perp = n$ and conclude $W \oplus W^\perp = \mathbb{R}^n$.

Exercise 5.36. Let K be a subspace of $\mathbb{C}^n$, and let $K^\perp$ denote the orthogonal complement of K with respect to the Hermitian inner product $(w, z) = \sum_{i=1}^n \overline{w_i} z_i$. Show that

(i) $K^\perp$ is a complex subspace of $\mathbb{C}^n$;

(ii) we have $\dim K + \dim K^\perp = n$; and

(iii) $\mathbb{C}^n = K \oplus K^\perp$.

Exercise 5.37. Suppose $\mathbb{F}$ is an arbitrary field and let W be a subspace of $\mathbb{F}^n$. Then $W^\perp$ can be defined in exactly the same way as in the real case.

(i) Show that W is a subspace of $\mathbb{F}^n$.

(ii) Show that $\dim W + \dim W^\perp = n$.

(iii) Show by example that it isn't necessarily true that $W + W^\perp = \mathbb{F}^n$.

Exercise 5.38. This exercise is used in the proof of Proposition 5.22. Show that if a vector space V is the sum of subspaces $V_1, \ldots, V_k$, then

$$\dim V \leq \sum_{i=1}^{k} \dim V_i.$$

Exercise 5.39. Prove the if statement of Proposition 5.22.

Exercise 5.40. Prove Proposition 5.23.

5.5 Vector Space Quotients

We conclude this chapter with the construction of the quotient of a vector space V by a subspace W. This construction requires that we first introduce the general notion of a quotient space, which is based on the concept of an equivalence relation. The particular vector space we are going to study is denoted by V/W. Its elements are called cosets. The reader will see that in a certain sense, the vector space V/W can be thought of as the result of subtracting W from V (not dividing V by W), but not too much should be read into this statement.

5.5.1 Equivalence Relations and Quotient Spaces

The first step in defining V/W is to introduce the concept of an equivalence relation on a set. As the reader will see, this notion equivalence is simply a generalization of the notion of equality. First, we need to recall what a relation on a set is. Let S be a set, and recall that $S \times S$ denotes the product set consisting of all pairs (s, t) with $s, t \in S$.

Definition 5.7. Let S be a non-empty 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 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.

It follows from (ii) and (iii) that any two elements in an equivalence class are equivalent.

Example 5.21. As mentioned above, an obvious example of an equivalence relation on an arbitrary set S is equality. That is, sEt if and only if $s = t$. The equivalence classes consist of the singletons $\{s\}$, as s varies over S.

Example 5.22. The notion of an equivalence relation gives a new way of defining a prime field. Let p be prime. Define an equivalence relation on $\mathbb{Z}$ by saying that two integers m and n are equivalent if and only if $m - n$ is divisible by p. It can readily be checked that this defines an equivalence relation. (See the proof of Proposition 5.26 below.)

Let $\mathbb{E}_p$ denote the set of equivalence classes. There are, in fact, p classes, which are represented by $0, 1, \ldots, p-1$. Let $[m]$ denote the equivalence class of m. Then we can define an addition of equivalence classes by adding any two representatives: $[m] + [n] = [m + n]$. Similarly, we can define a multiplication of equivalence classes by putting $[m][n] = [mn]$. One can easily check that these binary operations are well defined. The point is that they make $\mathbb{E}_p$ into a field, In fact, this field is another way of defining the prime field $\mathbb{F}_p$.

The above Example also gives a nice illustration of the following Proposition.

Proposition 5.24. *Let E is an equivalence relation on a set S. Then every element $a \in S$ is in an 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 $a \in C_1 \cap C_2$. By definition, every element of C_1 is equivalent to a, so $C_1 \subset C_2$ since $a \in C_2$. Similarly, $C_2 \subset C_1$, hence they are equal. This finishes the proof. □

Definition 5.8. The set of equivalence classes in S of an equivalence relation E is called the *quotient of S by E*. Sometimes this quotient will be denoted as S/E.

5.5.2 Cosets

Suppose V be a vector space over $\mathbb{F}$, and let W be a subspace. We are now going to define an equivalence relation on V whose equivalence classes are called *cosets of W in V*. The set of cosets will be denoted as V/W, and the main result is that V/W can be made into a vector space over $\mathbb{F}$. The definition is given in the following Proposition.

Proposition 5.25. *Let V be a vector space over $\mathbb{F}$ and let W be a subspace. Given $\mathbf{v}$ and $\mathbf{y}$ in V, let us say that $\mathbf{v} E_W \mathbf{y}$ if and only if $\mathbf{v} - \mathbf{y} \in W$. Then E_W is an equivalence relation on V.*

Proof. Clearly $\mathbf{v} E_W \mathbf{v}$ since $\mathbf{v} - \mathbf{v} = \mathbf{0} \in W$. If $\mathbf{v} E_W \mathbf{y}$, then $\mathbf{y} E_W \mathbf{v}$ since W is closed under scalar multiplication, and $(\mathbf{y} - \mathbf{v}) = (-1)(\mathbf{v} - \mathbf{y})$. Finally, if $\mathbf{v} E_W \mathbf{y}$ and $\mathbf{y} E_W \mathbf{z}$, then $\mathbf{v} E_W \mathbf{z}$ since $\mathbf{v} - \mathbf{z} = (\mathbf{v} - \mathbf{y}) + (\mathbf{y} - \mathbf{z})$ and W is closed under addition. Hence E_W is an equivalence relation on V. □

Let $\mathbf{v} \in V$ be fixed, and define $\mathbf{v} + W$ to be the set of all sums $\mathbf{v} + \mathbf{w}$, where $\mathbf{w}$ varies through W. That is,

$$\mathbf{v} + W = \{\mathbf{v} + \mathbf{w} \mid \mathbf{w} \in W\}. \tag{5.9}$$

Proposition 5.26. *The cosets of W in V are precisely the sets of the form $\mathbf{v} + W$, where $\mathbf{v}$ varies through V. In particular, $\mathbf{v} + W = \mathbf{y} + W$ if and only if $\mathbf{v} - \mathbf{y} \in W$.*

Proof. Let C denote the equivalence class of $\mathbf{v}$ and consider the coset $\mathbf{v} + W$. If $\mathbf{y} E_W \mathbf{v}$, then $\mathbf{y} - \mathbf{v} = \mathbf{w} \in W$. Hence $\mathbf{y} = \mathbf{v} + \mathbf{w}$, so $\mathbf{y} \in \mathbf{v} + W$. Therefore $C \subset \mathbf{v} + W$. Arguing in reverse, we likewise conclude that $\mathbf{v} + W \subset C$. □

We will denote the quotient space V/E_W simply by V/W. We refer to V/W as V *modulo* W .

Example 5.23. The geometric interpretation of a coset is straightforward. For example, if $V = \mathbb{R}^3$ 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 properties of Proposition 5.24 are all illustrated by the properties of parallel planes.

Our next goal is to show that V/W has a well defined addition and scalar multiplication, which makes it into a vector spaceover $\mathbb{F}$. Given two cosets $(\mathbf{v} + W)$ and $(\mathbf{y} + W)$, define their sum by putting

$$(\mathbf{v} + W) + (\mathbf{y} + W) = (\mathbf{v} + \mathbf{y}) + W. \tag{5.10}$$

In order that this addition be a binary operation on V/W, we have to show that the rule (5.10) 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 due to the fact that

$$(\mathbf{v} + \mathbf{y}) - (\mathbf{v}' + \mathbf{y}') = (\mathbf{v} - \mathbf{v}') + (\mathbf{y} - \mathbf{y}'),$$

$\mathbf{v} - \mathbf{v}' \in W$, $\mathbf{y} - \mathbf{y} \in W$ W is a subspace. Therefore, addition on V/W is well defined. Scalar multiplication on cosets is defined by

$$a(\mathbf{v} + W) = a\mathbf{v} + W, \tag{5.11}$$

and, by a similar argument, scalar multiplication is also well defined.

There are two limiting cases of V/W which are easy to understand. If $W = V$, then V/W has exactly one element, namely $\mathbf{0}$. On the other hand, if $W = \{\mathbf{0}\}$, then V/W is V, since the cosets have the form $\mathbf{v} + \{\mathbf{0}\}$. In this case, defining V/W is nothing more than V, since each coset consists of a unique element of V.

We can now prove

Theorem 5.27. *Let V be a vector space over a field $\mathbb{F}$ and suppose W is a subspace of V. Define V/W to be the set of cosets of W in V with addition and scalar multiplication defined as in* (5.10) *and* (5.11). *Then V/W is a vector space over $\mathbb{F}$. If V is finite dimensional, then*

$$\dim V/W = \dim V - \dim W. \tag{5.12}$$

Proof. The fact that V/W satisfies the vector space axioms is straightforward, so we will omit most of the details. 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.

To check the dimension formula (5.12), let $\dim V = n$. Choose a basis $\mathbf{w}_1, \ldots, \mathbf{w}_k$ 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 the cosets $\mathbf{v}_1 + W, \ldots, \mathbf{v}_{n-k} + W$ are a basis of V/W. To see they are independent, put $\mathbf{v}_i + W = \mathbf{y}_i$ if $1 \le i \le n-k$, and suppose there exist $a_1, \ldots, a_{n-k} \in \mathbb{F}$ such that $\sum_{i=1}^{n-k} a_i \mathbf{y}_i = \mathbf{0} + W$. 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 $\mathbf{w}_j$ comprise a basis of V implies that all a_i and b_j are zero. Therefore $\mathbf{y}_1, \ldots, \mathbf{y}_{n-k}$ are independent. We leave the fact that they span V/W as an exercise. □

One thing to notice about the quotient space V/W is that although the above Theorem tells us its dimension, it doesn't tell us there is natural choice of a basis. In order to find a basis in the above proof, we first needed a basis of W, which was then extended to a basis of V. Furthermore, the quotient space V/W is an abstract construction. It is not a subspace of V or W or some other natural vector space. For example, if $m < n$, then $\mathbb{F}^m$ can be considered in a natural way to be a subspace of subset $\mathbb{F}^n$, namely the subspace consisting of all n-tuples whose last $n - m$ components are zero. Now $\mathbb{F}^n/\mathbb{F}^m$ is vector space over $\mathbb{F}$ of dimension $n - m$, yet there is no natural identification between $\mathbb{F}^n/\mathbb{F}^m$ and $\mathbb{F}^{n-m}$.

Cosets will appear in a concrete setting in Chapter 6 as an essential ingredient in the construction of standard decoding tables for linear codes.

Exercises

Exercise 5.41. Prove that the cosets $\alpha_1, \ldots, \alpha_{n-k}$ defined in the proof of Theorem 5.27 span V/W.

Exercise 5.42. Let W be the subspace of $V = V(4,2)$ 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 5.43. 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 5.44. 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 5.45. 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 5.46. Let V be a vector space over $\mathbb{F}_p$ of dimension n, and let W be a subspace of dimension k.

(1) Show that every coset of W has p^k elements. (Hint: find a bijection from W to $\mathbf{x} + W$.)

(2) Show that the number of cosets of W is $p^{(n-k)}$.

5.6 Summary

In the previous chapter, we introduced the notion of a vector space V over an arbitrary field. The purpose of this chapter is to consider the basic theory of finite dimensional vector spaces: those with a finite spanning set. We first considered were the concepts of bases and dimension. A basis of a finite dimensional vector space V is a subset $\mathcal{B}$ of V such that every vector in V can be uniquely expressed as a linear combination of elements of $\mathcal{B}$. A basis has two properties: it spans V and is linearly independent. There are two other ways of thinking about a basis. A basis is a minimal spanning set and a maximal linearly independent subset of V.

We proved that every finite dimensional vector space V has a finite basis, and any two bases of V have the same number of vectors. The dimension of V is be defined as the number of elements in a basis of V. We also showed that every independent subset of V is contained in a basis, and every spanning set contains a basis.

After we introducing and proving the proerties of dimension, we considered several examples such as the row and column spaces of a matrix. These turn out to have the same dimension, a somewhat surprising fact. We also constructed some new vector spaces and computed their dimensions. For example, if U and W are subspaces of V, we defined the sum $U + W$, the smallest subspace of V containg both U and W and computed $\dim(U + W)$. The answer is given by the Hausdorff Intersection Formula: $\dim(U + W) = \dim U + \dim W - \dim(U \cap W)$. We also defined what it means to say V is the direct sum of subspaces U and W and gave examples.

Finally, we introduced the concept of the quotient of a vector space V by a subspace W. The quotient space V/W is obtained by first defining the notion of equivalence relation, and then considering a certain equivalence relation on V whose elements are the cosets of W in V. The dimension of V/W is $\dim V - \dim W$, as long as V is finite dimensional. The quotient V/W is an abstract vector space. It doesn't have a standard geometric model or interpretation, even if V and W are concrete vector spaces.

Chapter 6

Linear Coding Theory

Coding theory is a relatively new branch of mathematics which has had an enormous impact on the electronic revolution. The fundamental idea in the subject is that it is possible to design codes (that is, packages of binary strings) with the property that a random binary string close enough to one of the code's strings is only near one of the code strings. These are called error-correcting codes. They have been extremely important for data transmission from the time of the Mariner space probes to Venus in the nineteen sixties and seventies to the present, when electronic devices such as PCs, CDs, modems etc. have an enormous impact. To cite an example of the power of error-correcting codes, one of the error-correcting codes used by NASA for the Mariner space probes consisted of 64 strings of 32 zeros and ones, that is 32-bit strings. This code is able to figure out which string was transmitted even if the received string has up to seven errors. In other words, about a quarter of the digits of a received string could be incorrrect, and yet the correct string could be identified.

We will concentrate on a special class of codes called linear codes. A linear codes is simply a finite dimensional vector space over a prime field. The foundations of linear coding theory are applications of the concepts treated in Chapter 5.

6.1 Linear Codes

The purpose of this section is to introduce some of the basic concepts in linear coding theory. We will begin with the idea of a code. Recall that $\mathbb{F}_p$ denotes the field with p elements, where p is a prime, and $V(n,p)$ denotes $\mathbb{F}^n$, where $\mathbb{F} = \mathbb{F}_p$.

6.1.1 The Notion of a Code

Let p be a prime.

Definition 6.1. A *p-ary code* is just a nonempty subset C of some $V(n,p)$. The integer n is called the *length* of C, and the elements of C are called the *codewords*. The number of elements of C is denoted by $|C|$.

Since $V(n,p)$ is finite, every p-ary code of length n is finite.

Proposition 6.1. *The number of p-ary codes of length n is 2^{p^n}.*

Proof. By definition, the number of p-ary codes of length n is just the number of subsets of $V(n,p)$. But by Exercise 6.1, a set with k elements has exactly 2^k subsets, so the result follows from the fact that $|V(n,p)| = p^n$. □

We now define the notion of a linear code.

Definition 6.2. Let C be a p-ary code of length n. Then we say that the code C is *linear* if C is a linear subspace of $V(n,p)$.

Note that we will usually denote elements of $V(n,p)$ as strings $v_1 \dots v_n$ of elements of $\mathbb{F}_p$ of length n. Thus $0 \le v_i \le p-1$ for each index i.

Binary codes are the simplest and most pleasant to compute with, so we will concentrate primarily on them. The elements of $V(n,2)$ are known as binary strings of length n, or often *n-bit strings* . For example, there are 4 two-bit strings: 00, 01, 10, and 11.

In order to check that a binary code $C \subset V(n,2)$ is be linear, it suffices to check that the sum of any two codewords is a codeword. (For example, since C contains a codeword $\mathbf{c}$ and $\mathbf{c} + \mathbf{c} = \mathbf{0}$, the nullword $\mathbf{0} \in C$.) Having the structure of a vector space means that linear codes have much more structure than non-linear codes.

Notice that a linear code is completely determined once a set of codewords which spans it is given. In particular, the elements of a basis of a linear code are called *basic codewords*. If $\dim C = k$, there are k basic codewords. We use the special notation *p-ary $[n,k]$-code* when referring to a linear C subspace of $V(n,p)$ having dimension k. Of course, once one know a set of basic codewords for a linear code, one know exactly how many codewords C contains.

Proposition 6.2. *If C is a p-ary $[n,k]$-code, then $|C| = p^k$.*

Proof. Just apply Proposition 5.9. □

Example 6.1. The equation $x_1 + x_2 + x_3 + x_4 = 0$ over $\mathbb{F}_2$ defines a 4-bit linear code of dimension 3, hence a binary $[4, 3]$-code. Thus there are $8 = 2^3$ codewords. Rewriting the defining equation as $x_1 + x_2 + x_3 = x_4$, we see that x_4 can be viewed as a check digit since it's uniquely determined by x_1, x_2, x_3. Here, the codewords are the 4-bit strings with an even number of 1's. A particular set of basic codewords is $\{1001, 0101, 0011\}$, although there are also other choices. (How many?)

Example 6.2. Let

$$A = \begin{pmatrix} 1 & 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 \end{pmatrix},$$

and let C be the binary linear code of length 6 spanned by the rows of A. That is, $C = \text{row}(A)$. Since A is in row reduced form, its rows form a set of basic codewords for C. Thus C is a three dimensional subspace of $V(6, 2)$, so $|C| = 8$. The 7 non zero codewords are

$$100111,\ 010101,\ 001011,\ 110010,\ 101100,\ 011110,\ 111001.$$

Notice that every possible combination of 0's and 1's occurs in the first three positions, and in fact the first three letters tell you which linear combination of the basic codewords is being used. The last three letters serve as check digits. That is, the way to see if a 6-bit string is in the code, just find the (unique) codeword that has the first three digits of the string to be tested, and check if the last three digits agree. For example, 111111 isn't a codeword because its last three digits are 111, not 001.

6.1.2 Linear Codes Defined by Generating Matrices

A nice way to present a linear code is to display a *generating matrix* for the code, such as the matrix A of the previous example.

Definition 6.3. A *generating matrix* for a p-ary linear $[n, k]$-code C is a $k \times n$ matrix over $\mathbb{F}_p$ of the form $M = (I_k \mid A)$ such that $C = \text{row}(M)$.

Notice that the rows of a generating matrix are a set of basic codewords. Not every linear code is given by a generating matrix, but there is always a permutation of the components of C for which a generating matrix exists. Since a generating matrix is in reduced row echelon form, a code which has a generating matrix can only have one generating matrix.

Example 6.3. Let C be the binary [4,2]-code with generating matrix

$$M = \begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \end{pmatrix}.$$

Since M has rank two, $|C| = 2^2$. Taking the totality of linear combinations of the rows, we find that

$$C = \{0000, 1011, 0101, 1110\}.$$

This agrees with the claim $|C| = 4$.

Let us expand on the comment in Example 6.2 about check digits. If M is a generating matrix, then every element of the linear code $C = \mathrm{row}(M)$ can be expressed as a matrix product of the form $(x_1 \ \dots \ x_k)M$ for a suitable choice of the x_i. (To see this, transpose the fact that the column space of M^T consists of all vectors of the form $M^T(y_1 \ \dots \ y_n)^T$.) Let $\mathbf{x} = (x_1 \dots x_k)$, and put $\mathbf{c}(\mathbf{x}) = (x_1 \ \dots \ x_k)M$. Thus,

$$\mathbf{c}(\mathbf{x}) = (x_1 \dots x_k \sum_{i=1}^{k} a_{i1}x_i \cdots \sum_{i=1}^{k} a_{i(n-k)}x_i).$$

Since $x_1, \dots, x_k$ are completely arbitrary, the first k entries $x_1 \dots x_k$ are called the *message digits* and the last $n-k$ digits are called the *check digits*.

6.1.3 The International Standard Book Number

Ever since 1969, the world's publishers have issued an International Standard Book Number, or ISBN, to every book they have published. An ISBN is a 10 digit string of integers $a_1 \cdots a_9 \ a_{10}$ such that $a_1, \dots, a_9$ can take any values between 0 and 9 but a_{10} is also allowed to take the value 10, which is denoted by X for convenience. (Note that X is the Roman numeral standing for 10.)

For example, the book *Fermat's Enigma* by Simon Singh, published in 1997 by Penguin Books, has ISBN 0-14-026869-3. The first digit 0 indicates that the book is in English, the next block of digits identify the publisher as Penguin, and the next block is the set of digits that Penguin has assigned to the title. The last digit is the check digit, which is obtained as explained below. The major publishing companies are given shorter blocks (Penguin's is 14), which allows them to assign more titles. Small publishing companies

are assigned longer blocks, hence less titles. The check digit a_{10} is defined by requiring that $a_1 \cdots a_9\ a_{10}$ be a solution of the homogeneous linear equation

$$a_1 + 2a_2 + 3a_3 + \cdots + 9a_9 + 10a_{10} = 0$$

over $\mathbb{F}_{11}$. Hence the set of ISBNs is a subset of the 11-ary linear code [10,9]-code defined by the above homogeneous linear equation. The generating matrix of C is (1 2 3 4 5 6 7 8 9 10). Since $10+1=0$ in $\mathbb{F}_{11}$, the check digit a_{10} can be expressed explicitly as

$$a_{10} = \sum_{i=1}^{9} ia_i.$$

Note that the ISBN's are only a subset of C.

Example 6.4. For example, 0-15-551005-3 is an ISBN since $0+2+15+20+25+6+0+0+453 \equiv 3 \mathrm{mod}(11)$, as is 0-14-026869-3 from the above example.

Example 6.5. Suppose that an ISBN is entered as 0-19-432323-1. With a minimum amount of technology, the machine in which the numbers are being entered will warn the librarian that 0-19-432323-1 is not an ISBN: that is, $(0,1,9,4,3,2,3,2,3,1)$ doesn't satisfy $\sum_{i=1}^{10}\ ia_i = 0$ over $\mathbb{F}_{11}$. An error has been detected, but the place where the error is hasn't. There may be a single incorrect digit, or two digits may have been transposed. These two possibilities are probably the most common types of error. The next result says something about them.

Proposition 6.3. *A vector $\mathbf{a} = (a_1, \ldots, a_{10}) \in V(10,11)$ that differs from an element of C in exactly one place cannot belong to C; in particular it cannot be an ISBN. Similarly, an element of $V(10,11)$ obtained by transposing two unequal letters of an ISBN cannot be an ISBN.*

Proof. We will prove the first assertion but leave the second as an exercise. Suppose $\mathbf{c} = (c_1, \ldots, c_{10})$ is a codeword which differs from $\mathbf{a} \in V(10,11)$ in one exactly component, say $c_i = a_i$ if $i \neq j$, but $c_j \neq a_j$. Then

$$\mathbf{v} := \mathbf{a} - \mathbf{c} = (0, \ldots, 0, a_j - c_j, 0, \ldots, 0).$$

If $\mathbf{a} \in C$, then $\mathbf{v} - \mathbf{a} \in C$ too, hence $j(a_j - c_j) = 0$ in $\mathbb{F}_{11}$. But since neither j nor $a_j - c_j$ is zero in $\mathbb{Z}_{11}$, this contradicts the fact that $\mathbb{F}_{11}$ is a field. Hence $\mathbf{a} \notin C$. □

Suppose you know all but the kth digit of an ISBN. Can you find the missing digit? Try this with an example, say 0-13-832x44-3. This is a sure way to mildly astound your friends and relatives and maybe win a few bets. But don't bet with a librarian.

Exercises

Exercise 6.1. Show that the number of subsets of a finite set S is $2^{|S|}$. (Hint: use the fact that the number of subsets of a set with n elements is, by combinatorial reasoning, $\sum_{k=0}^{n} \binom{n}{k}$. Now use the binomial theorem to get the result.)

Exercise 6.2. Let $A \in \mathbb{F}^{m\times n}$, where $\mathbb{F} = \mathbb{F}_2$. How many elements does $\mathcal{N}(A)$ contain? How many elements does $\mathrm{col}(A)$ contain?

Exercise 6.3. Suppose C is the code $V(n, p)$. What is the generating matrix of C?

Exercise 6.4. Consider the generating matrix

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

Let us view both $\mathcal{N}(A)$ and $\mathrm{row}(A)$ as subspaces of $V(6, 2)$. What is $\dim(\mathcal{N}(A) \cap \mathrm{row}(A))$?

Exercise 6.5. Determine all x such that 0-13-832x4-4 is an ISBN.

Exercise 6.6. Determine all x and y such that both 1-2-3832xy4-4 and 3-33-x2y377-6 are ISBNs.

Exercise 6.7. Prove the second assertion of Proposition 6.3.

6.2 Error-Correcting Codes

We now come to one of the fundamental ideas in coding theory, the notion of error correcting. This is based on an extremely simple idea, namely that one can put a distance function on codewords. The distance between two codewords is just the number of places in which they differ. We begin by explaining this notion in much more detail.

6.2.1 Hamming Distance

One defines the distance between $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ to be $|\mathbf{u} - \mathbf{v}|$. It's also possible to define the distance between two elements of $V(n, p)$ in a somewhat similar (but different) manner.

Definition 6.4. 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

$$d(\mathbf{u}, \mathbf{v}) = \omega(\mathbf{u} - \mathbf{v}).$$

The Hamming distance $d(\mathbf{u}, \mathbf{v})$ will usually be refered to simply as the distance between $\mathbf{u}$ and $\mathbf{v}$.

Example 6.6. For example, $\omega(1010111) = 5$. The distance

$$d(1010111, 1111000) = 5.$$

Note that the only vector of weight zero is the zero vector. Therefore $\mathbf{u} = \mathbf{v}$ exactly when $\omega(\mathbf{u} - \mathbf{v}) = 0$. What makes the Hamming distance d so useful are the properties listed in the next Proposition.

Proposition 6.4. *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*) $d(\mathbf{u}, \mathbf{w}) \leq d(\mathbf{u}, \mathbf{v}) + d(\mathbf{v}, \mathbf{w})$.

Properties (i),(ii) and (iii) are well known for the usual distance on $\mathbb{R}^n$. Property (iii) is known as the *triangle inequality* , so named because in $\mathbb{R}^n$ it says that the length of any side of a triangle can't exceed the sum of the lengths of the other two sides. In general, if S is a 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 called the distance between $s, t \in S$.

The first two properties of the Hamming distance are easy to see, but the triangle inequality requires proof.

Proof of the triangle inequality. First consider the case where $\mathbf{u}$ 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 vectors obtained by dropping the components where $\mathbf{u}$ and $\mathbf{v}$ agree. Thus we are in the previous case, so

$$d(\mathbf{u},\mathbf{v}) = d(\mathbf{u}',\mathbf{v}') \leq d(\mathbf{u}',\mathbf{w}') + d(\mathbf{v}',\mathbf{w}').$$

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 taking fewer 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. □

Let $d(C)$ denote the minimum distance between any two distinct codewords. That is,

$$d(C) = \min\{d(\mathbf{c},\mathbf{c}') \mid \mathbf{c},\mathbf{c}' \in C,\ \mathbf{c} \neq \mathbf{c}'\}.$$

In general, $d(C)$ has to be computed by finding the distance between every pair of distinct codewords. If there are m codewords, this requires

$$\binom{m}{2} = \frac{m(m-1)}{2}$$

calculations (check this). But if C is linear, finding $d(C)$ requires a lot fewer operations.

Proposition 6.5. *If $C \subset V(n,p)$ is linear, then $d(C)$ is the minimum of the weights of all the non zero codewords. That is,*

$$d(C) = \min\{\omega(\mathbf{c}) \mid \mathbf{c} \in C,\ \mathbf{c} \neq \mathbf{0}\}.$$

Proof. We will leave this as an exercise. □

6.2.2 The Key Result

We first mention some common notation incorporating the minimal distance. Coding theorists refer to an arbitrary code $C \subset V(n,p)$ such that $|C| = M$ and $d(C) = d$ as a p-ary (n, M, d)-*code* . If C is linear, we know that $|C| = M = p^k$ where k is the dimension of C. In that case, they say C is a said to be p-*ary* $[n, k, d]$-*code*. As the next result shows, in designing codes, 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 6.6. *An (n, M, d)-code C can detect up to $d-1$ errors, i.e. if $\mathbf{c} \in C$ and $d(\mathbf{v}, \mathbf{c}) \leq d-1$ for some $\mathbf{v} \in V(n,p)$, then either $\mathbf{v} = \mathbf{c}$ or $\mathbf{v} \notin C$. Moreoever, C corrects up to $e = [(d-1)/2]$ errors, where $[r]$ denotes the greatest integer less than r. That is, if $\mathbf{v}$ is not a codeword, then there is at most one codeword $\mathbf{c}$ such that $d(\mathbf{v}, \mathbf{c}) \leq e$.*

Proof. We will leave the first assertion as an exercise but prove the harder second assertion. Assume $\mathbf{v}$ is not a codeword and $d(\mathbf{v}, \mathbf{c}) \leq (d-1)/2$, for some codeword $\mathbf{c}$. Suppose also that there exists a $\mathbf{c}' \in C$ such that $d(\mathbf{v}, \mathbf{c}') \leq d(\mathbf{v}, \mathbf{c})$. If $\mathbf{c} \neq \mathbf{c}'$, then, by definition, $d(\mathbf{c}, \mathbf{c}') \geq d$. But, by the triangle inequality,

$$d(\mathbf{c}, \mathbf{c}') \leq d(\mathbf{c}, \mathbf{v}) + d(\mathbf{v}, \mathbf{c}') \leq (d-1)/2 + (d-1)/2 = d-1.$$

This is impossible, so we conclude $\mathbf{c} = \mathbf{c}'$. □

If $\mathbf{v} \notin C$, but $d(\mathbf{v}, \mathbf{c}) \leq e$, then we say $\mathbf{c}$ is *error-correcting* for $\mathbf{v}$. The conclusion about error-correction means that if all but e digits of a codeword $\mathbf{c}$ are known, then every digit of $\mathbf{c}$ is known.

Example 6.7. Suppose C is a 6-bit code with $d = 3$. Then $e = 1$. If $\mathbf{c} = 100110$ is a codeword, then $\mathbf{v} = 000110$ can't be, but 100110 is the unique codeword within Hamming distance 1 of 000110.

Example 6.8. For the binary [4,3]-code 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 we don't know that it can correct any errors because $(d-1)/2 = 1/2 < 1$. However, if some additional information is known, such as the unique component where an error occurs, then the error can be corrected using the linear equation defining the code.

Exercises

Exercise 6.8. Consider the binary code $C \subset V(6,2)$ which consists of 000000 and the following nonzero codewords:

$$(100111), (010101), (001011), (110010), (101100), (011110), (111001).$$

(i) Determine whether or not C is linear.

(ii) Compute $d(C)$.

(iii) How many elements of C are nearest to (011111)?

(iv) Determine whether or not 111111 is a codeword. If not, is there a codeword nearest 111111?

Exercise 6.9. Show that for every k between 0 and n, there exists a linear code $C \subset V(n,2)$ such that that $d(C) = k$.

Exercise 6.10. Prove the first part of Proposition 6.6.

Exercise 6.11. Compute $d(C)$ for the code C of Example 6.2.

Exercise 6.12. 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.

(iii) Find $d(\mathcal{N}(A))$.

Exercise 6.13. Prove Proposition 6.5. That is, show that if $C \subset V(n,p)$ is linear, then $d(C)$ is the minimum of the weights of all the non zero codewords.

Exercise 6.14. Can Proposition 6.3 be deduced from Proposition 6.6?

6.3 Codes With Large Minimal Distance

The purpose of this section is to give some examples of codes with a large minimal distance. We will also prove an inequality which shows that $d(C)$ cannot be arbitrarily large.

6.3.1 Hadamard Codes

We will begin with an interesting construction of a family of binary codes based on a class of matrices named after the French mathematician, J. Hadamard. These codes which have the property that $d(C)$ is impressively large compared with $|C|$.

Definition 6.5. An $n \times n$ matrix H over $\mathbb{R}$ whose only entries are $\pm$ 1 is said to be *Hadamard* if and only if $HH^T = nI_n$.

It is easy to see that $HH^T = nI_n$ if and only if $H^T H = nI_n$.

Example 6.9. Examples of $n \times n$ Hadamard matrices for $n = 2, 4, 8$ are

$$H_2 = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}, \quad H_4 = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & -1 & 1 & -1 \end{pmatrix},$$

and

$$H_8 = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \end{pmatrix}.$$

After this point, it is no longer instructive to write them down. One can produce other Hadamard matrices from these by the transformation $H \mapsto PHQ$, where P and Q are permutation matrices.

Proposition 6.7. *If H is an $n \times n$ Hadamard matrix, then:*

(*i*) *any two distinct rows or any two distinct columns are orthogonal;*

(*ii*) *if $n > 1$, then any two rows of H agree in exactly $n/2$ places; and*

(iii) n is either 1, 2 or a multiple of 4.

The first two parts are easy. We will leave the third part as a challenging exercise. An interesting point is the fact that it's still an open problem as to whether there is a $4k \times 4k$ Hadamard matrix for every $k > 0$. Hadamard matrices exist for $k \leq 106$, but, at this time, it doesn't seem to be known whether there is a 428×428 Hadamard matrix. On the other hand, it is also known that if n is a power of two, then there exists an $n \times n$ Hadamard matrix, so aritrarily large Hadamard matrices exist also.

Hadamard codes are defined as follows. Let H be Hadamard, and consider the $n \times 2n$ matrix $(H|-H)$. Define $\mathcal{H}$ to be the binary matrix obtained by replacing all -1's by 0's. That is, $\mathcal{H} \in (\mathbb{F}_2)^{n \times 2n}$. The *Hadamard code* C associated to H is by definition the set of columns of $\mathcal{H}$. Thus a Hadamard code is a binary n-bit code with $2n$-codewords.

Proposition 6.8. *Let C be an n-bit Hadamard code. Then $d(C) = n/2$.*

Proof. Recall that n is a multiple of 4, so $n/2$ is an even integer. The fact that the ith and jth columns of H are orthogonal if $i \neq j$ implies they must differ in exactly $n/2$ components since all the components are ± 1. But the ith and jth columns of H and $-H$ are also orthogonal if $i \neq j$, so they differ in $n/2$ places too. Moreover, the ith columns of H and $-H$ differ in n places. This proves $d(C) = n/2$, as asserted. □

In the notation of Section 6.2.2, C is a binary $(n, 2n, n/2)$-code. For example, the Hadamard matrix H_2 gives the $(2, 4, 1)$-code

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

The code that was used in the transmission of data from the Mariner space probes to Venus in the 1970's was a binary $(32, 64, 16)$ Hadamard code. Since $(16 - 1)/2 = 7.5$, this code corrects 7 errors.

6.3.2 A Maximization Problem

Designing codes such as Hadamard codes, where $d(C)$ is large compared to $|C|$, is one of the basic problems in coding theory. Consider the following question. What is the maximum value of $|C|$ among all binary codes $C \subset V(n, 2)$ of length n such that $d(C) \geq m$ for a given integer m? If we impose the condition that C is linear, we're actually seeking to maximize $\dim C$, since $|C| = p^{\dim(C)}$. The binary $[8, 4, 4]$-code C_8 of the next example is such a code.

Example 6.10. Let

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

The row space C_8 of A is called the *extended Hamming code*. Notice that every row of A has weight 4, so the minimum distance of C_8 is at most 4. It can be seen (for example by enumerating the elements) that $d(C_8) = 4$. Hence C_8 is a binary $[8, 4, 4]$-code.

Proposition 6.9. *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 $\dim C \geq 5$. 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 changing C. By reordering the columns for convenience, we can suppose A_{red} has the form $(I_r|M)$, where $r \geq 5$. Hence M has at most three columns. But the only way we can have $d(C) \geq 4$ is if all entries of M are 1. Then subtracting the second row of A_{red} from the first gives an element of C having weight 2, contradicting $d(C) \geq 4$. Thus 4 is the maximum dimension of any such C in $V(8, 2)$. □

A similar argument gives the well known *singleton bound*.

Proposition 6.10. *If C is a p-ary $[n, k]$-code, then*

$$d(C) \leq n - k + 1.$$

Put another way, a linear code C of length n satisfies

$$\dim C + d(C) \leq n + 1.$$

We leave the proof as an exercise. A linear $[n, k]$-code C with $d(C) = n - k + 1$ is said to be *maximal distance separating.*

Exercises

Exercise 6.15. Suppose $H \in \mathbb{R}^{n\times n}$ is a ± 1 matrix such that $H^T H = nI_n$. Show that H is Hadamard.

Exercise 6.16. Prove parts (i and (ii) of Proposition 6.7.

Exercise 6.17. * Prove that the order of a Hadamard matrix is 1, 2 or a multiple of 4.

Exercise 6.18. Let r be a positive integer and define the ball $B_r(\mathbf{a}) \subset V(n,2)$ about $\mathbf{a} \in V(n,2)$ to be

$$B_r(\mathbf{a}) = \{\mathbf{x} \in V(n,2) \mid d(\mathbf{x},\mathbf{a}) \leq r\}.$$

Show that

$$|B_r(\mathbf{x})| = \sum_{i=0}^{r} \binom{n}{i}.$$

Exercise 6.19. Generalize Exercise 6.18 from $V(n,2)$ to $V(n,p)$.

Exercise 6.20. * Show that if C is a linear binary $[n,k]$-code and C is e-error-correcting , then

$$\sum_{i=0}^{e} \binom{n}{i} \leq 2^{(n-k)}.$$

In particular, if C is 1-error-correcting , then $|C| \leq 2^n/(1+n)$.

Exercise 6.21. Show that if P is a permutation matrix, then right multiplication by P defines a transformation $T : V(n,p) \to V(n,p)$ which preserves the Hamming distance.

Exercise 6.22. Prove the singleton bound, Proposition 6.10. (Suggestion: try to imitate the proof of Proposition 6.9.)

Exercise 6.23. For each n and k, find an example of a maximal distance separating p-ary $[n,k]$-code.

6.4 Perfect Linear Codes

In this section, we will consider an important class of codes, called perfect codes. These are codes where the codewords are uniformly distributed throughout $V(n,p)$ in some sense. Chief among these codes are the Hamming codes, which are the linear codes with minimal distance 3 or 4 hence $e = 1$, such that every element of $V(n,p)$ is within Hamming distance 1 of a codeword (which is unique, by Proposition 6.6).

6.4.1 A Geometric Problem

The basic result Proposition 6.6 on error-correcting codes can be given a pretty geometric interpretation as follows. If $r > 0$, define the *ball of radius* r *centred at* $\mathbf{v} \in V(n,p)$ to be

$$B_r(\mathbf{v}) = \{\mathbf{w} \in V(n,p) \mid d(\mathbf{w},\mathbf{v}) \leq r\}. \tag{6.1}$$

Suppose we put $r = e = [(d(C)-1)/2]$. Then Proposition 6.6 implies

Proposition 6.11. *If a code $C \subset V(n,p)$ satisfies $d(C) = d$, then for every $\mathbf{c} \in C$,*

$$B_{d-1}(\mathbf{c}) \cap C = \{\mathbf{c}\}.$$

Furthermore, if $d \geq 3$ so that $e \geq 1$, then an element $\mathbf{v} \in V(n,p)$ which lies in one of the balls $B_e(\mathbf{c})$ lies in exactly one of them. In particular, if $\mathbf{c}, \mathbf{c}' \in C$ and $\mathbf{c} \neq \mathbf{c}'$, then $B_e(\mathbf{c}) \cap B_e(\mathbf{c}') = \emptyset$.

The union of the balls $B_e(\mathbf{c})$ as $\mathbf{c}$ varies over C is the set of elements of $V(n,p)$ which are within e of a (unique) codeword. The nicest situation is when these balls cover $V(n,p)$; that is,

$$V(n,p) = \bigcup_{\mathbf{c} \in C} B_e(\mathbf{c}). \tag{6.2}$$

Now let $C \subset V(n,p)$ be a code, but not necessarily a linear code.

Definition 6.6. A code $C \subset V(n,p)$ is said to be *perfect* if (6.2) holds. That is, C is perfect if and only if every element of $V(n,p)$ is within e of a (unique) codeword.

We will give a number of examples of binary Hamming codes, and we will show below that in fact there exist infinitely many such codes. On the other hand, it turns out that perfect linear codes such that $e > 1$ are not so abundant. The only other possibilities are a binary $[23, 12]$-code with $e = 3$ and a ternary $[11, 6]$-code with $e = 2$.

Example 6.11. Consider the binary linear code $C_3 = \{000, 111\}$. Note $d = 3$ so $e = 1$. Now

$$V(3,2) = \{000, 100, 010, 001, 110, 101, 011, 111\}.$$

The first 4 elements are within 1 of 000 and the last 4 are within 1 of 111. Therefore C is perfect, so C is a Hamming code.

We will give a more convincing example of a perfect linear code after we give a numerical criterion for perfection.

6.4.2 How to Test for Perfection

It turns out that there is a simple test that tells when a linear code is perfect. For simplicity, we'll only consider the binary case.

Proposition 6.12. *Suppose $C \subset V(n,2)$ is a linear code with $\dim C = k$ and $d \geq 3$. Then C is perfect if and only if*

$$\sum_{i=0}^{e} \binom{n}{i} = 2^{(n-k)}. \tag{6.3}$$

In particular, if $e = 1$, then C is perfect if and only if

$$(1+n)2^k = 2^n. \tag{6.4}$$

Proof. Since $|V(n,2)| = 2^n$, C is perfect if and only if

$$\sum_{\mathbf{c} \in C} |B_e(\mathbf{c})| = 2^n.$$

By an obvious count,

$$|B_e(\mathbf{v})| = \sum_{i=0}^{e} \binom{n}{i}$$

for any $\mathbf{v} \in V(n,2)$. Since $|B_e(\mathbf{c}) \cap B_e(\mathbf{c}')| = 0$ for any pair of distinct codewords $\mathbf{c}$ and $\mathbf{c}'$ and $|C| = 2^k$, we infer that C is perfect if and only if

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

This gives the result. □

Notice that $|B_e(\mathbf{c})|$ actually has nothing to do with C. The problem of finding a perfect binary code actually reduces to finding a binary $[n,k]$-code such that $|B_e(\mathbf{0})| = 2^{(n-k)}$. If $d(C) = 3$ or 4, then C is perfect if and only if $n = 2^{n-k} - 1$, where $k = \dim C$. Some possible solutions for these conditions are $n = 3$, $k = 1$ and $n = 7$, $k = 4$. The first case is settled by $C = \{000, 111\}$. The next example shows that a perfect binary code with $n = 7$ and $k = 4$ can also be realized.

Example 6.12. Consider the 7-bit code C_7 defined as the row space of

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

By enumerating the 16 elements of C_7, one sees that $d(C_7) = 3$, so $e = 1$. Since $(7+1)2^4 = 2^7$, C_7 is indeed perfect.

6.4.3 The Existence of Binary Hamming Codes

In this Section, we will show there exist infinitely many binary Hamming codes. This is done by actually constructing them. Let $C \subset V(n,2)$ be such a code. Then since $e = 1$, we know that $1 + n = 2^{n-k}$, where $k = \dim(C)$. Putting $r = n - k$, we have $n = 2^r - 1$, and $k = n - r = 2^r - r - 1$. Thus the pairs (n,k) representing the length of C and its dimension take the form $(2^r - 1, 2^r - 1 - r)$, so some values are $(3,1)$, $(7,4)$, $(15,11)$ and $(31,26)$. Hamming codes for $(3,1)$ and $(7,4)$ have already been constructed. We will now show how to obtain the rest.

Let r be an arbitrary positive integer, and, as above, put $n = 2^r - 1$ and $k = n - r$. Form the $r \times n$ matrix $A = (I_r \mid B)$, where the columns of B are the nonzero elements of $V(r,2)$ of weight at least two written as column strings and enumerated in some order. Let $C \subset V(n,2)$ consist of the strings $\mathbf{c} = c_1 c_2 \cdots c_n$ such that $A\mathbf{c}^T = \mathbf{0}$. Since A has rank r, $\dim C = n - r = 2^r - 1 - r$. Thus, by construction, $1 + n = 2^{n-k}$, so C is a Hamming code as long as $d(C) = 3$.

We can see $d(C) = 3$ as follows: since A is in reduced row echelon form, the system $A\mathbf{x} = \mathbf{0}$ has fundamental solutions of weight three. For example, since the weight two vector $(110\cdots0)^T$ is one of the columns of A, say the qth column, then the string defined by $x_1 = x_2 = x_q = 1$ and $x_i = 0$ otherwise lies in C. On the other hand, no element of C can have weight one or two. One is obvious, and two cannot happen since for any i, j between 1

and n, there is a row of A with 1 in the ith column and 0 in the jth column. That is, if $\mathbf{c}$ is an n-bit string of weight two, then $A\mathbf{c}^T \neq \mathbf{0}$.

To see the above claims more concretely, suppose $r = 3$ and

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

Then the general solution of $A\mathbf{x} = \mathbf{0}$ has the form

$$\mathbf{x} = x_4 \begin{pmatrix} 1 \\ 1 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} + x_5 \begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} + x_6 \begin{pmatrix} 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} + x_7 \begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 1 \end{pmatrix},$$

so there exist codewords of weight 3 such as 1101000. However, by a direct check, there can't be any codewords of weight 1 or 2, so $d(C) = 3$.

The code C constructed above is a special type of code known as a *dual code* . In general, if A is an arbitrary $m \times n$ matrix over $\mathbb{F}_p$ and $C \subset V(n, p)$ is the code $\mathrm{row}(A)$, then the code

$$C^{\perp} = \{\mathbf{c} \in V(n, p) \mid A\mathbf{c}^T = \mathbf{0}\}$$

is called the code *dual* to C.

Summarizing the above discussion, we have

Proposition 6.13. *Let r be any positive integer, and put $n = 2^r - 1$. Suppose $A = (I_r \mid B)$, where B is the $r \times n$ matrix over $\mathbb{F}_2$ whose columns are the binary column vectors of weight at least two enumerated in some order. Then the code C dual to $\mathrm{row}(A)$ is a binary Hamming code of length n and dimension $k = n - r = 2^r - 1 - r$.*

This Proposition shows that there exists a binary Hamming code of length $2^r - 1$ for every $r > 0$. In particular, there are infinitely many binary Hamming codes, as claimed above.

6.4.4 Perfect Codes and Cosets

We will now find a connection between perfect linear codes and cosets which will be useful when we introduce the standard decoding table . The culmination of this section will be a characterization of when a linear code C is perfect in terms of the cosets of C.

The notion of a coset of a subspace W of a vector space V was introduced in Definition 5.5.2. In fact, we noticed that every coset of W, which is by definition an equivalence class of a certain equivalence relation on V, must have the form $\mathbf{v}+W$, for some $\mathbf{v} \in V$. A coset of W therefore has a concrete geometric interpretation, namely as the translation $\mathbf{v}+W$ of W by $\mathbf{v}$. Of course, it therefore makes sense to speak of the number of elements of a coset, which is obviously $|W|$ no matter what $\mathbf{v}$ is. (In fact, $\mathbf{w} \mapsto \mathbf{v}+\mathbf{w}$ is a bijection between W and $\mathbf{v}+W$. Since we know any two cosets of the same subspace W are either equal or disjoint (by Proposition 5.26), we have proved

Proposition 6.14. *Let V be a vector space over $\mathbb{F}_p$ of dimension n, and let W be a linear subspace of V of dimension k. Then the number of cosets of W in V is exactly $p^{(n-k)}$.*

Proof. We gave one proof just before the statement of the result. Here is another proof. By Theorem 5.27, the set of cosets V/W is a vector space over $\mathbb{F}_p$ of dimension $n-k$. Thus, $|V/W| = p^{(n-k)}$, by Proposition 5.9. □

Before getting to the connection between perfect codes and cosets, let us also record a useful property of the Hamming metric. Namely, the Hamming metric is *translation invariant*. That is, for all $\mathbf{w}, \mathbf{x}, \mathbf{y} \in V(n,p)$,

$$d(\mathbf{w}+\mathbf{y}, \mathbf{x}+\mathbf{y}) = d(\mathbf{w}, \mathbf{x}).$$

Indeed,

$$d(\mathbf{w}+\mathbf{y}, \mathbf{x}+\mathbf{y}) = \omega(\mathbf{w}+\mathbf{y}-(\mathbf{x}+\mathbf{y})) = \omega(\mathbf{w}-\mathbf{x}) = d(\mathbf{w}, \mathbf{x}).$$

Now let $C \subset V(n,p)$ be a linear code with $d \geq 3$. We can use the translation invariance property to show

Proposition 6.15. *No coset of C contains more than one element of $B_e(\mathbf{0})$. In particular, $|B_e(\mathbf{0})| \leq p^{(n-k)}$, where $k = \dim C$.*

Proof. Suppose to the contrary that $B_e(\mathbf{0})$ contains two elements of $\mathbf{x}+C$, say $\mathbf{x}+\mathbf{c}$ and $\mathbf{x}+\mathbf{c}'$. Then, by translation invariance,

$$d(\mathbf{x}+\mathbf{c}, \mathbf{x}+\mathbf{c}') = d(\mathbf{c}, \mathbf{c}') \leq d(\mathbf{c}, \mathbf{0}) + d(\mathbf{c}', \mathbf{0}) \leq 2e.$$

But this is impossible, so the first statement is proved. For the second, just apply Proposition 6.14. □

Now we come to the main result.

Theorem 6.16. *A linear code $C \subset V(n,p)$ with $d \geq 3$ is perfect if and only if every coset $\mathbf{x}+C$ of C contains an element of $B_e(\mathbf{0})$.*

Proof. Suppose C is perfect, and consider a coset $\mathbf{x}+C$. By (6.2), $\mathbf{x}+C$ meets some $B_e(\mathbf{c})$, hence there exists a $\mathbf{c}' \in C$ such that $\mathbf{x}+\mathbf{c}' \in B_e(\mathbf{c})$. By translation invariance, $d(\mathbf{x}+\mathbf{c}', \mathbf{c}) = d(\mathbf{x}+(\mathbf{c}'-\mathbf{c}), \mathbf{0}) \leq e$, so $\mathbf{x}+(\mathbf{c}'-\mathbf{c}) \in B_e(\mathbf{0})$. Hence $\mathbf{x}+C$ meets $B_e(\mathbf{0})$. For the sufficiency, suppose $\mathbf{y} \in V(n,p)$. Let $\mathbf{x} = \mathbf{y}+\mathbf{c}$ be the element of $\mathbf{y}+C$ contained in $B_e(\mathbf{0})$. Then $d(\mathbf{x}, \mathbf{0}) = d(\mathbf{x}-\mathbf{c}, -\mathbf{c}) = d(\mathbf{y}, -\mathbf{c}) \leq e$, so $\mathbf{y} \in B_e(-\mathbf{c})$. Thus we have (6.2), so C is perfect. □

Of course, by the previous Proposition, the element of $\mathbf{x}+C$ in $B_e(\mathbf{0})$ is unique.

6.4.5 The hat problem

The hat problem is an example of an instance where the existence of a particular mathematical structure, in this case a Hamming code, has a surprising application. In the hat game, there are three players each wearing either a white hat or a black hat. Each player can see the hats of the other two players, but cannot their own hat is. Furthermore the players aren't allowed to communicate with each other during the game. Each player has three buttons marked B,W and A (for abstain or no guess). At the sound of the buzzer, each player presses one of their buttons. The rule is that if at least one player guesses their own hat color correctly and nobody guesses incorrectly, they share a $1,000,000 prize.

The players are allowed to formulate a strategy before the game starts. The question is how should they proceed to maximize their chances of winning. One strategy would be to have two players abstain and to have the third make a random guess. The probability of winning with this strategy is a not bad 1/2. But this strategy doesn't make any use of fact that each player can see the hats of the other two players.

Consider the following strategy: when a player sees that the other two players have the same colored hats, they guess theirs is the opposite color. If they see different colored hats, they abstain. With this strategy, the only losing hat configurations are BBB or WWW, so they win six out of eight times. Here the probability of winning is a much improved 3/4.

What does this strategy have to do with perfect codes? Recall that $C = \{000, 111\}$ is a Hamming code. If 0 represents Black and 1 represents White, then the various hat arrays are represented by the 2^3 3-bit strings. Since $e = 1$, the above strategy amounts to the following. The three contestants agree

ahead of time to assume that the hat configuration isn't in C. That is, the hats aren't all white or all black. Suppose this assumption is correct. Then two players will see a black and a white. They should automatically abstain since there is no way of telling what their own hat colors are. The third will see either two blacks or two whites. If two blacks, then by assumption, she knows her hat is white and she hits the W button. If she sees two whites, then she hits the B. This strategy fails only when all hats are the same color, i.e. the hat configuration lies in C.

Next, suppose there are 7 players instead of 3. Recall, we displayed a perfect linear binary code (namely C_7) of length 7 with 16 codewords and minimum distance 3. We continue to represent Black by 0 and White by 1, so the possible hat arrays are represented by the 2^7 7-bit strings. Let's see if the strategy for three hats works with seven hats. First, all players need to know all 16 codewords. They agree beforehand to assume the hat array isn't in C_7. This happens in 7/8 cases, since the probability that the hat array is in C_7 is $2^4/2^7 = 1/8$. Thus what they need to do is devise a winning strategy for those times their array isn't in C_7.

Suppose their assumption is correct. Let $x_1 \dots x_7$ denote their the hat array. Since C_7 is perfect with $e = 1$, they know that $x_1 \dots x_7$ differs in exactly one place from a codeword. Let's see what happens if the difference occurs in the first bit. Then $c_1 x_2 \dots x_7 \in C_7$, where $c_1 \neq x_1$. The first player sees $x_2 \dots x_7$, and, knowing $c_1 x_2 \dots x_7 \in C_7$, guesses correctly that her hat color is x_1. The second player sees $x_1 x_3 \dots x_7$ and realizes that $x_1 y_2 x_3 \dots x_7$ isn't in C for either value of y_2. Therefore, she has to pass. The other five contestants face similar situations and reason in the same way. With this strategy, their probability of winning the million bucks is 7/8.

Can you devise a strategy for how to proceed if there are 4,5 or 6 players? What about 8 or 9? More information about this problem and other related (and more serious) problems can be found in the article *The hat problem and Hamming codes* by M. Bernstein in Focus Magazine, November 2001.

6.4.6 The Standard Decoding Table

To end this section, we introduce a decoding scheme called the standard decoding table which is closely related to some of the topics of the previous Section.

Let C be a binary linear code of length n, which is being used to transmit data from a satellite. If a codeword $\mathbf{c} = c_1 \dots c_n$ is transmitted and an n-bit string $\mathbf{d} \neq \mathbf{c}$ is received, the problem is to discover the value of the error $\mathbf{e} = \mathbf{d} - \mathbf{c}$. We will now consider a scheme for doing this, known as the

nearest neighbour decoding scheme , which uses the cosets of C to organize $V(n,2)$ into a *standard decoding table* (SDT for short).

To construct the standard decoding table we write the down the cosets $\mathbf{x}+C$ of C in a specific manner as follows. In the first row of the table, list the elements of C in some order, being sure to put $\mathbf{0}$ in the first column. Say the result is $\mathbf{0}, \mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_m$, where $m = |C| - 1$. The next step is to choose an element $\mathbf{a}_1 \in V(n,p)$ which isn't in C. The second row of the table is then the coset $\mathbf{a}_1, \mathbf{a}_1 + \mathbf{c}_1, \mathbf{a}_1 + \mathbf{c}_2, \ldots, \mathbf{a}_1 + \mathbf{c}_m$. For the third row, choose another $\mathbf{a}_2$ which hasn't already appeared in the first two rows, and proceed as before. Since the cosets of C in $V(n,2)$ are pairwise disjoint and their union is $V(n,2)$, this construction leads to a table with 2^{n-k} rows and 2^k columns, where $k = \dim C$, which contains all elements of $V(n,2)$.

The final step is to impose the condition that any two elements in the same row (i.e. coset) have the same error, namely the element in the leading column.

Example 6.13. The following table is a standard decoding table for the linear code $C \subset V(4,2)$ of Example 6.3.

$$\begin{array}{cccc} 0000 & 1011 & 0101 & 1110 \\ 1000 & 0011 & 1101 & 0110 \\ 0100 & 1111 & 0001 & 1010 \\ 0010 & 1001 & 0111 & 1100 \end{array} .$$

Note that since $|C| = 4$ and $|V(4,2)| = 16$, the table is 4×4. If the transmitted codeword $\mathbf{c}$ is received for example as $\mathbf{d} = 0111$, the decoding procedure is to scan the standard decoding table until 0111 is located. Since 0111 occurs in the last row, the error is the leading element 0010 and $\mathbf{c}$ is the codeword in the first row of the table above 0111.

A typical strategy for choosing the coset representatives $\mathbf{a}_i$ is to always choose elements for the first column of minimal weight among the possible choices. This reflects the assumption that the error isn't too serious. Notice that it can happen that a row of a standard decoding table contains more than one element of minimal weight, so the choice is of representatives isn't necessarily unique. This happens in the third row of the above table, for example. This row has two elements of weight one, and there is no reason to prefer decoding 1111 as 1011 rather than 1110.

The non zero elements of least nonzero weight in $V(n,2)$ are the standard basis vectors $\mathbf{e}_1, \ldots, \mathbf{e}_n$. If $\dim C = k$, then at most k of the standard basis vectors can lie in C. Those which don't are therefore natural candidates for the leading column.

It seems desirable to seek codes C for which there is a standard decoding table such that in each row, there is a unique vector of weight one. The codes with this property are exactly the Hamming codes, since as we just saw, Hamming codes have the property that every coset contains exactly one element of $B_e(\mathbf{0})$. But $B_e(\mathbf{0})$ consists of the null word and the standard basis vectors (i.e. elements of weight one) since $e = 1$. Since the third row of the table in Example 6.13 contains two elements of weight one, C isn't perfect.

Exercises

Exercise 6.24. Consider the code $C = \{00000, 11111\} \subset V(5, 2)$.

(i) Determine e.

(ii) Show that C is perfect.

(iii) Does C present any possibilities for a five player hat game?

Exercise 6.25. Show that any binary $[2^k - 1, 2^k - 1 - k]$-code with $d = 3$ is perfect. Notice C_7 is of this type.

Exercise 6.26. Can there exist a perfect binary $[5, k, 5]$-code?

Exercise 6.27. Suppose $n \equiv 2 \; mod(4)$. Show that there cannot be a perfect binary $[n, k]$-code with $e = 2$.

Exercise 6.28. Show that any binary $[23, 12]$-code with $d = 7$ is perfect.

Exercise 6.29. Let $\mathbb{F}$ be any field, and let $M = (I_k \; A)$, where $A \in \mathbb{F}^{k\times(n-k)}$. A *parity check matrix* for M is an $n \times (n-k)$ matrix H over $\mathbb{F}$ of rank $n-k$ such that $MH = O$.

(*i*) Show that $H = \begin{pmatrix} -A \\ I_{n-k} \end{pmatrix}$ is a parity check matrix for M.

(*ii*) Conclude that a parity check matrix has the property that

$$\mathcal{N}(M) = \text{col}(H)$$

as subspaces of $\mathbb{F}^n$.

Exercise 6.30. In this Exercise, we will view $\mathbb{F}^n$ as the set of row vectors with n components in $\mathbb{F}$. Let M and H be as in Exercise 6.29, and let $C \subset \mathbb{F}^n$ denote the row space of M. Show that if $\mathbf{d}_1$ and $\mathbf{d}_2$ are in $\mathbb{F}^n$, then

$$\mathbf{d}_1 H = \mathbf{d}_2 H$$

if and only if $\mathbf{d}_1 + C = \mathbf{d}_2 + C$. Thus the parity check matrix H takes the same value $\mathbf{d}H$ on each coset $\mathbf{d} + C$ of C and different values on different cosets.

Exercise 6.31. The previous two exercises can be applied to standard decoding tables as follows. Consider the linear code $C \subset V(n, p)$ defined as the row space of $M = (I_k \; A)$, where $A \in (\mathbb{F}_p)^{k\times(n-k)}$. Define the *syndrome of the coset* $\mathbf{d} + C$ to be $\mathbf{d}H$, where $H = \begin{pmatrix} -A \\ I_{n-k} \end{pmatrix}$ is the above parity check

matrix for M. By Exercise 6.30, the syndrome of $\mathbf{d} + C$ is well defined. By the definition of a standard decoding table , this means that two elements of $V(n, p)$ are on the same row if and only if they have the same syndrome. A *standard decoding table with syndromes* for C is a standard decoding table for C with an extra column which gives the syndromes of the rows. The advantage of knowing the syndromes of the cosets is that to find which row $\mathbf{d} \in V(n, p)$ is in, one just scans the last column to find the row containing $\mathbf{d}H$. Construct the standard decoding table with syndromes for the linear code of Example 6.3 starting from the standard decoding table of Example 6.13.

Exercise 6.32. Construct the parity check matrix and syndromes for C_7.

Exercise 6.33. Construct the standard decoding table with syndromes for the binary code C with generating matrix

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

(Note, this is rather complicated since C has 8 elements. Thus the standard decoding table with syndromes is an 8×9 table (counting the syndromes as one column) since $V(6, 2)$ has 64 elements.)

Exercise 6.34. Construct the standard decoding table with syndromes for the binary code with generating matrix

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

Exercise 6.35. Let C be a binary linear n-bit code with $n \geq 3$ with parity check matrix H. Show that if no two columns of H are dependent (i.e. equal), then $d(C) \geq 3$.

Exercise 6.36. Generalize Exercise 6.35 by showing that if $C \subset V(n, p)$ is a linear code with a parity check matrix H having the property that no m columns of H are linearly dependent, then $d(C) \geq m + 1$.

Exercise 6.37. Show that any coset of the linear code which contains the ISBNs contains a unique standard basis vector. In particular, any element of $V(10, 11)$ differs form an element of C in just one component.

Exercise 6.38. Find an upper bound on the number of operations required to scan the standard decoding table with syndromes associated to an *p-ary* $[n, k]$ -code to find any $\mathbf{d} \in V(n, p)$, and compare this result to the number of operations needed to find $\mathbf{d}$ before adding the column of syndromes.

6.5 Summary

This chapter gives a brief introodution to the subject of coding theory, which is a branch of mathematics with close ties to computer science and electrical engineering. A code is a subset C of some $V(n,p)$, and a linear code is a linear subspace. The elements of the code are the codewords. Linear coding theory provides an excellent (not to mention beautiful) realization of many of the topics we've already studied in elementary linear algebra: row operations, dimension, bases, cosets etc. The key concept is the notion of the Hamming distance, which is a natural metric on $V(n,p)$. The Hamming distance between two elements of $V(n,p)$ is the number of components where the two elements differ. The minimal distance $d(C)$ of a code C is the least distance between any two codewords. The main goal in coding theory is to devise codes C where $d(C)$ is large. In fact, the first basic result on error-correcting codes says that an element $\mathbf{x}$ of $V(n,p)$ can have distance $d(\mathbf{x},\mathbf{c}) \leq e = (d(C)-1)/2$ from at most one codeword $\mathbf{c}$.

A linear code such that every element of $V(n,p)$ is within e of a codeword is called perfect. A perfect code with $e = 1$ is called a Hamming code. Hamming codes have the property that every element of $V(n,p)$ is either a codeword or one unit from a codeword. We show there are infinitely many Hamming codes. (Surprisingly, there are only two examples of perfect codes where the minimum distance is greater than 4.) We show that the question of whether or not a linear code is perfect comes down to solving a combinatorial condition (6.3), and we also give an amusing (and surprising) application of Hamming codes to the question of whether you know what color the hat you're wearing is.

We also treat some other topics. Cosets come up in the idea of the nearest neighbor decoding scheme. We introduce a class of matrices, called Hadamard matrices, which enable one to construct codes (non linear codes unfortunately) such that $d(C)$ is very large compared. These codes were used for communication with the Mariner space probes during the 1960s and 70s.

For the reader who wishes to pursue coding theory more deeply, there are several elementary texts, such as *Introduction to Coding Theory* by R. Hill and *Introduction to the Theory of Error-Correcting Codes* by V. Pless. The more advanced book, *Applied Abstract Algebra* by R. Lidl and G. Pilz, gives many interesting applications of linear algebra besides coding theory. The web is also an excellent source of information.

Chapter 7

Linear Transformations

The purpose of this Chapter is to introduce linear transformations, a way of moving from one vector space to another. In particular, the linear transformations between two vector spaces V and W (over the same field) themselves form a vector space $L(V, W)$. If V and W are finite dimensional vector spaces, then a linear transformation is given by a matrix, so the theory of linear transformations is part of matrix theory. We will also study the geometric properties of linear transformations.

7.1 Definitions and Examples

7.1.1 Mappings

Recall that a mapping form a set X called the domain to a set Y called the target is a rule which assigns to each element of X a unique element $F(x) \in Y$. Such a mapping is denoted as $F : X \to Y$. The *range* of F is the set $F(X) = \{y \in Y \mid y = F(x) \ \exists \ x \in X\}$.

When the domain and target of a mapping F are vector spaces, the mapping is often called a *transformation*. A transformation $F : \mathbb{F}^n \to \mathbb{F}^m$ is completely determined by expressing $F(\mathbf{x})$ in terms of components:

$$F(\mathbf{x}) = \begin{pmatrix} f_1(\mathbf{x}) \\ \vdots \\ f_m(\mathbf{x}) \end{pmatrix} = \sum_{i=1}^{m} f_i(\mathbf{x})\mathbf{e}_i, \tag{7.1}$$

where $\mathbf{e}_1, \ldots, \mathbf{e}_m$ is the standard basis of $\mathbb{F}^m$. The functions $f_1, \ldots, f_m$ are called the *components* of F with respect to the standard basis. If W is a

finite dimensional vector space , we can define the components $f_1, \ldots, f_m$ of F with respect to a given basis $\mathbf{w}_1, \ldots, \mathbf{w}_m$ of W by writing

$$F(\mathbf{v}) = \sum_{i=1}^{m} f_i(\mathbf{v})\mathbf{w}_i.$$

The component functions are uniquely determined by F and this choice of basis.

7.1.2 The Definition of a Linear Transformation

From the linear algebraic viewpoint, the most important transformations are those which preserve linear combinations. These are either called *linear transformations* or *linear mappings*.

Definition 7.1. Suppose V and W are vector spaces over a field $\mathbb{F}$. Then a transformation $T : V \to W$ is said to be *linear* if

(1) $T(\mathbf{x} + \mathbf{y}) = T(\mathbf{x}) + T(\mathbf{y})$ for all $\mathbf{x},\ \mathbf{y} \in V$, and

(2) $T(r\mathbf{x}) = rT(\mathbf{x})$ for all $r \in \mathbb{F}$ and all $\mathbf{x} \in V$.

Clearly, a linear transformation T preserves linear combinations:

$$T(r\mathbf{x} + s\mathbf{y}) = rT(\mathbf{x}) + sT(\mathbf{y}),$$

for all $r, s \in \mathbb{F}$ and all $\mathbf{x}, \mathbf{y} \in V$.

Conversely, any transformation that preserves linear combinations is a linear transformation. Another obvious property is that any linear transformation maps $\mathbf{0}$ to $\mathbf{0}$: $T(\mathbf{0}) = \mathbf{0}$. This follows, for example, from the fact that

$$T(\mathbf{x}) = T(\mathbf{x} + \mathbf{0}) = T(\mathbf{x}) + T(\mathbf{0})$$

for any $\mathbf{x} \in V$, which can only happen if $T(\mathbf{0}) = \mathbf{0}$.

7.1.3 Some Examples

Example 7.1 (The identity transformation)**.** The transformation

$$Id : V \to V$$

defined by $Id(\mathbf{x}) = \mathbf{x}$ is called the *identity transformation*. This transformation is clearly linear.

Example 7.2. If $\mathbf{a} \in \mathbb{R}^n$, the dot product with $\mathbf{a}$ defines a linear transformation $T_{\mathbf{a}} : \mathbb{R}^n \to \mathbb{R}$ by $T_{\mathbf{a}}(\mathbf{x}) = \mathbf{a} \cdot \mathbf{x}$. It turns out that any linear transformation $T : \mathbb{R}^n \to \mathbb{R}$ has the form $T_{\mathbf{a}}$ for some $\mathbf{a} \in \mathbb{R}^n$.

Example 7.3. Consider the linear mapping $T : \mathbb{R}^2 \to \mathbb{R}^2$ given by

$$T\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \lambda x \\ \mu y \end{pmatrix},$$

where λ and μ are scalars. Since $T(\mathbf{e}_1) = \lambda \mathbf{e}_1$ and $T(\mathbf{e}_2) = \mu \mathbf{e}_2$, it follows that if both λ and μ are nonzero, then T maps a rectangle with sides parallel to $\mathbf{e}_1$ and $\mathbf{e}_2$ onto another such rectangle whose sides have been dilated by λ and μ and whose area has been changed by the factor $|\lambda\mu|$. If $|\lambda| \neq |\mu|$, then T maps circles to ellipses. For example, let C denote the unit circle $x^2 + y^2 = 1$, and put $w = \lambda x$ and $z = \mu y$. Then the image of T is the ellipse

$$(\frac{w}{\lambda})^2 + (\frac{z}{\mu})^2 = 1.$$

In general, a linear transformation $T : V \to V$ is called *semi-simple* if there exists a basis $\mathbf{v}_1, \ldots, \mathbf{v}_n$ of V such that $T(\mathbf{v}_i) = \lambda_i \mathbf{v}_i$ for each index i, for where each $\lambda_i \in \mathbb{F}$. It turns out that one of the main problems in the theory of linear transformations is how to determine when a linear transformation is semi-simple. A nonzero vector $\mathbf{v}$ such that $T(\mathbf{v}) = \lambda \mathbf{v}$ is called an *eigenvector* of T and λ is called the corresponding *eigenvalue*. The pair $(\lambda, \mathbf{v})$ is called an *eigenpair* for T. Eigenvalues and eigenvectors are the basis of the subject of eigentheory, which will be taken up in Chapter 9.

Example 7.4. The cross product gives a pretty example of a linear transformation on $\mathbb{R}^3$. Let $\mathbf{a} \in \mathbb{R}^3$ and define $C_{\mathbf{a}} : \mathbb{R}^3 \to \mathbb{R}^3$ by

$$C_{\mathbf{a}}(\mathbf{v}) = \mathbf{a} \times \mathbf{v}.$$

Notice that $C_{\mathbf{a}}(\mathbf{a}) = \mathbf{0}$, and that $C_{\mathbf{a}}(\mathbf{x})$ is orthogonal to $\mathbf{a}$ for any $\mathbf{x}$. The transformation $C_{\mathbf{a}}$ is used in mechanics to express angular momentum.

Example 7.5. Suppose $V = C[a, b]$, the space of continuous real valued functions on the closed interval $[a, b]$. The definite integral over $[a, b]$ defines a linear transformation

$$\int_a^b : V \to \mathbb{R}$$

by the rule $f \mapsto \int_a^b f(t)\mathrm{d}t$. The assertion that $\int_a^b$ is a linear transformation is just the fact that for all $r, s \in \mathbb{R}$ and $f, g \in V$,

$$\int_a^b (rf + sg)(t)\mathrm{d}t = r \int_a^b f(t)\mathrm{d}t + s \int_a^b g(t)\mathrm{d}t.$$

This example is the analogue for $C[a,b]$ of the linear transformation $T_{\mathbf{a}}$ on $\mathbb{R}^n$ defined in Example 7.2 above, where $\mathbf{a}$ is taken to be the constant function 1. For, by definition,

$$\int_a^b f(t)\mathrm{d}t = (f,1).$$

Example 7.6. Let V be a vector space over $\mathbb{F}$, and let W be a subspace of V. Let V/W be the quotient of V by W introduced in Chapter 5. Let $\pi : V \to V/W$ be the *quotient map* defined by

$$\pi(\mathbf{v}) = \mathbf{v} + W.$$

Then π is a linear map. We leave the details of this as an exercise.

7.1.4 General Properties of Linear Transformations

Suppose V is a finite dimensional vector space over $\mathbb{F}$, and W is another (not necessarily finite dimensional) vector space over $\mathbb{F}$. What we will now prove is that given a basis of V, there exists a unique linear transformation $T : V \to W$ taking whatever values we wish on the given basis of V, and, furthermore, its values on the basis of V uniquely determine the linear transformation. This is a property of linear transformations which we will use throughout this Chapter.

Proposition 7.1. *Let V and W be vector spaces over $\mathbb{F}$. Then every linear transformation $T : V \to W$ is uniquely determined by its values on a basis of V. Moreover, if $\mathbf{v}_1, \ldots, \mathbf{v}_n$ is a basis of V and $\mathbf{w}_1, \ldots, \mathbf{w}_n$ are arbitrary vectors in W, then there exists a unique linear transformation $T : V \to W$ such that $T(\mathbf{v}_i) = \mathbf{w}_i$ for each i. In other words, there is a unique linear transformation with any given values on a basis.*

Proof. The proof is surprisingly simple. Suppose $\mathbf{v}_1, \ldots, \mathbf{v}_n$ is a basis of V and $T : V \to W$ is linear. Since every $\mathbf{v} \in V$ has a unique expression

$$\mathbf{v} = \sum_{i=1}^{n} r_i \mathbf{v}_i,$$

where $r_1, \ldots r_n \in \mathbb{F}$, then

$$T(\mathbf{v}) = \sum_{i=1}^{n} r_i T(\mathbf{v}_i).$$

Thus T is uniquely determined by its values on $\mathbf{v}_1, \ldots, \mathbf{v}_n$.

Now suppose $\mathbf{w}_1, \ldots, \mathbf{w}_n$ are arbitrary vectors in W. We want to show there exists a unique linear transformation $T : V \to W$ such that $T(\mathbf{v}_i) = \mathbf{w}_i$ for each i. If $\mathbf{v} = \sum_{i=1}^n r_i \mathbf{v}_i$, put

$$T(\mathbf{v}) = \sum_{i=1}^{n} r_i T(\mathbf{v}_i).$$

This certainly defines a unique transformation, so all we have to do is show that T is linear. Let $\mathbf{v} = \sum \alpha_i \mathbf{v}_i$ and $\mathbf{w} = \sum \beta_i \mathbf{v}_i$. Then

$$\mathbf{v} + \mathbf{w} = \sum (\alpha_i + \beta_i) \mathbf{v}_i,$$

so, by definition,

$$T(\mathbf{v} + \mathbf{w}) = \sum (\alpha_i + \beta_i) T(\mathbf{v}_i) = \sum \alpha_i T(\mathbf{v}_i) + \sum \beta_i T(\mathbf{v}_i).$$

Hence $T(\mathbf{v} + \mathbf{w}) = T(\mathbf{v}) + T(\mathbf{w})$. Similarly, $T(r\mathbf{v}) = rT(\mathbf{v})$. Hence T is linear, so the proof is finished. □

Transformations $F : V \to W$ can be added using pointwise addition, and they can be multiplied by scalars in a similar way. That is, if $F, G : V \to W$ are any two transformations, we form their sum $F + G$ 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. Thus, we can take arbitrary linear combinations of transformations.

Proposition 7.2. *Let V and W be vector spaces over $\mathbb{F}$. Then any linear combination of linear transformations with domain V and target W is also linear. Thus the set $L(V, W)$ of all linear transformations $T : V \to W$ is a vector space over $\mathbb{F}$.*

The dimension of $L(V, W)$ can easily be deduced from Proposition 7.1.

Proposition 7.3. *Suppose* $\dim V = n$ *and* $\dim W = m$. *Then $L(V, W)$ has dimension mn.*

Proof. We will define a basis of $L(V, W)$, and leave the proof that it is a basis as an exercise. Choose any bases $\mathbf{v}_1, \ldots, \mathbf{v}_n$ of V and $\mathbf{y}_1, \ldots, \mathbf{y}_m$ of W. By Proposition 7.1, if $1 \leq i \leq n$ and $1 \leq j \leq m$, there exists a unique linear transformation $T_{ij} : V \to W$ defined by the condition that $T_{ij}(\mathbf{v}_j) = \mathbf{w}_i$. I claim that the mn linear transformations T_{ij} are a basis of $L(V, W)$. □

Exercises

Exercise 7.1. Show that every linear function $T : \mathbb{R} \to \mathbb{R}$ has the form $T(x) = ax$ for some $a \in \mathbb{R}$.

Exercise 7.2. Determine whether the following are linear or not:

(i) $f(x_1, x_2) = x^2 - x_2$.

(ii) $g(x_1, x_2) = x_1 - x_2$.

(iii) $f(x) = e^x$.

Exercise 7.3. Suppose $T : \mathbb{F}^n \to \mathbb{F}^m$ is an arbitrary transformation and write

$$T(\mathbf{v}) = \begin{pmatrix} f_1(\mathbf{v}) \\ f_2(\mathbf{v}) \\ \vdots \\ f_m(\mathbf{v}) \end{pmatrix}.$$

Show that T is linear if and only if each component f_i is a linear function. In particular, T is linear if and only if there exist $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_m$ in $\mathbb{F}^n$ such that for $f_i(\mathbf{v}) = \mathbf{a}_i \cdot \mathbf{v}$ for all $\mathbf{v} \in \mathbb{F}^n$.

Exercise 7.4. Suppose $T : V \to V$ is linear. Show that if $T(\mathbf{v}) = \mathbf{0}$ for some $\mathbf{v} \neq \mathbf{0}$, then 0 is an eigenvalue and $(\mathbf{v}, 0)$ is an eigenpair for T.

Exercise 7.5. Suppose $S, T : V \to W$ are two linear transformations. Show that if $\mathbf{v}_1, \ldots, \mathbf{v}_n$ is a basis of V and $S(\mathbf{v}_i) = T(\mathbf{v}_i)$ for all $i = 1, \ldots, n$, then $S = T$. That is, $S(\mathbf{v}) = T(\mathbf{v})$ for all $\mathbf{v} \in V$.

Exercise 7.6. Show that the cross product linear transformation $C_{\mathbf{a}} : \mathbb{R}^3 \to \mathbb{R}^3$ defined in Example 7.4 has 0 as an eigenvalue. Is this the only eigenvalue?

Exercise 7.7. What are the eigenvalues of the identity transformation?

Exercise 7.8. 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 7.9. Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be a linear map with matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$. 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 $T(\alpha\beta) = \alpha T(\beta)$ for all $\alpha, \beta \in \mathbb{C}$. Show that a necessary and sufficient condition is that $a = d$ and $b = -c$.

Exercise 7.10. Finish the proof of Proposition 7.3 by showing the T_{ij} are indeed a basis of $L(V, W)$.

Exercise 7.11. Let X and Y be sets and $\phi : X \to Y$ a mapping. Recall from Definition 4.5 that ϕ is injective if and only if for $x \in X$, $\phi(x) = \phi(x')$ implies $x = x'$, and ϕ is surjective if and only if $\phi(X) = Y$. 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 $id : X \to X$ defined by $id(x) = x$ for all $x \in X$.

(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 $id : Y \to Y$

(iii) Conclude that ϕ is a bijection if and only if there exists a mapping $\psi : Y \to X$ such that $\phi \circ \psi$ is the identity mapping $id : Y \to Y$ and $\psi \circ \phi$ is the identity mapping $id : X \to X$.

Exercise 7.12. Suppose that $T : V \to W$ is a linear transformation. Show that T is injective if and only if $T(\mathbf{x}) = \mathbf{0}$ only if $\mathbf{x} = \mathbf{0}$.

Exercise 7.13. Let $\mathbb{F}$ be a finite field of characteristic p. Let $T : \mathbb{F} \to \mathbb{F}$ be the transformation defined by $T(x) = x^p$. Recall that the set of multiples of $m1$ in $\mathbb{F}$, where $m - 0, 1, \ldots, p - 1$ form a subfield $\mathbb{F}' = \mathbb{F}_p$ of $\mathbb{F}$ and that $\mathbb{F}$ is a vector space over this subfield. Show that T is a linear transformation of $\mathbb{F}$ with respect to this vector space structure. The transformation T is called the *Frobenius map*.

7.2 Linear Transformations on $\mathbb{F}_n$ and Matrices

The purpose of this section is to make the connection between matrix theory and linear transformations on $\mathbb{F}^n$. In fact that they are every $m \times n$ matrix on $\mathbb{F}$ determines a linear transformation and conversely. This gives a powerful method for studying linear transformations. Moreover, looking at a matrix as a linear transformation gives us a natural explanation for the definition of matrix multiplication. It turns out that matrix multiplication is just the operation of composing two linear transformations.

7.2.1 Matrix Linear Transformations

Our first observation is

Proposition 7.4. *Suppose $A \in \mathbb{F}^{m\times n}$. Then A defines a linear transformation $T_A : \mathbb{F}^n \to \mathbb{F}^m$ by the rule $T_A(\mathbf{x}) = A\mathbf{x}$. If we express A in terms of its columns as $A = (\mathbf{a}_1\ \mathbf{a}_2\ \cdots\ \mathbf{a}_n)$, then*

$$T_A(\mathbf{x}) = A\mathbf{x} = \sum_{i=1}^{n} x_i\mathbf{a}_i.$$

Hence the value of T_A at $\mathbf{x}$ is a linear combination of the columns of A.

Proof. The fact that T_A is linear follows immediately from the distributive and scalar multiplication laws for matrix multiplication. □

The converse of the above Proposition is also true. All linear transformations from $\mathbb{F}^n$ to $\mathbb{F}^m$ come from matrices; that is, they are all matrix linear transformations.

Proposition 7.5. *Every linear transformation $T : \mathbb{F}^n \to \mathbb{F}^m$ is of the form T_A for a unique $m \times n$ matrix A. The ith column of A is $T(\mathbf{e}_i)$, where $\mathbf{e}_i$ is the ith standard basis vector, i.e. the ith column of I_n.*

Proof. Let $\mathbf{a}_i = T(\mathbf{e}_i)$ for all indices i. The point is that any $\mathbf{x} \in \mathbb{F}^n$ has the unique expansion

$$\mathbf{x} = \sum_{i=1}^{n} x_i\mathbf{e}_i,$$

so,

$$T(\mathbf{x}) = T\big(\sum_{i=1}^{n} x_i\mathbf{e}_i\big) = \sum_{i=1}^{n} x_iT(\mathbf{e}_i) = \sum_{i=1}^{n} x_i\mathbf{a}_i.$$

Thus,

$$T(\mathbf{x}) = A\mathbf{x},$$

where $A = (\mathbf{a}_1 \ \cdots \ \mathbf{a}_n)$. To see that A is unique, suppose A and B are in $\mathbb{F}^{m\times n}$ and $A \neq B$. Then $A\mathbf{e}_i \neq B\mathbf{e}_i$ for some i, so $T_A(\mathbf{e}_i) \neq T_B(\mathbf{e}_i)$. Hence different matrices define different linear transformations, so the proof is complete. □

Example 7.7. For example, the matrix of the identity transformation $Id : \mathbb{F}^n \to \mathbb{F}^n$ is the identity matrix I_n. That is $Id = T_{I_n}$.

A linear transformation $T : \mathbb{F}^n \to \mathbb{F}$ is called a *linear function* . If $a \in \mathbb{F}$, then the function $T_a(x) := ax$ is a linear function $T : \mathbb{F} \to \mathbb{F}$. More generally, we have

Proposition 7.6. *Every linear function $T : \mathbb{F}^n \to \mathbb{F}$ has the form $T(\mathbf{x}) = \mathbf{a}^T\mathbf{x}$ for some $\mathbf{a} \in \mathbb{F}^n$. That is, there exist $a_1, a_2, \ldots, a_n \in \mathbb{F}$ such that*

$$T\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \sum_{i=1}^{n} a_i x_i.$$

Proof. Just set $a_i = T(\mathbf{e}_i)$. □

Example 7.8. Let $\mathbf{a} = (1, 2, 0, 1)^T$. The associated linear function $T : \mathbb{F}^4 \to \mathbb{F}$ has the explicit form

$$T\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = x_1 + 2x_2 + x_4.$$

Finally we connect linear functions and linear transformations.

Proposition 7.7. *A transformation $T : \mathbb{F}^n \to \mathbb{F}^m$*

$$T(\mathbf{x}) = \begin{pmatrix} f_1(\mathbf{x}) \\ \vdots \\ f_m(\mathbf{x}) \end{pmatrix} = \sum_{i=1}^{m} f_i(\mathbf{x})\mathbf{e}_i,$$

is linear if and only if each of the component functions $f_1, \ldots, f_m$ is a linear function.

Proof. This is left to the reader. □

7.2.2 Composition and Multiplication

So far, matrix multiplication has just been a convenient tool without a natural interpretation. We'll now provide one. For this, we need to consider the composition of two transformations. Suppose $S : \mathbb{F}^p \to \mathbb{F}^n$ and $T : \mathbb{F}^n \to \mathbb{F}^m$ are any two transformations. Since the target of S is the domain of T, the operation of composition can be defined. The *composition* of S and T is the transformation $T \circ S : \mathbb{F}^p \to \mathbb{F}^m$ defined by

$$T \circ S(\mathbf{x}) = T\big(S(\mathbf{x})\big).$$

The following Proposition describes the composition if S and T are linear.

Proposition 7.8. *Suppose $S : \mathbb{F}^p \to \mathbb{F}^n$ and $T : \mathbb{F}^n \to \mathbb{F}^m$ are linear transformations with matrices A and B respectively. That is, $S = T_A$ and $T = T_B$. Then the composition $T \circ S : \mathbb{F}^p \to \mathbb{F}^m$ is also linear, and the matrix of $T \circ S$ is BA. In other words,*

$$T \circ S = T_B \circ T_A = T_{BA}.$$

Letting $M(S) \in \mathbb{F}^{n\times p}$ and $M(T) \in \mathbb{F}^{m\times n}$ denote the matrix of T, we therefore have

$$M(T \circ S) = M(T)M(S)$$

in $\mathbb{F}^{m\times p}$.

Proof. To prove $T \circ S$ is linear, note that

$$\begin{aligned} T \circ S(r\mathbf{x} + s\mathbf{y}) &= T\big(S(r\mathbf{x} + s\mathbf{y})\big) \\ &= T\big(rS(\mathbf{x}) + sS(\mathbf{y})\big) \\ &= rT\big(S(\mathbf{x})\big) + sT\big(S(\mathbf{y})\big). \end{aligned}$$

Thus, $T \circ S(r\mathbf{x} + s\mathbf{y}) = rT \circ S(\mathbf{x}) + sT \circ S(\mathbf{y})$, so $T \circ S$ is linear as claimed. To find the matrix of $T \circ S$, we observe that

$$T \circ S(\mathbf{x}) = T(A\mathbf{x}) = B(A\mathbf{x}) = (BA)\mathbf{x},$$

by the associativity of multiplication. This implies that the matrix of $T \circ S$ is the product BA, as asserted. The rest of the proof now follows easily. □

Note that the associativity of matrix multiplication is a key fact. In fact, we could have used that $T \circ S(\mathbf{x}) = (BA)\mathbf{x}$ to conclude that $T \circ S$ is linear, so the first part of the proof was actually unnecessary.

7.2.3 An Example: Rotations of $\mathbb{R}^2$

A nice way of illustrating the connection between matrix multiplication and composition is with rotations of the plane. Let $\mathcal{R}_\theta : \mathbb{R}^2 \to \mathbb{R}^2$ stand for the counter-clockwise rotation of $\mathbb{R}^2$ 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 non-collinear vectors in $\mathbb{R}^2$, and let P be the parallelogram they span. Then $\mathcal{R}_\theta$ rotates the whole parallelogram P about the origin $\mathbf{0}$ to a new parallelogram $\mathcal{R}_\theta(P)$. The edges of $\mathcal{R}_\theta(P)$ at $\mathbf{0}$ are $\mathcal{R}_\theta(\mathbf{x})$ and $\mathcal{R}_\theta(\mathbf{y})$. The diagonal $\mathbf{x}+\mathbf{y}$ of P is rotated to the diagonal of $\mathcal{R}_\theta(P)$, which is $\mathcal{R}_\theta(\mathbf{x}) + \mathcal{R}_\theta(\mathbf{y})$. Thus

$$\mathcal{R}_\theta(\mathbf{x}+\mathbf{y}) = \mathcal{R}_\theta(\mathbf{x}) + \mathcal{R}_\theta(\mathbf{y}).$$

Similarly, for any scalar r,

$$\mathcal{R}_\theta(r\mathbf{x}) = r\mathcal{R}_\theta(\mathbf{x}).$$

Therefore $\mathcal{R}_\theta$ is linear, as claimed. Let R_θ denote the matrix of $\mathcal{R}_\theta$. Then $R_\theta = (\mathcal{R}_\theta(\mathbf{e}_1) \;\; \mathcal{R}_\theta(\mathbf{e}_2))$, so

$$R_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}. \tag{7.2}$$

Thus

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

Notice that the matrix (7.2) of a rotation is an orthogonal matrix (see Section 3.3.3).

Consider the composition of two rotations. The rotation $\mathcal{R}_\psi$ followed that by $\mathcal{R}_\theta$ is clearly the rotation $\mathcal{R}_{\theta+\psi}$. That is,

$$\mathcal{R}_{\theta+\psi} = \mathcal{R}_\theta \circ \mathcal{R}_\psi.$$

Therefore, by Proposition 7.8, we see that $R_{\theta+\psi} = R_\theta R_\psi$. Hence

$$\begin{pmatrix} \cos(\theta+\psi) & -\sin(\theta+\psi) \\ \sin(\theta+\psi) & \cos(\theta+\psi) \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} \cos\psi & -\sin\psi \\ \sin\psi & \cos\psi \end{pmatrix}.$$

Expanding the product gives the angle sum formulas for $\cos(\theta + \psi)$ and $\sin(\theta + \psi)$. Namely,

$$\cos(\theta + \psi) = \cos\theta\cos\psi - \sin\theta\sin\psi,$$

and

$$\sin(\theta + \psi) = \sin\theta\cos\psi + \cos\theta\sin\psi.$$

This is probably the simplest proof of these basic formulas.

Exercises

Exercise 7.14. Find the matrix of the following transformations:

(*i*) $F(x_1, x_2, x_3)^T = (2x_1 - 3x_3, x_1 + x_2 - x_3, x_1, x_2 - x_3)^T$.

(*ii*) $G(x_1, x_2, x_3, x_4)^T = (x_1 - x_2 + x_3 + x_4, x_2 + 2x_3 - 3x_4)^T$.

(*iii*) The matrix of $G \circ F$.

Exercise 7.15. Find the matrix of

(*i*) The rotation $R_{-\pi/4}$ of $\mathbb{R}^2$ through $-\pi/4$.

(*ii*) The reflection H of $\mathbb{R}^2$ through the line $x = y$.

(*iii*) The matrices of $H \circ R_{-\pi/4}$ and $R_{-\pi/4} \circ H$, where H is the reflection of part (ii).

(*iv*) The rotation of $\mathbb{R}^3$ through $\pi/3$ about the z-axis.

Exercise 7.16. Show that every rotation $\mathcal{R}_\theta$ also defines a $\mathbb{C}$-linear map $\mathcal{R}_\theta : \mathbb{C} \to \mathbb{C}$. Describe this map in terms of the complex exponential.

Exercise 7.17. Let $V = \mathbb{C}$, and consider the transformation $R : V \to V$ defined by $R(z) = \alpha z$, where α is a fixed constant in $\mathbb{C}$. Write out what R is as a linear transformation $R : \mathbb{R}^2 \to \mathbb{R}^2$.

Exercise 7.18. Find the matrix of the cross product transformation $C_{\mathbf{a}} : \mathbb{R}^3 \to \mathbb{R}^3$ with respect to the standard basis in the following cases:

(*i*) $\mathbf{a} = \mathbf{e}_1$, and

(*ii*) $\mathbf{a} = \mathbf{e}_1 + \mathbf{e}_2 + \mathbf{e}_3$.

Exercise 7.19. Prove Proposition 7.7.

Exercise 7.20. Suppose $T : \mathbb{F}^n \to \mathbb{F}^n$ is linear. Give a necessary and sufficient condition for the inverse transformation T^{-1} exist?

7.3 Geometry of Linear Transformations on $\mathbb{R}^n$

As illustrated by the example of rotations, linear transformations $T : \mathbb{R}^n \to \mathbb{R}^n$ have rich geometric properties. In this section we will illustrate some of these geometric aspects.

7.3.1 Transformations of the Plane

We know that a linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$ is determined by $T(\mathbf{e}_1)$ and $T(\mathbf{e}_2)$, and so if $T(\mathbf{e}_1)$ and $T(\mathbf{e}_2)$ are non-collinear, then T sends each one of the coordinate axes $\mathbb{R}\mathbf{e}_i$ to the line $\mathbb{R}T(\mathbf{e}_i)$. Furthermore, T transforms the square $\mathbf{S}$ spanned by $\mathbf{e}_1$ and $\mathbf{e}_2$ onto the parallelogram $\mathbf{P}$ with edges $T(\mathbf{e}_1)$ and $T(\mathbf{e}_2)$. Indeed,

$$\mathbf{P} = \{rT(\mathbf{e}_1) + sT(\mathbf{e}_2) \mid 0 \leq r, s \leq 1\},$$

and since $T(r\mathbf{e}_1 + s\mathbf{e}_2) = rT(\mathbf{e}_1) + sT(\mathbf{e}_2)$, $T(\mathbf{S}) = \mathbf{P}$. More generally, T sends the parallelogram with sides $\mathbf{x}$ and $\mathbf{y}$ to the parallelogram with sides $T(\mathbf{x})$ and $T(\mathbf{y})$.

We now consider a case where $T(\mathbf{e}_1)$ and $T(\mathbf{e}_2)$ are collinear.

Example 7.9 (Projections)**.** The purpose of this Example is to introduce projections.The general theory of projections is developed in Chapter 10, so we won't include all the proofs. Let $\mathbf{a} \in \mathbb{R}^2$ be non-zero. The transformation

$$P_{\mathbf{a}}(\mathbf{x}) = \big(\frac{\mathbf{a} \cdot \mathbf{x}}{\mathbf{a} \cdot \mathbf{a}}\big)\mathbf{a}$$

is called the *projection* on the line $\mathbb{R}\mathbf{a}$ spanned by $\mathbf{a}$. Clearly the projection $P_{\mathbf{a}}$ is linear. One checks easily that

$$\mathbf{x} = P_{\mathbf{a}}(\mathbf{x}) + (\mathbf{x} - P_{\mathbf{a}}(\mathbf{x})) \tag{7.3}$$

is the *orthogonal decomposition of* of $\mathbf{x}$ into the sum of a component parallel to $\mathbf{a}$ and a component orthogonal to $\mathbf{a}$.

By the formula for $P_{\mathbf{a}}$, we get the explicit expression

$$P_{\mathbf{a}} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} \big(\frac{a_1x_1 + a_2x_2}{a_1^2 + a_2^2}\big)a_1 \\ \big(\frac{a_1x_1 + a_2x_2}{a_1^2 + a_2^2}\big)a_2 \end{pmatrix},$$

where $\mathbf{a} = (a_1, a_2)^T$ and $\mathbf{x} = (x_1, x_2)^T$. Thus the matrix of $P_{\mathbf{a}}$ is

$$\frac{1}{a_1^2 + a_2^2} \begin{pmatrix} a_1^2 & a_1a_2 \\ a_1a_2 & a_2^2 \end{pmatrix}.$$

Of course, projections don't send parallelograms to parallelograms, since any two values $P_{\mathbf{a}}(\mathbf{x})$ and $P_{\mathbf{a}}(\mathbf{y})$ are collinear.

Notice that each vector on the line spanned by $\mathbf{a}$ is preserved by $P_{\mathbf{a}}$, and every vector orthogonal to $\mathbf{a}$ is mapped by $P_{\mathbf{a}}$ to $\mathbf{0}$. It follows that $\mathbf{a}$ is an eigenvector with eigenvalue one and any vector orthogonal to $\mathbf{a}$ is an eigenvector with eigenvalue zero.

7.3.2 Orthogonal Transformations

A transformation $T : \mathbb{R}^n \to \mathbb{R}^n$ is said to be an *orthogonal transformation* if and only if

$$T(\mathbf{x}) \cdot T(\mathbf{y}) = \mathbf{x} \cdot \mathbf{y} \tag{7.4}$$

for all $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{R}^n$. By elementary properties of the dot product, it follows that orthogonal transformations are characterized by the fact that they preserve angles and lengths. Hence this type of transformation is closely related with Euclidean geometry. Let us show

Proposition 7.9. *Every orthogonal transformation $T : \mathbb{R}^n \to \mathbb{R}^n$ is linear. Moreover, the matrix of an orthogonal transformation is an orthogonal matrix. Conversely, an orthogonal matrix defines an orthogonal transformation.*

Proof. To show that an orthogonal transformation T is linear, we must show $T(\mathbf{x}+\mathbf{y}) = T(\mathbf{x})+T(\mathbf{y})$ and $T(r\mathbf{x}) = rT(\mathbf{x})$. These conditions are equivalent to showing $|T(\mathbf{x}+\mathbf{y})-T(\mathbf{x})-T(\mathbf{y})|^2 = 0$ and $|T(r\mathbf{x})-rT(\mathbf{x})|^2 = 0$. For this just recall that $|\mathbf{x}|^2 = \mathbf{x} \cdot \mathbf{x}$ and use the use the definition of an orthogonal transformation. The calculations are straightforward, but somewhat long, so we will leave them as an exercise.

For the converse, recall that $Q \in \mathbb{R}^{n\times n}$ is orthogonal if and only if $Q^TQ = I_n$. To show $T(\mathbf{x}) = Q\mathbf{x}$ is orthogonal, note that

$$T(\mathbf{x}) \cdot T(\mathbf{y}) = \big(Q\mathbf{x}\big)^T Q\mathbf{x} = \mathbf{x}^T Q^T Q\mathbf{x} = \mathbf{x}^T\mathbf{x},$$

so indeed the claim is correct. □

In other words, orthogonal transformations are just orthogonal matrices. For example, rotations of $\mathbb{R}^2$, studied in the last Section, are examples of orthogonal transformations, and we will now consider another type.

Let's next consider reflections. In everyday terms, one's reflection is the image one sees in a mirror. The mirror can be thought of as a plane through which $\mathbb{R}^3$ is being reflected. To analyze how a reflection happens, let's first analyze how a reflection works in $\mathbb{R}^2$. Suppose ℓ is a line in $\mathbb{R}^2$, which we'll

assume passes through the origin. Then the reflection of $\mathbb{R}^2$ through ℓ acts as follows. Every point on ℓ is fixed, and every point on the line $\ell^\perp$ through the origin orthogonal to ℓ is sent to its negative.

To analyze what happens further, let $\mathbf{b}$ be any non-zero vector on $\ell^\perp$, and let $H_\mathbf{b} : \mathbb{R}^2 \to \mathbb{R}^2$ denote the reflection through ℓ. Choose an arbitrary $\mathbf{v} \in \mathbb{R}^2$, and consider its orthogonal decomposition

$$\mathbf{v} = P_\mathbf{b}(\mathbf{v}) + \mathbf{c}$$

with $\mathbf{c}$ on ℓ. By the definition of the reflection,

$$H_\mathbf{b}(\mathbf{v}) = -P_\mathbf{b}(\mathbf{v}) + \mathbf{c}.$$

Replacing $\mathbf{c}$ by $\mathbf{v} - P_\mathbf{b}(\mathbf{v})$ gives the formula

$$\begin{aligned} H_\mathbf{b}(\mathbf{v}) &= \mathbf{v} - 2P_\mathbf{b}(\mathbf{v}) \\ &= \mathbf{v} - 2\big(\frac{\mathbf{v}\cdot\mathbf{b}}{\mathbf{b}\cdot\mathbf{b}}\big)\mathbf{b}. \end{aligned}$$

Expressing this in terms of the unit vector $\widehat{\mathbf{b}}$ determined by $\mathbf{b}$ gives us the simpler expression

$$H_\mathbf{b}(\mathbf{v}) = \mathbf{v} - 2(\mathbf{v}\cdot\widehat{\mathbf{b}})\widehat{\mathbf{b}}. \tag{7.5}$$

Lets just check $H_\mathbf{b}$ has the properties we sought. First, $H_\mathbf{b}(\mathbf{v}) = \mathbf{v}$ if $\mathbf{b}\cdot\mathbf{v} = 0$. Second, $H_\mathbf{b}(\mathbf{b}) = -\mathbf{b}$. Notice that $H_\mathbf{b} = I_2 - 2P_\mathbf{b}$, so a reflection is a linear transformation.

The above expression of a reflection extends from $\mathbb{R}^2$ to $\mathbb{R}^3$ and even to $\mathbb{R}^n$ for any $n \geq 3$ as follows. Let $\mathbf{b}$ be any nonzero vector in $\mathbb{R}^n$, and let W be the hyperplane in $\mathbb{R}^n$ consisting of all the vectors orthogonal to $\mathbf{b}$. Then the transformation $H_\mathbf{b} : \mathbb{R}^n \to \mathbb{R}^n$ defined by (7.5) is the reflection of $\mathbb{R}^n$ through W.

Example 7.10. Let $\mathbf{b} = (1,1)^T$, so $\widehat{\mathbf{b}} = (\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})^T$. Then $H_\mathbf{b}$ is the reflection through the line $x = -y$. We have

$$\begin{aligned} H_\mathbf{b}\begin{pmatrix} a \\ b \end{pmatrix} &= \begin{pmatrix} a \\ b \end{pmatrix} - 2\Big(\begin{pmatrix} a \\ b \end{pmatrix} \cdot \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}\Big) \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} \\ &= \begin{pmatrix} a - (a+b) \\ b - (a+b) \end{pmatrix} \\ &= \begin{pmatrix} -b \\ -a \end{pmatrix}. \end{aligned}$$

There are several worthwhile consequences of formula (7.5). As noted, any reflection through a hyperplane is linear, and the composition of a reflection with itself is the identity (since reflecting any $\mathbf{v}$ twice returns $\mathbf{v}$ to itself). That is, $H_{\mathbf{b}} \circ H_{\mathbf{b}} = I_2$. Furthermore, reflections preserve inner products,

$$H_{\mathbf{b}}(\mathbf{v}) \cdot H_{\mathbf{b}}(\mathbf{w}) = \mathbf{v} \cdot \mathbf{w},$$

so indeed they are also orthogonal. We will leave these the proofs of these properties as an exercise.

We noted above that every rotation of $\mathbb{R}^2$ is orthogonal (for example, a rotation matrix is orthogonal). The next Proposition gives a somewhat surprising description of the orthogonal transformations of $\mathbb{R}^2$.

Proposition 7.10. *Every orthogonal transformation of $\mathbb{R}^2$ is either a reflection or a rotation. In fact, the reflections are those orthogonal transformations T given by a symmetric orthogonal matrix different from I_2. The rotations $\mathcal{R}_\theta$ are those orthogonal transformations whose matrix R_θ is either I_2 or not symmetric.*

Proof. It is not hard to check that any 2×2 orthogonal matrix has the form

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

or

$$H_\theta = \begin{pmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{pmatrix}.$$

The former are rotations (including I_2) and the latter are symmetric, but do not include I_2. The transformations H_θ are in fact reflections. In fact, H_θ is the reflection through the line spanned by $(\cos(\theta/2), \sin(\theta/2))^T$. □

Recall that the set of 2×2 orthogonal matrices is the matrix group denoted aby $O(2, \mathbb{R})$. The product of two orthogonal matrices is therefore orthogonal. It is an interesting exercise to identify what sort of transformation is obtained by multiplying two reflections and multiplying a reflection and a rotation (see Exercise 7.33).

The structure of orthogonal transformations in higher dimensions is quite a bit more complicated, even in the 3×3 case. That is, the rotations and reflections of $\mathbb{R}^3$ do not give all the possible orthogonal linear transformations of $\mathbb{R}^3$.

Exercises

Exercise 7.21. Let $\mathbf{a} \in \mathbb{R}^n$.

(*i*) Show that the projection

$$P_{\mathbf{a}}(\mathbf{x}) = \big(\frac{\mathbf{a} \cdot \mathbf{x}}{\mathbf{a} \cdot \mathbf{a}}\big)\mathbf{a}$$

is linear.

(*ii*) Verify from the formula that the projection $P_{\mathbf{a}}$ fixes every vector on the line spanned by $\mathbf{a}$ and sends every vector orthogonal to $\mathbf{a}$ to $\mathbf{0}$.

(*iii*) Verify that every projection matrix P satisfies $P^2 = P$.

Exercise 7.22. Let $H_{\mathbf{b}} : \mathbb{R}^2 \to \mathbb{R}^2$ be the reflection of $\mathbb{R}^2$ through the line orthogonal to $\mathbf{b}$.

(*i*) Find formulas for $H_{\mathbf{b}}((1,0))$ and $H_{\mathbf{b}}((0,1))$.

(*ii*) Find an explicit formula for the matrix A of $H_{\mathbf{b}}$.

(*iii*) Show also that $H_{\mathbf{b}}(H_{\mathbf{b}}(\mathbf{x})) = \mathbf{x}$.

(*iv*) Conclude from part (*iii*) that $A^2 = I_2$. Also, check this with your explicit formula for A.

Exercise 7.23. Let $V = \mathbb{C}$ and consider the transformation $H : V \to V$ defined by $H(z) = \overline{z}$. Interpret H as a transformation from $\mathbb{R}^2$ to $\mathbb{R}^2$. Is H orthogonal?

Exercise 7.24. Consider the transformation $C_{\mathbf{a}} : \mathbb{R}^3 \to \mathbb{R}^3$ 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.

(*ii*) Describe the set of vectors $\mathbf{x}$ such that $C_{\mathbf{a}}(\mathbf{x}) = \mathbf{0}$.

(*iii*) Give a formula for $|C_{\mathbf{a}}(\mathbf{v})|$ in terms of $|\mathbf{a}|$ and $|\mathbf{v}|$.

Exercise 7.25. Let $\mathbf{u}$ and $\mathbf{v}$ be two orthogonal unit length vectors in $\mathbb{R}^2$. Show 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*) $P_{\mathbf{u}}(P_{\mathbf{v}}(\mathbf{x})) = P_{\mathbf{v}}(P_{\mathbf{u}}(\mathbf{x})) = \mathbf{0}$.

(*iii*) Conclude from (a) that $\mathbf{x} = (\mathbf{x} \cdot \mathbf{u})\mathbf{u} + (\mathbf{x} \cdot \mathbf{v})\mathbf{v}$.

Exercise 7.26. Complete the proof of Proposition 7.9 .

Exercise 7.27. Suppose $T : \mathbb{R}^2 \to \mathbb{R}^2$ is a linear transformation which sends any two non collinear vectors to non collinear vectors. Suppose $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{R}^2$ are non collinear. Show that T sends any 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 7.28. Show that all reflections are orthogonal linear transformations. In other words, show that for all $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{R}^n$,

$$H_{\mathbf{b}}(\mathbf{x}) \cdot H_{\mathbf{b}}(\mathbf{y}) = \mathbf{x} \cdot \mathbf{y}.$$

Exercise 7.29. Show explicitly that every orthogonal linear transformation preserves lengths of vectors, and angles and distances between two distinct vectors.

Exercise 7.30. A linear transformation $T : \mathbb{R}^n \to \mathbb{R}^n$ which preserves angles is called *conformal*. Describe the set of conformal mappings $T : \mathbb{R}^2 \to \mathbb{R}^2$.

Exercise 7.31. Find the reflection of $\mathbb{R}^3$ through the plane P if:

(i) P is the plane $x + y + z = 0$; and

(ii) P is the plane $ax + by + cz = 0$.

Exercise 7.32. Which of the following statements are true? Explain.

(i) The composition of two rotations is a rotation.

(ii) The composition of two reflections is a reflection.

(iii) The composition of a reflection and a rotation is a rotation.

(iv) The composition of two reflections is a rotation.

Exercise 7.33. Let H_λ, H_μ and R_ν be 2×2 reflection, reflection and rotation matrices as defined in the proof of Proposition 7.10. Find and describe:

(i) $H_\lambda H_\mu$;

(ii) $H_\lambda R_\nu$; and

(iii) $R_\nu H_\lambda$.

7.4 The Matrix of a Linear Transformation

Now suppose V and W are finite dimensional vector spaces over $\mathbb{F}$, and suppose $T : V \to W$ is a linear transformation. The purpose of this section is to give a define the matrix of T with respect to arbitrary bases of the domain V and the target W.

7.4.1 The Matrix $\mathcal{M}_{\mathcal{B}'}^{\mathcal{B}}(T)$

As above, let V and W be finite dimensional vector spaces over $\mathbb{F}$, and suppose $T : V \to W$ is linear. Fix a basis

$$\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$$

of V and a basis

$$\mathcal{B}' = \{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_m\}$$

of W. We will now define the matrix $\mathcal{M}_{\mathcal{B}'}^{\mathcal{B}}(T)$ of T with respect to these bases. Since $\mathcal{B}'$ is a basis of W, each $T(\mathbf{v}_j)$ can be uniquely written

$$T(\mathbf{v}_j) = \sum_{i=1}^{m} c_{ij}\mathbf{w}_i. \tag{7.6}$$

Definition 7.2. The *matrix of T with respect to the bases $\mathcal{B}$ and $\mathcal{B}'$* is defined to be the $m \times n$ matrix $\mathcal{M}_{\mathcal{B}'}^{\mathcal{B}}(T) = (c_{ij})$.

What this definition says is that the jth column of the matrix $\mathcal{M}_{\mathcal{B}'}^{\mathcal{B}}(T)$ is the column vector consisting of the coefficients of $T(\mathbf{v}_j)$ with respect to the basis $\mathcal{B}'$. One can express (7.6) in matrix form by

$$(T(\mathbf{v}_1)\ T(\mathbf{v}_2) \cdots T(\mathbf{v}_n)) = (\mathbf{w}_1\ \mathbf{w}_2 \cdots \mathbf{w}_m)\mathcal{M}_{\mathcal{B}'}^{\mathcal{B}}(T). \tag{7.7}$$

It's important to note that if $V = \mathbb{F}^n$, $W = \mathbb{F}^m$ and $T = T_A$, where $A \in \mathbb{F}^{m\times n}$, then $\mathcal{M}_{\mathcal{B}'}^{\mathcal{B}}(T) = A$ if $\mathcal{B}$ and $\mathcal{B}'$ are the standard bases. For $T_A(\mathbf{e}_j)$ is the always jth column of A. We remark that (7.4.3) implies

$$\mathcal{M}_{\mathcal{B}'}^{\mathcal{B}}(Id) = \mathcal{M}_{\mathcal{B}}^{\mathcal{B}'}, \tag{7.8}$$

where $Id : V \to V$ is the identity.

Example 7.11. Let V be the space of real polynomials of degree at most three, and let W be the space of real polynomials of degree at most two. Then differentiation is a linear transformation $D : V \to W$. Now

$$D(ax^3 + bx^2 + cx + d) = 3ax^2 + 2bx + c.$$

Let $\mathcal{B}$ be the basis $\{1, x, x^2, x^3\}$ of V and $\mathcal{B}'$ the basis $\{1, x, x^2\}$ of W. Now,

$$D(1) = 0,\ D(x) = 1,\ D(x^2) = 2x,\ D(x^3) = 3x^2.$$

Thus

$$\mathcal{M}_{\mathcal{B}'}^{\mathcal{B}}(D) = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix}.$$

Now suppose that $T : V \to V$ is a diagonalizable (or semi-simple) linear transformation. Recall that this means there exists a basis $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ of V such that $(\mathbf{v}_i) = \lambda_i \mathbf{v}_i$ for each index i between 1 and n. Hence $\mathcal{B}$ is an eigenbasis of V for T. In this case, $\mathcal{M}_{\mathcal{B}}^{\mathcal{B}}(T)$ is the diagonal matrix

$$\begin{pmatrix} \lambda_1 & & O \\ & \ddots & \\ O & & \lambda_n \end{pmatrix}.$$

But what is the matrix of T with respect to some other basis $\mathcal{B}'$ of V? This is the question that will be answered next. The first step in answering this question is to find out how to relate the expansions of a given vector in V with respect to two different bases.

7.4.2 Coordinates With Respect to a Basis

Our next goal is to find how one passes from one set of coordinates to another. Let $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ be a basis of V. Then every $\mathbf{v} \in V$ has a unique expression

$$\mathbf{v} = r_1\mathbf{v}_1 + r_2\mathbf{v}_2 + \cdots + r_n\mathbf{v}_n.$$

Definition 7.3. We will call $r_1, r_2, \ldots, r_n$ the *coordinates* of $\mathbf{v}$ with respect to $\mathcal{B}$. We will sometimes write $\mathbf{w} =< r_1, r_2, \ldots, r_n >$.

Of course, finding the coordinates of a vector in $\mathbb{F}^n$ with respect to a basis is the familiar problem of solving a linear system. We now want to consider how to go from one set of coordinates to another. It shouldn't be surprising that this also involves a linear system.

Suppose $\mathcal{B}' = \{\mathbf{v'}_1, \mathbf{v'}_2, \ldots, \mathbf{v'}_n\}$ is another a basis of V. Then

$$\mathbf{v} = s_1\mathbf{v'}_1 + s_2\mathbf{v'}_2 + \cdots + s_n\mathbf{v'}_n.$$

How does one pass from the coordinates $< r_1, r_2, \ldots, r_n >$ with respect to $\mathcal{B}$ to the coordinates $< s_1, s_2, \ldots, s_n >$ with respect to $\mathcal{B}'$? The simplest

answer comes from using matrix notation as follows. Define

$$\begin{pmatrix}\mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n\end{pmatrix}\begin{pmatrix}r_1\\ r_2\\ \vdots\\ r_n\end{pmatrix} = \sum_{i=1}^{n} r_i\mathbf{v}_i. \tag{7.9}$$

Then

$$\begin{pmatrix}\mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n\end{pmatrix}\begin{pmatrix}r_1\\ r_2\\ \vdots\\ r_n\end{pmatrix} = \sum_{i=1}^{n} r_i\mathbf{v}_i = \begin{pmatrix}\mathbf{v}'_1 & \mathbf{v}'_2 & \cdots & \mathbf{v}'_n\end{pmatrix}\begin{pmatrix}s_1\\ s_2\\ \vdots\\ s_n\end{pmatrix}.$$

Suppose we write

$$\begin{pmatrix}\mathbf{v}'_1 & \mathbf{v}_2 & \cdots & \mathbf{v}'_n\end{pmatrix} = \begin{pmatrix}\mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n\end{pmatrix} A$$

for some $n \times n$ matrix A over $\mathbb{F}$. Then we can argue that

$$\begin{pmatrix}r_1\\ r_2\\ \vdots\\ r_n\end{pmatrix} = A\begin{pmatrix}s_1\\ s_2\\ \vdots\\ s_n\end{pmatrix}. \tag{7.10}$$

Notice that if $V = \mathbb{F}^n$, then $\begin{pmatrix}\mathbf{v}'_1 & \mathbf{v}'_2 & \cdots & \mathbf{v}'_n\end{pmatrix} \in \mathbb{F}^{n\times n}$, so we can write

$$A = \begin{pmatrix}\mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n\end{pmatrix}^{-1}\begin{pmatrix}\mathbf{v}'_1 & \mathbf{v}'_2 & \cdots & \mathbf{v}'_n\end{pmatrix}.$$

We will denote the change of basis matrix A by $\mathcal{M}^{\mathcal{B}}_{\mathcal{B}'}$ (see below).

Example 7.12. Suppose $\mathbb{F} = \mathbb{R}$, and consider two bases of $\mathbb{R}^2$, say

$$\mathcal{B} = \{\begin{pmatrix}1\\2\end{pmatrix}, \begin{pmatrix}0\\1\end{pmatrix}\} \quad \text{and} \quad \mathcal{B}' = \{\begin{pmatrix}1\\1\end{pmatrix}, \begin{pmatrix}1\\-1\end{pmatrix}\}.$$

Suppose for example that $\mathbf{v} = \mathbf{e}_1$. The two sets of coordinates are easily found 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 the coordinates of $\mathbf{e}_1$ with respect to $\mathcal{B}$ are $< 1, -2 >$, and with respect to $\mathcal{B}'$ they are $< \frac{1}{2}, \frac{1}{2} >$.

To see how the two different sets of coordinates are related, we set up the following system, which expresses $\mathcal{B}'$ in terms of $\mathcal{B}$:

$$\begin{pmatrix}1\\1\end{pmatrix} = a_{11}\begin{pmatrix}1\\2\end{pmatrix} + a_{21}\begin{pmatrix}0\\1\end{pmatrix},$$

and

$$\begin{pmatrix}1\\-1\end{pmatrix} = a_{12}\begin{pmatrix}1\\2\end{pmatrix} + a_{22}\begin{pmatrix}0\\1\end{pmatrix}.$$

These equations are expressed in matrix form as:

$$\begin{pmatrix}1 & 1\\1 & -1\end{pmatrix} = \begin{pmatrix}1 & 0\\2 & 1\end{pmatrix}\begin{pmatrix}a_{11} & a_{12}\\a_{21} & a_{22}\end{pmatrix}.$$

The change of basis matrix $\mathcal{M}_{\mathcal{B}'}^{\mathcal{B}}$ which expresses how to go between the two sets of coordinates is defined (see below) to be

$$\mathcal{M}_{\mathcal{B}'}^{\mathcal{B}} = \begin{pmatrix}a_{11} & a_{12}\\a_{21} & a_{22}\end{pmatrix}.$$

Of course, here

$$\mathcal{M}_{\mathcal{B}'}^{\mathcal{B}} = \begin{pmatrix}1 & 0\\2 & 1\end{pmatrix}^{-1}\begin{pmatrix}1 & 1\\1 & -1\end{pmatrix} = \begin{pmatrix}1 & 1\\-1 & -3\end{pmatrix}.$$

Now suppose the coordinates of $\mathbf{p}$ with respect to $\mathcal{B}$ are $< r, s >$ and those with respect to $\mathcal{B}'$ are $< x, y >$. Then

$$\mathbf{p} = \begin{pmatrix}1 & 0\\2 & 1\end{pmatrix}\begin{pmatrix}r\\s\end{pmatrix} = \begin{pmatrix}1 & 1\\1 & -1\end{pmatrix}\begin{pmatrix}x\\y\end{pmatrix}.$$

Hence

$$\begin{pmatrix}r\\s\end{pmatrix} = \begin{pmatrix}1 & 0\\2 & 1\end{pmatrix}^{-1}\begin{pmatrix}1 & 1\\1 & -1\end{pmatrix}\begin{pmatrix}x\\y\end{pmatrix},$$

so

$$\begin{pmatrix}r\\s\end{pmatrix} = \mathcal{M}_{\mathcal{B}'}^{\mathcal{B}}\begin{pmatrix}x\\y\end{pmatrix} = \begin{pmatrix}1 & 1\\-1 & -3\end{pmatrix}\begin{pmatrix}x\\y\end{pmatrix}. \tag{7.11}$$

To summarize this example, we outline the general procedure for the case $n = 2$. Suppose $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2\}$ and $\mathcal{B}' = \{\mathbf{v}_1', \mathbf{v}_2'\}$. Let

$$\mathbf{w} = x\mathbf{v}_1' + y\mathbf{v}_2' = r\mathbf{v}_1 + s\mathbf{v}_2.$$

Now there exist unique scalars $a_{ij} \in \mathbb{F}$ such that

$$\mathbf{v}_1' = a_{11}\mathbf{v}_1 + a_{21}\mathbf{v}_2$$

$$\mathbf{v}_2' = a_{12}\mathbf{v}_1 + a_{22}\mathbf{v}_2.$$

Putting this into matrix form gives

$$(\mathbf{v}_1' \ \mathbf{v}_2') = (\mathbf{v}_1 \ \mathbf{v}_2)\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = (\mathbf{v}_1 \ \mathbf{v}_2)\mathcal{M}_{\mathcal{B}'}^{\mathcal{B}}.$$

By substitution, we see that

$$\mathbf{w} = (\mathbf{v}_1' \ \mathbf{v}_2')\begin{pmatrix} x \\ y \end{pmatrix} = (\mathbf{v}_1 \ \mathbf{v}_2)\mathcal{M}_{\mathcal{B}'}^{\mathcal{B}}\begin{pmatrix} x \\ y \end{pmatrix} = (\mathbf{v}_1 \ \mathbf{v}_2)\begin{pmatrix} r \\ s \end{pmatrix}.$$

Since $\mathbf{v}_1$ and $\mathbf{v}_2$ are independent, we get the final result:

$$\begin{pmatrix} r \\ s \end{pmatrix} = \mathcal{M}_{\mathcal{B}'}^{\mathcal{B}}\begin{pmatrix} x \\ y \end{pmatrix}. \tag{7.12}$$

Now consider 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

$$\mathbf{v}_j' = \sum_{i=1}^{n} a_{ij}\mathbf{v}_i.$$

This expression uniquely determines $\mathcal{M}_{\mathcal{B}'}^{\mathcal{B}}$, and gives the basic matrix identity

$$(\mathbf{v}_1' \ \mathbf{v}_2' \ \cdots \ \mathbf{v}_n') = (\mathbf{v}_1 \ \mathbf{v}_2 \ \cdots \ \mathbf{v}_n)\mathcal{M}_{\mathcal{B}'}^{\mathcal{B}}. \tag{7.13}$$

Notice that if $\mathcal{B}$ is any basis of V, then

$$(\mathbf{v}_1 \ \mathbf{v}_2 \ \cdots \ \mathbf{v}_n)A = (\mathbf{v}_1 \ \mathbf{v}_2 \ \cdots \ \mathbf{v}_n)B \ \Rightarrow \ A = B. \tag{7.14}$$

This is due to the fact that expressions in terms of bases are unique. Also, note

Proposition 7.11. *Suppose $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is a basis of V, and $A \in \mathbb{F}^{n\times n}$ is invertible. Then putting*

$$(\mathbf{w}_1\ \mathbf{w}_2 \ldots \mathbf{w}_n) = (\mathbf{v}_1\ \mathbf{v}_2 \ldots \mathbf{v}_n)A$$

defines a new basis of $\mathcal{B}' = \{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_n\}$ of V with change of basis matrix $\mathcal{M}_{\mathcal{B}'}^{\mathcal{B}} = A$.

Proof. Since $\dim V = n$, it suffices to show that $\mathbf{w}_1, \ldots, \mathbf{w}_n$ are independent. Suppose $\sum a_i \mathbf{w}_i = \mathbf{0}$. Then

$$\mathbf{0} = (\mathbf{w}_1\ \mathbf{w}_2 \ldots \mathbf{w}_n) \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} = (\mathbf{v}_1\ \mathbf{v}_2 \ldots \mathbf{v}_n)A \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}.$$

Putting $\mathbf{a} = (a_1, \ldots, a_n)^T$, it follows from the independence of $\mathcal{B}$ that $A\mathbf{a} = \mathbf{0}$. Since A is invertible, this implies $\mathbf{a} = \mathbf{0}$, giving the result. □

Now let us explore some of the properties of change of basis matrices.

Proposition 7.12. *Let $\mathcal{B}$ and $\mathcal{B}'$ be bases of V. Then*

$$\mathcal{M}_{\mathcal{B}}^{\mathcal{B}'} = (\mathcal{M}_{\mathcal{B}'}^{\mathcal{B}})^{-1}.$$

Proof. We have

$$(\mathbf{v}_1\ \ldots \mathbf{v}_n) = (\mathbf{v}'_1\ \ldots \mathbf{v}'_n)\mathcal{M}_{\mathcal{B}}^{\mathcal{B}'} = (\mathbf{v}_1\ \ldots \mathbf{v}_n)\mathcal{M}_{\mathcal{B}'}^{\mathcal{B}}\mathcal{M}_{\mathcal{B}}^{\mathcal{B}'}.$$

Thus, since $\mathcal{B}$ is a basis,

$$\mathcal{M}_{\mathcal{B}'}^{\mathcal{B}}\mathcal{M}_{\mathcal{B}}^{\mathcal{B}'} = I_n.$$

This shows $(\mathcal{M}_{\mathcal{B}'}^{\mathcal{B}})^{-1} = \mathcal{M}_{\mathcal{B}}^{\mathcal{B}'}$. □

What happens if a third basis $\mathcal{B}'' = \{\mathbf{v}''_1, \ldots, \mathbf{v}''_n\}$ is thrown in? If we iterate the expression in (7.16), we get

$$(\mathbf{v}''_1\ \ldots \mathbf{v}''_n) = (\mathbf{v}'_1\ \ldots \mathbf{v}'_n)\mathcal{M}_{\mathcal{B}''}^{\mathcal{B}'} = (\mathbf{v}_1\ \ldots \mathbf{v}_n)\mathcal{M}_{\mathcal{B}'}^{\mathcal{B}}\mathcal{M}_{\mathcal{B}''}^{\mathcal{B}'}.$$

Thus we get the final result on change of bases, namely

$$\mathcal{M}_{\mathcal{B}''}^{\mathcal{B}} = \mathcal{M}_{\mathcal{B}'}^{\mathcal{B}}\mathcal{M}_{\mathcal{B}''}^{\mathcal{B}'}. \tag{7.15}$$

7.4.3 Changing Basis for the Matrix of a Linear Transformation

The purpose of this section is to derive the change of basis forrmula for a linear transformation $T : V \to V$. Suppose

$$\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$$

and

$$\mathcal{B}' = \{\mathbf{v'}_1, \mathbf{v'}_2, \ldots, \mathbf{v'}_n\}$$

are bases of V. We want to derive the formula for $\mathcal{M}_{\mathcal{B}'}^{\mathcal{B}'}(T)$ in terms of $\mathcal{M}_{\mathcal{B}}^{\mathcal{B}}(T)$. Recall that

$$(T(\mathbf{v}_1)\ T(\mathbf{v}_2) \cdots T(\mathbf{v}_n)) = (\mathbf{v}_1\ \mathbf{v}_2 \cdots \mathbf{v}_n)\mathcal{M}_{\mathcal{B}}^{\mathcal{B}}(T).$$

Thus, by (7.14), $\mathcal{M}_{\mathcal{B}}^{\mathcal{B}}(T)$ is well defined.

We will now prove

Proposition 7.13. *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

$$\mathcal{M}_{\mathcal{B}'}^{\mathcal{B}'}(T) = P\mathcal{M}_{\mathcal{B}}^{\mathcal{B}}(T)P^{-1}. \tag{7.17}$$

Proof. For simplicity, let us assume $n = 2$. Hence, $(\mathbf{v}'_1\ \mathbf{v}'_2) = (\mathbf{v}_1\ \mathbf{v}_2)\mathcal{M}_{\mathcal{B}'}^{\mathcal{B}}$. Since T is linear,

$$T(\mathbf{v'}_1\ \mathbf{v'}_2) = T\big((\mathbf{v}_1\ \mathbf{v}_2)\mathcal{M}_{\mathcal{B}'}^{\mathcal{B}}\big) = \big(T(\mathbf{v}_1)\ T(\mathbf{v}_2)\big)\mathcal{M}_{\mathcal{B}'}^{\mathcal{B}}.$$

Thus,

$$\begin{aligned}
\big(T(\mathbf{v}'_1)\ T(\mathbf{v}'_2)\big) &= T(\mathbf{v'}_1\ \mathbf{v'}_2) \\
&= \big(T(\mathbf{v}_1)\ T(\mathbf{v}_2)\big)\mathcal{M}_{\mathcal{B}'}^{\mathcal{B}} \\
&= (\mathbf{v}_1\ \mathbf{v}_2)\mathcal{M}_{\mathcal{B}}^{\mathcal{B}}(T)\mathcal{M}_{\mathcal{B}'}^{\mathcal{B}} \\
&= (\mathbf{v}'_1\ \mathbf{v}'_2)\mathcal{M}_{\mathcal{B}}^{\mathcal{B}'}\mathcal{M}_{\mathcal{B}}^{\mathcal{B}}(T)\mathcal{M}_{\mathcal{B}'}^{\mathcal{B}}.
\end{aligned}$$

This implies, by definition, that

$$(\mathbf{v}'_1\ \mathbf{v}'_2)\mathcal{M}_{\mathcal{B}'}^{\mathcal{B}'}(T) = (\mathbf{v}'_1\ \mathbf{v}'_2)\mathcal{M}_{\mathcal{B}}^{\mathcal{B}'}\mathcal{M}_{\mathcal{B}}^{\mathcal{B}}(T)\mathcal{M}_{\mathcal{B}'}^{\mathcal{B}}.$$

Hence, by (7.14),

$$\mathcal{M}_{\mathcal{B}'}^{\mathcal{B}'}(T) = \mathcal{M}_{\mathcal{B}}^{\mathcal{B}'}\mathcal{M}_{\mathcal{B}}^{\mathcal{B}}(T)\mathcal{M}_{\mathcal{B}'}^{\mathcal{B}}.$$

Example 7.13. Consider the linear transformation T of $\mathbb{R}^2$ 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 $\mathcal{B}'$ of Example 7.12. Calling the standard basis $\mathcal{B}$, formula (7.16) says

$$B = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}^{-1} \begin{pmatrix} 1 & 0 \\ -4 & 3 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}.$$

Computing the product gives

$$B = \begin{pmatrix} 0 & -3 \\ 1 & 4 \end{pmatrix}.$$

The following example demonstrates a key application of the change of basis.

Example 7.14. Suppose $A \in \mathbb{F}^{n\times n}$ has an eigenbasis $\mathbf{v}_1, \ldots, \mathbf{v}_n$, which we will call $\mathcal{B}'$, and say $A\mathbf{v}_i = \lambda_i \mathbf{v}_i$ for all i. Let $P = (\mathbf{v}_1 \cdots \mathbf{v}_n)$, so $P = \mathcal{M}^{\mathcal{B}}_{\mathcal{B}'}$, where $\mathcal{B}$ is the standard basis. Then $AP = (\lambda_1 \mathbf{v}_1 \cdots \lambda_n \mathbf{v}_n) = PD$, where $D = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$. Note that $D = \mathcal{M}^{\mathcal{B}'}_{\mathcal{B}'}(T)$, where $T = T_A$, so this identity corresponds to (7.16), which says $D = P^{-1}AP$, since $A = \mathcal{M}^{\mathcal{B}}_{\mathcal{B}}(T_A)$. Note that this is equivalent to $A = PDP^{-1}$, which is immediate from $AP = PD$.

Definition 7.4. Let A and B be $n \times n$ matrices over $\mathbb{F}$. Then we say A *is similar to* B if and only if there exists an invertible $P \in \mathbb{F}^{n\times n}$ such that $B = PAP^{-1}$. We say that A is *diagonalizable* if and only if A is similar to a diagonal matrix.

Proposition 7.14. *If $A \in \mathbb{F}^{n\times n}$ is similar to a diagonal matrix D, say $A = QDQ^{-1}$, then the columns of $P = Q^{-1}$ are a basis of eigenvectors of A and the diagonal entries of D are the corresponding eigenvalues.*

Proof. Just argue as in the previous example. □

Note that two matrices representing the same linear transformation are similar, by equation (7.16). Conversely, if two matrices are similar, then they represent the same linear transformation with respect to suitable bases. Moreover, any linear transformation which is represented by a diagonalizable matrix is semi-simple and conversely.

It's not hard to see that similarity is an equivalence relation on $\mathbb{F}^{n\times n}$ (Exercise: check this). The equivalence class of a matrix A is called the *conjugacy class* of A. To summarize, we state

Proposition 7.15. *Let V be a vector space over $\mathbb{F}$ such that $\dim V = n$. Then the set of matrices representing a fixed linear transformation $T : V \to V$ form a conjugacy class in $\mathbb{F}^{n\times n}$. In addition, T is semi-simple if and only if its conjugacy class contains a diagonalizable matrix.*

Example 7.15. Let $\mathbb{F} = \mathbb{R}$ and suppose $\mathbf{v}_1$ and $\mathbf{v}_2$ denote $(1, 2)^T$ and $(0, 1)^T$ respectively. Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be the linear transformation such that such that $T(\mathbf{v}_1) = \mathbf{v}_1$ and $T(\mathbf{v}_2) = 3\mathbf{v}_2$. Note that Proposition 7.1 implies that T exists and is unique. Now the matrix of T with respect to the basis $\mathbf{v}_1, \mathbf{v}_2$ is

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

Thus T has a diagonal matrix in the $\mathbf{v}_1, \mathbf{v}_2$ basis. Let us find the matrix of T with respect to the standard basis. Let $\mathcal{B}$ denote $\{\mathbf{e}_1, \mathbf{e}_2\}$ and $\mathcal{B}' = \{\mathbf{v}_1, \mathbf{v}_2\}$. Now

$$\mathcal{M}_{\mathcal{B}}^{\mathcal{B}'} = \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix}^{-1} = \begin{pmatrix} 1 & 0 \\ -2 & 1 \end{pmatrix}.$$

We have

$$\mathcal{M}_{\mathcal{B}'}^{\mathcal{B}'}(T) = \begin{pmatrix} 1 & 0 \\ 0 & 3 \end{pmatrix} = \mathcal{M}_{\mathcal{B}}^{\mathcal{B}'}\mathcal{M}_{\mathcal{B}}^{\mathcal{B}}(T)\mathcal{M}_{\mathcal{B}'}^{\mathcal{B}}.$$

Thus

$$\mathcal{M}_{\mathcal{B}}^{\mathcal{B}}(T) = \mathcal{M}_{\mathcal{B}'}^{\mathcal{B}}\mathcal{M}_{\mathcal{B}'}^{\mathcal{B}'}(T)\mathcal{M}_{\mathcal{B}}^{\mathcal{B}'} = \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 3 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ -2 & 1 \end{pmatrix}.$$

Multiplying this out, we find

$$\mathcal{M}_{\mathcal{B}}^{\mathcal{B}}(T) = \begin{pmatrix} 1 & 0 \\ -4 & 3 \end{pmatrix}.$$

Exercises

Exercise 7.34. Find the coordinates of $\mathbf{e}_1,\ \mathbf{e}_2,\ \mathbf{e}_3$ of $\mathbb{R}^3$ in terms of the basis $\begin{pmatrix}1\\0\\0\end{pmatrix}, \begin{pmatrix}1\\1\\0\end{pmatrix}, \begin{pmatrix}1\\1\\1\end{pmatrix}$ of $\mathbb{R}^3$, and then find the matrix of the linear transformation

$$T\begin{pmatrix}x_1\\x_2\\x_3\end{pmatrix} = \begin{pmatrix}4x_1 + x_2 - x_3\\ x_1 + 3x_3\\ x_2 + 2x_3\end{pmatrix}$$

with respect to this basis.

Exercise 7.35. Again, consider the basis $\begin{pmatrix}1\\0\\0\end{pmatrix}, \begin{pmatrix}1\\1\\0\end{pmatrix}, \begin{pmatrix}1\\1\\1\end{pmatrix}$ of $\mathbb{R}^3$. Find the matrix of the linear transformation $T : \mathbb{R}^3 \to \mathbb{R}^3$ defined by $T(\mathbf{x}) = (1,1,1)^T \times \mathbf{x}$ with respect to this basis.

Exercise 7.36. Let $Id : V \to V$ be the identity transformation. Show that for any two bases $\mathcal{B}$ and $\mathcal{B}'$ of V, we have $\mathcal{M}^{\mathcal{B}}_{\mathcal{B}'}(Id) = \mathcal{M}^{\mathcal{B}'}_{\mathcal{B}}$.

Exercise 7.37. Let $H : \mathbb{R}^2 \to \mathbb{R}^2$ be the reflection through the line $x = y$. Find a basis of $\mathbb{R}^2$ such that the matrix of H is diagonal, and find the matrix of H with respect to this basis.

Exercise 7.38. Show that any projection $P_{\mathbf{a}} : \mathbb{R}^2 \to \mathbb{R}^2$ is diagonal by explicitly finding a basis for which the matrix of $P_{\mathbf{a}}$ is diagonal. Also, find the diagonal matrix?

Exercise 7.39. Let $\mathcal{R}_\theta$ be any 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 7.40. Let S and T be linear transformations from V to W, and suppose $\mathcal{B}$ is a basis of V and $\mathcal{B}'$ a basis of W. Show that

$$\mathcal{M}^{\mathcal{B}}_{\mathcal{B}'}(aS + bT) = \mathcal{M}^{\mathcal{B}}_{\mathcal{B}'}(S) + \mathcal{M}^{\mathcal{B}}_{\mathcal{B}'}(T)$$

for any $a,\ b \in \mathbb{F}$. In other words, assigning a matrix to a linear transformation is a linear transformation.

Exercise 7.41. Let $\mathcal{P}_n(\mathbb{R})$ denote the space of polynomials with real coefficients of degree n or less, and let $D : \mathcal{P}_n(\mathbb{R}) \to \mathcal{P}_{n-1}(\mathbb{R})$ be the derivative map. That is, $D(f) = f'$.

(i) Show that D is linear, and

(*ii*) find the matrix of D for the bases of $\mathcal{P}_n(\mathbb{R})$ and $\mathcal{P}_{n-1}(\mathbb{R})$ of your choice.

(*iii*) Find the matrix of $D^4 - 2D$ with respect to the same bases.

Exercise 7.42. Let $A : \mathcal{P}_{n-1}(\mathbb{R}) \to \mathcal{P}_n(\mathbb{R})$ be the map

$$A(f) = \int_0^x f(t)\mathrm{d}t.$$

Note that A stands for antiderivative.

(*i*) Show that A is linear and find the matrix of A with respect to the bases of $\mathcal{P}_{n-1}(\mathbb{R})$ and $\mathcal{P}_n(\mathbb{R})$ you used in Exercise 7.41.

(*ii*) Find the matrix of AD and the matrix of DA, where D is the derivative map.

Exercise 7.43. Let $T : V \to V$ be a linear transformation, and suppose $\mathcal{B}$ and $\mathcal{B}'$ are two bases of V. Express $\mathcal{M}_{\mathcal{B}}^{\mathcal{B}'}(T)$ in terms of $\mathcal{M}_{\mathcal{B}'}^{\mathcal{B}}(T)$, $\mathcal{M}_{\mathcal{B}'}^{\mathcal{B}}$ and $\mathcal{M}_{\mathcal{B}}^{\mathcal{B}'}$.

Exercise 7.44. Let $S : V \to V$ and $T : V \to V$ be linear transformations, and let $\mathcal{B}$ be a basis of V. Find $\mathcal{M}_{\mathcal{B}}^{\mathcal{B}}(T \circ S)$ in terms of $\mathcal{M}_{\mathcal{B}}^{\mathcal{B}}(S)$ and $\mathcal{M}_{\mathcal{B}}^{\mathcal{B}}(T)$.

Exercise 7.45. Let $S : V \to W$ and $T : W \to Y$ be linear transformations, and let $\mathcal{B}$, $\mathcal{B}'$ and $\mathcal{B}''$ be bases of V, W and Y respectively. Is it true or false that $\mathcal{M}_{\mathcal{B}''}^{\mathcal{B}}(TS) = \mathcal{M}_{\mathcal{B}''}^{\mathcal{B}'}(T)\mathcal{M}_{\mathcal{B}'}^{\mathcal{B}}(S)$?

Exercise 7.46. Let $\mathcal{B}$ be the basis $\begin{pmatrix}1\\0\\0\end{pmatrix}, \begin{pmatrix}1\\1\\0\end{pmatrix}, \begin{pmatrix}1\\1\\1\end{pmatrix}$ of $\mathbb{R}^3$, and let $T : \mathbb{R}^3 \to \mathbb{R}^3$ be the linear transformation such that

$$\mathcal{M}_{\mathcal{B}}^{\mathcal{B}}(T) = \begin{pmatrix} 2 & -3 & 3 \\ -2 & 1 & -7 \\ 5 & -5 & 7 \end{pmatrix}.$$

(*i*) Find $T \begin{pmatrix}1\\1\\1\end{pmatrix}$.

(*ii*) Calculate the matrix of T with respect to the standard basis $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$ of $\mathbb{R}^3$.

Exercise 7.47. Show that matrix similarity is an equivalence relation on $\mathbb{F}^{n\times n}$.

Exercise 7.48. Show that two similar matrices have the same eigenvalues by showing how their eigenvectors related.

Exercise 7.49. Prove Proposition 7.15.

7.5 Further Results on Linear Transformations

The purpose of this section is to develop some more of the basic facts about linear transformations.

7.5.1 The Kernel and Image of a Linear Transformation

Let $T : V \to W$ be a linear transformation. Recall that the range of T as a transformation is defined to be $T(V)$. In the case T is a linear transformation, we make the following definition.

Definition 7.5. The range of a linear transformation T is called the *image* of T and is denoted by $\text{im}(T)$. The *kernel* of T, is defined to be the set

$$\ker(T) = \{\mathbf{v} \in V \mid T(\mathbf{v}) = \mathbf{0}\}.$$

Here are two basic examples, which are quite different in nature.

Example 7.16. Suppose $V = \mathbb{F}^n$, $W = \mathbb{F}^m$ and $T = T_A$, i.e. $T(\mathbf{x}) = A\mathbf{x}$. Then $\ker(T) = \mathcal{N}(A)$, so the problem of finding $\ker(T)$ is simply to find the solution space of an $m \times n$ homogeneous linear system. Furthermore, $\text{im}(T) = \text{col}(A)$, since both $\text{im}(T)$ and $\text{col}(A)$ consist of all linear combinations of the columns of A. Hence $\ker(T)$ and $\text{im}(T)$ are subspaces of V and W. Moreover, $\dim \text{im}(T) = \text{rank}(A)$ and $\dim \ker(T) = n - \text{rank}(A)$.

Example 7.17. Let $\mathcal{P}(\mathbb{R})$ denote the space of polynomials with real coefficients. The derivative $D(f)$ of a polynomial is also a polynomial, and $D : \mathcal{P}(\mathbb{R}) \to \mathcal{P}(\mathbb{R})$ is a linear transformation. The kernel of D is the subspace $\mathbb{R}1$ of constant polynomials, while $\text{im}(D) = \mathcal{P}(\mathbb{R})$.

Example 7.18. Now let $C^\infty(\mathbb{R})$ denote the space of all real valued functions f on $\mathbb{R}$ such that $f^{(n)} = \dfrac{d^n f}{dx^n}$ exists for all $n > 0$. A linear differential operator of order n on $C^\infty(\mathbb{R})$ is a transformation $T : C^\infty(\mathbb{R}) \to C^\infty(\mathbb{R})$ of the form

$$T(f) = f^{(n)} + a_1 f^{(n-1)} + \cdots + a_{n-1} f' + a_n f,$$

where $a_1, \ldots, a_n \in C^\infty(\mathbb{R})$. In particular, if $a_1, \ldots, a_n$ are just real constants, then one says that T has constant coefficients. It is straightforward to check that a linear differential operator is a linear transformation. A basic result on differential equations says that if T is a linear differential operator of order n with constant coefficients, then $\dim \ker(T) = n$. The solution are of the form $e^{rx} f(x)$ where r is a (real or complex) root of

$$r^n + a_1 r^{(n-1)} + \cdots + a_{n-1} r + a_n = 0$$

and $f \in \mathcal{P}(\mathbb{R})$. The situation where T is an operator on an infinite dimensional vector space with a finite dimensional kernel is important in the theory of differential equations.

The following Proposition lays some of the groundwork for Theorem 7.17, which is our next main result.

Proposition 7.16. *The kernel and image of a linear transformation $T : V \to W$ are subspaces of V and W are respectively. Furthermore, T is one to one if and only if* $\ker(T) = \{\mathbf{0}\}$.

Proof. The first assertion is straightforward, so we will omit the proof. Suppose that T is one to one. Then, since $T(\mathbf{0}) = \mathbf{0}$, $\ker(T) = \{\mathbf{0}\}$. Conversely, suppose $\ker(T) = \{\mathbf{0}\}$. If $\mathbf{x}, \mathbf{y} \in V$ are such that $T(\mathbf{x}) = T(\mathbf{y})$, then

$$T(\mathbf{x}) - T(\mathbf{y}) = T(\mathbf{x} - \mathbf{y}) = \mathbf{0}.$$

Thus, $\mathbf{x} - \mathbf{y} \in \ker(T)$, so $\mathbf{x} - \mathbf{y} = \mathbf{0}$. Hence $\mathbf{x} = \mathbf{y}$, so T is one to one. □

The main result on the kernel and image is the following.

Theorem 7.17. *Suppose $T : V \to W$ is a linear transformation where* $\dim V = n$. *Then*

$$\dim \ker(T) + \dim \operatorname{im}(T) = n. \tag{7.18}$$

In fact, if $\dim \ker(T) = k > 0$, *then there exists a basis $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ of V so that $\mathbf{v}_1, \mathbf{v}_2, \ldots \mathbf{v}_k$ is a basis of* $\ker(T)$, *and, provided* $\ker(T) \neq V$, *$T(\mathbf{v}_{k+1}), T(\mathbf{v}_{k+2}), \ldots, T(\mathbf{v}_n)$ is a basis of* $\operatorname{im}(T)$.

Proof. Choose any basis $\mathbf{v}_1, \mathbf{v}_2, \ldots \mathbf{v}_k$ of $\ker(T)$. If $\ker(T) = V$, clearly $\operatorname{im}(T) = \{\mathbf{0}\}$, and we're through. Otherwise, we may, by the Dimension Theorem, extend the basis of $\ker(T)$ to any basis $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ of V. I claim that $T(\mathbf{v}_{k+1}), T(\mathbf{v}_{k+2}), \ldots, T(\mathbf{v}_n)$ span $\operatorname{im}(T)$. Indeed, if $\mathbf{w} \in \operatorname{im}(T)$, then $\mathbf{w} = T(\mathbf{v})$ for some $\mathbf{v} \in V$. But $\mathbf{v} = \sum a_i \mathbf{v}_i$, so

$$T(\mathbf{v}) = \sum_{i=1}^{n} a_i T(\mathbf{v}_i) = \sum_{i=k+1}^{n} a_i T(\mathbf{v}_i),$$

since $\mathbf{v}_1, \mathbf{v}_2, \ldots \mathbf{v}_k$ lie in $\ker(T)$. To see that $T(\mathbf{v}_{k+1}), T(\mathbf{v}_{k+2}), \ldots, T(\mathbf{v}_n)$ are independent, let

$$\sum_{i=k+1}^{n} a_i T(\mathbf{v}_i) = \mathbf{0}.$$

Then $T(\sum_{i=k+1}^{n} a_i \mathbf{v}_i) = \mathbf{0}$, so $\sum_{i=k+1}^{n} a_i \mathbf{v}_i \in \ker(T)$. But if $\sum_{i=k+1}^{n} a_i \mathbf{v}_i \neq \mathbf{0}$, the $\mathbf{v}_i$ $(1 \leq i \leq n)$ cannot be a basis, since every vector in $\ker(T)$ is a

linear combination of the $\mathbf{v}_i$ with $1 \leq i \leq k$. This shows that $\sum_{i=k+1}^{n} a_i \mathbf{v}_i = \mathbf{0}$, so each $a_i = 0$. Therefore, $T(\mathbf{v}_{k+1}), T(\mathbf{v}_{k+2}), \ldots, T(\mathbf{v}_n)$ give a basis of $\mathrm{im}(T)$. □

Theorem 7.17 should be viewed as the final version of the basic principle that in a linear system, the number of free variables plus the number of corner variables is the total number of variables, which was originally formulated in Chapter 2 (see (2.6)).

7.5.2 Vector Space Isomorphisms

One application of the existence result Proposition 7.1 is that if V and W are two vector spaces over $\mathbb{F}$ having the same dimension, then there exists a one to one linear transformation $T : V \to W$ such that $T(V) = W$. Hence, two finite dimensional vector spaces over the same field with the same dimension are in a sense indistinguishable. A linear transformation $S : V \to W$ which is both one to one and onto (i.e. bijective) is said to be an *isomorphism* between V and W.

Proposition 7.18. *Two finite dimensional vector spaces V and W over the same field are isomorphic if and only if they have the same dimension.*

Proof. We leave this as an exercise. □

Example 7.19 ($L(V, W)$)**.** Suppose $\dim V = n$ and $\dim W = m$. We know that the dimension of the space $L(V, W)$ of linear transformations with domain V and target W is mn. We also know $\dim \mathbb{F}^{m\times n} = mn$. Let us find an isomorphism from $L(V, W)$ to $\mathbb{F}^{m\times n}$. In fact, if we choose bases $\mathcal{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ for V and $\mathcal{B}' = \{\mathbf{w}_1, \ldots, \mathbf{w}_m\}$ of W, we get a mapping $\Phi : L(V, W) \to \mathbb{F}^{m\times n}$ by assigning to any $T \in L(V, W)$ its matrix $\mathcal{M}_{\mathcal{B}}^{\mathcal{B}}(T)$. I claim Φ is an isomorphism. In fact, Φ is clearly linear (see Exercise 7.40). It is one to one, since if $\Phi(T) = \mathcal{M}_{\mathcal{B}}^{\mathcal{B}'}(T) = O$, then $T = O$. It's also clear that Φ is onto. (For example, we could note that since $\dim \mathbb{F}^{m\times n} = mn$ and $\dim \ker(T) = 0$, $\dim \mathrm{im}(T) = mn$ as well (why?). Therefore, Φ is indeed an isomorphism.

7.5.3 Complex Linear Transformations

The purpose of this section is to explain when a linear transformation $S : \mathbb{R}^{2n} \to \mathbb{R}^{2n}$ is actually a complex linear transformation from $\mathbb{C}^n$ to itself. Every vector $\mathbf{z}$ in complex n-space $\mathbb{C}^n$ is uniquely decomposable as $\mathbf{z} = \mathbf{x}+i\mathbf{y}$, where $\mathbf{x}$ and $\mathbf{y}$ are both in $\mathbb{R}^n$. The transformation $T : \mathbb{C}^n \to \mathbb{R}^n \times \mathbb{R}^n$

defined by $T(\mathbf{z}) = (\mathbf{x}, \mathbf{y})^T$ is an isomorphism of real vector spaces, provided we think of $\mathbb{C}^n$ as a real vector space by restricting the scalars to the real numbers. For example, if $n = 2$, then

$$T\begin{pmatrix} a_1 + ib_1 \\ a_2 + ib_2 \end{pmatrix} = \begin{pmatrix} a_1 \\ a_2 \\ b_1 \\ b_2 \end{pmatrix}.$$

Now suppose $S : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n \times \mathbb{R}^n$ is a real linear transformation. Our plan is to determine when S complex linear in the sense that there is a complex linear transformation $\mathcal{S} : \mathbb{C}^n \to \mathbb{C}^n$ such that $\mathcal{S} = T^{-1}ST$.

The answer is quite simple. Suppose $\mathcal{S}$ is complex linear. Then $\mathcal{S}(i\mathbf{z}) = i\mathcal{S}(\mathbf{z})$ for all $\mathbf{z} \in \mathbb{C}^n$. Now $T(i\mathbf{z}) = T(-\mathbf{y} + i\mathbf{x}) = (-\mathbf{y}, \mathbf{x})$, so this suggests we consider the real linear transformation $J : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n \times \mathbb{R}^n$ defined by $J(\mathbf{x}, \mathbf{y}) = (-\mathbf{y}, \mathbf{x})$. Now let $S : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n \times \mathbb{R}^n$ be a real linear transformation, and define a real linear transformation $\mathcal{S} : \mathbb{C}^n \to \mathbb{C}^n$ by $\mathcal{S} = T^{-1}ST$. Then we have

Proposition 7.19. *The transformation $\mathcal{S} : \mathbb{C}^n \to \mathbb{C}^n$ is $\mathbb{C}$-linear if and only if $JS = SJ$.*

Proof. Suppose $\mathbf{z} = \mathbf{x} + i\mathbf{y}$. Note first that $\mathcal{S}$ is $\mathbb{C}$-linear if and only if $\mathcal{S}(i\mathbf{z}) = i\mathcal{S}(\mathbf{z})$ (Exercise 7.78). So it sufficies to show $\mathcal{S}(i\mathbf{z}) = i\mathcal{S}(\mathbf{z})$ if and only if $SJ = JS$. Write $S(\mathbf{x}, \mathbf{y}) = (S_1(\mathbf{x}, \mathbf{y}), S_2(\mathbf{x}, \mathbf{y}))$. Since $T\mathcal{S} = \mathcal{S}T$ and T is injective, we can infer $\mathcal{S}(\mathbf{z}) = S_1(\mathbf{x}, \mathbf{y}) + iS_2(\mathbf{x}, \mathbf{y})$. Now,

$$\begin{aligned} \mathcal{S}(i\mathbf{z}) &= T^{-1}ST(-\mathbf{y} + i\mathbf{x}) \\ &= T^{-1}S(-\mathbf{y}, \mathbf{x}) \\ &= T^{-1}SJ(\mathbf{x}, \mathbf{y}) \end{aligned}$$

On the other hand,

$$\begin{aligned} i\mathcal{S}(\mathbf{z}) &= -S_2(\mathbf{x}, \mathbf{y}) + iS_1(\mathbf{x}, \mathbf{y}) \\ &= T^{-1}(-S_2(\mathbf{x}, \mathbf{y}), S_1(\mathbf{x}, \mathbf{y})) \\ &= T^{-1}J(S_1(\mathbf{x}, \mathbf{y}), S_2(\mathbf{x}, \mathbf{y})) \\ &= T^{-1}JS(\mathbf{x}, \mathbf{y}) \end{aligned}$$

Thus $\mathcal{S}(i\mathbf{z}) = i\mathcal{S}(\mathbf{z})$ if and only if $SJ(\mathbf{x}, \mathbf{y}) = JS(\mathbf{x}, \mathbf{y})$, which is just what we needed to show.

□

Example 7.20. Let's illustrate the above Proposition in the case n=1. Here, $\mathbb{C} = \mathbb{R} \times \mathbb{R}$. Thus $J : \mathbb{R}^2 \to \mathbb{R}^2$ is the transformation

$$J\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} -y \\ x \end{pmatrix}.$$

Then the matrix of J with respect to to standard basis is

$$M(J) = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

Suppose $z = x + iy$ and $\mathcal{S}(z) = S_1(x, y) + iS_2(x, y)$. (Here we write vectors as rows.) Now let $S(x, y) = (ax + by, cx + dy)^T$, so the matrix of S with respect to to standard basis is

$$M(S) = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

Then $JS(x, y) = SJ(x, y)$ if and only if

$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

Carrying out the multiplication, we see that this means

$$\begin{pmatrix} -c & -d \\ a & b \end{pmatrix} = \begin{pmatrix} b & -a \\ d & -c \end{pmatrix}.$$

This is equivalent to $a = d$ and $b = -c$. Hence a linear transformation $S : \mathbb{R}^2 \to \mathbb{R}^2$ defines a $\mathbb{C}$-linear transformation if and only if $S = T_M$, where M has the form

$$M = \begin{pmatrix} a & -b \\ b & a \end{pmatrix}.$$

The astute reader will notice that

$$M = \begin{pmatrix} a & -b \\ b & a \end{pmatrix} = (\sqrt{a^2 + b^2})R_\theta,$$

where $\theta = \cos^{-1}(a/\sqrt{a^2 + b^2})$. Hence, every complex linear map of $\mathbb{C} = \mathbb{R}^2$ to itself is a multiple of a rotation.

More generally, if $S : \mathbb{R}^{2n} = \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^{2n} = \mathbb{R}^n \times \mathbb{R}^n$ is any linear transformation, we can analogously write the matrix of S in block form as

$$M(S) = \begin{pmatrix} A & B \\ C & D \end{pmatrix},$$

where $A,\ B,\ C,\ D \in \mathbb{R}^{n\times n}$. Now the matrix $M(J)$ of J is

$$M(J) = \begin{pmatrix} O & -I_n \\ I_n & O \end{pmatrix}.$$

Thus, imitating the previous example, we have $SJ = JS$ if and only if $M(S)M(J) = M(J)M(S)$ if and only if

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}\begin{pmatrix} O & -I_n \\ I_n & O \end{pmatrix} = \begin{pmatrix} O & -I_n \\ I_n & O \end{pmatrix}\begin{pmatrix} A & B \\ C & D \end{pmatrix}.$$

As above, this means that

$$\begin{pmatrix} B & -A \\ D & -C \end{pmatrix} = \begin{pmatrix} -C & -D \\ A & B \end{pmatrix},$$

whence $A = D$ and $C = -B$. Thus $SJ = JS$ if and only if

$$M(S) = \begin{pmatrix} A & B \\ -B & A \end{pmatrix}.$$

Thus we have

Proposition 7.20. *A linear transformation $S : \mathbb{R}^{2n} = \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^{2n} = \mathbb{R}^n \times \mathbb{R}^n$ with matrix $M(S) = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$, with respect to the standard basis, defines a complex linear transformation $\mathcal{S} = T^{-1}ST : \mathbb{C}^n \to \mathbb{C}^n$ if and only if $C = -B$ and $A = D$. In this case,*

$$M(S) = \begin{pmatrix} A & B \\ -B & A \end{pmatrix}.$$

Exercises

Exercise 7.50. Suppose $T : V \to V$ is a linear transformation, where V is finite dimensional over $\mathbb{F}$. Find the relationship between between $\mathcal{N}(\mathcal{M}_{\mathcal{B}}^{\mathcal{B}}(T))$ and $\mathcal{N}(\mathcal{M}_{\mathcal{B}'}^{\mathcal{B}'}(T))$, where $\mathcal{B}$ and $\mathcal{B}'$ are any two bases of V.

Exercise 7.51. Describe both the column space and null space of the matrix

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

Exercise 7.52. Prove the first part of Proposition 7.16 using only the basic definition of a linear transformation. That is, show that the kernel and image of a linear transformation T are subspaces of the domain and target of T respectively. Also, show that if V is a finite dimensional vector space , then $\dim T(V) \leq \dim V$.

Exercise 7.53. Let A and B be $n \times n$ matrices.

(a) Explain why the null space of A is contained in the null space of BA.

(b) Explain why the column space of A contains the column space of AB.

(c) If $AB = O$, show that $\text{col}(B)$ is a subspace of $\mathcal{N}(A)$.

Exercise 7.54. Assume $\mathbf{a}, \mathbf{b} \in \mathbb{R}^3$ and $\mathbf{a} \neq \mathbf{0}$. Describe, in terms of the transformation $C_{\mathbf{a}}$,

(i) $\{\mathbf{x} \in \mathbb{R}^3 \mid \mathbf{a} \times \mathbf{x} = \mathbf{b}\}$, and

(ii) $\{\mathbf{y} \in \mathbb{R}^3 \mid \mathbf{a} \times \mathbf{x} = \mathbf{y}\ \exists\ \mathbf{x} \in \mathbb{R}^3\}$.

Exercise 7.55. Suppose A is any 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, e.g. $\mathbb{F}_2$ or $\mathbb{C}$?

Exercise 7.56. Show that if A is any symmetric real $n \times n$ matrix, then $\text{col}(A) \cap \mathcal{N}(A) = \{\mathbf{0}\}$.

Exercise 7.57. Suppose A is a square matrix over an arbitrary field. Show that $A^2 = O$ if and only if $\text{col}(A) \subset \mathcal{N}(A)$.

Exercise 7.58. Find an example of a 2×2 matrix A over $\mathbb{F}_2$ such that $\text{col}(A) = \mathcal{N}(A)$.

Exercise 7.59. Find an example of a 3×3 matrix A over $\mathbb{F}_2$ such that $\mathrm{col}(A) = \mathcal{N}(A)$ or explain why no such example exists.

Exercise 7.60. 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 7.61. Suppose A is a symmetric real matrix so that $A^2 = O$. Show that $A = O$. In fact, show that $\mathrm{col}(A) \cap \mathcal{N}(A) = \{\mathbf{0}\}$.

Exercise 7.62. Find a non zero 2×2 symmetric matrix A over $\mathbb{C}$ such that $A^2 = O$. Show that no such a matrix exists if we replace $\mathbb{C}$ by $\mathbb{R}$.

Exercise 7.63. For two vectors $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{R}^n$, the dot product $\mathbf{x} \cdot \mathbf{y}$ can be expressed as $\mathbf{x}^T\mathbf{y}$. Use this to prove that for any real matrix A, A^TA and A have the same nullspace. Conclude that A^TA and A have the same rank. (Hint: consider $\mathbf{x}^TA^TA\mathbf{x}$.)

Exercise 7.64. Consider the subspace W of $\mathbb{R}^4$ spanned by $(1, 1, -1, 2)^T$ and $(1, 1, 0, 1)^T$. Find a system of homogeneous linear equations whose solution space is W.

Exercise 7.65. What are the null space and image of

(i) a projection $P_{\mathbf{b}} : \mathbb{R}^2 \to \mathbb{R}^2$,

(ii) the cross product map $T(\mathbf{x}) = \mathbf{x} \times \mathbf{v}$.

Exercise 7.66. What are the null space and image of a reflection $H_{\mathbf{b}} : \mathbb{R}^2 \to \mathbb{R}^2$. Ditto for a rotation $\mathcal{R}_\theta : \mathbb{R}^2 \to \mathbb{R}^2$.

Exercise 7.67. Ditto for the projection $P : \mathbb{R}^3 \to \mathbb{R}^3$ defined by

$$P \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix}.$$

Exercise 7.68. 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.

(a) Show that $\mathcal{N}(A)$ is either a line through the origin or a plane containing the origin.

(b) 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 7.69. Let $T : \mathbb{F}^n \to \mathbb{F}^n$ be a linear transformation such that $\ker(T) = \mathbf{0}$ and $\text{im}(T) = \mathbb{F}^n$. Prove the following statements.

(a) There exists a transformation $S : \mathbb{F}^n \to \mathbb{F}^n$ 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.

(b) Show that in addition, S is also a linear transformation.

(c) If A is the matrix of T and B is the matrix of S, then $BA = AB = I_n$.

Exercise 7.70. Prove Proposition 7.18. That is, show two finite dimensional vector spaces V and W over the same field are isomorphic if and only if they have the same dimension.

Exercise 7.71. Let $F : \mathbb{R}^2 \to \mathbb{R}^2$ be the linear transformation given by $F\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} y \\ x - y \end{pmatrix}$.

(a) Show that F has an inverse and find it.

(b) Verify that if A is the matrix of F, then $AB = BA = I_2$ if B is a matrix of the inverse of F.

Exercise 7.72. Let $S : \mathbb{F}^n \to \mathbb{F}^m$ and $T : \mathbb{F}^m \to \mathbb{F}^p$ be two linear transformations both of which are one to one. Show that the composition $T \circ S$ is also one to one. Conclude that if $A \in \mathbb{F}^{m\times n}$ has $\mathcal{N}(A) = \{\mathbf{0}\}$ and $B \in \mathbb{F}^{n\times p}$ has $\mathcal{N}(B) = \{\mathbf{0}\}$, then $\mathcal{N}(BA) = \{\mathbf{0}\}$ too.

Exercise 7.73. Show that if $T : V \to W$ is a linear transformation that sends a basis of V to a basis of W, then T is an isomorphism.

Exercise 7.74. Let W be a subspace of a finite dimensional vector space V. Show that there exists a linear transformation $T : V \to V$ such that $\ker(T) = W$.

Exercise 7.75. Suppose V and W are finite dimensional vector spaces over $\mathbb{F}$ which have the same dimension, and let $T : V \to W$ be linear. Show that T is an isomorphism if and only if either T is injective or T is surjective.

Exercise 7.76. Suppose $T : V \to W$ is linear. Show that there exists a unique linear transformation $\overline{T} : V/\ker(T) \to W$ such that $\overline{T}(\mathbf{v} + \ker(T)) = T(\mathbf{v})$.

Exercise 7.77. 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 the following.

(a) TS is injective if and only if S is injective and $\text{im}(S) \cap \ker(T) = \{\mathbf{0}\}$.

(*b*) TS is surjective if and only if T is surjective and $V = \mathrm{im}(S) + \ker(T)$.

(*c*) Concude that TS is an isomorphism if and only if S is injective, T is surjective and $\dim U = \dim W$.

Exercise 7.78. Suppose $\mathcal{S} : \mathbb{C}^n \to \mathbb{C}^n$ is real linear. Show that $\mathcal{S}$ is complex linear if and only if $\mathcal{S}(i\mathbf{z}) = i\mathcal{S}(\mathbf{z})$ for all $\mathbf{z} \in \mathbb{C}^n$.

Exercise 7.79. Let $\mathcal{S} : \mathbb{R}^2 \to \mathbb{R}^2$ be the real linear transformation defined by $\mathcal{S}(z) = \overline{z}$. Determine whether or not $\mathcal{S}$ is complex linear.

Exercise 7.80. Prove Proposition 7.18.

Exercise 7.81. Let $W = \mathbb{F}^{n\times n}$, and for each $A \in W$, define a transformation $\mathrm{ad}_A : W \to W$ by $\mathrm{ad}_A(X) = AX - XA$.

(*i*) Show that ad_A is indeed linear for each $A \in \mathbb{F}^{n\times n}$.

(*ii*) Describe the kernel of ad_A for any A.

(*iii*) Suppose $n = 2$ and $A = \mathrm{diag}(a_1, a_2)$. Compute the matrix of ad_A with respect to the basis E_{ij} $(1 \le i, j \le 2)$ of W. Note that this matrix is 4×4 .

(*iv*) Find the kernel of ad_A if $a_1 \neq a_2$.

Exercise 7.82. Prove Proposition 7.18.

7.6 Summary

A linear transformation between two vector spaces V and W over the same field (called the domain and target) is a transformation $T : V \to W$ which has the property that $T(a\mathbf{x} + b\mathbf{y}) = aT(\mathbf{x}) + bT(\mathbf{y})$ for all $\mathbf{x}, \mathbf{y}$ in V and a, b in $\mathbb{F}$. In other words, a transformation is linear if it preserves all linear combinations. Linear transformations are a way of using the linear properties of V to study W. The set of all linear transformations with domain V and target W is another vector space over $\mathbb{F}$ denoted by $L(V, W)$. For example, if V and W are real inner product spaces, the linear transformations which preserve the inner product are called orthogonal.

Matrix theory is enters into the theory of linear transformations because every linear transformation can represented as a matrix. In particular, if $V = \mathbb{F}^n$ and $W = \mathbb{F}^m$, then a linear transformation $T : V \to W$ is nothing but an $m \times n$ matrix over $\mathbb{F}$, i.e. an element of $\mathbb{F}^{m\times n}$ which we will usually denote by $M(T)$. Conversely, every element of $A \in \mathbb{F}^{m\times n}$ defines such a linear transformation$T_A : \mathbb{F}^n \to \mathbb{F}^m$. If V and W are finite dimensional, then whenever we are given bases $\mathcal{B}$ of V and $\mathcal{B}'$ of W, we can associate a unique matrix $\mathcal{M}_{\mathcal{B}'}^{\mathcal{B}}(T)$ to T. There are certain rules for manipulating these matrices explained in the text. They amount to the rule $M(T \circ S) = M(T)M(S)$ when $S : \mathbb{F}^p \to \mathbb{F}^n$ and $T : \mathbb{F}^n \to \mathbb{F}^m$ are both linear. If we express a linear transformation $T : V \to V$ in terms of two bases $\mathcal{B}$ and $\mathcal{B}'$ of V, then $\mathcal{M}_{\mathcal{B}'}^{\mathcal{B}'}(T) = P\mathcal{M}_{\mathcal{B}}^{\mathcal{B}}(T)P^{-1}$ where $P = \mathcal{M}_{\mathcal{B}}^{\mathcal{B}'}$ is a certain change of basis matrix. This means that two matrices representing the same linear transformation $T : V \to V$ are similar. Thus, two similar matrices represent the same single linear transformation.

There are two fundamental spaces associated with a linear transformation: its kernel $\ker(T)$ and its image $\mathrm{im}(T)$. The kernel and image of a linear transformation T correspond to the null space and column space of any matrix representing T. The fundamental relationship from Chapter 2 which said that in a linear system, the number of variables equals the number of free variables plus the number of corner variables takes its final form for a linear transformation $T : V \to W$ in the identity $\dim V = \dim \ker(T) + \dim \mathrm{im}(T)$. If $\dim \ker(T) = 0$ and $\dim \mathrm{im}(T) = \dim W$, then T is one to one and onto. In this case, it is called an isomorphism. linear transformation $T : V \to W$ taking arbitrarily preassigned values

One of the main general questions about linear transformations is this: when is a linear transformation $T : V \to V$ semi-simple? That is, when does there exist a basis $\mathcal{B}$ of V for which $\mathcal{M}_{\mathcal{B}}^{\mathcal{B}}(T)$ is diagonal. Such a basis is called an eigenbasis. Put another way, when is $\mathcal{M}_{\mathcal{B}}^{\mathcal{B}}(T)$ similar to a diagonal

matrix? We will provide an answer in Chapter 13 for the case when $\mathbb{F}$ is algebraically closed.

Chapter 8

An Introduction to the Theory of Determinants

The determinant of a square matrix A is a fundamental scalar associated to A with a long mathematical history. Most students first encounter it in the statement of Cramer's Rule, which has as a special case, the formula for the inverse of A (see Section 8.3.3 and Exercise 2.3). The determinant seems to have first appeared in a paper of Liebniz published in 1683, and it has since acquired a long and distinguished list of applications. We will use it in Chapter 10 to define the characteristic polynomial of A, which is the basic tool for finding A's eigenvalues.

The purpose of this chapter is to introduce the determinant and derive its basic properties.

8.1 The Definition of the Determinant

Let $\mathbb{F}$ denote an arbitrary field, and assume $A \in \mathbb{F}^{n\times n}$. The purpose of this Section is to define a scalar $\det(A) \in \mathbb{F}$, called the determinant of A. This scalar has a number of remarkable properties. For example, if $A, B \in \mathbb{F}^{n\times n}$, then $\det(AB) = \det(A)\det(B)$. Moreover, $\det(A) \neq 0$ if and only if A is nonsingular, i.e. A^{-1} exists. In particular, the nullspace of any $A \in \mathbb{F}^{n\times n}$ has positive dimension if and only if $\det(A) = 0$. This is a standard criterion for determining when a matrix is singular.

The definition of $\det(A)$ given in (8.2) below is a sum with $n!$ terms, so, at first glance, any hope of finding a general method for computing it is useless. Fortunately, as we will eventually see, row operations enable $\det(A)$ to be computed far more efficiently.

8.1.1 The 1×1 and 2×2 Cases

If A is 1×1, say $A = (a)$, we will put $\det(A) = a$. This definition clearly gives the two properties we want. Namely, if A and B are 1×1, then $\det(AB) = \det(A)\det(B)$, and A^{-1} exists if and only if $\det(A) \neq 0$.

The 2×2 case requires more cleverness. Here we put

$$\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc.$$

To see what this quantity measures, note that $ad - bc = 0$ if and only if the rows of A are proportional. Thus, A has rank zero or one if and only if $\det(A) = 0$ and rank 2 if and only if $ad - bc \neq 0$. Thus, A is invertible if and only if $\det(A) \neq 0$. In fact, recall that if A^{-1} exists, then

$$A^{-1} = \frac{1}{(ad - bc)} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

This is a special case of Cramer's Rule.

Proposition 8.1. *For any field $\mathbb{F}$, the determinant function $\det : \mathbb{F}^{2\times 2} \to \mathbb{F}$ has the following properties:*

(1) $\det(A) \neq 0$ *if and only if A is invertible;*

(2) $\det(I_2) = 1$*; and*

(3) $\det(AB) = \det(A)\det(B)$ *for any $A, B \in \mathbb{F}^{2\times 2}$.*

Proof. We indicated the proof of (1) above. Statement (2) is obvious, and (3) can be checked by a direct calculation. □

8.1.2 Some Combinatorial Preliminaries

Unlike many situations where we need to generalize a concept from the two dimensional case to the n dimensional case, the definition of $\det(A)$ for 2×2 matrices gives only the slightest hint of what to do for $n \times n$ matrices so that the three properties listed Proposition 8.1 still hold. The general definition will require us to introduce some elementary combinatorics.

Let X_n denote $\{1, 2, \ldots, n\}$. A bijection $\sigma : X_n \to X_n$ is called a *permutation* on n letters, and the set of all permutations on n letters is denoted by $S(n)$. One usually calls $S(n)$ the symmetric group on n letters. The study of $S(n)$ (for any n) is one of the basic topics in the subject of combinatorics. We will presently see that $S(n)$ can be identified with the matrix group of $n \times n$ permutation matrices. To begin, we state some of the basic properties of $S(n)$.

Lemma 8.2. *The symmetric group $S(n)$ has exactly $n!$ elements. Moreover, if $\pi, \sigma \in S(n)$, then the composition $\sigma\pi \in S(n)$. Finally, the inverse σ^{-1} of any $\sigma \in S(n)$ is also in $S(n)$.*

Proof. The first statement is an application of elementary combinatorics. The other two claims are obvious consequences of the definition of a bijection. □

In order to define the determinant, we also need to define the signature of a permutation, which goes as follows.

Definition 8.1. Let $\sigma \in S(n)$. Define the *signature* of σ to be the rational number

$$\operatorname{sgn}(\sigma) = \prod_{i<j} \frac{\sigma(i) - \sigma(j)}{i - j}.$$

Note that $\operatorname{sgn}(\sigma)$ is a nonzero rational number since $\sigma(i) \neq \sigma(j)$ if $i \neq j$.

Example 8.1. The identity permutation of $S(n)$ will be denoted by id_n. Since $id_n(j) = j$ for all $j \in X_n$, $\operatorname{sgn}(id_n) = 1$. To see another example, let $\sigma(1) = 2$, $\sigma(2) = 1$ and let $\sigma(i) = i$ if $i > 2$. Then if $i < j$,

$$\frac{\sigma(i) - \sigma(j)}{i - j} > 0$$

except when $i = 1$ and $j = 2$. Thus $\operatorname{sgn}(\sigma) < 0$. In fact, as we will now see, $\operatorname{sgn}(\sigma) = -1$.

Proposition 8.3. *For any $\sigma \in S(n)$, $\operatorname{sgn}(\sigma) = \pm 1$.*

Proof. Since σ is a bijection of X_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, σ defines a bijection from the set $T = \{(i, j) \mid 1 \leq i, j \leq n,\ i \neq j\}$ to itself by $(i, j) \mapsto (\sigma(i), \sigma(j))$. (Reason: it is clear that this mapping is injective, hence it's also surjective by the Pigeon Hole Principle.) Thus

$$\prod_{i \neq j} (\sigma(i) - \sigma(j))^2 = \prod_{i \neq j} (i - j)^2,$$

so $\operatorname{sgn}(\sigma)^2 = 1$. □.

Notice that if $i < j$, then

$$\frac{\sigma(i) - \sigma(j)}{i - j} > 0$$

exactly when $\sigma(i) < \sigma(j)$. Therefore

$$\operatorname{sgn}(\sigma) = (-1)^{m(\sigma)},$$

where

$$m(\sigma) = |\{(i, j) \mid i < j,\, \sigma(i) > \sigma(j)\}|.$$

Most facts about permutations can be deduced by considering transpositions. A *transposition* is an element σ of $S(n)$ which interchanges two elements of X_n and fixes all the others. The transposition which interchanges $a \neq b$ in X_n is denoted by σ_{ab}.

Example 8.2. For example, σ_{12} interchanges 1 and 2 and fixes every integer between 3 and n . I claim $m(\sigma_{12}) = 1$. For, the only pair (i, j) such that $i < j$ for which $\sigma_{12}(i) > \sigma_{12}(j)$ is the pair $(1, 2)$. Hence $\operatorname{sgn}(\sigma_{12}) = -1$, Note that σ_{12} is the permutation σ of the previous example.

We will need the explicit value for the signature of an arbitrary transposition. This is one of the results in the main theorem on the signature, which we now state and prove.

Theorem 8.4. *The signature mapping* $\operatorname{sgn} : S(n) \to \{\pm 1\}$ *satisfies the following properties:*

(1) *for all* $\sigma, \tau \in S(n)$, $\operatorname{sgn}(\tau\sigma) = \operatorname{sgn}(\tau)\operatorname{sgn}(\sigma)$;

(2) *if* σ *is a transposition, then* $\operatorname{sgn}(\sigma) = -1$; *and*

(3) *if* σ *is the identity, then* $\operatorname{sgn}(\sigma) = 1$.

Proof. First consider $\operatorname{sgn}(\tau\sigma)$. We have

$$\begin{aligned}
\operatorname{sgn}(\tau\sigma) &= \prod_{i<j} \frac{\tau\sigma(i) - \tau\sigma(j)}{i - j} \\
&= \prod_{i<j} \frac{\tau(\sigma(i)) - \tau(\sigma(j))}{\sigma(i) - \sigma(j)} \prod_{i<j} \frac{\sigma(i) - \sigma(j)}{i - j} \\
&= \prod_{r<s} \frac{\tau(r) - \tau(s)}{r - s} \prod_{i<j} \frac{\sigma(i) - \sigma(j)}{i - j} \\
&= \operatorname{sgn}(\tau)\operatorname{sgn}(\sigma).
\end{aligned}$$

Here, the third equality follows since σ is a permutation and

$$\frac{\tau(\sigma(i)) - \tau(\sigma(j))}{\sigma(i) - \sigma(j)} = \frac{\tau(\sigma(j)) - \tau(\sigma(i))}{\sigma(j) - \sigma(i)}.$$

Thus we get (1).

The proof of (2) uses the result of Example 8.1. Consider an arbitrary transposition σ_{ab}, where $1 \leq a < b \leq n$. I claim

$$\sigma_{ab} = \sigma_{1b}\sigma_{2a}\sigma_{12}\sigma_{2a}\sigma_{1b}. \tag{8.1}$$

We leave this as an exercise. By (1) and the fact that $\mathrm{sgn}(\sigma)^2 = 1$ for all σ,

$$\mathrm{sgn}(\sigma_{ab}) = \mathrm{sgn}(\sigma_{1b})\mathrm{sgn}(\sigma_{2a})\mathrm{sgn}(\sigma_{12})\mathrm{sgn}(\sigma_{2a})\mathrm{sgn}(\sigma_{1b}) = \mathrm{sgn}(\sigma_{12}).$$

But $\mathrm{sgn}(\sigma_{12}) = -1$, so we get also (2). The last statement is obvious, so the proof is finished. □

8.1.3 The Definition of the Determinant

We now give the general definition. Let $\mathbb{F}$ be any field, and let $A \in \mathbb{F}^{n\times n}$.

Definition 8.2. The determinant $\det(A)$ of A is defined to be

$$\det(A) := \sum_{\pi\in S(n)} \mathrm{sgn}(\pi)\ a_{\pi(1)1}a_{\pi(2)2}\cdots a_{\pi(n)n} \tag{8.2}$$

If A is 1×1, say $A = (a)$, then we already defined $\det(A) = a$. This agrees with the formula since $S(1)$ consists of the identity map, and the signature of the identity is 1. Suppose next that A is 2×2. There are only two elements $\sigma \in S(2)$, namely the identity id_2 and σ_{12}. Thus, by definition,

$$\det\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = +a_{11}a_{22} + \mathrm{sgn}(\sigma_{12})a_{21}a_{12} = a_{11}a_{22} - a_{21}a_{12}.$$

Thus the expression (8.2) agrees with the original definition.

Example 8.3 (3×3 determinants)**.** For the 3×3 case, we begin by listing the elements $\sigma \in S(3)$ and their signatures. We will denote σ by the triple $[\sigma(1), \sigma(2), \sigma(3)]$. Thus the signatures are given by the following table:

π	$[1,2,3]$	$[2,1,3]$	$[3,2,1]$	$[1,3,2]$	$[3,1,2]$	$[2,3,1]$
$\mathrm{sgn}(\pi)$	1	-1	-1	-1	$+1$	1

Hence,

$$\begin{aligned}\det(A) = {} & a_{11}a_{22}a_{33} - a_{21}a_{12}a_{33} - a_{31}a_{22}a_{13} \\ & -a_{11}a_{32}a_{23} + a_{31}a_{12}a_{23} + a_{21}a_{32}a_{13},\end{aligned}$$

which is a standard formula for a 3×3 determinant.

8.1.4 Permutations and Permutation Matrices

Permutations and permutation matrices are closely related. Since a permutation matrix P is a square matrix of zeros and ones so that each row and each column contains exactly one non-zero entry, P is uniquely determined by a permutation σ. Namely, if the ith column of P contains a 1 in the jth row, we put $\sigma(i) = j$. If P is $n \times n$, this defines a unique element $\sigma = \sigma(P) \in S(n)$. Conversely, given $\sigma \in S(n)$, define a permutation matrix P_σ by putting

$$P_\sigma = (\mathbf{e}_{\sigma(1)}\ \mathbf{e}_{\sigma(2)}\ \cdots\ \mathbf{e}_{\sigma(n)}),$$

where $\mathbf{e}_{\sigma(i)}$ is the standard basis vector whose $\sigma(i)$th-component is 1.

Proposition 8.5. *The mapping $\sigma \to P_\sigma$ defines a one to one corresponence between $S(n)$ and the set $P(n)$ of $n \times n$ permutation matrices.*

Proof. This is obvious consequence of the definitions. □

Example 8.4. As in the 3×3 case, we can represent any $\sigma \in S(n)$ by the symbol

$$[\sigma(1), \sigma(2), \ldots, \sigma(n)].$$

For example, the identity permutation of $S(3)$ is $[1, 2, 3]$. It corresponds to the permutation matrix I_3. The permutation $[2, 3, 1]$ corresponds to

$$P_{[2,3,1]} = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix},$$

and so forth. Note that the non-zero element of the ith column of P_σ is $p_{\sigma(i)i}$. Let's see what what happens when we form the product $P_{[2,3,1]}A$. We have

$$P_{[2,3,1]}A = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \mathbf{a}_3 \end{pmatrix} = \begin{pmatrix} \mathbf{a}_3 \\ \mathbf{a}_1 \\ \mathbf{a}_2 \end{pmatrix}.$$

More generally, let $P_{[i,j,k]}$ be the 3×3 permutation matrix in which the first row of I_3 is in the ith row, the second in the jth row and the third in the kth row. Then $P_{[i,j,k]}A$ is obtained from A by permutating the rows of A by the permutation $[i, j, k]$.

8.1.5 The Determinant of a Permutation Matrix

We now find the determinants of permutation matrices.

Proposition 8.6. *If $P \in P(n)$ has the form P_σ, then* $\det(P) = \operatorname{sgn}(\sigma)$.

Proof. We know that the only nonzero entries of P_σ are the entries of the form $p_{\sigma(i)i}$, all of which are 1 (see Example 8.4). Since

$$\det(P_\sigma) = \sum_{\pi \in S(n)} \operatorname{sgn}(\pi)\ p_{\pi(1)1} p_{\pi(2)2} \cdots p_{\pi(n)n},$$

the only non-zero term is

$$\operatorname{sgn}(\sigma) p_{\sigma(1)1} p_{\sigma(2)2} \cdots p_{\sigma(n)n} = \operatorname{sgn}(\sigma).$$

Therefore, $\det(P_\sigma) = \operatorname{sgn}(\sigma)$, as claimed. □

We now prove an important preliminary step in the our treatment of the determinant. We know that row swap matrices are permutation matrices. In fact, they are the permutation matrices coming from transpositions. We now want to show

Proposition 8.7. *If S is an $n \times n$ row swap matrix and P is an $n \times n$ permutation matrix, then* $\det(SP) = -\det(P)$. *In particular, if a permutation matrix P is a product of row swaps matrices, say $P = S_1 S_2 \cdots S_m$, then* $\det(P) = (-1)^m$.

Proof. Let $P = P_\tau$, where

$$P_\tau = (\mathbf{e}_{\tau(1)}\ \mathbf{e}_{\tau(2)}\ \cdots\ \mathbf{e}_{\tau(n)}),$$

and let $S = P_\sigma$, where σ is the transposition sending i to j and fixing all other $k \in X_n$. That is, $\sigma = \sigma_{ij}$. We need to calculate SP. In fact, I claim

$$SP = (\mathbf{e}_{\sigma\tau(1)}\ \mathbf{e}_{\sigma\tau(2)}\ \cdots\ \mathbf{e}_{\sigma\tau(n)}). \tag{8.3}$$

Indeed, suppose $\tau(k) = i$ and $\tau(\ell) = j$. Thus $\tau(k) = \sigma(j)$, and $\tau(\ell) = \sigma(i)$. This means that $\sigma^{-1}\tau(k) = j$ and $\sigma^{-1}\tau(\ell) = i$. But as σ is a transposition, $\sigma^{-1} = \sigma$, so $\sigma\tau(k) = j$ and $\sigma\tau(\ell) = i$. We now compute SP directly. For this, we may suppose without loss of generality that $i < j$. Then

$$\begin{aligned} SP &= S(\mathbf{e}_{\tau(1)}\ \cdots \mathbf{e}_{\tau(k)=i}\ \cdots\ \mathbf{e}_{\tau(\ell)=j}\ \cdots\ \mathbf{e}_{\tau(n)}) \\ &= (\mathbf{e}_{\tau(1)}\ \cdots \mathbf{e}_{\tau(\ell)=j}\ \cdots\ \mathbf{e}_{\tau(k)=i}\ \cdots\ \mathbf{e}_{\tau(n)}) \\ &= (\mathbf{e}_{\sigma\tau(1)=\tau(1)}\ \cdots \mathbf{e}_{\sigma\tau(k)=j}\ \cdots\ \mathbf{e}_{\sigma\tau(\ell)=i}\ \cdots\ \mathbf{e}_{\sigma\tau(n)=\tau(n)}) \end{aligned}$$

This gives us (8.3) as claimed. The upshot of this calculation is that

$$P_\sigma P_\tau = P_{\sigma\tau} \tag{8.4}$$

for all $\tau \in S(n)$ if $\sigma \in S(n)$ is any transposition. The Proposition now follows since

$$\det(SP) = \det(P_{\sigma\tau}) = \mathrm{sgn}(\sigma\tau) = \mathrm{sgn}(\sigma)\mathrm{sgn}(\tau) = -\mathrm{sgn}(\tau) = -\det(P),$$

since $\det(P) = \mathrm{sgn}(\tau)$ and $\mathrm{sgn}(\sigma) = -1$. □

Formula (8.4) in fact holds in general. That is, for all $\sigma, \tau \in S(n)$, we have $P_{\sigma\tau} = P_\sigma P_\tau$. We invite the reader to prove this in Exercise 8.9 below. What it says is in terms of abstract algebra is that, as groups, $P(n)$ and $S(n)$ are isomorphic.

In the next Section, we will apply these results on permutation matrices to obtain the rules for computing determinants via row operations.

Exercises

Exercise 8.1. Write down two 4×4 matrices A, B each with at most two zero entries such that $\det(A)\det(B) \neq 0$.

Exercise 8.2. Prove Proposition 8.2.

Exercise 8.3. If A is $n \times n$ and r is a scalar, find a formula for $\det(rA)$.

Exercise 8.4. Let $A = \begin{pmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \mathbf{a}_3 \end{pmatrix}$ and $A' = \begin{pmatrix} \mathbf{a}'_1 \\ \mathbf{a}_2 \\ \mathbf{a}_3 \end{pmatrix}$ be two elements of $\mathbb{F}^{3\times 3}$ written in row form. Find a formula for $\det(B)$ in terms of $\det(A)$ and $\det(A')$ if

$$B = \begin{pmatrix} \mathbf{a}_1 + r\mathbf{a}'_1 \\ \mathbf{a}_2 \\ \mathbf{a}_3 \end{pmatrix}$$

for some $r \in \mathbb{F}$.

Exercise 8.5. Find the determinant of a 2×2 reflection matrix.

Exercise 8.6. Find the determinant of a 2×2 rotation matrix.

Exercise 8.7. Classify the elements of $O(2, \mathbb{R})$ according to their determinants.

Exercise 8.8. Prove Equation (8.1).

Exercise 8.9. Show that every permutation $\tau \in S(n)$ can be written as a product of transpositions. (This isn't hard, but it's somewhat long to write out. For some hints, consult a book on elementary algebra.) Then use this to prove that $P_{\sigma\tau} = P_\sigma P_\tau$ for all $\sigma, \tau \in S(n)$.

8.2 Determinants and Row Operations

In the last section, we defined the determinant function and showed that the determinant of a permutation matrix is the signature of its permutation. This gives some insight as to how determinants are computed in general. But it's clear that trying to compute a determinant from the definition won't get us very far, so we need more tools. We will now show that elementary row operations once again turn out to give us what we need.

We begin by slightly reformulating the definition. This will simplify matters, especially for studying the Laplace expansion. For any $A \in \mathbb{F}^{n\times n}$, put

$$\delta(A) = \prod_{i=1}^{n} a_{ii}.$$

That is, $\delta(A)$ is the product of all the diagonal entries of A.

For example, it can be seen (by staring long enough at the definition of $\det(A)$) that if A is upper or lower triangular, then $\det(A) = \delta(A)$. This is also stated in Proposition 8.9 below. Our recasting of the definition is the formula in the next Proposition.

Proposition 8.8. *If $A \in \mathbb{F}^{n\times n}$, then*

$$\det(A) = \sum_{\sigma\in S(n)} \det(P_\sigma)\delta(P_\sigma A), \tag{8.5}$$

Before giving the proof, let's calculate a 3×3 example. Consider $\delta(PA)$, where P is the matrix $P_{[2,3,1]}$. Thus

$$P = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \quad \text{and} \quad A = \begin{pmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \mathbf{a}_3 \end{pmatrix}.$$

Recall from Example 8.4 that

$$P_{[2,3,1]}A = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \mathbf{a}_3 \end{pmatrix} = \begin{pmatrix} \mathbf{a}_3 \\ \mathbf{a}_1 \\ \mathbf{a}_2 \end{pmatrix}.$$

Hence

$$\delta(P_{[2,3,1]}A) = a_{31}a_{12}a_{23}.$$

Since $\sigma(1) = 2$, $\sigma(2) = 3$ and $\sigma(3) = 1$, we see that $\sigma^{-1}(1) = 3$, $\sigma^{-1}(2) = 1$ and $\sigma^{-1}(3) = 2$. Thus

$$\delta(P_\sigma A) = a_{\sigma^{-1}(1)1}a_{\sigma^{-1}(2)2}a_{\sigma^{-1}(3)3}.$$

Let's now give the proof of Proposition8.8.

Proof. From the above calculation,

$$\delta(P_\sigma A) = a_{\sigma^{-1}(1)1}a_{\sigma^{-1}(2)2}\cdots a_{\sigma^{-1}(n)n}.$$

As $\mathrm{sgn}(\sigma) = \mathrm{sgn}(\sigma^{-1})$ (why?), we have

$$\mathrm{sgn}(\sigma)\delta(P_\sigma A) = \mathrm{sgn}(\sigma^{-1})a_{\sigma^{-1}(1)1}a_{\sigma^{-1}(2)2}\cdots a_{\sigma^{-1}(n)n}.$$

Now, as σ varies over all of $S(n)$, so does σ^{-1}, hence we see that

$$\begin{aligned}
\det(A) &= \sum_{\sigma\in S(n)} \mathrm{sgn}(\sigma)\ a_{\sigma(1)1}a_{\sigma(2)2}\cdots a_{\sigma(n)n} \\
&= \sum_{\sigma\in S(n)} \mathrm{sgn}(\sigma^{-1})\ a_{\sigma^{-1}(1)1}a_{\sigma^{-1}(2)2}\cdots a_{\sigma^{-1}(n)n} \\
&= \sum_{\sigma\in S(n)} \mathrm{sgn}(\sigma)\delta(P_\sigma A) \\
&= \sum_{\sigma\in S(n)} \det(P_\sigma)\delta(P_\sigma A).
\end{aligned}$$

□

8.2.1 The Main Result

The strategy for computing determinants is explained by first considering the triangular case.

Proposition 8.9. *Suppose A is $n \times n$ and upper triangular, that is, every element in A below the diagonal is zero. Then*

$$\det(A) = \delta(A) = a_{11}a_{22}\cdots a_{nn}.$$

The same formula also holds for a lower triangular matrix.

Proof. The point is that the only nonzero term in $\det(A)$ is $a_{11}a_{22}\cdots a_{nn}$. For if P is a permutation matrix different from the identity, then there has to be an index i so that the ith row of PA is different from the ith row of A. But that means PA has to have a 0 on the diagonal, so $\delta(PA) = 0$. □

Hence the key to computing higher order determinants is to use row operations to bring A into triangular form. Thus we need to investigate how $\det(A)$ changes after a row operation is performed on A. Our main result is that the determinant function obeys the following rules with respect to row operations.

Theorem 8.10. *Let $\mathbb{F}$ be an arbitrary field, and suppose $n \geq 2$. Then the following four properties hold for all $A, B \in \mathbb{F}^{n\times n}$.*

(Det I) *If B is obtained from A by a row swap, then*

$$\det(B) = -\det(A). \tag{8.6}$$

(Det II) *If B is obtained from A by multiplying a row of A by a (possibly zero) scalar r, then*

$$\det(B) = r\det(A). \tag{8.7}$$

(Det III) *If B is obtained from A by replacing the ith row by itself plus a multiple of the jth row, $i \neq j$, then*

$$\det(B) = \det(A). \tag{8.8}$$

(Det IV) $\det(I_n) = 1$.

Proof. To prove **(Det I)**, suppose $B = SA$, where S swaps two rows, but leaves all the other rows alone. By the previous Proposition,

$$\det(B) = \det(SA) = \sum_{P\in P(n)} \det(P)\delta(P(SA)).$$

Since S and P are permutation matrices, so is $Q = PS$. But the map $P \to PS$ is an injection of $S(n)$ to itself (why?). Thus the Pigeon Hole Principle (Proposition 4.8) tells us that $P \to PS$ is a bijection of $S(n)$. Hence,

$$\begin{aligned}
\det(SA) &= \sum_{P\in P(n)} \det(PS)\delta((PS)SA) \\
&= \sum_{P\in P(n)} -\det(P)\delta(PS^2A) \\
&= -\sum_{P\in P(n)} \det(P)\delta(PA) \\
&= -\det(A),
\end{aligned}$$

since $S^2 = I_n$, and $\det(PS) = -\det(P)$ (Proposition8.8). Putting $B = SA$ gives us **(Det I)**.

To prove **(Det II)**, suppose E multiplies the i-th row of A by the scalar r. Then for every $P \in P(n)$, $\delta(P(EA)) = r\delta(PA)$. Thus $\det(EA) = r\det(A)$.

(Det III) follows from two facts. First of all, suppose $A, A' \in \mathbb{F}^{n\times n}$ coincide in all but one row, say the kth row. That is, if $A = (a_{ij})$ and $A' = (a'_{ij})$, then $a_{ij} = a'_{ij}$ as long as $i \neq k$. Now define a matrix $B = (b_{ij})$ by $b_{ij} = a_{ij} = a'_{ij}$ if $i \neq k$, and $b_{kj} = a_{kj} + a'_{kj}$. We claim

$$\det(B) = \det(A) + \det(A'). \tag{8.9}$$

To prove (8.9), fix $\pi \in S(n)$ and consider

$$\operatorname{sgn}(\pi)\ b_{\pi(1)1} \cdots b_{\pi(n)n}.$$

There exists exactly one index j such that $\pi(j) = k$. Then, by the definition of B, $b_{\pi(j)j} = a_{\pi(j)j} + a'_{\pi(j)j}$. Hence,

$$\operatorname{sgn}(\pi)\ b_{\pi(1)1} \cdots b_{\pi(n)n} = \operatorname{sgn}(\pi)\ a_{\pi(1)1} \cdots a_{\pi(n)n} + \operatorname{sgn}(\pi)\ a'_{\pi(1)1} \cdots a'_{\pi(n)n},$$

hence we get (8.9).

Now suppose E is the elementary matrix of type III is obtained from I_n by replacing the ith row of I_n by itself plus a times the jth row, where $i \neq j$. Thus $B = EA$. Let C be the matrix which is the same as A except that the ith row of C is a times the jth row of A. Then from (8.9), we know that $\det(B) = \det(A) + \det(C)$. Furthermore, by **(Det II)**, $\det(C) = a\det(C')$, where C' is the result of factoring a from the ith row of C. Thus

$$\det(B) = \det(A) + a\det(C'),$$

where C' is a matrix whose ith and jth rows coincide. Thus **(Det III)** will be proved once we establish

Lemma 8.11. *If $C \in \mathbb{F}^{n\times n}$ has two equal rows, then* $\det(C) = 0$.

There is an easy way to see this if the characteristic of $\mathbb{F}$ is different from 2. For, if the rth and sth rows of C coincide, then by **(Det I)**, $\det(C) = -\det(C)$. This implies $2\det(C) = 0$, so $\det(C) = 0$. We will give a proof of this which works for any characteristic, but we first need to introduce the Laplace expansion. This will be done in Section 8.3.1.

(Det IV) follows immediately from Proposition 8.9, so the Theorem is proved, modulo the above Lemma. □

In particular, since $\det(I_n) = 1$, **Det I-Det III** imply that if E is an elementary matrix, then

$$\det(E) = \begin{cases} -1 & \text{if } E \text{ is of type I,} \\ r & \text{if } E \text{ is of type II,} \\ 1 & \text{if } E \text{ is of type III} \end{cases}$$

Therefore we can summarize **Det I-Det III** in the following way.

Corollary 8.12. *If $A \in \mathbb{F}^{n\times n}$ and $E \in \mathbb{F}^{n\times n}$ is elementary, then*

$$\det(EA) = \det(E)\det(A). \tag{8.10}$$

In particular, if $A = E_1 \cdots E_k$, where $E_1, \ldots, E_k$ are elementary, then

$$\det(A) = \det(E_1)\cdots\det(E_k). \tag{8.11}$$

Proof. If E is a row swap , then $\det(EA) = -\det(A)$ by **Det I**. If E is of type II, say $\det(E) = r$, then by **Det II**, $\det(EA) = r\det(A) = \det(E)\det(A)$. If E is of type III, then $\det(E) = 1$ while $\det(EA) = \det(A)$ by **Det III**. Hence we get (8.10). Identity (8.11) follows by iterating the result of (8.10). □

8.2.2 Consequences

Corollary 8.12 suggests a definite method for evaluating $\det(A)$. First find elementary matrices $E_1, \ \ldots \ , E_k$ such that $U = E_k \cdots E_1 A$ is upper triangular. By Proposition 8.9 and Corollary 8.12,

$$\det(U) = \det(E_1)\cdots\det(E_k)\det(A) = u_{11}u_{22}\cdots u_{nn},$$

where the u_{ii} are the diagonal entries of U. Since no $\det(E_i) = 0$,

$$\det(A) = \frac{u_{11}u_{22}\cdots u_{nn}}{\det(E_1)\cdots\det(E_k)}. \tag{8.12}$$

Example 8.5. 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 $\det(A) = -1$.

Example 8.6. Let us next compute $\det(A)$, where A is the matrix of the previous example, this time taking the field of coefficients to be $\mathbb{Z}_2$. 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 $\mathbb{Z}_2$, 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

$$\det(A) = \det \begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} = 1.$$

Note that switching rows doesn't change the determinant since $-1 = 1$ in $\mathbb{F}_2$. In fact, we could also deduce that $\det(A) = 1$ using the steps in the previous example.

One can simplify evaluating $\det(A)$ even more in some special cases. For example, if A has the form

$$\begin{pmatrix} B & C \\ O & D \end{pmatrix}, \tag{8.13}$$

where the submatrices B and C are square, then $\det(A) = \det(B)\det(D)$. The proof is similar to the proof of (8.12).

To see what the determinant sees, notice that (8.12) implies $\det(A) \neq 0$ if and only if each $u_{ii} \neq 0$. Since this is the case precisely when A has maximal rank, we get

Proposition 8.13. *If A is $n \times n$, then $\det(A) \neq 0$ if and only if A has rank n.*

We can now prove the product formula, which was proved in Proposition 8.1 in the 2×2 case.

Theorem 8.14. *If A and B are any $n \times n$ matrices over $\mathbb{F}$, then*

$$\det(AB) = \det(A)\det(B). \tag{8.14}$$

Proof. If A and B both have rank n, each one of them can be expressed as a product of elementary matrices, say $A = E_1 \cdots E_k$ and $B = E_{k+1} \cdots E_m$. Then $AB = (E_1 \cdots E_k)(E_{k+1} \cdots E_m)$, so by Corollary 8.12,

$$\det(AB) = \det(E_1) \cdots \det(E_k)\det(E_{k+1}) \cdots \det(E_m) = \det(A)\det(B).$$

If either A or B has rank less than n, then $\det(A)\det(B) = 0$. But we already know that, in this case, the rank of AB is less that n too, so $\det(AB) = 0$. Thus the proof is done. □

Corollary 8.15. *If A is invertible, then $\det(A^{-1}) = \det(A)^{-1}$.*

Proof. This follows from the product formula since $AA^{-1} = I_n$ implies

$$1 = \det(I_n) = \det(AA^{-1}) = \det(A)\det(A^{-1}).$$

□

The following Proposition gives another remarkable about the determinant.

Proposition 8.16. *If A is any square matrix, then $\det(A) = \det(A^T)$.*

Proof. We know that A and A^T have the same rank, so the result is true if $\det(A) = 0$. Hence we can suppose A has maximal rank. Express A as a product of elementary matrices, say $A = E_1 E_2 \cdots E_k$. Then

$$A^T = (E_1 E_2 \cdots E_k)^T = E_k^T E_{k-1}^T \cdots E_1^T.$$

Thus it is sufficient to show $\det(E^T) = \det(E)$ for any elementary matrix E. This is clear if E is of type II, since elementary matrices of type II are symmetric. If E is of type III, then so is E^T, so $\det(E) = \det(E^T) = 1$. Finally, if E is of type I, that is E is a row swap, $E^T = E^{-1}$. In this case, $\det(E) = \det(E^T) = -1$. □

Exercises

Exercise 8.10. Two square matrices A and B are said to be similar if there exists a matrix M so that $B = MAM^{-1}$. Show that similar matrices have the same determinants.

Exercise 8.11. Suppose P is an $n \times n$ matrix so that $PP = P$. What is $\det(P)$? What if $P^4 = P^{-1}$?

Exercise 8.12. Suppose that $Q \in \mathbb{R}^{n \times n}$ is orthogonal. Find the possible values of $\det(Q)$.

Exercise 8.13. Which of the following statements are true and which are false? Give your reasoning.

(a) The determinant of a real symmetric matrix is always non negative.

(b) If A is any 2×3 real matrix, then $\det(AA^T) \geq 0$.

(c) If A is a square real matrix, then $\det(AA^T) \geq 0$.

Exercise 8.14. An $n \times n$ matrix A 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 8.15. A complex $n \times n$ matrix U is called *unitary* if $U^{-1} = \overline{U}^T$, where $\overline{U}$ is the matrix obtained by conjugating each entry of U. What are the possible values of the determinant of $\det(U)$ of a unitary matrix U.

Exercise 8.16. Compute

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

in two cases: first when the field is $\mathbb{Q}$ and secondly when the field is $\mathbb{Z}_5$.

8.3 Some Further Properties of the Determinant

In this section, we will obtain some further properties of the determinant. We will begin with the Laplace expansion, which is the classical way of expressing an $n \times n$ determinant as a sum of $(n-1) \times (n-1)$ determinants. Laplace expansion is an important theoretical tool, though not a very helpful computational technique. We will also give a characterization of the determinant as a function on $\mathbb{F}^{n\times n}$ and define the determinant of a linear transformation $T : V \to V$.

8.3.1 The Laplace Expansion

Suppose A is $n \times n$, and let A_{ij} denote the $(n-1) \times (n-1)$ submatrix obtained from A by deleting its ith row and jth column.

Theorem 8.17. *For any $A \in \mathbb{F}^{n\times n}$, we have*

$$\det(A) = \sum_{i=1}^{n} (-1)^{i+j} a_{ij} \det(A_{ij}). \tag{8.15}$$

This is the Laplace expansion along the jth column. The corresponding Laplace expansion of along the ith row is

$$\det(A) = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} \det(A_{ij}). \tag{8.16}$$

Proof. Since $\det(A) = \det(A^T)$, it suffices to prove (8.15) . For simplicity, we will assume $j = 1$, the other cases being similar. Now,

$$\begin{aligned}
\det(A) &= \sum_{\sigma\in S(n)} \operatorname{sgn}(\sigma)\ a_{\sigma(1)1} a_{\sigma(2)2} \cdots a_{\sigma(n)n} \\
&= a_{11} \sum_{\substack{\sigma\in S(n)\\ \sigma(1)=1}} \operatorname{sgn}(\sigma) a_{\sigma(2)2} \cdots a_{\sigma(n)n} \ + \\
&\quad + a_{21} \sum_{\substack{\sigma\in S(n)\\ \sigma(1)=2}} \operatorname{sgn}(\sigma) a_{\sigma(2)2} \cdots a_{\sigma(n)n} + \\
&\quad + \cdots + a_{n1} \sum_{\substack{\sigma\in S(n)\\ \sigma(1)=n}} \operatorname{sgn}(\sigma) a_{\sigma(2)2} \cdots a_{\sigma(n)n}
\end{aligned}$$

If $\sigma \in S(n)$, 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 $\sigma(1)$st row. Since $p_{\sigma(i)i} = 1$, it follows

that $P'_\sigma \in P(n-1)$ (why?). Note that $\det(P_\sigma) = (-1)^{(\sigma(1)-1)} \det(P'_\sigma)$, since if bringing P'_σ to I_{n-1} by row swaps uses t steps, an additional $\sigma(1) - 1$ adjacent row swaps bring P_σ to the identity. Next, recall that

$$\det(A) = \sum_{\sigma \in S(n)} \det(P_\sigma)\delta(P_\sigma A).$$

Since $\det(P_\sigma) = (-1)^{(\sigma(1)-1)} \det(P'_\sigma)$, we see that for each r with $1 \leq r \leq n$,

$$\sum_{\substack{\sigma \in S(n) \\ \sigma(1)=r}} \det(P_\sigma)\delta(P_\sigma A) = (-1)^{(r-1)} a_{r1} \sum_{\substack{\sigma \in S(n) \\ \sigma(1)=r}} \det(P'_\sigma)\delta(P'_\sigma A_{r1}).$$

But the right hand side is certainly $(-1)^{(r-1)} a_{r1} \det(A_{r1})$, since every element of $P(n-1)$ is P'_σ for exactly one $\sigma \in S(n)$ with $\sigma(1) = r$. Therefore,

$$\det(A) = \sum_{i=1}^{n} (-1)^{i-1} a_{i1} \det(A_{i1}),$$

which is the desired formula. □

Example 8.7. If A is 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 8.8. The Laplace expansion is useful for evaluating $\det(A)$ when A has entries which 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. 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 $x^3 - 4x + 7 = 0$.

REMARK: In many algebra texts, the determinant is actually defined inductively via the Laplace expansion. The problem with this approach is that in order to do this rigorously, one has to show all possible Laplace expansions have the same value. There is no simple way to do this, so it is usually taken for granted. All that the Laplace expansion does is systematically organize the terms of the determinant. In general, the Laplace expansion isn't even in the same ballpark as row operations as an efficient way of computing $\det(A)$. Using Laplace to evaluate even a 20×20 determinant is impractical except possibly for a super computer (note $20! = 2432902008176640000$). Yet in fairly mundane applications of linear algebra to biotechnology, one might need to evaluate a 2000×2000 determinant. Calculations involving genomes routinely require evaluating much larger determinants. On the other hand, the Laplace expansion is convenient for matrices which have very few nonzero entries.

8.3.2 The Case of Equal Rows

We still need to finish the proof of assertion **Det III** Theorem 8.10. Recall that this requires we establish

Proposition 8.18. *Suppose $\mathbb{F}$ is a field of arbitrary characteristic and $n > 1$. Then if $C \in \mathbb{F}^{n\times n}$ has two equal rows or two equal columns, then* $\det(C) = 0$.

Proof. Since $\det(A) = \det(A^T)$ (Proposition 8.16), it suffices to stick to the case of equal rows. Recall that we used a row switching argument in the proof of Theorem 8.10 to prove the special case where the characteristic of $\mathbb{F}$ is different from two. Suppose in general that two rows of A are equal, where $A \in \mathbb{F}^{n\times n}$ and $n > 1$. If $n = 2$, the result is clear by the explicit formula for $\det(A)$. The strategy will be to use induction and Laplace together. So assume that the result is true for all $A \in \mathbb{F}^{m\times m}$, where $2 \leq m \leq n-1$, and let $A \in \mathbb{F}^{n\times n}$ have the same the ith and jth rows. We may in fact suppose, without any loss of generality, that $i, j \neq 1$ (why?). That being the case, apply the Laplace expansion along the first row. We then see that $\det(A)$ is a sum of $(n-1) \times (n-1)$ determinants, each of which has two equal rows. By the inductive hypothesis, each of these $(n-1)\times(n-1)$ determinants is 0. Therefore $\det(A) = 0$ too. This completes the induction, so the Proposition is established. □

8.3.3 Cramer's Rule

Another reason the Laplace expansion is important, at least from a theoretical point of view, is that it gives a closed formula known as Cramer's Rule for the inverse of a matrix. Recall that if A is 2×2,

$$A^{-1} = \frac{1}{\det(A)} \begin{pmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{pmatrix}.$$

Inspecting this formula may suggest the correct formula for A^{-1} in the general case.

Definition 8.3. 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

$$Cof(A) = \big((-1)^{i+j} \det(A_{ji})\big) \tag{8.17}$$

is called the *cofactor* of A.

Proposition 8.19. *Suppose $A \in \mathbb{F}^{n\times n}$. Then $Cof(A)A = \det(A)I_n$. Thus, if $\det(A) \neq 0$, then*

$$A^{-1} = \frac{1}{\det(A)} Cof(A).$$

Proof. The essential ideas are all contained in the 3×3 case, so, for simplicity, let $n = 3$. By definition,

$$Cof(A) = \begin{pmatrix} \det(A_{11}) & -\det(A_{21}) & \det(A_{31}) \\ -\det(A_{12}) & \det(A_{22}) & -\det(A_{23}) \\ \det(A_{13}) & -\det(A_{23}) & \det(A_{33}) \end{pmatrix}.$$

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 8.17 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 two equal columns. Thus $c_{21} = 0$ by Proposition 8.18. Using similar reasoning on the other c_{ij} with $i \neq j$, we get $Cof(A)A = \det(A)I_3$, which is the first claim. If $\det(A) \neq 0$, dividing by $Cof(A)$ by $\det(A)$ gives a left inverse of A, hence A^{-1} (by Section 3.3), so the proof is complete. □

8.3.4 The Inverse of a Matrix Over $\mathbb{Z}$

Most of the matrices we've inverted have integral entries. By Proposition 8.19, these inverses have rational entries. But we also know that some of these inverses have integral entries, namely those where $\det(A) = \pm 1$. The question is whether these are all. The answer is given by

Proposition 8.20. *Suppose A is an invertible matrix with integral entries. Then A^{-1} also has integral entries if and only if* $\det(A) = \pm 1$.

Proof. We just proved the if statement. 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 8.21. *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 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 only multiply a row by ± 1 and replace it by itelf plus an integral multiple of another row.

8.3.5 A Characterization of the Determinant

Although it's a little tedious to write out all the details, one can infer directly from the definition that $\det(A)$ is an $\mathbb{F}$-linear function of the rows of A. That is, if all but one of the rows of A fixed, then the determinant is a linear

function of the remaining row. Now we can ask: is the determinant the only such function? This is answered by

Theorem 8.22. *The only function $F : \mathbb{F}^{n\times n} \to \mathbb{F}$ satisfying:*

(1) *F is $\mathbb{F}$-linear in each row,*

(2) *$F(A) = 0$ if two rows of A are equal, and*

(3) *$F(I_n) = 1$*

is the determinant function $F(A) = \det(A)$.

Proof. In fact, these conditions tell us that for any elementary matrix E, $F(EA)$ is computed from $F(A)$ in exactly the same way $\det(EA)$ is computed from $\det(A)$. We will omit the details. □

8.3.6 The determinant of a linear transformation

The product theorem enables us to define the determinant of a linear transformation $T : V \to V$, provided V is finite dimensional. The definition goes as follows.

Definition 8.4. The determinant $\det(T)$ of T is defined to be $\det(A)$, where $A \in \mathbb{F}^{n\times n}$ is the matrix representation of T with respect to some basis of V.

We need to check that $\det(T)$ is well defined. That is, we have to check the if B is another matrix representation of T, then $\det(A) = \det(B)$. But Proposition 7.13 says that if A and B are matrices of T with respect to different bases, then A and B are similar, i.e. there exists an invertible $P \in \mathbb{F}^{n\times n}$ such that $B = PAP^{-1}$. Thus

$$\det(B) = \det(PAP^{-1}) = \det(P)\det(A)\det(P^{-1}) = \det(A),$$

so $\det(T)$ is indeed well defined.

Exercises

Exercise 8.17. Find all values $x \in \mathbb{R}$ for which

$$A(x) = \begin{pmatrix} 1 & x & 2 \\ x & 1 & x \\ 2 & 3 & 1 \end{pmatrix}$$

is singular, that is, not invertible.

Exercise 8.18. Repeat the previous exercise for the matrix

$$B(x) = \begin{pmatrix} 1 & x & 1 & x \\ 1 & 0 & x & 1 \\ 0 & x & 1 & 1 \\ 1 & 0 & 1 & 0 \end{pmatrix}.$$

(Suggestion: use the Laplace expansion to evaluate.)

Exercise 8.19. Recall the cross product mapping $C_{\mathbf{a}} : \mathbb{R}^3 \to \mathbb{R}^3$ given by $C_{\mathbf{a}}(\mathbf{x}) = \mathbf{a} \times \mathbf{x}$. Find $\det(C_{\mathbf{a}})$.

Exercise 8.20. Why does condition (2) in Theorem 8.22 imply that the determinant changes sign under a row swap?

8.4 Geometric Applications of the Determinant

Now let us mention some of the geometric applications of determinants.

8.4.1 Cross and Vector Products

The *cross product* $\mathbf{x} \times \mathbf{y}$ of two vectors $\mathbf{x},\ \mathbf{y} \in \mathbb{R}^3$ is defined by the following rule:

$$\mathbf{x} \times \mathbf{y} = \det \begin{pmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_3 \\ x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \end{pmatrix}.$$

It is linear in each variable and has the nice property that $\mathbf{x} \times \mathbf{y}$ is orthogonal to both $\mathbf{x}$ and $\mathbf{y}$. In fact,

$$\mathbf{x} \times \mathbf{y} = |\mathbf{x}||\mathbf{y}| \sin\theta \mathbf{z},$$

where $\mathbf{z}$ is the unit vector such that $\det \begin{pmatrix} \mathbf{x} & \mathbf{y} & \mathbf{z} \end{pmatrix} > 0$. When this determinant is positive, we call $(\mathbf{x}, \mathbf{y}, \mathbf{z})$ a right handed triple.

There is an n dimensional generalization of the cross product called the *vector product*, which assigns to any $(n-1)$ vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{n-1} \in \mathbb{R}^n$ a vector

$$[\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{n-1}] \in \mathbb{R}^n$$

orthogonal to each of the $\mathbf{x}_i$. The vector product is defined by the following determinantal expression:

$$[\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{n-1}] = \det \begin{pmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \ldots & \mathbf{e}_n \\ x_{11} & x_{12} & \ldots & x_{1n} \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ x_{n-1,1} & x_{n-1,2} & \ldots & x_{n-1,n} \end{pmatrix}.$$

The fact that

$$\mathbf{x}_i \cdot [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{n-1}] = 0$$

for each i, $1 \leq i \leq n-1$, is due to the fact that a determinant with two equal rows is 0.

8.4.2 Determinants and Volumes

The absolute value of a real determinant has an interesting geometric interpretation. Note that a basis $\mathbf{w}_1, \ \ldots \ , \ \mathbf{w}_n$ of $\mathbb{R}^n$ spans an n-dimensional solid parallelogram $< \mathbf{w}_1, \ \ldots \ , \ \mathbf{w}_n >$. By definition

$$< \mathbf{w}_1, \ \ldots \ , \ \mathbf{w}_n >= \{\sum_{i=1}^{n} t_i \mathbf{w}_i \mid 0 \leq t_i \leq 1\}.$$

It can be shown that the volume of $< \mathbf{w}_1, \ \ldots \ , \ \mathbf{w}_n >$ is given by the formula

$$\mathbf{Vol}(< \mathbf{w}_1, \ \ldots \ , \ \mathbf{w}_n >) = |\det \begin{pmatrix} \mathbf{w}_1 & \mathbf{w}_2 & \cdots & \mathbf{w}_n \end{pmatrix}|.$$

To connect this with matrices, consider the linear transformation of $T : \mathbb{R}^n \to \mathbb{R}^n$ such that $T(\mathbf{e}_i) = \mathbf{w}_i$. In other words, the ith column of the matrix of T is $\mathbf{w}_i$. Thus $|\det(T)|$ is the volume of the image under T of the unit cube spanned by the standard basis; i.e.

$$|\det(T)| = \mathbf{Vol}(< \mathbf{w}_1, \ \ldots \ , \ \mathbf{w}_n >).$$

Hence the linear transformation associated to a real matrix having determinant of absolute value 1 preserves the volume of a cube, although the image is certainly not necessarily a cube. A linear transformation $T : \mathbb{R}^n \to \mathbb{R}^n$ of determinant 1 is said to be *unimodular*. We say that the matrix of T is unimodular. The set of all unimodular real $n \times n$ matrices is denoted by $SL(n, \mathbb{R})$ and called the *special linear group*.

Proposition 8.23. *Products and inverses of unimodular real matrices are also unimodular, and I_n is unimodular. Hence, $SL(n, \mathbb{R})$ is a matrix group.*

Unimodular matrices preserve volumes. Indeed, if $\mathbf{w}_1, \ \ldots \ , \ \mathbf{w}_n$ is a basis of $\mathbb{R}^n$, S is unimodular and T is the linear transformation defined in the previous paragraph, we have

$$\begin{aligned} \mathbf{Vol}(< S(\mathbf{w}_1), \ \ldots \ , \ S(\mathbf{w}_n) >) &= \mathbf{Vol}(< ST(\mathbf{e}_1), \ \ldots, \ ST(\mathbf{e}_n) >) \\ &= |\det(ST)| \\ &= |\det(S)\det(T)| \\ &= |\det(T)| \\ &= \mathbf{Vol}(< \mathbf{w}_1, \ \ldots \ , \ \mathbf{w}_n >). \end{aligned}$$

Using the above argument, one gets a geometric proof of the abolute product formula: for any $A, B \in \mathbb{R}^{n \times n}$,

$$|\det(AB)| = |\det(A)||\det(B)|.$$

This is of course weaker than the product formula itself.

8.4.3 Lewis Carroll's identity

Here's an obvious question. Why doesn't the formula for a 2×2 determinant work in some form in the $n \times n$ case? This question was answered by a professor of mathematics at Oxford University named Charles Dodgson, better known as Lewis Carroll, the author of Alice in Wonderland and Through the Looking Glass. Dodgson discoverered and published a proof of the following amusing identity. Given an $n \times n$ matrix A, where $n > 2$, let A_C be the $(n-2) \times (n-2)$ submatrix in the middle of A obtained by deleting the first and last rows and the first and last columns. If $n = 2$, we will define $\det(A_C) = 1$. Also, let A_{NW}, A_{NE}, A_{SW} and A_{SE} be the $(n-1) \times (n-1)$ submatrices of A in the upper left corner, upper right corner, lower left corner and lower right corner respectively.

Lewis Carroll's Identity says:

$$\det(A)\det(A_C) = \det(A_{NW})\det(A_{SE}) - \det(A_{NE})\det(A_{SW}) \qquad (8.18)$$

(see C.L. Dodgson, Proc. Royal Soc. London **17**, 555-560 (1860)). Interestingly, Lewis Carroll's Identity has recently reappeared in the modern setting of semi-simple Lie algebras. The reader is encouraged to try some examples, and to give a proof in the 3×3 case.

Exercises

Exercise 8.21. Verify Lewis Carroll's Identity for the matrix

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

Exercise 8.22. Under what condition does Lewis Carroll's Identity make it possible to evaluate $\det(A)$?

Exercise 8.23. Prove Lewis Carroll's Identity in the 3×3 case.

Exercise 8.24. Suppose A and B are $n \times n$ matrices over $\mathbb{R}$. Verify the absolute product formula using the discussion of volumes. That is, show

$$|\det(AB)| = |\det(A)||\det(B)|.$$

Exercise 8.25. Determine which the 3×3 permutation matrices that lie in $SL(3, \mathbb{R})$.

Exercise 8.26. Prove Proposition 8.23.

Exercise 8.27. Prove that $SL(3, \mathbb{R})$ is a subgroup of $GL(3, \mathbb{R})$. (Recall $GL(n, \mathbb{R}) = \{A \in \mathbb{R}^{n \times n} \mid \det(A) \neq 0\}$.)

Exercise 8.28. * If G and H are groups, then a mapping $\varphi : G \to H$ is called a *group homomorphism* if for any $a, b \in G$, $\varphi(ab) = \varphi(a)\varphi(b)$. Explain how the determinant can be viewed as a group homomorphism if we choose the group G to be $GL(n, \mathbb{F})$, where $\mathbb{F}$ is any field.

Exercise 8.29. Show that the mapping sending $\sigma \in S(n)$ to $P_\sigma \in P(n)$ is a bijective group homomorphism.

8.5 A Concise Summary of Determinants

Let $\mathbb{F}$ be a field. The determinant is a function $\det : \mathbb{F}^{n\times n} \to \mathbb{F}$ with the following properties:

(1) $\det(AB) = \det(A)\det(B)$, and

(2) $\det(I_n) = 1$.

Furthermore, if E is an elementary matrix, then:

(3) $\det(E) = -1$ if E is obtained by swapping two rows of I_n;

(4) $\det(E) = r$ if E is obtained by multiplying some row of I_n by r; and

(5) $\det(E) = 1$ if E is obtained by adding a multiple of some row of I_n to another row.

The following properties of $\det(A)$ are consequences of (1) through (5).

(6) $\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc$.

(7) $\det(A) \neq 0$ if and only if A is invertible.

(8) $\det(A^T) = \det(A)$.

(9) If A is upper triangular, then $\det(A) = a_{11} \cdots a_{nn}$. That is, $\det(A)$ is the product of the diagonal entries of A.

8.6 Summary

Determinants have a long history in mathematics because they give an explicit expression for the solution of a nonsingular system of n equations in n variables. (This is known as Cramer's Rule.) They seem to have first been defined by Liebniz in the 2×2 case. Matrices themselves didn't explicitly appear until the 19-th century. The definition of the determinant considered here is a certain sum of terms, each one associated to an element of the symmetric group. Hence the definition of the determinant requires some preliminary facts about the symetric group: namely, the definition of the signature of a permutation.

If $A \in \mathbb{F}^{n\times n}$, then $\det(A)$ is an element of $\mathbb{F}$ such that $\det(AB) = \det(A)\det(B)$ for all $B \in \mathbb{F}^{n\times n}$, $\det(I_n) = 1$ and $\det(A) \neq 0$ if and only if A is invertible. All the properties of determinants are rigorously proved without too much effort. The main problem is to understand how the determinant changes with the application of a row operation. In fact, modern computers find determinants via row operations, never by applying the definition. As an application, we derive the various Laplace expansions of $\det(A)$.

The determinant of a linear transformation $T : V \to V$ is also defined, as long as V is finite dimensional: $\det(T)$ is just the determinant any matrix representing T. In addition to its importance in algebra, which is amply demonstrated in eigentheory, the determinant also has many geometric applications. These stem from the fact that $|\det(A)|$ is the n-dimensional volume of the solid spanned by $A\mathbf{e}_1, A\mathbf{e}_2, \ldots, A\mathbf{e}_n$. This is the reason the determinant appears in the change of variables theorem for multiple integrals, which one studies in vector analysis.

Chapter 9

Eigentheory

Consider a linear transformation $T : V \to V$, where V is a finite dimensional vector space over a field $\mathbb{F}$. We know from Chapter 7 that one can study T from the standpoint of matrix theory by choosing a basis $\mathcal{B}$ of V and replacing T by the $n \times n$ matrix $\mathcal{M}_{\mathcal{B}}^{\mathcal{B}}(T)$ over $\mathbb{F}$, where $n = \dim V$. The nicest case is when T is semi-simple. That is, that there exists an eigenbasis $\mathcal{B}$ of V. This is due to the fact that $\mathcal{B}$ is an eigenbasis if and only if $\mathcal{M}_{\mathcal{B}}^{\mathcal{B}}(T)$ is a diagonal matrix. Thus, T is semi-simple if and only if $\mathcal{M}_{\mathcal{B}}^{\mathcal{B}}(T)$ is similar to a diagonal matrix for any basis $\mathcal{B}$ of V.

Of course, being able to view T as a matrix allows us to switch from a geometric concept (the transformation) to an algebraic object (its matrix), which may be easier to work with. For example, diagonal matrices are very easy to manipulate, since they commute with each other and their powers are easy to calculate. Hence, semi-simple operators are by far the easiest to work with. Thus, one of our fundamental goals in this Chapter is to determine when a matrix is diagonalizable, that is, similar to a diagonal matrix. The first step is to develop the tools for finding the eigenpairs for T. Along the way, we will give some examples of how eigentheory is used. The culmination of all this will be the Principal Axis Theorem(Chapter 11) and the Jordan Decomposition Theorem (Chapter 13).

9.1 Dynamical Systems

The purpose of this section is to give some motivation and an overview of the eigenvalue problem for a linear transformation. We will consider the dynamical system associated to a matrix A, and see how to to describe such a system by diagonalizing A. Eigentheory is the tool which enables one to

do this. Along the way, we will introduce several ideas and terms to be formally defined in the next section.

9.1.1 The Fibonacci Sequence

To illustrate, we consider the Fibonacci sequence. Let (a_k) denote the sequence defined by the putting $a_k = a_{k-1} + a_{k-2}$ if $k \geq 2$, starting with two arbitrary integers a_0 and a_1. The Fibonacci sequence can be expressed recursively as a matrix identity

$$\begin{pmatrix} a_{k+1} \\ a_k \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} a_k \\ a_{k-1} \end{pmatrix}$$

where $k \geq 1$. Hence putting

$$F = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix},$$

we see that

$$\begin{pmatrix} a_{k+1} \\ a_k \end{pmatrix} = F \begin{pmatrix} a_k \\ a_{k-1} \end{pmatrix} = F^2 \begin{pmatrix} a_{k-1} \\ a_{k-2} \end{pmatrix} = \cdots = F^k \begin{pmatrix} a_1 \\ a_0 \end{pmatrix}.$$

Letting

$$\mathbf{v}_0 = \begin{pmatrix} a_1 \\ a_0 \end{pmatrix} \quad \text{and} \quad \mathbf{v}_k = \begin{pmatrix} a_{k+1} \\ a_k \end{pmatrix},$$

we can therefore express the Fibonacci sequence in the form $\mathbf{v}_k = F^k \mathbf{v}_0$.

This is an example of a *dynamical system*. Suppose in general that $A \in \mathbb{R}^{n \times n}$, and fix an arbitrary vector $\mathbf{v}_0 \in \mathbb{R}^n$. Then the dynamical system associated to A having initial value $\mathbf{v}_0$ is the sequence $(\mathbf{v}_k)$ with

$$\mathbf{v}_k = A\mathbf{v}_{k-1}, \quad k = 1, 2, \ldots$$

This sequence is easy to analyze if A is diagonal, say $A = \operatorname{diag}(d_1, d_2, \ldots, d_n)$. Indeed, A's Nth power A^N is just the diagonal matrix

$$A^N = \operatorname{diag}(d_1^N, d_2^N, \ldots, d_n^N).$$

Thus, if $\mathbf{v}_0 = (v_1, v_2, \ldots, v_n)^T$, then

$$\mathbf{v}_N = ((d_1)^N v_1, (d_2)^N v_2, \ldots, (d_n)^N v_n)^T.$$

Fortunately, this isn't the only situation where we can compute A^N. Suppose $\mathbf{v} \in \mathbb{R}^n$ is nonzero and satisfies the condition $A\mathbf{v} = \lambda \mathbf{v}$ for some

scalar $\lambda \in \mathbb{R}$. Thus $(\lambda, \mathbf{v})$ is an eigenpair for A (see Example 7.3). Using this eigenpair, we get that

$$A^N\mathbf{v} = A^{N-1}(A\mathbf{v}) = A^{N-1}(\lambda\mathbf{v}) = \lambda A^{N-1}\mathbf{v}.$$

By iterating, we obtain

$$A^N\mathbf{v} = \lambda^N\mathbf{v}, \tag{9.1}$$

for all $N > 0$. Thus we completely know how the system behaves for any eigenpair $(\lambda, \mathbf{v})$.

Now suppose A admits an eigenbasis. That is, there exists a basis $\mathbf{v}_1, \mathbf{v}_2, \ldots \mathbf{v}_n$ of $\mathbb{F}^n$ such that $A\mathbf{v}_i = \lambda_i\mathbf{v}_i$ for $1 \leq i \leq n$. Writing an arbitrary $\mathbf{v} \in \mathbb{F}^n$ as $\mathbf{v} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_n\mathbf{v}_n$, it follows that

$$A\mathbf{v} = \sum a_i A\mathbf{v}_i = \sum a_i\lambda\mathbf{v}_i.$$

But by (9.1), an eigenbasis for A is also one for A^N. Hence,

$$A^N\mathbf{v} = \sum a_i A^N\mathbf{v}_i = \sum a_i\lambda^N\mathbf{v}_i. \tag{9.2}$$

The key to understanding a dynamical system with arbitrary initial value $\mathbf{v}_0$ is therefore to find an eigenbasis for A, expand $\mathbf{v}_0$ in terms of this basis and apply (9.2).

9.1.2 The Eigenvalue Problem

Let A be a square matrix over the field $\mathbb{F}$, i.e. $A \in \mathbb{F}^{n\times n}$. Suppose we want to find all eigenpairs $(\lambda, \mathbf{v})$ for A with $\lambda \in \mathbb{F}$. Now in the equation $A\mathbf{v} = \lambda\mathbf{v}$, the variables are λ and the components of $\mathbf{v}$, so the right hand side is a nonlinear equation. But by slightly reformulating the problem, we can see what we need to do more clearly. The equation $A\mathbf{v} = \lambda\mathbf{v}$ is equivalent to

$$(A - \lambda I_n)\mathbf{v} = \mathbf{0}. \tag{9.3}$$

This is a homogeneous linear system which has a nontrivial solution if and only if $A - \lambda I_n$ has a nontrivial null space, which is the case if and only if

$$\det(A - \lambda I_n) = 0. \tag{9.4}$$

This equation is called the *characteristic equation* of A. Thus $\lambda \in \mathbb{F}$ belongs to an eigenpair $(\lambda, \mathbf{v})$ with $\mathbf{v} \in \mathbb{F}^n$ if and only if λ is a root of $\det(A-\lambda I_n) = 0$. Thus the nonlinear part of the eigenvalue problem is to find the roots in $\mathbb{F}$ of the *characteristic polynomial* $\det(A - \lambda I_n)$. Once we have a $\lambda \in \mathbb{F}$ satisfying (9.4), the second problem is the straightforward linear problem of finding the null space $\mathcal{N}(A - \lambda I_n)$.

Example 9.1. Let's consider the real matrix

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

The 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. We therefore seek the $\lambda \in \mathbb{R}$ such that $\det(A - \lambda I_2) = 0$. Now

$$\det(A - \lambda I_2) = (1-\lambda)^2 - 2 \cdot 2 = \lambda^2 - 2\lambda - 3 = 0.$$

Since $\lambda^2 - 2\lambda - 3 = (\lambda - 3)(\lambda + 1)$, the eigenvalues of A are 3 and -1, both real. We can now proceed to finding corresponding eigenvectors by finding the null spaces $\mathcal{N}(A - 3I_2)$ and $\mathcal{N}(A + I_2)$. Clearly,

$$\mathcal{N}(A - 3I_2) = \mathcal{N}(\begin{pmatrix} -2 & 2 \\ 2 & -2 \end{pmatrix}) = \mathbb{R} \begin{pmatrix} 1 \\ 1 \end{pmatrix},$$

and

$$\mathcal{N}(A + I_2) = \mathcal{N}(\begin{pmatrix} 2 & 2 \\ 2 & 2 \end{pmatrix}) = \mathbb{R} \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

A consequence of this calculation is that everything combines into a single matrix equation

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

This says $AP = PD$. But since the columns of P are independent (by inspection), P is invertible, and we get the factorization $A = PDP^{-1}$. At this point, we say A has been diagonalized.

Obtaining a factorization $A = PDP^{-1}$ is the key to the problem of computing the powers A^N of A. For example,

$$A^2 = (PDP^{-1})(PDP^{-1}) = PDI_2DP^{-1} = PD^2P^{-1},$$

and, generalizing this to any positive integer N,

$$A^N = PD^NP^{-1}.$$

In the above examples, we solved the diagonalization problem. That is, given A we constructed a matrix P such that $A = PDP^{-1}$. Recall from Definition 7.4 that two matrices A and B in $\mathbb{F}^{n\times n}$ are said to be *similar* if there exists an invertible $M \in \mathbb{F}^{n\times n}$ so that $B = MAM^{-1}$. Thus the diagonalization problem for A is to find a diagonal matrix similar to A.

9.1.3 Fibonacci Revisted

Let us now find the Fibonacci sequence. The characteristic equation of F is $\lambda^2 - \lambda + 1 = 0$, so its eigenvalues are

$$\phi = \frac{1+\sqrt{5}}{2}, \quad \mu = \frac{1-\sqrt{5}}{2}.$$

One checks that $\mathcal{N}(F-\phi I_2) = \mathbb{R}\begin{pmatrix}\phi\\1\end{pmatrix}$, and $\mathcal{N}(F-\mu I_2) = \mathbb{R}\begin{pmatrix}\mu\\1\end{pmatrix}$. Therefore,

$$F = \begin{pmatrix}\phi & \mu\\1 & 1\end{pmatrix}\begin{pmatrix}\phi & 0\\0 & \mu\end{pmatrix}\begin{pmatrix}\phi & \mu\\1 & 1\end{pmatrix}^{-1}.$$

Hence

$$\begin{pmatrix}a_{m+1}\\a_m\end{pmatrix} = F^m\begin{pmatrix}a_1\\a_0\end{pmatrix} = \begin{pmatrix}\phi & \mu\\1 & 1\end{pmatrix}\begin{pmatrix}\phi^m & 0\\0 & \mu^m\end{pmatrix}\begin{pmatrix}\phi & \mu\\1 & 1\end{pmatrix}^{-1}\begin{pmatrix}a_1\\a_0\end{pmatrix}.$$

To take a special case, let $a_0 = 0$ and $a_1 = 1$. Then this leaves us with the identity

$$a_m = \frac{\phi^m - \mu^m}{\phi - \mu} = \frac{1}{\sqrt{5}\cdot 2^m}((1+\sqrt{5})^m - (1-\sqrt{5})^m). \tag{9.5}$$

Thus

$$\lim_{m\to\infty}\frac{a_{m+1}}{a_m} = \lim_{m\to\infty}\frac{\phi^{m+1}-\mu^{m+1}}{\phi^m - \mu^m} = \phi,$$

since $\lim_{m\to\infty}(\mu/\phi)^m = 0$. Therefore, for large m, the ratio a_{m+1}/a_m is approximately ϕ. Some 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.

9.1.4 An Infinite Dimensional Example

Now suppose that V is an arbitrary vector space over $\mathbb{R}$ and $T : V \to V$ is a linear transformation. In the infinite dimensional setting, it is customary to call a linear transformation a *linear operator*. The eigenvalue problem for T is still the same: to find scalars $\lambda \in \mathbb{R}$ so that there exists a non zero $\mathbf{v} \in V$ such that $T(\mathbf{v}) = \lambda\mathbf{v}$. However, since V is not assumed to be finite dimensional, the above method for solving the characteristic equation doesn't

work. An alternate procedure is to look for finite dimensional subspaces W of V so that $T(W) \subset W$. But in the infinite dimensional setting, there is in general no simple technique for finding eigenvalues. This has lead to the development of much more sophisticated techniques.

In the following example, we consider a linear operator which arises in differential equations.

Example 9.2. Let $V = C^\infty(\mathbb{R})$ be the space of real-valued functions on $\mathbb{R}$ which have derivatives of all orders. Since the derivative of such a function also has derivatives of all orders, differentiation defines a linear operator $D : C^\infty(\mathbb{R}) \to C^\infty(\mathbb{R})$. That is, $D(f) = f'$. It is clear that the exponential function $f(x) = e^{rx}$ is an eigenvector of D with corresponding eigenvalue r. Thus (r, e^{rx}) is an eigenpair for D. In this context, eigenvectors are usually called *eigenfunctions*. Considering D^2 instead of D, we easily see that for any integers m and n, $\cos mx$ and $\sin nx$ are also eigenfunctions with corresponding eigenvalues $-m^2$ and $-n^2$ respectively.

This is a fundamental example, and we will return to it in Chapter 11.

9.2 Eigentheory: the Basic Definitions

We will now begin to study eigentheory in earnest. A few of the definitions made in the previous informal section will be repeated here.

9.2.1 Eigenpairs for Linear Transformations and Matrices

Let $\mathbb{F}$ be a field, and suppose V is a finite dimensional vector space over $\mathbb{F}$.

Definition 9.1. Suppose $T : V \to V$ is a linear map. Then a pair $(\lambda, \mathbf{v})$, where $\lambda \in \mathbb{F}$ and $\mathbf{v} \in V$, is called an *eigenpair* for T if $\mathbf{v} \neq \mathbf{0}$ and

$$T(\mathbf{v}) = \lambda \mathbf{v}. \tag{9.6}$$

If $(\lambda, \mathbf{v})$ is an eigenpair for T, we call λ an $\mathbb{F}$-*eigenvalue*, or, simply, an *eigenvalue* of T and $\mathbf{v}$ an *eigenvector* of T corresponding to λ. An eigenpair for an $n \times n$ matrix $A \in \mathbb{F}^{n \times n}$ is just an eigenpair for the linear transformation $T_A : \mathbb{F}^n \to \mathbb{F}^n$.

Thus, $(\lambda, \mathbf{v})$ is an eigenpair for A if and only if $\lambda \in \mathbb{F}$, $\mathbf{v} \in \mathbb{F}^n$ is nonzero and $A\mathbf{v} = \lambda \mathbf{v}$. The following Proposition gives some basic general facts about eigenpairs for A. The reader can easily reformulate them for linear transformations.

Proposition 9.1. *Suppose A is a square matrix over $\mathbb{F}$ and $(\lambda, \mathbf{v})$ is an eigenpair for A. Then for any scalar $r \in \mathbb{F}$, $(r\lambda, \mathbf{v})$ is an eigenpair for rA. Moreover, for any 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 the null space $\mathcal{N}(A)$ is nontrivial.*

Proof. The proof is left as an exercise. □

9.2.2 The Characteristic Polynomial

To attack the general eigenpair problem for a linear transformation $T : V \to V$, we will first concentrate on the case where $V = \mathbb{F}^n$ and T is a matrix linear transformation. The next definition is motivated by the approach we used in Section 9.1.2.

Definition 9.2. Given $A \in \mathbb{F}^{n \times n}$, we call $p_A(\lambda) = \det(A - \lambda I_n)$ the *characteristic polynomial* of A. The *characteristic equation* of $A \in \mathbb{F}^{n \times n}$ is defined to be the equation $p_A(\lambda) = 0$.

In order to show that the definition makes sense, we should in fact show that $\det(A - \lambda I_n)$ is a polynomial. We now do this and somewhat more.

Proposition 9.2. *If $A \in \mathbb{F}^{n\times n}$, the function $p_A(\lambda) = \det(A - \lambda I_n)$ is a polynomial of degree n in λ having its coefficients $\mathbb{F}$. Its leading term is $(-1)^n\lambda^n$ and its constant term is $\det(A)$. The $\mathbb{F}$-eigenvalues of A are the roots of $p_A(\lambda) = 0$ in $\mathbb{F}$.*

Proof. The first two statements are more or less obvious consequences of the definition of the determinant. For the third, note that $\lambda \in \mathbb{F}$ is a root of $\det(A - \lambda I_n) = 0$ if and only if there exists a nonzero $\mathbf{x} \in \mathbb{F}^n$ such that $(A - \lambda I_n)\mathbf{x} = \mathbf{0}$. □

We will see in the examples below that sometimes not every root of $p_A(\lambda) = 0$ lies in the field $\mathbb{F}$ over which A is defined. Nevertheless, it is convenient to refer to all the roots of the characteristic polynomial of A as eigenvalues of A, even though we may only be interested in the $\mathbb{F}$-eigenvalues. In fact, there is a theorem in algebra which says that there exists a field containing $\mathbb{F}$ which also contains all the roots of $p_A(\lambda)$. In particular, when $A \in \mathbb{R}^{n\times n}$, the Fundamental Theorem of Algebra (Theorem 4.18) tells us that the characteristic equation $p_A(\lambda) = 0$ has n complex roots. Thus A has n complex eigenvalues. However, only the real roots, which happen to occur in conjugate pairs, are relevant for A if one is studying A as a real matrix.

Here is an example of a linear transformation on $\mathbb{R}^2$, whose matrix is diagonalizable over $\mathbb{C}$, but not over $\mathbb{R}$.

Example 9.3. The characteristic equation of the matrix

$$J = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

is $\lambda^2 + 1 = 0$. Hence J has no real eigenvalues since the roots of $\lambda^2 + 1 = 0$ are $\pm\, i$, hence pure imaginary. Of course, it is geometrically obvious J has no real eigenvalues. In fact, J is the rotation of $\mathbb{R}^2$ through $\pi/2$, and no nonzero vector can be rotated through $\pi/2$ into a multiple of itself.

Example 9.4. On the other hand, if we think of J as a 2×2 complex matrix, it has complex eigenvalues $\pm\, i$, illustrating why the term $\mathbb{F}$-eigenvalues is useful. Solving for the corresponding eigenvectors gives eigenpairs $(i, \begin{pmatrix} 1 \\ -i \end{pmatrix})$ and $(-i, \begin{pmatrix} 1 \\ i \end{pmatrix})$. Notice that the eigenpairs are conjugate in an obvious sense. This is because the matrix J is real.

Example 9.5. If $A = \left(\begin{smallmatrix} 1 & 2 \\ 2 & -1 \end{smallmatrix}\right)$, then $A - \lambda I_2 = \left(\begin{smallmatrix} 1-\lambda & 2 \\ -2 & -1-\lambda \end{smallmatrix}\right)$, so the characteristic polynomial of A is $|A - \lambda I_2| = (1-\lambda)(-1-\lambda) - (2)(2) = \lambda^2 - 5$. The eigenvalues of A are $\pm\sqrt{5}$. Both eigenvalues are real.

Example 9.6. Let $K = \left(\begin{smallmatrix} 0 & -i \\ i & 0 \end{smallmatrix}\right)$. The characteristic polynomial of K is $\lambda^2 - 1$, so the eigenvalues of K are ± 1. Thus K is a complex matrix that has real eigenvalues. Notice that $K = iJ$, where J is the rotation through $\pi/2$, so Proposition 9.1 in fact tells us that its eigenvalues are i times those of J.

Example 9.7 (Triangular Matrices)**.** Suppose $U = (u_{ij}) \in \mathbb{F}^{n\times n}$ is upper triangular. Then

$$p_U(\lambda) = (u_{11} - \lambda)(u_{22} - \lambda)\cdots(u_{nn} - \lambda).$$

Hence the eigenvalues of an upper triangular matrix are its diagonal entries. A similar remark holds for lower triangular matrices. We will show below that every $A \in \mathbb{C}^{n\times n}$ is similar to an upper triangular matrix.

9.2.3 Further Remarks on Linear Transformations

So far we haven't looked at the relationship between the eigenvalues of a linear transformation and the eigenvalues of a matrix representing the linear transformation. The first step in this direction is

Proposition 9.3. *Two similar matrices have the same characteristic polynomial.*

Proof. Suppose A and B are similar, say $B = MAM^{-1}$. Then

$$\begin{aligned} \det(B - \lambda I_n) &= \det(MAM^{-1} - \lambda I_n) \\ &= \det(M(A - \lambda I_n)M^{-1}) \\ &= \det(M)\det(A - \lambda I_n)\det(M^{-1}) \end{aligned}$$

Since $\det(M^{-1}) = \det(M)^{-1}$, the proof is done. □

This Proposition allows us to extend the definition of the characteristic polynomial to an arbitrary linear transformation $T : V \to V$, provided V is finite dimensional. Indeed, suppose $\mathcal{B}$ is a basis of V and $A = \mathcal{M}_{\mathcal{B}}^{\mathcal{B}}(T)$ is the matrix of T with respect to this basis.

Definition 9.3. If V is a finite dimensional vector space over $\mathbb{F}$ and $T : V \to V$ is linear, then we define the *characteristic polynomial* of T to be $p_T(\lambda) = p_A(\lambda)$. The *characteristic equation of* T is defined to be the equation $p_T(\lambda) = 0$.

We need to show the definition makes sense. But any two matrices $A = \mathcal{M}_{\mathcal{B}}^{\mathcal{B}}(T)$ and $A' = \mathcal{M}_{\mathcal{B}'}^{\mathcal{B}'}(T)$ representing T with respect to to different bases $\mathcal{B}$ and $\mathcal{B}'$ are similar. That is, $A' = PAP^{-1}$. Thus, by Proposition 9.3, $p_A(\lambda) = p_{A'}(\lambda)$. Hence the characteristic polynomial of T is unambiguously defined.

We will now show how the eigenpairs for $\mathcal{M}_{\mathcal{B}}^{\mathcal{B}}(T)$ are related to the eigenpairs for T.

Proposition 9.4. *Let V be a finite dimensional vector space over $\mathbb{F}$ and suppose $T : V \to V$ is linear. Then any root $\mu \in \mathbb{F}$ of the characteristic polynomial of T is an eigenvalue of T, and conversely.*

Proof. Let $A = \mathcal{M}_{\mathcal{B}}^{\mathcal{B}}(T)$ where $\mathcal{B}$ is the basis given by $\mathbf{v}_1, \ldots, \mathbf{v}_n$ of V. Suppose $\mu \in \mathbb{F}$ is a root of the characteristic equation $\det(A - \mu I_n) = 0$, and let $(\mu, \mathbf{x})$ be an eigenpair for A. Thus $A\mathbf{x} = \mu\mathbf{x}$. Let $\mathbf{x} = (x_1, \ldots, x_n)^T$ and put $\mathbf{v} = \sum_i x_i \mathbf{v}_i$. I claim $(\mu, \mathbf{v})$ is an eigenpair for T. For

$$\begin{aligned} T(\mathbf{v}) &= \sum_i x_i T(\mathbf{v}_i) \\ &= \sum_{i,j} x_i a_{ji} \mathbf{v}_j \\ &= \sum_j \mu x_j \mathbf{v}_j \\ &= \mu\mathbf{v} \end{aligned}$$

Since some $x_j \neq 0$, it follows that $\mathbf{v}$ is non zero, so $(\mu, \mathbf{v})$ is indeed an eigenpair for T. For the converse, just reverse the argument. □

The upshot of the previous two Propositions is that the eigentheory of linear transformations reduces to the eigentheory of matrices.

9.2.4 Formulas for the Characteristic Polynomial

The characteristic polynomial $p_A(\lambda)$ of a 2×2 matrix $A = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ over an arbitrary field $\mathbb{F}$ has a nice form, which we will explain and then generalize. First, define the *trace* of an $n \times n$ matrix to be the sum of its diagonal entries. Thus, in the 2×2 case above, $\mathrm{Tr}(A) = a + d$. Note that

$$p_A(\lambda) = (a - \lambda)(d - \lambda) - bc = \lambda^2 - (a + d)\lambda + (ad - bc).$$

Hence,

$$p_A(\lambda) = \lambda^2 + \mathrm{Tr}(A)\lambda + \det(A). \tag{9.7}$$

The quadratic formula therefore gives the eigenvalues of A in the form

$$\lambda = \frac{1}{2}\big(-\mathrm{Tr}(A) \pm \sqrt{\mathrm{Tr}(A)^2 - 4\det(A)}\big). \tag{9.8}$$

Hence if A is real, it has real eigenvalues if and only if the discriminant $\Delta(A) := \mathrm{Tr}(A)^2 - 4\det(A)$ is non negative: i.e. $\Delta(A) \geq 0$. If $\Delta(A) = 0$, the roots are real and repeated. If $\Delta(A) < 0$, the roots are complex and unequal. In this case, the roots are conjugate complex numbers. That is, they have the form $\lambda,\ \overline{\lambda}$ for some $\lambda \neq 0$.

By factoring the characteristic polynomial as

$$(\lambda - \lambda_1)(\lambda - \lambda_2) = \lambda^2 - (\lambda_1 + \lambda_2)\lambda + \lambda_1\lambda_2$$

and comparing coefficients, we immediately see that:

(i) the trace of A is the sum of the eigenvalues of A:

$$\mathrm{Tr}(A) = \lambda_1 + \lambda_2,$$

(ii) the determinant of A is the product of the eigenvalues of A:

$$\det(A) = \lambda_1\lambda_2.$$

For $n > 2$, the characteristic polynomial is more difficult to compute, though it is still true that the trace is the sum of the roots of the characteristic polynomial and the determinant is their product. Using row operations to compute a characteristic polynomial isn't very practical (see Example 8.8), so when computing by hand, the only obvious way to proceed is to use the Laplace expansion.

There's an important warning that has to be issued here. **Whenever you are computing the characteristic polynomial of a matrix, never (repeat, never) row reduce the matrix (or even do a single row operation on the matrix) before computing its characteristic polynomial**. There is absolutely no reason the characteristic polynomials of A and EA should be the same. If they were, all invertible matrices would have the same eigenvalues.

On the other hand, we will now exhibit (without proof) a beautiful formula for the characteristic polynomial of an arbitrary matrix in the spirit

of (9.7) involving the so called *principal minors* of A. Since $p_A(\lambda)$ is a polynomial in λ of degree n with leading coefficient $(-1)^n\lambda^n$ and constant term $\det(A)$, we can write

$$p_A(\lambda) = (-1)^n\lambda^n + (-1)^{n-1}\sigma_1(A)\lambda^{n-1} + (-1)^{n-2}\sigma_2(A)\lambda^{n-2} +$$

$$+ \cdots + (-1)^1\sigma_{n-1}(A)\lambda + \det(A), \tag{9.9}$$

where the $\sigma_i(A)$, $1 \le i \le n-1$, are the remaining coefficients.

Theorem 9.5. *The coefficients $\sigma_i(A)$ for $1 \le i \le n$ are given by*

$$\sigma_i(A) := \sum \big(\textit{all principal } i \times i \textit{ minors of } A\big), \tag{9.10}$$

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 the same $(n-i)$ columns.

By definition, the principal 1×1 minors are just the diagonal entries of A, since deleting all but the ith row and column leaves just the diagonal entry a_{ii}. Hence

$$\sigma_1(A) = a_{11} + a_{22} + \cdots + a_{nn}$$

so

$$\sigma_1(A) = \mathrm{Tr}(A).$$

Clearly the constant term $\sigma_n(A) = \det(A)$. In general, the number of $j \times j$ minors of A is the binomial coefficient

$$\binom{n}{j} = \frac{n!}{j!(n-j)!}$$

counting the number of subsets with j elements contained in a set with n elements. 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. Nevertheless, using Theorem 9.5 is by an effective way to expand $d\!et(A - \lambda I_n)$. You should even be able to do the 3×3 case without pencil and paper (or a calculator).

Example 9.8. For example, let

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

Then

$$\det(A - \lambda I_3) = -\lambda^3 + (-1)^2(3 - 1 + 0)\lambda^2 +$$

$$(-1)^1(\det \begin{pmatrix} 3 & -2 \\ 3 & -1 \end{pmatrix} + \det \begin{pmatrix} -1 & -3 \\ -2 & 0 \end{pmatrix} + \det \begin{pmatrix} 3 & -2 \\ 1 & 0 \end{pmatrix})\lambda + \det(A).$$

Thus the characteristic polynomial of A is

$$p_A(\lambda) = -\lambda^3 + 2\lambda^2 + \lambda - 2.$$

The question of how to find the roots of a characteristic polynomial often arises, but there is no obvious answer. In the 2×2, 3×3 and 4×4 cases, there are general formulas, though they aretwo unwieldy to write down here. But otherwise, there aren't any general methods for finding the roots of a polynomial. Solving the eigenvalue problem for a given square matrix is a problem which is usually approached by other methods, such as Newton's method or the QR algorithm, which we will treat in Chapter 12.

For a matrix with integral entries, the **rational root test** is helpful. This test says that if

$$p(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$$

is a polynomial with integer coefficients $a_0, \ldots a_n$, then the only possible rational roots have the form p/q, where p and q are integers without any common factors, p divides a_0 and q divides a_n. In particular, if the leading coefficient $a_n = 1$, then $q = \pm 1$, so the only possible rational roots are the inegers which divide the constant term a_0. Therefore we obtain

Proposition 9.6. *If A is a matrix with integer entries, then the only possible rational eigenvalues of A are the integers dividing* $\det(A)$.

Since the characteristic polynomial of the matrix A in the previous example is $-\lambda^3 + 2\lambda^2 + \lambda - 2$ the only possible rational eigenvalues are the divisors of 2, that is $\pm$ 1 and $\pm$ 2. Checking these possibilities, we find that $\pm$ 1 and 2 are roots, so these are the eigenvalues of A.

Note that the coefficients of $p_A(\lambda)$ are certain explicit functions of its roots. For if $p_A(\lambda)$ has roots $\lambda_1, \ldots, \lambda_n$, then

$$\begin{aligned} p_A(\lambda) &= (\lambda_1 - \lambda)(\lambda_1 - \lambda)\ldots(\lambda_1 - \lambda) \\ &= (-1)^n(\lambda)^n + (-1)^{n-1}(\lambda_1 + \lambda_2 + \ldots \lambda_n)\lambda^{n-1} + \cdots + \lambda_1\lambda_2\cdots\lambda_n \end{aligned}$$

Thus we obtain a generalization of what we showed in the 2×2 case. For example, we have

Proposition 9.7. *The trace of a matrix A is the sum of the roots of its characteristic polynomial, and similarly, the determinant is the product of the roots of its characteristic polynomial.*

The functions $\sigma_i(\lambda_1, \ldots, \lambda_n)$ expressing the coefficients $\sigma_i(A)$ as functions of $\lambda_1, \ldots, \lambda_n$ are called the *elementary symmetric functions* . For example,

$$\sigma_2(\lambda_1, \ldots, \lambda_n) = \sum_{i<j} \lambda_i \lambda_j.$$

Exercises

Exercise 9.1. Find the characteristic polynomial and real eigenvalues of the following matrices:

(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 9.2. 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 9.3. The following matrix A appeared 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 try to decide how many real eigenvalues A has.

Exercise 9.4. Find the characteristic polynomial of a 4×4 matrix A if you know that three eigenvalues of A are ± 1 and 2 and you also know that $\det(A) = 6$.

Exercise 9.5. Using only the definitions, prove Proposition 9.1.

Exercise 9.6. 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}.

Exercise 9.7. Show that two similar matrices have the same trace and determinant.

Exercise 9.8. True or False: Two matrices with the same characteristic polynomial are similar.

Exercise 9.9. If A is a square matrix, determine whether or not A and A^T have the same characteristic polynomial, hence the same eigenvalues.

Exercise 9.10. Show that 0 is an eigenvalue of A if and only if A is singular, that is, A^{-1} does not exist.

Exercise 9.11. True or False: If λ is an eigenvalue of A and μ is an eigenvalue of B, then $\lambda + \mu$ is an eigenvalue of $A + B$.

Exercise 9.12. An $n \times n$ matrix such that $A^k = O$ for some positive integer k is called *nilpotent*.

(a) Show all eigenvalues of a nilpotent matrix A are 0.

(b) Hence conclude that the characteristic polynomial of A is $(-1)^n \lambda^n$. In particular, the trace of a nilpotent matrix is 0.

(c) Find a 3×3 matrix A so that $A^2 \neq O$, but $A^3 = O$. (Hint: look for an upper triangular example.)

Exercise 9.13. Find the characteristic polynomial of the 5×5 X-Files matrix

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

Exercise 9.14. 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 we gave for $n = 2$ does not extend. First observe that if $p(x)$ is a polynomial with real coefficients, then $\overline{p(x)} = p(\overline{x})$.)

Exercise 9.15. 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 9.16. Find eigenpairs for the two eigenvalues of the rotation R_θ of $\mathbb{R}^2$. (Note, the eigenvalues are complex.)

Exercise 9.17. Show that in general, the only possible real eigenvalues of an $n \times n$ real orthogonal matrix are ± 1.

Exercise 9.18. Let

$$J = \begin{pmatrix} 0 & -I_2 \\ I_2 & 0 \end{pmatrix}.$$

Show that J cannot have any real eigenvalues, and find all its complex eigenvalues.

Exercise 9.19. Suppose A is $n \times n$ and invertible. Show that for any $n \times n$ matrix B, AB and BA have the same characteristic polynomial.

Exercise 9.20. * Find the characteristic polynomial of

$$\begin{pmatrix} a & b & c \\ b & c & a \\ c & a & b \end{pmatrix},$$

where a, b, c are all real. (Note that the second matrix in Problem 2 is of this type. What does the fact that the trace is an eigenvalue say?)

Exercise 9.21. Find the elementary symmetric functions $\sigma_i(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$ for $i = 1, 2, 3, 4$ 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.

9.3 Eigenvectors and Diagonalizability

Let V be a finite dimensional vector space over the field $\mathbb{F}$. We now begin to study the fundamental question of when a linear transformation $T : V \to V$ is semi-simple, or, equivalently, when is its matrix diagonalizable. In other words, we wish to study the question of when V admits a basis consisting of eigenvectors of T (i.e. an eigenbasis). The purpose of this section is to give an elementary characterization. A complete answer as to which linear transformations are semi-simple is given in Chapter 13.

9.3.1 Semi-simple Linear Transformations and Diagonalizability

Recall that a basis of V consisting of eigenvectors of a linear transformation $T : V \to V$ is called an *eigenbasis associated to* T. A linear transformation T which admits an eigenbasis is said to be *semi-simple*. Similarly, an eigenbasis for a matrix $A \in \mathbb{F}^{n\times n}$ is an eigenbasis for the linear transformation T_A. A matrix $A \in \mathbb{F}^{n\times n}$ is said to be *diagonalizable over* $\mathbb{F}$ if and only if there exists an invertible matrix $P \in \mathbb{F}^{n\times n}$ and a diagonal matrix $D \in \mathbb{F}^{n\times n}$ such that $A = PDP^{-1}$.

Recall that we showed in Proposition 7.14 that if $A = PDP^{-1}$, where D is diagonal, then the ith diagonal entry λ_i of D is an eigenvalue of A and the ith column of P is a corresponding eigenvector. This is easily seen from first principles. Since $A = PDP^{-1}$, we have $AP = PD$. Letting $\mathbf{p}_i$ denote the ith column of P, and equating the ith columns of AP and PD, we get that $A\mathbf{p}_i = \lambda_i\mathbf{p}_i$ for each i.

Proposition 9.8. *Suppose $A \in \mathbb{F}^{n\times n}$ and $A = PDP^{-1}$, where P and D are also in $\mathbb{F}^{n\times n}$ and D is diagonal. Then the columns of P are an eigenbasis for $\mathbb{F}^n$ associated to A. Hence a diagonalizable matrix admits an eigenbasis. Conversely, a matrix $A \in \mathbb{F}^{n\times n}$ which admits an eigenbasis for $\mathbb{F}^n$ is diagonalizable. Finally, if $\mathcal{B} = \{\mathbf{p}_1, \ldots, \mathbf{p}_n\}$ is an eigenbasis of $\mathbb{F}^n$ for A and $P = \begin{pmatrix}\mathbf{p}_1 & \cdots & \mathbf{p}_n\end{pmatrix}$, then the identity $A = PDP^{-1}$ is equivalent to saying $\mathcal{M}_{\mathcal{B}}^{\mathcal{B}}(T_A) = D$.*

Proof. The first assertion was just proved. For the converse, let $\mathcal{B}$ be a basis of $\mathbb{F}^n$, and let P be the matrix whose columns are these basis vectors in some order. Then if $A\mathbf{p}_i = \lambda_i\mathbf{p}_i$, we have $AP = PD$, so $A = PDP^{-1}$ since P is invertible. The last assertion is just a restatement of the definitions. □

9.3.2 Eigenspaces

We now define the eigenspaces of a linear transformation $T : V \to V$.

Definition 9.4. Let λ be an eigenvalue of T. Then the *eigenspace* of λ corresponding to λ is the subspace

$$E_\lambda = \{\mathbf{v} \in V \mid T(\mathbf{v}) = \lambda \mathbf{v}\}$$

of V. In particular, if $V = \mathbb{F}^n$ and $T = T_A$ for some $A \in \mathbb{F}^{n\times n}$, then $E_\lambda = \mathcal{N}(A - \lambda I_n)$. The dimension of E_λ is called the *geometric multiplicity of* λ.

Example 9.9. Consider a simple example, say $T = T_A$ where

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

Then $p_A(\lambda) = \lambda^2 - 2\lambda - 3 = (\lambda - 3)(\lambda + 1)$, so the eigenvalues of A are $\lambda = 3$ and -1. Now

$$A - (-1)I_2 = \begin{pmatrix} 2 & 2 \\ 2 & 2 \end{pmatrix} \quad \text{and} \quad A - 3I_2 = \begin{pmatrix} -2 & 2 \\ 2 & -2 \end{pmatrix}.$$

Thus $E_{-1} = \mathcal{N}(A + I_2) = \mathbb{R}\begin{pmatrix} 1 \\ -1 \end{pmatrix}$ and $E_3 = \mathcal{N}(A - 3I_2) = \mathbb{R}\begin{pmatrix} 1 \\ 1 \end{pmatrix}$. Thus we have found an eigenbasis of $\mathbb{R}^2$, so A is diagonalizable (over $\mathbb{R}$).

Example 9.10. Here is another calculation. Let

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

Then $p_A(\lambda) = -\lambda^3 + 2\lambda^2 + \lambda - 2$. The eigenvalues of A are $\pm\, 1,\ 2$. One finds that $\mathcal{N}(A - I_3) = \mathbb{R}(1, 0, 1)^T$, $\mathcal{N}(A + I_3) = \mathbb{R}(1, 1, 1)^T$, and $\mathcal{N}(A - 2I_3) = \mathbb{R}(0, 1, -1)^T$, where $\mathbb{R}\mathbf{v}$ is the line spanned by $\mathbf{v}$.

In the above examples, the characteristic polynomial has simple roots. In general, a polynomial $p(x)$ is said to have *simple roots* if and only if it has no linear factors of the form $(x - r)^2$, where r is a scalar. The *algebraic multiplicity* of a root r of $p(x) = 0$ is the largest value $k > 0$ such that $(x - r)^k$ divides $p(x)$. Clearly a polynomial has simple roots if and only if each root has algebraic multiplicity one.

The following result gives a well known criterion for diagonalizability in terms of simple roots. We will omit the proof, since a more general result is proven in the next section (see Proposition 9.4).

Proposition 9.9. *An $n \times n$ matrix A over $\mathbb{F}$ with n distinct eigenvalues in $\mathbb{F}$ is diagonalizable. More generally, if V be a finite dimensional vector space over $\mathbb{F}$ and $T : V \to V$ is a linear transformation with $\dim V$ distinct eigenvalues in $\mathbb{F}$, then T is semi-simple.*

Consider another example.

Example 9.11. The counting matrix

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

has characteristic polynomial $p_C(x) = -\lambda^3 + 15\lambda^2 - 21\lambda$, hence eigenvalues 0 and $\frac{1}{2}(15 \pm \sqrt{151})$. The eigenvalues of C are real and distinct, hence C is diagonalizable over $\mathbb{R}$.

It is worth mentioning that there exists a well know test for simple roots. We thus have a criterion for determining whether A has distinct eigenvalues.

Proposition 9.10. *A polynomial $p(x)$ has simple roots if and only if the equations $p(x) = 0$ and $p'(x) = 0$ have no root in common. In particular, a square matrix A has simple eigenvalues exactly when $p_A(\lambda)$ and $(p_A)'(\lambda)$ have no common roots.*

Proof. We leave this as an exercise. □

Example 9.12. Recall the Good Will Hunting matrix

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

The characteristic polynomial of A is

$$p_A\lambda) = \lambda^4 - 7\lambda^2 - 2\lambda + 4.$$

This polynomial has -1 as a root and factors as

$$(\lambda + 1)(\lambda^3 - \lambda^2 - 6\lambda + 4).$$

To show that $p(\lambda)$ has four distinct roots, it suffices to show $q(\lambda) = \lambda^3 - \lambda^2 - 6\lambda + 4 = 0$ has distinct roots since -1 is obviously not a root of q. Now $q'(\lambda) = 3\lambda^2 - 2\lambda - 6$ which has roots

$$r = \frac{2 \pm \sqrt{76}}{6}.$$

Now

$$q(\frac{2 + \sqrt{76}}{6}) < 0,$$

while

$$q(\frac{2 - \sqrt{76}}{6}) > 0.$$

Since the points where $q' = 0$ are not zeros of q, q has three distinct roots. Therefore, p_A has simple roots, so A has 4 distinct eigenvalues. Furthermore, it's clear that all the roots of q are real, so A is diagonalizable over $\mathbb{R}$.

Exercises

Exercise 9.22. Diagonalize the following matrices if possible:

$$A = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} -3 & 0 & -4 & -4 \\ 0 & 2 & 1 & 1 \\ 4 & 0 & 5 & 4 \\ -4 & 0 & -4 & -3 \end{pmatrix}.$$

Exercise 9.23. Diagonalize the three matrices in Exercise 9.1.

Exercise 9.24. Decide whether the Good Will Hunting matrix (cf Exercise 9.3) can be diagonalized.

Exercise 9.25. Suppose A and B are similar and $(\lambda, \mathbf{v})$ is an eigenpair for A. Find an eigenpair for B.

Exercise 9.26. 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 9.27. Let A be a real 3×3 matrix so that A and $-A$ are similar. Show that

(a) $\det(A) = \mathrm{Tr}(A) = 0$,

(b) 0 is an eigenvalue of A, and

(c) if some eigenvalue of A is non-zero, then A is diagonalizable over $\mathbb{C}$.

Exercise 9.28. Find an example of two real matrices which have the same characteristic polynomial which are not similar.

Exercise 9.29. A 4×4 matrix has eigenvalues ± 1, trace 3 and determinant 0. Can A be diagonalized?

Exercise 9.30. Let A be a 3×3 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 9.31. What is the characteristic polynomial of A^3 in terms of that of A?

Exercise 9.32. Prove the test for simple eigenvalues given by Proposition 9.10.

Exercise 9.33. Diagonalize $J = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$. More generally do the same for R_θ for all $\theta \neq 0$.

Exercise 9.34. Let $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ be a real matrix such that $bc > 0$.

(a) Show that A has distinct real eigenvalues.

(b) Show that one eigenvector of A lies in the first quadrant and another in the second quadrant.

Exercise 9.35. Let $V = \mathbb{R}^{n\times n}$ and let $\mathbb{T} : V \to V$ be the linear map defined by sending $A \in V$ to A^T. That is, $\mathbb{T}(A) = A^T$. Show that the only eigenvalues of $\mathbb{T}$ are ± 1. Show also that $\mathbb{T}$ is semi-simple by finding an eigenbasis of V for $\mathbb{T}$.

Exercise 9.36. We say that two $n \times n$ matrices A and B are simultaneously diagonalizable if they are diagonalized by the same matrix M. Show that two simultaneously diagonalizable matrices A and B commute (i.e. $AB = BA$).

Exercise 9.37. This is the converse to Exercise 9.36. Suppose that two $n\times n$ matrices A and B commute. Show that if both A and B are diagonalizable, then they are simultaneously diagonalizable. That is, they share a common eigenbasis.

Exercise 9.38. If ϕ ad μ are the eigenvalues of the Fibonacci matrix F of Section 9.1.1, 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 9.1.

9.4 When is a Matrix Diagonalizable?

In the last section, we noted that a matrix $A \in \mathbb{F}^{n\times n}$ with n distinct eigenvalues in $\mathbb{F}$ is diagonalizable. The purpose of this Section is to sharpen this result so that we can say exactly which matrices are diagonalizable. We will also consider examples of nondiagonalizable matrices.

9.4.1 A Characterization

To characterize the diagonalizable matrices, we have to investigate what happens when a matrix A has repeated eigenvalues. The fact that allows us to do this is the following generalization of Proposition 9.9.

Proposition 9.11. *Suppose $A \in \mathbb{F}^{n\times n}$, and let $\lambda_1, \ldots, \lambda_k \in \mathbb{F}$ be distinct eigenvalues of A. Choose an eigenpair $(\lambda_i, \mathbf{w}_i)$ for each λ_i. Then $\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_k$ are linearly independent. Moreover, if we choose a set of linearly independent eigenvectors in E_{λ_i} for each λ_i, $1 \le i \le k$, then the union of these k linearly independent sets of eigenvectors is linearly independent.*

Proof. Let W be the subspace of $\mathbb{F}^n$ spanned by $\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_k$. If these vectors are dependent, let m be the first index such that $\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_m$ independent, but $\mathbf{w}_1, \ldots, \mathbf{w}_m, \mathbf{w}_{m+1}$ are dependent. Then

$$\mathbf{w}_{m+1} = \sum_{i=1}^{m} a_i \mathbf{w}_i. \tag{9.11}$$

Applying A, we obtain $A\mathbf{w}_{m+1} = \sum_{i=1}^{m} a_i A\mathbf{w}_i$, so

$$\lambda_{m+1}\mathbf{w}_{m+1} = \sum_{i=1}^{m} a_i \lambda_i \mathbf{w}_i. \tag{9.12}$$

Multiplying (9.11) by λ_{m+1} and subtracting (9.12) gives

$$\sum_{i=1}^{m} (\lambda_{m+1} - \lambda_i) a_i \mathbf{w}_i = \mathbf{0}.$$

As $\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_m$ are independent, we infer that $(\lambda_{m+1} - \lambda_i)a_i = 0$ for all $1 \le i \le m$. Since each $(\lambda_{m+1} - \lambda_i) \neq 0$, all $a_i = 0$, which contradicts the fact that $\mathbf{w}_{m+1} \neq \mathbf{0}$. Hence, $\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_k$ are independent.

To verify the second assertion, suppose we take a set of linearly independent vectors from each E_{λ_i}, and consider a linear combination of all these

vectors which gives $\mathbf{0}$. Let $\mathbf{v}_i$ be the part of this sum which lies in E_{λ_i}. Hence we have that

$$\sum_{i=1}^{k} \mathbf{v}_i = \mathbf{0}.$$

It follows from the what we just proved that each $\mathbf{v}_i = \mathbf{0}$. Indeed, the first assertion of our Proposition tells us that the non zero $\mathbf{v}_i$ are independent, which would contradict the above identity. Therefore, for every i, the coefficients in the part of the sum involving the independent vectors from E_{λ_i} are all zero. Thus, all coefficients are zero, proving the second claim and finishing the proof. □

We can now characterize the diagonalizable matrices in terms of their eigenspaces.

Proposition 9.12. *Let A be an $n \times n$ matrix over $\mathbb{F}$, and suppose that $\lambda_1, \ \ldots \ , \lambda_k$ are the distinct eigenvalues of A in $\mathbb{F}$. Then A is diagonalizable if and only if*

$$\sum_{i=1}^{k} \dim E_{\lambda_i} = n. \tag{9.13}$$

In that case, the union of the bases of the E_{λ_i} is an eigenbasis of $\mathbb{F}^n$, and we have the direct sum decomposition

$$\mathbb{F}^n = \bigoplus_{i=1}^{k} E_{\lambda_i}.$$

Proof. If A is diagonalizable, then there exists an eigenbasis, and (9.13) holds. By the criterion for a direct sum (Proposition 5.22), we get that $\mathbb{F}^n = \bigoplus_{i=1}^{k} E_{\lambda_i}$. Conversely, if we have (9.13), then Proposition 9.11 implies there are n linearly independent eigenvectors. This impies $\mathbb{F}^n$ admits an eigenbasis. □

In particular, if an $n \times n$ matrix over $\mathbb{F}$ has n distinct eigenvalues in $\mathbb{F}$, then there exist n linearly independent eigenvectors. Thus we get an eigenbasis of $\mathbb{F}^n$, so we obtain the result on distinct eigenvalues mentioned in the previous section.

The above result also applies to a linear transformation $T : V \to V$ provided V is finite dimensional. In particular, T is semi-simple if and only if $\dim V = \sum_{i=1}^{k} \dim E_{\lambda_i}$, where $\lambda_1, \ \ldots \ , \lambda_k$ are the distinct eigenvalues of T in $\mathbb{F}$ and E_{λ_i} is the eigenspace of T corresponding to λ_i.

Thus, repeated eigenvalues do not preclude diagonalizability. Here is an example that is rather fun to analyze.

Example 9.13. 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 $E_0 = \mathcal{N}(B)$ has dimension three. Three independent eigenvectors for 0 are $\mathbf{f}_1 = (-1, 1, 0, 0)^T$, $\mathbf{f}_2 = (-1, 0, 1, 0)^T$, $\mathbf{f}_3 = (-1, 0, 0, 1)^T$. Another eigenvalue can be found by inspection, if we notice a special property of B. Every row of B adds up to 4. Thus, $\mathbf{f}_4 = (1, 1, 1, 1)^T$ is another eigenvector for $\lambda = 4$. By Proposition 9.11, we now have four linearly independent eigenvectors, hence an eigenbasis. Therefore B is diagonalizable; in fact, B is similar to $D = \text{diag}(0, 0, 0, 4)$. Note that one can also the characteristic polynomial of B by inspection. In fact, the all principal minors of B are zero save the 1×1 principal minors. Hence, $p_B(\lambda) = \lambda^4 - 4\lambda^3$.

In the above example, the fourth eigenvalue of B was found by noticing a special property of B. A better way to find λ_4 would have been to use the fact that the trace of a square matrix is the sum of its eigenvalues. Hence if all but one eigenvalue is known, the final eigenvalue can be found immediately. In our case, three eigenvalues are 0, hence the fourth must be the trace, which is 4.

9.4.2 Do Non-diagonalizable Matrices Exist?

We now have a criterion to answer the question of whether there exist non-diagonalizable matrices. Of course, we have seen that there exist real matrices which aren't diagonalizable over $\mathbb{R}$ since some of their eigenvalues are complex. But these matrices might all be diagonalizable over $\mathbb{C}$. However, it turns out that there are matrices for which (9.13) isn't satisfied. Such matrices can't be diagonalizable. In fact, examples are quite easy to find.

Example 9.14. Consider the real 2×2 matrix

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

Clearly $p_A(\lambda) = \lambda^2$, so 0 is the only eigenvalue of A. Clearly, $E_0 = \mathcal{N}(A) = \mathbb{R}\mathbf{e}_1$, so E_0 has dimension one. Therefore, A cannot have two linearly independent eigenvectors hence cannot be diagonalized over $\mathbb{R}$. For the same reason, it can't be diagonalized over $\mathbb{C}$ either.

Another way of seeing A isn't diagonalizable is to suppose it is. Then $A = MDM^{-1}$ for some invertible M. Since A's eigenvalues are both 0 and two similar matrices have he same eigenvalues, $D = \text{diag}(0, 0)$. This leads to the equation $A = MDM^{-1} = O$, where O is the zero matrix. But $A \neq O$.

The above example is an illustration of a nontrivial fact about eigenvalues. Namely, the dimension of the eigenspace of an eigenvalue (i.e. its geometric multiplicity) is at most the algebraic multiplicity of the eigenvalue. This is proved in the next Section.

9.4.3 Tridiagonalization of Complex Matrices

We will now show that every complex matrix is similar to an upper triangular matrix. This is known as *tridiagonalization*. It is the best general result available general result about diagonalization which we can prove now and a precursor to the Jordan Normal Form which will be explained in Chapter 13. The proof of tridiagonalization is a good illustration of the usefulness of quotient spaces.

Proposition 9.13. *Let $V = \mathbb{C}^n$ and let $T : V \to V$ be a linear transformation, say $T = T_A$ where $A \in \mathbb{C}^{n\times n}$. Then there exists a basis $\mathcal{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ of V such that for each index j,*

$$T(\mathbf{v}_j) = \sum_{i=1}^{j} b_{ij}\mathbf{v}_i. \tag{9.14}$$

Consequently, the matrix $\mathcal{M}_{\mathcal{B}}^{\mathcal{B}}(T)$ of T with respect to $\mathcal{B}$ is upper triangular. In particular, every $A \in \mathbb{C}^{n\times n}$ is similar to an upper triangular matrix.

Proof. Let us induct on n. The case $n = 1$ is trivial, so suppose the Proposition holds for $n-1$. Since $\mathbb{C}$ is algebraically closed, T has an eigenpair, say $(\lambda, \mathbf{w})$. Let $W = \mathbb{C}\mathbf{w}$ be the line in V spanned by $\mathbf{w}$. Now consider the quotient space V/W. We know from Theorem 5.27 that $\dim(V/W) = n-1$, and, as shown the proof of this result, if $\mathbf{v}_1, \ldots, \mathbf{v}_n$ is a basis of V with $\mathbf{v}_n = \mathbf{w}$, then the cosets $\mathbf{v}_1 + W, \ldots, \mathbf{v}_{n-1} + W$ form a basis of V/W. Now define a linear transformation $T_1 : V/W \to V/W$ by putting $T_1(\mathbf{v}_i+W) = T(\mathbf{v}_i)+W$ for $i = 1, \ldots, n-1$. These values uniquely define T_1 and since $T(W) \subset W$, it is easy to see from the definition of the vector space structure of V/W and the fact that T is linear that $T_1(\mathbf{v} + W) = T(\mathbf{v}) + W$ for all $\mathbf{v} \in V$. By the inductive assumption, there exists a basis $\overline{\mathbf{w}}_1, \ldots, \overline{\mathbf{w}}_{n-1}$ of V/W on

which (9.14) holds, say

$$T_1(\overline{\mathbf{w}}_j) = \sum_{i=1}^{j} c_{ij}\overline{\mathbf{w}}_i.$$

Letting $\mathbf{w}_i \in V$ be chosen so that $\overline{\mathbf{w}}_i = \mathbf{w}_i + W$, we see from this identity that $T(\mathbf{w}_1) - c_{11}\mathbf{w}_1 \in W$, $T(\mathbf{w}_2) - c_{12}\mathbf{w}_1 - c_{22}\mathbf{w}_2 \in W$ and so on. Hence, if we put $\mathbf{v}_1 = \mathbf{w}, \mathbf{v}_2 = \mathbf{w}_1, \ldots, \mathbf{v}_n = \mathbf{w}_{n-1}$, then (9.14) is clearly satisfied. It remains to show that $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are independent, hence a basis of V. This is left to the reader. □

The existence of a flag basis in fact holds for any linear transformation $T : V \to V$, provided V is defined over an arbitrary algebraically closed field, since the only property of $\mathbb{C}$ we used in the proof is that the characteristic polynomial of T has $\dim V$ complex roots. If $V = \mathbb{R}^n$, then the same proof works for a linear transformation T all of whose eigenvalues are real. In particular, every $A \in \mathbb{R}^{n\times n}$ such that all the eigenvalues of A are real is similar over the reals to an upper triangular matrix. That is, there exists an invertible P and a diagonal D, both in $\mathbb{R}^{n\times n}$, such that $A = PDP^{-1}$.

Of course, the eigenvalues of T are the coefficients b_{ii} on the diagonal in (9.14). The above argument shows that the eigenvalues of T_1 are exactly the eigenvalues of T except for λ (though λ may still be an eigenvalue of T_1). A basis such that (9.14) holds is called a *flag basis of V for T*. Scur's Theorem, which is proved in Chapter 11 gives a proof that every complex matrix is tridiagonalizable without using the notion of a quotient space. Schur's Theorem also proves more: namely that the flag basis can be chosen to be orthonormal.

By a similar, but more complicated argument, it can also be shown that two commuting complex matrices can be simultaneously diagonalized. Moreover, if both matrices are also diagonalizable, then they can be simultaneously diagonalized. Of course, two simultaneously diagonalizable matrices $A = PDP^{-1}$ and $A' = PD'P^{-1}$ matrices commute since $DD' = D'D$.

We will now prove

Proposition 9.14. *For any complex matrix, the geometric multiplicity of an eigenvalue is at most the algebraic multiplicity of the eigenvalue.*

Proof. Let $A \in \mathbb{C}^{n\times n}$ be arbitrary, and let λ be an eigenvalue of A. Since A can be tridiagonalized and since the eigenvalues of similar matrices have the same algebraic and geometric multiplicities, it suffices to assume A is upper triangular. But in this case, the result follows by inspection. For example, if λ is the only eigenvalue of A, then the rank of $A - \lambda I_n$ can take any value

between 0 and $n-1$, so the geometric multiplicity takes any value between n and 1 since the geometric multiplicity is $n - \text{rank}(A - \lambda I_n)$. □

Here is an example.

Example 9.15. Consider the matrices

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

In A, the geometric multiplicity of the eigenvalue 2 is one, while the algebraic multiplicity is one. In B, the geometric multiplicity of 2 is two. The matrix B is diagonalizable, while A is not.

9.4.4 The Cayley-Hamilton Theorem

We conclude this topic with a famous result called the Cayley-Hamilton Theorem, which gives an important relationship between a matrix and its characteristic polynomial.

Theorem 9.15. *Let A be an $n \times n$ matrix over an arbitrary field $\mathbb{F}$. Then $p_A(A) = O$. That is, A satisfies its own characteristic polynomial.*

Note that by $p_A(A) = O$ we mean

$$(-1)^n A^n + (-1)^{n-1}\text{Tr}(A)A^{n-1} + \cdots + \det(A)I_n = O.$$

Here we have put $A^0 = I_n$. For example, the characteristic polynomial of the matrix $J = \left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$ is $\lambda^2 + 1$. By Cayley-Hamilton, $J^2 + I_2 = O$, which is easy to check directly.

We will give a complete proof of the Cayley-Hamilton Theorem in Chapter 13. Let us outline a proof for the case $\mathbb{F} = \mathbb{C}$. The first thing to notice is that if A is diagonal, say $A = \text{diag}(d_1, \ldots, d_n)$, then $p_A(A) = \text{diag}(p(d_1), p(d_2), \ldots, p(d_n))$. But the diagonal entries of a diagonal matrix are the eigenvalues, so in this case, the conclusion $p_A(A) = O$ is clear. Now if A is diagonalizable, say $A = MDM^{-1}$, then

$$p_A(A) = p_A(MDM^{-1}) = Mp_A(D)M^{-1} = MOM^{-1} = O.$$

Thus we are done if A is diagonalizable. To finish the proof, one can use limits. If A is any matrix over $\mathbb{R}$ or $\mathbb{C}$, one can show there is a sequence of diagonalizable matrices A_k such that

$$\lim_{k \to \infty} A_k = A.$$

Letting p_k denote the characteristic polynomial of A_k, we then have

$$\lim_{k\to\infty} p_k(A_k) = p(A) = O$$

since each $p_k(A_k) = O$.

Exercises

Exercise 9.39. Show that if Q is orthogonal, then its only real eigenvalues are ± 1. Conclude that if Q is diagonalizable (over $\mathbb{R}$), say $Q = PDP^{-1}$, then the diagonal entries of D are ± 1.

Exercise 9.40. Determine which of the following matrices are diagonalizable over the reals:

(i) $A = \begin{pmatrix} 1 & 0 & -1 \\ -1 & 1 & 1 \\ 2 & -1 & -2 \end{pmatrix}$,

(ii) $B = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & -1 & 1 \end{pmatrix}$

(iii) $C = \begin{pmatrix} 2 & -1 & 1 \\ 1 & 0 & 1 \\ 1 & -1 & -2 \end{pmatrix}$,

(iv) $D = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}$.

(v) $E = \begin{pmatrix} 1 & 2 & 1 \\ 0 & 3 & 1 \\ 0 & 0 & 1 \end{pmatrix}$.

Exercise 9.41. Does

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

have distinct eigenvalues? Is it diagonalizable?

Exercise 9.42. Determine whether or not the following two matrices are similar:

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

Exercise 9.43. Show from first principals that if λ and μ are distinct eigenvalues of A, then $E_\lambda \cap E_\mu = \{\mathbf{0}\}$.

Exercise 9.44. Find an example of a non-diagonalizable 3×3 matrix A with real entries which is not upper or lower triangular such that every eigenvalue of A is 0.

Exercise 9.45. Determine whether or not the following two matrices are similar:

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

Exercise 9.46. Exactly two of the following three matrices are similar. Assuming this, determine which matrix is not similar to the others.:

$$A = \begin{pmatrix} 0 & 1 & 0 \\ 0 & -2 & -1 \\ 0 & 2 & 2 \end{pmatrix}, \quad B = \begin{pmatrix} 1 & 1 & 0 \\ 0 & -1 & 1 \\ -1 & -1 & 0 \end{pmatrix}, \quad C = \begin{pmatrix} 1 & 1 & 0 \\ 0 & -1 & 1 \\ 1 & 1 & 0 \end{pmatrix}.$$

Exercise 9.47. Assume B is a 3×3 real matrix.

(i) If the rank of B is 2, its trace is 4 and we know that B isn't diagonalizable. What are its eigenvalues?

(ii) Suppose the trace of A is 0, the determinant of A is 0 and the sum of each row of A is 3. What are the eigenvalues of A?

(iii) Is A of (ii) diagonalizable? Why?

Exercise 9.48. Recall that a square matrix A is called *nilpotent* if $A^k = O$ for some integer $k > 0$.

(i) Show that if A is nilpotent, then all eigenvalues of A are 0.

(ii) Prove that a nonzero nilpotent matrix cannot be diagonalizable.

(iii) Show, conversely, that if all the eigenvalues of A are 0, then A is nilpotent. (Hint: Consider the characteristic polynomial.)

Exercise 9.49. Show that if an $n \times n$ matrix A is nilpotent, then in fact $A^n = O$.

Exercise 9.50. Let A be a 3×3 matrix with eigenvalues 0,0,1. Show that $A^3 = A^2$.

Exercise 9.51. Let A be a 2×2 matrix so that $A^2 + 3A + 2I_2 = O$. Show that $-1, -2$ are eigenvalues of A.

Exercise 9.52. Suppose A is a 2×2 matrix so that $A^2 + A - 3I_2 = O$. Show that A is diagonalizable.

Exercise 9.53. Let U be an upper triangular matrix over $\mathbb{F}$ with distinct entries on its diagonal. Show that U is diagonalizable.

Exercise 9.54. Suppose that a 3×3 matrix A with real entries satisfies the equation $A^3 + A^2 - A + 2I_3 = O$.

(i) Find the eigen-values of A.

(ii) Is A diagonalizable? Explain.

Exercise 9.55. Is the matrix $A = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 2 \end{pmatrix}$ diagonalizable? What about $B = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{pmatrix}$?

Exercise 9.56. Let U be an arbitrary upper triangular matrix over $\mathbb{F}$ possibly having repeated diagonal entries. Try to find a condition to guarantee U is diagonalizable. (Hint: look at Exercise 9.55. Where do the repetitions on the diagonal occur?)

Exercise 9.57. Give the proof Proposition 9.12.

Exercise 9.58. Prove the Cayley-Hamilton Theorem for all diagonal matrices A by proving that if $p(x)$ is any polynomial, then

$$p(\mathrm{d}iag(\lambda_1, \lambda_2, \ldots, \lambda_n)) = \mathrm{d}iag(p(\lambda_1), p(\lambda_2), \ldots, p(\lambda_n)).$$

Deduce the Cayley-Hamilton Theorem for all diagonalizable matrices from the above identity.

Exercise 9.59. Prove the Cayley-Hamilton Theorem for upper triangular matrices, and then deduce the general case from Schur's Theorem (which you will have to look up).

Exercise 9.60. There is a deceptively attractive proof of Cayley-Hamilton that goes as follows. Consider the characteristic equation $\det(A - \lambda I_n) = 0$ of A. Since setting $\lambda = A$ in $p(\lambda) = \det(A - \lambda I_n)$ gives $\det(A - AI_n) = 0$, it follows that $p(A) = O$. Is this really a proof or is there a flaw in the argument?

Exercise 9.61. Use the Cayley-Hamilton Theorem to deduce that a 2×2 matrix A is nilpotent if and only if $\mathrm{Tr}(A) = \det(A) = 0$. Generalize this result to 3×3 matrices.

Exercise 9.62. * Fill in the details of the proof of the Cayley-Hamilton Theorem suggested above using sequences. That is, show that any real or complex matrix is the limit of a sequence of diagonalizable matrices.

9.5 The Exponential of a Matrix

We now return to the powers of a square matrix A, expanding on the remarks in §9.1. We will also define the exponential of a matrix, by extending the ordinary the *exponential* function e^x to be a function on $\mathbb{R}^{n\times n}$. We could also deal with complex matrices, but some care must be taken in defining the derivative of the exponential in the complex case. Thus we will omit it.

9.5.1 Powers of Matrices

Suppose $A \in \mathbb{R}^{n\times n}$ can be diagonalized, say $A = MDM^{-1}$. Then we saw that for any $k > 0$, $A^k = MD^kM^{-1}$. In particular, we can make a number of statements.

(1) If A is a diagonalizable real matrix with non negative eigenvalues, then A has a square root; in fact kth roots of A for all positive integers k are given by $A^{\frac{1}{k}} = MD^{\frac{1}{k}}M^{-1}$;

(2) If A is diagonalizable and none of the eigenvalues of A are 0, then the negative powers of A are found from the formula $A^{-k} = MD^{-k}M^{-1}$. Here, A^{-k} means $(A^{-1})^k$.

(3) If all the eigenvalues λ of A satisfy $0 \leq \lambda \leq 1$, then $\lim_{m\to\infty} A^m$ exists, and if no $\lambda = 1$, this limit is O.

We can in general obtain kth roots of a real matrix A as long as A is diagonalizable. One can't expect these matrices to be real however. For example, if A is diagonalizable but has a negative eigenvalue, then A cannot have any real square roots. However, these comments aren't valid for non-diagonalizable matrices.

9.5.2 The Exponential

Let $A \in \mathbb{R}^{n\times n}$. The *exponential* $\exp(A)$ of A is defined to be the matrix obtained by plugging A into the usual exponential series

$$e^x = 1 + x + \frac{1}{2!}x^2 + \frac{1}{3!}x^3 + \dots.$$

Thus the exponential $\exp(A)$ of A is given by the infinite series

$$\exp(A) = I_n + A + \frac{1}{2!}A^2 + \frac{1}{3!}A^3 + \cdots = I_n + \sum_{m=1}^{\infty} \frac{1}{m!}A^m. \tag{9.15}$$

It can be shown that for any A, every component in the exponential series for A converges, a fact we will simply assume. The matrix exponential behaves just like the ordinary exponential e^x in a number of ways, but the identity $e^{(x+y)} = e^x e^y$ no longer always holds. The reason for this is that although real number multiplication is commutative, matrix multiplication definitely isn't. In fact, we have

Proposition 9.16. *If $AB = BA$, then $\exp(A+B) = \exp(A)\exp(B)$.*

If A is diagonalizable, then the matter of finding $\exp(A)$ is easily settled. Suppose $D = \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$. Then $\exp(D) = \operatorname{diag}(e^{\lambda_1}, \ldots, e^{\lambda_n})$, and we have

Proposition 9.17. *Suppose A is a diagonalizable $n \times n$ matrix, say $A = PDP^{-1}$, where $D = \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$. Then*

$$\exp(A) = P\exp(D)P^{-1} = P\operatorname{diag}(e^{\lambda_1}, \ldots, e^{\lambda_n})P^{-1}.$$

In particular, if $\mathbf{v}_1, \ldots, \mathbf{v}_n$ is an eigenbasis of $\mathbb{R}^n$ for A, then it is also an eigenbasis for $\exp(A)$, and the eigenvector $\mathbf{v}_i$ has eigenvalue e^{λ_i}.

Hence, if $\mathbf{w} = \sum_{i=1}^n a_i \mathbf{v}_i$, then

$$\exp(A)\mathbf{w} = \sum_{i=1}^n a_i e^{\lambda_i} \mathbf{v}_i.$$

9.5.3 Uncoupling systems

One of the main applications of the exponential is to solve first order linear systems of differential equations by uncoupling. A typical application is the exponential growth problem solved in calculus. Assume $a(t)$ denotes the amount of a substance at time t that obeys the law $a'(t) = ka(t)$, where k is a constant. Then $a(t) = a_0 e^{kt}$ for all t, where a_0 is the initial amount of a.

The general form of this problem is the *first order linear system*

$$\frac{\mathrm{d}}{\mathrm{dt}}\mathbf{x}(t) = A\mathbf{x}(t),$$

where $A \in \mathbb{R}^{n\times n}$ and $\mathbf{x}(t) = (x_1(t), \ldots, x_n(t))^T$.

The geometric interpretation of this is that $\mathbf{x}(t)$ traces out a curve in $\mathbb{R}^n$, whose velocity vector at every time t is $A\mathbf{x}(t)$. It turns out that to solve this system, we consider the derivative with respect to t of $\exp(tA)$. First,

notice that by Proposition 9.16, $\exp((s+t)A) = \exp(sA)\exp(tA)$. Thus

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{dt}}\exp(tA) &= \lim_{s\to 0}\frac{1}{s}(\exp((t+s)A) - \exp(tA)) \\ &= \exp(tA)\lim_{s\to 0}\frac{1}{s}(\exp(sA) - I_n).\end{aligned}$$

It follows from the definition of $\exp(tA)$ that

$$\lim_{s\to 0}\frac{1}{s}(\exp(sA) - I_n) = A,$$

so we have the (not unexpected) formula

$$\frac{\mathrm{d}}{\mathrm{dt}}\exp(tA) = \exp(tA)A = A\exp(tA).$$

This implies that if we set $\mathbf{x}(t) = \exp(tA)\mathbf{v}$, then

$$\frac{\mathrm{d}}{\mathrm{dt}}\mathbf{x}(t) = \frac{\mathrm{d}}{\mathrm{dt}}\exp(tA)\mathbf{v} = A\exp(tA)\mathbf{v} = A\mathbf{x}(t).$$

Hence $\mathbf{x}(t)$ is a solution curve or trajectory of our given first order system. Since $\mathbf{x}(0) = \mathbf{v}$ is the initial value of $\mathbf{x}(t)$, it follows that the initial value of the trajectory $\mathbf{x}(t)$ can be arbutrarily prescribed, so a solution curve $\mathbf{x}(t)$ can be found passing through any given initial point $\mathbf{x}(0)$.

Example 9.16. Consider the system

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}.$$

The matrix $A = \left(\begin{smallmatrix} 2 & 1 \\ 1 & 2 \end{smallmatrix}\right)$ can be written $A = M\mathrm{diag}(3,1)M^{-1}$, so

$$\exp(tA) = M\exp\begin{pmatrix} 3t & 0 \\ 0 & t \end{pmatrix}M^{-1} = M\begin{pmatrix} e^{3t} & 0 \\ 0 & e^{t} \end{pmatrix}M^{-1}.$$

Therefore, using the value of M already calculated, our solution to the system is

$$\mathbf{x}(t) = \begin{pmatrix} x(t) \\ y(t) \end{pmatrix} = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}\begin{pmatrix} e^{3t} & 0 \\ 0 & e^{t} \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}\begin{pmatrix} x(0) \\ y(0) \end{pmatrix}.$$

The final expression for $\mathbf{x}(t)$ is therefore

$$\begin{pmatrix} x(t) \\ y(t) \end{pmatrix} = \frac{1}{2}\begin{pmatrix} e^{3t}+e^{t} & e^{3t}-e^{t} \\ e^{3t}+e^{t} & e^{3t}-e^{t} \end{pmatrix}\begin{pmatrix} x(0) \\ y(0) \end{pmatrix}.$$

If the matrix A is nilpotent, then the system $\mathbf{x}'(t) = A\mathbf{x}(t)$ is still solved by exponentiating tA. The only difference is that A is no longer diagonalizable unless $A = O$. However, since A is nilpotent, $A^k = O$ for some $k > 0$, and so the infinite series is actually a finite sum. More generally, if A is an arbitrary real $n \times n$ matrix, it turns out that A is similar to a matrix of the form $D + N$, where D is diagonal, N is upper triangular and $DN = ND$. But then the exponential of $D+N$ is easily computed from Proposition 9.16. Namely,

$$\exp(D + N) = \exp(D)\exp(N).$$

This factorization $A = P(D+N)P^{-1}$ is known as the *Jordan Decomposition* of A. We will establish the existence of this decomposition in Chapter 13.

Example 9.17. Consider the system

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}.$$

Notice that the matrix of the system is already in the $D + N$ form above. Now

$$\exp(t \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}) = \exp(t \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix})\exp(t \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}).$$

Thus

$$\exp(t \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}) = \begin{pmatrix} e^t & 0 \\ 0 & e^t \end{pmatrix} \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} e^t & te^t \\ 0 & e^t \end{pmatrix}.$$

Finally,

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} e^t & te^t \\ 0 & e^t \end{pmatrix} \begin{pmatrix} x_0 \\ y_0 \end{pmatrix},$$

where the point $\begin{pmatrix} x_0 \\ y_0 \end{pmatrix}$ gives the initial position of the solution. Therefore

$$\begin{aligned} x(t) &= x_0 e^t + y_0 t e^t \\ y(t) &= y_0 e^t \end{aligned}$$

Exercises

Exercise 9.63. Find all possible square roots of the following matrices if any exist:

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

Exercise 9.64. Do the same as in Problem 9.63 for the 4×4 all 1's matrix.

Exercise 9.65. Calculate the exponentials of the matrices of Problem 9.63. What are the eigenvalues of their exponentials?

Exercise 9.66. Suppose $A \in \mathbb{C}^{n \times n}$ is diagonalizable. Show that

$$\det(\exp(A)) = e^{\mathrm{Tr}(A)},$$

and conclude that $\det(\exp(A)) > 0$ for any diagonalizable $A \in \mathbb{R}^{n \times n}$.

Exercise 9.67. Verify that $\exp(A + B) = \exp(A)\exp(B)$ if A and B are diagonal matrices. Use this formula to find the inverse of $\exp(A)$ for any square matrix A over $\mathbb{R}$.

Exercise 9.68. Recall that a square matrix A is called *nilpotent* if $A^k = O$ for some integer $k > 0$. Find a formula for the exponential of a nilpotent 3×3 matrix A such that $A^3 = O$.

Exercise 9.69. Solve the first order system $\mathbf{x}'(t) = A\mathbf{x}(t)$ with $\mathbf{x}(0) = \begin{pmatrix} a \\ b \end{pmatrix}$ for the following matrices A:

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

Exercise 9.70. Compute the nth power of all the matrices of Exercise 9.63 and also the 3×3 all 1's matrix.

Exercise 9.71. Show that if A is skew symmetric (i.e. $A^T = -A$), then $\exp(A)$ is orthogonal.

Exercise 9.72. Compute $\exp(A)$ in the case $A = \begin{pmatrix} 0 & \theta \\ -\theta & 0 \end{pmatrix}$,

Exercise 9.73. * Show that every $A \in \mathbb{C}^{n \times n}$ is similar to an upper triangular matrix.

9.6 Summary

Let A be an $n \times n$ matrix over $\mathbb{F}$. An eigenpair for A consists of a pair $(\lambda, \mathbf{v})$, where $\lambda \in \mathbb{F}$, $\mathbf{v} \in \mathbb{F}^n - \{\mathbf{0}\}$ and $A\mathbf{v} = \lambda\mathbf{v}$. The scalar λ is called an eigenvalue, and $\mathbf{v}$ is called an eigenvector associated to λ. Eigenvalues and eigenvectors are similarly defined for a linear transformation $T : V \to V$ in exactly the same manner. The eigenvalues of a square matrix A over $\mathbb{F}$ are the roots of the characteristic polynomial $p_A(\lambda) = \det(A - \lambda I_n)$.

Similar matrices have the same characteristic polynomial. Hence if V is a finite dimensional vector space , the characteristic polynomial of T may be defined as $p_A(\lambda)$ for any matrix representing T with respect to a basis of V. That is, the eigenvalues for T are the eigenvalues of any matrix representing T. Among the basic questions considered in this chapter are: given $A \in \mathbb{F}^{n\times n}$, how do we find its eigenpairs, how does this tell us what the eigenpairs of related matrices are, and, most importantly, does there exist a basis of $\mathbb{F}^n$ consisting of eigenvectors? Matrices which admit eigenbases are called diagonalizable, and linear transformations which admit eigenbases are said to be semi-simple.

The eigenvectors of A corresponding to a given eigenvalue λ, together with $\mathbf{0}$, form the subspace E_λ of $\mathbb{F}^n$ called the eigenspace of λ. The answer to the question of whether A is diagonalizable is yes if the sum of the dimensions of all the eigenspaces of A is n, in which case $\mathbb{F}^n$ is the direct sum of all the E_λ. The question of which matrices are diagonalizable deserves more attention, however, and it will be studied further in subsequent chapters. An interesting fact known as the Cayley-Hamilton Theorem states that every square matrrix satisfies its own characteristic polynomial.

The powers A^m of a diagonalizable matrix A are easy to compute. This allows one to determine the dynamical system associated to a diagonalizable matrix (or more generally, the dynamical system $A^m\mathbf{v}$ where $\mathbf{v}$ is an eigenvalue of A. A well known example, the Fibonacci sequence, is shown to come from the dynamical system associated to the matrix $F = \left(\begin{smallmatrix} 1 & 1 \\ 1 & 0 \end{smallmatrix}\right)$, hence large Fibonacci numbers are determined by the dominant eigenvalue ϕ of F, which is known as the Golden Mean.

The final topic covered in this chapter is the exponential of a square matrix over $\mathbb{R}$ or $\mathbb{C}$. The exponential has many useful applications such as to uncouple a linear systems of first order systems of differential equations. It also has many much deeper applications in matrix theory.

Chapter 10

The Inner Product Spaces $\mathbb{R}^n$ and $\mathbb{C}^n$

The goal of this chapter is to study some geometric problems arising in $\mathbb{R}^n$ and $\mathbb{C}^n$ which can be treated with the techniques of linear algebra. We will concentrate mainly on the case of $\mathbb{R}^n$, since the case of $\mathbb{C}^n$ is handled similarly. The first is to find the minimal distance from a point of $\mathbb{R}^n$ to a subspace. This is called the *least squares problem*. It leads naturally to the notions of projections and pseudo-inverses. We will also consider orthonormal bases of both $\mathbb{R}^n$ and $\mathbb{C}^n$, which will be also needed in our treatment of the Principal Axis Theorem, proved in Chapter 11. In particular, we will show that every subspace of either $\mathbb{R}^n$ or $\mathbb{C}^n$ has an orthonormal basis. We will also introduce the notion of an isometry, which will allow us to extend the above results to any finite dimensional inner product space. The last Section is devoted to studying the rotations of $\mathbb{R}^3$ and to giving examples of rotation groups.

10.1 The Orthogonal Projection on a Subspace

The purpose of this Section is to consider the problem of finding the distance from a point of $\mathbb{R}^n$ to a subspace. There is an analogous problem for $\mathbb{C}^n$, and we could treat both problems simultaneously. However, to simplify the exposition, we will concentrate on the real case.

Thus suppose W is a subspace of $\mathbb{R}^n$ and $\mathbf{x} \in \mathbb{R}^n$. The *distance from* $\mathbf{x}$ *to* W is defined as the minimum distance $d(\mathbf{x}, \mathbf{w}) = |\mathbf{x} - \mathbf{w}|$ as $\mathbf{w}$ varies over W. We will show that a solution to the problem of finding the minimal distance exists by actually constructing it. This general question turns out

to lead to many interesting ideas, including orthogonal complements and projections.

10.1.1 The Orthogonal Complement of a Subspace

Consider a subspace W of $\mathbb{R}^n$. Recall from Example 5.34 that the *orthogonal complement* of W is the subspace $W^\perp$ of $\mathbb{R}^n$ defined as the set of all vectors $\mathbf{v} \in \mathbb{R}^n$ which are orthogonal to every vector in W. That is,

$$W^\perp = \{\mathbf{v} \in \mathbb{R}^n \mid \mathbf{v} \cdot \mathbf{w} = 0 \ \forall \ \mathbf{w} \in W\} \tag{10.1}$$

Note that $W^\perp$ is defined whether or not W is a subspace. Furthermore, $W^\perp$ is always a subspace of $\mathbb{R}^n$ irregardless of whether or not W is.

It's instructive to visualize $W^\perp$ in matrix terms. Suppose W is the column space of an $n \times k$ real matrix A. Then $W^\perp = \mathcal{N}(A^T)$. In other words, ignoring the distinction between row and column vectors, the row space and null space of a matrix are each other's the orthogonal complements. Since A^T is $k \times n$, we know, from an often repeated fact, that $\text{rank}(A^T) + \dim \mathcal{N}(A^T) = n$. But by Section 5.3.2, the rank of A^T is the dimension of the row space of A^T, which is clearly the same as the dimension of the column space of A, namely $\dim W$. Therefore,

$$\dim W + \dim W^\perp = n. \tag{10.2}$$

This assertion is part of the following basic result.

Proposition 10.1. *Let W be a subspace of $\mathbb{R}^n$ and $W^\perp$ its orthogonal complement. Then*

(*i*) $W \cap W^\perp = \{\mathbf{0}\}$.

(*ii*) *Every $\mathbf{x} \in \mathbb{R}^n$ can be orthogonally decomposed in exactly one way as $\mathbf{x} = \mathbf{w} + \mathbf{y}$, where $\mathbf{w} \in W$ and $\mathbf{y} \in W^\perp$. In particular, we have the direct sum decomposition*

$$W \oplus W^\perp = \mathbb{R}^n.$$

(*iii*) $(W^\perp)^\perp = W$.

Proof. Part (*i*) follows immediately from the fact that if $\mathbf{v} \in W \cap W^\perp$, then $\mathbf{v} \cdot \mathbf{v} = 0$, so $\mathbf{v} = \mathbf{0}$. The proof of (*ii*) is harder. If $W = \mathbb{R}^n$, there is nothing

to prove. Otherwise, both W and $W^\perp$ have positive dimensions. Now the Hausdorff Intersection Formula says that

$$\dim(W + W^\perp) = \dim W + \dim W^\perp - \dim(W \cap W^\perp).$$

By (i) and (10.2), $\dim(W + W^\perp) = n$, so $W + W^\perp = \mathbb{R}^n$. Hence the expression $\mathbf{x} = \mathbf{w} + \mathbf{y}$ in (ii) exists. It's unique since if $\mathbf{x} = \mathbf{w}' + \mathbf{y}'$, then $(\mathbf{x} - \mathbf{x}') = -(\mathbf{y} - \mathbf{y}')$, so $\mathbf{x} - \mathbf{x}' = \mathbf{y}' - \mathbf{y} = \mathbf{0}$ by (i). We leave (iii) as an exercise. □

Definition 10.1. Let $\mathbf{x} = \mathbf{w}+\mathbf{y}$ be the decomposition of $\mathbf{x}$ as in Proposition 10.1 (ii). Then $\mathbf{w}$ is called the *component* of $\mathbf{x}$ in W.

10.1.2 The Subspace Distance Problem

We now consider the least squares problem for a subspace W of $\mathbb{R}^n$. Let $\mathbf{x} \in \mathbb{R}^n$ be arbitrary. We want to find the minimum distance $d(\mathbf{x}, \mathbf{y})$, as $\mathbf{y}$ varies over all of W.

The minimum is called the *distance from* $\mathbf{x}$ *to* W. Observe that minimizing the distance $d(\mathbf{x}, \mathbf{y})$ is equivalent to minimizing the sum of squares $|\mathbf{x} - \mathbf{y}|^2$. This is a convenient simplification so that we can use the Pythagorean property of the inner product. By part (ii) of Proposition 10.1, we may decompose $\mathbf{x}$ uniquely as $\mathbf{x} = \mathbf{w} + \mathbf{v}$ with $\mathbf{w} \in W$ and $\mathbf{v} \in W^\perp$. Then for any $\mathbf{y} \in W$, $(\mathbf{x} - \mathbf{w}) \cdot (\mathbf{w} - \mathbf{y}) = 0$, so, by Pythagoras,

$$\begin{aligned} |\mathbf{x} - \mathbf{y}|^2 &= |(\mathbf{x} - \mathbf{w}) + (\mathbf{w} - \mathbf{y})|^2 \\ &= |\mathbf{x} - \mathbf{w}|^2 + |\mathbf{w} - \mathbf{y}|^2 \\ &\geq |\mathbf{x} - \mathbf{w}|^2. \end{aligned}$$

Thus the minimum distance is realized by the component $\mathbf{w}$ of $\mathbf{x}$ in W. To summarize this, we state

Proposition 10.2. *The distance from* $\mathbf{x} \in \mathbb{R}^n$ *to the subspace* W *is* $|\mathbf{x}-\mathbf{w}|$, *where* $\mathbf{w}$ *is the component of* $\mathbf{x}$ *in* W. *Put another way, the distance from* $\mathbf{x}$ *to* W *is* $|\mathbf{y}|$, *where* $\mathbf{y}$ *is the component of* $\mathbf{x}$ *in* $W^\perp$.

10.1.3 The Projection on a Subspace

The above solution of the least squares problem now requires us to find a way to compute the expression for the component $\mathbf{w}$ of $\mathbf{x}$ in W. To do so, we begin with the following definition.

Definition 10.2. The *orthogonal projection* of $\mathbb{R}^n$ onto a subspace W is the transformation $P_W : \mathbb{R}^n \to W$ defined by $P_W(\mathbf{x}) = \mathbf{w}$, where $\mathbf{w}$ is the component of $\mathbf{x}$ in W.

Thus $P_W(\mathbf{x})$ is the unique solution of the least squares problem for the subspace W. One usually calls P_W simply the *projection* of $\mathbb{R}^n$ onto W. We now derive a method for finding P_W.

Let $\mathbf{w}_1, \ldots, \mathbf{w}_m$ be a basis of W, and put $A = (\mathbf{w}_1 \ \cdots \ \mathbf{w}_m)$. Then W is the column space $\mathrm{col}(A)$ of A. Note that $A \in \mathbb{R}^{n\times m}$. Putting $\mathbf{w} = P_W(\mathbf{x})$, notice that $\mathbf{w}$ satisfies the *normal equations*:

$$\mathbf{w} = A\mathbf{v} \quad \text{and} \quad A^T(\mathbf{x} - \mathbf{w}) = 0, \tag{10.3}$$

for some $\mathbf{v} \in \mathbb{R}^m$. The first equation simply says $\mathbf{w}$ is a linear combination of $\mathbf{w}_1, \ldots, \mathbf{w}_m$, while the second says that $\mathbf{x}-\mathbf{w} \in W^\perp$, since the rows of A^T span W. By Proposition 10.1 (ii), both equations have solutions. The only question is whether the solution can be expressed elegantly and usefully. Multiplying the equation $\mathbf{w} = A\mathbf{v}$ by A^T leads to the single equation

$$A^T\mathbf{x} = A^T\mathbf{w} = A^TA\mathbf{v}.$$

We now need the following fact.

Proposition 10.3. *Let $A \in \mathbb{R}^{n\times m}$. Then A and A^TA have the same rank. In particular, if $\mathrm{rank}(A) = m$, then A^TA is invertible.*

The proof is outlined in Exercise 10.14. Hence we can uniquely solve for $\mathbf{v}$. Indeed,

$$\mathbf{v} = (A^TA)^{-1}A^T\mathbf{x}.$$

Multiplying by A, we get an expression for $\mathbf{w}$ in the following elegant form:

$$\mathbf{w} = A\mathbf{v} = A(A^TA)^{-1}A^T\mathbf{x}. \tag{10.4}$$

Therefore,

$$P_W(\mathbf{x}) = A(A^TA)^{-1}A^T\mathbf{x}. \tag{10.5}$$

This equation says that P_W is the linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^n$ with matrix $A(A^TA)^{-1}A^T \in \mathbb{R}^{n\times n}$. Whether we write $P_W : \mathbb{R}^n \to \mathbb{R}^n$ or $P_W : \mathbb{R}^n \to W$ is immaterial. We will frequently write $P_W = A(A^TA)^{-1}A^T$ below.

Let's next consider some examples.

Example 10.1. Recall that the case of a projection from $\mathbb{R}^2$ onto a line was studied in Example 7.9. We now generalize it to the case where W is a line in $\mathbb{R}^n$, say $W = \mathbb{R}\mathbf{w}$. Then the matrix of P_W is given by $\mathbf{w}\left(\mathbf{w}^T\mathbf{w}\right)^{-1}\mathbf{w}^T$, so

$$P_W(\mathbf{x}) = \frac{\mathbf{w}^T\mathbf{x}}{\mathbf{w}^T\mathbf{w}}\mathbf{w}.$$

This verifies the formula of Example 7.9.

Example 10.2. Let

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

Then A has rank 2, and

$$A^TA = \begin{pmatrix} 7 & 2 \\ 2 & 3 \end{pmatrix}.$$

Hence, by a direct computation,

$$P_W = A(A^TA)^{-1}A^T = \frac{1}{17}\begin{pmatrix} 14 & 1 & 5 & 4 \\ 1 & 11 & 4 & 7 \\ 5 & 4 & 3 & 1 \\ 4 & 7 & 1 & 6 \end{pmatrix}.$$

Example 10.3. Suppose $W = \mathbb{R}^n$. Then clearly, $P_W = I_n$. For, in this case, A has rank n, so A and A^T are both invertible. Thus

$$P_W = A(A^TA)^{-1}A^T = A(A^{-1}(A^T)^{-1})A^T = I_n$$

as claimed.

The following Proposition summarizes what we can now say about projections. Assume as above that $A \in \mathbb{R}^{n\times m}$ has rank m and $W = \mathrm{col}(A)$. The reader should compare this result with Example 7.9.

Proposition 10.4. *The projection $P_W : \mathbb{R}^n \to W$ is a linear transformation, and*

$$\mathbf{x} = P_W(\mathbf{x}) + (\mathbf{x} - P_W(\mathbf{x}))$$

is the orthogonal sum decomposition of $\mathbf{x}$ into the sum of a component in W and a component in $W^\perp$. In addition, P_W has the following properties:

(*i*) *if $\mathbf{w} \in W$, then $P_W(\mathbf{w}) = \mathbf{w}$;*

(*ii*) $P_W P_W = P_W$; *and finally,*

(*iii*) *the matrix* $A(A^TA)^{-1}A^T$ *of* P_W *(with respect to the standard basis) is symmetric.*

Proof. The fact that $P_W(\mathbf{x}) = A(A^TA)^{-1}A^T\mathbf{x}$ immediately implies P_W is linear. That the decomposition $\mathbf{x} = P_W(\mathbf{x}) + (\mathbf{x} - P_W(\mathbf{x}))$ is the orthogonal sum decomposition of $\mathbf{x}$ follows from the normal equations. It remains to show $(i) - (iii)$. If $\mathbf{w} \in W$, then $\mathbf{w} = \mathbf{w} + \mathbf{0}$ is the orthogonal sum decomposition of $\mathbf{w}$ with one component in W and the other in $W^\perp$, so it follows from the uniqueness of such a decomposition that $P_W(\mathbf{w}) = \mathbf{w}$. One can also see $A(A^TA)^{-1}A^T\mathbf{w} = \mathbf{w}$ by applying the fact that $\mathbf{w} = A\mathbf{v}$. Part (ii) follows immediately from (i) by setting $\mathbf{w} = P_W(\mathbf{x})$. It can also be shown by a direct computation. We leave part (iii) as a simple exercise. □

In the next section, we express P_W in another way using an orthonormal basis. This expression is theoretically important, because it is tied up with Fourier series.

10.1.4 Inconsistent Systems and the Pseudoinverse

Suppose $A \in \mathbb{R}^{n\times m}$ has independent columns (equivalently rank m), and consider the linear system $A\mathbf{x} = \mathbf{b}$. By the independence of the columns, this system either has a unique solution or no solution at all (when $\mathbf{b}$ isn't in $\mathrm{col}(A)$). In the inconsistent case, is there a consistent system which we can replace it with? The intuitive choice is to replace $\mathbf{b}$ by the element of $\mathrm{col}(A)$ closest to $\mathbf{b}$, which, by the above analysis, is simply $P_W(\mathbf{b})$.

Let us make a remark about linear transformation $T_A : \mathbb{R}^m \to \mathbb{R}^n$ defined by A. We know T_A is injective since $\mathcal{N}(A) = \{\mathbf{0}\}$. Hence, by the result of Exercise 10.18, there exists a left inverse $S : \mathbb{R}^n \to \mathbb{R}^m$ which is also linear. If $m \neq n$, then S isn't unique, but, nevertheless, we can ask if there is a natural choice for S. The answer is supplied by the pseudo-inverse, which we will now define in terms of matrices. The *pseudo-inverse* A^+ of a matrix $A \in \mathbb{R}^{n\times m}$ of rank m is by definition

$$A^+ = (A^TA)^{-1}A^T. \tag{10.6}$$

Proposition 10.5. *The pseudo-inverse* A^+ *of an element* A *of* $\mathbb{R}^{n\times m}$ *of rank* m *satisfies* $A^+A = I_m$ *and* $AA^+ = P_W$, *where* $W = \mathrm{col}(A)$.

Proof. This follows immediately from the expression for P_W derived in the previous section. □

Now suppose $A\mathbf{x} = \mathbf{b}$ is consistent. Then $\mathbf{x} = A^+A\mathbf{x} = A^+\mathbf{b}$. (Note: if $m = n$, then $A^+ = A^{-1}$.) On the other hand, if $A\mathbf{x} = \mathbf{b}$ is inconsistent, then the vector $\mathbf{c}$ in $\mathbb{R}^n$ nearest $\mathbf{b}$ such that $A\mathbf{x} = \mathbf{c}$ is consistent is given by putting $\mathbf{c} = P_W(\mathbf{b})$, by Proposition 10.4. Putting $\mathbf{x} = A^+\mathbf{b}$ and multiplying by A gives $A\mathbf{x} = AA^+\mathbf{b} = P_W(\mathbf{b})$, so again we have that the solution is $A^+\mathbf{b}$.

In summary, we have

Proposition 10.6. *If $A \in \mathbb{R}^{n\times m}$ has rank m, then the possibly inconsistent system $A\mathbf{x} = \mathbf{b}$ has unique least squares solution $\mathbf{x} = A^+\mathbf{b}$. That is, $A^+\mathbf{b}$ is the unique solution of the consistent system $A\mathbf{x} = \mathbf{c}$, where $\mathbf{c}$ is the element of the column space of A closest to $\mathbf{b}$.*

Thus the pseudo-inverse finds the optimal solution of an inconsistent system. The system $A\mathbf{x} = \mathbf{b}$, whether consistent or inconsistent, is solved in the above sense by $\mathbf{x} = A^+\mathbf{b}$.

10.1.5 Applications of the Pseudoinverse

Let's consider a typical application. Suppose that one has m points (a_i, b_i) in $\mathbb{R}^2$, which represent the outcome of an experiment, and the problem is to find the line $y = cx + d$ fitting these points as well as possible. If the points (a_i, b_i) are on a line, then there exist $c, d \in \mathbb{R}$ such that $b_i = ca_i + d$ for all $i = 1, \ldots, m$. When this happens, we get the equation

$$A\mathbf{x} = \begin{pmatrix} a_1 & 1 \\ a_2 & 1 \\ \vdots & \vdots \\ a_m & 1 \end{pmatrix} \begin{pmatrix} c \\ d \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix},$$

with the unknowns c and d in the components of $\mathbf{x}$. The natural move is to find A^+, and replace we replace $(b_1 \ldots, b_m)^T$ by $AA^+(b_1 \ldots, b_m)^T = (c_1, \ldots, c_m)^T$. By the theory of the pseudo-inverse, all (a_i, c_i) lie on a line $y = cx + d$ and the sum

$$\sum_{i=1}^{m} (b_i - c_i)^2$$

is minimized. Thus

$$\mathbf{x} = \begin{pmatrix} c \\ d \end{pmatrix} = A^+\mathbf{b} = (A^TA)^{-1}A^T\mathbf{b},$$

and carrying out the computation gives c and d in the form

$$\begin{pmatrix} c \\ d \end{pmatrix} = \begin{pmatrix} \sum a_i^2 & \sum a_i \\ \sum a_i & m \end{pmatrix}^{-1} \begin{pmatrix} \sum a_i b_i \\ \sum b_i \end{pmatrix}.$$

The c_i are then given by $c_i = ca_i + d$. Note that the 2×2 matrix in this solution is invertible just as long as we don't have all $a_i = 0$ or all $a_i = 1$.

The problem of fitting a set of points (a_i, b_i, c_i) to a plane is similar. The method can also be adapted to the problem of fitting a set of points in $\mathbb{R}^2$ to a nonlinear curve, such as an ellipse. This is apparently the origin of the least squares method. The inventor of this method was the renowned mathematician Gauss. In 1801, Gauss astonished the astronomical world by predicting (based on at most 9° of the observed orbit) where astronomers would be able to find an obscure astroid named Ceres in a full 11 months time. Nowadays, the activity of tracking astroids is known as astrometry and is carried out, to a large extent, by amateur astronomers. It is considered an extremely crucial, though unglamorous, activity.

Exercises

Exercise 10.1. Find:

(i) the component of $\mathbf{x} = (1, 1, 2)^T$ on the line $\mathbb{R}(2, -1, 0)^T$, and

(ii) the minimum distance from $\mathbf{x} = (1, 1, 2)^T$ to this line.

Exercise 10.2. Find:

(i) the component of $\mathbf{x} = (1, 1, 2, 1)^T$ in the subspace of $\mathbb{R}^4$ spanned by $(1, 1, -1, 1)^T$ and $(2, -1, 0, 1)^T$, and

(ii) the minimum distance from $\mathbf{x} = (1, 1, 2, 1)^T$ to this subspace.

Exercise 10.3. Show that the distance from a point $(x_0, y_0, z_0)^T \in \mathbb{R}^3$ to the plane $ax + by + cz = d$ is given by

$$\frac{|ax_0 + by_0 + cz_0 - d|}{(a^2 + b^2 + c^2)^{1/2}}.$$

Exercise 10.4. Generalize the formula in Exercise 10.3 to the case of a hyperplane $a_1x_1 + \cdots + a_nx_n = b$ in $\mathbb{R}^n$.

Exercise 10.5. Let A be the matrix

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

(i) Find the projection P_W of $\mathbb{R}^4$ onto the column space W of A.

(ii) Find the projection of $(2, 1, 1, 1)^T$ onto W.

(iii) Find the projection of $(2, 1, 1, 1)^T$ onto $W^\perp$.

Exercise 10.6. Let W just be a subset of $\mathbb{R}^n$. Show that $W^\perp$ is nevertheless a subspace of $\mathbb{R}^n$, and give a formula for its dimension.

Exercise 10.7. Show that every projection matrix is symmetric.

Exercise 10.8. What are the eigenvalues of a projection P_W?

Exercise 10.9. True or False: The matrix of a projection can always be diagonalized, i.e. there always exists an eigenbasis of $\mathbb{R}^n$ for every P_W.

Exercise 10.10. Assuming the pseudo-inverse A^+ of A is defined and $W = \mathrm{col}(A)$, show that $A^+P_W = A^+$. Use this to find the kernel of A^+.

Exercise 10.11. Show directly that every projection matrix is symmetric.

Exercise 10.12. Diagonalize (if possible) the matrix P_W in Example 10.6.

Exercise 10.13. Prove the Pythagorean relation used to prove Proposition 10.1. That is, show that if $\mathbf{p} \cdot \mathbf{q} = 0$, then

$$|\mathbf{p} + \mathbf{q}|^2 = |\mathbf{p} - \mathbf{q}|^2 = |\mathbf{p}|^2 + |\mathbf{q}|^2.$$

Conversely, if this identity holds for $\mathbf{p}$ and $\mathbf{q}$, show that $\mathbf{p}$ and $\mathbf{q}$ are orthogonal.

Exercise 10.14. The purpose of this exercise is to prove Proposition 10.3.

(*i*) Show that if $A \in \mathbb{R}^{m\times n}$, then $\mathcal{N}(A) = \mathcal{N}(A^TA)$. (Hint: note that $\mathbf{x}^TA^TA\mathbf{x} = |A\mathbf{x}|^2$ for any $\mathbf{x} \in \mathbb{R}^n$.)

(*ii*) Use part (*i*) to make the conclusion that the ranks of A^TA and A coincide.

Exercise 10.15. Find the pseudo-inverse of the matrix

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

Exercise 10.16. Show that the result of Exercise 10.14 does not always hold if $\mathbb{R}$ is replaced with $\mathbb{Z}_2$ (or another $\mathbb{Z}_p$) by giving an explicit example of a 3×2 matrix A over $\mathbb{Z}_2$ of rank 2 so that A^TA has rank 0 or 1.

Exercise 10.17. Show that if A has independent columns, then any left inverse of A has the form $A^+ + C$, where $CA = O$. (Note: $CA = O$ is equivalent to $\text{col}(A) \subset \mathcal{N}(C)$. If $CA = O$, what is $(A^+ + C)A$? And conversely?)

Exercise 10.18. Let $\mathbb{F}$ be any field. Show that every one to one linear map $T : \mathbb{F}^m \to \mathbb{F}^n$ has at least one left inverse $S : \mathbb{F}^n \to \mathbb{F}^m$ which is linear. (Hint: Let $W = \text{im}(T)$, and choose a basis of W. Then extend this basis to a basis of $\mathbb{F}^n$, and define S on this basis in an appropriate way.)

Exercise 10.19. This Exercise is a continuation of Exercise 10.18. Suppose $A \in \mathbb{F}^{n\times m}$ has rank m. Show there exists a nonzero matrix $B \in \mathbb{F}^{m\times n}$ such that $BA = O$ and conclude from this that the linear transformation S in Exercise 10.18 isn't unique.

Exercise 10.20. Suppose H is a hyperplane in $\mathbb{R}^n$ with normal line L. Interpret each of $P_H + P_L$, $P_H P_N$ and $P_N P_H$ by giving a formula for each.

Exercise 10.21. Find the line that best fits the points $(-1, 1)$, $(0, .5)$, $(1, 2)$, and $(1.5, 2.5)$.

Exercise 10.22. Suppose coordinates have been put on the universe so that the sun's position is $(0, 0, 0)$. Four observations of a planet orbiting the sun tell us that the planet passed through the points $(5, .1, 0)$, $(4.2, 2, 1.4)$, $(0, 4, 3)$, and $(-3.5, 2.8, 2)$. Find the plane (through the origin) that best fits the planet's orbit. (Note: Kepler's laws tell us that the orbits of the planets lie in a plane. So if the above points don't lie on a plane, it is because the measurements are inaccurate.)

10.2 Orthonormal Sets

The purpose of this section is to study orthonormal sets. In particular, we'll show that every subspace of $\mathbb{R}^n$ has an orthonormal basis and find an alternative way to express P_W.

10.2.1 Some General Properties

Suppose $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k$ are unit vectors in $\mathbb{R}^n$. Then we say $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k$ are *orthonormal* if and only if $\mathbf{u}_i \cdot \mathbf{u}_j = 0$ if $i \neq j$. The following Proposition is very basic.

Proposition 10.7. *Suppose $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k$ are orthonormal, and assume $\mathbf{x} = \sum_{i=1}^k a_i \mathbf{u}_i$. Then:*

(*i*) $a_i = \mathbf{x} \cdot \mathbf{u}_i$ *for each* i;

(*ii*) $|\mathbf{x}|^2 = \sum_{i=1}^k a_i^2$; *and*

(*iii*) *in particular,* $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k$ *are linearly independent.*

Proof. For (*i*), note that

$$\mathbf{x} \cdot \mathbf{u}_j = (\sum_{i=1}^k a_i \mathbf{u}_i) \cdot \mathbf{u}_j = \sum_{i=1}^k a_i (\mathbf{u}_i \cdot \mathbf{u}_j) = a_j,$$

by the definition of an orthonormal set. For (*ii*), we have

$$|\mathbf{x}|^2 = \mathbf{x} \cdot \mathbf{x} = \sum_{i,j=1}^k a_i a_j \mathbf{u}_i \cdot \mathbf{u}_j = \sum_{i=1}^k a_i^2.$$

Part (*iii*) is an immediate consequence of (*ii*). □

10.2.2 Orthonormal Bases

An *orthonormal basis* of a subspace W of $\mathbb{R}^n$ is an orthonormal subset of W which spans W. The following result is an immediate consequence of the previous Proposition.

Proposition 10.8. *If $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k$ is an orthonormal basis of a subspace W of $\mathbb{R}^n$, then every $\mathbf{x} \in W$ has the unique expression*

$$\mathbf{x} = \sum_{i=1}^k (\mathbf{x} \cdot \mathbf{u}_i) \mathbf{u}_i. \tag{10.7}$$

In particular, every vector in W is the sum of its projections on an arbitrary orthonormal basis of W.

The coefficients of $\mathbf{x}$ in (10.7) are called the *Fourier coefficients* of $\mathbf{x}$ with respect to the orthonormal basis.

Example 10.4. Here are some examples.

(a) The standard basis $\mathbf{e}_1, \dots, \mathbf{e}_n$ is an orthonormal basis of $\mathbb{R}^n$.

(b) $\mathbf{u}_1 = \frac{1}{\sqrt{3}}(1,1,1)^T$, $\mathbf{u}_2 = \frac{1}{\sqrt{6}}(1,-2,1)^T$, $\mathbf{u}_3 = \frac{1}{\sqrt{2}}(1,0,-1)^T$ are an orthonormal basis of $\mathbb{R}^3$. The first two basis vectors are an orthonormal basis of the plane $x - z = 0$.

(c) The columns of an orthogonal matrix $Q \in \mathbb{R}^{n\times n}$ are always an orthonormal basis of $\mathbb{R}^n$ and conversely. For example, the matrix

$$Q = \frac{1}{2}\begin{pmatrix} 1 & 1 & 1 & 1 \\ -1 & 1 & -1 & 1 \\ 1 & -1 & -1 & 1 \\ 1 & 1 & -1 & -1 \end{pmatrix}.$$

Using this fact, produce another orthonormal bases of $\mathbb{R}^4$.

Example 10.5. For example, using the orthonormal basis of $\mathbb{R}^4$ formed by the columns of the matrix Q in the previous example, we have

$$\begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} = \frac{1}{4}\begin{pmatrix} 1 \\ -1 \\ 1 \\ 1 \end{pmatrix} + \frac{1}{4}\begin{pmatrix} 1 \\ 1 \\ -1 \\ 1 \end{pmatrix} + \frac{1}{4}\begin{pmatrix} 1 \\ -1 \\ -1 \\ -1 \end{pmatrix} + \frac{1}{4}\begin{pmatrix} 1 \\ 1 \\ 1 \\ -1 \end{pmatrix}.$$

We now come to an important result.

Proposition 10.9. *A non trivial subspace of $\mathbb{R}^n$ admits an orthonormal basis.*

Proof. Suppose W is a subspace and $\dim W > 0$. Weuse by induction on $\dim W$. If $\dim W = 1$, the result is obvious (why?). Thus suppose $\dim W = m > 1$ and the result is true for any subspace of $\mathbb{R}^n$ of dimension at most $m-1$. Let $\mathbf{u}$ be a unit vector in W and let $H = (\mathbb{R}\mathbf{u})^\perp \cap W$. Then $H \subset W$, but $H \neq W$ since $\mathbf{u} \notin W$. Thus $\dim H < m$. Hence H admits an orthonormal basis, call it O. Now let $\mathbf{x}$ be an arbitrary element of W. Then $\mathbf{y} = \mathbf{x} - (\mathbf{x}\cdot\mathbf{u})\mathbf{u} \in H$, so $\mathbf{y}$ is a linear combination of the elements of O.

Thus $\mathbf{u}$ and O form a spanning set for W. But by definition, $\mathbf{u}$ is orthogonal to every element of O, so we get an orthonormal basis of W. □

Note that the previous argument shows that for any nonzero $\mathbf{u} \in W$, $\dim\left((\mathbb{R}\mathbf{u})^\perp \cap W\right) = \dim W - 1$.

10.2.3 Projections and Orthonormal Bases

We now show that orthonormal bases give an alternative way to approach projections. In fact, we have

Proposition 10.10. *Let $\mathbf{u}_1, \ldots, \mathbf{u}_m$ be an orthonormal basis for a subspace W of $\mathbb{R}^n$. Then the projection $P_W : \mathbb{R}^n \to W$ is given by*

$$P_W(\mathbf{x}) = \sum_{i=1}^{m} (\mathbf{x} \cdot \mathbf{u}_i)\mathbf{u}_i. \tag{10.8}$$

If $Q \in \mathbb{R}^{n\times m}$ is the matrix $Q = (\mathbf{u}_1 \ \cdots \ \mathbf{u}_m)$ with orthonormal columns, then

$$P_W(\mathbf{x}) = QQ^T\mathbf{x}. \tag{10.9}$$

Put another way,

$$P_W = \sum_{i=1}^{m} \mathbf{u}_i\mathbf{u}_i^T. \tag{10.10}$$

Proof. Let $\mathbf{w} = \sum_{i=1}^{m}(\mathbf{x} \cdot \mathbf{u}_i)\mathbf{u}_i$. By the definition of $P_W(\mathbf{x})$, it suffices to show that $\mathbf{x} - \mathbf{w} \in W^\perp$. For then, $\mathbf{x} = \mathbf{w} + (\mathbf{x} - \mathbf{w})$ is the decomposition of $\mathbf{x}$ into its component in W and its component in $W^\perp$, which we know characterizes $P_W(\mathbf{x})$. But clearly $(\mathbf{x} - \mathbf{w}) \cdot \mathbf{u}_i = 0$ for each i, so indeed, $\mathbf{x} - \mathbf{w} \in W^\perp$ as claimed.

It remains to (10.9). But since the columns of Q are a basis of W, the matrix of P_W is $Q(QQ^T)^{-1}Q^T$. But, as the columns of Q are orthonormal, we have $Q^TQ = I_m$. Therefore, we have the result. □

Equation (10.9) gives the simplest expression for P_W, but it requires an orthonormal basis. Formula (10.8) is sometimes called the *projection formula*.

Example 10.6. Let $W = \operatorname{span}\{(1,1,1,1)^T, (1,-1,-1,1)^T\}$. To find the matrix of P_W, observe that $\mathbf{u}_1 = 1/2(1,1,1,1)^T$ and $\mathbf{u}_2 = 1/2(1,-1,-1,1)^T$ form an orthonormal basis of W, so

$$P_W = QQ^T = 1/4 \begin{pmatrix} 1 & 1 \\ 1 & -1 \\ 1 & -1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & -1 & 1 \end{pmatrix}.$$

Carrying out the calculation, we find that

$$P_W = \frac{1}{2} \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}.$$

Here is an important remark that will be expanded in Section 10.4. The formula $P_W = A(A^TA)^{-1}A^T$ only applies when W is a subspace of $\mathbb{R}^n$. On the other hand, formula (10.10) for P_W works as long as W is any finite dimensional subspace of an arbitrary inner product space.

Exercises

Exercise 10.23. Expand $(1,0,0)^T$ using the orthonormal basis consisting of the columns of the matrix Q of Example 10.4(b). Do the same for $(1,0,0,0)$ using the rows of U.

Exercise 10.24. Find an orthonormal basis for the plane $x - 2y + 3z = 0$ in $\mathbb{R}^3$. Now extend this orthonormal set in $\mathbb{R}^3$ to an orthonormal basis of $\mathbb{R}^3$.

Exercise 10.25. Suppose $n = 3$ and W is a line or a plane. Is it True or False that there exists an orthonormal eigenbasis for P_W?

Exercise 10.26. Finish proving Proposition 10.7 by showing that any orthonormal set in $\mathbb{R}^n$ is linearly independent.

Exercise 10.27. Suppose $\mathbf{u}_1, \ \ldots \ , \mathbf{u}_{j-1}, \mathbf{u}_n$ is a basis of $\mathbb{R}^n$ such that every vector $\mathbf{w}$ in $\mathbb{R}^n$ satisfies (10.7). Prove that $\mathbf{u}_1, \ \ldots \ , \mathbf{u}_{j-1}, \mathbf{u}_n$ form an orthonormal basis.

Exercise 10.28. Let $Q = (\mathbf{u}_1 \mathbf{u}_2 \cdots \mathbf{u}_n)$ be orthogonal. If Q is not symmetric, show how to produce a new orthonormal basis of $\mathbb{R}^n$ from the columns of Q. What new orthonormal basis of $\mathbb{R}^4$ does one obtain from the orthonormal basis in Example 10.4, part (c)?

Exercise 10.29. Let $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n$ be an orthonormal basis of $\mathbb{R}^n$. Show that $I_n = \sum_{1=i}^{n} \ \mathbf{u}_i \mathbf{u}_i^T$. This verifies the identity in (10.7).

Exercise 10.30. Show that the reflection $H_{\mathbf{u}}$ through the hyperplane orthogonal to a unit vector $\mathbf{u} \in \mathbb{R}^n$ is given by the formula $H_{\mathbf{u}} = I_n - 2P_W$, where $W = \mathbb{R}\mathbf{u}$ (so $H = W^{\perp}$).

Exercise 10.31. Using the formula of Exercise 10.30 for the reflection $H_{\mathbf{u}}$ through the hyperplane orthogonal to a unit vector $\mathbf{u} \in \mathbb{R}^n$, find the matrix of $H_{\mathbf{u}}$ in the following cases:

(a) $\mathbf{u}$ is a unit normal to the hyperplane $x_1 + \sqrt{2}x_2 + x_3 = 0$ in $\mathbb{R}^3$.

(b) $\mathbf{u}$ is a unit normal to the hyperplane $x_1 + x_2 + x_3 + x_4 = 0$ in $\mathbb{R}^4$.

Exercise 10.32. Suppose $A = QDQ^{-1}$ with Q orthogonal and D diagonal. Show that A is always symmetric and that A is orthogonal if and only if all diagonal entries of D are either ± 1. Show that A is the matrix of a reflection $H_{\mathbf{u}}$ precisely when $D = \mathrm{diag}(-1, 1, \ldots, 1)$, that is exactly one diagonal entry of D is -1 and all others are $+1$.

Exercise 10.33. How would you define the reflection H through a subspace W of $\mathbb{R}^n$? What properties should the matrix of H have? For example, what should the eigenvalues of H be?

Exercise 10.34. Show that the matrix of the reflection $H_{\mathbf{u}}$ is always a symmetric orthogonal matrix such that $H_{\mathbf{u}}\mathbf{u} = -\mathbf{u}$ and $H_{\mathbf{u}}\mathbf{x} = \mathbf{x}$ if $\mathbf{x}\cdot\mathbf{u} = 0$.

Exercise 10.35. Let Q be the matrix of the reflection $H_{\mathbf{b}}$.

(a) What are the eigenvalues of Q?

(b) Use the result of (a) to show that $\det(Q) = -1$.

(c) Show that Q can be diagonalized by explicitly finding an eigenbasis of $\mathbb{R}^n$ for Q. Hence, every reflection is semi-simple.

Exercise 10.36. Check directly that if $R = I_n - P_W$, then $R^2 = R$. Verify also that the eigenvalues of R are 0 and 1 and that $E_0 = W$ and $E_1 = W^{\perp}$.

Exercise 10.37. Let $\mathbf{u}$ be a unit vector in $\mathbb{R}^n$. Show that the reflection $H_{\mathbf{u}}$ through the hyperplane H orthogonal to $\mathbf{u}$ admits an orthonormal eigenbasis.

10.3 Gram-Schmidt and the QR-Factorization

10.3.1 The Gram-Schmidt Method

We showed in the last Section that any subspace of $\mathbb{R}^n$ admits an orthonormal basis. Moreover, the proof essentially showed how such a basis can be obtained. Now, suppose W is a subspace of $\mathbb{R}^n$ with a given basis $\mathbf{w}_1, \ \dots \ , \mathbf{w}_m$. We will now give a constructive procedure, called the Gram-Schmidt method, which shows how to construct an orthonormal basis $\mathbf{u}_1, \ \dots \ , \mathbf{u}_m$ of W such that

$$\text{span}\{\mathbf{u}_1, \ \dots \ , \mathbf{u}_k\} = \text{span}\{\mathbf{w}_1, \ \dots \ , \mathbf{w}_k\} \tag{10.11}$$

for each index for $k = 1, \dots m$. In fact, as we will note later, the Gram-Schmidt method works in an arbitrary inner product space such as $C[a, b]$,

We will prove

Proposition 10.11. *For each $1 \leq j \leq m$, put $W_j = span\{\mathbf{w}_1, \dots, \mathbf{w}_j\}$. Let $\mathbf{v}_1 = \mathbf{w}_1$ and, for $j \geq 2$, put*

$$\mathbf{v}_j = \mathbf{w}_j - P_{W_{j-1}}(\mathbf{w}_j). \tag{10.12}$$

Then each $\mathbf{v}_j \in W_j$, $\mathbf{v}_j \neq \mathbf{0}$ and $\mathbf{v}_j \cdot \mathbf{v}_k = 0$ if $j \neq k$. Putting

$$\mathbf{u}_i = |\mathbf{v}_i|^{-1}\mathbf{v}_i,$$

we therefore obtain an orthonormal basis of W satisfying (10.11)*. Moreover, if $j \geq 2$,*

$$\mathbf{v}_j = \mathbf{w}_j - (\mathbf{w}_j \cdot \mathbf{u}_1)\mathbf{u}_1 - (\mathbf{w}_j \cdot \mathbf{u}_2)\mathbf{u}_2 - \cdots - (\mathbf{w}_j \cdot \mathbf{u}_{j-1})\mathbf{u}_{j-1}. \tag{10.13}$$

Proof. Clearly $\mathbf{v}_j \in W_j$ for all j. If $\mathbf{v}_j = \mathbf{0}$, then $\mathbf{w}_j \in W_{j-1}$, contradicting the linear independence of $\mathbf{w}_1, \dots, \mathbf{w}_m$. In addition, by Proposition 10.4, $\mathbf{v}_j$ is orthogonal to W_{j-1} if $j \geq 2$. Since $W_j \subset W_{j+1}$, it follows that $\mathbf{v}_j \cdot \mathbf{v}_k = 0$ if $j < k$. Hence $\mathbf{u}_1, \dots, \mathbf{u}_m$ is an orthonormal basis of W satisfying (10.11). Finally, equation (10.13) is just an application of the projection formula. □

10.3.2 The QR-Decomposition

The Gram-Schmidt method can be stated in a form which is convenient for computations. This form is also the basis of the the QR-algorithm. This expression is called the QR-*decomposition*.

The main point in the QR-decomposition becomes clear by considering an example. Suppose $A = (\mathbf{w}_1\ \mathbf{w}_2\ \mathbf{w}_3)$ is a real $n\times 3$ matrix with independent columns. Applying Gram-Schmidt to $\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3$ gives an orthonormal basis $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3$ of the column space $W = \mathrm{col}(A)$. Moreover, by construction,

$$\mathbf{w}_1 = \mathbf{u}_1,$$

$$\mathbf{w}_2 = (\mathbf{w}_2 \cdot \mathbf{u}_1)\mathbf{u}_1 + (\mathbf{w}_2 \cdot \mathbf{u}_2)\mathbf{u}_2,$$

and finally

$$\mathbf{w}_3 = (\mathbf{w}_3 \cdot \mathbf{u}_1)\mathbf{u}_1 + (\mathbf{w}_3 \cdot \mathbf{u}_2)\mathbf{u}_2 + (\mathbf{w}_3 \cdot \mathbf{u}_3)\mathbf{u}_3.$$

Thus,

$$(\mathbf{w}_1\ \mathbf{w}_2\ \mathbf{w}_3) = (\mathbf{u}_1\ \mathbf{u}_2\ \mathbf{u}_3) \begin{pmatrix} \mathbf{w}_1 \cdot \mathbf{u}_1 & \mathbf{w}_2 \cdot \mathbf{u}_1 & \mathbf{w}_3 \cdot \mathbf{u}_1 \\ 0 & \mathbf{w}_2 \cdot \mathbf{u}_2 & \mathbf{w}_3 \cdot \mathbf{u}_2 \\ 0 & 0 & \mathbf{w}_3 \cdot \mathbf{u}_3 \end{pmatrix}.$$

In general, if $A = (\mathbf{w}_1\ \cdots\ \mathbf{w}_m)$ is an $n \times m$ matrix over $\mathbb{R}$ with linearly independent columns, and $Q = (\mathbf{u}_1\ \cdots\ \mathbf{u}_m)$ is the $n \times m$ matrix with orthonormal columns produced by the Gram-Schmidt method, then the matrix R of Fourier coefficients of the $\mathbf{w}_i$ in terms of the orthonormal basis given by Gram-Schmidt is upper triangular. Moreover, R is invertible since its diagonal entries $\mathbf{w}_i \cdot \mathbf{u}_i \neq 0$. Summarizing, we have

Proposition 10.12. *Every $A \in \mathbb{R}^{n\times m}$ of rank m can be factored*

$$A = QR. \tag{10.14}$$

where $Q \in \mathbb{R}^{n\times m}$ has orthonormal columns and $R \in \mathbb{R}^{m\times m}$ is invertible and upper triangular.

Note that the QR-decomposition isn't unique, but it can always be arranged so that the diagonal entries of R are positive. If A is square, then both Q and R are square. In particular, Q is an orthogonal matrix. The factorization $A = QR$ is the first step in the QR-algorithm, which is an important method for approximating the eigenvalues of A.

Exercises

Exercise 10.38. Let $W \subset \mathbb{R}^4$ be the span of $(1,0,1,1)^T$, $(-1,1,0,0)^T$, and $(1,0,1,-1)^T$.

(i) Find an orthonormal basis of W.

(ii) Expand $(0,0,0,1)^T$ and $(1,0,0,0)^T$ in terms of this basis.

Exercise 10.39. Find an orthonormal basis of the plane W in $\mathbb{R}^4$ spanned by $(0,1,0,1)^T$ and $(1,-1,0,0)^T$. Do the same for $W^\perp$. Now find an orthonormal basis of $\mathbb{R}^4$ containing the orthonormal bases of W and $W^\perp$.

Exercise 10.40. Let

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

Find the QR-factorization of A such that R has a positive diagonal.

Exercise 10.41. Find a 4×4 orthogonal matrix Q whose first three columns are the columns of A in the previous problem.

Exercise 10.42. What would happen if the Gram-Schmidt method were applied to a set of vectors that were not lineary independent? In other words, why can't we produce an orthonormal basis from nothing?

Exercise 10.43. In the QR-decomposition, the diagonal entries of R are non zero. What would happen if there were a zero on R's diagonal?

Exercise 10.44. Show that for any subspace W of $\mathbb{R}^n$, P_W can be expressed as $P_W = QDQ^T$, where D is diagonal and Q is orthogonal. Find the diagonal entries of D, and describe Q.

Exercise 10.45. Let A have independent columns. Verify the formula $P = QQ^T$ using $A = QR$.

Exercise 10.46. Suppose A has independent columns and let $A = QR$ be the QR-factorization of A.

(i) Find the pseudo-inverse A^+ of A in terms of Q and R; and

(ii) Find a left inverse of A in terms of Q and R.

10.4 Further Remarks on Inner Product Spaces

Recall from Chapter 4 that an inner product space is a real vector space V with an inner product usually denoted by $(\mathbf{x}, \mathbf{y})$. The main example of an inner product space is Euclidean space $\mathbb{R}^n$, the inner product being the dot product $\mathbf{x} \cdot \mathbf{y}$. The purpose of this section is to consider some aspects of general inner product spaces, such as Hermitian inner product spaces and isometries. We will begin with some examples.

10.4.1 The Space of Linear Transformations $L(\mathbb{R}^n, \mathbb{R}^n)$

Recall that $L(\mathbb{R}^n, \mathbb{R}^n)$ denotes the space of linear transformations $T : \mathbb{R}^n \to \mathbb{R}^n$. This space has an inner product. This isn't surprising since we know that $L(\mathbb{R}^n, \mathbb{R}^n)$ is isomorphic with $\mathbb{R}^{n\times n}$ (a linear transformation T being sent to its matrix), which in turn is isomorphic with the inner product space with $\mathbb{R}^{n^2}$. There is an elegant way to define an inner product on $\mathbb{R}^{n\times n}$. If $A, B \in \mathbb{R}^{n\times n}$, put

$$(A, B) = \text{Trace}(A^T B).$$

It can easily be seen that if $A = (\mathbf{a}_1\ \mathbf{a}_2\ \cdots\ \mathbf{a}_n)$ and $B = (\mathbf{b}_1\ \mathbf{b}_2\ \cdots\ \mathbf{b}_n)$. then

$$(A, B) = \sum_{i=1}^{n} \mathbf{a}_i \cdot \mathbf{b}_i = \sum_{i=1}^{n} \mathbf{a}_i^T \mathbf{b}_i.$$

The axioms for an inner product follow readily. For example, since $\text{Trace}(A+B) = \text{Trace}(A) + \text{Trace}(B)$ and $\text{Trace}(rA) = r\text{Trace}(A)$, it follows that the inner product is linear in each component. Moreover, $(A, B) = (B, A)$ so it is symmetric. Finally, if $A \neq O$, then

$$(A, A) = \sum_{i=1}^{n} \mathbf{a}_i^T \mathbf{a}_i = \sum_{i=1}^{n} |\mathbf{a}_i|^2 > 0,$$

since some $\mathbf{a}_i \neq \mathbf{0}$.

Proposition 10.13. *The matrices E_{ij}, $1 \leq i, j \leq n$ form an orthonormal basis of $\mathbb{R}^{n\times n}$.*

Proof. First note $E_{ij}E_{rs} = E_{is}$ if $j = r$ and is $E_{ij}E_{rs} = O$ otherwise. Furthermore, $\text{Trace}(E_{is}) = 0$ or 1 depending on whether $i = s$ or not. Consequently,

$$(E_{ij}, E_{rs}) = \text{Trace}(E_{ij}^T E_{rs}) = \text{Trace}(E_{ji}E_{rs}) = \begin{cases} 1 \text{ if } \ r = i, j = s \\ 0 \text{ otherwise} \end{cases}.$$

This gives the result. □

10.4.2 The Space $C[a, b]$

We now consider the space $C[a, b]$ of continuous real valued functions on a closed interval $[a, b]$ in $\mathbb{R}$ which was already introduced in Example 4.16. This is an example of an infinite dimensional inner product space. Recall that the inner product on $C[a, b]$ is defined by

$$(f, g) = \int_a^b f(t)g(t)\mathrm{d}t. \tag{10.15}$$

The notion of orthogonality in a general inner product space V is the same as for $\mathbb{R}^n$. Two vectors $\mathbf{x}$ and $\mathbf{y}$ are orthogonal if and only if $(\mathbf{x}, \mathbf{y}) = 0$. Thus we can always try to extend the results we obtained for $\mathbb{R}^n$ on least squares and projections to an arbitrary V. There is no difficulty in doing this if V is a finite dimensional inner product space. The reason for this is explained in Section 10.4.4 below.

Problems arise in the infinite dimensional setting, however, due to the fact that there isn't a well defined notion of the projection of V to an arbitrary subspace W. The difficulty is in trying to imitate the result in the finite dimensional case which says that $V = W + W^{\perp}$. Moreover, in the infinite dimensional case, one has to reformulate the notion of an orthonormal basis, which we will not attempt to do here.

Nevertheless, if W is a finite dimensional subspace of an inner product space V, then W has an orthonormal basis of W. In fact, such a basis can be constructed from any starting basis of W by applying Gram-Schmidt. Thus the Projection Formula gives a well defined projection of V onto W.

In particular, a set of functions $S \subset C[a, b]$ is orthonormal if and only if for any $f, g \in S$,

$$(f, g) = \begin{cases} 0 & \text{if } f \neq g \\ 1 & \text{if } f = g \end{cases}$$

Suppose $\{f_1, f_2, \dots\}$ is an orthonormal set in $C[a, b]$. As above, the Fourier coefficients of a function $f \in C[a, b]$ are the numbers (f, f_i), $i = 1, 2, \dots$.

Example 10.7. Let us use Gram-Schmidt to find an orthonormal basis of the three dimensional subspace W of $C[-1, 1]$ spanned by 1, x and x^2. We leave it as an exercise to check that $(1, 1) = 2$, $(1, x) = 0$, $(x, x) = (1, x^2) = 2/3$, $(x, x^2) = 0$ and $(x^2, x^2) = 2/5$. Now $1/\sqrt{2}$ and $x/\sqrt{\frac{2}{3}}$ are orthonormal. The component of x^2 on the subspace they span is $1/3$. But,

$$||x^2 - 1/3||^2 = \int_{-1}^{1} (x^2 - 1/3)^2 \mathrm{d}x = \frac{8}{45}.$$

Thus the desired orthonormal basis is $f_1 = 1/\sqrt{2}$, $f_2 = x/\sqrt{\frac{2}{3}}$ and $f_3 = \alpha(x^2 - 1/3)$, where $\alpha = \sqrt{\frac{45}{8}}$.

Example 10.8. To continue the previous example, suppose we want to minimize the integral

$$\int_{-1}^{1} (\cos x - (a_1 + a_2 x + a_3 x^2))^2 \mathrm{d}x.$$

In other words, we want to minimize the square of the distance in $C[-1, 1]$ from $\cos x$ to the subspace spanned by 1, x and x^2.

We can realize the solution by proceeding exactly as in the case of $\mathbb{R}^n$, except that we replace $\mathbb{R}^n$ by the subspace of $C[-1, 1]$ generated by 1, x, x^2 and $\cos x$. Thus we need to compute the Fourier coefficients

$$\alpha_i = (\cos x, f_i) = \int_{-1}^{1} \cos x f_i(x) \mathrm{d}x.$$

It turns out that $\alpha_1 = \frac{2\cos 1}{\sqrt{2}}$, $\alpha_2 = 0$ (since $\cos x$ is even and x is odd) and $\alpha_3 = \frac{\sqrt{45}}{\sqrt{8}}(4\cos 1 - \frac{8}{3}\sin 1)$. Thus the best least squares approximation is

$$f(x) = \cos 1 + \frac{\sqrt{45}}{\sqrt{8}}(4\cos 1 - \frac{8}{3}\sin 1)(x^2 - \frac{1}{3}).$$

We leave it to the reader to finish the example by computing the explict value of $\int_{-1}^{1}(\cos x - f(x))^2 \mathrm{d}x$. (Just kidding.)

10.4.3 Hermitian Inner Products

As we also saw in Chapter 4, complex n-space $\mathbb{C}^n$ admits a Hermitian inner product $\mathbf{w} \cdot \mathbf{z}$ for all $\mathbf{w}, \mathbf{z} \in \mathbb{C}^n$. Here

$$\mathbf{w} \bullet \mathbf{z} = \sum_{i=1}^{n} \overline{w}_i z_i. \tag{10.16}$$

This can be put in matrix form by defining $\mathbf{w}^H$ to be $\overline{\mathbf{w}}^T$. Then the Hermitian inner product can be expressed as $\mathbf{w} \bullet \mathbf{z} = \mathbf{w}^H \mathbf{z}$.

The Hermitian inner product satisfies the following properties with respect to scalar multiplication by $\mathbb{C}$.

$$(\alpha \mathbf{w}) \bullet \mathbf{z} = \overline{\alpha}(\mathbf{w} \bullet \mathbf{z}),$$

and

$$\mathbf{w} \bullet (\alpha \mathbf{z}) = \alpha(\mathbf{w} \bullet \mathbf{z}),$$

for all $\alpha \in \mathbb{C}$. Thus

$$(r\mathbf{w}) \bullet \mathbf{z} = \mathbf{w} \bullet (r\mathbf{z}) = r(\mathbf{w} \bullet \mathbf{z})$$

if $r \in \mathbb{R}$. Note, however, that a Hermitian inner product is not an inner productin the usual sense, since, in general, $\mathbf{w} \bullet \mathbf{z} \neq \mathbf{z} \cdot \mathbf{w}$. Nevertheless, the Hermitian length squared $|\mathbf{w}|^2 = \mathbf{w} \bullet \mathbf{w} = \mathbf{w}^H \mathbf{w}$ coincides with the usual Euclidean length.

Recall that a general Hermitian inner product space is a complex vector space with a Hermitian inner product (see Definition 4.14). Such a space is called a *Hilbert space*.

Definition 10.3. Let V be a Hermitian inner product space, for example $\mathbb{C}^n$ with the above Hermitianinner product. A basis of V which is orthonormal for the Hermitian inner product is called a *Hermitian basis*.

For example, the standard basis $e_1, \ldots, e_n$ of $\mathbb{C}^n$ is Hermitian. The basis $\frac{1}{\sqrt{2}}(1, i)^T$, $\frac{1}{\sqrt{2}}(1, -i)^T$ of $\mathbb{C}^2$ is another example of a Hermitian basis.

We next consider the obvious extension of the Gram-Schmidt method to the Hermitian case.

Proposition 10.14. *Every finite dimensional Hermitian inner product space V admits a Hermitian orthonormal basis. In fact, given a basis $\mathbf{w}_1, \ldots, \mathbf{w}_n$ of V, there exists a Hermitian orthonormal basis $\mathbf{z}_1, \ldots, \mathbf{z}_n$ with the property that $\mathbf{z}_1, \ldots, \mathbf{z}_k$ and $\mathbf{w}_1, \ldots, \mathbf{w}_k$ span the same subspace of V for each index k.*

Proof. Notice that the obvious modification of the Gram-Schmidt method can be applied here. □

The reader should be aware that all the results about least squares and projections go through in the Hermitian case in the same way as the real case, although slight modifications are needed to take into account the difference between the real and the Hermitian inner product.

Just as orthonormal bases give rise to orthogonal matrices, Hermitian bases give rise to a class of matrices known as unitary matrices.

Definition 10.4. An $n \times n$ matrix U over $\mathbb{C}$ is said to be *unitary* if and only if $U^H U = I_n$. In other words, U is unitary if and only if the columns of U form a Hermitian basis of $\mathbb{C}^n$.

We will denote the set of all $n \times n$ unitary matrices by $U(n, \mathbb{C})$. Notice that $O(n, \mathbb{R}) \subset U(n, \mathbb{C})$.

Proposition 10.15. *$U(n, \mathbb{C})$ is a matrix group.*

Some of the other properties of unitary matrices are mentioned in the next Proposition.

Proposition 10.16. *Let $U \in \mathbb{C}^{n\times n}$. Then U is unitary if and only if the columns of U form a Hermitian basis of $\mathbb{C}^n$. Moreover,*

(i) every eigenvalue of a unitary matrix has absolute value 1, and

(ii) the determinant of a unitary matrix has absolute value 1, and

Proof. This is left as an exercise. □

10.4.4 Isometries

We now introduce the notion of an isometry. Isometries enable one to transfer properties of one inner product space to another.

Definition 10.5. Let U and V be inner product spaces, and suppose $\rho : U \to V$ is a transformation such that

$$(\rho(\mathbf{x}), \rho(\mathbf{y})) = (\mathbf{x}, \mathbf{y})$$

for all $\mathbf{x}, \mathbf{y} \in U$. Then ρ is called an *isometry*.

A similar definition holds if U and V are Hermitian inner product spaces.

Proposition 10.17. *Let U and V be any inner product spaces, and assume $\rho : U \to V$ is an isometry. Then ρ is a one to one linear transformation.*

Proof. To show ρ is linear, we have to show that for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ and $r \in \mathbb{R}$,

$$|\rho(\mathbf{x} + \mathbf{y}) - \rho(\mathbf{x}) - \rho(\mathbf{y})|^2 = 0, \tag{10.17}$$

and

$$|\rho(r\mathbf{x}) - r\rho(\mathbf{x})|^2 = 0. \tag{10.18}$$

For the first, after expanding the left hand side and using the fact that ρ is an isometry, we get

$$|\rho(\mathbf{x} + \mathbf{y}) - \rho(\mathbf{x}) - \rho(\mathbf{y})|^2 = \big((\mathbf{x} + \mathbf{y}), (\mathbf{x} + \mathbf{y})\big)$$

$$-2\big((\mathbf{x} + \mathbf{y}), \mathbf{x}\big) - 2\big((\mathbf{x} + \mathbf{y}), \mathbf{y}\big) - 2(\mathbf{x}, \mathbf{y}) - \big(\mathbf{x}, \mathbf{x}\big) - \big(\mathbf{y}, \mathbf{y}\big).$$

But expanding the right hand side gives 0, so we get (10.17). The proof that $|\rho(r\mathbf{x}) - r\rho(\mathbf{x})|^2 = 0$ is similar so we will omit it. Hence ρ is indeed linear.

To show ρ is injective, it suffices to show that $\rho(\mathbf{x}) = \mathbf{0}$ implies $\mathbf{x} = \mathbf{0}$. But $|\rho(\mathbf{x})|^2 = |\mathbf{x}|^2$, so this is clear. □

Proposition 10.18. *Suppose U and V are finite dimensional inner product spaces. Then every isometry $\rho : U \to V$ carries each orthonormal basis of U to an orthonormal basis of $\rho(U)$. Conversely, a linear transformation $\rho : U \to V$ which carries some orthonormal basis of U to an orthonormal basis of $\rho(U)$ is an isometry.*

The proof is an exercise. There is also a Hermitian version, which the reader may readily formulate. Thus a isometry of $\mathbb{R}^n$ is simply an orthogonal matrix Q. The set of isometries of $\mathbb{R}^n$ is the matrix group $O(n, \mathbb{R})$, the orthogonal group of $\mathbb{R}^n$. Similarly, the set of all linear isometries of $\mathbb{C}^n$ is the matrix group $U(n, \mathbb{C})$ consisting of all $n \times n$ unitary matrices.

The reader should note that any two finite dimensional inner product spaces of the same dimension are isometric (see Exercise 10.51). Hence if V is a finite dimensional inner product space of dimension n, then V is isometric to $\mathbb{R}^n$. If V is

Exercises

Exercise 10.47. This problem concerns the Gram-Schmidt method for the inner product on $C[-1,1]$.

(a) Apply Gram-Schmidt to the functions $1,\ x,\ x^2$ on the interval $[-1,1]$ to produce an orthonormal basis of the set of polynomials on $[-1,1]$ of degree at most two. The resulting functions P_0, P_1, P_2 are the first three normalized *orthogonal polynomials* of Legendre type.

(b) Show that your nth polynomial P_n satisfies the differential equation

$$(1-x^2)y'' - 2xy' + n(n+1)y = 0.$$

(c) The nth degree Legendre polynomial satisfies this second order differential equation equation for all $n \geq 0$. This and the orthogonality condition can be used to generate all the Legendre polynomials. Find P_3 and P_4 without GS.

Exercise 10.48. Using the result of the previous exercise, find the projection of $x^4 + x$ on the subspace of $C[-1,1]$ spanned by $1,\ x,\ x^2$.

Exercise 10.49. Find approprate constants $a, b \in \mathbb{R}$ which minimize the value of

$$\int_{-\pi}^{\pi} \big(\tan x - a - b\cos x\big)^2 dx.$$

Don't evaluate the integral.

Exercise 10.50. Show that every finite dimensional real vector space admits an inner product.

Exercise 10.51. Show that any two finite dimensional inner product spaces of the same dimension are isometric.

Exercise 10.52. Prove Proposition 10.18.

Exercise 10.53. Suppose V is an inner product space with a given basis $\mathbf{v}_1, \ldots, \mathbf{v}_n$. Using this basis, we can introduce coordinates on V, as in Chapter 7. Does the formula $A(A^TA)^{-1}A^T$ for projecting still work when the columns of A are the coordinates of some independent subset of V with respect to the basis consisting of the $\mathbf{v}_i$? Discuss this.

Exercise 10.54. Show that $U(n,\mathbb{C})$ is a matrix group.

Exercise 10.55. Suppose $U \in U(n,\mathbb{C})$. Show that $|\mathrm{Trace}(U)| \leq n$.

10.5 The Group of Rotations of $\mathbb{R}^3$

In crystallography, the study of the molecular structure of crystals, one of the basic problems is to determine the set of rotational symmetries of a particular crystal. A general problem of this type is determine the set of all rotational symmetries $\mathrm{Rot}(S)$ of an arbitrary solid S in $\mathbb{R}^3$. We will show that $\mathrm{Rot}(S)$ is in fact a matrix group. This isn't a trivial assertion, since it isn't clear from the definition of a rotation given below that the composition of two rotations is a rotation. Note that we will always consider the identity to be a rotation. Hence, $\mathrm{Rot}(S)$ is always non empty, although frequently the identity is its only element.

A simple, down to earth problem is to describe $\mathrm{Rot}(S)$ when S is a Platonic solid in $\mathbb{R}^3$. Note that a *Platonic solid* is a solid whose boundary is a union of plane polygons, all of which are congruent. It has been known since the ancient Greeks that there are exactly five types of Platonic solids: a cube, a regular quadrilateral, a regular tetrahedron, a regular dodecahedron and a regular icosahedron.

10.5.1 Rotations of $\mathbb{R}^3$

The first problem is how to define a what rotation of $\mathbb{R}^3$ should be. We will use a definition due to Euler, which intuitively makes good sense. Namely, a *rotation of* $\mathbb{R}^3$ is a transformation $\rho : \mathbb{R}^3 \to \mathbb{R}^3$ which fixes every point on some line ℓ through $\mathbf{0}$ and which rotates every plane orthogonal to ℓ through the same fixed angle θ.

Using this as the basic definition, we will first give a concrete description of the set of rotations of $\mathrm{Rot}(\mathbb{R}^3)$ and prove that it is indeed a matrix group. Suppose $\rho \in \mathrm{Rot}(\mathbb{R}^3)$, and let $\mathbf{x},\ \mathbf{y}$ be a pair of arbitrary nonzero elements of $\mathbb{R}^3$. Clearly $|\rho(\mathbf{x})| = |\mathbf{x}|$, $|\rho(\mathbf{y})| = |\mathbf{y}|$ and the angle between $\rho(\mathbf{x})$ and $\rho(\mathbf{y})$ is the same as the angle between $\mathbf{x}$ and $\mathbf{y}$. It follows that

$$\rho(\mathbf{x}) \cdot \rho(\mathbf{y}) = \mathbf{x} \cdot \mathbf{y}. \tag{10.19}$$

Therefore, by Proposition 10.17, ρ is linear, and, by Proposition 10.18, its matrix $M(\sigma)$ with respect to the standard basis is orthogonal. Hence, identifying $\sigma \in \mathrm{Rot}(\mathbb{R}^3)$ with its matrix $M(\sigma)$, we therefore see that $\mathrm{Rot}(\mathbb{R}^3) \subset O(3, \mathbb{R})$, the set of 3×3 orthogonal matrices.

Our next claim is that every $\rho \in \mathrm{Rot}(\mathbb{R}^3)$ has a determinant one. Indeed, by definition, ρ leaves a line ℓ through the origin pointwise fixed, so ρ has eigenvalue 1. Moreover, the plane P through $\mathbf{0}$ orthogonal to ℓ is rotated

through an angle θ, hence there exists an orthonormal basis of $\mathbb{R}^3$ for which the matrix R of ρ has the form

$$R = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{pmatrix}.$$

Obviously, $\det(R) = 1$, so $\det(\rho) = 1$ also since $\det(\sigma)$ is the determinant of any matrix representing §.

To complete the description of $\mathrm{Rot}(\mathbb{R}^3)$, we need to bring in the matrix group $SO(3)$. Recall that $SL(3,\mathbb{R})$ denotes the matrix group of all 3×3 real matrices of determinant 1. As mentioned in Chapter 8, $SL(3,\mathbb{R})$ consists of 3×3 real matrices A whose linear transformation T_A preserves both volumes and orientation. Put $SO(3) = SL(3,\mathbb{R}) \cap O(3,\mathbb{R})$. We call $SL(3,\mathbb{R})$ the *special linear group* and $SO(3)$ the *special orthogonal group*. By the above, $\mathrm{Rot}(\mathbb{R}^3) \subset SO(3)$. In fact, we will now show

Theorem 10.19. *After identifying each $\sigma \in \mathrm{Rot}(\mathbb{R}^3)$ with its matrix $M(\sigma)$, we have $\mathrm{Rot}(\mathbb{R}^3) = SO(3)$.*

Proof. It suffices to show $SO(3) \subset \mathrm{Rot}(\mathbb{R}^3)$, i.e. every element of $SO(3)$ defines a rotation. I claim that if $\sigma \in SO(3)$, then 1 is an eigenvalue of σ, and moreover, if $\sigma \neq I_3$, the eigenspace E_1 of 1 has dimension exactly one. That is, E_1 is a line. To see this, recall that every 3×3 real matrix has at least one real eigenvalue. Also, since we also know that the real eigenvalues of an orthogonal matrix are either 1 or -1, it follows that the eigenvalues of σ satisfy one of the following possibilities:

(i) 1 of multiplicity three,

(ii) $1, -1$, where -1 has multiplicity two, and

(iii) 1, λ, $\overline{\lambda}$, where $\lambda \neq \overline{\lambda}$.

The last possibility is due to the fact that the complex roots of a real polynomial occur in conjugate pairs, and $\lambda\overline{\lambda} > 0$ if $\lambda \neq 0$.

In all three cases, 1 is an eigenvalue of any $\sigma \in SO(3)$, so $\dim E_1 \geq 1$. I claim that if $\sigma \in SO(3)$ and $\sigma \neq I_3$, then $\dim E_1 = 1$. Indeed, if $\dim E_1 = 3$, then $\sigma = I_3$, so we only have to eliminate the possibility that $\dim E_1 = 2$. But if $\dim E_1 = 2$, then σ fixes the plane E_1 pointwise. Since σ also preserves angles, it also has to send the line $L = E_1^{\perp}$ to itself. Thus L is also an eigenspace. The only possible real eigenvalues of σ being 1 or -1, we deduce from the assumption that $\dim E_1 = 2$ that there exists a basis of $\mathbb{R}^3$ so that

the matrix of σ (i.e. T_σ) is

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

But that implies $\det(\sigma) = -1$, so $\dim E_1 = 2$ cannot happen. This gives the claim that $\dim E_1 = 1$ if $\sigma \neq I_3$.

Therefore σ fixes every point on a unique line ℓ through the origin and maps the plane $\ell^\perp$ orthogonal to ℓ into itself. It remains to show σ rotates $\ell^\perp$. Let $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3$ be an orthonormal basis in $\mathbb{R}^3$ such that $\mathbf{u}_1, \mathbf{u}_2 \in \ell^\perp$ and $\mathbf{u}_3 \in \ell$ (so $\sigma(\mathbf{u}_3) = \mathbf{u}_3$). Since $\sigma\mathbf{u}_1$ and $\sigma\mathbf{u}_2$ are orthogonal unit vectors on $\ell^\perp$, we can choose an angle θ such that

$$\sigma\mathbf{u}_1 = \cos\theta\mathbf{u}_1 + \sin\theta\mathbf{u}_2$$

and

$$\sigma\mathbf{u}_2 = \pm(\sin\theta\mathbf{u}_1 - \cos\theta\mathbf{u}_2).$$

Let Q be the matrix of σ with respect to the basis $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3$. Then

$$Q = \begin{pmatrix} \cos\theta & \pm\sin\theta & 0 \\ \sin\theta & \pm(-\cos\theta) & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

But $\det(\sigma) = 1$ so $\det(Q) = 1$ also. Hence the only possibility is that

$$Q = \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{10.20}$$

Consequently, σ fixes ℓ pointwise and rotates the plane $\ell^\perp$ through the angle θ. This implies σ defines a rotation of $\mathbb{R}^3$, so the proof is finished. □

We get a pleasant conclusion.

Corollary 10.20. $\mathrm{Rot}(\mathbb{R}^3)$ *is a matrix group. In particular, the composition of two rotations of* $\mathbb{R}^3$ *fixing* $\mathbf{0}$ *is another rotation fixing* $\mathbf{0}$.

Proof. Since $SO(3) = SL(3,\mathbb{R}) \cap O(3,\mathbb{R})$, it is the intersection of two matrix groups. But the intersection of two matrix groups is also a matrix group, so the result follows. □

The fact that the composition of two rotations about $\mathbf{0}$ is a rotation about $\mathbf{0}$ is certainly not obvious from the definition. It depends heavily

on the fact that $\det(AB) = \det(A)\det(B)$ if A and B lie in $\mathbb{R}^{3\times 3}$. But this is verifiable by a straightforward calculation. The question of how one describes the unique line fixed pointwise by the composition of two rotations about $\mathbf{0}$ can be answered, but the result is not worth mentioning here.

The above argument gives another result.

Proposition 10.21. *The matrix of a rotation $\sigma \in SO(3)$ is similar via another rotation to a matrix of the form*

$$Q = \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

10.5.2 Rotation Groups of Solids

One of the nicest application of rotations is to study the symmetries of solids such as crystals. As usual, we will continue to identify a matrix A and its transformation T_A. We begin with

Definition 10.6. Let S be a solid in $\mathbb{R}^3$. The *rotation group* of S is defined to be the set of all $\sigma \in SO(3)$ such that $\sigma(S) = S$. We denote the rotation group of S by $\mathrm{Rot}(S)$.

Proposition 10.22. *Let S be a solid in $\mathbb{R}^3$. If σ and τ are rotations of S, then so are $\sigma\tau$ and σ^{-1}. Hence the rotation group $\mathrm{Rot}(S)$ of S is a matrix group.*

Proof. Clearly σ^{-1} and $\sigma\tau$ are both in $SO(3)$. It's also clear that $\sigma^{-1}(S) = S$ as well as $\sigma\tau(S) = S$. Since $I_3 \in Rot(S)$, $\mathrm{Rot}(S)$ is a matrix group. □

Let us now determine the group of rotations of a cube.

Example 10.9. Let S denote the cube with vertices at the points (A, B, C), where $A, B, C = \pm 1$. Let us find $\mathrm{Rot}(S)$. Every rotation of $\mathbb{R}^3$ which maps S to itself maps each one of its six faces to another face. Moreover, since any face contains a basis of $\mathbb{R}^3$, each $\sigma \in \mathrm{Rot}(S)$ is completely determined by how it acts on any face. Let F denote one of the faces. Given any one of the six faces F' of S, there is at least one σ such that $\sigma(F) = F'$. Furthermore, each face has four rotations, so by Proposition 10.22, so $\mathrm{Rot}(S)$ has at least 24 distinct elements.

Next consider the 4 diagonals of S, i.e. the segments which join a vertex (A, B, C) to $(-A, -B, -C)$. Every rotation of S permutes these segments, and two rotations which define the same permutation of the diagonals coincide (why?). Since the number of permutations of 4 objects is $4! = 24$, it

follows that $\mathrm{Rot}(S)$ also has at most 24 elements. Therefore, $\mathrm{Rot}(S)$ contains exactly 24 rotations. Moreover, every permutation of the diagonals is given by a rotation.

Example 10.10. It turns out that it is interesting to know how to define the rotation which fixes two of the diagonal and keep the other two diagonals the same. This is because of the general fact every permutation of n objects can be written as a product of permutations that switch two of the objects and leave all the others fixed. (Recall that in Chapter 8, we called such permutations transpositions. In the context of row operations, they are simply the row swaps.) Let the vertices on the top face of the cube in the previous example be labelled a, b, c, d in clockwise order looking down, where $a = (1, 1, 1)$. The diagonals meet the cube in opposite points such as a and $-a$. Let's denote this diagonal by $\{a, -a\}$, and denote the other diagonals in a similar way. Suppose we want to interchange the opposite diagonals $\{a, -a\}$ and $\{c, -c\}$ and leave the other two diagonals fixed. We do this in two steps. First let H_1 be the reflection of $\mathbb{R}^3$ through the plane containing $\mathbf{0}$, a and $c = (-1, -1, 1)$. Clearly $H_1(S) = S$. Next, let H_2 be the reflection of $\mathbb{R}^3$ through the xy-plane. That is,

$$H_2 \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x \\ y \\ -z \end{pmatrix}.$$

Since every reflection is orthogonal, $H_2H_1 \in O(3, \mathbb{R})$. Moreover, $H_2H_1(S) = S$. But $\det(H_2H_1) = \det(H_2)\det(H_1) = (-1)(-1) = 1$ since the determinant of the reflection through a hyperplane is -1, so $H_2H_1 \in SO(3)$. Hence H_2H_1 is a rotation. You can check that this rotation interchanges the diagonals $\{a, -a\}$ and $\{c, -c\}$ and fixes the other two diagonals. The line about which $\mathbb{R}^3$ is rotated is easy to find. Since H_1 and H_2 each leave a plane pointwise fixed, their product leaves the intersection of these two planes pointwise fixed. Hence, H_2H_1 is the rotation of S is about the intersection of the xy-plane and the plane through the diagonals $\{a, -a\}$ and $\{c, -c\}$. This is the line $x = y$, $z = 0$.

Example 10.11. Consider the set consisting of the midpoints of the 6 faces of the cube S. The solid polygon S' determined by these 6 points is called the regular octahedron. It is a solid with 8 triangular faces all congruent to each other. The cube and the regular octahedron are two of the 5 Platonic solids. Since each element of $\mathrm{Rot}(S)$ must also send midpoint to another midpoint, it follows that $\mathrm{Rot}(S) \subset \mathrm{Rot}(S')$. The other containment clearly also holds, so we deduce that $\mathrm{Rot}(S) = \mathrm{Rot}(S')$.

Notice that this discussion shows there are exactly 24 3×3 orthogonal matrices of determinant one whose entries are 0 or ± 1.

Exercises

Exercise 10.56. Let S be a regular quadrilateral in $\mathbb{R}^3$, that is S has 4 faces made up of congruent triangles. How many elements does $Rot(S)$ have?

Exercise 10.57. Compute $\mathrm{Rot}(S)$ in the following cases:

(a) S is the half ball $\{x^2 + y^2 + z^2 \leq 1, z \geq 0\}$, and

(b) S is the solid rectangle $\{-1 \leq x \leq 1, -2 \leq y \leq 2, -1 \leq z \leq 1\}$.

Exercise 10.58. 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 H reflects $\mathbb{R}^2$ through.)

Exercise 10.59. Prove that there are exactly 24 3×3 orthogonal matrices of determinant one whose entries are 0 or ± 1.

Exercise 10.60. Compute the number of elements of the rotation group of a regular tetrahedron S centered at the origin.

10.6 Summary

The purpose of this chapter was to study inner product spaces, most notably $\mathbb{R}^n$. We began by showing that if W is any subspace of $\mathbb{R}^n$, then $\mathbb{R}^n$ is the direct sum $W \oplus W^\perp$, where $W^\perp$ is the orthogonal complement to W consisting of all $\mathbf{x} \in \mathbb{R}^n$ such that $\mathbf{x} \cdot \mathbf{w} = 0$ for all $\mathbf{w} \in W$. Thus any $\mathbf{x} \in \mathbb{R}^n$ can be expressed uniquely as $\mathbf{x} = \mathbf{w} + \mathbf{y}$ where $\mathbf{w} \in W$ and $\mathbf{w} \cdot \mathbf{y} = 0$. Then $\mathbf{w}$ is called the projection of $\mathbf{x}$ on W or the component of $\mathbf{x}$ in W, and we write $P_W(\mathbf{x}) = \mathbf{w}$. We then solved the subspace distance problem by showing that the minimal distance from $\mathbf{x}$ to W is $|\mathbf{y}|$.

The projection P_W can be expressed in two ways. One expression has the form $P_W = A(AA^T)^{-1}A^T$, where A is any matrix of maximal ranksuch that $\text{col}(A) = W$. The second expression requires an orthonormal basis, i.e. a basis of W consisting of mutually orthogonal unit vectors. This expression is called the projection formula. It uses the Fourier coefficients $\mathbf{x} \cdot \mathbf{u}_i$, where the $\mathbf{u}_i$ are the orthonormal basis vectors.

Another basic result we proved is that every subspace of $\mathbb{R}^n$ has an orthonormal basis. Of course, this also holds in any finite dimensional inner product space. We first gave an existence proof and then wrote down an explicit method for calculating an orthonormal basis called the Gram-Schmidt method. Two other topics covered in this Section were pseudo-inverse and the QR-factorization of a matrix.

Finally, we applied eigentheory to classify the rotations of $\mathbb{R}^3$ in the sense of Euler. These turn out to coincide with the matrix group $SO(3)$ consisting of all orthogonal matrices of determinant one. After this, we discussed the rotations of a cube and a regular octahedron both centered at the origin, showing that each solid has exactly 24 rotations.

Chapter 11

Unitary Diagonalization Theorems

The purpose of this Chapter is to classify the unitarily diagonalizable complex matrices. Put another way, we will describe the complex linear transformations admitting a Hermitian orthonormal eigenbasis. The first result we'll prove is that every $n \times n$ matrix A over $\mathbb{C}$ is unitarily triagonalizable. That is, there exists a unitary matrix U and an upper triangular matrix T such that $A = UTU^{-1}$. Since, by the definition of a unitary matrix, $U^{-1} = U^H$, where $U^H = (\overline{U})^T$.

After proving this preliminary result, we'll introduce the notion of a normal matrix and show that the normal matrices are exactly the unitarily diagonalizable complex matrices. After that, we'll discuss self adjoint operators and the Principal Axis Theorem.

11.1 The Normal Matrix Theorem

Let A be an $n \times n$ matrix over $\mathbb{C}$. Then, by the Fundamental Theorem of Algebra, the characteristic polynomial of A has n complex roots. Recall (see Proposition 9.13) that we already showed that every complex $n \times n$ matrix A is similar to an upper triangular matrix. We will now show that in fact there exists a unitary U such that $U^H AU$ is upper triangular. This is a the main step in the classification of unitarily diagonalizable matrices.

Theorem 11.1. *Let A be an $n \times n$ complex matrix. Then there exists an $n \times n$ unitary matrix U and an upper triangular matrix T so that $A = UTU^H = UTU^{-1}$.*

Proof. We will induct on n. As the result is trivial if $n = 1$, let us suppose $n > 1$ and that Schur's Theorem is true for all $k \times k$ matrices over $\mathbb{C}$ whenever $k < n$. By the fact that A has n eigenvalues in $\mathbb{C}$, there exists an eigenpair $(\lambda_1, \mathbf{u}_1)$ for A. Applying the Hermitian version of the Gram-Schmidt process, we may include $\mathbf{u}_1$ in a Hermitian orthonormal basis $\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_n$ of $\mathbb{C}^n$. Let $U_1 = (\mathbf{u}_1 \ \mathbf{u}_2 \ \cdots \mathbf{u}_n)$ be the corresponding unitary matrix. By construction,

$$AU_1 = (A\mathbf{u}_1 \ A\mathbf{u}_2 \ \cdots A\mathbf{u}_n) = (\lambda_1\mathbf{u}_1 \ A\mathbf{u}_2 \ \cdots A\mathbf{u}_n).$$

Hence

$$U_1^H AU_1 = \begin{pmatrix} \mathbf{u}_1^H \\ \mathbf{u}_2^H \\ \vdots \\ \mathbf{u}_n^H \end{pmatrix} (\lambda_1\mathbf{u}_1 \ A\mathbf{u}_2 \ \cdots A\mathbf{u}_n) = \begin{pmatrix} \lambda_1\mathbf{u}_1^H\mathbf{u}_1 & * & \cdots & * \\ \lambda_1\mathbf{u}_2^H\mathbf{u}_1 & * & \cdots & * \\ \vdots & \vdots & & \vdots \\ \lambda_1\mathbf{u}_3^H\mathbf{u}_1 & * & \cdots & * \end{pmatrix},$$

and since $\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_n$ are Hermitian orthonormal,

$$U_1^H AU_1 = \begin{pmatrix} \lambda_1 & * & \cdots & * \\ 0 & * & \cdots & * \\ \vdots & \vdots & & \vdots \\ 0 & * & \cdots & * \end{pmatrix}. \tag{11.1}$$

Now concentrate on the $(n-1) \times (n-1)$ matrix in the lower right hand corner of $U_1^H AU_1$, Calling this matrix B, the induction assumption applies to the $(n-1) \times (n-1)$ matrix B in the lower right hand corner of $U_1^H AU_1$, so there exists an $(n-1) \times (n-1)$ unitary matrix U' so that $(U')^H BU'$ is an upper triangular matrix, say T. The matrix

$$U_2 = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & U' & \\ 0 & & & \end{pmatrix}$$

is obviously unitary, and

$$U_2^H(U_1^H AU_1)U_2 = \begin{pmatrix} \lambda_1 & * & \cdots & * \\ 0 & & & \\ \vdots & & T & \\ 0 & & & \end{pmatrix}.$$

Since T is upper triangular, $U_2^H(U_1^H A U_1)U_2$ is also upper triangular. We are therefore done if $U = U_1U_2$ is unitary. But , by Proposition 10.15, the product of two unitary matrices is unitary, so indeed A has been put into upper triangular form by a unitary matrix. This completes the induction step, so the Theorem is proven. □

11.1.1 Normal Matrices and the Main Theorem

We will now prove the Normal Matrix Theorem.

Definition 11.1. A matrix $N \in \mathbb{C}^{n\times n}$ is said to be *normal* if and only if

$$NN^H = N^H N. \tag{11.2}$$

The Normal Matrix Theorem goes as follows.

Theorem 11.2. *A matrix $A \in \mathbb{C}^{n\times n}$ is unitarily diagonalizable if and only if A is normal.*

Proof. The only if part of the Theorem is straightforward and is left as an exercise. Suppose A is normal. By Schur's Theorem, we can 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 $TT^H = T^HT$ (why?). Hence we need to show that an upper triangular normal matrix is diagonal. The key is to compare the diagonal entries of TT^H and TT^H. Let t_{ii} be the ith diagonal entry of T, and let $\mathbf{a}_i$ denote the ith column of T. Now the diagonal entries of T^HT are $|\mathbf{a}_1|^2, |\mathbf{a}_2|^2, \ldots, |\mathbf{a}_n|^2$. On the other hand, the diagonal entries of T^HT are $|t_{11}|^2, |t_{22}|^2, \ldots, |t_{nn}|^2$. Therefore, $|\mathbf{a}_i|^2 = |t_{ii}|^2$ for each i, so T is diagonal. It follows that A is unitarily diagonalizable, so the proof is complete. □

The reader should note also that if $A \in \mathbb{C}^{n\times n}$ is unitarily diagonalizable, then so is A^H.

11.1.2 Examples

Let us now give some examples of normal matrices. Clearly, all real symmetric matrices are normal. A more general class of normal matrices is obtained by considering matrices K such that $K^H = K$.

Example 11.1 (Hermitian Matrices)**.** A matrix $K \in \mathbb{C}^{n\times n}$ such that $K^H = K$ is called *Hermitian*. Clearly the diagonal of a Hermitian matrix is real.

Notice that a real Hermitian matrix is just a symmetric matrix. For example, the matrix

$$K = \begin{pmatrix} 0 & 1+i & -i \\ 1-i & 0 & 2 \\ i & 2 & 0 \end{pmatrix}$$

is Hermitian. Its characteristic polynomialis $p_K(\lambda) = -\lambda^3 + 7\lambda - 4$. This polynomial has 3 real roots, but they are difficult to express.

The following Proposition gives a useful fact about both Hermitian and symmetric matrices.

Proposition 11.3. *A matrix $K \in \mathbb{C}^{n\times n}$ is Hermitian if and only if K is normal and all its eigenvalues are real. In particular, all eigenvalues of a real symmetric matrix are real.*

Proof. Suppose K is Hermitian. Then we can write $K = UDU^H$, with U unitary and D diagonal. Since $K = K^H = UD^HU^H$, it follows that $D^H = D$, so D is real. Hence all eigenvalues of K are real. Conversely, if K is normal with real eigenvalues, we can write $K = UDU^H$ as above with D real. Thus $K^H = K$, so K is Hermitian. □

One can also obtain other classes of normal matrices by writing $A = UDU^H$ and imposing conditions on the form of the eigenvalues, i.e. on D.

Example 11.2 (Skew Symmetric Matrices)**.** A matrix J is said to be *skew Hermitian* iff iJ is Hermitian. Since Hermitian matrices are normal, so are skew Hermitian matrices. Also, the nonzero eigenvalues of a skew Hermitian matrix are purely imaginary, i.e. they have the form $i\lambda$ for a nonzero $\lambda \in \mathbb{R}$. A real skew Hermitian matrix is called *skew symmetric*. Thus a real matrix S is skew symmetric if and only if $S^T = -S$. 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 11.1 below). Thus a skew symmetric matrix of odd order has 0 as an eigenvalue. But as the matrix J in the above example shows, the determinant of a skew symmetric matrix of even order needn't be zero.

The characteristic polynomial of the matrix S above is $-\lambda^3 - 9\lambda$, so the eigenvalues of S are $0, \pm 3i$, confirming the observation in the previous

example that all nonzero eigenvalues of a skew Hermitian matrix are purely imaginary. Moreover, the eigenvalues of S are also conjugate since S is a real matrix.

The Normal Matrix Theorem 11.2 enables us to deduce the following

Proposition 11.4. *Suppose $A \in \mathbb{R}^{n\times n}$ is skew symmetric matrix and n is even, say $n = 2m$, and suppose 0 is not an eigenvalue of A. Then there exists an orthonormal basis $\mathbf{x}_1,\ \mathbf{y}_1, \ldots, \mathbf{x}_m,\ \mathbf{y}_m$ 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*

$$J_k = \begin{pmatrix} 0 & \lambda_k \\ -\lambda_k & 0 \end{pmatrix},$$

where $\lambda_k \in \mathbb{R}$ is a nonzero real number such that $i\lambda_k$ is an eigenvalue of A. (This basis is not necessarily unique, however.)

Proof. Choose a Hermitian orthonormal basis of $\mathbb{C}^n$ consisting of eigenvectors of A. Since A is real, its eigenvalues occur in conjugate pairs, so as 0 is not an eigenvalue, the eigenvalues can be sorted into pairs $\pm\, i\lambda_k$, where λ_k is a nonzero real number and k varies from 1 to m. This sorts the above basis into pairs $\mathbf{u}_k$ and $\mathbf{u}'_k$ where, say, $A\mathbf{u}_k = i\lambda_k\mathbf{u}_k$, and $A\mathbf{u}'_k = -i\lambda_k\mathbf{u}'_k$. In fact, since A is real, we can assume $\mathbf{u}'_k = \overline{\mathbf{u}}_k$. Since $\lambda \neq 0$, the $\mathbf{u}_j$ and the $\overline{\mathbf{u}}_k$ are mutually Hermitian orthogonal. Now, let $\mathbf{x}_k = (\mathbf{u}_k + \overline{\mathbf{u}}_k)/\sqrt{2}$ and $\mathbf{y}_k = i(\overline{\mathbf{u}}_k - \mathbf{u}_k)/\sqrt{2}$. Clearly $\mathbf{x}_k$ and $\mathbf{y}_k$ are real; that is, they lie in $\mathbb{R}^n$. Moreover, the $\mathbf{x}_j$ and $\mathbf{y}_k$ are mutually orthogonal, as one can check using the fact that the $\mathbf{u}_j$ and the $\overline{\mathbf{u}}_k$ are mutually Hermitian orthogonal. By a direct computation, $A\mathbf{x}_k = -\lambda_k\mathbf{y}_k$, and $A\mathbf{y}_k = \lambda_k\mathbf{x}_k$. Hence we have the result. □

This Proposition shows that the 2×2 matrix $J_\lambda = \begin{pmatrix} 0 & \lambda \\ -\lambda & 0 \end{pmatrix}$ determines how nonsingular skew symmetric matrices of even order are produced. If zero is allowed to be an eigenvalue, the only change is that some of the λ are zero. The general $2m \times 2m$ skew symmetric matrix S can be thought of as being orthogonally similar to a direct sum of matrices of this type. That is, there exist $\lambda_1, \ldots, \lambda_m$ such that

$$S = Q\mathrm{diag}(J_{\lambda_1}, \ldots, J_{\lambda_m})Q^T,$$

for some orthogonal Q, where $\mathrm{diag}(J_{\lambda_1}, \ldots, J_{\lambda_m})$ is the $2m\times 2m$ matrix with the J_{λ_i}s down the diagonal and zeros elsewhere.

If $S \in \mathbb{R}^{n\times n}$ is skew symmetric, then its characteristic polynomial$p_S(x)$ has the property that

$$p_S(x) = p_{S^T}(x) = p_{-S}(x) = (-1)^n p_S(-x).$$

Thus if n is even, $p_S(x)$ is an even polynomial in the sense that $p_S(-x) = p_S(x)$. This is equivalent to the property that only even powers of x occur. Similarly, if n is odd, then $p_S(-x) = -p_S(x)$, and $p_S(x)$ is an odd polynomial: only odd powers of x occur. In particular, $p_S(0) = -p_S(0)$, so 0 is an eigenvalue of S as claimed above.

Another condition on the eigenvalues is that they all have modulus one. This leads to the following example.

Example 11.3. Let $A = UDU^H$, where every diagonal entry of D is a unit complex number. Then D is unitary, hence so is A. Conversely, every unitary matrix is normal and the eigenvalues of a unitary matrix have modulus one (see Exercise 11.3), so 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 orthogonal, hence unitary. 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 $\mathbf{u}_1 = \dfrac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ i \end{pmatrix}$ and $\mathbf{u}_1' = \dfrac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -i \end{pmatrix}$. The way U acts as a complex linear transformation of $\mathbb{C}^2$ can be interpreted geometrically as follows. U rotates vectors on the principal axis $\mathbb{C}(1,i)^T$ spanned by $(1,i)^T$ (thought of as a real two plane) through $\frac{\pi}{2}$ and rotates vectors on the orthogonal principal axis by $-\frac{\pi}{2}$. Of course, as a transformation on $\mathbb{R}^2$, U is simply the rotation $R_{\pi/2}$.

Exercises

Exercise 11.1. Unitarily diagonalize the skew symmetric matrix J of Example 11.2.

Exercise 11.2. Let S be a skew Hermitian $n \times n$ matrix. Show the following:

(i) If n is odd, then $\det(S)$ is pure imaginary, 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 11.3. Let U be any unitary matrix. Show that:

(i) every eigenvalue of U also has modulus 1, and

(ii) $\det(U)$ has modulus 1.

Exercise 11.4. Suppose A is skew Hermitian. Show that e^A is unitary.

Exercise 11.5. Suppose Q be an orthogonal $n \times n$ real matrix with no real eigenvalues. True or False: $\det(Q) = 1$.

Exercise 11.6. Suppose all eigen-values of of a unitary matrix Q are 1. True or false: $Q = I_n$.

Exercise 11.7. Are all complex matrices normal?

Exercise 11.8. Let $\mathcal{N}(n) \subset \mathbb{C}^{n \times n}$ be the set of normal $n \times n$ complex matrices. Is $\mathcal{N}(n)$ a subspace of $\mathbb{C}^{n \times n}$?

Exercise 11.9. Formulate the notion of a normal operator on a Hermitian inner product space.

Exercise 11.10. Which of the following statements are always true, which are sometimes true and which are always false? Discuss your reasoning.

(i) If A is normal and U is unitary, the UAU^H is normal.

(ii) If A is normal and invertible, then A^{-1} is normal.

(ii) If A is normal and k is a positive integer, then A^k is normal.

(iv) If $A \in \mathbb{R}^{n \times n}$ is invertible, the every eigenvalue of $A^T A$ is positive.

(v) If $A \in \mathbb{R}^{n \times n}$, then AA^T and $A^T A$ have the same eigenvalues.

(vi) If A is skew symmetric, e^A is orthogonal.

Exercise 11.11. Show that if A and B are normal and $AB = BA$, then AB is normal.

11.2 The Principal Axis Theorem

The purpose of this Section is to discuss the Principal Axis Theorem for matrices. This is one of the most classical results in linear algebra. Of course, the Theorem itself is almost an immediate consequence of the Normal Matrix Theorem, but it also has some other aspects which we will mention, such as the fact that every symmetric (or more generally, Hermitian) matrix is a linear combination of projection matrices. A more geometric treatment will be given in the next Section, where the notion of a Hermitian matrix is finally given a geometric interpretation.

Theorem 11.5 (Matrix version)**.** *Let $A \in \mathbb{C}^{n\times n}$ be Hermitian. Then there exist real numbers $\lambda_1, \ldots, \lambda_n$ and Hermitian orthonormal $\mathbf{u}_1, \ldots, \mathbf{u}_n$ such that each $(\lambda_i, \mathbf{u}_i)$ is an eigenpair for A. Consequently, A is unitarily diagonalizable. More precisely,*

$$A = UDU^{-1} = UDU^H,$$

where $U = \begin{pmatrix}\mathbf{u}_1 & \ldots & \mathbf{u}_n\end{pmatrix}$ and $D = diag(\lambda_1, \ldots, \lambda_n)$. If A is real, hence symmetric, then there exists a real orthogonal Q such that

$$A = QDQ^{-1} = QDQ^T.$$

Hence any real symmetric matrix is orthogonally diagonalizable.

Proof. The Hermitian case follows immediately from the Normal Matrix Theorem. For the symmetric case, we need to prove that U may be chosen to be real. Examining the proof of Schur's Theorem, we see that since the eigenvalues of a symmetric A are real, we may assume that the unitary matrix U such that U^HAU is upper triangular may in fact be taken to be orthogonal. For this we have to know that the product of two orthogonal matrices is orthogonal. Thus there exists an orthogonal matrix Q such that Q^TAQ is a real diagonal matrix D. Therefore, $A = QDQ^{-1} = QDQ^T$, so any real symmetric matrix can be orthogonally diagonalized. □

Note that the converse of the Principal Axis Theorem is also true. Any matrix of the form UDU^{-1}, where U is unitary and D is real diagonal is Hermitian, and any matrix of the form QDQ^{-1}, where Q is orthogonal is symmetric.

The principal axes are of course the lines spanned by orthonormal the eigenbasis vectors. In the Hermitian case, they are copies of $\mathbb{C}$, and hence are actually real two planes.

One of the simple consequences of the Principal Axis Theorem is that any two eigenspaces of a Hermitian matrix which correspond to two different eigenvalues are Hermitian orthogonal. In particular, distinct eigenvalues of a real symmetric matrix have orthogonal eigenspaces. The reader can also deduce this directly from the definition of a Hermitian or self adjoint operator.

11.2.1 Examples

Example 11.4. Let H denote a 2×2 reflection matrix. Then H has eigenvalues ± 1. Either unit vector $\mathbf{u}$ on the reflecting line together with either unit vector $\mathbf{v}$ orthogonal to the reflecting line form an orthonormal eigenbasis of $\mathbb{R}^2$ for H. Thus $Q = (\mathbf{u}\ \mathbf{v})$ is orthogonal and

$$H = Q \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} Q^{-1} = Q \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} Q^T.$$

Note that there are only four possible choices for Q. All 2×2 reflection matrices are similar to $\mathrm{diag}[1, -1]$. The only thing that can vary is Q.

Here is another example.

Example 11.5. Let B be 4×4 the all ones matrix. The rank of B is one, so 0 is an eigenvalue and $\mathcal{N}(B) = E_0$ has dimension three. In fact $E_0 = (\mathbb{R}(1,1,1,1)^T)^\perp$. Another eigenvalue is 4. Indeed, $(1,1,1,1)^T \in E_4$, so we know there exists an eigenbasissince $\dim E_0 + \dim E_4 = 4$. To produce an orthonormal basis, we simply need to find an orthonormal basis of $E_0^\perp$. We will do this by inspection rather than Gram-Schmidt, since it is easy to find vectors orthogonal to $(1,1,1,1)^T$. In fact, $\mathbf{v}_1 = (1,-1,0,0)^T$, $\mathbf{v}_2 = (0,0,1,-1)^T$, and $\mathbf{v}_3 = (1,1,-1,-1)^T$ give an orthonormal basis after we normalize. We know that our fourth eigenvector, $\mathbf{v}_4 = (1,1,1,1)^T$, is orthogonal to E_0, so we can for example express B as QDQ^T where $Q = \begin{pmatrix} \frac{\mathbf{v}_1}{\sqrt{2}} & \frac{\mathbf{v}_2}{\sqrt{2}} & \frac{\mathbf{v}_3}{2} & \frac{\mathbf{v}_4}{2} \end{pmatrix}$ and

$$D = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 4 \end{pmatrix}.$$

11.2.2 A Projection Formula for Symmetric Matrices

One of the nice applications of the Principal Axis Theorem is that it enables us to express any symmetric matrix as a linear combination of orthogonal projections. For example, suppose $A \in \mathbb{R}^{n\times n}$ is symmetric, and let $\mathbf{u}_1, \ldots, \mathbf{u}_n$ be an orthonormal eigenbasis of $\mathbb{R}^n$ for A. Also, let λ_i denote the eigenvalue for $\mathbf{u}_i$. Then, if $\mathbf{x} \in \mathbb{R}^n$, the projection formula (10.8) allows us to write

$$\mathbf{x} = (\mathbf{u}_1^T\mathbf{x})\mathbf{u}_1 + \cdots + (\mathbf{u}_n^T\mathbf{x})\mathbf{u}_n.$$

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

$$A = \lambda_1\mathbf{u}_1\mathbf{u}_1^T + \cdots + \lambda_n\mathbf{u}_n\mathbf{u}_n^T. \tag{11.3}$$

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 (11.3) indeed expresses A as a linear combination of orthogonal projections. This formula holds in the Hermitian case as well.

This formula can also be put in a more elegant form. If $\mu_1, \ldots, \mu_k$ are the distinct eigenvalues of A, then

$$A = \mu_1 P_{E_{\mu_1}} + \mu_2 P_{E_{\mu_2}} + \cdots + \mu_k P_{E_{\mu_k}}. \tag{11.4}$$

For example, in the case of the all ones matrix of Example 11.5,

$$A = 0P_{E_0} + 4P_{E_4}.$$

Exercises

Exercise 11.12. Let V be a real finite dimensional inner product space. Show that a linear transformation $T : V \to V$ is self adjoint if and only if the matrix of T with respect to an arbitrary orthonormal basis is symmetric. Formulate and prove the corresponding result for the Hermitian case.

Exercise 11.13. Show directly from the definition of self adjointness that all eigenvalues of a self adjoint operator are real.

Exercise 11.14. Orthogonally diagonalize the following matrices:

$$\begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 1 & 3 \\ 1 & 3 & 1 \\ 3 & 1 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix}.$$

I claim that you can diagonalize the first and third matrices without pencil and paper. You can also find an eigenvalue of the second by inspection.

Exercise 11.15. Show directly from the definition that the eigenspaces for different eigenvalues of a self adjoint operator are mutually orthogonal.

Exercise 11.16. Consider the 2×2 real symmetric matrix $A = \begin{pmatrix} a & b \\ b & c \end{pmatrix}$.

(i) Show directly that both roots of the characteristic polynomialof A are real.

(ii) Prove that A is orthogonally diagonalizable without appealing to the Principal Axis Theorem.

Exercise 11.17. Suppose B is a real, symmetric 3×3 matrix such that $(1,\ 0,\ 1)^T \in \text{Null}(B - I_3)$, and $(1,\ 1,\ -1)^T \in \text{Null}(B - 2I_3)$. If $\det(B) = 12$, find B.

Exercise 11.18. Answer either T or F. If T, give a brief reason. If F, give a counter example.

(a) The sum and product of two symmetric matrices is symmetric.

(b) For any real matrix A, the eigenvalues of $A^T A$ are all real.

(c) For A as in (b), the eigenvalues of $A^T A$ are all non negative.

(d) If two symmetric matrices A and B have the same eigenvalues, counting multiplicities, then A and B are orthogonally similar ($A = QBQ^T$ where Q is orthogonal).

Exercise 11.19. 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 which 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 11.20. Recall that two matrices A and B with a common eigenbasis commute. Conclude that if A and B have a common eigenbasis and are symmetric, then AB is symmetric.

Exercise 11.21. Suppose A, B and AB are symmetric. Show that A and B are simultaneously diagonalizable.

Exercise 11.22. Show that if N is normal, then N and N^H are simultaneously diagonalizable.

Exercise 11.23. Let W be a subspace of $\mathbb{R}^n$. Show that the projection P_W is self adjoint.

Exercise 11.24. Let W be a hyperplane in $\mathbb{R}^n$, and let H be the reflection through W. Show that H is self adjoint; and explicitly describe how to orthogonally diagonalize its matrix.

Exercise 11.25. Let W be a subspace of $\mathbb{R}^n$. Simultaneously orthogonally diagonalize P_W and $P_{W^\perp}$.

Exercise 11.26. * Diagonalize

$$\begin{pmatrix} a & b & c \\ b & c & a \\ c & a & b \end{pmatrix},$$

where a, b, c are all real. (Note that the second matrix in Problem 1 is of this type.) What does the fact that the trace is an eigenvalue say?

Exercise 11.27. * 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: thimk!)

Exercise 11.28. Prove that a real symmetric matrix A whose only eigenvalues are ± 1 is orthogonal.

Exercise 11.29. Suppose $A \in \mathbb{R}^{n\times n}$ is symmetric. Show the following.

(i) $\mathcal{N}(A)^\perp = \mathrm{im}(A)$.

(ii) $\mathrm{im}(A)^\perp = \mathcal{N}(A)$.

(iii) $\mathrm{col}(A) \cap \mathcal{N}(A) = \{\mathbf{0}\}$.

(iv) Conclude form (iii) that if $A^k = O$ for some $k > 0$, then $A = O$.

Exercise 11.30. Give a direct proof of the Principal Axis Theorem in the 2×2 Hermitian case.

Exercise 11.31. Show that two 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 11.32. Let $A \in \mathbb{R}^{n\times n}$ be symmetric, and let λ_m and λ_M be its minimum and maximum eigenvalues respectively.

(a) Use formula (11.3) to show that for every $\mathbf{x} \in \mathbb{R}^n$, we have

$$\lambda_m \mathbf{x}^T\mathbf{x} \le \mathbf{x}^T A\mathbf{x} \le \lambda_M \mathbf{x}^T\mathbf{x}.$$

(b) Use this inequality to find the maximum and minimum value of $|A\mathbf{x}|$ on the ball $|\mathbf{x}| \le 1$.

(c) Show that if $A \in \mathbb{R}^{n\times n}$ is symmetric, then the maximum and minimum values of $\mathbf{x}^T A\mathbf{x}$ for $|\mathbf{x}| = 1$ are eigenvalues of A.

Exercise 11.33. Show that if $Q \in \mathbb{R}^{n\times n}$ is orthogonal and symmetric, then $Q^2 = I_n$. Moreover, if 1 is not an eigenvalue of Q, then $Q = -I_n$.

Exercise 11.34. Find the eigen-values of $K = \left(\begin{smallmatrix} 2 & 3+4i \\ 3-4i & -2 \end{smallmatrix}\right)$ and diagonalize K.

Exercise 11.35. Unitarily diagonalize $R_\theta = \left(\begin{smallmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{smallmatrix}\right)$.

Exercise 11.36. Show that the trace and determinant of a Hermitian matrix are real. In fact, show that the characteristic polynomial of a Hermitian matrix has real coefficients.

Exercise 11.37. Prove that the Hermitian matrices are exactly the complex matrices with real eigen-values that can be diagonalized using a unitary matrix.

Exercise 11.38. Show that $U(n)$ is a matrix group, and give a description of $U(2)$.

Exercise 11.39. What is the relationship between $U(1)$ and $SO(2)$?

Exercise 11.40. Show that two unit vectors in $\mathbb{C}^n$ coincide if and only if their Hermitian inner product is 1.

Exercise 11.41. Consider a 2×2 unitary matrix U such that one of U's columns is in $\mathbb{R}^2$. Is U orthogonal?

Exercise 11.42. Supppose W is a complex subspace of $\mathbb{C}^n$. Show that the projection P_W is Hermitian.

Exercise 11.43. How does one adjust the formula $P_W = A(AA^T)^{-1}A^T$ to get the formula for the projection of a complex subspace W of $\mathbb{C}^n$?

11.3 Diagonalization of Self Adjoint Operators

The Principal Axis Theorem says that every Hermitian matrix is unitarily diagonalizable. This is, of course, a special case of the Normal Matrix Theorem. On the other hand, since Hermitian matrices are quite fundamental, we will make some additional comments. In particular, we will give the operator theoretic meaning of the Hermitian condition.

11.3.1 Self Adjoint Operators

Let V be either a real or Hermitian inner product space. Consider a linear transformation $T : V \to V$ such that

$$(T(\mathbf{x}), \mathbf{y}) = (\mathbf{x}, T(\mathbf{y}))$$

for all $\mathbf{x}, \mathbf{y} \in V$. In the real case, we will say T is *self adjoint* , and in the complex case, we will say T is *Hermitian self adjoint*, or, for brevity, that T is a *Hermitian operator*.

Suppose, to begin, that $V = \mathbb{R}^n$. Here, self adjoint operators turn out to be very familiar objects.

Proposition 11.6. *A linear transformation $T : \mathbb{R}^n \to \mathbb{R}^n$ is self adjoint if and only if $T = T_A$, where A is symmetric.*

Proof. We will leave this as an exercise. □

Similarly, in the Hermitian case we have

Proposition 11.7. *A linear transformation $T : \mathbb{C}^n \to \mathbb{C}^n$ is Hermitian self adjoint if and only if the matrix of T is Hermitian.*

Proof. This is also an exercise. □

In a general, if V is a finite dimensional inner product space, then a linear transformation $T : V \to V$ is self adjoint if and only if the matrix of T with respect to an orthonormal basis of V is Hermitian, in the complex case, and symmetric, in the real case. It is not hard to show directly from the definition that all eigenvalues of a self adjoint operator are real.

11.3.2 The Geometry of Self Adjointness

The geometric meaning of the condition that a linear transformation is self adjoint is summed up in the following.

Proposition 11.8. *Suppose $T : V \to V$ is a self adjoint linear transformation on a finite dimensional inner product space V. Then:*

(i) if W is a subspace such that $T(W) \subset W$, then $T(W^\perp) \subset W^\perp$; and

(ii) the image $\mathrm{im}(T)$ of T is $\ker(T)^\perp$.

Proof. For (i), 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 $T(\mathbf{y}) \in W^\perp$, so (i) is proved. For (ii), let $W = \ker(T)$. I claim that $\mathrm{im}(T) \subset W^\perp$. Indeed, if $\mathbf{y} = T(\mathbf{x})$ and $\mathbf{w} \in W$, then $(\mathbf{w}, \mathbf{y}) = (T(\mathbf{w}), \mathbf{x}) = 0$ since $T(\mathbf{w}) = \mathbf{0}$. By Proposition 10.1,

$$\dim V = \dim W + \dim W^\perp.$$

We also know from Theorem 7.17 that

$$\dim V = \dim \ker(T) + \dim \mathrm{im}(T).$$

Hence, $\dim \mathrm{im}(T) = \dim \ker(T)^\perp$. But $\mathrm{im}(T) \subset W^\perp$, so (ii) is proven. □

Let us next return to a couple of familiar examples.

Example 11.6. Let W be a subspace of $\mathbb{R}^n$. Then the projection P_W is self adjoint. In fact, we know that its matrix with respect to the standard basis has the form $C(CC^T)^{-1}C^T$, which is clearly symmetric. Another way to see the self adjointness is to choose an orthonormal basis $\mathbf{u}_1, \ldots, \mathbf{u}_n$ of $\mathbb{R}^n$ so that $\mathbf{u}_1, \ldots, \mathbf{u}_m$ span W. Then, by the projection formula, $P_W(\mathbf{x}) = \sum_{i=1}^k (\mathbf{x} \cdot \mathbf{u}_i)\mathbf{u}_i$. It follows immediately that $P_W(\mathbf{u}_i) \cdot \mathbf{u}_j = \mathbf{u}_i \cdot P_W(\mathbf{u}_j)$ for all indices i and j, and this implies P_W is self adjoint.

11.3.3 Another Proof of the Principal Axis Theorem

We will now give a proof of the Principal Axis Theorem which takes advantage of the geometric properties of self adjoint operators.

Theorem 11.9. *Let V be a Hermitian finite dimensional inner product space, and let $T : V \to V$ be self adjoint. Then there exists an Hermitian orthonormal eigenbasis $\mathcal{Q}$ of V consisting of eigenvectors of T. Thus T is semi-simple, and the matrix $\mathcal{M}_{\mathcal{Q}}^{\mathcal{Q}}(T)$ is diagonal.*

Proof. Just as in the first proof, we will induct on $\dim V$. The case $\dim V = 1$ is clear, so suppose the theorem is true if $\dim V = k$ whenever $k < n$, where $n > 1$. Let $(\lambda, \mathbf{w})$ be an eigenpair for T, where $\mathbf{w}$ is a unit vector, and put $W = \mathbb{C}\mathbf{w}$. Obviously, $T(W) \subset W$, so the previous Proposition implies

that $T(W^\perp) \subset W^\perp$. Thus, T defines a self adjoint operator on $W^\perp$. But $\dim W^\perp = \dim V - 1$, so, by induction, $W^\perp$ admits an orthonormal eigenbasis. Combining this orthonormal basis of $W^\perp$ with $\mathbf{w}$ gives n orthonormal elements of V, hence an orthonormal basis of V. This gives the sought after eigenbasis of V for T, so the induction proof is done. □

Note that one can deduce that the eigenvalues of a Hermitian self adjoint map T are real directly from the self adjointness condition $(T(\mathbf{x}), \mathbf{x}) = (\mathbf{x}, T(\mathbf{x}))$. The proof of the Principal Axis Theorem in the real case is just the same, except that one needs to use the additional fact just mentioned the eigenvalues of a real self adjoint operator are real.

Example 11.7. Consider the linear transformation $T : \mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}$ defined by $T(A) = A^T$. (This is indeed a linear transformation.) Recall that the inner product is given by $(A, B) = \text{Trace}(A^T B)$.

I claim T is self adjoint. In other words, we claim that $(A^T, B) = (A, B^T)$ for all A, B. To see this directly, one has to show $\text{Trace}(AB) = \text{Trace}(A^T B^T)$, which we leave as an exercise.

Another way to show T is self adjoint is to show that it admits an orthonormal eigenbasis. Notice that since $T^2 = I_m$, where $m = n^2$, the eigenvalues of T are plus and minus one. Such an orthonormal eigenbasis consists of the symmetric matrices E_{ii} and $(E_{ij} + E_{ji})/\sqrt{2}$ where $i \neq j$, and the skew symmetric matrices $(E_{ij} - E_{ji})/\sqrt{2}$ for $i \neq j$. This shows that every real matrix can be expressed uniquely as the sum of a symmetric matrix and a skew symmetric matrix. (This fact was proved over an arbitrary field of characteristic different from two in Chapter 7.)

11.3.4 An Infinite Dimensional Self Adjoint Operator

We now give an example of a self adjoint operator (i.e. linear transformation) in an infinite dimensional context. Such self adjoint operators are frequently encountered in mathematical, as well as physical, problems. Note that although our finite dimensional methods don't help in finding eigenvalues in the infinite dimensional case, it is still true, by the same argument, that the eigenvalues of a self adjiont operator are real.

Our inner product space will be a certain subspace of function space $C[0, 2\pi]$, the continuous functions $f : [0, 2\pi] \to \mathbb{R}$ with the usual inner product

$$(f, g) = \int_0^{2\pi} f(x)g(x)\mathrm{d}x.$$

A linear transformation $T : C[0, 2\pi] \to C[0, 2\pi]$ is self adjoint if and only if $(Tf, g) = (f, Tg)$ for all f, g, that is

$$\int_0^{2\pi} T(f)(x)g(x)\mathrm{d}x = \int_0^{2\pi} f(x)T(g)(x)\mathrm{d}x.$$

Consider the subspace $\mathcal{P}$ of $C[0, 2\pi]$ consisting of all the functions f which have derivatives of all orders on $[0, 2\pi]$ and satisfy the condition that

$$f^{(i)}(0) = f^{(i)}(2\pi) \qquad \text{if} \qquad i = 0, 1, 2, \ldots,$$

where $f^{(i)}$ denotes the ith derivative of f. These functions can be viewed as the smooth functions on the unit circle. Among them are the trigonometric functions $\cos \lambda x$ and $\sin \lambda x$, where $\lambda \in \mathbb{R}$. We will show below that $\mathcal{P}$ is infinite dimensional.

Our aim is to give an example of a self adjoint operator on $\mathcal{P}$. By definition, if $f \in \mathcal{P}$, then $f^{(i)} \in \mathcal{P}$ for all $i \geq 1$. Hence the ith derivative operator $D^i(f) = f^{(i)}$ is a linear transformation $D^i : \mathcal{P} \to \mathcal{P}$ for all $i > 0$. I claim the second derivative operator $D^2(f) = f''$ is self adjoint. This is shown using integration by parts. For

$$\begin{aligned}(D^2(f), g) &= \int_0^{2\pi} f''(t)g(t)\mathrm{d}t \\ &= f'(2\pi)g(2\pi) - f'(0)g(0) - \int_0^{2\pi} f'(t)g'(t)\mathrm{d}t.\end{aligned}$$

But by the definition of $\mathcal{P}$, $f'(2\pi)g(2\pi) - f'(0)g(0) = 0$, so

$$(D^2(f), g) = -\int_0^{2\pi} f'(t)g'(t)\mathrm{d}t.$$

Since this expression for $(D^2(f), g)$ is symmetric in f and g, it follows that

$$(D^2(f), g) = (f, D^2(g)),$$

so D^2 is self adjoint, as claimed.

Since D^2 is self adjoint, one is naturally curious about its eigenvalues. In general, there is no method for finding the eigenvalues of a linear operator on an infinite dimensional space. Certainly, the characteristic polynomialdoesn't make sense in this setting. But one can easily see that the trig functions $\cos \lambda x$ and $\sin \lambda x$ are eigenfunctions for $-\lambda^2$ if $\lambda \neq 0$. In fact, by a

general theorem in differential equations, if $\mu \geq 0$, then any solution of the equation

$$D^2(f) + \mu f = 0$$

has the form $f = a\cos\sqrt{\mu}x + b\sin\sqrt{\mu}x$ for some $a, b \in \mathbb{R}$. (Note that although $D^2(x) = 0$, $x \notin \mathcal{P}$.)

Thus, D^2 is a self adjoint operator on $\mathcal{P}$ such that every non positive real number λ is an eigenvalue. The corresponding eigenspaces are $E_0 = \mathbb{R}$ and $E_{-\lambda} = \mathbb{R}\cos\sqrt{\lambda}x + \mathbb{R}\sin\sqrt{\lambda}x$ if $\lambda > 0$. There are some other consequences too. For example, if $\lambda_1, \ldots \lambda_k > 0$ and $f_i \in E_{-\lambda_i^2}$, then $f_1, \ldots, f_k$ are linearly independent (in fact, orthogonal): it cannot be spanned by finitely many functions. Therefore, as mentioned above, $\mathcal{P}$ is infinite dimensional. Furthermore, by Exercise 11.15, distinct eigenvalues of a self adjoint linear operator have orthogonal eigenspaces. Hence if f_λ and f_μ are eigenfunctions for $-\lambda^2 \neq -\mu^2$, then

$$\int_0^{2\pi} f_\lambda(t) f_\mu(t) \mathrm{d}t = 0,$$

where f_λ and f_μ are any eigenfunctions for $-\lambda$ and $-\mu$ respectively. In particular, we can conclude

$$\int_0^{2\pi} \sin\sqrt{\lambda}t \sin\sqrt{\mu}t \mathrm{d}t = 0,$$

with corresponding identities for the other pairs of eigenfunctions f_λ and f_μ. In addition, $\cos\sqrt{\lambda}x$ and $\sin\sqrt{\lambda}x$ are also orthogonal.

After normalization, we also obtain an orthonormal set

$$\frac{1}{\sqrt{2\pi}}, \quad \frac{1}{\sqrt{\pi}}\cos\sqrt{\lambda}x, \quad \frac{1}{\pi}\sin\sqrt{\lambda}x,$$

in $\mathcal{P}$. In Fourier analysis, one considers the eigenfunctions $\cos\sqrt{\lambda}x$ and $\sin\sqrt{\lambda}x$, where λ is a positive integer. The *Fourier series* of a function $f \in C[0, 2\pi]$ such that $f(0) = f(2\pi)$ is the infinite series

$$\frac{1}{\pi}\sum_{m=1}^{\infty} a_m \cos mx + \frac{1}{\pi}\sum_{m=1}^{\infty} b_m \sin mx, \tag{11.5}$$

where a_m and b_m are the Fourier coefficients encountered in §33. In particular,

$$a_m = \frac{1}{\sqrt{\pi}}\int_0^{2\pi} f(t)\cos mt \mathrm{d}t$$

and

$$b_m = \frac{1}{\sqrt{\pi}} \int_0^{2\pi} f(t) \sin mt \mathrm{d}t.$$

The basic question is in what sense (11.5) represents f. For this, we refer to a text on Fourier series. Note that the sum F of a finite number of terms of (11.5) give the projection of f onto a corresponding finite dimensional subspace of $\mathcal{P}$. This amounts to minimizing the integral $\int_0^{2\pi}(f - F)^2 \mathrm{d}t$, where F is varies through the corresponding subspace.

Exercises

Exercise 11.44. Show that if V is a finite dimensional inner product space, then $T \in L(V)$ is self adjoint if and only if for every orthonormal basis $\mathbf{u}_1, \ldots, \mathbf{u}_n$ OF V, $(T(\mathbf{u}_i), \mathbf{u}_j) = (\mathbf{u}_i, T(\mathbf{u}_j))$ for all indices i and j.

Exercise 11.45. Let U and V be inner product spaces of the same dimension. Show that a linear transformation $\Phi : U \to V$ is an isometry if and only if Φ carries some orthonormal basis of U onto an orthonormal basis of V.

Exercise 11.46. Suppose $\Phi : \mathbb{R}^2 \to \mathbb{R}^2$ is an arbitrary isomorphism. Show there exist a pair of inner products on $\mathbb{R}^2$ such that Φ is an isometry.

Exercise 11.47. 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 11.48. Let U be an inner product space. Show that the matrix of an isometry $\Phi : U \to U$ with respect to an orthonormal basis is orthogonal. Conversely show that if $\mathcal{B}$ is an orthonormal basis of U, then every orthogonal matrix Q defines an isometry Φ on U such that $Q = \mathcal{M}_{\beta}^{\beta}(\Phi)$.

Exercise 11.49. Give the proof of Propositions 11.6 and 11.7.

Exercise 11.50. Show how to reduce the proof of the Principal Axis Theorem (Theorem 11.9) to the case of a symmetric $n \times n$ matrix by choosing an isometry $\Phi : V \to \mathbb{R}^n$ ($n = \dim V$).

Exercise 11.51. Let $V = \mathbb{R}^{n \times n}$ and recall the transpose transformation $\mathbb{T} : V \to V$ defined by $\mathbb{T}(A) = A^T$ (see Exercise 9.35). Show that there exists an inner product on V for which $\mathbb{T}$ is self adjoint.

Exercise 11.52. Using the notation of Exercise 11.51, find an orthonormal basis $\mathcal{B}$ of V such that the matrix of $\mathbb{T}$ with respect to $\mathcal{B}$ is symmetric.

Exercise 11.53. Let V be a finite dimensional inner product space with inner product $(\ ,\ )$, 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 transformation.

(ii) If $V = \mathbb{R}^n$, find the matrix of T^*.

Exercise 11.54. Find the first six coefficients a_0, a_1, a_2 and b_0, b_1, b_2 in the Fourier expansion of $\sin^2 x$. Explain the significance of the approximation of $\sin^2 x$ these coefficients give.

Exercise 11.55. Show that the functions e^{imx}, $m = 0, \pm 1, \pm 2, \ldots$ form an orthogonal set. What is the associated orthonormal set?

11.4 Summary

The object of this chapter is to classify the unitarily diagonalizable complex matrices. First of all, we show that every real and complex square matrix is similar to an upper triangular matrix via a unitary matrix. In particular, if V is a finite dimensional Hermitian inner product space, then any linear transformation $T : V \to V$ admits a Hermitianorthonormal basis for which T's matrix is upper triangular. This is used to give a simple proof of the Normal Matrix Theorem, which says every normal matrix is unitarily diagonalizable, where a matrix N is said to be normal if and only if $N^H N = NN^H$. A matrix $A \in \mathbb{C}^{n\times n}$ is Hermitian if and only if $A^H = A$, so we immediately conclude the Principal Axis Theorem, which tells us that every Hermitian matrix is unitarily diagonalizable, and, moreover, any real Hermitian matrix, that is any symmetric matrix, is orthogonally diagonalizable. Among the other classes of matrices which are normal are unitary matrices, orthogonal matrices, skew symmetric and skew Hermitian matrices.

The geometric interpretation of a Hermitian matrix is a self adjoint operator. We give a geometric proof of the Principal Axis Theorem using this interpretation. We also give an example of a self adjoint operator in an infinite dimensional setting.

Chapter 12

Some Applications of Eigentheory

The purpose of this Chapter is to present an assortment of results and applications of eigentheory. For real symmetric matrices, we will study the notion of a quadratic form, prove Sylvester's Theorem and classify positive definite matrices. Quadratic forms arise in many mathematical contexts, but the one most students encounter first is in answering the question of whether a critical point of a function of several variables gives a maximum or minimum. By a local study of the behavior of a function all of whose partial derivatives vanish at a point $\mathbf{p}$ (i.e. $\mathbf{p}$ is a critical point), we arrive at the second derivative test, which rests on the critierion for a real symmetric matrix to be positive definite. In the second section, we give a brief introduction to graph theory and show how a symmetric matrix is associated to a graph. In the third section, we study the QR-algorithm and the power method, which two well known methods for approximating eigenvalues. We also discuss why the QR-algorithm converges.

12.1 Quadratic Forms

12.1.1 The Definition

Polynomials in several variables are the most important class of functions in algebra. Of course, linear functions are encountered throughout linear algebra. They make up the most basic class of of polynomials. The next most important class is the polynomials in which every term has degree two. Such a polynomial is called a *quadratic form*. An arbitrary quadratic form

over a field $\mathbb{F}$ is a polynomial of the form

$$q(x_1, \ldots, x_n) = \sum_{i,j=1}^{n} h_{ij} x_i x_j, \tag{12.1}$$

where each $h_{ij} \in \mathbb{F}$. Notice that since $x_i x_j = x_j x_i$, putting

$$q_{ij} = 1/2(h_{ij} + h_{ji})$$

gives another expression for q where the coefficients are symmetric in the sense that $q_{ij} = q_{ji}$ for all indices i, j.

The following Proposition points to the connection with symmetric matrices.

Proposition 12.1. *For every quadratic form q over $\mathbb{F}$, there exists a unique symmetric matrix $A \in \mathbb{F}^{n\times n}$ such that*

$$q(\mathbf{x}) = \mathbf{x}^T A \mathbf{x},$$

for all $\mathbf{x} \in \mathbb{F}^n$. Conversely, every symmetric matrix $A \in \mathbb{F}^{n\times n}$ defines a unique quadratic form over $\mathbb{F}$ by $q(\mathbf{x}) = \mathbf{x}^T A \mathbf{x}$.

Proof. This is left as an exercise. □

12.1.2 Critical Point Theory

Suppose $f(x_1, \ldots, x_n)$ is a smooth real valued function which has a critical point at $\mathbf{r} \in \mathbb{R}^n$. That is,

$$\frac{\partial f}{\partial x_i}(\mathbf{r}) = 0$$

for all i. The *Hessian* H_f at $\mathbf{r}$ is defined to be the symmetric $n \times n$ matrix

$$H_f = \frac{1}{2}\big(\frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{r})\big).$$

One of the basic results in critical point theory says that if $\mathbf{x}^T H_f \mathbf{x} > 0$ for all nonzero $\mathbf{x} \in \mathbb{R}^n$, then f has a local minimum at $\mathbf{r}$. The contribution of linear algebra is to provide the theorem which classifies when a general quadratic form q satisfies $q(\mathbf{x}) > 0$ if $\mathbf{x} \neq \mathbf{0}$ (see Proposition 12.6). In critical point theory, this result is called the second derivative test. The following example is a preview of the general result.

Example 12.1. Consider the function $f(x,y) = ax^2 + 2bxy + cy^2$, where $a, b, c \in \mathbb{R}$. The origin is a critical point, and H_f at $(0,0)$ is

$$H_f = \begin{pmatrix} a & b \\ b & c \end{pmatrix}.$$

Note that f is just the quadratic form associated to H_f. The second derivative test says that if $\det(H_f) = ac - b^2 > 0$, then f has a local minimum at $(0,0)$ if $a > 0$ and a local maximum if $a < 0$. Moreover, if $ac - b^2 < 0$, $(0,0)$ gives neither a local max or min. One can easily see what these conditions mean in terms of the eigenvalues of H_f. In fact, $f(x,y) > 0$ for all $(x,y) \neq (0,0)$ if and only if both eigenvalues of H_f are positive. Similarly, $f(x,y) < 0$ for all $(x,y) \neq (0,0)$ if and only if both eigenvalues of H_f are negative. If $ac - b^2 < 0$, then one eigenvalue is positive and the other is negative, and neither inequality holds for all $(x,y) \neq (0,0)$. These claims follow easily by writing $H_f = QDQ^T$, where Q is orthogonal, and putting $(u,v) = (x,y)Q$. Then

$$\begin{pmatrix} x & y \end{pmatrix} H_f \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} u & v \end{pmatrix} \begin{pmatrix} \lambda & 0 \\ 0 & \mu \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} = \lambda u^2 + \mu v^2,$$

where λ, μ are the eigenvalues. Hence the signs of the eigenvalues of H_f determine the behavior of $f(x,y)$ near the origin. For example, the conditions $a > 0$ and $ac - b^2 > 0$ are equivalent to $Tr(H_f) > 0$ and $\det(H_f) > 0$, which are, in turn, equivalent to $\lambda + \mu > 0$ and $\lambda\mu > 0$, i.e. $\lambda,\ \mu > 0$. The condition $\det(H_f) < 0$ tells us that $\lambda\mu < 0$, so the eigenvalues have different signs, and hence $f(x,y)$ assumes both positive and negative values near $(0,0)$.

We remark that the above example only treats the algebraic aspect of the second derivative test. The harder part is establishing the claim that the eigenvalues of H_f, if nonzero, determine the nature of the critical point.

Example 12.2. Consider the quadratic form $q(x,y) = x^2 + xy + y^2$. Its associated symmetric matrix is

$$A = \begin{pmatrix} 1 & 1/2 \\ 1/2 & 1 \end{pmatrix},$$

and $q(x,y) = (x\ y)A(x\ y)^T$. The trace of A is two and its determinant is $3/4$. Hence both eigenvalues are positive. They are in fact $3/2$ and $1/2$. Thus A can be expressed as QDQ^T, where $Q = \frac{1}{\sqrt{2}} \left(\begin{smallmatrix} 1 & -1 \\ 1 & 1 \end{smallmatrix}\right)$ and $D = \mathrm{diag}(3/2, 1/2)$. Putting $(u\ v) = (x\ y)Q$, gives $q(x,y) = 3/2u^2 + 1/2v^2$, so $q(x,y)$ can be expressed as the sum of positive squares.

Example 12.3. The above remarks are useful for determining the nature of a curve $ax^2 + 2bxy + cy^2 = d$ in $\mathbb{R}^2$. In the previous example,

$$x^2 + xy + y^2 = 3/2u^2 + 1/2v^2,$$

where x, y, u and v are related by

$$(u\ v) = (x\ y)Q,$$

where Q is orthogonal. Thus $\begin{pmatrix} x \\ y \end{pmatrix} = Q \begin{pmatrix} u \\ v \end{pmatrix}$. Now if we consider the orthonormal basis of $\mathbb{R}^2$ defined by the columns $\mathbf{q}_1$ and $\mathbf{q}_2$ of Q, this tells us

$$\begin{pmatrix} x \\ y \end{pmatrix} = u\mathbf{q}_1 + v\mathbf{q}_2.$$

In other words, u and v are the coordinates of $\begin{pmatrix} x \\ y \end{pmatrix}$ with respect to the orthonormal basis $\mathbf{q}_1$ and $\mathbf{q}_2$. Moreover, since $\det(Q) = 1$, $\mathbf{q}_1$ and $\mathbf{q}_2$ are obtained by rotating $\mathbf{e}_1$ and $\mathbf{e}_2$. Letting $< u, v >$ be coordinates on $\mathbb{R}^2$ (with axes along $\mathbf{q}_1$ and $\mathbf{q}_2$), we see that the curve is the ellipse $3/2u^2 + 1/2v^2 = 1$.

Let us summarize the above discussion and examples.

Proposition 12.2. *Every real quadratic form $q(x_1, \ldots, x_n)$ can be written as a sum of squares*

$$q(x_1, \ldots, x_n) = \sum_{i=1}^{n} \lambda_i r_i^2,$$

where $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of the matrix associated to q. The coordinates $r_1, \ldots, r_n$ are obtained from the orthogonal change of variables $(r_1\ \cdots\ r_n) = (x_1\ \cdots\ x_n)Q$, i.e. $\mathbf{r} = Q^T\mathbf{x}$.

12.1.3 Sylvester's Theorem

Two symmetric matrices A and B are said to be *congruent* if there exists an invertible $C \in \mathbb{R}^{n\times n}$ such that $B = CAC^T$. Congruence is an equivalence relation, and it is easy to see that if two symmetric matrices are congruent, they the same quadratic form with respect to different bases. Thus, congruence is analogous to similarity in the context of linear transformations. If C is orthogonal, then of course A and B have the same eigenvalues, but in general congruent matrices have different characteristic polynomials. The purpose of this section is to prove Sylvester's Theorem, which describes a

common property of the eigenvalues of an arbitrary pair of congruent matrices.

Before stating the theorem, suppose A is orthogonally diagonalized as $A = QDQ^T$. Let π and ν denote the number of positive and negative eigenvalues of A respectively. Since congruent matrices have the same rank, it follows that they have the same $\pi + \nu$. What isn't obvious is that they have the same π and the same ν. One way of saying this is to define the *index* of A by putting $\text{ind}(A) = \pi - \nu$. Sylvester's Theorem is the following:

Theorem 12.3. *Congruent matrices have the same index. In particular, congruent matrices have the same number of positive and negative eigenvalues.*

Proof. Let A and $B = CAC^T$ be two congruent matrices. To show $\text{ind}(A) = \text{ind}(B)$, it suffices to show they have the same π, since $\pi + \nu$ is the same for both. Choose an orthogonal matrix Q such that $A = QDQ^T$, where D is diagonal. Now CAC^T is still symmetric, so we may also write $CAC^T = PEP^T$, where P is orthogonal and E is diagonal. This gives

$$CAC^T = CQDQ^TC^T = PEP^T,$$

so $E = MDM^T$, where $M = P^TCQ$. Since A and D have the same π, as do B and E, it suffices to show that the diagonal matrices D and E have the same π.

Suppose E has more positive entries than D. For convenience, assume that the first r diagonal entries of $e_1, \ldots, e_r$ of E are positive and the rest are negative or zero. Suppose D has s positive entries, and let $f_i(\mathbf{x})$ denote the ith component of $M^T\mathbf{x}$. Then, since $\mathbf{x}^TE\mathbf{x} = \mathbf{x}^TMDM^T\mathbf{x}$,

$$\sum_{i=1}^{r} e_i x_i^2 = \sum_{j=1}^{n} d_j f_j(x_1, \ldots, x_r, 0, \ldots, 0)^2. \tag{12.2}$$

As $s < r$, there exists a nontrivial $\mathbf{x} \in \mathbb{R}^n$ of the form $(x_1, \ldots, x_r, 0, \ldots, 0)^T$ such that $f_j(x_1, \ldots, x_r, 0, \ldots, 0) = 0$ for each j for which $d_j > 0$. But, for this choice of $\mathbf{x}$, the left hand side of (12.2) is positive, while the right side is nonpositive. This is impossible, hence, $r \leq s$. By symmetry, $r = s$, so the proof is finished. □

Sylvester's Theorem has the following Corollary.

Corollary 12.4. *Suppose $A \in \mathbb{R}^{n\times n}$ is an invertible symmetric matrix with LDU decomposition $A = LDL^T$. Then the number of positive eigenvalues of A is the number of positive entries in D, and similarly for the number of*

negative eigenvalues. Thus the signs of the eigenvalues of A are the same as the signs of the pivots of A.

Proof. This follows immediately from Sylvester's Theorem. □

Notice that this Corollary applies to even if A has a zero eigenvalue. For, we can replace A by $A + rI_n$ for any r for which $A + rI_n$ is invertible.

Consider an example.

Example 12.4. Let $q(x, y, z) = x^2 + 2xy + 4xz + 2y^2 + 6yz + 2z^2$. The symmetric matrix associated to q is

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

A routine calculation gives

$$L^*A = DU = \begin{pmatrix} 1 & 1 & 2 \\ 0 & 1 & 1 \\ 0 & 0 & -3 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -3 \end{pmatrix} \begin{pmatrix} 1 & 1 & 2 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix},$$

so we see immediately that the pivots of A are 1, 1, -3. Thus A has two positive and one negative eigenvalue. The surface

$$x^2 + 2xy + 4xz + 2y^2 + 6yz + 2z^2 = 1$$

is thus a hyperboloid of one sheet.

12.1.4 Positive Definite Matrices

The examples dealing with critical points are a motivation for asking when a quadratic form takes only positive values away from $\mathbf{0}$. Let's make a basic definition.

Definition 12.1. Suppose $A \in \mathbb{R}^{n\times n}$ is symmetric and let $q(\mathbf{x}) = \mathbf{x}A\mathbf{x}^T$ be its associated quadratic form. Then we say A is *positive definite* if and only if $q(\mathbf{x}) > 0$ whenever $\mathbf{x} \neq \mathbf{0}$. Similarly, we say q is *negative definite* if and only if $q(\mathbf{x}) < 0$ whenever $\mathbf{x} \neq \mathbf{0}$. Otherwise, we say that q is *indefinite*.

Clearly the quadratic form q has a minimum at the origin if A is positive definite, and a maximum if A is negative definite. As the above examples show, we have

Proposition 12.5. *Let $A \in \mathbb{R}^{n\times n}$ be symmetric. Then the associated quadratic form $q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$ is positive definite if and only if all the eigenvalues of A are positive and negative definite if and only if all the eigenvalues of A are negative.*

Proof. This is an exercise. □

To obtain a satisfactory test to determine whether A is positive definite, we return to the LDU decomposition. One thing we would like to know is when a symmetric matrix can be written LDU: that is, $LPDU$ isn't necessary.

12.1.5 A Test For Positivity

Let $A \in \mathbb{R}^{n\times n}$ be symmetric and invertible, and suppose $A = LPDU$, where $U = L^T$, P is a permutation matrix and PD is symmetric (cf. Exercise 3.60). The trick is to notice that since L and U are respectively lower and upper triangular, the matrices in the upper left corner of A can be expressed in a very nice way. Let A_m denote the $m \times m$ block in A's upper left hand corner. Then $A_m = L_m(PD)_m U_m$. (This isn't hard to show, but it's a little messy, so we will leave it as an exercise.) Since $\det(L_m) = \det(U_m) = 1$ for all m, we have that $\det(A_m) = \det((PD)_m)$. We can now establish the following test.

Proposition 12.6. *A real symmetric $n \times n$ matrix A has an LDU decomposition if and only if $\det(A_m) \neq 0$ for all m. Moreover, if A has an LDU decomposition, then the i pivot d_i is given by the formula*

$$d_i = \frac{\det(A_i)}{\det(A_{i-1})}, \tag{12.3}$$

where by definition, we put $\det(A_0) = 1$.

Proof. As usual, put $A = LPDU$, where P is a permutation matrix. As noted above, if $m \geq 1$, then $A_m = L_m(PD)_m U_m$. Hence,

$$\det(A_m) = \det(L_m)\det((PD)_m)\det(U_m) = \det((PD)_m),$$

since L_m and U_m are unipotent. But, as D is diagonal, $(PD)_m = P_m D_m$. Notice that P_m could in fact be the zero matrix, but since $\det(A_m) \neq 0$, $\det(P_m) \neq 0$ for all m. The only way this can happen is that $P = I_n$. For $\det(P_1) = p_{11} \neq 0$, so $\det(P_2) = p_{11}p_{22}$ which is nonzero, hence $p_{22} \neq 0$, and so on. Hence if every $\det(A_m) \neq 0$, A has an LDU decomposition. The

converse is obvious, so we are done with the first claim. We will leave the proof of (12.3) as an exercise. □

We now show

Proposition 12.7. *A real symmetric $n \times n$ matrix A is positive definite if and only if* $\det(A_m) > 0$ *for all m.*

Proof. The only if statement follows immediately from Proposition 12.6. To prove the if statement, suppose $\det(A_m) > 0$ for all m. Then A has an LDU decomposition, and its pivots are all positive by (12.3). Hence A is positive definite. □

Similarly, one can also show

Proposition 12.8. *A symmetric $A \in \mathbb{R}^{n\times n}$ is negative definite if and only if $(-1)^m det(A_m) > 0$ for all $m \leq n$.*

Proof. The proof is yet another an exercise. □

Here is another example.

Example 12.5. Consider the matrix

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

One finds that $\det(A_1) = \det(A_2) = \det(A_3) = 1$ and $\det(A_4) = -1$. Hence A has an LDU decomposition with pivots 1, 1, 1, -1. Thus A has three positive and one negative eigenvalue.

Exercises

Exercise 12.1. Show that if $A = \begin{pmatrix} a & b \\ b & c \end{pmatrix}$ satisfies $a > 0$ and $ac - b^2 > 0$, then both eigenvalues of A are positive. In other words, justify the second derivative test.

Exercise 12.2. Decide whether $g(x, y, z) = x^2 + 6xy + 2xz + 3y^2 - xz + z^2$ has a max, min or neither at $(0, 0, 0)$.

Exercise 12.3. Formulate a second derivative test for functions of three variables.

Exercise 12.4. Suppose A is a symmetric matrix such that $\det(A) \neq 0$ and A has both positive and negative diagonal entries. Explain why A must be indefinite.

Exercise 12.5. Show that if A is a positive definite 3×3 symmetric 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 12.6. 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 define. (Your answer could be a diagonal matrix.)

Exercise 12.7. For the following pairs A, B of symmetric matrices, determine whether A and B are congruent or not.

(i) A and B have the same characteristic polynomial and distinct eigenvalues.

(ii) $\det(A) < 0$, $\det(B) > 0$.

(iii) $A = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$ and $B = \begin{pmatrix} 1 & 2 \\ 2 & -1 \end{pmatrix}$.

Exercise 12.8. Show that if $A \in \mathbb{R}^{n \times n}$ is positive definite, then every diagonal entry of A is positive. Also show that rA is positive definite if $r > 0$ and negative definite if $r < 0$.

Exercise 12.9. 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 SAS^T positive definite?

Exercise 12.10. Prove (12.3). That is, show that if A has an LDU decomposition with nonzero pivots d_i, then $d_i = \dfrac{\det(A_i)}{\det(A_{i-1})}$ for all indices i. (Note: by definition, $\det(A_0) = 1$.)

Exercise 12.11. Prove Proposition 12.8.

Exercise 12.12. Prove the converse statement in Proposition 12.6. That is, show that if A is positive definite, then $\det(A_m) > 0$ for all $m \leq n$.

Exercise 12.13. Describe the surface $(x\ y\ z)A(x\ y\ z)^T = 1$ for the following choices of A:

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

Exercise 12.14. Suppose $A_i = 0$ for some $i < n$. Does this mean A has a zero eigenvalue?

Exercise 12.15. Show that if A is positive definite or negative definite, then A has an LDU decomposition.

Exercise 12.16. When is e^A positive definite? Can e^A ever be negative definite or indefinite?

Exercise 12.17. A symmetric real matrix A is called *positive semi-definite* if its quadratic form q satisfies $q(\mathbf{x}) \geq 0$ for all $\mathbf{x} \in \mathbb{R}^n$. Prove that A is positive semi-definite if and only if every eigenvalue of A is non-negative.

12.2 Symmetric Matrices and Graph Theory

The purpose of this section is to give a very brief introduction to the subject of graph theory and to show how symmetric matrices play a fundamental role. A *graph* is a structure that consists of a finite set of vertices $v_1, v_2, \ldots, v_n$ and a finite set of bonds or edges $e_1, e_2, \ldots, e_N$. An edge e_i always joins two distinct vertices v_r and v_s say, and we will also denote such an edge by $v_r v_s$. We will always assume that two vertices are joined by at most one edge, but two distinct vertices need not be joined by any edge.

Graphs arise in all sorts of situations. For example, one version of the travelling salesman problem poses the following question: suppose a travelling salesman has to visit n cities any one of which is conected to any other city by a road. Assuming the cost of driving between any two cities is the same, find the least expensive route. Note that the edges may intersect, but the driver has to stay on the edge she started on. For example, suppose there are four cities located at the corners of a square. The shortest path through all four cities is the path around the edge avoiding the diagonals.

A problem which goes back to the 18th century is the question of whether there exists a path which allows one to cross all seven bridges over the Prugel River in the city of Königsberg without ever crossing the same bridge twice. It was shown to be impossible by L. Euler in 1736. For more information and examples, I suggest consulting *Introduction to Graph Theory* by B. Bollobás.

12.2.1 The Adjacency Matrix and Regular Graphs

Every graph has an associated symmetric matrix with 0,1 entries called the *adjacency matrix* of the graph. If the graph is Γ and its vertices are $v_1, v_2, \ldots, v_n$, then the adjacency matrix A_Γ is the $n \times n$ matrix a_{ij}, where a_{ij} is 1 if there exists an edge joining v_i and v_j and 0 if not. It's clear from the definition that $a_{ij} = a_{ji}$, so A is symmetric as claimed. Moreover, the graph can be reconstructed from its adjacency matrix.

Here are some examples of graphs and their matrices.

Example 12.6. Let Γ_1 be the graph with vertices v_1, v_2, v_3 and edges v_1v_2, v_1v_3 and v_2v_3 . Then

$$A_{\Gamma_1} = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}.$$

If Γ_2 has vertices $v_1, \ldots, v_4$ and edges v_1v_2, v_1v_3, v_1v_4, and v_2v_4, then

$$A_{\Gamma_2} = \begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \end{pmatrix}.$$

Clearly, the number of 1's in the ith row is the number of edges containing v_i. The number of edges at a vertex v_i is denoted by $d(v_i)$ and called the *degree* of v_i. Clearly, $d(v_i)$ is the sum of the entries in the ith row of A_Γ.

Many graphs have the property that any two vertices have the same degree. That is, $d(v_i) = d(v_j)$ for all i, j. These graphs are called *regular*. More particularly, a graph is called *k-regular* if any every vertex has degree k. To test k-regularity, it is convenient to single out the vector $\mathbf{1_n} \in \mathbb{R}^{\mathbf{n}}$ all of whose components are 1. The next Proposition characterizes the k-regular graphs.

Proposition 12.9. *A graph Γ with n vertices is k-regular if and only if $(k, \mathbf{1_n})$ is an eigenpair for A_Γ.*

Proof. A graph Γ is k-regular if and only if every row of A_Γ has exactly k 1's, i.e. each row sum is k. This is the same as saying $(k, \mathbf{1_n})$ is an eigenpair. □

This result can be improved in the following way. A graph Γ is said to be *connected* if any two of its vertices v, v' can be joined by a *path* in Γ, a path being a sequence of edges x_0x_1, x_1x_2, …, $x_{s-1}x_s$, where $x_0, \ldots, x_s$ are not necessarily distinct vertices. Let $\Delta(\Gamma)$ denote the largest value of $d(v_i)$, and let $\delta(\Gamma)$ denote the smallest value of $d(v_i)$. Then we have

Theorem 12.10. *Let Γ be a connected graph with adjacency matrix A_Γ. Then we have the following.*

(*i*) *Every eigenvalue λ of A_Γ satisfies $|\lambda| \leq \Delta(\Gamma)$.*

(*ii*) *The largest eigenvalue λ_M satisfies $\delta(\Gamma) \leq \lambda_M \leq \Delta(\Gamma)$.*

(*iii*) *Γ is $\Delta(\Gamma)$-regular if and only if $\lambda_M = \Delta(\Gamma)$.*

(*iv*) *Finally, if Γ is regular, then multiplicity of $\lambda_M = \Delta(\Gamma)$ as an eigenvalue is 1.*

Proof. Let λ be an eigenvalue and choose an eigenvector $\mathbf{u}$ for λ with the property the $|u_j| \leq 1$ for each component while $u_s = 1$ for some component u_s. Then

$$|\lambda| = |\lambda u_s| = |\sum_j a_{sj}u_j| \leq \sum_j a_{sj}|u_j| \leq \sum_j a_{sj} \leq \Delta(\Gamma).$$

This proves (i). To show (ii), it suffices to show $\lambda_M \geq \delta(\Gamma)$. Recall the projection formula for symmetric matrices (11.3). Namely, for all $\mathbf{u} \in \mathbb{R}^n$,

$$A\mathbf{u} = \lambda_1(\mathbf{u}_1^T\mathbf{u})\mathbf{u}_1 + \cdots + \lambda_n(\mathbf{u}_n^T\mathbf{u})\mathbf{u}_n,$$

where $\mathbf{u}_1, \ldots, \mathbf{u}_n$ is an orthonormal eigenbasis for A_Γ. Thus,

$$\mathbf{u}^T A\mathbf{u} \leq \lambda_M \mathbf{u}^T\mathbf{u}.$$

In particular, $\mathbf{1}_n^T A_\Gamma \mathbf{1}_n \leq n\lambda_M$. On the other hand,

$$\mathbf{1}_n^T A_\Gamma \mathbf{1}_n = \sum_{i,j} a_{ij} \geq n\delta(\Gamma).$$

Hence $n\lambda_M \geq n\delta(\Gamma)$, which proves (ii).

We now prove (iii). (This is the only place where the hypothesis that Γ is connected is used.) The claim that if Γ is $\Delta(\Gamma)$-regular, then $\lambda_M = \Delta(\Gamma)$ is obvious. Suppose that $\lambda_M = \Delta(\Gamma)$. Using the eigenvector $\mathbf{u}$ chosen in the first paragraph of the proof, we have

$$\Delta(\Gamma) = \Delta(\Gamma)u_s = \sum_j a_{sj}u_j \leq \sum_j a_{sj}|u_j| \leq \sum_j a_{sj} \leq \Delta(\Gamma).$$

Hence for every j such that $a_{sj} \neq 0$, we have $u_j = 1$. Since every vertex can be joined to v_s by a path, it follows that $\mathbf{u} = \mathbf{1}_n$. But this implies Γ is $\Delta(\Gamma)$-regular, giving (iii). It also follows that the multiplicity of $\Delta(\Gamma)$ as an eigenvalue is 1, which proves (iv). □

The adjacency matrix also answers the question of how many paths join two vertices v_i and v_j of Γ. Let us say that a path with r edges has length r.

Proposition 12.11. *The number of paths of length $r \geq 1$ between two not necessarily distinct vertices v_i and v_j of Γ is $(A_\Gamma^r)_{ij}$.*

Proof. This is just a matter of applying the definition of matrix multiplication. □

Example 12.7. Consider the connected graph with two vertices. Its adjacency matrix is $A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$. Now $A^m = I_2$ if m is even and $A^m = A$ if m is odd. Thus, as is easy to see directly, there is one path of any even length from each vertex to itself and one of any odd length from each vertex to the other.

12.3 Computing Eigenvalues

12.3.1 The QR-Algorithm

We will now discuss a method known as the QR-algorithm for approximating the eigenvalues of a nonsingular complex matrix $A \in \mathbb{C}^{n\times n}$. It is assumed here that the reader has a passing knowledge of sequences and limits.

Recall from Chapter 10 that A can be factored in the form $A = QR$, where Q is unitary and R is upper triangular with nonzero diagonals, i.e. $R \in GL(n, \mathbb{C})$. The QR-algorithm starts from the fact that QR and RQ are similar, and hence have the same eigenvalues. Putting $A = A_0$, and successively applying QR, we have

$$A_0 = Q_0R_0,$$

$$A_1 = Q_0^{-1}AQ_0 = R_0Q_0 = Q_1R_1,$$
$$A_2 = Q_1^{-1}Q_0^{-1}AQ_0Q_1 = R_1Q_1 = Q_2R_2,$$

and

$$A_3 = Q_2^{-1}Q_1^{-1}Q_0^{-1}AQ_0Q_1Q_2 = R_2Q_2 = Q_3R_3.$$

Continuing in this manner, we get a sequence of similar matrices

$$A_1 = Q_1R_1,\ A_2 = Q_2R_2,\ A_3 = Q_3R_3, \ldots,\ A_i = Q_iR_i, \ldots$$

where each Q_i is unitary and each R_i is upper triangular.

Of course it is entirely possible that the this process gives no new information at all. For example, if A itself is unitary, then $A_i = A$ for all i and each $R_i = I_n$. However, in some (but not all) cases, the sequences A_i, Q_i and R_i all converge. In particular, if limit $\mathbf{R} = \lim_{m\to\infty} R_m$ exists, then $\mathbf{R}$ is upper triangular. The point of the QR-algorithm, as we will see, is that in the eigenvalues of A are sometimes the diagonal entries of $\mathbf{R}$.

12.3.2 The QR-Convergence Theorem

To understand the QR-algorithm, we need to define what is meant by a convergent sequence of matrices.

Definition 12.2. A sequence $A_m = (a_{ij}^{(m)})$ of complex matrices is said to be *convergent* if and only if all the component sequences $a_{ij}^{(m)}$ converge in $\mathbb{C}$. Suppose $\lim_{m\to\infty} a_{ij}^{(m)} = x_{ij}$ for all i, j. Then we say that $\lim_{m\to\infty} A_m = X$, where $X = (x_{ij})$.

Proposition 12.12. *Let U_m be a sequence of unitary matrices such that $\lim_{m\to\infty} U_m$ exists, say $\lim_{m\to\infty} U_m = X \in \mathbb{C}^{n\times n}$. Then $X \in U(n,\mathbb{C})$.*

Proof. We will assume the fact that limits of sequences of matrices behave like limits of sequences of real or complex numbers. In particular, the product rule holds. (This isn't hard to verify, since the sum and product rules hold for sequences in $\mathbb{C}$.) Since $U_m^H U_m = I_n$ for all m, it follows that

$$I_n = \lim_{m\to\infty} U_m^H U_m = \lim_{m\to\infty} U_m^H \lim_{m\to\infty} U_m = X^H X.$$

Hence, X is unitary. □

Another pertinent fact is

Proposition 12.13. *Every sequence of $n \times n$ unitary matrices has a convergent subsequence.*

Proof. This follows from the fact that the components of a unitary matrix are bounded. In fact, since the columns of a unitary matrix U are unit vectors in $\mathbb{C}^n$, every component u_{ij} of U must satisfy $|u_{ij}| \leq 1$. Thus, every component sequence has a convergent subsequence, so every sequence of unitary matrices has a convergent subsequence. □

Now let us return to the QR-algorithm. Let $A = A_0,\ A_1,\ A_2, \ldots$ be the sequence of matrices similar to A defined above. The $(m+1)$st term of this sequence is

$$A_{m+1} = U_m^{-1} A U_m = Q_{m+1} R_{m+1}, \tag{12.4}$$

where

$$U_m = Q_0 Q_1 \cdots Q_m. \tag{12.5}$$

Hence, if $\lim_{m\to\infty} Q_m = I_n$, we have

$$\lim_{m\to\infty} U_m^{-1} A U_m = \mathbf{R}.$$

Note that all $U_m \in U(n,\mathbb{C})$, so a subsequence converges to some $X \in U(n,\mathbb{C})$. Assuming $\lim_{m\to\infty} Q_m = I_n$, we conclude that

$$X^{-1} A X = \mathbf{R}.$$

In other words, under the assumption that $\lim_{m\to\infty} Q_m = I_n$, we recover Schur's Theorem: there exists a unitary X such that $X^{-1}AX$ is upper triangular. In particular, the diagonal entries of R are the eigenvalues of the original matrix A.

We now state a general convergence theorem for the QR-algorithm. We will not prove it here. Notice that the hypotheses rule out the possibility that A is unitary.

Theorem 12.14. *Let $A \in \mathbb{C}^{n\times n}$ have the property that its eigenvalues λ_i satisfy the condition $|\lambda_i| \neq |\lambda_j|$ if $i \neq j$. Suppose U_m $(m \geq 1)$ is the sequence of unitary matrices defined in* (12.5)*. Then $\lim_{m\to\infty} U_m^{-1} A U_m$ exists and is upper triangular. Taking $X \in U(n, \mathbb{C})$ to be the limit of any subsequence of the sequence U_m, we obtain $X^{-1}AX = \mathbf{R}$.*

The hypotheses allow A to have a zero eigenvalue. But we can always replace A by $A - rI_n$, where r is chosen so that $A - rI_n$ is nonsingular. Employing such shifts is in fact one of the techniques used in numerical analysis to speed up the convergence.

12.3.3 The Power Method

There is a related method for approximating the eigenvalues of a complex matrix called the power method. Define a *flag* $\mathbf{F}$ in $\mathbb{C}^n$ to be a sequence $(F_1, F_2, \ldots, F_n)$ of subspaces of $\mathbb{C}^n$ such that $F_1 \subset F_2 \subset \cdots \subset F_n$ of $\mathbb{C}^n$ and $\dim F_i = i$ for each i. An ordered basis $(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n)$ of $\mathbb{C}^n$ is a *flag basis* for $\mathbf{F}$ if $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_i$ is a basis of F_i for each i. (Recall that a similar notion was defined in Chapter 11). Given $A = \begin{pmatrix}\mathbf{a}_1 & \cdots & \mathbf{a}_n\end{pmatrix} \in GL(n, \mathbb{C})$, the associated flag $\mathbf{F}(A)$ is the flag $(F_1, F_2, \ldots, F_n)$ such that F_i is the span of the first i columns of A. Since A is nonsingular, the condition $\dim F_i = i$ is automatically satisfied.

Using flag bases and column operations, it is not hard to see

Proposition 12.15. *Two matrices A and B in $GL(n, \mathbb{C})$ define the same flag if and only if there exists an upper triangular matrix $T \in GL(n, \mathbb{C})$ such that $A = BT$.*

Each $A \in GL(n, \mathbb{C})$ acts on a flag $\mathbf{F} = (F_1, F_2, \ldots, F_n)$ by

$$A\mathbf{F} = (A(F_1), A(F_2), \ldots, A(F_n)).$$

Thus elements of $GL(n, \mathbb{C})$ permute flags.

Definition 12.3. A flag $\mathbf{F}$ such that $A\mathbf{F} = \mathbf{F}$ is called a *fixed flag* or *eigenflag* for A.

Example 12.8. Let A be diagonal, say $A = \operatorname{diag}(\alpha_1, \alpha_2, \ldots \alpha_n)$, where $\alpha_i \neq \alpha_j$ if $i \neq j$ and all $\alpha_i \neq 0$, and let P be an $n \times n$ permutation matrix. Then the flag $\mathbf{F}(P)$ is fixed by A, and since the eigenvalues of A are distinct, the $\mathbf{F}(P)$, where P runs over all such permutation matrices, are in fact the only flags fixed by A. Hence A has exactly $n!$ fixed flags.

We will now introduce the notion of a limit of flags. Let $\mathbf{F}_m$ $(m \geq 1)$ be a sequence of flags. Then we say $\lim_{m\to\infty} \mathbf{F}_m = \mathbf{F}$ if for each m, there

exists a flag basis $\mathbf{v}_1^{(m)}, \mathbf{v}_2^{(m)}, \ldots, \mathbf{v}_n^{(m)}$ of $\mathbf{F}_m$ and a flag basis $\mathbf{v}_1, \ldots, \mathbf{v}_n$ of $\mathbf{F}$ such that for each index i, $\lim_{m\to\infty} \mathbf{v}_i^{(m)} = \mathbf{v}_i$.

There is an alternate way of formulating this definition using unitary matrices. Applying Gram-Schmidt to a flag basis gives an orthonormal flag basis (neat!), so every flag $\mathbf{F}$ can be written in the form $\mathbf{F}(U)$, where U is unitary. Given a sequence $\mathbf{F}_m$ of flags, $\lim_{m\to\infty} \mathbf{F}_m = \mathbf{F}$ if and only if for each $m \geq 0$, there exists a unitary matrix U_m such that:

(*i*) $\mathbf{F}(U_m) = \mathbf{F}_m$,

(*ii*) $\lim_{m\to\infty} U_m$ exists, and

(*iii*) $\mathbf{F}(U) = \mathbf{F}$, where $U = \lim_{m\to\infty} U_m$.

We can now explain the *power method*. Given $A \in GL(n, \mathbb{C})$, consider the sequence of flags $\mathbf{F}(A^m)$ $(m \geq 1)$. Suppose this sequence has a limit, say $\lim_{m\to\infty} \mathbf{F}(A^m) = \mathbf{F}$. Then $\mathbf{F}$ is a fixed flag for A. For,

$$\begin{aligned} A\mathbf{F} &= A \lim_{m\to\infty} \mathbf{F}(A^m) \\ &= \lim_{m\to\infty} A\mathbf{F}(A^m) \\ &= \lim_{m\to\infty} \mathbf{F}(A^{m+1}) \\ &= \mathbf{F}. \end{aligned}$$

Now, if $\mathbf{F} = \mathbf{F}(X)$, where X is unitary, then it follows that

$$A\mathbf{F} = A\mathbf{F}(X) = \mathbf{F}(AX) = \mathbf{F}(X).$$

Thus, we see again that there exists an upper triangular T such that $AX = XT$.

The power method suggests that to find the eigenvalues of A, compute A^m for a large value of m, and assume that the flag $\mathbf{F}(A^m)$ of this power of A is near a flag fixed by A. Thus if we apply Gram-Schmidt to get an orthonormal flag basis of $\mathbf{F}(A^m)$, this gives a unitary X such that $X^{-1}AX$ is (hypothetically) close to an upper triangular matrix T. The eigenvalues of A are therefore approximated by the diagonal of $X^{-1}AX$.

To connect this reasoning with the QR-algorithm, note that in the notation of Section 12.3.1,

$$A^{m+1} = (Q_0 Q_1 \cdots Q_{m-1} Q_m)(R_m R_{m-1} \cdots R_1 R_0). \tag{12.6}$$

(This is easily proved by induction with the help of the identities $R_i Q_i = Q_{i+1} R_{i+1}$ for $i \geq 0$.) But $T_m := R_m R_{m-1} \cdots R_1 R_0$ is upper triangular, $U_m = Q_0 Q_1 \cdots Q_{m-1} Q_m$ is unitary and

$$A^{m+1} = U_m T_m.$$

Therefore

$$\mathbf{F}(A^{m+1}) = \mathbf{F}(U_m T_m) = \mathbf{F}(U_m),$$

since T_m is upper triangular. Thus, we may assume that $\mathbf{F}(U_m) = \mathbf{F}(X)$.

Finally, we also note without proof the following.

Proposition 12.16. *Suppose all eigenvalues of $A \in GL(n, \mathbb{C})$ have different moduli. Then the flag sequence $\mathbf{F}(A^m)$ $(m \geq 1)$ converges, so the flag $\mathbf{F} = \lim_{m\to\infty} \mathbf{F}(A^m)$ is fixed by A. Moreover, the line F_1 is spanned by an eigenvector of A for the eigenvalue with the largest modulus.*

The above discussion of the power method is an illustration of what one might call a fixed point method. Every $A \in GL(n, \mathbb{C})$ has fixed flags, by Schur's Theorem; the problem of finding one is sometimes solved by the power method.

We end the section with a comment about the $LPDU$ decomposition (cf. Chapter 3). The set of all flags in $\mathbb{C}^n$ is often called the the *flag variety of* $\mathbb{C}^n$. Let us denote it by $\mathcal{F}(n)$. The $LPDU$ decomposition of a matrix $A \in GL(n, \mathbb{C})$ actually partitions $GL(n, \mathbb{C})$ into what are called "cells". The cells consist of all matrices $A = LPDU$ with the same permutation matrix P. (Recall, even though L and U are not in general uniquely determined by A, P and D are.) The subsets of the form $\mathbf{F}(LPDU) = \mathbf{F}(LP)$ in the flag variety $\mathcal{F}(n)$ are known as *Schubert cells*. They contain a great deal of interesting geometrical information about $\mathcal{F}(n)$ and are used in many ways to study its geometry.

12.4 Summary

The purpose of this chapter was to make some applications of eigentheory. The first application is real quadratic forms. A quadratic form is a function on $\mathbb{F}^n$ of the form $q(\mathbf{x}) = x^T A\mathbf{x}$, where A is a symmetric matrix. For example, the second order terms in the Taylor series of a smooth function $f(x_1, \ldots, x_n)$ is a real quadratic form. In the real case, the main question is whether or not $q(\mathbf{x}) > 0$ for all $\mathbf{x} \neq \mathbf{0}$. When the answer is yes, q is said to be positive definite. It turns out that q is positive definite if and only if all eigenvalues of A are positive. But a simpler criterion is that A is positive definite if and only if all its pivots are positive. Moreover, we gave a simple determinantal test for the positivity of the pivots.

We next considered defined the concept of a graph and introduced a certain real symmetric matrix called the adjacency matrix of the graph, and we discussed the significance of its largest eigenvalue.

The third application is the QR-algorithm. In fact, given $A \in \mathbb{C}^{n\times n}$, we showed how to construct a sequence of unitary matrices U_m such that $\lim_{m\to\infty} U_m A U_m^{-1}$ is an upper triangular matrix T. Moreover, we also showed that the sequence U_m can be assumed to converge to a unitary matrix X, so $XAX^{-1} = T$. This justifies the QR-algorithm and gives another way of approaching Schur's Theorem.

Chapter 13

The Jordan Decomposition Theorem

The purpose of this Chapter is to study some fundamental results about the linear transformations $T : V \to V$, where V is a finite dimensional vector space V over an algebraically closed field $\mathbb{F}$. Of course, the eigenvalues of a linear transformation are its main invariants, hence the restriction that $\mathbb{F}$ be algebraically closed. with linear transformations all of whose eigenvalues lie *a priori* in $\mathbb{F}$. The first result is the Jordan Decomposition Theorem, which says that every linear transformation $T : V \to V$, can be uniquely written as $T = S + N$, where S and N are linear transformations on V such that S is semi-simple (i.e. V admits an eigenbasis for T), N is nilpotent (i.e. $N^k = O$ for some positive integer k), and $SN = NS$. The linear transformations S and N are called the *semi-simple* and *nilpotent* parts of T respectively. Thus, T commutes with its semi-simple and nilpotent parts S and N. The expression $T = S+N$ is called the *Jordan decomposition* of T. It is very useful for understanding the structure of linear transformations, and moreover, it's also crucial in proving many of the deep results about the space $L(V)$ of all linear transformations $T : V \to V$.

Our proof of the Jordan Decomposition Theorem illustrates the importance of the Cayley-Hamilton Theorem, which we proved in Chapter 9, but only for diagonalizable matrices. This Theorem is in fact an immediate consequence of the proof of the Jordan Decomposition Theorem given below. Furthermore, the question of whether T is itself semi-simple, i.e. $N = O$, reduces to checking whether T satisfies an equation $q(T) = O$, where q is an explicit polynomial which divides the characteristic polynomial of T.

Finally, we explore the Jordan Canonical Form. Although we skip the

proof, as it is mainly combinatorial, we will state the result. Namely, every $A \in \mathbb{F}^{n\times n}$ is similar to a matrix of a certain type called Jordan Canonical Form. The number of such forms which are nilpotent corresponds to the number of expressions of n as a sum of non-negative integers written in decreasing order. Such decompositions are called partitions of n. The number of partitions of n is a famous function denoted as $\pi(n)$. Hence there are exactly $\pi(n)$ nilpotent conjugacy classes of $n \times n$ matrices.

13.1 The Main Result

Let $\mathbb{F}$ be an algebraically closed field (e.g. $\mathbb{C}$), and suppose V is a finite dimensional vector space over $\mathbb{F}$. We will now prove one of the most fundamental results on the structure of linear transformations $T : V \to V$. As mentioned above, T is semi-simple if V admits an eigenbasis and is nilpotent if $T^k = 0$ for some positive integer k. In particular, we will show T can be uniquely decomposed as $T = S + N$, where S is semi-simple, N is nilpotent and $SN = NS$. This is called the *Jordan decomposition* of T. S is called the *semi-simple part of* T and N is called the *nilpotent part of* T .

Theorem 13.1 (Jordan Decomposition Theorem)**.** *Let $\mathbb{F}$ be an arbitrary algebraically closed field, and consider a linear transformation $T : V \to V$, where V is a finite dimensional vector space over $\mathbb{F}$ of dimension n. Let $\lambda_1, \ldots, \lambda_m$ be the distinct eigenvalues of T, and suppose μ_i denotes the multiplicity of λ_i. Thus*

$$p_T(x) = (-1)^n(x - \lambda_1)^{\mu_1} \cdots (x - \lambda_m)^{\mu_m}. \tag{13.1}$$

Then there exist subspaces $C_1, \ldots, C_m$ of V with the following properties:

(1) *$T(C_i) \subset C_i$, $\dim C_i = \mu_i$ and V is the direct sum*

$$V = C_1 \oplus C_2 \oplus \cdots \oplus C_m. \tag{13.2}$$

(2) *Let $S : V \to V$ be the semi-simple linear transformation defined by the condition that $S(\mathbf{v}) = \lambda_i\mathbf{v}$ if $\mathbf{v} \in C_i$. Then, the linear transformation $N = T - S$ is nilpotent.*

(3) *S and N commute, hence both commute with T.*

(4) *The decomposition $T = S + N$ is the only decomposition of T into the sum of a semi-simple S and a nilpotent N such that $SN = NS$.*

Proof. First of all, notice that if $R : V \to V$ is linear, then $\ker(R^r) \subset \ker(R^s)$ if r and s are positive integers such that $r < s$. Since V is finite dimensional, it follows that for some $r > 0$, $\ker(R^r) = \ker(R^{r+1})$, and thus $\ker(R^r) = \ker(R^s)$ if $r \leq s$. In this case, we say that $\ker(R^r)$ is *stable*. Now define

$$C_i = \ker(R_i^{r_i}),$$

where $R_i = T - \lambda_i I_n$ and r_i is large enough so that $\ker(R_i^{r_i})$ is stable.

Lemma 13.2. *Let $\overline{U}_i = R_i^{r_i}$, and choose $k > 0$ such that $\ker(\overline{U}_1 \cdots \overline{U}_m)^k = \ker(\overline{U}_1 \cdots \overline{U}_m)^{k+1}$. Put $U_i = \overline{U}_i^k$. Then $U = U_1 \cdots U_m = O$, where O denotes the zero transformation.*

Proof. Let $W = \ker(U)$. To show that $W = V$, we will show $V/W = \{\mathbf{0}\}$. Obviously, $T(W) \subset W$, so, arguing as in the proof of Proposition 9.13, T induces a linear transformation $T' : V/W \to V/W$ such that $T'(\mathbf{v} + W) = T(\mathbf{v}) + W$. Suppose $W \neq V$, so that $\dim V/W > 0$. Since $\mathbb{F}$ is algebraically closed, it follows that T' has an eigenvalue μ in $\mathbb{F}$, hence an eigenvector $\mathbf{v} + W$ in V/W. By definition,

$$T'(\mathbf{v} + W) = \mu(\mathbf{v} + W) = (\mu\mathbf{v}) + W.$$

It follows that $\mathbf{x} = T(\mathbf{v}) - \mu\mathbf{v} \in W$. Clearly, $U(T - \mu I_n) = (T - \mu I_n)U$, so $(T - \mu I_n)(W) \subset W$.

Now suppose μ is not an eigenvalue of T. Then $(T - \mu I_n)$ is injective on V, so $(T - \mu I_n)^{-1}$ exists. Moreover, $(T - \mu I_n)(W) = W$, for dimensional reasons, since $(T - \mu I_n)(W) \subset W$. Therefore, $(T - \mu I_n)^{-1}(W) = W$ too. Consequently, since $\mathbf{x} \in W$,

$$(T - \mu I_n)^{-1}(\mathbf{x}) = (T - \mu I_n)^{-1}(T - \mu I_n)(\mathbf{v}) = \mathbf{v} \in W.$$

But this is a contradiction, since, by assumption, $\mathbf{v} + W \neq \mathbf{0} + W$ in V/W.

Hence, μ is an eigenvalue of T. Since $T(\mathbf{v}) - \mu\mathbf{v} \in W$, we deduce that $U^2(\mathbf{v}) = \mathbf{0}$. But, by definition, $W = \ker(U) = \ker(U^2)$, so, in fact, $\mathbf{v} \in W$. This is again a contradiction, so we conclude $V = W$. □

We will show next $\ker(U_i) \cap \ker(U_j) = \{\mathbf{0}\}$ if $i \neq j$. First, we show

Lemma 13.3. *If $i \neq j$, then $R_i(\ker(U_j)) \subset \ker(U_j)$. Moreover, R_i is one to one on $\ker(U_j)$.*

Proof. As $R_i R_j = R_j R_i$ for all i and j, it follows that $R_i(\ker(R_j^s)) \subset \ker(R_j^s)$ for any $s > 0$. Hence, $R_i(\ker(U_j)) \subset \ker(U_j)$. To see R_i is one to one on $\ker(U_j)$, it suffices to show $\ker(R_i) \cap \ker(U_j) = \{\mathbf{0}\}$. Let $\mathbf{v} \in \ker(R_i) \cap \ker(U_j)$.

By the kernel criterion for injectivity, it suffices to show $\mathbf{v} = \mathbf{0}$. Note first that $\ker(R_i) \cap \ker(R_j) = \{\mathbf{0}\}$, since different eigenspaces have trivial intersection. But since $R_i R_j = R_j R_i$, $R_j(\ker(R_i)) \subset \ker(R_i)$. Thus, R_j is a one to one linear transformation of $\ker(R_i)$ to itself. But the composition of any two injective injective linear transformations (if defined) is injective. Consequently, R_j^s is injective on $\ker(R_i)$ for any $s > 0$. Therefore, by the definition of U_j, $\ker(R_i) \cap \ker(U_j) = \{\mathbf{0}\}$. □

It follows immediately that U_i is one to one on $\ker(U_j)$, so $\ker(U_i) \cap \ker(U_j) = \{\mathbf{0}\}$ if $i \neq j$. We now come to the final Lemma.

Lemma 13.4. *Suppose $Q_1, \ldots, Q_m$ are linear transformations with domain and target V such that $Q_1 \cdots Q_m = O$. Suppose further that $Q_i Q_j = Q_j Q_i$ and $\ker(Q_i) \cap \ker(Q_j) = \{\mathbf{0}\}$ for all $i \neq j$. Then*

$$V = \bigoplus_{i=1}^{m} \ker(Q_i). \tag{13.3}$$

Proof. We will use induction on m, the case $m = 1$ being obvious. Hence, assume

$$\ker(Q_2 \cdots Q_m) = \bigoplus_{i=2}^{m} \ker(Q_i). \tag{13.4}$$

Suppose $P : V \to V$ and $Q : V \to V$ are linear transformations such that $PQ = O$ and $\ker(P) \cap \ker(Q) = \{\mathbf{0}\}$. Then I claim $V = \ker(P) \oplus \ker(Q)$. Since $\dim(\ker(P) \cap \ker(Q)) = 0$, the claim follows from Proposition 5.20 provided $\dim \ker(P) + \dim \ker(Q) \geq \dim V$. Now

$$\dim V = \dim \ker(Q) + \dim \operatorname{im}(Q).$$

But as $PQ = O$, $\operatorname{im}(Q) \subset \ker(P)$, so we indeed have the desired inequality.

To finish the proof, let $P = Q_1$ and $Q = Q_2 \cdots Q_m$. Then $PQ = O$, so, to apply the claim, we have to show $\ker(Q_1) \cap \ker(Q) = \{\mathbf{0}\}$. Suppose $\mathbf{v} \in \ker(Q_1) \cap \ker(Q)$. By (13.4), we can write $\mathbf{v} = \sum_{i=2}^{m} \mathbf{q}_i$, where $\mathbf{q}_i \in \ker(Q_i)$. Thus, $Q_1(\mathbf{v}) = \sum_{i=2}^{m} Q_1(\mathbf{q}_i) = \mathbf{0}$. Now $Q_1(\mathbf{q}_i) \in \ker(Q_i)$. For, as $Q_1 Q_i = Q_i Q_1$ for all i,

$$Q_i Q_1(\mathbf{q}_i) = Q_1 Q_i(\mathbf{q}_i) = Q_1(\mathbf{0}) = \mathbf{0}.$$

Therefore, by the induction hypothesis, $Q_1(\mathbf{q}_i) = \mathbf{0}$ for all $i > 1$. But $\ker(Q_1) \cap \ker(Q_i) = \{\mathbf{0}\}$ for $i > 1$, so each $\mathbf{q}_i = \mathbf{0}$. Hence $\mathbf{v} = \mathbf{0}$, so we conclude $V = \ker(Q_1) \oplus \ker(Q)$. Consequently, $\dim V = \dim \ker(Q_1) +$

$\dim\ker(Q)$. Furthermore, by the induction hypothesis, we conclude that $V = \sum_{i=1}^{m} \ker(Q_i)$.

To show this sum is direct, it suffices, by Proposition 5.22, to show that $\dim V = \sum_{i=1}^{m} \dim\ker(Q_i)$. But, by Proposition 5.22 and the induction hypothesis,

$$\begin{aligned} \dim V &= \dim\ker(Q_1) + \dim\ker(Q) \\ &= \ker(Q_1) + \sum_{i=2}^{m} \dim\ker(Q_i). \end{aligned}$$

Thus $V = \bigoplus_{i=1}^{m} \ker(Q_i)$, completing the proof. □

Since, by definition, $\ker(U_i) = C_i$, we have thus $V = \bigoplus_{i=1}^{m} C_i$, which proves (13.2). It is easy to see $T(C_i) \subset C_i$ for all i. To finish the proof of (1), we have to show that $\nu_i = \dim C_i$ is the multiplicity of λ_i as an eigenvalue, i.e. $\nu_i = \mu_i$.

Choosing a basis of each C_i, and using the fact that $T(C_i) \subset C_i$, we get a basis $\mathcal{B}$ of V for which the matrix $A = \mathcal{M}_{\mathcal{B}}^{\mathcal{B}}(T)$ has block form

$$A = \begin{pmatrix} A_1 & O & \cdots & O \\ O & A_2 & \cdots & O \\ \vdots & \vdots & \ddots & \vdots \\ O & O & \cdots & A_m \end{pmatrix}, \tag{13.5}$$

where A_i is $\nu_i \times \nu_i$, and each O is a zero matrix. It follows that

$$p_A(x) = p_{A_1}(x) \cdots p_{A_m}(x). \tag{13.6}$$

But the only eigenvalue of A_i is λ_i since $R_j = T - \lambda_j I_n$ is one to one on C_i if $i \neq j$. Thus $p_{A_i}(x) = (x - \lambda_i)^{\nu_i}$, so we conclude that the multiplicity of λ_i is ν_i, which finishes the proof of (1).

Note that since $p_T(x) = p_A(x)$, we have also shown (13.1). That is,

$$p_T(x) = (x - \lambda_1)^{\mu_1} \cdots (x - \lambda_m)^{\mu_m}.$$

We next prove (2). The transformation S is well defined, and, by definition, $S(C_i) \subset C_i$ for all i. Since $T(C_i) \subset C_i$, we have $N(C_i) \subset C_i$ too. To show N is nilpotent, we only need to show that N is nilpotent on each C_i. But for $\mathbf{v} \in C_i$, we have

$$N^{r_i}(\mathbf{v}) = (T - S)^{r_i}(\mathbf{v}) = (T - \lambda_i I_n)^{r_i}(\mathbf{v}) = \mathbf{0},$$

by the definition of C_i. Hence N is nilpotent.

To prove (3), it suffices to show that if $\mathbf{v} \in C_i$, then $NS(\mathbf{v}) = SN(\mathbf{v})$. But this is obvious, since $N(C_i) \subset C_i$ and $S(\mathbf{v}) = \lambda_i \mathbf{v}$ for all $\mathbf{v} \in C_i$.

To finish the proof, suppose $T = S' + N'$ is another decomposition of T, where S' is semi-simple, N' is nilpotent, and $S'N' = N'S'$.

Now as S' is semi-simple, we can write

$$V = \sum_{i=1}^{k} E_{\gamma_i}(S'), \tag{13.7}$$

where $\gamma_1, \ldots, \gamma_k$ are the distinct eigenvalues of S' and $E_{\gamma_i}(S')$ denotes the eigenspace of S' for γ_i. Since N' and S' commute, $N'(E_{\gamma_i}(S')) \subset E_{\gamma_i}(S')$. Therefore, we can assert that for any $\mathbf{v} \in E_{\gamma_i}(S')$,

$$N'^{r}(\mathbf{v}) = (T - S')^r(\mathbf{v}) = (T - \gamma_i I_n)^r(\mathbf{v}) = \mathbf{0}$$

if r is sufficiently large. But this says γ_i is an eigenvalue of T, say $\gamma_i = \lambda_j$. Thus, $E_{\gamma_i}(S') \subset C_j$. Hence, $S = S'$ on $E_{\gamma_i}(S')$, and therefore (13.7) implies $S' = S$ on V. Hence, $N = N'$ too, and the proof is complete. □

Definition 13.1. The subspaces $C_1, \ldots, C_m$ associated to $T : V \to V$ are called the *invariant subspaces* of T.

Corollary 13.5. *If $\mathbb{F}$ is algebraically closed, then any $n \times n$ matrix A over $\mathbb{F}$ can be expressed in one and only one way as the sum $A = S + N$ of two commuting matrices S and N in $\mathbb{F}^{n \times n}$, where S is diagonalizable and N is nilpotent.*

Proof. This follows immediately from the Theorem. □

Let's compute an example.

Example 13.1. Let $V = \mathbb{F}^3$, and let T be the matrix linear transformation

$$T = \begin{pmatrix} 5 & 12 & 6 \\ -2 & -5 & -3 \\ 1 & 4 & 4 \end{pmatrix}.$$

The characteristic polynomial of T is $-(x-1)^2(x-2)$, so the eigenvalues are 1 and 2, which is repeated. Now the matrices $T - I_3$ and $T - 2I_3$ row reduce to

$$\begin{pmatrix} 1 & 0 & -2 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 1 & 0 & -3 \\ 0 & 1 & 3/2 \\ 0 & 0 & 0 \end{pmatrix}$$

respectively. Hence T is not semi-simple. Eigenvectors for 2 and 1 are respectively

$$\begin{pmatrix} 2 \\ -1 \\ 1 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 3 \\ -3/2 \\ 1 \end{pmatrix}.$$

Let us find the invariant subspaces. Since 2 is a simple root, its invariant subspace is simply the line $E_2(T) = \mathbb{C}(2, -1, 1)^T$. Now

$$(T - I_3)^2 = \begin{pmatrix} -2 & 0 & 6 \\ 1 & 0 & -3 \\ -1 & 0 & 3 \end{pmatrix},$$

which clearly has rank one. Its kernel, which is spanned by

$$\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 3 \\ 0 \\ 1 \end{pmatrix},$$

is therefore the other invariant subspace of T. Hence the semi-simple linear transformation S is determined by

$$S(\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}) = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad S(\begin{pmatrix} 3 \\ 0 \\ 1 \end{pmatrix}) = \begin{pmatrix} 3 \\ 0 \\ 1 \end{pmatrix}, \quad S(\begin{pmatrix} 2 \\ -1 \\ 1 \end{pmatrix}) = 2 \begin{pmatrix} 2 \\ -1 \\ 1 \end{pmatrix}.$$

The matrix of S (as usual found by $SP = PD$) is therefore

$$M_S = \begin{pmatrix} -1 & 0 & 6 \\ 1 & 1 & -3 \\ -1 & 0 & 4 \end{pmatrix},$$

and we get N by subtraction:

$$N = \begin{pmatrix} 6 & 12 & 0 \\ -3 & -6 & 0 \\ 2 & 4 & 0 \end{pmatrix}.$$

Thus the decomposition of T as the sum of commuting diagonalizable and a nilpotent matrices is

$$\begin{pmatrix} 5 & 12 & 6 \\ -2 & -5 & -3 \\ 1 & 4 & 4 \end{pmatrix} = \begin{pmatrix} -1 & 0 & 6 \\ 1 & 1 & -3 \\ -1 & 0 & 4 \end{pmatrix} + \begin{pmatrix} 6 & 12 & 0 \\ -3 & -6 & 0 \\ 2 & 4 & 0 \end{pmatrix}.$$

Notice that if P is the matrix which diagonalizes S, i.e.

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

then

$$P^{-1}TP = \begin{pmatrix} -5 & -9 & 0 \\ 4 & 7 & 0 \\ 0 & 0 & 2 \end{pmatrix}.$$

This gives us the matrix $\mathcal{M}_{\mathcal{B}}^{\mathcal{B}}(T)$ of T consisting of blocks down the diagonal. We will also see that by choosing P more carefully, we can even guarantee that $P^{-1}TP$ is upper triangular.

Exercises

Exercise 13.1. Find the Jordan decomposition for each of the following 2×2 matrices:

(*i*) $A = \begin{pmatrix} 1 & -1 \\ 1 & -1 \end{pmatrix}$,

(*ii*) $B = \begin{pmatrix} 2 & 4 \\ 0 & 1 \end{pmatrix}$,

(*iii*) $C = \begin{pmatrix} 3 & 1 \\ -1 & 1 \end{pmatrix}$.

Exercise 13.2. Find the Jordan decomposition for each of the following 3×3 matrices:

(*i*) $A = \begin{pmatrix} 2 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 2 \end{pmatrix}$, and

(*ii*) $B = \begin{pmatrix} 2 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 2 \end{pmatrix}$.

Exercise 13.3. Suppose $T : V \to V$ is asemi-simple linear transformationon a finite dimensional vector space V, and suppose W is a subspace of V such that $T(W) \subset W$. Let $T' : W \to W$ be the linear transformation induced by restricting T to W. Show that T' is semi-simple.

Exercise 13.4. Suppose A and B are commuting $n \times n$ matrices over $\mathbb{C}$. Suppose A and B are both diagonalizable. Show that they are simultaneously diagonalizable. Hint: Exercise 13.3 may be helpful.

13.2 Further Results on Linear Transformations

The purpose of this Section is to derive some consequences of the Jordan Decomposition Theorem. We will first give a complete proof of the Cayley-Hamilton Theorem, and that will be followed by a discussion of the Jordan Canonical Form.

13.2.1 The Cayley-Hamilton Theorem

The Cayley-Hamilton Theorem, which was partially proved in Chapter 9, will now be proved completely. It follows almost immediately from the Jordan decomposition.

Theorem 13.6. *Let $\mathbb{F}$ be algebraically closed, and let V be finite dimensional over $\mathbb{F}$. Then every $T \in L(V)$ satisfies its own characteristic polynomial.*

Proof. Let $\lambda_1, \ldots, \lambda_m$ be the distinct eigenvalues of T, and let μ_i be the multiplicity of λ_i. We have to show that for every $\mathbf{v} \in V$,

$$(T - \lambda_1 I_n)^{\mu_1} \cdots (T - \lambda_m I_n)^{\mu_m}(\mathbf{v}) = \mathbf{0}. \tag{13.8}$$

Since $V = \sum C_i$, it suffices to show (13.8) if $\mathbf{v} \in C_i$ for some i. What we need is

Lemma 13.7. *Let W be a finite dimensional vector space and assume $T \in L(W)$ is nilpotent. Then $T^{\dim W} = O$.*

Proof. We will leave this as an exercise. □

To finish the proof of the Theorem, suppose $\mathbf{v} \in C_i$. In the proof of the Jordan Decomposition Theorem, we showed that $T - \lambda_i I_n$ is nilpotent on C_i. As $\dim C_i = \mu_i$, the Lemma says that if $\mathbf{v} \in C_i$, then $(T - \lambda_i I_n)^{\mu_i}(\mathbf{v}) = \mathbf{0}$. But this implies (13.8) for $\mathbf{v}$ since the operators $(T - \lambda_i I_n)^{\mu_i}$ commute. Hence the proof is complete. □

Corollary 13.8. *A linear transformation $T : V \to V$ is nilpotent if and only if every eigenvalue of T is 0.*

Proof. We leave this as an exercise. □

One of our big questions is how do you tell whether a linear transformation $T : V \to V$ is semi-simple. The simplest characterization seems to be as follows. As usual, $\lambda_1, \ldots, \lambda_m$ denote the distinct eigenvalues of T.

Theorem 13.9. *A linear transformation $T : V \to V$ is semi-simple if and only if*

$$(T - \lambda_1 I_n) \cdots (T - \lambda_m I_n) = O. \tag{13.9}$$

Proof. T is semi-simple if and only if T equals its semi-simple part S. Since $S = \lambda_i I_n$ on C_i (see the proof of Theorem 13.1), it follows that T is semi-simple if and only if $T - \lambda_i I_n = O$ on each C_i. Hence if T is semi-simple, then $(T - \lambda_1 I_n) \cdots (T - \lambda_m I_n) = O$ on V. On the other hand, suppose $(T - \lambda_1 I_n) \cdots (T - \lambda_m I_n) = O$. Then Lemma 13.4 says that $\dim V = \sum_{i=1}^{m} \dim E_{\lambda_i}(T)$, where $E_\lambda(T)$ is the eigenspace of T corresponding to the eigenvalue λ. By the diagonalizability criterion of Proposition 9.12, it follows that T is semi-simple. □

Example 13.2. Let's reconsider the linear transformation $T : \mathbb{C}^3 \to \mathbb{C}^3$ from Example 13.1. By direct computation,

$$(T - I_3)(T - 2I_2) = \begin{pmatrix} -6 & -12 & 0 \\ 3 & 6 & 0 \\ -2 & -4 & 0 \end{pmatrix}.$$

This tells us that T is not semi-simple, which of course, we already knew.

Let $\mathbb{F}[x]$ denote the ring of polynomials in a variable x with coefficients in $\mathbb{F}$. Notice that the Cayley-Hamilton Theorem tells us that there is always a polynomial $p(x) \in \mathbb{F}[x]$ for which $p(T) = O$. This is also guaranteed by the fact that $\dim L(V) = n^2$, so $I_n,\ T,\ T^2,\ \ldots$ can't all be linearly independent.

Definition 13.2. Let $T : V \to V$ be linear, and assume $T \neq O$. Then the polynomial $p(x) \in \mathbb{F}[x]$ of least positive degree and leading coefficient one such that $p(T) = O$ is called the *minimal polynomial* of T.

Of course, it isn't clear that a unique minimal polynomial exists. However, let p_1 and p_2 each be a minimal polynomial. By division with remainders, we can find polynomials $q(x)$ and $r(x)$ in $\mathbb{F}[x]$ such that

$$p_2(x) = q(x)p_1(x) + r(x),$$

where either $r = 0$ or the degree of r is less than the degree of p_1. But as $p_1(T) = p_2(T) = O$, it follows that $r(T) = O$ also. Since either $r = 0$ or the degree of r is smaller than the degree of p_2, we conclude $r = 0$. But then $q(x)$ is a constant since p_1 and p_2 have to have the same degree. Thus $q = 1$ since p_1 and p_2 each have leading coefficient one. Hence $p_1 = p_2$.

Proposition 13.10. *Suppose $T : V \to V$ is a nonzero linear transformationwith distinct eigenvalues are $\lambda_1, \ldots, \lambda_m$. The minimal polynomial $p(x)$ of T is unique, it divides the characteristic polynomial $p_T(x)$, and $(x - \lambda_1) \cdots (x - \lambda_m)$ divides $p(x)$.*

Proof. The uniqueness was already shown. By the Cayley-Hamilton Theorem, $p_T(T) = O$. Hence writing $p_T(x) = q(x)p(x) + r(x)$ as above and repeating the argument, we get $r = 0$. The fact that $(x - \lambda_1) \cdots (x - \lambda_m)$ divides $p(x)$ is clear from the proof of Theorem 13.1. Indeed, we can factor $p(x)$ into linear factors $p(x) = (x - a_1) \cdots (x - a_k)$ where all $a_i \in \mathbb{F}$. If a λ_j is not among the a_i, we know $p(T)$ cannot be zero on C_j. Hence each $(x - \lambda_j)$ has to be a factor. □

Corollary 13.11. *A nonzero linear transformation $T : V \to V$ is semi-simple if and only if its minimal polynomial is $(x - \lambda_1) \cdots (x - \lambda_m)$.*

Proof. Just apply Theroem 13.9 and Proposition 13.11. □

Example 13.3. Let's reconsider $T : \mathbb{C}^3 \to \mathbb{C}^3$ from Example 13.1. We have $P_T(x) = -(x - 1)^2(x - 2)$. Now by direct computation,

$$(T - I_3)(T - 2I_2) = \begin{pmatrix} -6 & -12 & 0 \\ 3 & 6 & 0 \\ -2 & -4 & 0 \end{pmatrix}.$$

This tells us that T is not semi-simple, which, of course, we already knew.

13.2.2 The Jordan Canonical Form

The Jordan decomposition $T = S + N$ of a $T \in L(V)$ can be extensively improved. The first step is to find a basis for which T is upper triangular. Recall that such a basis is known as a flag basis. In fact, it will suffice to find a flag basis for N on each invariant subspace C_i of T.

For this we may as well suppose $C_i = V$. In other words, we're assuming $T : V \to V$ is a linear transformation with a single eigenvalue. Let k be the least positive integer for which $N^k = O$, where N is the nilpotent part. Clearly

$$\ker(N) \subset \ker(N^2) \subset \cdots \subset \ker(N^k) = V.$$

Since k is the least integer such that $N^k = O$, each of the above inclusions is proper. Thus we can construct a basis of V by first selecting a basis of $\ker(N)$, extending this basis to a basis of $\ker(N^2)$, extending the second

basis to a basis of $\ker(N^3)$ and so forth until a basis $\mathcal{B}$ of $V = \ker(N^k)$ is obtained.

Notice that for each $r > 0$, $N(\ker(N^r)) \subset \ker(N^{r-1})$. Thus the basis just constructed gives us a flag basis for N, and so the matrix $\mathcal{M}_{\mathcal{B}}^{\mathcal{B}}(N)$ is strictly upper triangular with zeros on its diagonal since N is nilpotent. Since S is a multiple of $\lambda_i I_{\mu_i}$ on C_i, $\mathcal{M}_{\mathcal{B}}^{\mathcal{B}}(T)$ is upper triangular and each of its diagonal entries is equal to λ.

Example 13.4. Let

$$N = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}.$$

Then $N(\mathbf{e}_1) = \mathbf{0}$, $N(\mathbf{e}_2) = \mathbf{e}_1$ and $N(\mathbf{e}_3) = \mathbf{e}_2$. Thus

$$\ker(N) = \mathbb{F}\mathbf{e}_1 \subset \ker(N^2) = \mathbb{F}\mathbf{e}_1 + \mathbb{F}\mathbf{e}_2 \subset (N^3) = \mathbb{F}^3.$$

The matrix N is an example of a matrix in Jordan Canonical Form.

Theorem 13.12 (The Jordan Canonical Form)**.** *As usual, let V be a finite dimensional vector space over the algebraically closed field $\mathbb{F}$, and suppose $T \in L(V)$. Then there exists a basis $\mathcal{B}$ of V for which*

$$\mathcal{M}_{\mathcal{B}}^{\mathcal{B}}(T) = \begin{pmatrix} J_1 & O & \cdots & O \\ O & J_2 & \cdots & O \\ \vdots & \vdots & \ddots & \vdots \\ O & \cdots & O & J_s \end{pmatrix}, \tag{13.10}$$

where the matrices J_i (the Jordan blocks) have the form

$$J_i = \lambda I_{n_i} + N_i,$$

where N_i is the upper triangular $n_i \times n_i$ matrix with 0's on the diagonal, 1's on the super diagonal and 0's above the super diagonal (as in the example N above). Furthermore, we may suppose $n_1 \geq n_2 \geq \cdots \geq n_s$. In particular, when $V = \mathbb{F}^n$, we get the result that every $A \in \mathbb{F}^{n \times n}$ is similar to a matrix having the form (13.10).

The proof requires that we play around a bit more in the manner of the above discussion. We will skip the details, as they are quite messy.

Note that there is no connection between the n_i and the eigenvalues λ_j of T, except that if $J_i = \lambda I_{n_i} + N_i$, then n_i cannot exceed the multiplicity of λ as a root of $p_T(x)$. Note also that each eigenvalue λ_i of T appears μ_i times on the diagonal of $\mathcal{M}_{\mathcal{B}}^{\mathcal{B}}(T)$.

Example 13.5. Here are a couple of examples in the 4×4 case:

$$J_1 = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad J_2 = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \quad \text{and} \quad J_3 = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

The first matrix has one 4×4 Jordan block, the second has a 3×3 Jordan block and a 1×1 Jordan block and the third has two 2×2 Jordan blocks.

13.2.3 A Connection With Number Theory

One of the interesting connections between matrix theory and number theory comes from the Jordan Canonical Forms of nilpotent matrices. To elucidate this, we need to introduce the notion of a partition of n.

Definition 13.3. Let n be a positive integer. Then a *partition* is a non-increasing sequence of positive integers $a_1 \geq a_2 \geq \cdots \geq a_s$ such that $\sum_{i=1}^{s} a_i = n$. The partition function $\pi(n)$ is the function which counts the number of partitions of n.

Thus $\pi(1) = 1$, $\pi(2) = 2$, $\pi(3) = 3$ and $\pi(4) = 5$.

Example 13.6. The partitions of 6 are

$$6 = 1+1+1+1+1+1 = 2+1+1+1 = 2+2+1+1 = 3+2+1 =$$

$$2+2+2 = 3+3 = 4+1+1 = 4+2 = 5+1.$$

Thus there are 10 partition of 6, so $\pi(6) = 10$.

The partition function grows very rapidly. The upshot of the Jordan Canonical Form is that to each partition $(n_1, n_2, \ldots, n_s)$ of n, there is a nilpotent matrix of the form (13.10) (with only zeros on the diagonal, of course), and every $n \times n$ nilpotent matrix is similar to one of these matrices. This seemingly accidental connection has lead to some suprisingly deep results in algebra.

Exercises

Exercise 13.5. Let W be a finite dimensional vector space and assume $T \in L(W)$ is nilpotent. Show that $T^{\dim W} = O$.

Exercise 13.6. Use the Cayley-Hamilton Theorem to prove directly that the minimal polynomial of a linear transformation $T : V \to V$ divides the characteristic polynomial of T. (Hint: write $p_T(x) = a(x)p(x) + r(x)$ where either $r = 0$ or the degree of $r(x)$ is smaller than the degree of $p(x)$.)

Exercise 13.7. Prove Corollary 13.8 directly using the fact that

$$\ker(T) \subset \ker(T^2) \subset \ker(T^3) \subset \cdots .$$

Exercise 13.8. Compute the minimal polynomials of the following matrices:

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

Exercise 13.9. Show that if $\mathbb{F} = \mathbb{C}$, the Cayley-Hamilton Theorem follows from the fact any element of $L(V)$ is the limit of a sequence of semi-simple elements of $L(V)$. How does one construct such a sequence?

Exercise 13.10. Prove directly that if A is a 2×2 matrix over $\mathbb{F}$ which isn't diagonalizable, then A is similar to a matrix of the form

$$\begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix}.$$

Exercise 13.11. Describe the Jordan Canonical Form of a diagonalizable matrix.

Exercise 13.12. Deduce a Jordan Canonical Form for

$$M = \begin{pmatrix} 5 & 12 & 6 \\ -2 & -5 & -3 \\ 1 & 4 & 4 \end{pmatrix}.$$

Exercise 13.13. Deduce a Jordan Canonical Form for

$$N = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}.$$

Exercise 13.14. List all 4×4 and 5×5 nilpotent matrices in Jordan Canonical Form.

13.3 Summary

The Jordan Decomposition Theorem states that if V is a finite dimensional vector space over an algebraically closed field, then every linear transformation $T : V \to V$ admits a unique decomposition into the sum of a semi-simple linear transformation S and a nilpotent transformation N which commute: $SN = NS$. This result is closely related to the Cayley-Hamilton Theorem (T satisfies its characteristic polynomial), and, in fact, yields a short proof. The Jordan Decomposition Theorem gives a simple necessary and sufficient criterion for T to be semi-simple. Simply put, T is semi-simple if and only if its minimal polynomial is $(x - \lambda_1) \cdots (x - \lambda_m)$, where $\lambda_1, \ldots, \lambda_m$ are the distinct eigenvaluesof T.

Moreover, there exists an eigenbasis $\mathcal{B}$ for S such that the matrix of T with respect to $\mathcal{B}$ is upper triangular. A refinement known as the Jordan Canonical Form asserts further that there exists an eigenbasis for S for which the matrix M_N of N takes a particularly simple form: M_N has a decomposition into so called Jordan blocks, where each Jordan block has 1's on its super-diagonal and zeros everywhere else. It follows that the equivalence classes of nilpotent matrices under similarity are in one to one correspondence with the partitions of n, which gives a surprising connection between matrix theory and number theory.

Chapter 14

Appendix: $\mathbb{R}^2$ and $\mathbb{R}^3$

The purpose of this Appendix is to give a brief introduction to the fundamental concepts of vectors and their geometry in $\mathbb{R}^2$ and $\mathbb{R}^3$.

14.1 Basic Concepts

$\mathbb{R}^2$ and $\mathbb{R}^3$ consist respectively of all column vectors $\begin{pmatrix} x \\ y \end{pmatrix}$ and $\begin{pmatrix} x \\ y \\ z \end{pmatrix}$, where $x,\ y,\ z \in \mathbb{R}$. They represent a plane with a two dimensional coordinate system and ordinary space with a three dimensional coordinate system.

FIGURE 1 (Euclidean PLANE)

FIGURE 2 (Euclidean 3-space)

One can always visualize $\mathbb{R}^2$ as a subset of $\mathbb{R}^3$, namely as those vectors whose last component $z = 0$. Vectors in $\mathbb{R}^2$ or $\mathbb{R}^3$ are added by adding there corresponding components. For example, in $\mathbb{R}^3$ we have

$$\begin{pmatrix} r \\ s \\ t \end{pmatrix} + \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} r+x \\ s+y \\ t+z \end{pmatrix}. \tag{14.1}$$

This definition of addition satisfies a rule called the *Parallelogram Law.*
Parallelogram Law: *The sum* $\mathbf{a} + \mathbf{b}$ *of two vectors in* $\mathbb{R}^2$ *or* $\mathbb{R}^3$ *is the vector along the diagonal of the parallelogram with vertices at* $\mathbf{0}$, $\mathbf{a}$ *and* $\mathbf{b}$.

FIGURE 3
(PARALLELOGRAM LAW)

There is a second operation called *scalar multiplication*, where a vector $\mathbf{a}$ is dilated by a real number r. This is defined for $\mathbb{R}^3$ (in a rather obvious way) by

$$r\mathbf{a} = r\begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} = \begin{pmatrix} ra_1 \\ ra_2 \\ ra_3 \end{pmatrix}. \tag{14.2}$$

Scalar multiplication has an obvious geometric interpretation. Multiplying $\mathbf{a}$ by r stretches or shrinks $\mathbf{a}$ along itself by the factor $|r|$, changing its direction if $r < 0$. The geometric interpretation of addition is the Parallelogram Law.

14.2 Lines

We now discuss how to represent lines. We will discuss planes below. A line in $\mathbb{R}^2$ is cut out by a linear equation $ax + by = c$, but a line in $\mathbb{R}^3$ needs two equations $a_1x + b_1y + c_1z = d_1$ and $a_2x + b_2y + c_2z = d_2$. Intuitively, each equation cuts out a plane, and a line is the set of points on both planes. But a better approach to finding a line is to use the principle that a line is determined by two points.

So suppose we want to express the line through $\mathbf{a}$ and $\mathbf{b}$. Notice that the curve

$$\mathbf{x}(t) = \mathbf{a} + t(\mathbf{b} - \mathbf{a}) = (1 - t)\mathbf{a} + t\mathbf{b}, \tag{14.3}$$

where t varies through $\mathbb{R}$, has the property that $\mathbf{x}(0) = \mathbf{a}$, and $\mathbf{x}(1) = \mathbf{b}$. The motion depends linearly on t, and, as can be seen from the Parallelogram Law, $\mathbf{x}(t)$ traces out the line through $\mathbf{a}$ parallel to $\mathbf{b} - \mathbf{a}$.

Definition 14.1. The *line through* $\mathbf{a}$ *parallel to* $\mathbf{c}$ is defined to be the path traced out by the curve $\mathbf{x}(t) = \mathbf{a} + t\mathbf{c}$ as t takes on all real values. We will refer to $\mathbf{x}(t) = \mathbf{a} + t\mathbf{c}$ as the *vector form* of the line.

The vector form $\mathbf{x}(t) = \mathbf{a} + t\mathbf{c}$ leads directly to *parametric form* of the line. In the parametric form, the components $x_1,\ x_2,\ x_3$ of $\mathbf{x}$ are expressed as linear functions of t as follows:

$$x_1 = a_1 + tc_1,\ x_2 = a_2 + tc_2,\ x_3 = a_3 + tc_3. \tag{14.4}$$

Note that the definition for a line in the plane is essentially the same. Just forget the last coordinate.

Letting $\mathbf{a}$ vary while $\mathbf{b}$ is kept fixed gives all lines of the form $\mathbf{x} = \mathbf{a}+t\mathbf{b}$. Every point of $\mathbb{R}^3$ is on one of these lines, and two lines either coincide or don't meet at all. (The proof of this is an exercise.) We will say that two lines $\mathbf{a}+t\mathbf{b}$ and $\mathbf{a}'+t\mathbf{b}'$ are *parallel* if $\mathbf{b}$ and $\mathbf{b}'$ are collinear. That is, $\mathbf{b}' = a\mathbf{c}$ for some real number a. We will also say that the line $\mathbf{a} + t\mathbf{b}$ is parallel to $\mathbf{b}$.

Example 14.1. Let's find an expression for the line in $\mathbb{R}^3$ passing through $\begin{pmatrix} 3 \\ 4 \\ -1 \end{pmatrix}$ and $\begin{pmatrix} 1 \\ 0 \\ 9 \end{pmatrix}$. We apply the trick in (14.3). Consider

$$\mathbf{x} = (1-t)\begin{pmatrix} 3 \\ 4 \\ -1 \end{pmatrix} + t\begin{pmatrix} 1 \\ 0 \\ 9 \end{pmatrix}.$$

Clearly, when $t = 0$, $\mathbf{x} = \begin{pmatrix} 3 \\ 4 \\ -1 \end{pmatrix}$, and when $t = 1$, $\mathbf{x} = \begin{pmatrix} 1 \\ 0 \\ 9 \end{pmatrix}$. We can then express $\mathbf{x}$ in the vector form $\mathbf{x} = \mathbf{a} + t(\mathbf{b} - \mathbf{a})$ where $\mathbf{a} = \begin{pmatrix} 3 \\ 4 \\ -1 \end{pmatrix}$ and $\mathbf{b} = \begin{pmatrix} 1 \\ 0 \\ 9 \end{pmatrix}$. The parametric form is

$$x_1 = -2t + 1,\ x_2 = -4t + 4,\ x_3 = 10t + 1.$$

14.3 The Inner Product

Before discussing planes, we will introduce the inner product on $\mathbb{R}^2$ and $\mathbb{R}^3$. This is the fundamental tool on which all measurement is based.

Definition 14.2. The inner product of two vectors $\mathbf{a} = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix}$ and $\mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}$ in $\mathbb{R}^3$ is defined to be $\mathbf{a} \cdot \mathbf{b} := a_1b_1 + a_2b_2 + a_3b_3$.

The inner product satisfies several basic identities. Let $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$ be arbitrary vectors and r any scalar (i.e., $r \in \mathbb{R}$). Then

(1) $\mathbf{a} \cdot \mathbf{b} = \mathbf{b} \cdot \mathbf{a}$,

(2) $(\mathbf{a} + \mathbf{b}) \cdot \mathbf{c} = \mathbf{a} \cdot \mathbf{c} + \mathbf{b} \cdot \mathbf{c}$,

(3) $(r\mathbf{a}) \cdot \mathbf{b} = \mathbf{a} \cdot (r\mathbf{b}) = r(\mathbf{a} \cdot \mathbf{b})$, and

(4) $\mathbf{a} \cdot \mathbf{a} > 0$ unless $\mathbf{a} = \mathbf{0}$, in which case $\mathbf{a} \cdot \mathbf{a} = 0$.

These properties are all easy to prove, so we will leave them as exercises.

The *length* $|\mathbf{a}|$ of $\mathbf{a}$ is defined in terms of the dot product by putting

$$\begin{aligned} |\mathbf{a}| &= \sqrt{\mathbf{a} \cdot \mathbf{a}} \\ &= (\sum_{i=1}^{3} a_i^2)^{1/2}. \end{aligned}$$

Notice that

$$|r\mathbf{a}| = |r||\mathbf{a}|\,.$$

The *distance* $d(\mathbf{a}, \mathbf{b})$ between two vectors $\mathbf{a}$ and $\mathbf{b}$ is defined as the length of their difference $\mathbf{a} - \mathbf{b}$. Thus

$$\begin{aligned} d(\mathbf{a}, \mathbf{b}) &= |\mathbf{a} - \mathbf{b}| \\ &= \big((\mathbf{a} - \mathbf{b}) \cdot (\mathbf{a} - \mathbf{b})\big)^{1/2} \\ &= \big(\sum_{i=1}^{3} (a_i - b_i)^2\big)^{1/2}. \end{aligned}$$

The dot product is used to detect when two vectors are orthogonal. We say $\mathbf{a}$ and $\mathbf{b}$ are *orthogonal* (a fancy word for perpendicular) if $\mathbf{a} \cdot \mathbf{b} = 0$. The zero vector $\mathbf{0}$ is orthogonal to every vector, and by property (4) of the dot product, $\mathbf{0}$ is the only vector orthogonal to itself. Two vectors $\mathbf{a}$ and $\mathbf{b}$ in $\mathbb{R}^2$ are orthogonal if and only if and only if $a_1 b_1 + a_2 b_2 = 0$. Thus if $a_1, b_2 \neq 0$, then $\mathbf{a}$ and $\mathbf{b}$ are orthogonal if and only if $a_2/a_1 = -b_1/b_2$. Hence the slopes of orthogonal vectors in $\mathbb{R}^2$ are negative reciprocals.

Proposition 14.1. *Two vectors* $\mathbf{a}$ *and* $\mathbf{b}$ *are orthogonal if and only if* $|\mathbf{a} + \mathbf{b}| = |\mathbf{a} - \mathbf{b}|$.

To prove this, consider the triangle with vertices at $\mathbf{0}, \mathbf{a}, \mathbf{b}$. The hypotenuse of this triangle is a segment of length $|\mathbf{a} - \mathbf{b}|$. (This follows from the Parallelogram Law.) Next consider the triangle with vertices at $\mathbf{0}, \mathbf{a}, -\mathbf{b}$. The hypotenuse of this triangle is likewise a segment of length $|\mathbf{a} + \mathbf{b}|$. Now suppose $|\mathbf{a}+\mathbf{b}| = |\mathbf{a}-\mathbf{b}|$. Then by the side side side criterion for congruence,

which says that two triangles are congruent if and only if they have corresponding sides of equal length, the two triangles are congruent. It follows that $\mathbf{a}$ and $\mathbf{b}$ are orthogonal. For the converse direction, suppose $\mathbf{a}$ and $\mathbf{b}$ are orthogonal. Then the side angle side criterion for congruence applies, so our triangles are congruent. Thus $|\mathbf{a}+\mathbf{b}| = |\mathbf{a}-\mathbf{b}|$.

DIAGRAM FOR PROOF

It is much easier to prove this Proposition with algebra (namely the dot product). The point is that $\mathbf{a}\cdot\mathbf{b} = 0$ if and only if $|\mathbf{a}+\mathbf{b}|^2 = |\mathbf{a}-\mathbf{b}|^2$. For

$$(\mathbf{a}+\mathbf{b})\cdot(\mathbf{a}+\mathbf{b}) = (\mathbf{a}-\mathbf{b})\cdot(\mathbf{a}-\mathbf{b})$$

reduces to the equation $2\mathbf{a}\cdot\mathbf{b} = -2\mathbf{a}\cdot\mathbf{b}$, which holds if and only if $\mathbf{a}\cdot\mathbf{b} = 0$.

14.4 Planes

A *plane in* $\mathbb{R}^3$ is defined to be the solution set of a linear equation

$$ax + by + cz = d \tag{14.5}$$

in three variables x, y and z. The linear equation (14.5) expresses that the dot product of the vector $\mathbf{a} = \begin{pmatrix} a \\ b \\ c \end{pmatrix}$ and the variable vector $\mathbf{x} = \begin{pmatrix} x \\ y \\ z \end{pmatrix}$ is the constant d:

$$\mathbf{a}\cdot\mathbf{x} = d.$$

If $d = 0$, the plane passes through the origin, and its equation is said to be *homogeneous*. In this case it is easy to see how to interpret the plane equation. The plane $ax + by + cz = 0$ consists of all $\begin{pmatrix} r \\ s \\ t \end{pmatrix}$ orthogonal to $\mathbf{a} = \begin{pmatrix} a \\ b \\ c \end{pmatrix}$. For this reason, $\begin{pmatrix} a \\ b \\ c \end{pmatrix}$ is said to be a *normal* to the plane. (On a good day, one is normal to the plane of the floor.)

Example 14.2. Find the plane through $\begin{pmatrix}1\\2\\3\end{pmatrix}$ with nomal $\begin{pmatrix}2\\3\\5\end{pmatrix}$. Now $\mathbf{a} = \begin{pmatrix}2\\3\\5\end{pmatrix}$, so in the equation (14.5) we have $d = \begin{pmatrix}2\\3\\5\end{pmatrix} \cdot \begin{pmatrix}1\\2\\3\end{pmatrix} = 23$. Hence the plane is $2x + 3y + 5z = 23$.

Holding $\mathbf{a} \neq \mathbf{0}$ constant and varying d gives a family of planes $\mathbf{a} \cdot \mathbf{x} = d$ completely filling up $\mathbb{R}^3$ such that two planes either coincide or don't have any points in common. Hence the family of planes $ax + by + cz = d$ (a, b, c fixed and d arbitrary) are all *parallel*. By drawing a picture, one can see from the Parallelogram Law that every vector $\begin{pmatrix}r\\s\\t\end{pmatrix}$ on $ax + by + cz = d$ is the sum of a fixed vector $\begin{pmatrix}x_0\\y_0\\z_0\end{pmatrix}$ on $ax + by + cz = d$ and an arbitrary vector $\begin{pmatrix}x\\y\\z\end{pmatrix}$ on the parallel plane $ax + by + cz = 0$ through the origin.

Example 14.3. Suppose we want to find the plane through

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

The linear system

$$\begin{aligned} a + b + c &= d \\ a + 2b + c &= d \\ c &= d \end{aligned}$$

expresses the fact that the plane $ax + by + cz = d$ contains all three points. Subtracting the first equation from the second, we get $b = 0$. The last equation says $c = d$, so from the first, we get $a + c = c$. Hence $a = b = 0$ and $c = d$. The plane therefore has the form $cz = c$ or $z = 1$ since $c \neq 0$. (The astute observer would have noticed this at the beginning and saved themself having to do the calculation.)

14.5 Orthogonal Decomposition and Projection

One of the fundamental applications of the dot product is the *orthogonal decomposition* of a vector into two or more mutually orthogonal components.

Proposition 14.2. *Given* $\mathbf{a}, \mathbf{b}$ *such that* $\mathbf{b} \neq \mathbf{0}$, *there exists a unique scalar* r *so that* $\mathbf{a} = r\mathbf{b} + \mathbf{c}$ *where* $\mathbf{b}$ *and* $\mathbf{c}$ *are orthogonal. In fact,*

$$r = (\frac{\mathbf{a} \cdot \mathbf{b}}{\mathbf{b} \cdot \mathbf{b}}),$$

and

$$\mathbf{c} = \mathbf{a} - (\frac{\mathbf{a} \cdot \mathbf{b}}{\mathbf{b} \cdot \mathbf{b}})\mathbf{b}.$$

Proof. We see this as follows: since we want $r\mathbf{b} = \mathbf{a} - \mathbf{c}$, where $\mathbf{c}$ has the property that $\mathbf{b} \cdot \mathbf{c} = 0$, then

$$r\mathbf{b} \cdot \mathbf{b} = (\mathbf{a} - \mathbf{c}) \cdot \mathbf{b} = \mathbf{a} \cdot \mathbf{b} - \mathbf{c} \cdot \mathbf{b} = \mathbf{a} \cdot \mathbf{b}.$$

As $\mathbf{b} \cdot \mathbf{b} \neq 0$, it follows that $r = \mathbf{a} \cdot \mathbf{b}/\mathbf{b} \cdot \mathbf{b}$. The reader should check that $\mathbf{c} = \mathbf{a} - (\frac{\mathbf{a} \cdot \mathbf{b}}{\mathbf{b} \cdot \mathbf{b}})\mathbf{b}$ is in fact orthogonal to $\mathbf{b}$. Thus we get the desired orthogonal decomposition

$$\mathbf{a} = (\frac{\mathbf{a} \cdot \mathbf{b}}{\mathbf{b} \cdot \mathbf{b}})\mathbf{b} + \mathbf{c}.$$

□

FIGURE 3
ORTHOGONAL DECOMPOSITION

Definition 14.3. The vector

$$P_{\mathbf{b}}(\mathbf{a}) = (\frac{\mathbf{a} \cdot \mathbf{b}}{\mathbf{b} \cdot \mathbf{b}})\mathbf{b}$$

will be called the *orthogonal projection* of $\mathbf{a}$ on $\mathbf{b}$.

By the previous Proposition, another way to express the orthogonal decomposition of $\mathbf{a}$ into the sum of a component parallel to $\mathbf{b}$ and a component orthogonal to $\mathbf{b}$ is

$$\mathbf{a} = P_{\mathbf{b}}(\mathbf{a}) + (\mathbf{a} - P_{\mathbf{b}}(\mathbf{a})). \tag{14.6}$$

Example 14.4. Suppose $\mathbf{b}$ and $\mathbf{c}$ are any two nonzero orthogonal vectors in $\mathbb{R}^2$, so that $\mathbf{b} \cdot \mathbf{c} = 0$. I claim that any vector $\mathbf{a}$ orthogonal to $\mathbf{b}$ is a multiple of $\mathbf{c}$. Reason: if $\mathbf{a} \cdot \mathbf{b} = 0$, then $a_1 b_1 + a_2 b_2 = 0$. Assuming, for example, that $b_1 \neq 0$, then $c_2 \neq 0$ and

$$a_1 = -\frac{b_2}{b_1} a_2 = \frac{c_1}{c_2} a_2,$$

and so the claim follows from $a_1 = \dfrac{a_2}{c_2} c_1$ and $a_2 = \dfrac{a_2}{c_2} c_2$.

It follows that for any $\mathbf{a} \in \mathbb{R}^2$, there are scalars r and s so that $\mathbf{a} = r\mathbf{b} + s\mathbf{c}$. We can solve for r and s by using the dot product as before. For example, $\mathbf{a} \cdot \mathbf{b} = r\mathbf{b} \cdot \mathbf{b}$. Hence we can conclude that if $\mathbf{b} \neq \mathbf{0}$, then

$$r\mathbf{b} = P_{\mathbf{b}}(\mathbf{a}),$$

and similarly, if $\mathbf{c} \neq \mathbf{0}$, then

$$s\mathbf{c} = P_{\mathbf{c}}(\mathbf{a}).$$

Therefore, we have now proved a fundamental fact which we call the **projection formula** for $\mathbb{R}^2$.

Proposition 14.3. *If* $\mathbf{b}$ *and* $\mathbf{c}$ *are two non zero mutually orthogonal vectors in* $\mathbb{R}^2$*, then any vector* $\mathbf{a}$ *in* $\mathbb{R}^2$ *can be uniquely expressed as the sum of its projections on* $\mathbf{b}$ *and* $\mathbf{c}$*. In other words,*

$$\mathbf{a} = P_{\mathbf{b}}(\mathbf{a}) + P_{\mathbf{c}}(\mathbf{a}) = \left(\frac{\mathbf{a} \cdot \mathbf{b}}{\mathbf{b} \cdot \mathbf{b}}\right)\mathbf{b} + \left(\frac{\mathbf{a} \cdot \mathbf{c}}{\mathbf{c} \cdot \mathbf{c}}\right)\mathbf{c}. \tag{14.7}$$

Projections can be written much more simply if we bring in the notion of a unit vector. When $\mathbf{b} \neq \mathbf{0}$, the *unit vector along* $\mathbf{b}$ is defined to be the vector of length one given by the formula

$$\widehat{\mathbf{b}} = \frac{\mathbf{b}}{(\mathbf{b} \cdot \mathbf{b})^{1/2}} = \frac{\mathbf{b}}{|\mathbf{b}|}.$$

(Be sure to check that $\widehat{\mathbf{b}}$ is indeed of length one.) Unit vectors are also called *directions*. Keep in mind that a direction $\widehat{\mathbf{a}}$ exists only when $\mathbf{a} \neq \mathbf{0}$. It is obviously impossible to assigne a direction to the zero vector. If $\widehat{\mathbf{b}}$ and $\widehat{\mathbf{c}}$ are unit vectors, then the projection formula (14.7) takes the simpler form

$$\mathbf{a} = (\mathbf{a} \cdot \widehat{\mathbf{b}})\widehat{\mathbf{b}} + (\mathbf{a} \cdot \widehat{\mathbf{c}})\widehat{\mathbf{c}}. \tag{14.8}$$

Example 14.5. Let $\mathbf{b} = \begin{pmatrix} 3 \\ 4 \end{pmatrix}$ and $\mathbf{c} = \begin{pmatrix} 4 \\ -3 \end{pmatrix}$. Then $\widehat{\mathbf{b}} = \frac{1}{5}\begin{pmatrix} 3 \\ 4 \end{pmatrix}$ and $\widehat{\mathbf{c}} = \frac{1}{5}\begin{pmatrix} 4 \\ -3 \end{pmatrix}$. Let $\mathbf{a} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$. Thus, for example, $P_{\mathbf{b}}(\mathbf{a}) = \frac{7}{5}\begin{pmatrix} 3 \\ 4 \end{pmatrix}$, and $\mathbf{a} = \frac{7}{5}\begin{pmatrix} 3 \\ 4 \end{pmatrix} + \frac{1}{5}\begin{pmatrix} 4 \\ -3 \end{pmatrix}$.

14.6 The Cauchy-Schwartz Inequality and Cosines

If $\mathbf{a} = \mathbf{b} + \mathbf{c}$ is an orthogonal decomposition in $\mathbb{R}^3$ (which just means that $\mathbf{b} \cdot \mathbf{c} = 0$), then

$$|\mathbf{a}|^2 = |\mathbf{b}|^2 + |\mathbf{c}|^2.$$

This is known as Pythagoras's Theorem (see Exercise 4).

If we apply Pythagoras' Theorem to (14.6), for example, we get

$$|\mathbf{a}|^2 = |P_{\mathbf{b}}(\mathbf{a})|^2 + |\mathbf{a} - P_{\mathbf{b}}(\mathbf{a})|^2.$$

Hence,

$$|\mathbf{a}|^2 \geq |P_{\mathbf{b}}(\mathbf{a})|^2 = (\frac{\mathbf{a} \cdot \mathbf{b}}{\mathbf{b} \cdot \mathbf{b}})^2 |\mathbf{b}|^2 = \frac{(\mathbf{a} \cdot \mathbf{b})^2}{|\mathbf{b}|^2}.$$

Cross multiplying and taking square roots, we get a famous fact known as the Cauchy-Schwartz inequality.

Proposition 14.4. *For any* $\mathbf{a},\ \mathbf{b} \in \mathbb{R}^3$, *we have*

$$|\mathbf{a} \cdot \mathbf{b}| \leq |\mathbf{a}||\mathbf{b}|.$$

Moreover, if one of $\mathbf{a}$ *or* $\mathbf{b}$ *is nonzero, then equality holds if and only if* $\mathbf{a}$ *and* $\mathbf{b}$ *are collinear.*

Note that two vectors $\mathbf{a}$ and $\mathbf{b}$ are said to be *collinear* whenever one of them is a scalar multiple of the other. If either $\mathbf{a}$ and $\mathbf{b}$ is zero, then automatically they are collinear. If $\mathbf{b} \neq \mathbf{0}$ and the Cauchy-Schwartz inequality is an equality, then working backwards, one sees that $|\mathbf{a} - P_{\mathbf{b}}(\mathbf{a})|^2 = 0$, hence the validity of the second claim.

Cauchy-Schwartz says that for any two unit vectors $\widehat{\mathbf{a}}$ and $\widehat{\mathbf{b}}$, we have the inequality

$$-1 \leq \widehat{\mathbf{a}} \cdot \widehat{\mathbf{b}} \leq 1.$$

We can therefore define the angle θ between any two non zero vectors $\mathbf{a}$ and $\mathbf{b}$ in $\mathbb{R}^3$ by putting

$$\theta := \cos^{-1}(\hat{\mathbf{a}} \cdot \hat{\mathbf{b}}).$$

Note that we don't try to define the angle when either $\mathbf{a}$ or $\mathbf{b}$ is $\mathbf{0}$. (Recall that if $-1 \leq x \leq 1$, then $\cos^{-1} x$ is the unique angle θ such that $0 \leq \theta \leq \pi$ with $\cos\theta = x$.) With this definition, we have

$$\mathbf{a} \cdot \mathbf{b} = |\mathbf{a}||\mathbf{b}| \cos\theta \tag{14.9}$$

provided $\mathbf{a}$ and $\mathbf{b}$ are any two non-zero vectors in $\mathbb{R}^3$. Hence if $|\mathbf{a}| = |\mathbf{b}| = 1$, then the projection of $\mathbf{a}$ on $\mathbf{b}$ is

$$P_{\mathbf{b}}(\mathbf{a}) = (\cos\theta)\mathbf{b},$$

which justifies the definition. Thus another way of expressing the projection formula is

$$\widehat{\mathbf{a}} = (\cos\beta)\widehat{\mathbf{b}} + (\cos\gamma)\widehat{\mathbf{c}}.$$

Here β and γ are the angles between $\mathbf{a}$ and $\mathbf{b}$ and $\mathbf{c}$ respectively, and $\cos\beta$ and $\cos\gamma$ are called the corresponding direction cosines.

In the case of $\mathbb{R}^2$, there is already a notion of the angle between two vectors, defined in terms of arclength on a unit circle. Hence the expression $\mathbf{a} \cdot \mathbf{b} = |\mathbf{a}||\mathbf{b}| \cos\theta$ is often (especially in physics) taken as definition for the dot product, rather than as definition of angle, as we did here. However, defining $\mathbf{a} \cdot \mathbf{b}$ in this way has the disadvantage that it is not at all obvious that elementary properties such as the identity $(\mathbf{a}+\mathbf{b}) \cdot \mathbf{c} = \mathbf{a} \cdot \mathbf{c} + \mathbf{b} \cdot \mathbf{c}$ hold. Moreover, using this as a definition in $\mathbb{R}^3$ has the problem that the angle between two vectors must also be defined. The way to solve this is to use arclength, but this requires bringing in an unnecessary amount of machinery. On the other hand, the algebraic definition is easy to state and remember, and it works for any dimension. The Cauchy-Schwartz inequality, which is valid in $\mathbb{R}^3$, tells us that it possible two define the angle θ between $\mathbf{a}$ and $\mathbf{b}$ via (14.9) to be $\theta := \cos^{-1}(\hat{\mathbf{a}} \cdot \hat{\mathbf{b}})$.

14.7 Examples

Let us now consider a couple of typical applications of the ideas we just discussed.

Example 14.6. A film crew wants to shoot a car moving along a straight road with constant speed x km/hr. The camera will be moving along a straight track at y km/hr. The desired effect is that the car should appear to have exactly half the speed of the camera. At what angle to the road should the track be built?

Solution: Let θ be the angle between the road and the track. We need to find θ so that the projection of the velocity vector $\mathbf{v}_R$ of the car on the track is exactly half of the velocity vector $\mathbf{v}_T$ of the camera. Thus

$$\left(\frac{\mathbf{v}_R \cdot \mathbf{v}_T}{\mathbf{v}_T \cdot \mathbf{v}_T}\right)\mathbf{v}_T = \frac{1}{2}\mathbf{v}_T$$

and $\mathbf{v}_R \cdot \mathbf{v}_T = |\mathbf{v}_R||\mathbf{v}_T| \cos\theta$. Now $|\mathbf{v}_R| = x$ and $|\mathbf{v}_T| = y$ since speed is by definition the magnitude of velocity. Thus

$$\frac{xy}{y^2}\cos\theta = \frac{1}{2}$$

Consequently, $\cos\theta = y/2x$. In particular the camera's speed cannot exceed twice the car's speed.

Projection methods allow us to find formulas for distance. Consider the distance from a point to a plane P in $\mathbb{R}^3$. The problem becomes quite simple if we break it up into two cases. First, consider the case of a plane P through the origin, say with equation $ax + by + cz = 0$. Suppose $\mathbf{v}$ is an arbitrary vector in $\mathbb{R}^3$ whose distance to P is what we seek. Decompose $\mathbf{v}$ into orthogonal components as

$$\mathbf{v} = P_{\mathbf{n}}(\mathbf{v}) + (\mathbf{v} - P_{\mathbf{n}}(\mathbf{v})). \tag{14.10}$$

It's intuitively clear that the distance we're looking for is

$$d = |P_{\mathbf{n}}(\mathbf{v})| = |\mathbf{v} \cdot \mathbf{n}|/\sqrt{\mathbf{n} \cdot \mathbf{n}},$$

but let us check carefully. For one thing, we need to say what the distance from $\mathbf{v}$ to P is. We will assume it means the minimum value of $|\mathbf{v} - \mathbf{r}|$, where $\mathbf{r}$ is on P. To simplify notation, put $\mathbf{p} = P_{\mathbf{n}}(\mathbf{v})$ and $\mathbf{q} = \mathbf{v} - \mathbf{p}$. Since $\mathbf{v} = \mathbf{p} + \mathbf{q}$,

$$\mathbf{v} - \mathbf{r} = \mathbf{p} + \mathbf{q} - \mathbf{r}.$$

Since P contains the origin, $\mathbf{q} - \mathbf{r}$ lies on P since both $\mathbf{q}$ and $\mathbf{r}$ do. As $\mathbf{p}$ and $\mathbf{q} - \mathbf{r}$ are orthogonal,

$$|\mathbf{v} - \mathbf{r}|^2 = |\mathbf{p}|^2 + |\mathbf{q} - \mathbf{r}|^2.$$

But $\mathbf{p}$ is fixed, so $|\mathbf{v}-\mathbf{r}|^2$ is minimized when $|\mathbf{q}-\mathbf{r}|^2 = 0$. Thus $|\mathbf{v}-\mathbf{r}|^2 = |\mathbf{p}|^2$, and the distance $D(\mathbf{v}, P)$ from $\mathbf{v}$ to P is indeed

$$D(\mathbf{v}, P) = |\mathbf{p}| = \frac{|\mathbf{v} \cdot \mathbf{n}|}{(\mathbf{n} \cdot \mathbf{n})^{\frac{1}{2}}} = |\mathbf{v} \cdot \widehat{\mathbf{n}}|.$$

Also, the point on P nearest $\mathbf{v}$ is $\mathbf{q}$. If $\mathbf{v} = \begin{pmatrix} r \\ s \\ t \end{pmatrix}$ and $\mathbf{n} = \begin{pmatrix} a \\ b \\ c \end{pmatrix}$, then the distance is

$$D(\mathbf{v}, P) = \frac{|ar + bs + ct|}{\sqrt{a^2 + b^2 + c^2}}.$$

We now attack the general problem by reducing it to the first case. We want to find the distance $D(\mathbf{v}, Q)$ from $\mathbf{v}$ to an arbitrary plane Q in $\mathbb{R}^3$. Suppose the equation of Q is $ax + by + cz = d$, and let $\mathbf{c}$ be a vector on Q. I claim that the distance from $\mathbf{v}$ to Q is the same as the distance from $\mathbf{v} - \mathbf{c}$ to the plane P parallel to Q through the origin, i.e. the plane $ax + by + cz = 0$. Indeed, we already showed that every vector on Q has the form $\mathbf{w} + \mathbf{c}$ where $\mathbf{w}$ is on P. Thus let $\mathbf{r}$ be the vector on Q nearest $\mathbf{v}$. Since $d(\mathbf{v}, \mathbf{r}) = |\mathbf{v} - \mathbf{r}|$, it follows easily from $\mathbf{r} = \mathbf{w} + \mathbf{c}$ that $d(\mathbf{v}, \mathbf{r}) = d(\mathbf{v} - \mathbf{c}, \mathbf{w})$. Hence the problem amounts to minimizing $d(\mathbf{v} - \mathbf{c}, \mathbf{w})$ for $\mathbf{w} \in P$, which we already solved. Thus

$$D(\mathbf{v}, Q) = |(\mathbf{v} - \mathbf{c}) \cdot \widehat{\mathbf{n}}|,$$

which reduces to the formula

$$D(\mathbf{v}, Q) = \frac{|ar + bs + ct - d|}{\sqrt{a^2 + b^2 + c^2}},$$

since

$$\mathbf{c} \cdot \widehat{\mathbf{n}} = \frac{\mathbf{c} \cdot \mathbf{n}}{(\mathbf{n} \cdot \mathbf{n})^{\frac{1}{2}}} = \frac{d}{\sqrt{a^2 + b^2 + c^2}}.$$

In summary, we have

Proposition 14.5. *Let Q be the plane in $\mathbb{R}^3$ defined by $ax + by + cz = d$, and let $\mathbf{v}$ be any vector in $\mathbb{R}^3$, possibly lying on Q. Let $D(\mathbf{v}, Q)$ be the distance from $\mathbf{v}$ to Q. Then*

$$D(\mathbf{v}, Q) = \frac{|ar + bs + ct - d|}{\sqrt{a^2 + b^2 + c^2}}.$$

Exercises

14.8 The Cross Product

14.8.1 The Basic Definition

Sometimes one needs a vector $\mathbf{c}$ orthogonal to a pair of noncollinear vectors $\mathbf{a}$ and $\mathbf{b}$. This is provided courtesy of the cross product $\mathbf{a} \times \mathbf{b}$. This is defined geometrically as follows. Let P denote the unique plane through the origin containing both $\mathbf{a}$ and $\mathbf{b}$, and let $\mathbf{n}$ be the choice of unit vector normal to P so that the thumb, index finger and middle finger of your right hand can be lined up with the three vectors $\mathbf{a}, \mathbf{b}$ and $\mathbf{n}$ without dislocating any joints. In this case we call $(\mathbf{a}, \mathbf{b}, \mathbf{n})$ a *right handed triple*. (Otherwise, it's a left handed triple.) Let θ be the angle between $\mathbf{a}$ and $\mathbf{b}$, so $0 < \theta < \pi$. Then we put

$$\mathbf{a} \times \mathbf{b} = |\mathbf{a}||\mathbf{b}| \sin\theta \mathbf{n}. \tag{14.11}$$

If $\mathbf{a}$ and $\mathbf{b}$ are collinear, we set $\mathbf{a} \times \mathbf{b} = \mathbf{0}$.

While this definition is very elegant, and is useful for revealing the geometric properties of the cross product, it doesn't help much in computing $\mathbf{a} \times \mathbf{b}$. But if $\mathbf{a} \cdot \mathbf{b} = 0$, (since $\cos\theta = 0$) one sees that $|\mathbf{a} \times \mathbf{b}| = |\mathbf{a}||\mathbf{b}|$.

To see a couple of examples, note that $(\mathbf{i}, \mathbf{j}, \mathbf{k})$ and $(\mathbf{i}, -\mathbf{j}, -\mathbf{k})$ both are right handed triples, but $(\mathbf{i}, -\mathbf{j}, \mathbf{k})$ and $(\mathbf{j}, \mathbf{i}, \mathbf{k})$ are left handed. Thus $\mathbf{i} \times \mathbf{j} = \mathbf{k}$, while $\mathbf{j} \times \mathbf{i} = -\mathbf{k}$. Similarly, $\mathbf{j} \times \mathbf{k} = \mathbf{i}$ and $\mathbf{k} \times \mathbf{j} = -\mathbf{i}$. In fact, these examples point out two of the general properties of the cross product:

$$\mathbf{a} \times \mathbf{b} = -\mathbf{b} \times \mathbf{a},$$

and

$$(-\mathbf{a}) \times \mathbf{b} = \mathbf{a} \times (-\mathbf{b}) = -(\mathbf{a} \times \mathbf{b}).$$

The question is whether or not the cross product is computable. In fact, the answer to this is yes. First, let us make a temporary definition. If $\mathbf{a}, \mathbf{b} \in \mathbb{R}^3$, put

$$\mathbf{a} \wedge \mathbf{b} = \begin{pmatrix} a_2 b_3 - a_3 b_2 \\ a_3 b_1 - a_1 b_3 \\ a_1 b_2 - a_2 b_1 \end{pmatrix}.$$

Notice that $\mathbf{a} \wedge \mathbf{b}$ is defined without any restrictions on $\mathbf{a}$ and $\mathbf{b}$. It is not hard to verify by direct computation that $\mathbf{a} \wedge \mathbf{b}$ is orthogonal to both $\mathbf{a}$ and $\mathbf{b}$, so $\mathbf{a} \wedge \mathbf{b} = r(\mathbf{a} \times \mathbf{b})$ for some $r \in \mathbb{R}$.

The key fact is the following

Proposition 14.6. *For all* $\mathbf{a}$ *and* $\mathbf{b}$ *in* $\mathbb{R}^3$,

$$\mathbf{a} \times \mathbf{b} = \mathbf{a} \wedge \mathbf{b}.$$

This takes care of the computability problem since $\mathbf{a} \wedge \mathbf{b}$ is easily computed. An outline the proof goes as follows. First, note the following identity:

$$|\mathbf{a} \wedge \mathbf{b}|^2 + (\mathbf{a} \cdot \mathbf{b})^2 = (|\mathbf{a}||\mathbf{b}|)^2. \tag{14.12}$$

The proof is just a calculation, and we will omit it. Since $\mathbf{a} \cdot \mathbf{b} = |\mathbf{a}||\mathbf{b}| \cos\theta$, and since $\sin\theta \geq 0$, we deduce that

$$|\mathbf{a} \wedge \mathbf{b}| = |\mathbf{a}||\mathbf{b}| \sin\theta. \tag{14.13}$$

It follows that $\mathbf{a} \wedge \mathbf{b} = \pm|\mathbf{a}||\mathbf{b}| \sin\theta \mathbf{n}$. The fact that the sign is $+$ proven by showing that

$$(\mathbf{a} \wedge \mathbf{b}) \cdot \mathbf{n} > 0.$$

The proof of this step is a little tedious and we will omit it.

14.8.2 Some further properties

Before giving applications, we let us write down the algebraic properties of the cross product.

Proposition 14.7. *Suppose* $\mathbf{a}, \mathbf{b}, \mathbf{c} \in \mathbb{R}^3$. *Then:*

(*i*) $\mathbf{a} \times \mathbf{b} = -\mathbf{b} \times \mathbf{a}$,

(*ii*) $(\mathbf{a} + \mathbf{b}) \times \mathbf{c} = \mathbf{a} \times \mathbf{c} + \mathbf{b} \times \mathbf{c}$, *and*

(*iii*) *for any* $r \in \mathbb{R}$, $(r\mathbf{a}) \times \mathbf{b} = \mathbf{a} \times (r\mathbf{b}) = r(\mathbf{a} \times \mathbf{b})$.

Proof. The first and third identities are obvious from the original definition. The second identity, which says that the cross product is distributive, is not at all obvious from the definition. On the other hand, it is easy to check directly that

$$(\mathbf{a} + \mathbf{b}) \wedge \mathbf{c} = \mathbf{a} \wedge \mathbf{c} + \mathbf{b} \wedge \mathbf{c},$$

so (*ii*) has to hold also since $\mathbf{a} \times \mathbf{b} = \mathbf{a} \wedge \mathbf{b}$. □

14.8.3 Examples and applications

The first application is to use the cross product to find a normal $\mathbf{n}$ to the plane P through $\mathbf{p}, \mathbf{q}, \mathbf{r}$, assuming they don't all lie on a line. Once we have $\mathbf{n}$, it is easy to find the equation of P. We begin by considering the plane Q through the origin parallel to P. First put $\mathbf{a} = \mathbf{q} - \mathbf{p}$ and $\mathbf{b} = \mathbf{r} - \mathbf{p}$. Then $\mathbf{a}, \mathbf{b} \in Q$, so we can put $\mathbf{n} = \mathbf{a} \times \mathbf{b}$. Suppose

$$\mathbf{n} = \begin{pmatrix} a \\ b \\ c \end{pmatrix} \quad \text{and} \quad \mathbf{p} = \begin{pmatrix} p_1 \\ p_2 \\ p_3 \end{pmatrix}.$$

Then the equation of Q is $ax+by+cz = 0$, and the equation of P is obtained by noting that

$$\mathbf{n} \cdot (\begin{pmatrix} x \\ y \\ z \end{pmatrix} - \begin{pmatrix} p_1 \\ p_2 \\ p_3 \end{pmatrix}) = 0.$$

Therefore the equation of P is

$$ax + by + cz = ap_1 + bp_2 + cp_3.$$

Example 14.7. Let's find an equation for the plane in $\mathbb{R}^3$ through $\begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}$, $\begin{pmatrix} 0 \\ 3 \\ -1 \end{pmatrix}$ and $\begin{pmatrix} 2 \\ 0 \\ 0 \end{pmatrix}$. To find a normal, compute $\begin{pmatrix} -1 \\ 2 \\ 1 \end{pmatrix} \times \begin{pmatrix} -2 \\ 3 \\ -1 \end{pmatrix} = \begin{pmatrix} -5 \\ -3 \\ 1 \end{pmatrix}$. Thus the plane has equation $-5x - 3y + z = -10$.

The cross product also gives the area formula for a parallelogram.

Proposition 14.8. *Let* $\mathbf{a}$ *and* $\mathbf{b}$ *be two noncollinear vectors in* $\mathbb{R}^3$. *Then the area of the parallelogram spanned by* $\mathbf{a}$ *and* $\mathbf{b}$ *is* $|\mathbf{a} \times \mathbf{b}|$.

We can extend the area formula to 3-dimensional (i.e. solid) parallelograms. Any three noncoplanar vectors $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$ in $\mathbb{R}^3$ determine a solid parallelogram called a parallelepiped. This parallelepiped $\mathcal{P}$ can be explicitly defined as

$$\mathcal{P} = \{r\mathbf{a} + s\mathbf{b} + t\mathbf{c} \mid 0 \leq r, s, t \leq 1\}.$$

For example, the parallelepiped spanned by $\mathbf{i}$, $\mathbf{j}$ and $\mathbf{k}$ is the unit cube in $\mathbb{R}^3$ with vertices at $\mathbf{0}$, $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$, $\mathbf{i}+\mathbf{j}$, $\mathbf{i}+\mathbf{k}$, $\mathbf{j}+\mathbf{k}$ and $\mathbf{i}+\mathbf{j}+\mathbf{k}$. A parallelepiped has 8 vertices and 6 sides which are pairwise parallel.

To get the volume formula, we introduce the *triple product* $\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c})$ of $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$.

Proposition 14.9. *Let* $\mathbf{a}$, $\mathbf{b}$ *and* $\mathbf{c}$ *be three noncoplanar vectors in* $\mathbb{R}^3$. *Then the volume of the parallelepiped they span is* $|\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c})|$.

Proof. We leave this as a worthwhile exercise. □

By the definition of the triple product,

$$\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) = a_1(b_2c_3 - b_3c_2) - a_2(b_3c_1 - b_1c_3) + a_3(b_1c_2 - b_2c_1).$$

The right hand side of this equation is by definition the 3×3 determinant

$$\det \begin{pmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{pmatrix}.$$

Example 14.8. We next find the formula for the distance between two lines. Consider two lines ℓ_1 and ℓ_2 in $\mathbb{R}^3$ parameterized as $\mathbf{a}_1 + t\mathbf{b}_1$ and $\mathbf{a}_2 + t\mathbf{b}_2$ respectively. We claim that the distance between ℓ_1 and ℓ_2 is

$$d = \frac{|(\mathbf{a}_1 - \mathbf{a}_2) \cdot (\mathbf{b}_1 \times \mathbf{b}_2)|}{|\mathbf{b}_1 \times \mathbf{b}_2|}.$$

This formula is somewhat surprising because it says that one can choose any two initial points $\mathbf{a}_1$ and $\mathbf{a}_2$ to compute d. It's intuitively clear that $\mathbf{b}_1 \times \mathbf{b}_2$ is involved since $\mathbf{b}_1 \times \mathbf{b}_2$ is orthogonal to the directions of both lines. But one way to see this concretely is to take a tube of radius r centred along ℓ_1 and expand r until the tube touches ℓ_2. The point $\mathbf{v}_2$ of tangency on ℓ_2 and the center $\mathbf{v}_1$ on ℓ_1 of the disc (orthogonal to ℓ_1) touching ℓ_2 give the two points so that $d = d(\mathbf{v}_1, \mathbf{v}_2)$, and, by construction, $\mathbf{v}_1 - \mathbf{v}_2$ is parallel to $\mathbf{b}_1 \times \mathbf{b}_2$. Now let $\mathbf{v}_i = \mathbf{a}_i + t_i\mathbf{b}_i$ for $i = 1, 2$, and denote the unit vector in the direction of $\mathbf{b}_1 \times \mathbf{b}_2$ by $\widehat{\mathbf{u}}$. Then

$$\begin{aligned} d &= |\mathbf{v}_1 - \mathbf{v}_2| \\ &= (\mathbf{v}_1 - \mathbf{v}_2) \cdot \frac{(\mathbf{v}_1 - \mathbf{v}_2)}{|\mathbf{v}_1 - \mathbf{v}_2|} \\ &= |(\mathbf{v}_1 - \mathbf{v}_2) \cdot \widehat{\mathbf{u}}| \\ &= |(\mathbf{a}_1 - \mathbf{a}_2 + t_1\mathbf{b}_1 - t_2\mathbf{b}_2) \cdot \widehat{\mathbf{u}}| \\ &= |(\mathbf{a}_1 - \mathbf{a}_2) \cdot \widehat{\mathbf{u}}|. \end{aligned}$$

The last equality is due to the fact that $\mathbf{b}_1 \times \mathbf{b}_2$ is orthogonal to $t_1\mathbf{b}_1 - t_2\mathbf{b}_2$ plus the fact that the dot product is distributive. This is the formula we sought.

Exercises

Note that $\begin{pmatrix} a \\ b \\ c \end{pmatrix}$ is denoted throughout by $(a, b, c)^T$.

Exercise 14.1. Express the line $ax + by = c$ in $\mathbb{R}^2$ in parametric form.

Exercise 14.2. Express the line with vector form $(x, y)^T = (1, -1)^T + t(2, 3)^T$ in the form $ax + by = c$.

Exercise 14.3. Find the line through the points **a** and **b** in the following cases:

(i) $\mathbf{a} = (1, 1, -3)^T$ and $\mathbf{b} = (6, 0, 2)^T$, and

(ii) $\mathbf{a} = (1, 1, -3, 4)^T$ and $\mathbf{b} = (6, 0, 2, -3)^T$.

Exercise 14.4. Find the line of intersection of the planes $3x - y + z = 0$ and $x - y - z = 1$ in parametric form.

Exercise 14.5. Do the following:

(a) Find the equation in vector form of the line through $(1, -2, 0)^T$ parallel to $(3, 1, 9)^T$.

(b) Find the plane perpendicular to the line of part (a) passing through $(0, 0, 0)^T$.

(c) At what point does the line of part (a) meet the plane of part (b)?

Exercise 14.6. Determine whether or not the lines $(x, y, z)^T = (1, 2, 1)^T + t(1, 0, 2)^T$ and $(x, y, z)^T = (2, 2, -1)^T + t(1, 1, 0)^T$ intersect.

Exercise 14.7. Consider any two lines in $\mathbb{R}^3$. Suppose I offer to bet you they don't intersect. Do you take the bet or refuse it? What would you do if you knew the lines were in a plane?

Exercise 14.8. Find an equation for the plane in $\mathbb{R}^3$ through the points $(6, 1, 0)^T$, $(1, 0, 1)^T$ and $(3, 1, 1)^T$

Exercise 14.9. Compute the intersection of the line through $(3, -1, 1)^T$ and $(1, 0, 2)^T$ with the plane $ax + by + cz = d$ when

(i) $a = b = c = 1$, $d = 2$,

(ii) $a = b = c = 1$ and $d = 3$.

Exercise 14.10. Find the distance from the point $(1,1,1)^T$ to

(i) the plane $x+y+z=1$, and

(ii) the plane $x-2y+z=0$.

Exercise 14.11. Find the orthogonal decomposition $(1,1,1)^T = \mathbf{a}+\mathbf{b}$, where $\mathbf{a}$ lies on the plane P with equation $2x+y+2z=0$ and $\mathbf{a}\cdot\mathbf{b}=0$. What is the orthogonal projection of $(1,1,1)^T$ on P?

Exercise 14.12. Here's another bet. Suppose you have two planes in $\mathbb{R}^3$ and I have one. Furthermore, your planes meet in a line. I'll bet that all three of our planes meet. Do you take this bet or refuse it. How would you bet if the planes were all in $\mathbb{R}^4$ instead of $\mathbb{R}^3$?

Exercise 14.13. Show that two lines in $\mathbb{R}^3$ which meet in two points coincide.

Exercise 14.14. Verify that the union of the lines $\mathbf{x}=\mathbf{a}+t\mathbf{b}$, where $\mathbf{b}$ is fixed but $\mathbf{a}$ is arbitrary is $\mathbb{R}^3$. Also show that two of these lines are the same or have no points in common.

Exercise 14.15. Verify the Parallelogram Law (in $\mathbb{R}^3$) by computing where the line through $\mathbf{a}$ parallel to $\mathbf{b}$ meets the line through $\mathbf{b}$ parallel to $\mathbf{a}$.

Exercise 14.16. Verify the four properties of the dot product on $\mathbb{R}^3$.

Exercise 14.17. Verify the assertion that $\mathbf{b}\cdot\mathbf{c}=0$ in the proof of Theorem 14.2.

Exercise 14.18. Prove the second statement in the Cauchy-Schwartz inequality that $\mathbf{a}$ and $\mathbf{b}$ are collinear if and only if $|\mathbf{a}\cdot\mathbf{b}| = |\mathbf{a}||\mathbf{b}|$.

Exercise 14.19. A nice application of Cauchy-Schwartz is that if $\mathbf{a}$ and $\mathbf{b}$ are unit vectors in $\mathbb{R}^3$ such that $\mathbf{a}\cdot\mathbf{b}=1$, then $\mathbf{a}=\mathbf{b}$. Prove this.

Exercise 14.20. Show that $P_{\mathbf{b}}(r\mathbf{x}+s\mathbf{y}) = rP_{\mathbf{b}}(\mathbf{x})+sP_{\mathbf{b}}(\mathbf{y})$ for all $\mathbf{x},\mathbf{y}\in\mathbb{R}^3$ and $r,s\in\mathbb{R}$. Also show that $P_{\mathbf{b}}(\mathbf{x})\cdot\mathbf{y} = \mathbf{x}\cdot P_{\mathbf{b}}(\mathbf{y})$.

Exercise 14.21. Prove the vector version of Pythagoras's Theorem. If $\mathbf{c}=\mathbf{a}+\mathbf{b}$ and $\mathbf{a}\cdot\mathbf{b}=0$, then $|\mathbf{c}|^2 = |\mathbf{a}|^2+|\mathbf{b}|^2$.

Exercise 14.22. Show that for any $\mathbf{a}$ and $\mathbf{b}$ in $\mathbb{R}^3$,

$$|\mathbf{a}+\mathbf{b}|^2 - |\mathbf{a}-\mathbf{b}|^2 = 4\mathbf{a}\cdot\mathbf{b}.$$

Exercise 14.23. Use the formula of the previous problem to prove Proposition 2.1, that is to show that $|\mathbf{a}+\mathbf{b}| = |\mathbf{a}-\mathbf{b}|$ if and only if $\mathbf{a}\cdot\mathbf{b}=0$.

Exercise 14.24. Prove the law of cosines: If a triangle has sides with lengths $a,\ b,\ c$ and θ is the angle between the sides of lengths a and b, then $c^2 = a^2 + b^2 - 2ab\cos\theta$. (Hint: Consider $\mathbf{c} = \mathbf{b} - \mathbf{a}$.)

Exercise 14.25. Another way to motivate the definition of the projection $P_{\mathbf{b}}(\mathbf{a})$ is to find the minimum of $|\mathbf{a} - t\mathbf{b}|^2$. Find the minimum using calculus and interpret the result.

Exercise 14.26. Orthogonally decompose the vector $(1, 2, 2)$ in $\mathbb{R}^3$ as $\mathbf{p}+\mathbf{q}$ where $\mathbf{p}$ is required to be a multiple of $(3, 1, 2)$.

Exercise 14.27. Use orthogonal projection to find the vector on the line $3x + y = 0$ which is nearest to $(1, 2)$. Also, find the nearest point.

Exercise 14.28. How can you modify the method of orthogonal projection to find the vector on the line $3x + y = 2$ which is nearest to $(1, -2)$?

Exercise 14.29. Using the cross product, find the plane through the origin that contains the line through $(1, -2, 0)^T$ parallel to $(3, 1, 9)^T$.

Exercise 14.30. Using the cross product, find

(a) the line of intersection of the planes $3x+2y-z = 0$ and $4x+5y+z = 0$, and

(b) the line of intersection of the planes $3x+2y-z = 2$ and $4x+5y+z = 1$.

Exercise 14.31. Is $\mathbf{x} \times \mathbf{y}$ orthogonal to $2\mathbf{x} - 3\mathbf{y}$? Generalize this.

Exercise 14.32. Find the distance from $(1, 2, 1)^T$ to the plane containing $1, 3, 4)^T$, $(2, -2, -2)^T$, and $(7, 0, 1)^T$ using the cross product.

Exercise 14.33. Formulate a definition for the angle between two planes in $\mathbb{R}^3$. (Suggestion: consider their normals.)

Exercise 14.34. Find the distance from the line $\mathbf{x} = (1, 2, 3)^T + t(2, 3, -1)^T$ to the origin in two ways:

(i) using projections, and

(ii) using calculus, by setting up a minimization problem.

Exercise 14.35. Find the distance from the point $(1, 1, 1)^T$ to the line $x = 2 + t, y = 1 - t, z = 3 + 2t$,

Exercise 14.36. Show that in $\mathbb{R}^3$, the distance from a point $\mathbf{p}$ to a line $\mathbf{x} = \mathbf{a} + t\mathbf{b}$ can be expressed in the form

$$d = \frac{|(\mathbf{p} - \mathbf{a}) \times \mathbf{b}|}{|\mathbf{b}|}.$$

Exercise 14.37. Prove the identity

$$|\mathbf{a} \times \mathbf{b}|^2 + (\mathbf{a} \cdot \mathbf{b})^2 = (|\mathbf{a}||\mathbf{b}|)^2.$$

Deduce that if $\mathbf{a}$ and $\mathbf{b}$ are unit vectors, then

$$|\mathbf{a} \times \mathbf{b}|^2 + (\mathbf{a} \cdot \mathbf{b})^2 = 1.$$

Exercise 14.38. Show that

$$\mathbf{a} \times (\mathbf{b} \times \mathbf{c}) = (\mathbf{a} \cdot \mathbf{c})\mathbf{b} - (\mathbf{a} \cdot \mathbf{b})\mathbf{c}.$$

Deduce from this $\mathbf{a} \times (\mathbf{b} \times \mathbf{c})$ is not necessarily equal to $(\mathbf{a} \times \mathbf{b}) \times \mathbf{c}$. In fact, can you say when they are equal?