Essential Linear Algebra

Joel G. Broida
University of Colorado, Boulder

Copyright ©2009 by Joel G. Broida. All rights reserved.

August 17, 2009

Preface

This text grew out of the need to teach real (but practical and useful) linear algebra to students with a wide range of backgrounds, desires and goals. It is meant to provide a solid foundation in modern linear algebra as used by mathematicians, physicists and engineers. While anyone reading this book has probably had at least a passing exposure to the concepts of vector spaces and matrices, we do not assume any prior rigorous coursework on the subject. In the sense that the present text is a beginning treatment of linear algebra, it is admittedly not an extremely elementary treatment, and is probably best suited for an upper division course at most American universities. In other words, we assume the student has had at least three semesters of calculus, and hence possesses a certain amount of that intangible quantity called "mathematical maturity."

This book is not is a text on "applied linear algebra" or numerical methods. We feel that it is impossible to cover adequately in a single (or even two) semester course all of the linear algebra theory that students should know nowadays, as well as numerical methods. To try to cover both does a disservice to both. It is somewhat like trying to teach linear algebra and differential equations in a single course – in our experience the students end up not knowing either subject very well. We realize that there are only so many courses that the student has the time to take, but it is better to have a firm understanding of the basics rather than a cursory overview of many topics. If the student learns the foundations well, then specializing to systems of differential equations or other numerical techniques should be an easy transition.

As we just stated, it is our belief that many of the newer books on linear algebra try to cover too much material in the sense that they treat both theory as well as numerous applications. Unfortunately, the applications themselves aren't covered in sufficient detail for the student to learn them, and they may lie in a subject far removed from the interests of any one particular student. The net result is that the various applications become somewhat of a waste of time, and amount to blindly plugging into equations. Furthermore, covering these applications detracts from the amount of time necessary to cover the foundational material, all the more important since linear algebra is rarely more than a one semester course. As a result, most students finish the semester without having a real understanding of the fundamentals, and don't really understand how linear algebra aids in numerical calculations. Our opinion is that it is far better to thoroughly cover the fundamentals, because this then enables the student to later pick up a book on a more specialized subject or application and already understand the underlying theory.

For example, physics students learn about Clebsch-Gordon coefficients when studying the addition of angular momentum in quantum mechanics courses. This gives the impression that Clebsch-Gordon coefficients are somehow unique to quantum mechanics, whereas in reality they are simply the entries in the unitary transition matrix that represents a change of basis in a finite-dimensional space. Understanding this makes it far easier to grasp the concepts of just what is going on. Another example is the diagonalization of the inertia tensor in classical mechanics. The student should realize that finding the principal moments and principal axes of a solid object is just a straightforward application of finding the eigenvalues and eigenvectors of a real symmetric matrix.

The point we are trying to emphasize is that the student that understands the general mathematical framework will see much more clearly what is really going on in applications that are covered in many varied courses in engineering, physics and mathematics. By understanding the underlying mathematics thoroughly, it will make it much easier for the student to see how many apparently unrelated topics are in fact completely equivalent problems in different guises.

There are a number of ways in which this text differs from most, if not all, other linear algebra books on the market. We begin in Chapter 1 with a treatment of vector spaces rather than matrices, and there are at least two reasons for this. First, the concept of a vector space is used in many courses much more than the concept of a matrix is used, and the student will likely need to understand vector spaces as used in these other courses early in the semester. And second, various properties of matrices (such as the rank) developed in Chapter 2 are based on vector spaces. It seems to us that it is better to treat matrices after the student learns about vector spaces, and not have to jump back and forth between the topics. It is in Chapter 1 that we treat both the direct sum of vector spaces and define general inner product spaces. We have found that students don't have a problem with the elementary "dot product" that they learned in high school, but the concept of an abstract inner product causes a lot of confusion, as does even the more general bracket notation for the dot product.

The first really major difference is in our treatment of determinants given in Chapter 3. While definitely useful in certain situations, determinants in and of themselves aren't as important as they once were. However, by developing them from the standpoint of permutations using the Levi-Civita symbol, the student gains an extremely important calculational tool that appears in a wide variety of circumstances. The ability to work with this notation greatly facilitates an understanding of much of modern differential geometry, which now finds applications in engineering as well as many topics in modern physics, such as general relativity, quantum gravity and strings. Understanding this formalism will be particularly beneficial to those students who go on to graduate school in engineering or the physical sciences.

The second major difference is related to the first. In Chapter 8 we include a reasonably complete treatment of the fundamentals of multilinear mappings, tensors and exterior forms. While this is usually treated in books on differential geometry, it is clear that the underlying fundamentals do not depend on the concept of a manifold. As a result, after learning what is in this book, the student should have no trouble specializing to the case of tangent spaces and differential forms. And even without the more advanced applications of differential geometry, the basic concept of a tensor is used not only in classical physics (for example, the inertia tensor and the electromagnetic field tensor), but also in engineering (where second rank tensors are frequently called "dyadics.").

In Chapter 8 we also give a reasonably complete treatment of the volume of a parallelepiped in $\mathbb{R}^n$, and how this volume transforms under linear transformations. This also leads to the rather abstract concept of "orientation" which we try to motivate and explain in great detail. The chapter ends with a discussion of the metric tensor, and shows how the usual vector gradient is related to the differential of a function, working out the case of spherical polar coordinates in detail.

Otherwise, most of the subjects we treat are fairly standard, although our treatment is somewhat more detailed than most. Chapter 4 contains a careful but practical treatment of linear transformations and their matrix representations. We have tried to emphasize that the ith column of a matrix representation is just the image of the ith basis vector. And of course this then leads to a discussion of how the matrix representations change under a change of basis.

In Chapter 5 we give an overview of polynomials and their roots, emphasizing the distinction between algebraic and geometric multiplicities. From there we proceed to our treatment of eigenvalues and eigenvectors. Because they are so important in many applications, we give a careful discussion of invariant subspaces, and show how diagonalizing a linear transformation amounts to finding a new basis in which the matrix representation of a linear operator is the direct sum of the invariant eigenspaces. This material is directly applicable to physical applications such as quantum mechanics as well as more mathematical applications such as the representations of finite groups. Indeed, the famous Schur's lemmas are nothing more than very simple applications of the concept of invariant subspaces. And also in this chapter, we prove the extremely useful result (the Schur canonical form) that any complex matrix can be put into upper triangular form. This also easily leads to a proof that any normal matrix can be diagonalized by a unitary transformation.

Linear operators are treated in Chapter 6 which begins with a careful development of the operator adjoint. From this point, we give a more detailed treatment of normal operators in general, and hermitian (or orthogonal) operators in particular. We also discuss projections, the spectral theorem, positive operators, and the matrix exponential series.

Bilinear forms are covered in Chapter 7, and much of the chapter deals with the diagonalization of bilinear forms. In fact, we treat the simultaneous diagonalization of two real symmetric bilinear forms in quite a bit of detail. This is an interesting subject because there is more than one way to treat the problem, and this ultimately leads to a much better understanding of all of the approaches as well as clarifying what was really done when we covered the standard eigenvalue problem. As a specific example, we give a thorough and detailed treatment of coupled small oscillations. We develop the theory from the conventional standpoint of differential equations (Lagrange's equations for coupled oscillators), and then show how this is really an eigenvalue problem where simultaneously diagonalizing the kinetic and potential energy terms in the Lagrangian gives the general solution as a linear combination of eigenvectors with coefficients that are just the normal coordinates of the system.

Finally, we include an appendix that provides an overview of mappings, the real and complex numbers, and the process of mathematical induction. While most students should be familiar with this material, it is there as an easy reference for those who may need it.

Contents

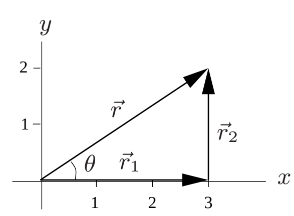
Chapter 1

Vector Spaces

Linear algebra is essentially a study of various transformation properties defined on a vector space, and hence it is only natural that we carefully define vector spaces. This chapter therefore presents a fairly rigorous development of (finite-dimensional) vector spaces, and a discussion of their most important fundamental properties. While many linear algebra texts begin with a treatment of matrices, we choose to start with vectors and vector spaces. Our reason for this is that any discussion of matrices quickly leads to defining row and column spaces, and hence an understanding of vector spaces is needed in order to properly characterize matrices.

1.1 Motivation

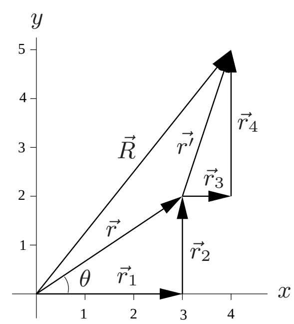
Basically, the general definition of a vector space is simply an axiomatization of the elementary properties of ordinary three-dimensional Euclidean space that you should already be familiar with. In order to motivate this definition we consider motion in the two-dimensional plane. Let us start at what we call the 'origin,' and let the positive x-axis be straight ahead and the y-axis to our left. If we walk three feet straight ahead and then two feet to the left, we can represent this motion graphically as follows. Let $\vec{r}_1$ represent our forward motion, let $\vec{r}_2$ represent our turn to the left, and let $\vec{r}$ represent our *net* movement from the origin to the end of our walk. Graphically we may display this as shown below.



We describe each segment of our walk by a pair of numbers (x, y) that describes how far we travel in each direction relative to the starting point of each segment. Using this notation we have the **vectors** $\vec{r}_1 = (3, 0)$ and $\vec{r}_2 = (0, 2)$, and our net motion can be described by the vector $\vec{r} = (3, 2)$. The number 3 is called the x-**coordinate** of $\vec{r}$, and the number 2 is called the y-**coordinate**. Since this is only motivation, we don't have to be too careful to define all of our manipulations precisely, and we note that we can obtain this net motion by adding the displacements in the x and y directions independently: $\vec{r}_1 + \vec{r}_2 = (3, 0) + (0, 2) = (3, 2) = \vec{r}$. In other words, $\vec{r} = \vec{r}_1 + \vec{r}_2$.

The distance we end up from the origin is the length (or **norm**) of the vector and is just $\sqrt{3^2 + 2^2} = \sqrt{13}$ as given by the Pythagorean theorem. However, in order to describe our exact location, knowing our distance from the origin obviously isn't enough — this would only specify that we are somewhere on a circle of radius $\sqrt{13}$. In order to precisely locate our position, we also need to give a direction relative to some reference direction which we take to be the x-axis. Thus our direction is given by the angle θ defined by $\tan\theta = 2/3$. This is why in elementary courses a vector is sometimes loosely described by saying it is an object that has length and direction.

Now suppose we take another walk from our present position. We travel first a distance in the x direction given by $\vec{r}_3 = (1, 0)$ and then a distance in the y direction given by $\vec{r}_4 = (0, 3)$. Relative to the start of this trip we are now at the location $\vec{r'} = \vec{r}_3 + \vec{r}_4 = (1, 3)$, and we have the following total path taken:



Observe that our final position is given by the vector $\vec{R} = (4, 5) = (3+1, 2+3) = (3, 2) + (1, 3) = \vec{r} + \vec{r'}$, and hence we see that arbitrary vectors in the plane can be added together to obtain another vector in the plane. It should also be clear that if we repeat the entire trip again, then we will be at the point $\vec{R} + \vec{R} = 2\vec{R} = (3, 0) + (0, 2) + (1, 0) + (0, 3) + (3, 0) + (0, 2) + (1, 0) + (0, 3) = 2[(3, 0) + (0, 2)] + 2[(1, 0) + (0, 3)] = 2\vec{r} + 2\vec{r'}$ and hence $2\vec{R} = 2(\vec{r} + \vec{r'}) = 2\vec{r} + 2\vec{r'}$.

In summary, vectors can be added together to obtain another vector, and multiplying the sum of two vectors by a number (called a **scalar**) is just the sum of the individual vectors each multiplied by the scalar. In this text, almost all of the scalars we shall deal with will be either elements of the real number field $\mathbb{R}$ or the complex number field $\mathbb{C}$. We will refer to these two fields by the generic symbol $\mathcal{F}$. Essentially, we think of a **field** as a set of 'numbers' that we can add and multiply together to obtain another 'number' (closure) in a way such that for all $a, b, c \in \mathcal{F}$ we have $(a + b) + c = a + (b + c)$ and $(ab)c = a(bc)$

(associativity), $a + b = b + a$ and $ab = ba$ (commutativity), $a(b + c) = ab + ac$ (distributivity over addition), and where every number has an additive inverse (i.e., for each $a \in \mathcal{F}$ there exists $-a \in \mathcal{F}$ such that $a + (-a) = 0$ where 0 is defined by $0 + a = a + 0 = a$ for all $a \in \mathcal{F}$), and every *nonzero* number has a multiplicative inverse (i.e., for every nonzero $a \in \mathcal{F}$ there exists $a^{-1} \in \mathcal{F}$ such that $aa^{-1} = a^{-1}a = 1$, where $1 \in \mathcal{F}$ is defined by $1a = a1 = a$). In other words, a field behaves the way we are used to the real numbers behaving. However, fields are much more general than $\mathbb{R}$ and $\mathbb{C}$, and the interested reader may wish to read some of the books listed in the bibliography for a more thorough treatment of fields in general.

With this simple picture as motivation, we now turn to a careful definition of vector spaces.

1.2 Definitions

A nonempty set V is said to be a **vector space** over a field $\mathcal{F}$ if: (i) there exists an operation called **addition** that associates to each pair $x, y \in V$ a new vector $x + y \in V$ called the **sum** of x and y; (ii) there exists an operation called **scalar multiplication** that associates to each $a \in \mathcal{F}$ and $x \in V$ a new vector $ax \in V$ called the **product** of a and x; (iii) these operations satisfy the following axioms:

(VS1) $x + y = y + x$ for all $x, y \in V$.
(VS2) $(x + y) + z = x + (y + z)$ for all $x, y, z \in V$.
(VS3) There exists an element $0 \in V$ such that $0 + x = x$ for all $x \in V$.
(VS4) For all $x \in V$ there exists an element $-x \in V$ such that $x + (-x) = 0$.
(VS5) $a(x + y) = ax + ay$ for all $x, y \in V$ and all $a \in \mathcal{F}$.
(VS6) $(a + b)x = ax + bx$ for all $x \in V$ and all $a, b \in \mathcal{F}$.
(VS7) $a(bx) = (ab)x$ for all $x \in V$ and all $a, b \in \mathcal{F}$.
(VS8) $1x = x$ for all $x \in V$ where 1 is the (multiplicative) identity in $\mathcal{F}$.

The members of V are called **vectors**, and the members of $\mathcal{F}$ are called **scalars**. The vector $0 \in V$ is called the **zero vector**, and the vector $-x$ is called the **negative** of the vector x.

Throughout this chapter, V will always denote a vector space, and the corresponding field $\mathcal{F}$ will be understood even if it is not explicitly mentioned. If $\mathcal{F}$ is the real field $\mathbb{R}$, then we obtain a **real vector space** while if $\mathcal{F}$ is the complex field $\mathbb{C}$, then we obtain a **complex vector space**.

Example 1.1. Probably the best known example of a vector space is the set $\mathcal{F}^n = \mathcal{F} \times \cdots \times \mathcal{F}$ of all n-tuples $(a_1, \ldots, a_n)$ where each $a_i \in \mathcal{F}$. (See Section A.1 of the appendix for a discussion of the Cartesian product of sets.) To make $\mathcal{F}^n$ into a vector space, we define the **sum** of two elements $(a_1, \ldots, a_n) \in \mathcal{F}^n$ and $(b_1, \ldots, b_n) \in \mathcal{F}^n$ by

$$(a_1, \ldots, a_n) + (b_1, \ldots, b_n) = (a_1 + b_1, \ldots, a_n + b_n)$$

and scalar multiplication for any $k \in \mathcal{F}$ by

$$k(a_1, \ldots, a_n) = (ka_1, \ldots, ka_n).$$

If $A = (a_1, \ldots, a_n)$ and $B = (b_1, \ldots, b_n)$, then we say that $A = B$ if and only if $a_i = b_i$ for each $i = 1, \ldots, n$. Defining $0 = (0, \ldots, 0)$ and $-A = (-a_1, \ldots, -a_n)$ as the identity and inverse elements respectively of $\mathcal{F}^n$, the reader should have no trouble verifying properties (VS1)–(VS8).

The most common examples of the space $\mathcal{F}^n$ come from considering the fields $\mathbb{R}$ and $\mathbb{C}$. For instance, the space $\mathbb{R}^3$ is (with the Pythagorean notion of distance defined on it) just the ordinary three-dimensional Euclidean space (x, y, z) of elementary physics and geometry. Furthermore, it is standard to consider vectors in $\mathcal{F}^n$ as columns. For example, the vector $X = (x, y, z) \in \mathbb{R}^3$ should really be written as

$$X = \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

but it is typographically easier to write them as rows, and we will continue with this practice unless we need to explicitly show their column structure.

We shall soon see that any finite-dimensional vector space V over a field $\mathcal{F}$ is essentially the same as the space $\mathcal{F}^n$. In particular, we will prove that V is isomorphic to $\mathcal{F}^n$ for some positive integer n. (The term "isomorphic" will be defined carefully in the next section. But to put it briefly, two sets are isomorphic if there is a one-to-one correspondence between them.)

Example 1.2. Another very useful vector space is the space $\mathcal{F}[x]$ of all polynomials in the indeterminate x over the field $\mathcal{F}$. In other words, every element in $\mathcal{F}[x]$ is a polynomial of the form $a_0 + a_1x + \cdots + a_nx^n$ where each $a_i \in \mathcal{F}$ and n is any positive integer (called the **degree** of the polynomial). Addition and scalar multiplication are defined in the obvious way by

$$\sum_{i=0}^{n} a_i x^i + \sum_{i=0}^{n} b_i x^i = \sum_{i=0}^{n} (a_i + b_i)\, x^i$$

and

$$c\sum_{i=0}^{n} a_i x^i = \sum_{i=0}^{n} (ca_i)\, x^i$$

(If we wish to add together two polynomials $\sum_{i=0}^{n} a_i x^i$ and $\sum_{i=0}^{m} b_i x^i$ where $m > n$, then we simply define $a_i = 0$ for $i = n + 1, \ldots, m$.)

Since we have not yet defined the multiplication of vectors, we ignore the fact that polynomials can be multiplied together. It should be clear that $\mathcal{F}[x]$ does indeed form a vector space.

Example 1.3. We can also view the field $\mathbb{C}$ as a vector space over $\mathbb{R}$. In fact, we may generally consider the set of n-tuples $(z_1, \ldots, z_n)$, where each $z_i \in \mathbb{C}$, to be a vector space over $\mathbb{R}$ by defining addition and scalar multiplication (by *real* numbers) as in Example 1.1. We thus obtain a *real* vector space that is quite distinct from the space $\mathbb{C}^n$ (in which we can multiply by complex numbers).

We now prove several useful properties of vector spaces.

Theorem 1.1. *Let V be a vector space over $\mathcal{F}$. Then for all $x, y, z \in V$ and every $a \in \mathcal{F}$ we have*
(i) $x + y = z + y$ implies $x = z$.
(ii) $ax = 0$ if and only if $a = 0$ or $x = 0$.
(iii) $-(ax) = (-a)x = a(-x)$.

Proof. We first remark that there is a certain amount of sloppiness in our notation since the symbol 0 is used both as an element of V and as an element of $\mathcal{F}$. However, there should never be any confusion as to which of these sets 0 lies in, and we will continue with this common practice.

(i) If $x + y = z + y$, then

$$(x + y) + (-y) = (z + y) + (-y)$$

implies

$$x + (y + (-y)) = z + (y + (-y))$$

which implies $x + 0 = z + 0$ and hence $x = z$. This is frequently called the (right) **cancellation law**. It is also clear that $x + y = x + z$ implies $y = z$ (left cancellation).

(ii) If a = 0, then

$$0x = (0 + 0)x = 0x + 0x.$$

But $0x = 0 + 0x$ so that $0 + 0x = 0x + 0x$, and hence (i) implies $0 = 0x$. If $x = 0$, then

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

But $a0 = 0 + a0$ so that $0 + a0 = a0 + a0$, and again we have $0 = a0$. Conversely, assume that $ax = 0$. If $a \neq 0$ then a^{-1} exists, and hence

$$x = 1x = (a^{-1}a)x = a^{-1}(ax) = a^{-1}0 = 0$$

by the previous paragraph.

(iii) By (VS4) we have $ax + (-(ax)) = 0$, whereas by (ii) and (VS6), we have

$$0 = 0x = (a + (-a))x = ax + (-a)x.$$

Hence $ax + (-(ax)) = ax + (-a)x$ implies $-(ax) = (-a)x$ by (i). Similarly, $0 = x + (-x)$ so that

$$0 = a0 = a(x + (-x)) = ax + a(-x).$$

Then $0 = ax + (-(ax)) = ax + a(-x)$ implies $-(ax) = a(-x)$. ∎

In view of this theorem, it makes sense to define **subtraction** in V by

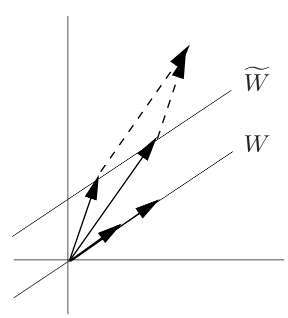
$$x - y = x + (-y).$$

It should then be clear that a vector space will also have the properties we expect, such as $a(x - y) = ax - ay$, and $-(x - y) = -x + y$.

If we take an arbitrary subset of vectors in a vector space then, in general, this subset will not be a vector space itself. The reason for this is that in general, even the addition of two vectors in the subset will not result in a vector that is again a member of the subset. Because of this, we make the following definition.

Suppose V is a vector space over $\mathcal{F}$ and $W \subset V$. Then, if $x, y \in W$ and $c \in \mathcal{F}$ implies that $x + y \in W$ and $cx \in W$, we say that W is a **subspace** of V. Indeed, if $c = 0$ then $0 = 0x \in W$ so that $0 \in W$, and similarly $-x = (-1)x \in W$ so that $-x \in W$ also. It is now easy to see that W obeys (VS1)–(VS8) if V does. It should also be clear that an equivalent way to define a subspace is to require that $cx + y \in W$ for all $x, y \in W$ and all $c \in \mathcal{F}$.

It is extremely important to realize that any subspace always contains the zero vector. As a simple example, consider a line W through the origin of the usual xy-plane. Then the sum of any two points lying in W will still lie in W. But if we consider a line $\widetilde{W}$ that does not pass through the origin, then the sum of two points on $\widetilde{W}$ will not lie on $\widetilde{W}$. Thus the subset W is a subspace of V but $\widetilde{W}$ is not.



If W is a subspace of V and $W \neq V$, then W is called a **proper** subspace of V. In particular, $W = \{0\}$ is a subspace of V, but it is not very interesting, and hence from now on we assume that any proper subspace contains more than simply the zero vector. (One sometimes refers to $\{0\}$ and V as **trivial** subspaces of V.)

Example 1.4. Consider the elementary Euclidean space $\mathbb{R}^3$ consisting of all triples (x, y, z) of real scalars. If we restrict our consideration to those vectors of the form $(x, y, 0)$, then we obtain a subspace of $\mathbb{R}^3$. In fact, this subspace is essentially just the space $\mathbb{R}^2$ which we think of as the usual xy-plane. We leave it as a simple exercise for the reader to show that this does indeed define a subspace of $\mathbb{R}^3$. Note that any other plane parallel to the xy-plane is *not* a subspace (why?).

Example 1.5. Let V be a vector space over $\mathcal{F}$, and let $S = \{x_1, \ldots, x_n\}$ be any n vectors in V. Given any set of scalars $\{a_1, \ldots, a_n\}$, the vector

$$\sum_{i=1}^{n} a_i x_i = a_1 x_1 + \cdots + a_n x_n$$

is called a **linear combination** of the n vectors $x_i \in S$, and the set $\mathscr{S}$ of all such linear combinations of elements in S is called the subspace **spanned** (or **generated**) by S. Indeed, if $A = \sum_{i=1}^{n} a_i x_i$ and $B = \sum_{i=1}^{n} b_i x_i$ are vectors in $\mathscr{S}$ and $c \in \mathcal{F}$, then both

$$A + B = \sum_{i=1}^{n} (a_i + b_i) x_i$$

and

$$cA = \sum_{i=1}^{n} (ca_i) x_i$$

are vectors in $\mathscr{S}$. Hence $\mathscr{S}$ is a subspace of V. $\mathscr{S}$ is sometimes called the **linear span** of S, and we say that S **spans** $\mathscr{S}$.

In view of this example, we might ask whether or not every vector space is in fact the linear span of some set of vectors in the space. In the next section we shall show that this leads naturally to the concept of the dimension of a vector space.

Exercises

1. Verify axioms (VS1)–(VS8) for the space $\mathcal{F}^n$.

2. Let S be any set, and consider the collection V of all mappings f of S into a field $\mathcal{F}$. For any $f, g \in V$ and $\alpha \in \mathcal{F}$, we define $(f + g)(x) = f(x) + g(x)$ and $(\alpha f)(x) = \alpha f(x)$ for every $x \in S$. Show that V together with these operations defines a vector space over $\mathcal{F}$.

3. Consider the two element set $\{x, y\}$ with addition and scalar multiplication by $c \in \mathcal{F}$ defined by

$$x + x = x \qquad x + y = y + x = y \qquad y + y = x \qquad cx = x \qquad cy = x.$$

Does this define a vector space over $\mathcal{F}$?

4. Let V be a vector space over $\mathcal{F}$. Show that if $x \in V$ and $a, b \in \mathcal{F}$ with $a \neq b$, then $ax = bx$ implies $x = 0$.

5. Let $(V, +, \star)$ be a real vector space with the addition operation denoted by $+$ and the scalar multiplication operation denoted by $\star$. Let $v_0 \in V$ be fixed. We define a new addition operation $\oplus$ on V by $x \oplus y = x + y + v_0$, and a new scalar multiplication operation $\circledast$ by $\alpha \circledast x = \alpha \star x + (\alpha - 1) \star v_0$. Show that $(V, \oplus, \circledast)$ defines a real vector space.

6. Let $F[\mathbb{R}]$ denote the space of all real-valued functions defined on $\mathbb{R}$ with addition and scalar multiplication defined as in Exercise 2. In other words, $f \in F[\mathbb{R}]$ means $f : \mathbb{R} \to \mathbb{R}$.
   (a) Let $C[\mathbb{R}]$ denote the set of all continuous real-valued functions defined on $\mathbb{R}$. Show that $C[\mathbb{R}]$ is a subspace of $F[\mathbb{R}]$.
   (b) Repeat part (a) with the set $D[\mathbb{R}]$ of all such differentiable functions.

7. Referring to the previous exercise, let $D^n[\mathbb{R}]$ denote the set of all n-times differentiable functions from $\mathbb{R}$ to $\mathbb{R}$. Consider the subset V of $D^n[\mathbb{R}]$ given by the set of all functions that satisfy the differential equation

$$f^{(n)}(x) + a_{n-1} f^{(n-1)}(x) + a_{n-2} f^{(n-2)}(x) + \cdots + a_1 f^{(1)}(x) + a_0 f(x) = 0$$

where $f^{(i)}(x)$ denotes the ith derivative of $f(x)$ and a_i is a fixed real constant. Show that V is a vector space.

8. Let $V = \mathbb{R}^3$. In each of the following cases, determine whether or not the subset W is a subspace of V:
   (a) $W = \{(x, y, 0) : x, y \in \mathbb{R}\}$ (see Example 1.4) .
   (b) $W = \{(x, y, z) \in \mathbb{R}^3 : z \geq 0\}$.
   (c) $W = \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 + z^2 \leq 1\}$.
   (d) $W = \{(x, y, z) \in \mathbb{R}^3 : x + y + z = 0\}$.
   (e) $W = \{(x, y, z) \in \mathbb{R}^3 : x, y, z \in \mathbb{Q}\}$.
   (f) $W = \{(x, y, z) \in \mathbb{R}^3 - \{0, 0, 0\}\}$.

9. Let S be a nonempty subset of a vector space V. In Example 1.5 we showed that the linear span $\mathscr{S}$ of S is a subspace of V. Show that if W is any other subspace of V containing S, then $\mathscr{S} \subset W$.

10. (a) Determine whether or not the intersection $\bigcap_{i=1}^{n} W_i$ of a finite number of subspaces W_i of a vector space V is a subspace of V.
   (b) Determine whether or not the union $\bigcup_{i=1}^{n} W_i$ of a finite number of subspaces W_i of a space V is a subspace of V.

11. Let W_1 and W_2 be subspaces of a space V such that $W_1 \cup W_2$ is also a subspace of V. Show that one of the W_i is subset of the other.

12. Let W_1 and W_2 be subspaces of a vector space V. If, for every $v \in V$ we have $v = w_1 + w_2$ where $w_i \in W_i$, then we write $V = W_1 + W_2$ and say that V is the **sum** of the subspaces W_i. If $V = W_1 + W_2$ and $W_1 \cap W_2 = \{0\}$, show that every $v \in V$ has a unique representation $v = w_1 + w_2$ with $w_i \in W_i$.

13. Let V be the set of all (infinite) real sequences. In other words, any $v \in V$ is of the form $(x_1, x_2, x_3, \dots)$ where each $x_i \in \mathbb{R}$. If we define the addition and scalar multiplication of distinct sequences componentwise exactly as in Example 1.1, then it should be clear that V is a vector space over $\mathbb{R}$. Determine whether or not each of the following subsets of V in fact forms a subspace of V:
    (a) All sequences containing only a finite number of nonzero terms.
    (b) All sequences of the form $\{x_1, x_2, \dots, x_N, 0, 0, \dots\}$ where N is fixed.
    (c) All **decreasing sequences**, i.e., sequences where $x_{k+1} \leq x_k$ for each $k = 1, 2, \dots$.
    (d) All **convergent sequences**, i.e., sequences for which $\lim_{k\to\infty} x_k$ exists.

14. For which value of k will the vector $v = (1, -2, k) \in \mathbb{R}^3$ be a linear combination of the vectors $x_1 = (3, 0, -2)$ and $x_2 = (2, -1, -5)$?

15. Write the vector $v = (1, -2, 5)$ as a linear combination of the vectors $x_1 = (1, 1, 1)$, $x_2 = (1, 2, 3)$ and $x_3 = (2, -1, 1)$.

1.3 Linear Independence and Bases

Let $x_1, \dots, x_n$ be vectors in a vector space V. We say that these vectors are **linearly dependent** if there exist scalars $a_1, \dots, a_n \in \mathcal{F}$, not all equal to 0, such that

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = \sum_{i=1}^{n} a_ix_i = 0.$$

The vectors x_i are said to be **linearly independent** if they are not linearly dependent. In other words, if $\{x_1, \dots, x_n\}$ is linearly independent, then $\sum_{i=1}^{n} a_ix_i = 0$ implies that $a_1 = \cdots = a_n = 0$. From these definitions, it follows that any set containing a linearly dependent subset must be linearly dependent, and any subset of a linearly independent set is necessarily linearly independent.

It is important to realize that a set of vectors may be linearly dependent with respect to one field, but independent with respect to another. For example, the set $\mathbb{C}$ of all complex numbers is itself a vector space over either the field of real numbers or over the field of complex numbers. However, the set $\{x_1 = 1, x_2 = i\}$ is linearly independent if $\mathcal{F} = \mathbb{R}$, but linearly dependent if $\mathcal{F} = \mathbb{C}$ since

$ix_1 + (-1)x_2 = 0$. We will always assume that a linear combination is taken with respect to the same field that V is defined over.

As a means of simplifying our notation, we will frequently leave off the limits of a sum when there is no possibility of ambiguity. Thus, if we are considering the set $\{x_1, \ldots, x_n\}$, then a linear combination of the x_i will often be written as $\sum a_i x_i$ rather than $\sum_{i=1}^{n} a_i x_i$. In addition, we will often denote a collection $\{x_1, \ldots, x_n\}$ of vectors simply by $\{x_i\}$.

Example 1.6. Consider the three vectors in $\mathbb{R}^3$ given by

$$\begin{aligned} e_1 &= (1,0,0) \\ e_2 &= (0,1,0) \\ e_3 &= (0,0,1) \end{aligned}$$

Using the definitions of addition and scalar multiplication given in Example 1.1, it is easy to see that these three vectors are linearly independent. This is because the zero vector in $\mathbb{R}^3$ is given by $(0,0,0)$, and hence

$$a_1 e_1 + a_2 e_2 + a_3 e_3 = (a_1, a_2, a_3) = (0,0,0)$$

implies that $a_1 = a_2 = a_3 = 0$.

On the other hand, the vectors

$$\begin{aligned} x_1 &= (1,0,0) \\ x_2 &= (0,1,2) \\ x_3 &= (1,3,6) \end{aligned}$$

are linearly dependent since $x_3 = x_1 + 3x_2$.

From a practical point of view, to say that a set of vectors is linearly dependent means that one of them is a linear combination of the rest. The formal proof of this fact is given in the following elementary result.

Theorem 1.2. *A finite set S of vectors in a space V is linearly dependent if and only if one vector in the set is a linear combination of the others. In other words, S is linearly dependent if one vector in S is in the subspace spanned by the remaining vectors in S.*

Proof. If $S = \{x_1, \ldots, x_n\}$ is a linearly dependent subset of V, then

$$a_1 x_1 + a_2 x_2 + \cdots + a_n x_n = 0$$

for some set of scalars $a_1, \ldots, a_n \in \mathcal{F}$ not all equal to 0. Suppose, to be specific, that $a_1 \neq 0$. Then we may write

$$x_1 = -(a_2/a_1)x_2 - \cdots - (a_n/a_1)x_n$$

which shows that x_1 is a linear combination of $x_2, \ldots, x_n$.

Conversely, if $x_1 = \sum_{i \neq 1} a_i x_i$ then

$$x_1 + (-a_2)x_2 + \cdots + (-a_n)x_n = 0$$

which shows that the collection $\{x_1, \ldots, x_n\}$ is linearly dependent. ∎

It is important to realize that no linearly independent set of vectors can contain the zero vector. To see this, note that if $S = \{x_1, \ldots, x_n\}$ and $x_1 = 0$, then $ax_1 + 0x_2 + \cdots + 0x_n = 0$ for all $a \in \mathcal{F}$, and hence by definition, S is a linearly dependent set.

Theorem 1.3. *Let $S = \{x_1, \ldots, x_n\} \subset V$ be a linearly independent set, and let $\mathscr{S}$ be the linear span of S. Then every $v \in \mathscr{S}$ has a unique representation*

$$v = \sum_{i=1}^{n} a_i x_i$$

where each $a_i \in \mathcal{F}$.

Proof. By definition of $\mathscr{S}$, we can always write $v = \sum a_i x_i$. As to uniqueness, it must be shown that if we also have $v = \sum b_i x_i$, then it follows that $b_i = a_i$ for every $i = 1, \ldots, n$. But this is easy since $\sum a_i x_i = \sum b_i x_i$ implies $\sum (a_i - b_i) x_i = 0$, and hence $a_i - b_i = 0$ (since $\{x_i\}$ is linearly independent). Therefore $a_i = b_i$ for each $i = 1, \ldots, n$. ∎

If S is a finite subset of a vector space V such that $V = \mathscr{S}$ (the linear span of S), then we say that V is **finite-dimensional**. However, we must define what is meant in general by the dimension of V. If $S \subset V$ is a linearly independent set of vectors with the property that $V = \mathscr{S}$, then we say that S is a **basis** for V. In other words, a **basis** for V is a linearly independent set that spans V. We shall see that the number of elements in a basis is what is meant by the **dimension** of V. But before we can state this precisely, we must be sure that such a number is well-defined. In other words, we must show that any basis has the same number of elements. We prove this (see the corollary to Theorem 1.6) in several steps.

Theorem 1.4. *Let $\mathscr{S}$ be the linear span of $S = \{x_1, \ldots, x_n\} \subset V$. If $k \leq n$ and $\{x_1, \ldots, x_k\}$ is linearly independent, then there exists a linearly independent subset of S of the form $\{x_1, \ldots, x_k, x_{i_1}, \ldots, x_{i_\alpha}\}$ whose linear span also equals $\mathscr{S}$.*

Proof. If $k = n$ there is nothing left to prove, so we assume that $k < n$. Since $x_1, \ldots, x_k$ are linearly independent, we let x_j (where $j > k$) be the first vector in S that is a linear combination of the preceding $x_1, \ldots, x_{j-1}$. If no such j exists, then take $(i_1, \ldots, i_\alpha) = (k+1, \ldots, n)$. Then the set of $n-1$ vectors $x_1, \ldots, x_{j-1}, x_{j+1}, \ldots, x_n$ has a linear span that must be contained in $\mathscr{S}$ (since this set is just a subset of S). However, if v is any vector in $\mathscr{S}$, we can write $v = \sum_{i=1}^n a_i x_i$ where x_j is just a linear combination of the first $j-1$ vectors. In other words, v is a linear combination of $x_1, \ldots, x_{j-1} x_{j+1}, \ldots, x_n$ and hence these $n-1$ vectors also span $\mathscr{S}$.

We now continue this process by picking out the first vector in this set of $n-1$ vectors that is a linear combination of the preceding vectors. An identical argument shows that the linear span of this set of $n-2$ vectors must also be $\mathscr{S}$. It is clear that we will eventually obtain a set $\{x_1, \ldots, x_k, x_{i_1}, \ldots, x_{i_\alpha}\}$ whose linear span is still $\mathscr{S}$, but in which no vector is a linear combination of the preceding ones. This means that the set must be linearly independent (Theorem 1.2). ∎

Corollary 1. *If V is a finite-dimensional vector space such that the set $S = \{x_1, \ldots, x_m\} \subset V$ spans V, then some subset of S is a basis for V.*

Proof. By Theorem 1.4, S contains a linearly independent subset that also spans V. But this is precisely the requirement that S contain a basis for V. ∎

Corollary 2. *Let V be a finite-dimensional vector space and let $\{x_1, \ldots, x_n\}$ be a basis for V. Then any element $v \in V$ has a unique representation of the form*

$$v = \sum_{i=1}^n a_i x_i$$

where each $a_i \in \mathcal{F}$.

Proof. Since $\{x_i\}$ is linearly independent and spans V, Theorem 1.3 shows us that any $v \in V$ may be written in the form $v = \sum_{i=1}^n a_i x_i$ where each $a_i \in \mathcal{F}$ is unique (for this particular basis). ∎

It is important to realize that Corollary 1 asserts the existence of a finite basis in any finite-dimensional vector space, but says nothing about the uniqueness of this basis. In fact, there are an infinite number of possible bases for any such space. However, by Corollary 2, once a particular basis has been chosen, then any vector has a unique expansion in terms of this basis.

Example 1.7. Returning to the space $\mathcal{F}^n$, we see that any $(a_1, \ldots, a_n) \in \mathcal{F}^n$ can be written as the linear combination

$$a_1(1, 0, \ldots, 0) + a_2(0, 1, 0, \ldots, 0) + \cdots + a_n(0, \ldots, 0, 1).$$

This means that the n vectors

$$\begin{aligned} e_1 &= (1, 0, 0, \ldots, 0) \\ e_2 &= (0, 1, 0, \ldots, 0) \\ &\vdots \\ e_n &= (0, 0, 0, \ldots, 1) \end{aligned}$$

span $\mathcal{F}^n$. They are also linearly independent since $\sum a_i e_i = (a_1, \ldots, a_n) = 0$ if and only if $a_i = 0$ for all $i = 1, \ldots, n$. The set $\{e_i\}$ is extremely useful, and will be referred to as the **standard basis** for $\mathcal{F}^n$.

This example leads us to make the following generalization. By an **ordered basis** for a finite-dimensional space V, we mean a finite sequence of vectors that is linearly independent and spans V. If the *sequence* $x_1, \ldots, x_n$ is an ordered basis for V, then the *set* $\{x_1, \ldots, x_n\}$ is a basis for V. In other words, the set $\{x_1, \ldots, x_n\}$ gives rise to $n!$ different ordered bases. Since there is usually nothing lost in assuming that a basis is ordered, we shall continue to assume that $\{x_1, \ldots, x_n\}$ denotes an ordered basis unless otherwise noted.

Given any (ordered) basis $\{x_1, \ldots, x_n\}$ for V, we know that any $v \in V$ has a unique representation $v = \sum_{i=1}^n a_i x_i$. We call the scalars $a_1, \ldots, a_n$ the **coordinates** of v relative to the (ordered) basis $\{x_1, \ldots, x_n\}$. In particular, we call a_i the ith **coordinate** of v. Moreover, we now proceed to show that these coordinates define a direct correspondence between V and $\mathcal{F}^n$ (or, as we shall define it below, an isomorphism).

Let V and W be vector spaces over $\mathcal{F}$. We say that a mapping $\phi : V \to W$ is a **vector space homomorphism** (or, as we shall call it later, a **linear transformation**) if

$$\phi(x + y) = \phi(x) + \phi(y)$$

and

$$\phi(ax) = a\phi(x)$$

for all $x, y \in V$ and $a \in \mathcal{F}$. If ϕ is injective (i.e., one-to-one), then we say that ϕ is an **isomorphism**, and if ϕ is bijective (i.e., injective and surjective, or one-to-one and onto), that V and W are **isomorphic**. (If necessary, the reader may wish to review mappings in Section A.2 of the appendix to understand some of these terms.)

As we now show, the set of vectors $x \in V$ that map into $0 \in W$ under ϕ gives us some very important information about ϕ. To show this, we define the

kernel of ϕ to be the set

$$\operatorname{Ker}\phi = \{x \in V : \phi(x) = 0 \in W\}.$$

If $x, y \in \operatorname{Ker}\phi$ and $c \in \mathcal{F}$ we have

$$\phi(x+y) = \phi(x) + \phi(y) = 0$$

and

$$\phi(cx) = c\phi(x) = c0 = 0.$$

This shows that both $x+y$ and cx are in $\operatorname{Ker}\phi$, and hence $\operatorname{Ker}\phi$ is a subspace of V. Note also that if $a = 0$ and $x \in V$ then

$$\phi(0) = \phi(ax) = a\phi(x) = 0.$$

Alternatively, we could also note that

$$\phi(x) = \phi(x+0) = \phi(x) + \phi(0)$$

and hence $\phi(0) = 0$. Finally, we see that

$$0 = \phi(0) = \phi(x + (-x)) = \phi(x) + \phi(-x)$$

and therefore

$$\phi(-x) = -\phi(x).$$

The importance of the kernel arises from the following result.

Theorem 1.5. *Let $\phi : V \to W$ be a vector space homomorphism. Then ϕ is an isomorphism if and only if* $\operatorname{Ker}\phi = \{0\}$.

Proof. If ϕ is injective, then the fact that $\phi(0) = 0$ implies that we must have $\operatorname{Ker}\phi = \{0\}$. Conversely, if $\operatorname{Ker}\phi = \{0\}$ and $\phi(x) = \phi(y)$, then

$$0 = \phi(x) - \phi(y) = \phi(x-y)$$

implies that $x - y = 0$, or $x = y$. ∎

Now let us return to the above notion of an ordered basis. For any finite-dimensional vector space V over $\mathcal{F}$ and any (ordered) basis $\{x_1, \ldots, x_n\}$, we define a mapping $\phi : V \to \mathcal{F}^n$ by

$$\phi(v) = \phi\left(\sum_{i=1}^n a_i x_i\right) = (a_1, \ldots, a_n)$$

for each

$$v = \sum_{i=1}^n a_i x_i \in V.$$

Since

$$\begin{aligned}\phi\Big(\sum a_i x_i + \sum b_i x_i\Big) &= \phi\Big(\sum (a_i + b_i) x_i\Big) \\ &= (a_1 + b_1, \dots, a_n + b_n) \\ &= (a_1, \dots, a_n) + (b_1, \dots, b_n) \\ &= \phi\Big(\sum a_i x_i\Big) + \phi\Big(\sum b_i x_i\Big)\end{aligned}$$

and

$$\begin{aligned}\phi(kv) &= \phi\Big(k \sum a_i x_i\Big) = \phi\Big(\sum (k a_i) x_i\Big) = (k a_1, \dots, k a_n) \\ &= k(a_1, \dots, a_n) = k\phi(v)\end{aligned}$$

we see that ϕ is a vector space homomorphism. Because the coordinates of any vector are unique for a fixed basis, we see that this mapping is indeed well-defined and one-to-one. (Alternatively, the identity element in the space $\mathcal{F}^n$ is $(0, \dots, 0)$, and the only vector that maps into this is the zero vector in V. Hence $\operatorname{Ker}\phi = \{0\}$ and ϕ is an isomorphism.) It is clear that ϕ is surjective since, given any ordered set of scalars $a_1, \dots, a_n \in \mathcal{F}$, we can define the vector $v = \sum a_i x_i \in V$. Therefore we have shown that V and $\mathcal{F}^n$ are isomorphic for some n, where n is the number of vectors in an ordered basis for V.

If V has a basis consisting of n elements, is it possible to find another basis consisting of $m \neq n$ elements? Intuitively we guess not, for if this were true then V would be isomorphic to $\mathcal{F}^m$ as well as to $\mathcal{F}^n$, which implies that $\mathcal{F}^m$ is isomorphic to $\mathcal{F}^n$ for $m \neq n$. That this is not possible should be obvious by simply considering the projection of a point in $\mathbb{R}^3$ down onto the plane $\mathbb{R}^2$. Any point in $\mathbb{R}^2$ is thus the image of an entire vertical line in $\mathbb{R}^3$, and hence this projection can not possibly be an isomorphism. Nevertheless, we proceed to prove this in detail beginning with our next theorem.

Theorem 1.6. *Let $\{x_1, \dots, x_n\}$ be a basis for V, and let $\{y_1, \dots, y_m\}$ be linearly independent vectors in V. Then $m \le n$.*

Proof. Since $\{x_1, \dots, x_n\}$ spans V, we may write each y_i as a linear combination of the x_j. In particular, choosing y_m, it follows that the set

$$\{y_m, x_1, \dots, x_n\}$$

is linearly dependent (Theorem 1.2) and spans V (since the x_k already do so). Hence there must be a proper subset $\{y_m, x_{i_1}, \dots, x_{i_r}\}$ with $r \le n - 1$ that forms a basis for V (Theorem 1.4). Now this set spans V so that y_{m-1} is a linear combination of this set, and hence

$$\{y_{m-1}, y_m, x_{i_1}, \dots, x_{i_r}\}$$

is linearly dependent and spans V. By Theorem 1.4 again, we can find a set $\{y_{m-1}, y_m, x_{j_1}, \dots, x_{j_s}\}$ with $s \leq n-2$ that is also a basis for V. Continuing our process, we eventually obtain the set

$$\{y_2, \dots, y_m, x_\alpha, x_\beta, \dots\}$$

which spans V and must contain at least one of the x_k (since y_1 is not a linear combination of the set $\{y_2, \dots, y_m\}$ by hypothesis). This set was constructed by adding $m-1$ vectors y_i to the original set of n vectors x_k, and deleting at least $m-1$ of the x_k along the way. However, we still have at least one of the x_k in our set, and hence it follows that $m-1 \leq n-1$, or $m \leq n$. ∎

Corollary. *Any two bases for a finite-dimensional vector space must consist of the same number of elements.*

Proof. Let $\{x_1, \dots, x_n\}$ and $\{y_1, \dots, y_m\}$ be bases for V. Since the y_i are linearly independent, Theorem 1.6 says that $m \leq n$. On the other hand, the x_j are linearly independent so that $n \leq m$. Therefore we must have $n = m$. ∎

We now return to the proof that $\mathcal{F}^m$ is isomorphic to $\mathcal{F}^n$ if and only if $m = n$. Let us first show that an isomorphism maps a basis to a basis.

Theorem 1.7. *Let $\phi : V \to W$ be an isomorphism of finite-dimensional vector spaces. Then a set of vectors $\{\phi(v_1), \dots, \phi(v_n)\}$ is linearly dependent in W if and only if the set $\{v_1, \dots, v_n\}$ is linearly dependent in V.*

Proof. If the set $\{v_1, \dots, v_n\}$ is linearly dependent, then for some set of scalars $\{a_1, \dots, a_n\}$, not all equal to 0, we have $\sum_{i=1}^n a_i v_i = 0$. Applying ϕ to both sides of this equation yields

$$0 = \phi(0) = \phi\Big(\sum a_i v_i\Big) = \sum \phi(a_i v_i) = \sum a_i \phi(v_i).$$

But since not all of the a_i are 0, this means that $\{\phi(v_i)\}$ must be linearly dependent.

Conversely, if $\phi(v_1), \dots, \phi(v_n)$ are linearly dependent, then there exists a set of scalars $b_1, \dots, b_n$ not all 0 such that $\sum b_i \phi(v_i) = 0$. But this means

$$0 = \sum b_i \phi(v_i) = \sum \phi(b_i v_i) = \phi\Big(\sum b_i v_i\Big)$$

which implies that $\sum b_i v_i = 0$ (since $\operatorname{Ker} \phi = \{0\}$). This shows that the set $\{v_i\}$ is linearly dependent. ∎

Corollary. *If $\phi : V \to W$ is an isomorphism of finite-dimensional vector spaces, then $\{\phi(x_i)\} = \{\phi(x_1), \ldots, \phi(x_n)\}$ is a basis for W if and only if $\{x_i\} = \{x_1, \ldots, x_n\}$ is a basis for V.*

Proof. Since ϕ is an isomorphism, for any vector $w \in W$ there exists a unique $v \in V$ such that $\phi(v) = w$. If $\{x_i\}$ is a basis for V, then $v = \sum_{i=1}^{n} a_i x_i$ and

$$w = \phi(v) = \phi\Big(\sum a_i x_i\Big) = \sum a_i \phi\left(x_i\right).$$

Hence the $\phi(x_i)$ span W, and they are linearly independent by Theorem 1.7.

On the other hand, if $\{\phi(x_i)\}$ is a basis for W, then there exist scalars $\{b_i\}$ such that for any $v \in V$ we have

$$\phi(v) = w = \sum b_i \phi(x_i) = \phi\Big(\sum b_i x_i\Big).$$

Since ϕ is an isomorphism, this implies that $v = \sum b_i x_i$, and hence $\{x_i\}$ spans V. The fact that it is linearly independent follows from Theorem 1.7. This shows that $\{x_i\}$ is a basis for V. ∎

Theorem 1.8. *$\mathcal{F}^n$ is isomorphic to $\mathcal{F}^m$ if and only if $n = m$.*

Proof. If $n = m$ the result is obvious. Now assume that $\mathcal{F}^n$ and $\mathcal{F}^m$ are isomorphic. We have seen in Example 1.7 that the standard basis of $\mathcal{F}^n$ consists of n vectors. Since an isomorphism carries one basis onto another (corollary to Theorem 1.7), any space isomorphic to $\mathcal{F}^n$ must have a basis consisting of n vectors. Hence, by the corollary to Theorem 1.6 we must have $m = n$. ∎

Corollary. *If V is a finite-dimensional vector space over $\mathcal{F}$, then V is isomorphic to $\mathcal{F}^n$ for a unique integer n.*

Proof. It was shown following Theorem 1.5 that V is isomorphic to $\mathcal{F}^n$ for some integer n, and Theorem 1.8 shows that n must be unique. ∎

The corollary to Theorem 1.6 shows us that the number of elements in any basis for a finite-dimensional vector space is fixed. We call this unique number n the **dimension** of V over $\mathcal{F}$, and we write $\dim V = n$. Our next result agrees with our intuition, and is quite useful in proving other theorems.

Theorem 1.9. *Every subspace W of a finite-dimensional vector space V is finite-dimensional, and* $\dim W \le \dim V$.

Proof. We must show that W has a basis, and that this basis contains at most $n = \dim V$ elements. If $W = \{0\}$, then $\dim W = 0 \le n$ and we are done. If W contains some $x_1 \neq 0$, then let $W_1 \subset W$ be the subspace spanned by x_1. If $W = W_1$, then $\dim W = 1$ and we are done. If $W \neq W_1$, then there exists some $x_2 \in W$ with $x_2 \notin W_1$, and we let W_2 be the subspace spanned by $\{x_1, x_2\}$. Again, if $W = W_2$, then $\dim W = 2$. If $W \neq W_2$, then choose some $x_3 \in W$ with $x_3 \notin W_2$ and continue this procedure. However, by Theorem 1.6, there can be at most n linearly independent vectors in V, and hence $\dim W \le n$. ∎

Note that the zero subspace is spanned by the vector 0, but $\{0\}$ is not linearly independent so it can not form a basis. Therefore the zero subspace is *defined* to have dimension zero.

Finally, let us show that any set of linearly independent vectors may be extended to form a complete basis.

Theorem 1.10. *Let V be finite-dimensional and $S = \{x_1, \dots, x_m\}$ any set of m linearly independent vectors in V. Then there exists a set $\{x_{m+1}, \dots, x_{m+r}\}$ of vectors in V such that $\{x_1, \dots, x_{m+r}\}$ is a basis for V.*

Proof. Since V is finite-dimensional, it has a basis $\{v_1, \dots, v_n\}$. Then the set $\{x_1, \dots, x_m, v_1, \dots, v_n\}$ spans V so, by Theorem 1.4, we can choose a subset $\{x_1, \dots, x_m, v_{i_1}, \dots, v_{i_r}\}$ of linearly independent vectors that span V. Letting $v_{i_1} = x_{m+1}, \dots, v_{i_r} = x_{m+r}$ proves the theorem. ∎

Exercises

1. Determine whether or not the three vectors $x_1 = (2, -1, 0)$, $x_2 = (1, -1, 1)$ and $x_3 = (0, 2, 3)$ form a basis for $\mathbb{R}^3$.

2. In each of the following, show that the given set of vectors is linearly independent, and decide whether or not it forms a basis for the indicated space:

   (a) $\{(1, 1), (1, -1)\}$ in $\mathbb{R}^2$.
   (b) $\{(2, 0, 1), (1, 2, 0), (0, 1, 0)\}$ in $\mathbb{R}^3$.
   (c) $\{(1, 0, 0, 0), (1, 1, 0, 0), (1, 1, 1, 0), (1, 1, 1, 1)\}$ in $\mathbb{R}^4$.

3. Extend each of the following sets to a basis for the given space:

   (a) $\{(1, 1, 0), (2, -2, 0)\}$ in $\mathbb{R}^3$.
   (b) $\{(1, 0, 0, 0), (1, 0, 1, 0), (1, 0, 0, 1)\}$ in $\mathbb{R}^4$.
   (c) $\{(1, 1, 0, 0), (1, -1, 0, 0), (1, 0, 1, 0)\}$ in $\mathbb{R}^4$.

4. Show that the vectors $u = (1 + i, 2i)$, $v = (1, 1 + i) \in \mathbb{C}^2$ are linearly dependent over $\mathbb{C}$, but linearly independent over $\mathbb{R}$.

5. Find the coordinates of the vector $(3, 1, -4) \in \mathbb{R}^3$ relative to the basis $x_1 = (1, 1, 1)$, $x_2 = (0, 1, 1)$ and $x_3 = (0, 0, 1)$.

6. Let $\mathbb{R}_3[x]$ be the space of all real polynomials of degree ≤ 3. Determine whether or not each of the following sets of polynomials is linearly independent:
   (a) $\{x^3 - 3x^2 + 5x + 1, x^3 - x^2 + 8x + 2, 2x^3 - 4x^2 + 9x + 5\}$.
   (b) $\{x^3 + 4x^2 - 2x + 3, x^3 + 6x^2 - x + 4, 3x^3 + 8x^2 - 8x + 7\}$.

7. Let V be a finite-dimensional space, and let W be any subspace of V. Show that there exists a subspace W' of V such that $W \cap W' = \{0\}$ and $V = W + W'$ (see Exercise 1.2.12 for the definition of $W + W'$).

8. Let $\phi : V \to W$ be a homomorphism of two vector spaces V and W.
   (a) Show that ϕ maps any subspace of V onto a subspace of W.
   (b) Let S' be a subspace of W, and define the set $S = \{x \in V : \phi(x) \in S'\}$. Show that S is a subspace of V.

9. Let V be finite-dimensional, and assume that $\phi : V \to V$ is a surjective homomorphism. Prove that ϕ is in fact an isomorphism of V onto V.

10. Let V have basis $x_1, x_2, \ldots, x_n$, and let $v_1, v_2, \ldots, v_n$ be any n elements in V. Define a mapping $\phi : V \to V$ by

$$\phi\left(\sum_{i=1}^{n} a_i x_i\right) = \sum_{i=1}^{n} a_i v_i$$

    where each $a_i \in \mathcal{F}$.
    (a) Show that ϕ is a homomorphism.
    (b) When is ϕ an isomorphism?

1.4 Direct Sums

We now present some useful ways of constructing a new vector space from several given spaces. The reader is advised to think carefully about these concepts, as they will become quite important later in this book. We also repeat our earlier remark that all of the vector spaces that we are discussing are considered to be defined over the same field $\mathcal{F}$.

Let A and B be subspaces of a finite-dimensional vector space V. Then we may define the **sum** of A and B to be the set $A + B$ given by

$$A + B = \{a + b : a \in A \text{ and } b \in B\}.$$

It is important to note that A and B must both be subspaces of the same space V, or else the addition of $a \in A$ to $b \in B$ is not defined. In fact, since A and B

are subspaces of V, it is easy to show that $A+B$ is also subspace of V. Indeed, given any a_1+b_1 and a_2+b_2 in $A+B$ and any $k \in \mathcal{F}$ we see that

$$(a_1+b_1)+(a_2+b_2)=(a_1+a_2)+(b_1+b_2) \in A+B$$

and

$$k(a_1+b_1)=ka_1+kb_1 \in A+B$$

as required. This definition can clearly be extended by induction to any finite collection $\{A_i\}$ of subspaces.

In addition to the sum of the subspaces A and B, we may define their **intersection** $A \cap B$ by

$$A \cap B=\{x \in V : x \in A \text{ and } x \in B\}.$$

Since A and B are subspaces, we see that for any $x, y \in A \cap B$ we have both $x+y \in A$ and $x+y \in B$ so that $x+y \in A \cap B$, and if $x \in A \cap B$ then $kx \in A$ and $kx \in B$ so that $kx \in A \cap B$. Since $0 \in A \cap B$, we then see that $A \cap B$ is a nonempty subspace of V. This can also be extended to any finite collection of subspaces of V.

Our next theorem shows that the dimension of the sum of A and B is just the sum of the dimensions of A and B minus the dimension of their intersection.

Theorem 1.11. *If A and B are subspaces of a finite-dimensional space V, then*

$$\dim(A+B)=\dim A+\dim B-\dim(A \cap B).$$

Proof. Since $A+B$ and $A\cap B$ are subspaces of V, it follows that both $A+B$ and $A\cap B$ are finite-dimensional (Theorem 1.9). We thus let $\dim A=m$, $\dim B=n$ and $\dim A \cap B=r$.

Let $\{u_1,\dots,u_r\}$ be a basis for $A \cap B$. By Theorem 1.10 there exists a set $\{v_1,\dots,v_{m-r}\}$ of linearly independent vectors in V such that

$$\{u_1,\dots,u_r,v_1,\dots,v_{m-r}\}$$

is a basis for A. Similarly, we have a basis

$$\{u_1,\dots,u_r,w_1,\dots,w_{n-r}\}$$

for B. It is clear that the set

$$\{u_1,\dots,u_r,v_1,\dots,v_{m-r},w_1,\dots,w_{n-r}\}$$

spans $A+B$ since any $a+b \in A+B$ (with $a \in A$ and $b \in B$) can be written as a linear combination of these $r+(m-r)+(n-r)=m+n-r$ vectors. To prove that they form a basis for $A+B$, we need only show that these $m+n-r$ vectors are linearly independent.

Suppose we have sets of scalars $\{a_i\}$, $\{b_j\}$ and $\{c_k\}$ such that

$$\sum_{i=1}^{r} a_i u_i + \sum_{j=1}^{m-r} b_j v_j + \sum_{k=1}^{n-r} c_k w_k = 0.$$

Then

$$\sum_{i=1}^{r} a_i u_i + \sum_{j=1}^{m-r} b_j v_j = -\sum_{k=1}^{n-r} c_k w_k.$$

Since the left side of this equation is an element of A while the right side is an element of B, their equality implies that they both belong to $A \cap B$, and hence

$$-\sum_{k=1}^{n-r} c_k w_k = \sum_{i=1}^{r} d_i u_i$$

for some set of scalars $\{d_i\}$. But $\{u_1, \dots, u_r, w_1, \dots, w_{n-r}\}$ forms a basis for B and hence they are linearly independent. Therefore, writing the above equation as

$$\sum_{i=1}^{r} d_i u_i + \sum_{k=1}^{n-r} c_k w_k = 0$$

implies that

$$d_1 = \cdots = d_r = c_1 = \cdots = c_{n-r} = 0.$$

We are now left with

$$\sum_{i=1}^{r} a_i u_i + \sum_{j=1}^{m-r} b_j v_j = 0.$$

But $\{u_1, \dots, u_r, v_1, \dots, v_{m-r}\}$ is also linearly independent so that

$$a_1 = \cdots = a_r = b_1 = \cdots = b_{m-r} = 0.$$

This proves that $\{u_1, \dots, u_r, v_1, \dots, v_{m-r}, w_1, \dots, w_{n-r}\}$ is linearly independent as claimed. The proof is completed by simply noting that we have shown

$$\dim(A + B) = m + n - r = \dim A + \dim B - \dim(A \cap B). \qquad \blacksquare$$

We now consider a particularly important special case of the sum. If A and B are subspaces of V such that $A \cap B = \{0\}$ and $V = A + B$, then we say that V is the **internal direct sum** of A and B. A completely equivalent way of defining the internal direct sum is given in the following theorem.

Theorem 1.12. *Let A and B be subspaces of a finite-dimensional vector space V. Then V is the internal direct sum of A and B if and only if every $v \in V$ can be uniquely written in the form $v = a + b$ where $a \in A$ and $b \in B$.*

Proof. Let us first assume that V is the internal direct sum of A and B. In other words, $V = A + B$ and $A \cap B = \{0\}$. Then by definition, for any $v \in V$ we have $v = a + b$ for some $a \in A$ and $b \in B$. Suppose we also have $v = a' + b'$ where $a' \in A$ and $b' \in B$. Then $a + b = a' + b'$ so that $a - a' = b' - b$. But note that $a - a' \in A$ and $b' - b \in B$, and hence the fact that $A \cap B = \{0\}$ implies that $a - a' = b' - b = 0$. Therefore $a = a'$ and $b = b'$ so that the expression for v is unique.

Conversely, suppose that every $v \in V$ may be written uniquely in the form $v = a + b$ with $a \in A$ and $b \in B$. This means that $V = A + B$, and we must still show that $A \cap B = \{0\}$. In particular, if $v \in A \cap B$ we may write $v = v + 0$ with $v \in A$ and $0 \in B$, or alternatively, we may write $v = 0 + v$ with $0 \in A$ and $v \in B$. But we are assuming that the expression for v is unique, and hence we must have $v = 0$ (since the contributions from A and B must be the same in both cases). Thus $A \cap B = \{0\}$ and the sum is direct. ∎

We emphasize that the internal direct sum is defined for two subspaces A and B of a given space V. As we stated above, this is because the addition of two vectors from distinct spaces is not defined. In spite of this, we now proceed to show that it is nevertheless possible to define the sum of two distinct vector spaces.

Let A and B be distinct vector spaces (over the same field $\mathcal{F}$, of course). While the sum of a vector in A and a vector in B makes no sense, we may relate these two spaces by considering the Cartesian product $A \times B$ defined as (see Section A.1)

$$A \times B = \{(a, b) : a \in A \text{ and } b \in B\}.$$

Using the ordered pairs (a, b), it is now easy to turn $A \times B$ into a vector space by making the following definitions (see Example 1.1).

First, we say that two elements (a, b) and (a', b') of $A \times B$ are equal if and only if $a = a'$ and $b = b'$. Next, we define addition and scalar multiplication in the obvious manner by

$$(a, b) + (a', b') = (a + a', b + b')$$

and

$$k(a, b) = (ka, kb).$$

We leave it as an exercise for the reader to show that with these definitions, the set $A \times B$ defines a vector space V over $\mathcal{F}$. This vector space is called the **external direct sum** of the spaces A and B, and is denoted by $A \oplus B$.

While the external direct sum was defined for arbitrary spaces A and B, there is no reason why this definition can not be applied to two subspaces of a larger space V. We now show that in such a case, the internal and external direct sums are isomorphic.

Theorem 1.13. *If V is the internal direct sum of A and B, then V is isomorphic to the external direct sum $A \oplus B$.*

Proof. If V is the internal direct sum of A and B, then any $v \in V$ may be written uniquely in the form $v = a + b$. This uniqueness allows us to define the mapping $\phi : V \to A \oplus B$ by

$$\phi(v) = \phi(a + b) = (a, b).$$

Since for any $v = a + b$ and $v' = a' + b'$, and for any scalar k we have

$$\phi(v + v') = (a + a', b + b') = (a, b) + (a', b') = \phi(v) + \phi(v')$$

and

$$\phi(kv) = (ka, kb) = k(a, b) = k\phi(v)$$

it follows that ϕ is a vector space homomorphism. It is clear that ϕ is surjective, since for any $(a, b) \in A \oplus B$ we have $\phi(v) = (a, b)$ where $v = a + b \in V$. Finally, if $\phi(v) = (0, 0)$ then we must have $a = b = 0 = v$ and hence $\operatorname{Ker} \phi = \{0\}$. This shows that ϕ is also injective (Theorem 1.5). In other words, we have shown that V is isomorphic to $A \oplus B$. ∎

Because of this theorem, we shall henceforth refer only to the **direct sum** of A and B, and denote this sum by $A \oplus B$. It follows trivially from Theorem 1.11 that

$$\dim(A \oplus B) = \dim A + \dim B.$$

Example 1.8. Consider the ordinary Euclidean three-space $V = \mathbb{R}^3$. Note that any $v \in \mathbb{R}^3$ may be written as

$$(v_1, v_2, v_3) = (v_1, v_2, 0) + (0, 0, v_3)$$

which is just the sum of a vector in the xy-plane and a vector on the z-axis. It should also be clear that the only vector in the intersection of the xy-plane with the z-axis is the zero vector. In other words, defining the space A to be the xy-plane $\mathbb{R}^2$ and the space B to be the z-axis $\mathbb{R}^1$, we see that $V = A \oplus B$ or $\mathbb{R}^3 = \mathbb{R}^2 \oplus \mathbb{R}^1$.

On the other hand, if we try to write $\mathbb{R}^3$ as the direct sum of the xy-plane A with say, the yz-plane B, then the intersection condition is violated since $A \cap B$ is the entire y-axis. In this case, any vector lying on the y-axis can be specified in terms of its components in either the xy-plane or in the yz-plane.

In many of our later applications we shall need to take the direct sum of several vector spaces. While it should be obvious that this follows simply by induction from the above case, we go through the details nevertheless. We say that a vector space V is the **direct sum** of the subspaces $W_1, \dots, W_r$ if the following properties are true:

(DS1) $W_i \neq \{0\}$ for each $i = 1, \dots, r$;
(DS2) $W_i \cap (W_1 + \cdots + W_{i-1} + W_{i+1} + \cdots + W_r) = \{0\}$ for $i = 1, \dots, r$;

(DS3) $V = W_1 + \cdots + W_r$.

If V is the direct sum of the W_i, then we write $V = W_1 \oplus \cdots \oplus W_r$. The generalization of Theorem 1.12 is the following.

Theorem 1.14. *If $W_1, \ldots, W_r$ are subspaces of V, then*

$$V = W_1 \oplus \cdots \oplus W_r$$

if and only if every $v \in V$ has a unique representation of the form

$$v = v_1 + \cdots + v_r$$

where $v_i \in W_i$ for each $i = 1, \ldots, r$.

Proof. First assume that V is the direct sum of $W_1, \ldots, W_r$. Given any $v \in V$, property (DS3) of the direct sum tells us that we have

$$v = v_1 + \cdots + v_r$$

where $v_i \in W_i$ for each $i = 1, \ldots, r$. If we also have another representation

$$v = v_1' + \cdots + v_r'$$

with $v_i' \in W_i$, then

$$v_1 + \cdots + v_r = v_1' + \cdots + v_r'$$

so that for any $i = 1, \ldots, r$ we have

$$v_i' - v_i = (v_1 - v_1') + \cdots + (v_{i-1} - v_{i-1}') + (v_{i+1} - v_{i+1}') + \cdots + (v_r - v_r').$$

Since $v_i' - v_i \in W_i$ and the right hand side of this equation is an element of $W_1 + \cdots + W_{i-1} + W_{i+1} + \cdots + W_r$, we see that (DS2) requires $v_i' - v_i = 0$, and hence $v_i' = v_i$. This proves the uniqueness of the representation.

Conversely, assume that each $v \in V$ has a unique representation of the form $v = v_1 + \cdots + v_r$ where $v_i \in W_i$ for each $i = 1, \ldots, r$. Since (DS3) is automatically satisfied, we must show that (DS2) is also satisfied. Suppose

$$v_1 \in W_1 \cap (W_2 + \cdots + W_r).$$

Since

$$v_1 \in W_2 + \cdots + W_r$$

we must also have

$$v_1 = v_2 + \cdots + v_r$$

for some $v_2 \in W_2, \ldots, v_r \in W_r$. But then

$$0 = -v_1 + v_2 + \cdots + v_r$$

and

$$0 = 0 + \cdots + 0$$

are two representations of the vector 0, and hence the uniqueness of the representations implies that $v_i = 0$ for each $i = 1, \ldots, r$. In particular, the case $i = 1$ means that

$$W_1 \cap (W_2 + \cdots + W_r) = \{0\}.$$

A similar argument applies to $W_i \cap (W_2 + \cdots + W_{i-1} + W_{i+1} + \cdots + W_r)$ for any $i = 1, \ldots, r$. This proves (DS2). ∎

If $V = W_1 \oplus \cdots \oplus W_r$, then it seems reasonable that we should be able to form a basis for V by adding up the bases of the subspaces W_i. This is indeed the case as we now show.

Theorem 1.15. *Let $W_1, \ldots, W_r$ be subspaces of V, and for each $i = 1, \ldots, r$ let W_i have basis $\mathcal{B}_i = \{w_{i1}, \ldots, w_{in_i}\}$. Then V is the direct sum of the W_i if and only if the union of bases*

$$\mathcal{B} = \bigcup_{i=1}^{r} \mathcal{B}_i = \{w_{11}, \ldots, w_{1n_1}, \ldots, w_{r1}, \ldots, w_{rn_r}\}$$

is a basis for V.

Proof. Suppose that $\mathcal{B}$ is a basis for V. Then for any $v \in V$ we may write

$$\begin{aligned} v &= (a_{11}w_{11} + \cdots + a_{1n_1}w_{1n_1}) + \cdots + (a_{r1}w_{r1} + \cdots + a_{rn_r}w_{rn_r}) \\ &= w_1 + \cdots + w_r \end{aligned}$$

where

$$w_i = a_{i1}w_{i1} + \cdots + a_{in_i}w_{in_i} \in W_i$$

and $a_{ij} \in \mathcal{F}$. Now let

$$v = w'_1 + \cdots + w'_r$$

be any other expansion of v, where each $w'_i \in W_i$. Using the fact that $\mathcal{B}_i$ is a basis for W_i we have

$$w'_i = b_{i1}w_{i1} + \cdots + b_{in_i}w_{in_i}$$

for some set of scalars b_{ij}. This means that we may also write

$$v = (b_{11}w_{11} + \cdots + b_{1n_1}w_{1n_1}) + \cdots + (b_{r1}w_{r1} + \cdots + b_{rn_r}w_{rn_r}).$$

However, since $\mathcal{B}$ is a basis for V, we may equate the coefficients of w_{ij} in these two expressions for v to obtain $a_{ij} = b_{ij}$ for all i, j. We have thus proved that the representation of v is unique, and hence Theorem 1.14 tells us that V is the direct sum of the W_i.

Now suppose that V is the direct sum of the W_i. This means that any $v \in V$ may be expressed in the unique form $v = w_1 + \cdots + w_r$ where $w_i \in W_i$ for each $i = 1, \ldots, r$. Given that $\mathcal{B}_i = \{w_{i1}, \ldots, w_{in_i}\}$ is a basis for W_i, we must show that $\mathcal{B} = \bigcup \mathcal{B}_i$ is a basis for V. We first note that each $w_i \in W_i$ may be expanded in terms of the members of $\mathcal{B}_i$, and therefore $\bigcup \mathcal{B}_i$ clearly spans V. It remains to show that the elements of $\mathcal{B}$ are linearly independent.

We first write

$$(c_{11}w_{11} + \cdots + c_{1n_1}w_{1n_1}) + \cdots + (c_{r1}w_{r1} + \cdots + c_{rn_r}w_{rn_r}) = 0$$

and note that

$$c_{i1}w_{i1} + \cdots + c_{in_i}w_{in_i} \in W_i.$$

Using the fact that $0+\cdots+0 = 0$ (where each $0 \in W_i$) along with the uniqueness of the representation in any direct sum, we see that for each $i = 1, \ldots, r$ we must have

$$c_{i1}w_{i1} + \cdots + c_{in_i}w_{in_i} = 0.$$

However, since $\mathcal{B}_i$ is a basis for W_i, this means that $c_{ij} = 0$ for every i and j, and hence the elements of $\mathcal{B} = \bigcup \mathcal{B}_i$ are linearly independent. ∎

Corollary. *If $V = W_1 \oplus \cdots \oplus W_r$, then*

$$\dim V = \sum_{i=1}^{r} \dim W_i.$$

Proof. Obvious from Theorem 1.15. This also follows by induction from Theorem 1.11. ∎

Exercises

1. Let W_1 and W_2 be subspaces of $\mathbb{R}^3$ defined by $W_1 = \{(x, y, z) : x = y = z\}$ and $W_2 = \{(x, y, z) : x = 0\}$. Show that $\mathbb{R}^3 = W_1 \oplus W_2$.

2. Let W_1 be any subspace of a finite-dimensional space V. Prove there exists a subspace W_2 of V such that $V = W_1 \oplus W_2$.

3. Let W_1, W_2 and W_3 be subspaces of a vector space V. Show that

$$(W_1 \cap W_2) + (W_1 \cap W_3) \subset W_1 \cap (W_2 + W_3).$$

Give an example in $V = \mathbb{R}^2$ for which equality does not hold.

4. Let $V = F[\mathbb{R}]$ be as in Exercise 1.2.6. Let W_+ and W_- be the subsets of V defined by $W_+ = \{f \in V : f(-x) = f(x)\}$ and $W_- = \{f \in V : f(-x) = -f(x)\}$. In other words, W_+ is the subset of all even functions, and W_- is the subset of all odd functions.

(a) Show that W_+ and W_- are subspaces of V.
(b) Show that $V = W_+ \oplus W_-$.

5. Let W_1 and W_2 be subspaces of a vector space V.

   (a) Show that $W_1 \subset W_1 + W_2$ and $W_2 \subset W_1 + W_2$.
   (b) Prove that $W_1 + W_2$ is the smallest subspace of V that contains both W_1 and W_2. In other words, if $\mathscr{S}(W_1, W_2)$ denotes the linear span of W_1 and W_2, show that $W_1 + W_2 = \mathscr{S}(W_1, W_2)$. [*Hint*: Show that $W_1 + W_2 \subset \mathscr{S}(W_1, W_2)$ and $\mathscr{S}(W_1, W_2) \subset W_1 + W_2$.]

6. Let V be a finite-dimensional vector space. For any $x \in V$, we define $\mathcal{F}x = \{ax : a \in \mathcal{F}\}$. Prove that $\{x_1, x_2, \ldots, x_n\}$ is a basis for V if and only if $V = \mathcal{F}x_1 \oplus \mathcal{F}x_2 \oplus \cdots \oplus \mathcal{F}x_n$.

7. If A and B are vector spaces, show that $A + B$ is the span of $A \cup B$.

1.5 Inner Product Spaces

Before proceeding with the general theory of inner products, let us briefly review what the reader should already know from more elementary courses. It is assumed that the reader is familiar with vectors in $\mathbb{R}^3$, and we show that for any $\vec{a}, \vec{b} \in \mathbb{R}^3$ the **scalar product** (also called the **dot product**) $\vec{a} \cdot \vec{b}$ may be written as either

$$\vec{a} \cdot \vec{b} = \sum_{i=1}^{3} a_i b_i$$

where $\{a_i\}$ and $\{b_i\}$ are the coordinates of $\vec{a}$ and $\vec{b}$ relative to the standard basis for $\mathbb{R}^3$ (see Example 1.7), or as

$$\vec{a} \cdot \vec{b} = \|\vec{a}\| \, \|\vec{b}\| \cos\theta$$

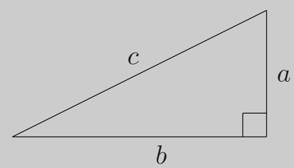
where $\theta = \angle(\vec{a}, \vec{b})$ and

$$\|\vec{a}\|^2 = \sum_{i=1}^{3} {a_i}^2$$

with a similar equation for $\|\vec{b}\|$. The symbol $\|\cdot\|$ is just the vector space generalization of the absolute value of numbers, and will be defined carefully below (see Example 1.9). For now, just think of $\|\vec{a}\|$ as meaning the length of the vector $\vec{a}$ in $\mathbb{R}^3$.

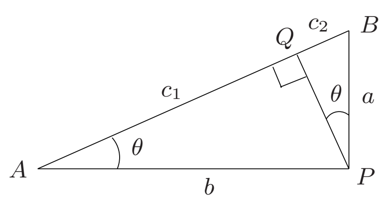
Just for fun, for the sake of completeness, and to show exactly what these equations depend on, we prove this as a series of simple lemmas. Our first lemma is known as the **Pythagorean theorem**.

Lemma 1.1. *Given a right triangle with sides a, b, and c as shown,*



we have $c^2 = a^2 + b^2$.

Proof. As shown in the figure below, we draw the line $\overline{PQ}$ perpendicular to the hypotenuse $c = \overline{AB}$. Note that we can now write c as the sum of the two parts $c_1 = \overline{AQ}$ and $c_2 = \overline{QB}$. First observe that $\triangle ABP$ is similar to $\triangle APQ$ because they are both right triangles and they have the angle θ in common (so they must have their third angle the same). Let us denote this similarity by $\triangle ABP \sim \triangle APQ$. If we let this third angle be $\alpha = \angle(ABP)$, then we also have $\alpha = \angle(APQ)$, and hence $\triangle ABP$, $\triangle APQ$ and $\triangle PBQ$ are all similar.



Using the fact that $\triangle APQ \sim \triangle ABP$ and $\triangle PBQ \sim \triangle ABP$ along with $c = c_1 + c_2$ we have

$$\frac{c_1}{b} = \frac{b}{c} \qquad \text{and} \qquad \frac{c_2}{a} = \frac{a}{c}$$

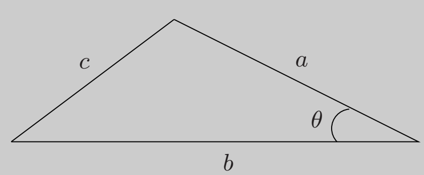
and therefore

$$c = c_1 + c_2 = \frac{a^2 + b^2}{c}$$

from which the lemma follows immediately. ∎

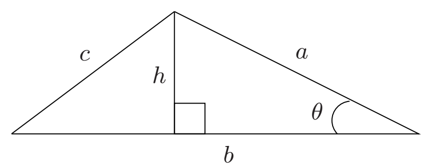
Our next lemma is known as the **law of cosines**. This result, together with Lemma 1.1, shows that for any triangle T with sides $c \le a \le b$, it is true that $a^2 + b^2 = c^2$ if and only if T is a right triangle.

Lemma 1.2. *For any triangle as shown,*



we have $c^2 = a^2 + b^2 - 2ab\cos\theta$.

Proof. Draw a perpendicular to side b as shown:



By the Pythagorean theorem we have

$$\begin{aligned} c^2 &= h^2 + (b - a\cos\theta)^2 \\ &= (a\sin\theta)^2 + (b - a\cos\theta)^2 \\ &= a^2\sin^2\theta + b^2 - 2ab\cos\theta + a^2\cos^2\theta \\ &= a^2 + b^2 - 2ab\cos\theta \end{aligned}$$

where we used $\sin^2\theta + \cos^2\theta = 1$ which follows directly from Lemma 1.1 with $a = c(\sin\theta)$ and $b = c(\cos\theta)$. ∎

We now *define* the scalar product $\vec{a} \cdot \vec{b}$ for any $\vec{a}, \vec{b} \in \mathbb{R}^3$ by

$$\vec{a} \cdot \vec{b} = \sum_{i=1}^{3} a_i b_i = \vec{b} \cdot \vec{a}$$

where $\vec{a} = (a_1, a_2, a_3)$ and $\vec{b} = (b_1, b_2, b_3)$. It is easy to see that

$$\vec{a} \cdot (\vec{b} + \vec{c}) = \sum_{i=1}^{3} a_i(b_i + c_i) = \sum_{i=1}^{3} (a_i b_i + a_i c_i) = \vec{a} \cdot b + \vec{a} \cdot \vec{c}$$

and similarly, it is easy to show that

$$(\vec{a} + \vec{b}) \cdot \vec{c} = \vec{a} \cdot \vec{c} + \vec{b} \cdot \vec{c}$$

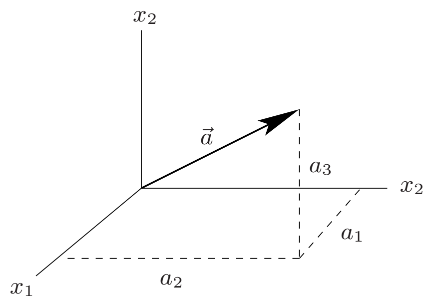
and

$$(k\vec{a}) \cdot \vec{b} = k(\vec{a} \cdot \vec{b})$$

where $k \in \mathbb{R}$.

From the figure below, we see the Pythagorean theorem also shows us that

$$\|\vec{a}\|^2 = \sum_{i=1}^{3} a_i a_i = \vec{a} \cdot \vec{a}.$$



This is the justification for writing $\|\vec{a}\|$ to mean the length of the vector $\vec{a} \in \mathbb{R}^3$.

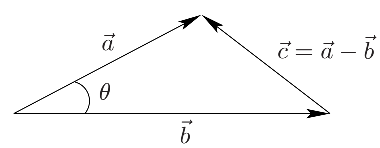
Noting that any two vectors (with a common origin) in $\mathbb{R}^3$ lie in a plane, we have the following well-known formula for the dot product.

Lemma 1.3. *For any $\vec{a}, \vec{b} \in \mathbb{R}^3$ we have*

$$\vec{a} \cdot \vec{b} = ab\cos\theta$$

where $a = \|\vec{a}\|$, $b = \|\vec{b}\|$ and $\theta = \angle(\vec{a}, \vec{b}\,)$.

Proof. Draw the vectors $\vec{a}$ and $\vec{b}$ along with their difference $\vec{c} = \vec{a} - \vec{b}$:

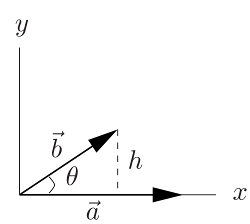


By the law of cosines we have $c^2 = a^2 + b^2 - 2ab\cos\theta$, while on the other hand

$$c^2 = \|\vec{a} - \vec{b}\|^2 = (\vec{a} - \vec{b}\,) \cdot (\vec{a} - \vec{b}\,) = a^2 + b^2 - 2\vec{a} \cdot \vec{b}.$$

Therefore we see that $\vec{a} \cdot \vec{b} = ab\cos\theta$. ∎

Another more intuitive way to see that $\vec{a} \cdot \vec{b} = ab\cos\theta$ is the following. Orient the coordinate system so that we have the vectors $\vec{a}$ and $\vec{b}$ in the xy-plane as shown below.



From this figure we see that $\vec{a} = (a, 0, 0)$ and $\vec{b} = (b\cos\theta, b\sin\theta, 0)$. But then $\vec{a}\cdot\vec{b} = a_1b_1 + a_2b_2 + a_3b_3 = ab\cos\theta$ as before. Since neither the length of a vector nor the angle between two vectors depends on the orientation of the coordinate system, this result must be true in general.

The main reason that we went through all of this is to motivate the generalization to arbitrary vector spaces. For example, if $u, v \in \mathbb{R}^n$, then to say that

$$u \cdot v = \sum_{i=1}^{n} u_i v_i$$

makes sense, whereas to say that $u \cdot v = \|u\| \, \|v\| \cos\theta$ leaves one wondering just what the "angle" θ means in higher dimensions. In fact, this will be used to *define* the angle θ.

We now proceed to define a general scalar (or inner) product $\langle u, v\rangle$ of vectors $u, v \in V$. Throughout this section, we let V be a vector space over either the real field $\mathbb{R}$ or the complex field $\mathbb{C}$. By way of motivation, we will want the inner product $\langle \cdot\,, \cdot\rangle$ applied to a single vector $v \in V$ to yield the length (or norm) of v, so that $\|v\|^2 = \langle v, v\rangle$. But $\|v\|$ must be a real number even if the field we are working with is $\mathbb{C}$. Noting that for any complex number $z \in \mathbb{C}$ we have $|z|^2 = zz^*$, we are led to make the following definition.

Let V be a vector space over $\mathcal{F}$ (where $\mathcal{F}$ is either $\mathbb{R}$ or $\mathbb{C}$). By an **inner product** on V (sometimes called the **Hermitian inner product**), we mean a mapping $\langle \cdot\,, \cdot\rangle : V \times V \to \mathcal{F}$ such that for all $u, v, w \in V$ and $a, b \in \mathcal{F}$ we have

(IP1) $\langle au + bv, w\rangle = a^*\langle u, w\rangle + b^*\langle v, w\rangle$;
(IP2) $\langle u, v\rangle = \langle v, u\rangle^*$;
(IP3) $\langle u, u\rangle \geq 0$ and $\langle u, u\rangle = 0$ if and only if $u = 0$.

Using these properties, we also see that

$$\begin{aligned} \langle u, av + bw\rangle &= \langle av + bw, u\rangle^* \\ &= (a^*\langle v, u\rangle + b^*\langle w, u\rangle)^* \\ &= a\langle u, v\rangle + b\langle u, w\rangle \end{aligned}$$

and hence, for the sake of reference, we call this

(IP1′) $\langle u, av + bw\rangle = a\langle u, v\rangle + b\langle u, w\rangle$.

(The reader should be aware that instead of $\langle au, v\rangle = a^*\langle u, v\rangle$, many authors define $\langle au, v\rangle = a\langle u, v\rangle$ and $\langle u, av\rangle = a^*\langle u, v\rangle$. This is particularly true in mathematics texts, whereas we have chosen the convention used by most physics texts. Of course, this has no effect on any of our results.)

Another remark that is worth pointing out is this. Our condition (IP3), that $\langle u, u\rangle \geq 0$ and $\langle u, u\rangle = 0$ if and only if $u = 0$ is sometimes called a **positive definite** inner product. If condition (IP3) is dropped entirely, we obtain an **indefinite** inner product, but this is rarely used (at least by physicists and engineers). However, if we replace (IP3) by the weaker requirement

(IP3′) $\langle u, v\rangle = 0$ for all $v \in V$ if and only if $u = 0$

then we obtain what is called a **nondegenerate** inner product. For example, the Minkowski space of special relativity has the property that any lightlike vector $v \neq 0$ still has the property that $\langle v, v\rangle = 0$.

A space V together with an inner product is called an **inner product space**. If V is an inner product space over the field $\mathbb{C}$, then V is called a **complex** inner product space, and if the field is $\mathbb{R}$, then V is called a **real** inner product space. A complex inner product space is frequently called a **unitary space**, and a real inner product space is frequently called a **Euclidean space**. Note that in the case of a real space, the complex conjugates in (IP1) and (IP2) are superfluous.

By (IP2) we have $\langle u, u\rangle \in \mathbb{R}$ so that we may define the **length** (or **norm**) of u to be the nonnegative real number

$$\|u\| = \langle u, u\rangle^{1/2}.$$

If $\|u\| = 1$, then u is said to be a **unit vector**. If $\|v\| \neq 0$, then we can normalize v by setting $u = v/\|v\|$. One sometimes writes $\hat{v}$ to mean the unit vector in the direction of v, i.e., $v = \|v\|\,\hat{v}$.

Example 1.9. Let $X = (x_1, \dots, x_n)$ and $Y = (y_1, \dots, y_n)$ be vectors in $\mathbb{C}^n$. We define

$$\langle X, Y\rangle = \sum_{i=1}^{n} x_i^* y_i$$

and leave it to the reader to show that this satisfies (IP1)–(IP3). In the case of the space $\mathbb{R}^n$, we have $\langle X, Y\rangle = X \cdot Y = \sum x_i y_i$. This inner product is called the **standard inner product** in $\mathbb{C}^n$ (or $\mathbb{R}^n$).

We also see that if $X, Y \in \mathbb{R}^n$ then

$$\|X - Y\|^2 = \langle X - Y, X - Y\rangle = \sum_{i=1}^{n} (x_i - y_i)^2.$$

Thus $\|X - Y\|$ is indeed just the distance between the points $X = (x_1, \dots, x_n)$ and $Y = (y_1, \dots, y_n)$ that we would expect by applying the Pythagorean theorem to points in $\mathbb{R}^n$. In particular, $\|X\|$ is simply the length of the vector X.

It is now easy to see why we defined the inner product as we did. For example, consider simply the space $\mathbb{C}^3$. Then with respect to the standard inner product on $\mathbb{C}^3$, the vector $X = (1, i, 0)$ will have norm $\|X\|^2 = \langle X, X\rangle = 1 + 1 + 0 = 2$, while if we had used the expression corresponding to the standard inner product on $\mathbb{R}^3$, we would have found $\|X\|^2 = 1 - 1 + 0 = 0$ even though $X \neq 0$.

Example 1.10. Let V be the vector space of continuous complex-valued functions defined on the real interval $[a,b]$. We may define an inner product on V by

$$\langle f,g\rangle = \int_a^b f^*(x)g(x)\,dx$$

for all $f,g \in V$. It should be obvious that this satisfies the three required properties of an inner product.

In Appendix A (see Theorem A.7) we proved an elementary result that essentially entailed taking the inner product of vectors in $\mathbb{C}^n$. We now generalize this to an important result known as the **Cauchy-Schwartz inequality**.

Theorem 1.16 (**Cauchy-Schwartz**). *Let V be an inner product space. Then for any $u,v \in V$ we have*

$$\left|\langle u,v\rangle\right| \le \|u\|\,\|v\|\,.$$

Proof. If either u or v is zero the theorem is trivially true. We therefore assume that $u \neq 0$ and $v \neq 0$. Then, for any real number c, we have (using (IP2) and the fact that $|z|^2 = zz^*$)

$$\begin{aligned}
0 &\le \|v - c\langle u,v\rangle u\|^2 \\
&= \langle v - c\langle u,v\rangle u, v - c\langle u,v\rangle u\rangle \\
&= \langle v,v\rangle - c\langle u,v\rangle\langle v,u\rangle - c\langle u,v\rangle^*\langle u,v\rangle + c^2\langle u,v\rangle^*\langle u,v\rangle\langle u,u\rangle \\
&= \|v\|^2 - 2c\left|\langle u,v\rangle\right|^2 + c^2\left|\langle u,v\rangle\right|^2\|u\|^2\,.
\end{aligned}$$

Now let $c = 1/\|u\|^2$ to obtain

$$0 \le \|v\|^2 - \frac{\left|\langle u,v\rangle\right|^2}{\|u\|^2}$$

or

$$\left|\langle u,v\rangle\right|^2 \le \|u\|^2\,\|v\|^2\,.$$

Taking the square root proves the theorem. ∎

We have seen that an inner product may be used to define a norm on V. In fact, the norm has several properties that may be used to define a normed vector space axiomatically as we see from the next theorem.

Theorem 1.17. *The norm in an inner product space V has the following properties for all $u,v \in V$ and $k \in \mathcal{F}$:*

(N1) $\|u\| \geq 0$ *and* $\|u\| = 0$ *if and only if* $u = 0$.
(N2) $\|ku\| = |k| \, \|u\|$.
(N3) $\|u + v\| \leq \|u\| + \|v\|$.

Proof. Since $\|u\| = \langle u, u \rangle^{1/2}$, (N1) follows from (IP3). Next, we see that

$$\|ku\|^2 = \langle ku, ku \rangle = |k|^2 \, \|u\|^2$$

and hence taking the square root yields (N2). Finally, using Theorem 1.16 and the fact that $z + z^* = 2 \operatorname{Re} z \leq 2 \, |z|$ for any $z \in \mathbb{C}$, we have

$$\begin{aligned}
\|u + v\|^2 &= \langle u + v, u + v \rangle \\
&= \langle u, u \rangle + \langle u, v \rangle + \langle v, u \rangle + \langle v, v \rangle \\
&= \|u\|^2 + \langle u, v \rangle + \langle u, v \rangle^* + \|v\|^2 \\
&\leq \|u\|^2 + 2 \, |\langle u, v \rangle| + \|v\|^2 \\
&\leq \|u\|^2 + 2 \, \|u\| \, \|v\| + \|v\|^2 \\
&= \big(\|u\| + \|v\| \big)^2 .
\end{aligned}$$

Taking the square root yields (N3). ∎

We note that property (N3) is frequently called the **triangle inequality** because in two or three dimensions, it simply says that the sum of two sides of a triangle is greater than the third. Furthermore, we note that properties (N1)–(N3) may be used to *define* a normed vector space. In other words, a **normed vector space** is defined to be a vector space V together with a mapping $\| \cdot \| : V \to \mathbb{R}$ that obeys properties (N1)–(N3). While a normed space V does not in general have an inner product defined on it, the existence of an inner product leads in a natural way (i.e., by Theorem 1.17) to the existence of a norm on V.

Example 1.11. Let us prove a simple but useful result dealing with the norm in any normed space V. From the properties of the norm, we see that for any $u, v \in V$ we have

$$\|u\| = \|u - v + v\| \leq \|u - v\| + \|v\|$$

and

$$\|v\| = \|v - u + u\| \leq \|u - v\| + \|u\| .$$

Rearranging each of these yields

$$\|u\| - \|v\| \leq \|u - v\|$$

and

$$\|v\| - \|u\| \leq \|u - v\| .$$

This shows that

$$\Big| \, \|u\| - \|v\| \, \Big| \leq \|u - v\| .$$

Example 1.12. Consider the space V of Example 1.10 and the associated inner product $\langle f, g\rangle$. Applying the Cauchy-Schwartz inequality (Theorem 1.16) we have

$$\left|\int_a^b f^*(x)g(x)\,dx\right| \le \left[\int_a^b |f(x)|^2 dx\right]^{1/2} \left[\int_a^b |g(x)|^2 dx\right]^{1/2}.$$

From property (N3) in Theorem 1.17 we see that $\|f+g\| \le \|f\| + \|g\|$ or $\langle f+g, f+g\rangle^{1/2} \le \langle f, f\rangle^{1/2} + \langle g, g\rangle^{1/2}$ which becomes

$$\left[\int_a^b |f(x)+g(x)|^2 dx\right]^{1/2} \le \left[\int_a^b |f(x)|^2 dx\right]^{1/2} + \left[\int_a^b |g(x)|^2 dx\right]^{1/2}.$$

The reader might try and prove either of these directly from the definition of the integral if he or she wants to gain an appreciation of the power of the axiomatic approach to inner products.

Finally, let us finish our generalization of Lemmas 1.1–1.3. If we repeat the proof of Lemma 1.3 using the inner product and norm notations, we find that for any $u, v \in \mathbb{R}^3$ we have $\langle u, v\rangle = \|u\|\,\|v\|\cos\theta$. Now let V be any real vector space. We define the **angle** θ between two nonzero vectors $u, v \in V$ by

$$\cos\theta = \frac{\langle u, v\rangle}{\|u\|\,\|v\|}.$$

Note that $\cos\theta \le 1$ by Theorem 1.16 so this definition makes sense. We say that u is **orthogonal** (or **perpendicular**) to v if $\langle u, v\rangle = 0$. If u and v are orthogonal, we often write this as $u \perp v$. From the basic properties of the inner product, it then follows that $\langle v, u\rangle = \langle u, v\rangle^* = 0^* = 0$ so that v is orthogonal to u also. Thus $u \perp v$ if and only if $\cos\theta = 0$. While $\cos\theta$ is only defined in a real vector space, our definition of orthogonality is valid in any space V over $\mathcal{F}$.

Exercises

1. Let $x = (x_1, x_2)$ and $y = (y_1, y_2)$ be vectors in $\mathbb{R}^2$, and define the mapping $\langle\cdot\,,\cdot\rangle : \mathbb{R}^2 \to \mathbb{R}$ by $\langle x, y\rangle = x_1y_1 - x_1y_2 - x_2y_1 + 3x_2y_2$. Show this defines an inner product on $\mathbb{R}^2$.

2. Let $x = (3, 4) \in \mathbb{R}^2$, and evaluate $\|x\|$ with respect to the norm induced by:
   (a) The standard inner product on $\mathbb{R}^2$.
   (b) The inner product defined in the previous exercise.

3. Let V be an inner product space, and let $x, y \in V$.

(a) Prove the **parallelogram law**:

$$\|x+y\|^2 + \|x-y\|^2 = 2\|x\|^2 + 2\|y\|^2 .$$

(The geometric meaning of this equation is that the sum of the squares of the diagonals of a parallelogram is equal to the sum of the squares of the sides.)

(b) Prove the **Pythagorean theorem**:

$$\|x+y\|^2 = \|x\|^2 + \|y\|^2 \quad \text{if } x \perp y.$$

4. Find a unit vector orthogonal to the vectors $x = (1,1,2)$ and $y = (0,1,3)$ in $\mathbb{R}^3$.

5. Let $u = (z_1, z_2)$ and $v = (w_1, w_2)$ be vectors in $\mathbb{C}^2$, and define the mapping $\langle\cdot\,,\cdot\rangle : \mathbb{C}^2 \to \mathbb{R}$ by

$$\langle u, v\rangle = z_1 w_1^* + (1+i) z_1 w_2^* + (1-i) z_2 w_1^* + 3 z_2 w_2^*.$$

Show this defines an inner product on $\mathbb{C}^2$.

6. Let $u = (1-2i, 2+3i) \in \mathbb{C}^2$ and evaluate $\|u\|$ with respect to the norm induced by:
   (a) The standard norm on $\mathbb{C}^2$.
   (b) The inner product defined in the previous exercise.

7. Let V be an inner product space. Verify the following **polar form identities**:
   (a) If V is a real space and $x, y \in V$, then

$$\langle x, y\rangle = \frac{1}{4}\big(\,\|x+y\|^2 - \|x-y\|^2\,\big).$$

   (b) If V is a complex space and $x, y \in V$, then

$$\langle x, y\rangle = \frac{1}{4}\big(\,\|x+y\|^2 - \|x-y\|^2\,\big) + \frac{i}{4}\big(\,\|ix+y\|^2 - \|ix-y\|^2\,\big)$$

(If we were using instead the inner product defined by $\langle \alpha x, y\rangle = \alpha\langle x, y\rangle$, then the last two terms in this equation would read $\|x \pm iy\|$.)

8. Let $V = C[0,1]$ be the space of continuous real-valued functions defined on the interval $[0,1]$. Define an inner product on $C[0,1]$ by

$$\langle f, g\rangle = \int_0^1 f(t) g(t)\, dt.$$

   (a) Verify that this does indeed define an inner product on V.
   (b) Evaluate $\|f\|$ where $f = t^2 - 2t + 3 \in V$.

9. Given a vector space V, we define a mapping $d : V \times V \to \mathbb{R}$ by $d(x, y) = \|x - y\|$ for all $x, y \in V$. Show that:

   (a) $d(x, y) \geq 0$ and $d(x, y) = 0$ if and only if $x = y$.
   (b) $d(x, y) = d(y, x)$.
   (c) $d(x, z) \leq d(x, y) + d(y, z)$ (triangle inequality).

   The number $d(x, y)$ is called the **distance** from x to y, and the mapping d is called a **metric** on V. Any arbitrary set S on which we have defined a function $d : S \times S \to \mathbb{R}$ satisfying these three properties is called a **metric space**.

10. Let $\{e_1, \ldots, e_n\}$ be an orthonormal basis for a complex space V, and let $x \in V$ be arbitrary. Show

    (a) $x = \sum_{i=1}^{n} e_i \langle e_i, x\rangle$.

    (b) $\|x\|^2 = \sum_{i=1}^{n} \left|\langle e_i, x\rangle\right|^2$.

11. Show equality holds in the Cauchy-Schwartz inequality if and only if one vector is proportional to the other.

1.6 Orthogonal Sets

If a vector space V is equipped with an inner product, then we may define a subspace of V that will turn out to be extremely useful in a wide variety of applications. Let W be any subset of such a vector space V. (Note that W need not be a subspace of V.) We define the **orthogonal complement** of W to be the set $W^\perp$ given by

$$W^\perp = \{v \in V : \langle v, w\rangle = 0 \text{ for all } w \in W\}.$$

Theorem 1.18. *Let W be any subset of a vector space V. Then $W^\perp$ is a subspace of V.*

Proof. We first note that $0 \in W^\perp$ since for any $v \in V$ we have

$$\langle 0, v\rangle = \langle 0v, v\rangle = 0\langle v, v\rangle = 0.$$

To finish the proof, we simply note that for any $u, v \in W^\perp$, for any scalars $a, b \in \mathcal{F}$, and for every $w \in W$ we have

$$\langle au + bv, w\rangle = a^*\langle u, w\rangle + b^*\langle v, w\rangle = a^*0 + b^*0 = 0$$

so that $au + bv \in W^\perp$. ∎

Consider the space $\mathbb{R}^3$ with the usual Cartesian coordinate system (x, y, z). If we let $W = \mathbb{R}^2$ be the xy-plane, then $W^\perp = \mathbb{R}^1$ is just the z-axis since the standard inner product on $\mathbb{R}^3$ shows that any $v \in \mathbb{R}^3$ of the form $(0, 0, c)$ is orthogonal to any $w \in \mathbb{R}^3$ of the form $(a, b, 0)$. Thus, in this case anyway, we see that $W \oplus W^\perp = \mathbb{R}^3$. We will shortly prove that $W \oplus W^\perp = V$ for any inner product space V and subspace $W \subset V$. Before we can do this however, we must first discuss orthonormal sets of vectors.

A set $\{v_i\}$ of nonzero vectors in a space V is said to be an **orthogonal set** (or to be **mutually orthogonal**) if $\langle v_i, v_j\rangle = 0$ for $i \neq j$. If in addition, each v_i is a unit vector, then the set $\{v_i\}$ is said to be an **orthonormal set** and we write

$$\langle v_i, v_j\rangle = \delta_{ij}$$

where the very useful symbol δ_{ij} (called the **Kronecker delta**) is defined by

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j. \end{cases}$$

Theorem 1.19. *Any orthonormal set of vectors $\{v_i\}$ is linearly independent.*

Proof. If $\sum_i a_i v_i = 0$ for some set of scalars $\{a_i\}$, then

$$0 = \langle v_j, 0\rangle = \left\langle v_j, \sum_i a_i v_i \right\rangle = \sum_i a_i \langle v_j, v_i\rangle = \sum_i a_i \delta_{ij} = a_j$$

so that $a_j = 0$ for each j, and hence $\{v_i\}$ is linearly independent. ∎

Note that in the proof of Theorem 1.19 it was not really necessary that each v_i be a unit vector. Any orthogonal set would work just as well.

Theorem 1.20. *If $\{v_1, v_2, \dots, v_n\}$ is an orthonormal set in V and if $w \in V$ is arbitrary, then the vector*

$$u = w - \sum_i \langle v_i, w\rangle v_i$$

is orthogonal to each of the v_i.

Proof. We simply compute $\langle v_j, u\rangle$:

$$\begin{aligned} \langle v_j, u\rangle &= \left\langle v_j, w - \sum_i \langle v_i, w\rangle v_i \right\rangle \\ &= \langle v_j, w\rangle - \sum_i \langle v_i, w\rangle \langle v_j, v_i\rangle \end{aligned}$$

$$= \langle v_j, w\rangle - \sum_i \langle v_i, w\rangle \delta_{ij}$$
$$= \langle v_j w\rangle - \langle v_j, w\rangle = 0. \qquad \blacksquare$$

The numbers $c_i = \langle v_i, w\rangle$ are frequently called the **Fourier coefficients** of w with respect to v_i. In fact, we leave it as an exercise for the reader to show that the expression $\|w - \sum_i a_i v_i\|$ achieves its minimum precisely when $a_i = c_i$ (see Exercise 1.6.4). Furthermore, we also leave it to the reader (see Exercise 1.6.5) to show that

$$\sum_{i=1}^{n} |c_i|^2 \le \|w\|^2$$

which is called **Bessel's inequality**.

As we remarked earlier, most mathematics texts write $\langle u, av\rangle = a^*\langle u, v\rangle$ rather than $\langle u, av\rangle = a\langle u, v\rangle$. In this case, Theorem 1.20 would be changed to read that the vector

$$u = w - \sum_i \langle w, v_i\rangle v_i$$

is orthogonal to each v_j.

Example 1.13. The simplest and best known example of an orthonormal set is the set $\{e_i\}$ of standard basis vectors in $\mathbb{R}^n$. Thus

$$\begin{aligned} e_1 &= (1, 0, 0, \ldots, 0) \\ e_2 &= (0, 1, 0, \ldots, 0) \\ &\vdots \\ e_n &= (0, 0, 0, \ldots, 1) \end{aligned}$$

and clearly

$$\langle e_i, e_j\rangle = e_i \cdot e_j = \delta_{ij}$$

since for any $X = (x_1, \ldots, x_n)$ and $Y = (y_1, \ldots, y_n)$ in $\mathbb{R}^n$, we have

$$\langle X, Y\rangle = X \cdot Y = \sum_{i=1}^{n} x_i y_i.$$

(It would perhaps be better to write the unit vectors as $\hat{e}_i$ rather than e_i, but this will generally not cause any confusion.)

Example 1.14. Let V be the space of continuous complex-valued functions defined on the real interval $[-\pi, \pi]$. As in Example 1.10, we define

$$\langle f, g\rangle = \int_{-\pi}^{\pi} f^*(x) g(x)\, dx$$

for all $f, g \in V$. We show that the set of functions

$$f_n = \left(\frac{1}{2\pi}\right)^{1/2} e^{inx}$$

for $n = 1, 2, \dots$ forms an orthonormal set.

If $m = n$, then

$$\langle f_m, f_n \rangle = \langle f_n, f_n \rangle = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-inx} e^{inx}\, dx = \frac{1}{2\pi} \int_{-\pi}^{\pi} dx = 1.$$

If $m \neq n$, then we have

$$\begin{aligned} \langle f_m, f_n \rangle &= \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-imx} e^{inx}\, dx = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{i(n-m)x}\, dx \\ &= \frac{1}{2\pi} \frac{e^{i(n-m)x}}{i(n-m)} \bigg|_{-\pi}^{\pi} \\ &= \frac{\sin(n-m)\pi}{\pi(n-m)} = 0 \end{aligned}$$

since $\sin n\pi = 0$ for any integer n. (Note that we also used the fact that

$$\sin\theta = \frac{e^{i\theta} - e^{-i\theta}}{2i}$$

which follows from the Euler formula mentioned in Appendix A.) Therefore $\langle f_m, f_n \rangle = \delta_{mn}$. That the set $\{f_n\}$ is orthonormal is of great use in the theory of Fourier series.

We now wish to show that every finite-dimensional vector space with an inner product has an orthonormal basis. The proof is based on the famous Gram-Schmidt orthogonalization process, the precise statement of which we present as a corollary following the proof.

Theorem 1.21. *Let V be a finite-dimensional inner product space. Then there exists an orthonormal set of vectors that forms a basis for V.*

Proof. Let $\dim V = n$ and let $\{u_1, \dots, u_n\}$ be a basis for V. We will construct a new basis $\{w_1, \dots, w_n\}$ such that $\langle w_i, w_j \rangle = \delta_{ij}$. To begin, we choose

$$w_1 = \frac{u_1}{\|u_1\|}$$

so that

$$\|w_1\|^2 = \langle w_1, w_1 \rangle = \langle u_1/\|u_1\|, u_1/\|u_1\| \rangle = \langle u_1, u_1 \rangle / \|u_1\|^2$$

$$= \|u_1\|^2 \,/\, \|u_1\|^2 = 1$$

and hence w_1 is a unit vector. We now take u_2 and subtract off its "projection" along w_1. This will leave us with a new vector v_2 that is orthogonal to w_1. Thus, we define

$$v_2 = u_2 - \langle w_1, u_2 \rangle w_1$$

so that

$$\langle w_1, v_2 \rangle = \langle w_1, u_2 \rangle - \langle w_1, u_2 \rangle \langle w_1, w_1 \rangle = 0$$

(this also follows from Theorem 1.20). If we let

$$w_2 = \frac{v_2}{\|v_2\|}$$

then $\{w_1, w_2\}$ is an orthonormal set . (That $v_2 \neq 0$ will be shown below.)

We now go to u_3 and subtract off its projection along w_1 and w_2. In other words, we define

$$v_3 = u_3 - \langle w_2, u_3 \rangle w_2 - \langle w_1, u_3 \rangle w_1$$

so that $\langle w_1, v_3 \rangle = \langle w_2, v_3 \rangle = 0$. Choosing

$$w_3 = \frac{v_3}{\|v_3\|}$$

we now have an orthonormal set $\{w_1, w_2, w_3\}$.

It is now clear that given an orthonormal set $\{w_1, \ldots, w_k\}$, we let

$$v_{k+1} = u_{k+1} - \sum_{i=1}^{k} \langle w_i, u_{k+1} \rangle w_i$$

so that v_{k+1} is orthogonal to $w_1, \ldots, w_k$ (Theorem 1.20), and hence we define

$$w_{k+1} = \frac{v_{k+1}}{\|v_{k+1}\|}.$$

It should now be obvious that we can construct an orthonormal set of n vectors from our original basis of n vectors. To finish the proof, we need only show that $w_1, \ldots, w_n$ are linearly independent.

To see this, note first that since u_1 and u_2 are linearly independent, w_1 and u_2 must also be linearly independent, and hence $v_2 \neq 0$ by definition of linear independence. Thus w_2 exists and $\{w_1, w_2\}$ is linearly independent by Theorem 1.19. Next, $\{w_1, w_2, u_3\}$ is linearly independent since w_1 and w_2 are in the linear span of u_1 and u_2. Hence $v_3 \neq 0$ so that w_3 exists, and Theorem 1.19 again shows that $\{w_1, w_2, w_3\}$ is linearly independent.

In general then, if $\{w_1, \ldots, w_k\}$ is linearly independent, it follows that the set $\{w_1, \ldots, w_k, u_{k+1}\}$ is also independent since $\{w_1, \ldots, w_k\}$ is in the linear span of $\{u_1, \ldots, u_k\}$. Hence $v_{k+1} \neq 0$ and w_{k+1} exists. Then $\{w_1, \ldots, w_{k+1}\}$ is linearly independent by Theorem 1.19. Thus $\{w_1, \ldots, w_n\}$ forms a basis for V, and $\langle w_i, w_j \rangle = \delta_{ij}$. ∎

Corollary (Gram-Schmidt process). *Let $\{u_1, \dots, u_n\}$ be a linearly independent set of vectors in an inner product space V. Then there exists a set of orthonormal vectors $w_1, \dots, w_n \in V$ such that the linear span of $\{u_1, \dots, u_k\}$ is equal to the linear span of $\{w_1, \dots, w_k\}$ for each $k = 1, \dots, n$.*

Proof. This corollary follows by a careful inspection of the proof of Theorem 1.21. ∎

We emphasize that the Gram-Schmidt algorithm (the "orthogonalization process" of Theorem 1.21) as such applies to any inner product space, and is not restricted to only finite-dimensional spaces.

Example 1.15. Consider the following basis vectors for $\mathbb{R}^3$:

$$u_1 = (3, 0, 4) \qquad u_2 = (-1, 0, 7) \qquad u_3 = (2, 9, 11).$$

Let us apply the Gram-Schmidt process (with the standard inner product on $\mathbb{R}^3$) to obtain a new orthonormal basis for $\mathbb{R}^3$.

Since $\|u_1\| = \sqrt{9 + 16} = 5$, we define

$$w_1 = u_1/5 = (3/5, 0, 4/5).$$

Next, using $\langle w_1, u_2 \rangle = -3/5 + 28/5 = 5$ we let

$$v_2 = (-1, 0, 7) - (3, 0, 4) = (-4, 0, 3).$$

Since $\|v_2\| = 5$, we have

$$w_2 = (-4/5, 0, 3/5).$$

Finally, using $\langle w_1, u_3 \rangle = 10$ and $\langle w_2, u_3 \rangle = 5$ we let

$$v_3 = (2, 9, 11) - (-4, 0, 3) - (6, 0, 8) = (0, 9, 0)$$

and hence, since $\|v_3\| = 9$, our third basis vector becomes

$$w_3 = (0, 1, 0).$$

We leave it to the reader to show that $\{w_1, w_2, w_3\}$ does indeed form an orthonormal basis for $\mathbb{R}^3$.

We are now ready to prove our earlier assertion. Note that here we require W to be a subspace of V.

Theorem 1.22. *Let W be a subspace of a finite-dimensional inner product space V. Then $V = W \oplus W^\perp$.*

Proof. By Theorem 1.9, W is finite-dimensional. Therefore, if we choose a basis $\{v_1, \dots, v_k\}$ for W, it may be extended to a basis $\{v_1, \dots, v_n\}$ for V (Theorem 1.10). Applying Theorem 1.21 to this basis, we construct a new orthonormal basis $\{u_1, \dots, u_n\}$ for V where

$$u_r = \sum_{j=1}^{r} a_{rj} v_j$$

for $r = 1, \dots, n$ and some coefficients a_{rj} (determined by the Gram-Schmidt process). In particular, we see that $u_1, \dots, u_k$ are all in W, and hence they form an orthonormal basis for W.

Since $\{u_1, \dots, u_n\}$ are orthonormal, it follows that $u_{k+1}, \dots, u_n$ are in $W^\perp$ (since $\langle u_i, u_j \rangle = 0$ for all $i \le k$ and any $j = k+1, \dots, n$). Therefore, given any $x \in V$ we have

$$x = a_1 u_1 + \cdots + a_n u_n$$

where

$$a_1 u_1 + \cdots + a_k u_k \in W$$

and

$$a_{k+1} u_{k+1} + \cdots + a_n u_n \in W^\perp.$$

This means that $V = W + W^\perp$, and we must still show that $W \cap W^\perp = \{0\}$. But if $y \in W \cap W^\perp$, then $\langle y, y \rangle = 0$ since $y \in W^\perp$ implies that y is orthogonal to any vector in W, and in particular, $y \in W$. Hence $y = 0$ by (IP3), and it therefore follows that $W \cap W^\perp = \{0\}$. ∎

Corollary. *If V is finite-dimensional and W is a subspace of V, then $(W^\perp)^\perp = W$.*

Proof. Given any $w \in W$ we have $\langle w, v \rangle = 0$ for all $v \in W^\perp$. This implies that $w \in (W^\perp)^\perp$ and hence $W \subset (W^\perp)^\perp$. By Theorem 1.22, $V = W \oplus W^\perp$ and hence

$$\dim V = \dim W + \dim W^\perp$$

(Theorem 1.11). But $W^\perp$ is also a subspace of V, and hence $V = W^\perp \oplus (W^\perp)^\perp$ (Theorem 1.22) which implies

$$\dim V = \dim W^\perp + \dim (W^\perp)^\perp.$$

Therefore, comparing these last two equations shows that $\dim W = \dim (W^\perp)^\perp$. This result together with $W \subset (W^\perp)^\perp$ implies that $W = (W^\perp)^\perp$. ∎

Finally, note that if $\{e_i\}$ is an orthonormal basis for V, then any $x \in V$ may be written as $x = \sum_i x_i e_i$ where

$$\langle e_j, x\rangle = \left\langle e_j, \sum_i x_i e_i \right\rangle = \sum_i x_i \langle e_j, e_i\rangle = \sum_i x_i \delta_{ij} = x_j .$$

Therefore we may write

$$x = \sum_i \langle e_i, x\rangle e_i$$

which is a very useful expression.

We will have much more to say about inner product spaces after we have treated linear transformations in detail. For the rest of this book, unless explicitly stated otherwise, all vector spaces will be assumed to be finite-dimensional. In addition, the specific scalar field $\mathcal{F}$ will generally not be mentioned, but it is to be understood that all scalars are elements of $\mathcal{F}$.

Exercises

1. Let W be a subset of a vector space V. Prove the following:
   (a) $0^\perp = V$ and $V^\perp = 0$.
   (b) $W \cap W^\perp = \{0\}$.
   (c) $W_1 \subset W_2$ implies $W_2^\perp \subset W_1^\perp$.

2. Let U and W be subspaces of a finite-dimensional inner product space V. Prove the following:
   (a) $(U + W)^\perp = U^\perp \cap W^\perp$.
   (b) $(U \cap W)^\perp = U^\perp + W^\perp$.

3. Let $\{e_1, \dots, e_n\}$ be an orthonormal basis for an arbitrary inner product space V. If $u = \sum_i u_i e_i$ and $v = \sum_i v_i e_i$ are any vectors in V, show that

$$\langle u, v\rangle = \sum_{i=1}^n u_i^* v_i$$

   (this is just the generalization of Example 1.9).

4. Suppose $\{e_1, \dots, e_n\}$ is an orthonormal set in a vector space V, and x is any element of V. Show that the expression

$$\left\| x - \sum_{k=1}^n a_k e_k \right\|$$

   achieves its minimum value when each of the scalars a_k is equal to the Fourier coefficient $c_k = \langle e_k, x\rangle$. [*Hint*: Using Theorem 1.20 and the

Pythagorean theorem (see Exercise 1.5.3), add and subtract the term $\sum_{k=1}^{n} c_k e_k$ in the above expression to conclude that

$$\left\| x - \sum_{k=1}^{n} c_k e_k \right\|^2 \leq \left\| x - \sum_{k=1}^{n} a_k e_k \right\|^2$$

for any set of scalars a_k.]

5. Let $\{e_1, \ldots, e_n\}$ be an orthonormal set in an inner product space V, and let $c_k = \langle e_k, x\rangle$ be the Fourier coefficient of $x \in V$ with respect to e_k. Prove **Bessel's inequality**:

$$\sum_{k=1}^{n} |c_k|^2 \leq \|x\|^2$$

[*Hint*: Use the definition of the norm along with the obvious fact that $0 \leq \|x - \sum_{k=1}^{n} c_k e_k\|^2$.]

6. Find an orthonormal basis (relative to the standard inner product) for the following subspaces:
   (a) The subspace W of $\mathbb{C}^3$ spanned by the vectors $u_1 = (1, i, 0)$ and $u_2 = (1, 2, 1 - i)$.
   (b) The subspace W of $\mathbb{R}^4$ spanned by $u_1 = (1, 1, 0, 0)$, $u_2 = (0, 1, 1, 0)$ and $u_3 = (0, 0, 1, 1)$.

7. Consider the space $\mathbb{R}^3$ with the standard inner product.
   (a) Convert the vectors $u_1 = (1, 0, 1)$, $u_2 = (1, 0, -1)$ and $u_3 = (0, 3, 4)$ to an orthonormal basis $\{e_1, e_2, e_3\}$ of $\mathbb{R}^3$.
   (b) Write the components of an arbitrary vector $x = (x_1, x_2, x_3) \in \mathbb{R}^3$ in terms of the basis $\{e_i\}$.

8. Let V be the space of all polynomials of degree ≤ 3 defined on the interval $[-1, 1]$. Define an inner product on V by

$$\langle f, g\rangle = \int_{-1}^{1} f(t)g(t)\, dt.$$

Find an orthonormal basis for V generated by the functions $\{1, x, x^2, x^3\}$.

9. Let V and W be isomorphic inner product spaces under the vector space homomorphism $\phi : V \to W$, and assume that ϕ has the additional property that

$$\|\phi(x) - \phi(y)\| = \|x - y\|.$$

Such a ϕ is called an **isometry**, and V and W are said to be **isometric** spaces. (We also note that the norm on the left side of this equation is in W, while the norm on the right side is in V. We shall rarely distinguish between norms in different spaces unless there is some possible ambiguity.) Let V have orthonormal basis $\{v_1, \ldots, v_n\}$ so that any $x \in V$ may be written as $x = \sum x_i v_i$. Prove that the mapping $\phi : V \to \mathbb{R}^n$ defined by $\phi(x) = (x_1, \ldots, x_n)$ is an isometry of V onto $\mathbb{R}^n$ (with the standard inner product).

10. Let $\{e_1, e_2, e_3\}$ be an orthonormal basis for $\mathbb{R}^3$, and let $\{u_1, u_2, u_3\}$ be three mutually orthonormal vectors in $\mathbb{R}^3$. Let ${u_\lambda}^i$ denote the ith component of u_λ with respect to the basis $\{e_i\}$. Prove the **completeness relation**

$$\sum_{\lambda=1}^{3} {u_\lambda}^i {u_\lambda}^j = \delta_{ij}.$$

11. Let W be a finite-dimensional subspace of a possibly infinite-dimensional inner product space V. Prove that $V = W \oplus W^\perp$. [*Hint*: Let $\{w_1, \ldots, w_k\}$ be an orthonormal basis for W, and for any $x \in V$ define

$$x_1 = \sum_{i=1}^{k} \langle w_i, x\rangle w_i$$

and $x_2 = x - x_1$. Show that $x_1 + x_2 \in W + W^\perp$, and that $W \cap W^\perp = \{0\}$.]

Chapter 2

Linear Equations and Matrices

In this chapter we introduce matrices via the theory of simultaneous linear equations. This method has the advantage of leading in a natural way to the concept of the reduced row echelon form of a matrix. In addition, we will formulate some of the basic results dealing with the existence and uniqueness of systems of linear equations. In Chapter 4 we will arrive at the same matrix algebra from the viewpoint of linear transformations.

In order to introduce the idea of simultaneous linear equations, suppose we have two lines in the plane $\mathbb{R}^2$, and we ask whether or not they happen to intersect anywhere. To be specific, say the lines have the equations

$$\begin{aligned} x_2 &= -(1/2)x_1 + 5/2 \\ x_2 &= x_1 - 1/2. \end{aligned} \tag{2.1}$$

If these lines intersect, then there exists a point $(x_1, x_2) \in \mathbb{R}^2$ that satisfies both of these equations, and hence we would like to solve the pair of equations

$$\begin{aligned} x_1 + 2x_2 &= 5 \\ x_1 - x_2 &= 1/2. \end{aligned} \tag{2.2}$$

In this particular case, the easiest way to solve these is to use equation (2.1) directly and simply equate $-(1/2)x_1 + 5/2 = x_1 - 1/2$ to obtain $x_1 = 2$ and hence $x_2 = x_1 - 1/2 = 3/2$. But a more general approach is to use equation (2.2) as follows. Multiply the first of equations (2.2) by -1 and add to the second to obtain a new second equation $-3x_2 = -9/2$. This again yields $x_2 = 3/2$ and hence also $x_1 = 5 - 2x_2 = 2$.

We now turn our attention to generalizing this situation to more than two variables. This leads to systems of m linear equations in n unknowns.

2.1 Systems of Linear Equations

Let $a_1, \ldots, a_n, y$ be elements of a field $\mathcal{F}$, and let $x_1, \ldots, x_n$ be **unknowns** (also called **variables** or **indeterminates**). Then an equation of the form

$$a_1x_1 + \cdots + a_nx_n = y$$

is called a **linear equation in n unknowns** (over $\mathcal{F}$). The scalars a_i are called the **coefficients** of the unknowns, and y is called the **constant term** of the equation. A vector $(c_1, \ldots, c_n) \in \mathcal{F}^n$ is called a **solution vector** of this equation if and only if

$$a_1c_1 + \cdots + a_nc_n = y$$

in which case we say that $(c_1, \ldots, c_n)$ **satisfies** the equation. The set of all such solutions is called the **solution set** (or the **general solution**).

Now consider the following **system of m linear equations in n unknowns**:

$$\begin{aligned} a_{11}x_1 + \cdots + a_{1n}x_n &= y_1 \\ a_{21}x_1 + \cdots + a_{2n}x_n &= y_2 \\ &\vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n &= y_m \end{aligned}$$

We abbreviate this system by

$$\sum_{j=1}^{n} a_{ij}x_j = y_i, \qquad i = 1, \ldots, m.$$

If we let S_i denote the solution set of the equation $\sum_j a_{ij}x_j = y_i$ for each i, then the solution set S of the system is given by the intersection $S = \bigcap S_i$. In other words, if $(c_1, \ldots, c_n) \in \mathcal{F}^n$ is a solution of the system of equations, then it is a solution of each of the m equations in the system.

Example 2.1. Consider this system of two equations in three unknowns over the real field $\mathbb{R}$:

$$\begin{aligned} 2x_1 - 3x_2 + x_3 &= 6 \\ x_1 + 5x_2 - 2x_3 &= 12 \end{aligned}$$

The vector $(3, 1, 3) \in \mathbb{R}^3$ is not a solution of this system because

$$2(3) - 3(1) + 3 = 6$$

while

$$3 + 5(1) - 2(3) = 2 \neq 12.$$

However, the vector $(5, 1, -1) \in \mathbb{R}^3$ is a solution since

$$2(5) - 3(1) + (-1) = 6$$

and

$$5 + 5(1) - 2(-1) = 12.$$

Associated with a system of linear equations are two rectangular arrays of elements of $\mathcal{F}$ that turn out to be of great theoretical as well as practical significance. For the system $\sum_j a_{ij}x_j = y_i$, we define the **matrix of coefficients** A as the array

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

and the **augmented matrix** as the array $\operatorname{aug} A$ given by

$$\operatorname{aug} A = \left[\begin{array}{cccc|c} a_{11} & a_{12} & \cdots & a_{1n} & y_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & y_2 \\ \vdots & \vdots & & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & y_n \end{array}\right]$$

In general, we will use the term **matrix** to denote any array such as the array A shown above. This matrix has m **rows** and n **columns**, and hence is referred to as an $m \times n$ matrix, or a matrix of **size** $m \times n$. By convention, an element $a_{ij} \in \mathcal{F}$ of A is labeled with the first index referring to the row and the second index referring to the column. The scalar a_{ij} is usually called the i, jth **entry** (or **element**) of the matrix A. We will frequently denote the matrix A by the symbol (a_{ij}).

Before proceeding with the general theory, let us give a specific example demonstrating how to solve a system of linear equations.

Example 2.2. Let us attempt to solve the following system of linear equations:

$$\begin{aligned} 2x_1 + x_2 - 2x_3 &= -3 \\ x_1 - 3x_2 + x_3 &= 8 \\ 4x_1 - x_2 - 2x_3 &= 3 \end{aligned}$$

That our approach is valid in general will be proved in our first theorem below.

Multiply the first equation by $1/2$ to get the coefficient of x_1 equal to 1:

$$\begin{aligned} x_1 + (1/2)x_2 - x_3 &= -3/2 \\ x_1 - 3x_2 + x_3 &= 8 \\ 4x_1 - x_2 - 2x_3 &= 3 \end{aligned}$$

Multiply this first equation by -1 and add it to the second to obtain a new second equation, then multiply this first equation by -4 and add it to the third to obtain a new third equation:

$$\begin{aligned} x_1 + (1/2)x_2 - x_3 &= -3/2 \\ -(7/2)x_2 + 2x_3 &= 19/2 \\ -3x_2 + 2x_3 &= 9 \end{aligned}$$

Multiply this second by $-2/7$ to get the coefficient of x_2 equal to 1, then multiply this new second equation by 3 and add to the third:

$$\begin{aligned} x_1 + (1/2)x_2 - x_3 &= -3/2 \\ x_2 - (4/7)x_3 &= -19/7 \\ (2/7)x_3 &= 6/7 \end{aligned}$$

Multiply the third by $7/2$, then add $4/7$ times this new equation to the second:

$$\begin{aligned} x_1 + (1/2)x_2 - x_3 &= -3/2 \\ x_2 &= -1 \\ x_3 &= 3 \end{aligned}$$

Add the third equation to the first, then add $-1/2$ times the second equation to the new first to obtain

$$\begin{aligned} x_1 &= 2 \\ x_2 &= -1 \\ x_3 &= 3 \end{aligned}$$

This is now a solution of our system of equations. While this system could have been solved in a more direct manner, we wanted to illustrate the systematic approach that will be needed below.

Two systems of linear equations are said to be **equivalent** if they have equal solution sets. That each successive system of equations in Example 2.2 is indeed equivalent to the previous system is guaranteed by the following theorem. Note that this theorem is nothing more than a formalization of the above example.

Theorem 2.1. *The system of two equations in n unknowns over a field $\mathcal{F}$*

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \end{aligned} \tag{2.3}$$

with $a_{11} \neq 0$ is equivalent to the system

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a'_{22}x_2 + \cdots + a'_{2n}x_n &= b'_2 \end{aligned} \tag{2.4}$$

in which

$$a'_{2i} = a_{11}a_{2i} - a_{21}a_{1i}$$

for each $i = 1, \ldots, n$ and

$$b'_2 = a_{11}b_2 - a_{21}b_1$$

Proof. Let us define

$$L_i = \sum_{j=1}^{n} a_{ij} x_j$$

so that equations (2.3) may be written as the system

$$\begin{aligned} L_1 &= b_1 \\ L_2 &= b_2 \end{aligned} \tag{2.5}$$

while equations (2.4) are just

$$\begin{aligned} L_1 &= b_1 \\ -a_{21}L_1 + a_{11}L_2 &= -a_{21}b_1 + a_{11}b_2 \end{aligned} \tag{2.6}$$

If $(x_1, \ldots, x_n) \in \mathcal{F}^n$ is a solution of equations (2.5), then the two equations

$$\begin{aligned} a_{21}L_1 &= a_{21}b_1 \\ a_{11}L_2 &= a_{11}b_2 \end{aligned}$$

and hence also

$$-a_{21}L_1 + a_{11}L_2 = -a_{21}b_1 + a_{11}b_2$$

are all true equations. Therefore every solution of equations (2.5) also satisfies equations (2.6).

Conversely, suppose that we have a solution $(x_1, \ldots, x_n)$ to the system (2.6). Then clearly

$$a_{21}L_1 = a_{21}b_1$$

is a true equation. Hence, adding this to the second of equations (2.6) gives us

$$a_{21}L_1 + (-a_{21}L_1 + a_{11}L_2) = a_{21}b_1 + (-a_{21}b_1 + a_{11}b_2)$$

or

$$a_{11}L_2 = a_{11}b_2.$$

Thus $L_2 = b_2$ is also a true equation. This shows that any solution of equations (2.6) is a solution of equations (2.5) also. ∎

It should now be reasonably clear why we defined the matrix $\operatorname{aug} A$ — we want to perform the above operations on $\operatorname{aug} A$ to end up (if possible) with a matrix of the form

$$\left[\begin{array}{cccc|c} x_1 & 0 & \cdots & 0 & c_1 \\ 0 & x_2 & \cdots & 0 & c_2 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & x_n & c_n \end{array}\right].$$

From here we see that the solution to our system is simply $x_i = c_i$.

We point out that in the proof of Theorem 2.1 (as well as in Example 2.2), it was only the coefficients themselves that were of any direct use to us. The unknowns x_i were never actually used in any of the manipulations. This is the reason that we defined the matrix of coefficients (a_{ij}). What we now proceed to do is to generalize the above method of solving systems of equations in a manner that utilizes this matrix explicitly.

Exercises

1. For each of the following systems of equations, find a solution if it exists:

(a)
$$\begin{aligned} x + 2y - 3z &= -1 \\ 3x - y + 2z &= 7 \\ 5x + 3y - 4z &= 2 \end{aligned}$$

(b)
$$\begin{aligned} 2x + y - 2z &= 10 \\ 3x + 2y + 2z &= 1 \\ 5x + 4y + 3z &= 4 \end{aligned}$$

(c)
$$\begin{aligned} x + 2y - 3z &= 6 \\ 2x - y + 4z &= 2 \\ 4x + 3y - 2z &= 14 \end{aligned}$$

2. Determine whether or not the each of the following two systems is equivalent (over $\mathbb{C}$):

(a)
$$\begin{aligned} x - y &= 0 \\ 2x + y &= 0 \end{aligned} \qquad \text{and} \qquad \begin{aligned} 3x + y &= 0 \\ x + y &= 0 \end{aligned}$$

(b)
$$\begin{aligned} -x + y + 4z &= 0 \\ x + 3y + 8z &= 0 \\ (1/2)x + y + (5/2)z &= 0 \end{aligned} \qquad \text{and} \qquad \begin{aligned} x - z &= 0 \\ y + 3z &= 0 \end{aligned}$$

(c)
$$\begin{aligned} 2x + (-1+i)y + t &= 0 \\ 3y - 2iz + 5t &= 0 \end{aligned}$$

and

$$\begin{aligned} (1 + i/2)x + 8y - iz - t &= 0 \\ (2/3)x - (1/2)y + z + 7t &= 0 \end{aligned}$$

2.2 Elementary Row Operations

The important point to realize in Example 2.2 is that we solved a system of linear equations by performing some combination of the following operations:

(a) Change the order in which the equations are written.
(b) Multiply each term in an equation by a nonzero scalar.
(c) Multiply one equation by a nonzero scalar and then add this new equation to another equation in the system.

Note that (a) was not used in Example 2.2, but it would have been necessary if the coefficient of x_1 in the first equation had been 0. The reason for this is that we want the equations put into echelon form as defined below.

We now see how to use the matrix $\operatorname{aug} A$ as a tool in solving a system of linear equations. In particular, we define the following so-called **elementary row operations** (or **transformations**) as applied to the augmented matrix:

(α) Interchange two rows.
(β) Multiply one row by a nonzero scalar.
(γ) Add a scalar multiple of one row to another.

It should be clear that operations (α) and (β) have no effect on the solution set of the system and, in view of Theorem 2.1, that operation (γ) also has no effect.

The next two examples show what happens both in the case where there is no solution to a system of linear equations, and in the case of an infinite number of solutions. In performing these operations on a matrix, we will let R_i denote the ith row. We leave it to the reader to repeat Example 2.2 using this notation.

Example 2.3. Consider this system of linear equations over the field $\mathbb{R}$:

$$\begin{aligned} x + 3y + 2z &= 7 \\ 2x + \ \ y - \ \ z &= 5 \\ -x + 2y + 3z &= 4 \end{aligned}$$

The augmented matrix is

$$\left[\begin{array}{rrr|r} 1 & 3 & 2 & 7 \\ 2 & 1 & -1 & 5 \\ -1 & 2 & 3 & 4 \end{array}\right]$$

and the reduction proceeds as follows.

We first perform the following elementary row operations:

$$\begin{array}{r} \\ R_2 - 2R_1 \to \\ R_3 + \ \ R_1 \to \end{array} \left[\begin{array}{rrr|r} 1 & 3 & 2 & 7 \\ 0 & -5 & -5 & -9 \\ 0 & 5 & 5 & 11 \end{array}\right]$$

Now, using this matrix, we obtain

$$\begin{array}{r} \\ -\ R_2 \to \\ R_3 + R_2 \to \end{array} \left[\begin{array}{rrr|r} 1 & 3 & 2 & 7 \\ 0 & 5 & 5 & 9 \\ 0 & 0 & 0 & 2 \end{array}\right]$$

It is clear that the equation $0z = 2$ has no solution, and hence this system has no solution.

Example 2.4. Let us solve the following system over the field $\mathbb{R}$:

$$\begin{aligned} x_1 - 2x_2 + 2x_3 - \ \ x_4 &= -14 \\ 3x_1 + 2x_2 - \ \ x_3 + 2x_4 &= \ \ 17 \\ 2x_1 + 3x_2 - \ \ x_3 - \ \ x_4 &= \ \ 18 \\ -2x_1 + 5x_2 - 3x_3 - 3x_4 &= \ \ 26 \end{aligned}$$

We have the matrix aug A given by

$$\left[\begin{array}{rrrr|r} 1 & -2 & 2 & -1 & -14 \\ 3 & 2 & -1 & 2 & 17 \\ 2 & 3 & -1 & -1 & 18 \\ -2 & 5 & -3 & -3 & 26 \end{array}\right]$$

and hence we obtain the sequence

$$\begin{array}{r} \\ R_2 - 3R_1 \to \\ R_3 - 2R_1 \to \\ R_4 + 2R_1 \to \end{array} \left[\begin{array}{rrrr|r} 1 & -2 & 2 & -1 & -14 \\ 0 & 8 & -7 & 5 & 59 \\ 0 & 7 & -5 & 1 & 46 \\ 0 & 1 & 1 & -5 & -2 \end{array}\right]$$

$$\begin{array}{r} \\ R_4 \to \\ R_2 - 8R_4 \to \\ R_3 - 7R_4 \to \end{array} \left[\begin{array}{rrrr|r} 1 & -2 & 2 & -1 & -14 \\ 0 & 1 & 1 & -5 & -2 \\ 0 & 0 & -15 & 45 & 75 \\ 0 & 0 & -12 & 36 & 60 \end{array}\right]$$

$$\begin{array}{r} \\ \\ (-1/15)R_3 \to \\ (-1/12)R_4 \to \end{array} \left[\begin{array}{rrrr|r} 1 & -2 & 2 & -1 & -14 \\ 0 & 1 & 1 & -5 & -2 \\ 0 & 0 & 1 & -3 & -5 \\ 0 & 0 & 1 & -3 & -5 \end{array}\right]$$

We see that the third and fourth equations are identical, and hence we have three equations in four unknowns:

$$\begin{aligned} x_1 - 2x_2 + 2x_3 - x_4 &= -14 \\ x_2 + x_3 - 5x_4 &= -2 \\ x_3 - 3x_4 &= -5 \end{aligned}$$

It is now apparent that there are an infinite number of solutions because, if we let $c \in \mathbb{R}$ be any real number, then our solution set is given by $x_4 = c$, $x_3 = 3c - 5$, $x_2 = 2c + 3$ and $x_1 = -c + 2$.

Two $m \times n$ matrices are said to be **row equivalent** if one can be transformed into the other by a finite number of elementary row operations. As we stated just prior to Example 2.3, these elementary row operations have no effect on the solution set of a system of linear equations. The formal statement of this is contained in our next theorem.

Theorem 2.2. *Let A and B be the augmented matrices of two systems of m linear equations in n unknowns. If A is row equivalent to B, then both systems have the same solution set.*

Proof. If A is row equivalent to B, then we can go from the system represented by A to the system represented by B by a succession of the operations (a), (b) and (c) described above. It is clear that operations (a) and (b) have no effect on the solutions, and the method of Theorem 2.1 shows that operation (c) also has no effect. ∎

In order to describe the desired form of the augmented matrix after performing the elementary row operations, we need some additional terminology.

A matrix is said to be in **row echelon form** if successive rows of the matrix start out (from the left) with more and more zeros. In particular, a matrix is said to be in **reduced row echelon form** if it has the following properties (which are more difficult to state precisely than they are to understand):

(a) All zero rows (if any) occur below all nonzero rows.
(b) The first nonzero entry (reading from the left) in each row is equal to 1.
(c) If the first nonzero entry in the ith row is in the j_ith column, then every other entry in the j_ith column is 0.
(d) If the first nonzero entry in the ith row is in the j_ith column, then $j_1 < j_2 < \cdots$.

Loosely put, the reduced row echelon form has more and more zeros as you go down the rows, the first element of each nonzero row is a 1, and every other element above and below that first 1 is a zero.

We will call the first (or **leading**) nonzero entries in each row of a row echelon matrix the **distinguished elements** of the matrix. (The leading entry of a row that is added to another row is also frequently referred to as a **pivot**.) Thus, a matrix is in reduced row echelon form if the distinguished elements are each equal to 1, and they are the only nonzero entries in their respective columns.

Example 2.5. The matrix

$$\begin{bmatrix} 1 & 2 & -3 & 0 & 1 \\ 0 & 0 & 5 & 2 & -4 \\ 0 & 0 & 0 & 7 & 3 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

is in row echelon form but not in reduced row echelon form. However, the matrix

$$\begin{bmatrix} 1 & 0 & 5 & 0 & 2 \\ 0 & 1 & 2 & 0 & 4 \\ 0 & 0 & 0 & 1 & 7 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

is in reduced row echelon form. Note that the distinguished elements of the first matrix are the numbers 1, 5 and 7, and the distinguished elements of the second matrix are the numbers 1, 1 and 1.

It should be clear from Example 2.4 that every matrix can be put into reduced row echelon form. Our next theorem proves this in detail by outlining an algorithm generally known as **Gaussian elimination**. (Sometimes this refers to reducing to row echelon form, and **Gauss-Jordan elimination** refers to reducing all the way to reduced row echelon form.)

Theorem 2.3. *Every $m \times n$ matrix A is row equivalent to a reduced row echelon matrix.*

Proof. Suppose that we first put A into the form where the leading entry in each nonzero row is equal to 1, and where every other entry in the column containing this first nonzero entry is equal to 0. (This is called simply the **row-reduced** form of A.) If this can be done, then all that remains is to perform a finite number of row interchanges to achieve the final desired reduced row echelon form.

To obtain the row-reduced form we proceed as follows. First consider row 1. If every entry in row 1 is equal to 0, then we do nothing with this row. If row 1 is nonzero, then let j_1 be the smallest positive integer for which $a_{1j_1} \neq 0$ and multiply row 1 by $(a_{1j_1})^{-1}$. Next, for each $i \neq 1$ we add $-a_{ij_1}$ times row 1 to row i. This leaves us with the leading entry a_{1j_1} of row 1 equal to 1, and every other entry in the j_1th column equal to 0.

Now consider row 2 of the matrix we are left with. Again, if row 2 is equal to 0 there is nothing to do. If row 2 is nonzero, assume that the first nonzero entry occurs in column j_2 (where $j_2 \neq j_1$ by the results of the previous paragraph). Multiply row 2 by $(a_{2j_2})^{-1}$ so that the leading entry in row 2 is equal to 1, and then add $-a_{ij_2}$ times row 2 to row i for each $i \neq 2$. Note that these operations have no effect on either column j_1, or on columns $1, \dots, j_1$ of row 1.

It should now be clear that we can continue this process a finite number of times to achieve the final row-reduced form. We leave it to the reader to take an arbitrary matrix (a_{ij}) and apply successive elementary row transformations to achieve the desired final form. ∎

For example, I leave it to you to show that the reduced row echelon form of the matrix in Example 2.4 is

$$\begin{bmatrix} 1 & 0 & 0 & 1 & 2 \\ 0 & 1 & 0 & -2 & 3 \\ 0 & 0 & 1 & -3 & -5 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

While we have shown that every matrix is row equivalent to at least one reduced row echelon matrix, it is not obvious that this equivalence is unique. However, we shall show in the next section that this reduced row echelon matrix is in fact unique. Because of this, the reduced row echelon form of a matrix is often called the **row canonical form**.

Exercises

1. Show that row equivalence defines an equivalence relation on the set of all matrices.

2. For each of the following matrices, first reduce to row echelon form, and then to row canonical form:

(a) $\begin{bmatrix} 1 & -2 & 3 & -1 \\ 2 & -1 & 2 & 2 \\ 3 & 1 & 2 & 3 \end{bmatrix}$ (b) $\begin{bmatrix} 1 & 2 & -1 & 2 & 1 \\ 2 & 4 & 1 & -2 & 3 \\ 3 & 6 & 2 & -6 & 5 \end{bmatrix}$

(c) $\begin{bmatrix} 1 & 3 & -1 & 2 \\ 0 & 1 & -5 & 3 \\ 2 & -5 & 3 & 1 \\ 4 & 1 & 1 & 5 \end{bmatrix}$

3. For each of the following systems, find a solution or show that no solution exists:

(a) $\begin{aligned} x + y + z &= 1 \\ 2x - 3y + 7z &= 0 \\ 3x - 2y + 8z &= 4 \end{aligned}$ (b) $\begin{aligned} x - y + 2z &= 1 \\ x + y + z &= 2 \\ 2x - y + z &= 5 \end{aligned}$

(c) $\begin{aligned} x - y + 2z &= 4 \\ 3x + y + 4z &= 6 \\ x + y + z &= 1 \end{aligned}$ (d) $\begin{aligned} x + 3y + z &= 2 \\ 2x + 7y + 4z &= 6 \\ x + y - 4z &= 1 \end{aligned}$

(e) $\begin{aligned} x + 3y + z &= 0 \\ 2x + 7y + 4z &= 0 \\ x + y - 4z &= 0 \end{aligned}$ (f) $\begin{aligned} 2x - y + 5z &= 19 \\ x + 5y - 3z &= 4 \\ 3x + 2y + 4z &= 5 \end{aligned}$

(g) $\begin{aligned} 2x - y + 5z &= 19 \\ x + 5y - 3z &= 4 \\ 3x + 2y + 4z &= 25 \end{aligned}$

4. Let f_1, f_2 and f_3 be elements of $F[\mathbb{R}]$ (i.e., the space of all real-valued functions defined on $\mathbb{R}$).
   (a) Given a set $\{x_1, x_2, x_3\}$ of real numbers, define the 3×3 matrix $F(x) = (f_i(x_j))$ where the rows are labeled by i and the columns are labeled by j. Prove that the set $\{f_i\}$ is linearly independent if the rows of the matrix $F(x)$ are linearly independent.
   (b) Now assume that each f_i has first and second derivatives defined on some interval $(a, b) \subset \mathbb{R}$, and let $f_i^{(j)}$ denote the jth derivative of f_i (where $f_i^{(0)}$ is just f_i). Define the matrix $W(x) = (f_i^{(j-1)}(x))$ where $1 \le i, j \le 3$. Prove that $\{f_i\}$ is linearly independent if the rows of $W(x)$ are independent for some $x \in (a, b)$. (The determinant of $W(x)$

is called the **Wronskian** of the set of functions $\{f_i\}$.)

Show that each of the following sets of functions is linearly independent:

(c) $f_1(x) = -x^2 + x + 1,\ f_2(x) = x^2 + 2x,\ f_3(x) = x^2 - 1.$
(d) $f_1(x) = \exp(-x),\ f_2(x) = x,\ f_3(x) = \exp(2x).$
(e) $f_1(x) = \exp(x),\ f_2(x) = \sin x,\ f_3(x) = \cos x.$

5. Let
$$A = \begin{bmatrix} 3 & -1 & 2 \\ 2 & 1 & 1 \\ 1 & -3 & 0 \end{bmatrix}.$$
Determine the values of $Y = (y_1, y_2, y_3)$ for which the system $\sum_i a_{ij}x_j = y_i$ has a solution.

6. Repeat the previous problem with the matrix
$$A = \begin{bmatrix} 3 & -6 & 2 & -1 \\ -2 & 4 & 1 & 3 \\ 0 & 0 & 1 & 1 \\ 1 & -2 & 1 & 0 \end{bmatrix}$$

2.3 Row and Column Spaces

We now forget about systems of equations, and instead focus our attention directly on the matrices themselves. This will be absolutely essential in discussing the properties of linear transformations.

First of all, it will be extremely useful to consider the rows and columns of an arbitrary $m \times n$ matrix as vectors in their own right. In particular, the rows of A are to be viewed as vector n-tuples $A_1, \dots, A_m$ where each $A_i = (a_{i1}, \dots, a_{in}) \in \mathcal{F}^n$. Similarly, the columns of A are to be viewed as vector m-tuples $A^1, \dots, A^n$ where each $A^j = (a_{1j}, \dots, a_{mj}) \in \mathcal{F}^m$. As we mentioned earlier, for notational clarity we should write A^j as the column vector
$$\begin{bmatrix} a_{1j} \\ \vdots \\ a_{mj} \end{bmatrix}$$
but it is typographically easier to write this horizontally whenever possible. *Note that we label the row vectors of A by subscripts, and the columns of A by superscripts.*

Since each row A_i is an element of $\mathcal{F}^n$, the set of all rows of a matrix can be used to generate a new vector space V over $\mathcal{F}$. In other words, V is the space spanned by the rows A_i, and hence any $v \in V$ may be written as
$$v = \sum_{i=1}^{m} c_i A_i$$

where each $c_i \in \mathcal{F}$. The space V (which is apparently a subspace of $\mathcal{F}^n$) is called the **row space** of A. The dimension of V is called the **row rank** of A, and will be denoted by $\operatorname{rr}(A)$. Since V is a subspace of $\mathcal{F}^n$ and $\dim \mathcal{F}^n = n$, it follows that $\operatorname{rr}(A) = \dim V \leq n$. On the other hand, V is spanned by the m vectors A_i, so that we must have $\dim V \leq m$. It then follows that $\operatorname{rr}(A) \leq \min\{m, n\}$.

In an exactly analogous manner, we define the **column space** W of a matrix A as that subspace of $\mathcal{F}^m$ spanned by the n column vectors A^j. Thus any $w \in W$ is given by

$$w = \sum_{j=1}^{n} b_j A^j$$

The **column rank** of A, denoted by $\operatorname{cr}(A)$, is given by $\operatorname{cr}(A) = \dim W$ and, as above, we must have $\operatorname{cr}(A) \leq \min\{m, n\}$.

We will sometimes denote the row space of A by $\operatorname{row}(A)$ and the column space by $\operatorname{col}(A)$.

An obvious question is whether a sequence of elementary row operations changes either the row space or the column space of a matrix. What we will show is that the row space itself remains unchanged, and the column space at least maintains its dimension. In other words, both the row rank and column rank remain unchanged by a sequence of elementary row operations. We will then show that in fact the row and column ranks are the same, and therefore we are justified in defining the rank of a matrix as either the row or column rank. Let us now verify these statements.

Under elementary row operations, it should be clear that the row space won't change because all we are doing is taking different linear combinations of the same vectors. In other words, the elementary row operations simply result in a new basis for the row space. In somewhat more formal terms, suppose A is row-equivalent to $\tilde{A}$. Then the rows of $\tilde{A}$ are linear combinations of the rows of A, and therefore the row space of $\tilde{A}$ is a subspace of the row space of A. On the other hand, we can reverse the order of row operations so that $\tilde{A}$ is row equivalent to A. Then the rows of A are linear combinations of the rows of $\tilde{A}$ so that the row space of A is a subspace of the row space of $\tilde{A}$. Therefore the row spaces are the same for A and $\tilde{A}$ so that $\operatorname{rr}(A) = \operatorname{rr}(\tilde{A})$.

However, what happens to the column space is not so obvious. The elementary row transformations interchange and mix up the components of the column vectors, so the column spaces are clearly not the same in A and $\tilde{A}$. But the interesting point, and what makes all of this so useful, is that the dimension of the column space hasn't changed. In other words, we still have $\operatorname{cr}(A) = \operatorname{cr}(\tilde{A})$.

Probably the easiest way to see this is to consider those columns of A that are linearly *dependent*; and with no loss of generality we can call them $A^1, \ldots, A^r$. Then their linear dependence means there are nonzero scalars $x_1, \ldots, x_r$ such that $\sum_{i=1}^{r} A^i x_i = 0$. In full form this is

$$\begin{bmatrix} a_{11} \\ \vdots \\ a_{m1} \end{bmatrix} x_1 + \cdots + \begin{bmatrix} a_{1r} \\ \vdots \\ a_{mr} \end{bmatrix} x_r = 0.$$

But this is a system of m linear equations in r unknowns, and we have seen that the solution set doesn't change under row equivalence. In other words, $\sum_{i=1}^{r} \tilde{A}^i x_i = 0$ for the *same coefficients* x_i. Then the same r columns of $\tilde{A}$ are linearly *dependent*, and hence both A and $\tilde{A}$ have the same $(n-r)$ *independent* columns, i.e., $\operatorname{cr}(A) = \operatorname{cr}(\tilde{A})$. (There can't be more dependent columns of $\tilde{A}$ than A because we can apply the row operations in reverse to go from $\tilde{A}$ to A. If $\tilde{A}$ had more dependent columns, then when we got back to A we would have more than we started with.)

Let us summarize what we have just said as a theorem for ease of reference.

Theorem 2.4. *Let A and $\tilde{A}$ be row equivalent $m \times n$ matrices. Then the row space of A is equal to the row space of $\tilde{A}$, and hence $\operatorname{rr}(A) = \operatorname{rr}(\tilde{A})$. Furthermore, we also have $\operatorname{cr}(A) = \operatorname{cr}(\tilde{A})$. (However, note that the column space of A is not necessarily the same as the column space of $\tilde{A}$.)*

Now look back at the reduced row echelon form of a matrix A (as in Example 2.5). The number of nonzero rows of A is just $\operatorname{rr}(A)$, and all of these rows begin with a 1 (the distinguished elements). But all other entries in each column containing these distinguished elements are 0, and the remaining columns are linear combinations of these. In other words, the number of linearly independent *columns* in the reduced row echelon form of A is the same as the row rank of A.

This discussion proves the following very important result.

Theorem 2.5. *If $A = (a_{ij})$ is any $m \times n$ matrix over a field $\mathcal{F}$, then $\operatorname{rr}(A) = \operatorname{cr}(A)$.*

In view of this theorem, we define the **rank** of a matrix A as the number $\operatorname{rank}(A)$ given by

$$\operatorname{rank}(A) = \operatorname{rr}(A) = \operatorname{cr}(A).$$

The concept of rank is extremely important in many branches of mathematics (and hence physics and engineering). For example, the inverse and implicit function theorems, surface theory in differential geometry, and the theory of linear transformations (which we will cover in a later chapter) all depend on rank in a fundamental manner.

Combining Theorem 2.5 with the discussion just prior to it, we have the basis for a practical method of finding the rank of a matrix.

Theorem 2.6. *If A is any matrix, then $\operatorname{rank}(A)$ is equal to the number of nonzero rows in the (reduced) row echelon matrix row equivalent to A. (Alternatively, $\operatorname{rank}(A)$ is the number of nonzero columns in the (reduced) column-echelon matrix column equivalent to A.)*

There is one special case that is worth pointing out. By way of terminology, if A is an $n \times n$ matrix such that $a_{ij} = 0$ for $i \neq j$ and $a_{ii} = 1$, then we say that A is the **identity matrix** of **size** n, and write this matrix as I_n. Since the size is usually understood, we will generally simply write I. If $I = (I_{ij})$, then another useful way of writing this is in terms of the Kronecker delta as $I_{ij} = \delta_{ij}$. Written out, I has the form

$$I = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$

If A is an $n \times n$ matrix and $\operatorname{rank}(A) = n$, then the reduced row echelon form of A is just the identity matrix I_n, and we have the next result.

Theorem 2.7. *If A is an $n \times n$ matrix of rank n, then the reduced row echelon matrix row equivalent to A is the identity matrix I_n.*

An $n \times n$ matrix of rank n is said to be **nonsingular**, and if $\operatorname{rank}(A) < n$, then A is said to be **singular**. As we will see shortly, if a matrix is nonsingular, we will be able to define an "inverse." But to do so, we first have to define matrix multiplication. We will return to this after the next section.

Example 2.6. Let us find the rank of the matrix A given by

$$A = \begin{bmatrix} 1 & 2 & -3 \\ 2 & 1 & 0 \\ -2 & -1 & 3 \\ -1 & 4 & -2 \end{bmatrix}.$$

To do this, we will apply Theorem 2.6 to columns instead of rows (just for variety). Proceeding with the elementary transformations, we obtain the following sequence of matrices:

$$\begin{array}{ccc} & & \\ \multicolumn{3}{c}{\begin{bmatrix} 1 & 0 & 0 \\ 2 & -3 & 6 \\ -2 & 3 & -3 \\ -1 & 6 & -5 \end{bmatrix}} \\ & \uparrow & \uparrow \\ & A^2 - 2A^1 & A^3 + 3A^1 \end{array}$$

$$\begin{array}{ccc} & & \\ \multicolumn{3}{c}{\begin{bmatrix} 1 & 0 & 0 \\ 2 & -1 & 0 \\ -2 & 1 & 1 \\ -1 & 2 & 7/3 \end{bmatrix}} \\ & \uparrow & \uparrow \\ & (1/3)A^2 & (1/3)(A^3 + 2A^2) \end{array}$$

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 3 & 1/3 & 7/3 \end{bmatrix}$$
$$\begin{array}{cc} \uparrow & \uparrow \\ A^1 + 2A^2 & -(A^2 - A^3) \end{array}$$

Thus the reduced column-echelon form of A has three nonzero columns, so that $\operatorname{rank}(A) = \operatorname{cr}(A) = 3$. We leave it to the reader (see Exercise 2.3.1) to show that the row canonical form of A is

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

and hence $\operatorname{rank}(A) = \operatorname{cr}(A) = \operatorname{rr}(A)$ as it should.

Exercises

1. Verify the row-canonical form of the matrix in Example 2.6.

2. Let A and B be arbitrary $m \times n$ matrices. Show that $\operatorname{rank}(A + B) \leq \operatorname{rank}(A) + \operatorname{rank}(B)$.

3. Using elementary row operations, find the rank of each of the following matrices:

(a) $\begin{bmatrix} 1 & 3 & 1 & -2 & -3 \\ 1 & 4 & 3 & -1 & -4 \\ 2 & 3 & -4 & -7 & -3 \\ 3 & 8 & 1 & -7 & -8 \end{bmatrix}$ (b) $\begin{bmatrix} 1 & 2 & -3 \\ 2 & 1 & 0 \\ -2 & -1 & 3 \\ -1 & 4 & -2 \end{bmatrix}$

(c) $\begin{bmatrix} 1 & 3 \\ 0 & -2 \\ 5 & -1 \\ -2 & 3 \end{bmatrix}$ (d) $\begin{bmatrix} 5 & -1 & 1 \\ 2 & 1 & -2 \\ 0 & -7 & 12 \end{bmatrix}$

4. Repeat the previous problem using elementary column operations.

2.4 Solutions to Systems of Linear Equations

We now apply the results of the previous section to the determination of some general characteristics of the solution set to systems of linear equations. We will have more to say on this subject after we have discussed determinants in the next chapter.

To begin with, a system of linear equations of the form

$$\sum_{j=1}^{n} a_{ij}x_j = 0, \qquad i = 1, \dots, m$$

is called a **homogeneous system** of m linear equations in n unknowns. It is obvious that choosing $x_1 = x_2 = \cdots = x_n = 0$ will satisfy this system, but this is not a very interesting solution. It is called the **trivial** (or **zero**) **solution**. Any other solution, if it exists, is referred to as a **nontrivial solution**.

A more general type of system of linear equations is of the form

$$\sum_{j=1}^{n} a_{ij}x_j = y_i, \qquad i = 1, \dots, m$$

where each y_i is a given scalar. This is then called an **inhomogeneous system** (or sometimes a **nonhomogeneous** system) of linear equations. Let us define the *column* vector

$$Y = (y_1, \dots, y_m) \in \mathcal{F}^m$$

and also note that $a_{ij}x_j$ is just x_j times the ith component of the jth column $A^j \in \mathcal{F}^m$. Thus our system of inhomogeneous equations may be written in the form

$$\sum_{j=1}^{n} A^j x_j = \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{bmatrix} x_1 + \begin{bmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{bmatrix} x_2 + \cdots + \begin{bmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{bmatrix} x_n = Y$$

where this vector equation is to be interpreted in terms of its components. (In the next section, we shall see how to write this as a product of matrices.) It should also be obvious that a homogeneous system may be written in this notation as

$$\sum_{j=1}^{n} A^j x_j = 0.$$

Let us now look at some elementary properties of the solution set of a homogeneous system of equations.

Theorem 2.8. *The solution set S of a homogeneous system of m equations in n unknowns is a subspace of $\mathcal{F}^n$.*

Proof. Let us write our system as $\sum_j a_{ij}x_j = 0$. We first note that $S \neq \varnothing$ since $(0, \dots, 0) \in \mathcal{F}^n$ is the trivial solution of our system. If $u = (u_1, \dots, u_n) \in \mathcal{F}^n$ and $v = (v_1, \dots, v_n) \in \mathcal{F}^n$ are both solutions, then

$$\sum_j a_{ij}(u_j + v_j) = \sum_j a_{ij}u_j + \sum_j a_{ij}v_j = 0$$

so that $u + v \in S$. Finally, if $c \in \mathcal{F}$ then we also have

$$\sum_j a_{ij}(cu_j) = c \sum_j a_{ij} u_j = 0$$

so that $cu \in S$. ∎

If we look back at Example 2.4, we see that a system of m equations in $n > m$ unknowns will necessarily result in a nonunique, and hence nontrivial, solution. The formal statement of this fact is contained in our next theorem.

Theorem 2.9. *Let a homogeneous system of m equations in n unknowns have the $m \times n$ matrix of coefficients A. Then the system has a nontrivial solution if and only if* $\operatorname{rank}(A) < n$.

Proof. By writing the system in the form $\sum_j x_j A^j = 0$, it is clear that a nontrivial solution exists if and only if the n column vectors $A^j \in \mathcal{F}^m$ are linearly dependent. Since the rank of A is equal to the dimension of its column space, we must therefore have $\operatorname{rank}(A) < n$. ∎

It should now be clear that if an $n \times n$ (i.e., square) matrix of coefficients A (of a homogeneous system) has rank equal to n, then the only solution will be the trivial solution since reducing the augmented matrix (which then has the last column equal to the zero vector) to reduced row echelon form will result in each variable being set equal to zero (see Theorem 2.7).

Theorem 2.10. *Let a homogeneous system of linear equations in n unknowns have a matrix of coefficients A. Then the solution set S of this system is a subspace of $\mathcal{F}^n$ with dimension*

$$\dim S = n - \operatorname{rank}(A).$$

Proof. Assume that S is a nontrivial solution set, so that by Theorem 2.9 we have $\operatorname{rank}(A) < n$. Assume also that the unknowns $x_1, \dots, x_n$ have been ordered in such a way that the first $k = \operatorname{rank}(A)$ columns of A span the column space (this is guaranteed by Theorem 2.4). Then the remaining columns $A^{k+1}, \dots, A^n$ may be written as

$$A^i = \sum_{j=1}^{k} b_{ij} A^j, \qquad i = k+1, \dots, n$$

where each $b_{ij} \in \mathcal{F}$. If we define $b_{ii} = -1$ and $b_{ij} = 0$ for $j \neq i$ and $j > k$, then we may write this as

$$\sum_{j=1}^{n} b_{ij} A^j = 0, \qquad i = k+1, \dots, n$$

(note the upper limit on this sum differs from the previous equation). Next we observe that the solution set S consists of all vectors $x \in \mathcal{F}^n$ such that

$$\sum_{j=1}^{n} x_j A^j = 0$$

and hence in particular, the $n-k$ vectors

$$b^{(i)} = (b_{i1}, \ldots, b_{in})$$

for each $i = k+1, \ldots, n$ must belong to S. We show that they in fact form a basis for S, which is then of dimension $n-k$.

To see this, we first write out each of the $b^{(i)}$:

$$\begin{aligned} b^{(k+1)} &= (b_{k+1,1}, \ldots, b_{k+1,k}, -1, 0, 0, \ldots, 0) \\ b^{(k+2)} &= (b_{k+2,1}, \ldots, b_{k+2,k}, 0, -1, 0, \ldots, 0) \\ &\ \vdots \\ b^{(n)} &= (b_{n1}, \ldots, b_{nk}, 0, 0, \ldots, 0, -1). \end{aligned}$$

Hence for any set $\{c_i\}$ of $n-k$ scalars we have

$$\sum_{i=k+1}^{n} c_i b^{(i)} = \left(\sum_{i=k+1}^{n} c_i b_{i1}, \ldots, \sum_{i=k+1}^{n} c_i b_{ik}, -c_{k+1}, \ldots, -c_n \right)$$

and therefore

$$\sum_{i=k+1}^{n} c_i b^{(i)} = 0$$

if and only if $c_{k+1} = \cdots = c_n = 0$. This shows that the $b^{(i)}$ are linearly independent. (This should have been obvious from their form shown above.)

Now suppose that $d = (d_1, \ldots, d_n)$ is any solution of

$$\sum_{j=1}^{n} x_j A^j = 0.$$

Since S is a vector space (Theorem 2.8), any linear combination of solutions is a solution, and hence the vector

$$y = d + \sum_{i=k+1}^{n} d_i b^{(i)}$$

must also be a solution. In particular, writing out each component of this expression shows that

$$y_j = d_j + \sum_{i=k+1}^{n} d_i b_{ij}$$

and hence the definition of the b_{ij} shows that $y = (y_1, \dots, y_k, 0, \dots, 0)$ for some set of scalars y_i. Therefore, we have

$$0 = \sum_{j=1}^{n} y_j A^j = \sum_{j=1}^{k} y_j A^j$$

and since $\{A^1, \dots, A^k\}$ is linearly independent, this implies that $y_j = 0$ for each $j = 1, \dots, k$. Hence $y = 0$ so that

$$d = - \sum_{i=k+1}^{n} d_i b^{(i)}$$

and we see that any solution may be expressed as a linear combination of the $b^{(i)}$.

Since the $b^{(i)}$ are linearly independent and we just showed that they span S, they must form a basis for S. ∎

Suppose that we have a homogeneous system of m equations in $n > m$ unknowns, and suppose that the coefficient matrix A is in row echelon form and has rank m. Then each of the m successive equations contains fewer and fewer unknowns, and since there are more unknowns than equations, there will be $n - m = n - \operatorname{rank}(A)$ unknowns that do not appear as the first entry in any of the rows of A. These $n - \operatorname{rank}(A)$ unknowns are called **free variables**. We may arbitrarily assign any value we please to the free variables to obtain a solution of the system.

Let the free variables of our system be $x_{i_1}, \dots, x_{i_k}$ where $k = n - m = n - \operatorname{rank}(A)$, and let v_s be the solution vector obtained by setting x_{i_s} equal to 1 and each of the remaining free variables equal to 0. (This is essentially what was done in the proof of Theorem 2.10.) We claim that $v_1, \dots, v_k$ are linearly independent and hence form a basis for the solution space of the (homogeneous) system (which is of dimension $n - \operatorname{rank}(A)$ by Theorem 2.10).

To see this, we basically follow the proof of Theorem 2.10 and let B be the $k \times n$ matrix whose rows consist of the solution vectors v_s. For each s, our construction is such that we have $x_{i_s} = 1$ and $x_{i_r} = 0$ for $r \neq s$ (and the remaining $m = n - k$ unknowns are in general nonzero). In other words, the solution vector v_s has a 1 in the position of x_{i_s}, while for $r \neq s$ the vector v_r has a 0 in this same position. This means that each of the k columns corresponding to the free variables in the matrix B contains a single 1 and the rest zeros. We now interchange column 1 and column i_1, then column 2 and column $i_2, \dots,$ and finally column k and column i_k. This yields the matrix

$$C = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & 0 & b_{1,k+1} & \cdots & b_{1n} \\ 0 & 1 & 0 & \cdots & 0 & 0 & b_{2,k+1} & \cdots & b_{2n} \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 1 & b_{k,k+1} & \cdots & b_{kn} \end{bmatrix}$$

where the entries $b_{i,k+1}, \ldots, b_{in}$ are the values of the remaining m unknowns in the solution vector v_i. Since the matrix C is in row echelon form, its rows are independent and hence $\operatorname{rank}(C) = k$. However, C is column-equivalent to B, and therefore $\operatorname{rank}(B) = k$ also (by Theorem 2.4 applied to columns). But the rows of B consist precisely of the k solution vectors v_s, and thus these solution vectors must be independent as claimed.

Example 2.7. Consider the homogeneous system of linear equations

$$\begin{aligned} x + 2y - 4z + 3w - t &= 0 \\ x + 2y - 2z + 2w + t &= 0 \\ 2x + 4y - 2z + 3w + 4t &= 0 \end{aligned}$$

If we reduce this system to row echelon form, we obtain

$$\begin{aligned} x + 2y - 4z + 3w - t &= 0 \\ 2z - w + 2t &= 0 \end{aligned} \tag{2.7}$$

It is obvious that the rank of the matrix of coefficients is 2, and hence the dimension of the solution space is $5 - 2 = 3$. The free variables are clearly y, w and t. The solution vectors v_s are obtained by choosing $(y = 1, w = 0, t = 0)$, $(y = 0, w = 1, t = 0)$ and $(y = 0, w = 0, t = 1)$. Using each of the these in equation 2.7, we obtain the solutions

$$\begin{aligned} v_1 &= (-2, 1, 0, 0, 0) \\ v_2 &= (-1, 0, 1/2, 1, 0) \\ v_3 &= (-3, 0, -1, 0, 1) \end{aligned}$$

Thus the vectors v_1, v_2 and v_3 form a basis for the solution space of the homogeneous system.

We emphasize that the corollary to Theorem 2.4 shows us that the solution set of a homogeneous system of equations is unchanged by elementary row operations. It is this fact that allows us to proceed as we did in Example 2.7.

We now turn our attention to the solutions of an inhomogeneous system of equations $\sum_j a_{ij}x_j = y_i$.

Theorem 2.11. *Let an inhomogeneous system of linear equations have matrix of coefficients A. Then the system has a solution if and only if* $\operatorname{rank}(A) = \operatorname{rank}(\operatorname{aug} A)$.

Proof. Let $c = (c_1, \ldots, c_n)$ be a solution of $\sum_j a_{ij}x_j = y_i$. Then writing this as

$$\sum_j c_j A^j = Y$$

shows us that Y is in the column space of A, and hence

$$\operatorname{rank}(\operatorname{aug} A) = \operatorname{cr}(\operatorname{aug} A) = \operatorname{cr}(A) = \operatorname{rank}(A).$$

Conversely, if $\operatorname{cr}(\operatorname{aug} A) = \operatorname{rank}(\operatorname{aug} A) = \operatorname{rank}(A) = \operatorname{cr}(A)$, then Y is in the column space of A, and hence $Y = \sum c_j A^j$ for some set of scalars c_j. But then the vector $c = (c_1, \ldots, c_n)$ is a solution since it obviously satisfies $\sum_j a_{ij} x_j = y_i$. ∎

Using Theorem 2.10, it is easy to describe the general solution to an inhomogeneous system of equations.

Theorem 2.12. *Let*

$$\sum_{j=1}^{n} a_{ij} x_j = y_i$$

be a system of inhomogeneous linear equations. If $u = (u_1, \ldots, u_n) \in \mathcal{F}^n$ is a solution of this system, and if S is the solution space of the associated homogeneous system, then the set

$$u + S = \{u + v : v \in S\}$$

is the solution set of the inhomogeneous system.

Proof. If $w = (w_1, \ldots, w_n) \in \mathcal{F}^n$ is any other solution of $\sum_j a_{ij} x_j = y_i$, then

$$\sum_j a_{ij}(w_j - u_j) = \sum_j a_{ij} w_j - \sum_j a_{ij} u_j = y_i - y_i = 0$$

so that $w - u \in S$, and hence $w = u + v$ for some $v \in S$. Conversely, if $v \in S$ then

$$\sum_j a_{ij}(u_j + v_j) = \sum_j a_{ij} u_j + \sum_j a_{ij} v_j = y_j + 0 = y_j$$

so that $u + v$ is a solution of the inhomogeneous system. ∎

Theorem 2.13. *Let A be an $n \times n$ matrix of rank n. Then the system*

$$\sum_{j=1}^{n} A^j x_j = Y$$

has a unique solution for arbitrary vectors $Y \in \mathcal{F}^n$.

Proof. Since $Y = \sum A^j x_j$, we see that $Y \in \mathcal{F}^n$ is just a linear combination of the columns of A. Since $\operatorname{rank}(A) = n$, it follows that the columns of A are linearly independent and hence form a basis for $\mathcal{F}^n$. But then any $Y \in \mathcal{F}^n$ has a unique expansion in terms of this basis (Theorem 1.4, Corollary 2) so that the vector X with components x_j must be unique. ∎

Example 2.8. Let us find the complete solution set over the real numbers of the inhomogeneous system

$$\begin{aligned} 3x_1 + x_2 + 2x_3 + 4x_4 &= 1 \\ x_1 - x_2 + 3x_3 - x_4 &= 3 \\ x_1 + 7x_2 - 11x_3 + 13x_4 &= -13 \\ 11x_1 + x_2 + 12x_3 + 10x_4 &= 9 \end{aligned}$$

We assume that we somehow found a particular solution $u = (2, 5, 1, -3) \in \mathbb{R}^4$, and hence we seek the solution set S of the associated homogeneous system. The matrix of coefficients A of the homogeneous system is given by

$$A = \begin{bmatrix} 3 & 1 & 2 & 4 \\ 1 & -1 & 3 & -1 \\ 1 & 7 & -11 & 13 \\ 11 & 1 & 12 & 10 \end{bmatrix}$$

The first thing we must do is determine $\operatorname{rank}(A)$. Since the proof of Theorem 2.10 dealt with columns, we choose to construct a new matrix B by applying elementary column operations to A. Thus we define

$$B = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -1 & 4 & 5 & 3 \\ 7 & -20 & -25 & -15 \\ 1 & 8 & 10 & 6 \end{bmatrix}$$

where the columns of B are given in terms of those of A by $B^1 = A^2$, $B^2 = A^1 - 3A^2$, $B^3 = A^3 - 2A^2$ and $B^4 = A^4 - 4A^2$. It is obvious that B^1 and B^2 are independent, and we also note that $B^3 = (5/4)B^2$ and $B^4 = (3/4)B^2$. Then $\operatorname{rank}(A) = \operatorname{rank}(B) = 2$, and hence we have $\dim S = 4 - 2 = 2$.

(An alternative method of finding $\operatorname{rank}(A)$ is as follows. If we interchange the first two rows of A and then add a suitable multiple the new first row to eliminate the first entry in each of the remaining three rows, we obtain

$$\begin{bmatrix} 1 & -1 & 3 & -1 \\ 0 & 4 & -7 & 7 \\ 0 & 8 & -14 & 14 \\ 0 & 12 & -21 & 21 \end{bmatrix}$$

It is now clear that the first two rows of this matrix are independent, and that the third and fourth rows are each multiples of the second. Therefore $\operatorname{rank}(A) = 2$ as above.)

We now follow the first part of the proof of Theorem 2.10. Observe that since $\operatorname{rank}(A) = 2$ and the first two columns of A are independent, we may write

$$A^3 = (5/4)A^1 - (7/4)A^2$$

and

$$A^4 = (3/4)A^1 + (7/4)A^2.$$

We therefore define the vectors

$$b^{(3)} = (5/4, -7/4, -1, 0)$$

and

$$b^{(4)} = (3/4, 7/4, 0, -1)$$

which are independent solutions of the homogeneous system and span the solution space S. Therefore the general solution set to the inhomogeneous system is given by

$$\begin{aligned} u + S &= \{u + \alpha b^{(3)} + \beta b^{(4)}\} \\ &= \{(2, 5, 1, -3) + \alpha(5/4, -7/4, -1, 0) + \beta(3/4, 7/4, 0, 1)\} \end{aligned}$$

where $\alpha, \beta \in \mathbb{R}$ are arbitrary.

Exercises

1. Find the dimension and a basis for the solution space of each of the following systems of linear equations over $\mathbb{R}$:

(a) $$\begin{aligned} x + 4y + 2z &= 0 \\ 2x + y + 5z &= 0 \end{aligned}$$

(b) $$\begin{aligned} x + 3y + 2z &= 0 \\ x + 5y + z &= 0 \\ 3x + 5y + 8z &= 0 \end{aligned}$$

(c) $$\begin{aligned} x + 2y + 2z - w + 3t &= 0 \\ x + 2y + 3z + w + t &= 0 \\ 3x + 6y + 8z + w + t &= 0 \end{aligned}$$

(d) $$\begin{aligned} x + 2y - 2z - 2w - t &= 0 \\ x + 2y - z + 3w - 2t &= 0 \\ 2x + 4y - 7z + w + t &= 0 \end{aligned}$$

2. Consider the subspaces U and V of $\mathbb{R}^4$ given by

$$\begin{aligned} U &= \{(a, b, c, d) \in \mathbb{R}^4 : b + c + d = 0\} \\ V &= \{(a, b, c, d) \in \mathbb{R}^4 : a + b = 0 \text{ and } c = 2d\}. \end{aligned}$$

(a) Find the dimension and a basis for U.
(b) Find the dimension and a basis for V.
(c) Find the dimension and a basis for $U \cap V$.

3. Find the complete solution set of each of the following systems of linear equations over $\mathbb{R}$:

(a) $$\begin{aligned} 3x - y &= 7 \\ 2x + y &= 1 \end{aligned}$$

(b) $$\begin{aligned} 2x - y + 3z &= 5 \\ 3x + 2y - 2z &= 1 \\ 7x + 4z &= 11 \end{aligned}$$

(c) $$\begin{aligned} 5x + 2y - z &= 0 \\ 3x + 5y + 3z &= 0 \\ x + 8y + 7z &= 0 \end{aligned}$$

(d) $$\begin{aligned} x - y + 2z + w &= 3 \\ 2x + y - z - w &= 1 \\ 3x + y + z - 3w &= 2 \\ 3x - 2y + 6z &= 7 \end{aligned}$$

2.5 Matrix Algebra

We now introduce the elementary algebraic operations on matrices. These operations will be of the utmost importance throughout the remainder of this text. In Chapter 4 we will see how these definitions arise in a natural way from the algebra of linear transformations.

Given two $m \times n$ matrices $A = (a_{ij})$ and $B = (b_{ij})$, we define their **sum** $A + B$ to be the matrix with entries

$$(A + B)_{ij} = a_{ij} + b_{ij}$$

obtained by adding the corresponding entries of each matrix. Note that both A and B must be of the same size. We also say that A **equals** B if $a_{ij} = b_{ij}$ for all i and j. It is obvious that

$$A + B = B + A$$

and that

$$A + (B + C) = (A + B) + C$$

for any other $m \times n$ matrix C. We also define the **zero matrix** 0 as that matrix for which $A + 0 = A$. In other words, $(0)_{ij} = 0$ for every i and j. Given a matrix $A = (a_{ij})$, we define its **negative** (or **additive inverse**)

$$-A = (-a_{ij})$$

such that $A + (-A) = 0$. Finally, for any scalar c we define the product of c and A to be the matrix

$$cA = (ca_{ij}).$$

Since in general the entries a_{ij} in a matrix $A = (a_{ij})$ are independent of each other, it should now be clear that the set of all $m \times n$ matrices forms a vector space of dimension mn over a field $\mathcal{F}$ of scalars. In other words, any $m \times n$ matrix A with entries a_{ij} can be written in the form

$$A = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} E_{ij}$$

where the $m \times n$ *matrix* E_{ij} is defined as having a 1 in the (i, j)th position and 0's elsewhere, and there are clearly mn such matrices. Observe that another way of describing the matrix E_{ij} is to say that it has entries $(E_{ij})_{rs} = \delta_{ir}\delta_{js}$.

We denote the space of all $m \times n$ matrices over the field $\mathcal{F}$ by $M_{m\times n}(\mathcal{F})$. The particular case of $m = n$ defines the space $M_n(\mathcal{F})$ of all **square** matrices of **size** n. We will often refer to a matrix in $M_n(\mathcal{F})$ as an **n-square** matrix.

Now let $A \in M_{m\times n}(\mathcal{F})$ be an $m \times n$ matrix, $B \in M_{r\times m}(\mathcal{F})$ be an $r \times m$ matrix, and consider the two systems of linear equations

$$\sum_{j=1}^{n} a_{ij}x_j = y_i, \qquad i = 1, \dots, m$$

and

$$\sum_{j=1}^{m} b_{ij}y_j = z_i, \qquad i = 1, \dots, r$$

where $X = (x_1, \dots, x_n) \in \mathcal{F}^n$, $Y = (y_1, \dots, y_m) \in \mathcal{F}^m$ and $Z = (z_1, \dots, z_r) \in \mathcal{F}^r$. Substituting the first of these equations into the second yields

$$z_i = \sum_{j=1}^{m} b_{ij}y_j = \sum_{j=1}^{m} b_{ij} \sum_{k=1}^{n} a_{jk}x_k = \sum_{k=1}^{n} c_{ik}x_k$$

where we defined the **product** of the $r \times m$ matrix B and the $m \times n$ matrix A to be the $r \times n$ matrix $C = BA$ whose entries are given by

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

Thus the (i, k)th entry of $C = BA$ is given by the standard scalar product

$$(BA)_{ik} = B_i \cdot A^k$$

of the ith row of B with the kth column of A (where both A^k and B_i are considered as vectors in $\mathcal{F}^m$). Note that matrix multiplication is generally not commutative, i.e., $AB \neq BA$. Indeed, the product AB may not even be defined.

Example 2.9. Let A and B be given by

$$A = \begin{bmatrix} 1 & 6 & -2 \\ 3 & 4 & 5 \\ 7 & 0 & 8 \end{bmatrix} \qquad B = \begin{bmatrix} 2 & -9 \\ 6 & 1 \\ 1 & -3 \end{bmatrix}.$$

Then the product of A and B is given by

$$C = AB = \begin{bmatrix} 1 & 6 & -2 \\ 3 & 4 & 5 \\ 7 & 0 & 8 \end{bmatrix} \begin{bmatrix} 2 & -9 \\ 6 & 1 \\ 1 & -3 \end{bmatrix}$$

$$= \begin{bmatrix} 1\cdot 2+6\cdot 6-2\cdot 1 & -1\cdot 9+6\cdot 1+2\cdot 3 \\ 3\cdot 2+4\cdot 6+5\cdot 1 & -3\cdot 9+4\cdot 1-5\cdot 3 \\ 7\cdot 2+0\cdot 6+8\cdot 1 & -7\cdot 9+0\cdot 1-8\cdot 3 \end{bmatrix}$$

$$= \begin{bmatrix} 36 & 3 \\ 35 & -38 \\ 22 & -87 \end{bmatrix}.$$

Note that it makes no sense to evaluate the product BA.

It is also easy to see that if we have the matrices

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

then

$$AB = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 4 & 3 \end{bmatrix}$$

while

$$BA = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} = \begin{bmatrix} 3 & 4 \\ 1 & 2 \end{bmatrix} \neq AB.$$

Example 2.10. Two other special cases of matrix multiplication are worth explicitly mentioning. Let $X \in \mathbb{R}^n$ be the column vector

$$X = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}.$$

If A is an $m \times n$ matrix, we may consider X to be an $n \times 1$ matrix and form the product AX:

$$AX = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} a_{11}x_1 + \cdots + a_{1n}x_n \\ \vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n \end{bmatrix} = \begin{bmatrix} A_1 \cdot X \\ \vdots \\ A_m \cdot X \end{bmatrix}.$$

As expected, the product AX is an $m \times 1$ matrix with entries given by the standard scalar product $A_i \cdot X$ in $\mathbb{R}^n$ of the ith row of A with the vector X. Note that this may also be written in the form

$$AX = \begin{bmatrix} a_{11} \\ \vdots \\ a_{m1} \end{bmatrix} x_1 + \cdots + \begin{bmatrix} a_{1n} \\ \vdots \\ a_{mn} \end{bmatrix} x_n$$

which clearly shows that AX is just a linear combination of the columns of A.

Now let $Y \in \mathbb{R}^m$ be the row vector $Y = (y_1, \dots, y_m)$. If we view this as a $1 \times m$ matrix, then we may form the $1 \times n$ matrix product YA given by

$$\begin{aligned} YA &= (y_1, \dots, y_m) \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \\ &= (y_1 a_{11} + \cdots + y_m a_{m1}, \dots, y_1 a_{1n} + \cdots + y_m a_{mn}) \\ &= (Y \cdot A^1, \dots, Y \cdot A^n) \end{aligned}$$

This again yields the expected form of the product with entries $Y \cdot A^i$.

This example suggests the following commonly used notation for systems of linear equations. Consider the system

$$\sum_{j=1}^{n} a_{ij} x_j = y_i$$

where $A = (a_{ij})$ is an $m \times n$ matrix. Suppose that we define the column vectors

$$X = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \in \mathcal{F}^n \quad \text{and} \quad Y = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix} \in \mathcal{F}^m.$$

If we consider X to be an $n \times 1$ matrix and Y to be an $m \times 1$ matrix, then we may write this system in matrix notation as

$$AX = Y.$$

Note that the ith row vector of A is $A_i = (a_{i1}, \dots, a_{in})$, so if $\mathcal{F} = \mathbb{R}$ the expression $\sum_j a_{ij} x_j = y_i$ may be written as the standard scalar product

$$A_i \cdot X = y_i.$$

We leave it to the reader to show that if A is an $n \times n$ matrix, then

$$AI_n = I_n A = A.$$

Even if A and B are both square matrices (i.e., matrices of the form $m \times m$), the product AB will not generally be the same as BA unless A and B are diagonal matrices (see Exercise 2.5.4). However, we do have the following.

Theorem 2.14. *For matrices of proper size (so these operations are defined), we have:*

(i) $(AB)C = A(BC)$ *(associative law).*
(ii) $A(B+C) = AB + AC$ *(left distributive law).*
(iii) $(B+C)A = BA + CA$ *(right distributive law).*
(iv) $k(AB) = (kA)B = A(kB)$ *for any scalar* k.

Proof. (i) $[(AB)C]_{ij} = \sum_k (AB)_{ik}c_{kj} = \sum_{r,k}(a_{ir}b_{rk})c_{kj} = \sum_{r,k} a_{ir}(b_{rk}c_{kj})$
$= \sum_r a_{ir}(BC)_{rj} = [A(BC)]_{ij}$.

(ii) $[A(B+C)]_{ij} = \sum_k a_{ik}(B+C)_{kj} = \sum_k a_{ik}(b_{kj}+c_{kj})$
$= \sum_k a_{ik}b_{kj} + \sum_k a_{ik}c_{kj} = (AB)_{ij} + (AC)_{ij}$
$= [(AB)+(AC)]_{ij}$.

(iii) Left to the reader (Exercise 2.5.1).
(iv) Left to the reader (Exercise 2.5.1). ∎

Given a matrix $A = (a_{ij})$, we define the **transpose** of A, denoted by $A^T = (a^T_{ij})$ to be the matrix with entries given by $a^T_{ij} = a_{ji}$. In other words, if A is an $m \times n$ matrix, then A^T is an $n \times m$ matrix whose columns are just the rows of A. Note in particular that a column vector is just the transpose of a row vector.

Example 2.11. If A is given by

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$$

then A^T is given by

$$\begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}.$$

Theorem 2.15. *The transpose has the following properties:*

(i) $(A+B)^T = A^T + B^T$.
(ii) $(A^T)^T = A$.
(iii) $(cA)^T = cA^T$ *for any scalar* c.
(iv) $(AB)^T = B^TA^T$.

Proof. (i) $[(A+B)^T]_{ij} = [(A+B)]_{ji} = a_{ji} + b_{ji} = a^T_{ij} + b^T_{ij} = (A^T + B^T)_{ij}$.
(ii) $(A^T)^T_{ij} = (A^T)_{ji} = a_{ij} = (A)_{ij}$.

(iii) $(cA)^T_{ij} = (cA)_{ji} = ca_{ij} = c(A^T)_{ij}$.
(iv) $(AB)^T_{ij} = (AB)_{ji} = \sum_k a_{jk}b_{ki} = \sum_k b^T_{ik}a^T_{kj} = (B^T A^T)_{ij}$. ∎

Our last basic property of the transpose is the following theorem, the easy proof of which is left as an exercise.

Theorem 2.16. *For any matrix A we have* $\operatorname{rank}(A^T) = \operatorname{rank}(A)$.

Proof. This is Exercise 2.5.2. ∎

We now wish to relate this matrix algebra to our previous results dealing with the rank of a matrix. Before doing so, let us first make some elementary observations dealing with the rows and columns of a matrix product. Assume that $A \in M_{m\times n}(\mathcal{F})$ and $B \in M_{n\times r}(\mathcal{F})$ so that the product AB is defined. Since the (i,j)th entry of AB is given by $(AB)_{ij} = \sum_k a_{ik}b_{kj}$, we see that the ith row of AB is given by a linear combination of the rows of B:

$$(AB)_i = \left(\sum_k a_{ik}b_{k1}, \dots, \sum_k a_{ik}b_{kr}\right) = \sum_k a_{ik}\,(b_{k1}, \dots, b_{kr}) = \sum_k a_{ik}B_k.$$

This shows that the row space of AB is a subspace of the row space of B. Another way to write this is to observe that

$$\begin{aligned}(AB)_i &= \left(\sum_k a_{ik}b_{k1}, \dots, \sum_k a_{ik}b_{kr}\right)\\ &= (a_{i1}, \dots, a_{in})\begin{bmatrix} b_{11} & \cdots & b_{1r} \\ \vdots & & \vdots \\ b_{n1} & \cdots & b_{nr}\end{bmatrix} = A_iB.\end{aligned}$$

Similarly, for the columns of a product we find that the jth column of AB is a linear combination of the columns of A:

$$(AB)^j = \begin{bmatrix} \sum_k a_{1k}b_{kj} \\ \vdots \\ \sum_k a_{mk}b_{kj}\end{bmatrix} = \sum_{k=1}^n \begin{bmatrix} a_{1k} \\ \vdots \\ a_{mk}\end{bmatrix} b_{kj} = \sum_{k=1}^n A^k b_{kj}$$

and therefore the column space of AB is a subspace of the column space of A. We also have the result

$$(AB)^j = \begin{bmatrix} \sum_k a_{1k}b_{kj} \\ \vdots \\ \sum_k a_{mk}b_{kj}\end{bmatrix} = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn}\end{bmatrix}\begin{bmatrix} b_{1j} \\ \vdots \\ b_{nj}\end{bmatrix} = AB^j.$$

These formulas will be quite useful to us in a number of theorems and calculations.

Theorem 2.17. *If A and B are any matrices for which the product AB is defined, then the row space of AB is a subspace of the row space of B, and the column space of AB is a subspace of the column space of A.*

Proof. As we saw above, using $(AB)_i = \sum_k a_{ik} B_k$ it follows that the ith row of AB is in the space spanned by the rows of B, and hence the row space of AB is a subspace of the row space of B.

As to the column space, this was also shown above. Alternatively, note that the column space of AB is just the row space of $(AB)^T = B^T A^T$, which is a subspace of the row space of A^T by the first part of the theorem. But the row space of A^T is just the column space of A. ∎

Corollary. $\operatorname{rank}(AB) \leq \min\{\operatorname{rank}(A), \operatorname{rank}(B)\}$.

Proof. Let $\operatorname{row}(A)$ be the row space of A, and let $\operatorname{col}(A)$ be the column space of A. Then

$$\operatorname{rank}(AB) = \dim(\operatorname{row}(AB)) \leq \dim(\operatorname{row}(B)) = \operatorname{rank}(B)$$

while

$$\operatorname{rank}(AB) = \dim(\operatorname{col}(AB)) \leq \dim(\operatorname{col}(A)) = \operatorname{rank}(A).$$ ∎

Let us now prove some very useful properties of the row and column spaces of a matrix. To begin with, suppose $A \in M_{m\times n}(\mathcal{F})$. We define the **kernel** of A to be the set

$$\ker A = \{X \in \mathcal{F}^n : AX = 0\}.$$

(In the context of matrices, this is usually called the **null space** of A and denoted by $\operatorname{nul}(A)$. The dimension of $\operatorname{nul}(A)$ is then called the **nullity** of A and is denoted $\operatorname{nullity}(A)$ or $\operatorname{null}(A)$. Since we think this is somewhat confusing, and in Chapter 4 we will use $\operatorname{nul}(A)$ to denote the dimension of the kernel of a linear transformation, we chose the notation as we did.)

It is easy to see that $\ker A$ is a subspace of $\mathcal{F}^n$. Indeed, if $X, Y \in \ker A$ and $k \in \mathcal{F}$, then clearly $A(kX+Y) = kAX + AY = 0$ so that $kX + Y \in \ker A$ also. In fact, $\ker A$ is just the solution set to the homogeneous system $AX = 0$, and therefore $\dim(\ker A)$ is just the dimension $\dim S$ of the solution set. In view of Theorem 2.10, this proves the following very useful result, known as the **rank theorem** (or the **dimension theorem**).

Theorem 2.18. *Let $A \in M_{m\times n}(\mathcal{F})$. Then*

$$\operatorname{rank} A + \dim(\ker A) = n.$$

Another very useful result we will need in a later chapter comes by considering the orthogonal complement of the row space of a real matrix $A \in M_{m\times n}(\mathbb{R})$. This is by definition the set of all $X \in \mathbb{R}^n$ that are orthogonal to every row of A. In other words, using the standard inner product on $\mathbb{R}^n$ we have

$$(\mathrm{row}(A))^\perp = \{X \in \mathbb{R}^n : A_i \cdot X = 0 \text{ for all } i = 1, \dots, m\}.$$

But this is just the homogeneous system $AX = 0$ and hence $(\mathrm{row}(A))^\perp = \ker A$. Applying this relation to A^T we have $(\mathrm{row}(A^T))^\perp = \ker A^T$. But $\mathrm{row}(A^T) = \mathrm{col}(A)$ and therefore $(\mathrm{col}(A))^\perp = \ker A^T$.

We state this result as a theorem for future reference.

Theorem 2.19. *Let $A \in M_{m\times n}(\mathbb{R})$. Then*

$$(\mathrm{row}(A))^\perp = \ker A \quad \textit{and} \quad (\mathrm{col}(A))^\perp = \ker A^T.$$

Exercises

1. Complete the proof of Theorem 2.14.
2. Prove Theorem 2.16.
3. Let A be any $m \times n$ matrix and let X be any $n \times 1$ matrix, both with entries in $\mathcal{F}$. Define the mapping $f : \mathcal{F}^n \to \mathcal{F}^m$ by $f(X) = AX$.
   (a) Show that f is a linear transformation (i.e., a vector space homomorphism).
   (b) Define the set $\mathrm{Im}\, f = \{AX : X \in \mathcal{F}^n\}$. Show that $\mathrm{Im}\, f$ is a subspace of $\mathcal{F}^m$.
   (c) Let U be the column space of A. Show that $\mathrm{Im}\, f = U$. [*Hint*: Use Example 2.10 to show that $\mathrm{Im}\, f \subset U$. Next, use the equation $(AI)^j = AI^j$ to show that $U \subset \mathrm{Im}\, f$.]
   (d) Let N denote the solution space to the system $AX = 0$. In other words, $N = \{X \in \mathcal{F}^n : AX = 0\}$. ($N$ is just the null space of A.) Show that
   $$\dim N + \dim U = n.$$
   [*Hint*: Suppose $\dim N = r$, and extend a basis $\{x_1, \dots, x_r\}$ for N to a basis $\{x_i\}$ for $\mathcal{F}^n$. Show that U is spanned by the vectors $Ax_{r+1}, \dots, Ax_n$, and then that these vectors are linearly independent. Note that this exercise is really just another proof of Theorem 2.10.]
4. A square matrix of the form
   $$\begin{bmatrix} a_{11} & 0 & \cdots & 0 \\ 0 & a_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{nn} \end{bmatrix}$$

is called a **diagonal** matrix. In other words, $A = (a_{ij})$ is diagonal if $a_{ij} = 0$ for $i \neq j$. If A and B are both square matrices, we may define the **commutator** $[A, B]$ of A and B to be the matrix $[A, B] = AB - BA$. If $[A, B] = 0$, we say that A and B **commute**.

(a) Show that any diagonal matrices A and B commute.
(b) Prove that the only $n \times n$ matrices which commute with every $n \times n$ diagonal matrix are diagonal matrices.

5. Given the matrices

$$A = \begin{bmatrix} 2 & -1 \\ 1 & 0 \\ -3 & 4 \end{bmatrix} \qquad B = \begin{bmatrix} 1 & -2 & -5 \\ 3 & 4 & 0 \end{bmatrix}$$

compute the following:

(a) AB.
(b) BA.
(c) AA^T.
(d) $A^T A$.
(e) Verify that $(AB)^T = B^T A^T$.

6. Consider the matrix $A \in M_n(\mathcal{F})$ given by

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 \end{bmatrix}$$

Thus A has zero entries everywhere except on the **superdiagonal** where the entries are 1's. Let $A^2 = AA$, $A^3 = AAA$, and so on. Show that $A^n = 0$ but $A^{n-1} \neq 0$.

7. Given a matrix $A = (a_{ij}) \in M_n(\mathcal{F})$, the sum of the diagonal elements of A is called the **trace** of A, and is denoted by $\operatorname{tr} A$. Thus

$$\operatorname{tr} A = \sum_{i=1}^{n} a_{ii}.$$

(a) Prove that $\operatorname{tr}(A + B) = \operatorname{tr} A + \operatorname{tr} B$ and that $\operatorname{tr}(\alpha A) = \alpha(\operatorname{tr} A)$ for any scalar α.
(b) Prove that $\operatorname{tr}(AB) = \operatorname{tr}(BA)$.

8. Prove that it is impossible to find matrices $A, B \in M_n(\mathbb{R})$ such that their **commutator** $[A, B] = AB - BA$ is equal to 1.

9. A matrix $A = (a_{ij})$ is said to be **upper triangular** if $a_{ij} = 0$ for $i > j$. In other words, every entry of A below the main diagonal is zero. Similarly, A is said to be **lower triangular** if $a_{ij} = 0$ for $i < j$. Prove that the product of upper (lower) triangular matrices is an upper (lower) triangular matrix.

10. Consider the so-called **Pauli spin matrices**

$$\sigma_1 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \qquad \sigma_2 = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix} \qquad \sigma_3 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

and define the **permutation symbol** ε_{ijk} by

$$\varepsilon_{ijk} = \begin{cases} 1 & \text{if } (i, j, k) \text{ is an even permutation of } (1, 2, 3) \\ -1 & \text{if } (i, j, k) \text{ is an odd permutation of } (1, 2, 3) \\ 0 & \text{if and two indices are the same} \end{cases} .$$

The **commutator** of two matrices $A, B \in M_n(\mathcal{F})$ is defined by $[A, B] = AB - BA$, and the **anticommutator** is given by $[A, B]_+ = AB + BA$.

(a) Show that $[\sigma_i, \sigma_j] = 2i \sum_k \varepsilon_{ijk}\sigma_k$. In other words, show that $\sigma_i\sigma_j = i\sigma_k$ where (i, j, k) is an even permutation of $(1, 2, 3)$.
(b) Show that $[\sigma_i, \sigma_j]_+ = 2I\delta_{ij}$.
(c) Using part (a), show that $\operatorname{tr}\sigma_i = 0$.
(d) For notational simplicity, define $\sigma_0 = I$. Show that $\{\sigma_0, \sigma_1, \sigma_2, \sigma_3\}$ forms a basis for $M_2(\mathbb{C})$. [*Hint*: Show that $\operatorname{tr}(\sigma_\alpha\sigma_\beta) = 2\delta_{\alpha\beta}$ where $0 \le \alpha, \beta \le 3$. Use this to show that $\{\sigma_\alpha\}$ is linearly independent.]
(e) According to part (d), any $X \in M_2(\mathbb{C})$ may be written in the form $X = \sum_\alpha x_\alpha\sigma_\alpha$. How would you find the coefficients x_α?
(f) Show that $\langle\sigma_\alpha, \sigma_\beta\rangle = (1/2)\operatorname{tr}(\sigma_\alpha\sigma_\beta)$ defines an inner product on $M_2(\mathbb{C})$.
(g) Show that any matrix $X \in M_2(\mathbb{C})$ that commutes with all of the σ_i (i.e., $[X, \sigma_i] = 0$ for each $i = 1, 2, 3$) must be a multiple of the identity matrix.

11. A square matrix S is said to be **symmetric** if $S^T = S$, and a square matrix A is said to be **skewsymmetric** (or **antisymmetric**) if $A^T = -A$.

(a) Show that $S \neq 0$ and $A \neq 0$ are linearly independent in $M_n(\mathcal{F})$.
(b) What is the dimension of the space of all $n \times n$ symmetric matrices?
(c) What is the dimension of the space of all $n \times n$ antisymmetric matrices?

12. Find a basis $\{A_i\}$ for the space $M_n(\mathcal{F})$ that consists only of matrices with the property that ${A_i}^2 = A_i$ (such matrices are called **idempotent** or **projections**). [*Hint*: The matrices

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

will work in the particular case of $M_2(\mathcal{F})$.]

13. Show that it is impossible to find a basis for $M_n(\mathcal{F})$ such that every pair of matrices in the basis commutes with each other.

14. (a) Show that the set of all nonsingular $n \times n$ matrices forms a spanning set for $M_n(\mathcal{F})$. Exhibit a basis of such matrices.
    (b) Repeat part (a) with the set of all singular matrices.

15. Show that the set of all matrices of the form $AB-BA$ do not span $M_n(\mathcal{F})$. [*Hint*: Use the trace.]

16. Is it possible to span $M_n(\mathcal{F})$ using powers of a single matrix A? In other words, can $\{I_n, A, A^2, \dots, A_n, \dots\}$ span $M_n(\mathcal{F})$? [*Hint*: Consider Exercise 4 above.]

2.6 Invertible Matrices

As mentioned earlier, we say that a matrix $A \in M_n(\mathcal{F})$ is **nonsingular** if $\operatorname{rank}(A) = n$, and **singular** if $\operatorname{rank}(A) < n$. Given a matrix $A \in M_n(\mathcal{F})$, if there exists a matrix $B \in M_n(\mathcal{F})$ such that $AB = BA = I_n$, then B is called an **inverse** of A, and A is said to be **invertible**.

Technically, a matrix B is called a **left inverse** of A if $BA = I$, and a matrix B' is a **right inverse** of A if $AB' = I$. Then, if $AB = BA = I$, we say that B is a **two-sided inverse** of A, and A is then said to be **invertible**. Furthermore, if A has a left inverse B and a right inverse B', then it is easy to see that $B = B'$ since $B = BI = B(AB') = (BA)B' = IB' = B'$. We shall now show that if B is either a left or a right inverse of A, then A is invertible. (We stress that this result is valid only in finite dimensions. In the infinite dimensional case, either a left or right inverse alone is not sufficient to show invertibility. This distinction is important in fields such as functional analysis and quantum mechanics.)

Theorem 2.20. *A matrix $A \in M_n(\mathcal{F})$ has a right (left) inverse if and only if A is nonsingular. This right (left) inverse is also a left (right) inverse, and hence is an inverse of A.*

Proof. Suppose A has a right inverse B. Then $AB = I_n$ so that $\operatorname{rank}(AB) = \operatorname{rank}(I_n)$. Since $\operatorname{rank}(I_n)$ is clearly equal to n (Theorem 2.6), we see that $\operatorname{rank}(AB) = n$. But then from the corollary to Theorem 2.17 and the fact that both A and B are $n \times n$ matrices (so that $\operatorname{rank}(A) \le n$ and $\operatorname{rank}(B) \le n$), it follows that $\operatorname{rank}(A) = \operatorname{rank}(B) = n$, and hence A is nonsingular.

Now suppose that A is nonsingular so that $\operatorname{rank}(A) = n$. If we let E^j be the jth column of the identity matrix I_n, then for each $j = 1, \dots, n$ the system of equations

$$\sum_{i=1}^{n} A^i x_i = AX = E^j$$

has a unique solution which we denote by $X = B^j$ (Theorem 2.13). Now let B be the matrix with columns B^j. Then the jth column of AB is given by

$$(AB)^j = AB^j = E^j$$

and hence $AB = I_n$. It remains to be shown that $BA = I_n$.

To see this, note that $\operatorname{rank}(A^T) = \operatorname{rank}(A) = n$ (Theorem 2.16) so that A^T is nonsingular also. Hence applying the same argument shows there exists a unique $n \times n$ matrix C^T such that $A^T C^T = I_n$. Since $(CA)^T = A^T C^T$ and ${I_n}^T = I_n$, this is the same as $CA = I_n$. We now recall that it was shown prior to the theorem that if A has both a left and a right inverse, then they are the same. Therefore $B = C$ so that $BA = AB = I_n$, and hence B is an inverse of A. Clearly, the proof remains valid if "right" is replaced by "left" throughout. ∎

This theorem has several important consequences which we state as corollaries.

Corollary 1. *A matrix $A \in M_n(\mathcal{F})$ is nonsingular if and only if it has an inverse. Furthermore, this inverse is unique.*

Proof. As we saw above, if B and C are both inverses of A, then $B = BI = B(AC) = (BA)C = IC = C$. ∎

In view of this corollary, the unique inverse to a matrix A will be denoted by A^{-1} from now on.

Corollary 2. *If A is an $n \times n$ nonsingular matrix, then A^{-1} is nonsingular and $(A^{-1})^{-1} = A$.*

Proof. If A is nonsingular, then (by Theorem 2.20) A^{-1} exists so that $A^{-1}A = AA^{-1} = I$. But this means that $(A^{-1})^{-1}$ exists and is equal to A, and hence A^{-1} is also nonsingular. ∎

Corollary 3. *If A and B are nonsingular then so is AB, and $(AB)^{-1} = B^{-1}A^{-1}$.*

Proof. The fact that A and B are nonsingular means that A^{-1} and B^{-1} exist. We therefore see that

$$(B^{-1}A^{-1})(AB) = B^{-1}IB = B^{-1}B = I$$

and similarly $(AB)(B^{-1}A^{-1}) = I$. It then follows that $B^{-1}A^{-1} = (AB)^{-1}$. Since we have now shown that AB has an inverse, Theorem 2.20 tells us that AB must be nonsingular. ∎

Corollary 4. *If A is nonsingular then so is A^T, and $(A^T)^{-1} = (A^{-1})^T$.*

Proof. That A^T is nonsingular is a direct consequence of Theorem 2.16. Next we observe that

$$(A^{-1})^T A^T = (AA^{-1})^T = I^T = I$$

so the uniqueness of the inverse tells us that $(A^T)^{-1} = (A^{-1})^T$. Note this also shows that A^T is nonsingular. ∎

Corollary 5. *A system of n linear equations in n unknowns has a unique solution if and only if its matrix of coefficients is nonsingular.*

Proof. Consider the system $AX = Y$. If A is nonsingular, then a unique A^{-1} exists, and therefore we have $X = A^{-1}Y$ as the unique solution. (Note that this is essentially the content of Theorem 2.13.)

Conversely, if this system has a unique solution, then the solution space of the associated homogeneous system must have dimension 0 (Theorem 2.12). Then Theorem 2.10 shows that we must have $\operatorname{rank}(A) = n$, and hence A is nonsingular. ∎

A major problem that we have not yet discussed is how to actually find the inverse of a matrix. One method involves the use of determinants as we will see in the next chapter. However, let us show another approach based on the fact that a nonsingular matrix is row-equivalent to the identity matrix (Theorem 2.7). This method has the advantage that it is algorithmic, and hence is easily implemented on a computer.

Since the jth column of a product AB is AB^j, we see that considering the particular case of $AA^{-1} = I$ leads to

$$(AA^{-1})^j = A(A^{-1})^j = E^j$$

where E^j is the jth column of I. What we now have is the inhomogeneous system

$$AX = Y$$

(or $\sum_j a_{ij}x_j = y_i$) where $X = (A^{-1})^j$ and $Y = E^j$. As we saw in Section 2.2, we may solve for the vector X by reducing the augmented matrix to reduced row echelon form. For the particular case of $j = 1$ we have

$$\operatorname{aug} A = \left[\begin{array}{ccc|c} a_{11} & \cdots & a_{1n} & 1 \\ a_{21} & \cdots & a_{2n} & 0 \\ \vdots & & \vdots & \vdots \\ a_{n1} & \cdots & a_{nn} & 0 \end{array}\right]$$

and hence the reduced form will be

$$\left[\begin{array}{ccccc|c} 1 & 0 & 0 & \cdots & 0 & c_{11} \\ 0 & 1 & 0 & \cdots & 0 & c_{21} \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & c_{n1} \end{array}\right]$$

for some set of scalars c_{ij}. This means that the solution to the system is $x_1 = c_{11}$, $x_2 = c_{21}$, $\ldots$, $x_n = c_{n1}$. But $X = (A^{-1})^1 =$ the first column of A^{-1}, and therefore this last matrix may be written as

$$\left[\begin{array}{ccc|c} 1 & \cdots & 0 & a^{-1}{}_{11} \\ \vdots & & \vdots & \vdots \\ 0 & \cdots & 1 & a^{-1}{}_{n1} \end{array}\right]$$

Now, for each $j = 1, \ldots, n$ the system $AX = A(A^{-1})^j = E^j$ always has the same matrix of coefficients, and only the last column of the augmented matrix depends on j. Since finding the reduced row echelon form of the matrix of coefficients is independent of this last column, it follows that we may solve all n systems simultaneously by reducing the single matrix

$$\left[\begin{array}{ccc|ccc} a_{11} & \cdots & a_{1n} & 1 & \cdots & 0 \\ \vdots & & \vdots & \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn} & 0 & \cdots & 1 \end{array}\right]$$

In other words, the reduced form will be

$$\left[\begin{array}{ccc|ccc} 1 & \cdots & 0 & a^{-1}{}_{11} & \cdots & a^{-1}{}_{1n} \\ \vdots & & \vdots & \vdots & & \vdots \\ 0 & \cdots & 1 & a^{-1}{}_{n1} & \cdots & a^{-1}{}_{nn} \end{array}\right]$$

where the matrix $A^{-1} = (a^{-1}{}_{ij})$ satisfies $AA^{-1} = I$ since $(AA^{-1})^j = A(A^{-1})^j = E^j$ is satisfied for each $j = 1, \ldots, n$.

Example 2.12. Let us find the inverse of the matrix A given by

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

We leave it as an exercise for the reader to show that the reduced row echelon form of

$$\left[\begin{array}{ccc|ccc} -1 & 2 & 1 & 1 & 0 & 0 \\ 0 & 3 & -2 & 0 & 1 & 0 \\ 2 & -1 & 0 & 0 & 0 & 1 \end{array}\right]$$

is

$$\left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 1/6 & 1/12 & 7/12 \\ 0 & 1 & 0 & 1/3 & 1/6 & 1/6 \\ 0 & 0 & 1 & 1/2 & -1/4 & 1/4 \end{array}\right]$$

and hence A^{-1} is given by

$$\begin{bmatrix} 1/6 & 1/12 & 7/12 \\ 1/3 & 1/6 & 1/6 \\ 1/2 & -1/4 & 1/4 \end{bmatrix}$$

Exercises

1. Verify the reduced row echelon form of the matrix given in Example 2.12.

2. Find the inverse of a general 2×2 matrix. What constraints are there on the entries of the matrix?

3. Show that a matrix is not invertible if it has any zero row or column.

4. Find the inverse of each of the following matrices:

   (a) $\begin{bmatrix} 1 & 0 & 2 \\ 2 & -1 & 3 \\ 4 & 1 & 8 \end{bmatrix}$ (b) $\begin{bmatrix} 1 & 3 & 4 \\ 3 & -1 & 6 \\ -1 & 5 & 1 \end{bmatrix}$ (c) $\begin{bmatrix} 1 & 2 & 1 \\ 2 & 5 & 2 \\ 1 & 3 & 3 \end{bmatrix}$

5. Use the inverse of the matrix in Exercise 4(c) above to find the solutions of each of the following systems:

   (a) $\begin{aligned} x + 2y + z &= 10 \\ 2x + 5y + 2z &= 14 \\ x + 3y + 3z &= 30 \end{aligned}$ (b) $\begin{aligned} x + 2y + z &= 2 \\ 2x + 5y + 2z &= -1 \\ x + 3y + 3z &= 6 \end{aligned}$

6. What is the inverse of a diagonal matrix?

7. (a) Prove that an upper triangular matrix is invertible if and only if every entry on the main diagonal is nonzero (see Exercise 2.5.9 for the definition of an upper triangular matrix).
   (b) Prove that the inverse of a lower (upper) triangular matrix is lower (upper) triangular.

8. Find the inverse of the following matrix:

$$\begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & 2 & 3 & 4 \\ 0 & 0 & 3 & 4 \\ 0 & 0 & 0 & 4 \end{bmatrix}$$

9. (a) Let A be any 2×1 matrix, and let B be any 1×2 matrix. Prove that AB is not invertible.
   (b) Repeat part (a) where A is any $m \times n$ matrix and B is any $n \times m$ matrix with $n < m$.

10. Summarize several of our results by proving the equivalence of the following statements for any $n \times n$ matrix A:
    (a) A is invertible.
    (b) The homogeneous system $AX = 0$ has only the zero solution.
    (c) The system $AX = Y$ has a solution X for every $n \times 1$ matrix Y.

11. Let A and B be square matrices of size n, and assume that A is nonsingular. Prove that $\operatorname{rank}(AB) = \operatorname{rank}(B) = \operatorname{rank}(BA)$.

12. A matrix A is called a **left zero divisor** if there exists a nonzero matrix B such that $AB = 0$, and A is called a **right zero divisor** if there exists a nonzero matrix C such that $CA = 0$. If A is an $m \times n$ matrix, prove that:
    (a) If $m < n$, then A is a left zero divisor.
    (b) If $m > n$, then A is a right zero divisor.
    (c) If $m = n$, then A is both a left and a right zero divisor if and only if A is singular.

13. Let A and B be nonsingular symmetric matrices for which $AB - BA = 0$. Show that AB, $A^{-1}B$, AB^{-1} and $A^{-1}B^{-1}$ are all symmetric.

2.7 Elementary Matrices

Recall the elementary row operations α, β, γ described in Section 2.2. We now let e denote any one of these three operations, and for any matrix A we define $e(A)$ to be the result of applying the operation e to the matrix A. In particular, we define an **elementary matrix** to be any matrix of the form $e(I)$. The great utility of elementary matrices arises from the following theorem.

Theorem 2.21. *If A is any $m \times n$ matrix and e is any elementary row operation, then*

$$e(A) = e(I_m)A.$$

Proof. We must verify this equation for each of the three types of elementary row operations. First consider an operation of type α. In particular, let α be the interchange of rows i and j. Then

$$[e(A)]_k = A_k \quad \text{for } k \neq i, j$$

while

$$[e(A)]_i = A_j \quad \text{and} \quad [e(A)]_j = A_i.$$

On the other hand, using $(AB)_k = A_k B$ we also have

$$[e(I)A]_k = [e(I)]_k A.$$

If $k \neq i, j$ then $[e(I)]_k = I_k$ (the kth row of I, not the $k \times k$ identity matrix) so that

$$[e(I)]_k A = I_k A = A_k.$$

Written out in full, the rules of matrix multiplication make it easy to see what it going on:

$$I_k A = [\, 0 \cdots 1 \cdots 0 \,] \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} = [\, a_{k1} \cdots a_{kn} \,] = A_k. \tag{2.8}$$

If $k = i$, then $[e(I)]_i = I_j$ and

$$[e(I)]_i A = I_j A = A_j.$$

Similarly, we see that

$$[e(I)]_j A = I_i A = A_i.$$

This verifies the theorem for transformations of type α.

For a type β transformation that multiplies row k by the scalar c, we just have $[e(I)]_i = I_i$ for $i \neq k$ and $[e(I)]_k = cI_k$. Then looking at equation (2.8) should make it obvious that $e(A) = e(I)A$.

We now go on to transformations of type γ. So, let e be the addition of c times row j to row i. Then

$$[e(I)]_k = I_k \qquad \text{for } k \neq i$$

and

$$[e(I)]_i = I_i + cI_j.$$

Therefore

$$[e(I)]_i A = (I_i + cI_j)A = A_i + cA_j = [e(A)]_i$$

and for $k \neq i$ we have

$$[e(I)]_k A = I_k A = A_k = [e(A)]_k. \qquad \blacksquare$$

If e is of type α, then rows i and j are interchanged. But this is readily undone by interchanging the same rows again, and hence e^{-1} is defined and is another elementary row operation. For type β operations, some row is multiplied by a scalar c, so in this case e^{-1} is simply multiplication by $1/c$. Finally, a type γ operation adds c times row j to row i, and hence e^{-1} adds $-c$ times row j to row i. Thus all three types of elementary row operations have inverses which are also elementary row operations.

By way of nomenclature, a square matrix $A = (a_{ij})$ is said to be **diagonal** if $a_{ij} = 0$ for $i \neq j$ (see Exercise 2.5.4). The most common example of a diagonal matrix is the identity matrix.

Theorem 2.22. *Every elementary matrix is nonsingular, and*

$$[e(I)]^{-1} = e^{-1}(I).$$

Furthermore, the transpose of an elementary matrix is an elementary matrix.

Proof. By definition, $e(I)$ is row equivalent to I and hence has the same rank as I (Theorem 2.4). Thus $e(I)$ is nonsingular since I is nonsingular, and hence $e(I)^{-1}$ exists. Since it was shown above that e^{-1} is an elementary row operation, we apply Theorem 2.21 to the matrix $e(I)$ to obtain

$$e^{-1}(I)e(I) = e^{-1}(e(I)) = I.$$

Similarly, applying Theorem 2.21 to $e^{-1}(I)$ yields

$$e(I)e^{-1}(I) = e(e^{-1}(I)) = I.$$

This shows that $e^{-1}(I) = [e(I)]^{-1}$.

Now let e be a type α transformation that interchanges rows i and j (with $i < j$). Then the ith row of $e(I)$ has a 1 in the jth column, and the jth row has a 1 in the ith column. In other words,

$$[e(I)]_{ij} = 1 = [e(I)]_{ji}$$

while for $r, s \neq i, j$ we have

$$[e(I)]_{rs} = 0 \quad \text{if } r \neq s$$

and

$$[e(I)]_{rr} = 1.$$

Taking the transpose shows that

$$[e(I)]^T{}_{ij} = [e(I)]_{ji} = 1 = [e(I)]_{ij}$$

and

$$[e(I)]^T{}_{rs} = [e(I)]_{sr} = 0 = [e(I)]_{rs}.$$

Thus $[e(I)]^T = e(I)$ for type α operations.

Since I is a diagonal matrix, it is clear that for a type β operation which simply multiplies one row by a nonzero scalar, we have $[e(I)]^T = e(I)$.

Finally, let e be a type γ operation that adds c times row j to row i. Then $e(I)$ is just I with the additional entry $[e(I)]_{ij} = c$, and hence $[e(I)]^T$ is just I with the additional entry $[e(I)]_{ji} = c$. But this is the same as c times row i added to row j in the matrix I. In other words, $[e(I)]^T$ is just another elementary matrix. ∎

We now come to our main result dealing with elementary matrices. For ease of notation, we denote an elementary matrix by E rather than by $e(I)$. In other words, the result of applying the elementary row operation e_i to I will be denoted by the matrix $E_i = e_i(I)$.

Theorem 2.23. *Every nonsingular $n \times n$ matrix may be written as a product of elementary $n \times n$ matrices.*

Proof. It follows from Theorem 2.7 that any nonsingular $n \times n$ matrix A is row equivalent to I_n. This means that I_n may be obtained by applying r successive elementary row operations to A. Hence applying Theorem 2.21 r times yields

$$E_r \cdots E_1 A = I_n$$

so that

$$A = {E_1}^{-1} \cdots {E_r}^{-1} I_n = {E_1}^{-1} \cdots {E_r}^{-1}.$$

The theorem now follows if we note that each ${E_i}^{-1}$ is an elementary matrix according to Theorem 2.22 (since ${E_i}^{-1} = [e(I)]^{-1} = e^{-1}(I)$ and e^{-1} is an elementary row operation). ∎

Corollary. *If A is an invertible $n \times n$ matrix, and if some sequence of elementary row operations reduces A to the identity matrix, then the same sequence of row operations reduces the identity matrix to A^{-1}.*

Proof. By hypothesis we may write $E_r \cdots E_1 A = I$. But then multiplying from the right by A^{-1} shows that $A^{-1} = E_r \cdots E_1 I$. ∎

Note this corollary provides another proof that the method given in the previous section for finding A^{-1} is valid.

There is one final important property of elementary matrices that we will need in a later chapter. Let E be an $n \times n$ elementary matrix representing any of the three types of elementary row operations, and let A be an $n \times n$ matrix. As we have seen, multiplying A from the left by E results in a new matrix with the same rows that would result from applying the elementary row operation to A directly. We claim that multiplying A from the right by E^T results in a new matrix whose columns have the same relationship as the rows of EA. We will prove this for a type γ operation, leaving the easier type α and β operations to the reader (see Exercise 2.7.1).

Let γ be the addition of c times row j to row i. Then the rows of E are given by $E_k = I_k$ for $k \neq i$, and $E_i = I_i + cI_j$. Therefore the columns of E^T are given by

$$(E^T)^k = I^k \quad \text{for } k \neq i$$

and

$$(E^T)^i = I^i + cI^j.$$

Now recall that the kth column of AB is given by $(AB)^k = AB^k$. We then have

$$(AE^T)^k = A(E^T)^k = AI^k = A^k \qquad \text{for } k \neq i$$

and

$$(AE^T)^i = A(E^T)^i = A(I^i + cI^j) = AI^i + cAI^j = A^i + cA^j.$$

This is the same relationship as that found between the rows of EA where $(EA)_k = A_k$ and $(EA)_i = A_i + cA_j$ (see the proof of Theorem 2.21).

Exercises

1. Let A be an $n \times n$ matrix, and let E be an $n \times n$ elementary matrix representing a type α or β operation. Show that the columns of AE^T have the same relationship as the rows of EA.

2. Write down 4×4 elementary matrices that will induce the following elementary operations in a 4×4 matrix when used as left multipliers. Verify that your answers are correct.
   (a) Interchange the $2nd$ and $4th$ rows of A.
   (b) Interchange the $2nd$ and $3rd$ rows of A.
   (c) Multiply the $4th$ row of A by 5.
   (d) Add k times the $4th$ row of A to the $1st$ row of A.
   (e) Add k times the $1st$ row of A to the $4th$ row of A.

3. Show that any $e_\alpha(A)$ may be written as a product of $e_\beta(A)$'s and $e_\gamma(A)$'s. (The notation should be obvious.)

4. Pick any 4×4 matrix A and multiply it from the *right* by each of the elementary matrices found in the previous problem. What is the effect on A?

5. Prove that a matrix A is row equivalent to a matrix B if and only if there exists a nonsingular matrix P such that $B = PA$.

6. Reduce the matrix
$$A = \begin{bmatrix} 1 & 0 & 2 \\ 0 & 3 & -1 \\ 2 & 3 & 3 \end{bmatrix}$$
to the reduced row echelon form R, and write the elementary matrix corresponding to each of the elementary row operations required. Find a nonsingular matrix P such that $PA = R$ by taking the product of these elementary matrices.

7. Let A be an $n \times n$ matrix. Summarize several of our results by proving that the following are equivalent:

(a) A is invertible.
(b) A is row equivalent to I_n .
(c) A is a product of elementary matrices.

8. Using the results of the previous problem, prove that if $A = A_1 A_2 \cdots A_k$ where each A_i is a square matrix, then A is invertible if and only if each of the A_i is invertible.

   The remaining problems are all connected, and should be worked in the given order.

9. Suppose that we define elementary column operations exactly as we did for rows. Prove that every elementary column operation on A can be achieved by multiplying A on the *right* by an elementary matrix. [*Hint*: You can either do this directly as we did for rows, or by taking transposes and using Theorem 2.22.]

10. Show that an $m \times n$ reduced row echelon matrix R of rank k can be reduced by elementary column operations to an $m \times n$ matrix C of the form

$$C = \begin{bmatrix} 1 & 0 & \cdots & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & 0 & 0 \end{bmatrix}$$

    where the first k entries on the main diagonal are 1's, and the rest are 0's.

11. From the previous problem and Theorem 2.3, show that every $m \times n$ matrix A of rank k can be reduced by elementary row and column operations to the form C. We call the matrix C the **canonical form** of A.

12. We say that a matrix A is **row-column-equivalent** (abbreviated r.c.e.) to a matrix B if A can be transformed into B by a finite number of elementary row and column operations. Prove:

    (a) If A is a matrix, e is an elementary row operation, and e' is an elementary column operation, then $(eA)e' = e(Ae')$.
    (b) r.c.e. is an equivalence relation.
    (c) Two $m \times n$ matrices A and B are r.c.e. if and only if they have the same canonical form, and hence if and only if they have the same rank.

13. If A is any $m \times n$ matrix of rank k, prove there exists a nonsingular $m \times m$ matrix P and a nonsingular $n \times n$ matrix Q such that $PAQ = C$ (the canonical form of A).

14. Prove that two $m \times n$ matrices A and B are r.c.e. if and only if there exists a nonsingular $m \times m$ matrix P and a nonsingular $n \times n$ matrix Q such that $PAQ = B$.

2.8 The *LU* Factorization*

We now show how elementary matrices can be used to "factor" a matrix into the product of a lower triangular matrix times an upper triangular matrix. This factorization can then be used to easily implement the solution to a system of inhomogeneous linear equations. We will first focus our attention on type β (multiply a row by a nonzero scalar) and type γ (add a scalar multiple of one row to another) transformations. Afterwards, we will discuss how to handle the additional complications introduced by the interchange of two rows (the type α transformations).

Before beginning, let us introduce some very common and useful terminology. As we mentioned in the last section, a square matrix $A = (a_{ij}) \in M_n(\mathcal{F})$ is said to be **diagonal** if the only nonzero entries are those on the main diagonal. In other words, $a_{ij} = 0$ if $i \neq j$ and each a_{ii} may or may not be zero. A typical diagonal matrix thus looks like

$$\begin{bmatrix} a_{11} & 0 & \cdots & 0 \\ 0 & a_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{nn} \end{bmatrix}.$$

Referring to Exercise 2.5.9, a square matrix $A = (a_{ij})$ is said to be **upper triangular** if $a_{ij} = 0$ for $i > j$. In other words, every entry of A below the main diagonal is zero. Similarly, A is said to be **lower triangular** if $a_{ij} = 0$ for $i < j$. Thus a general lower triangular matrix is of the form

$$\begin{bmatrix} a_{11} & 0 & 0 & \cdots & 0 \\ a_{21} & a_{22} & 0 & \cdots & 0 \\ a_{31} & a_{32} & a_{33} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \cdots & a_{nn} \end{bmatrix}.$$

And in that exercise, you were asked to prove that the product of lower (upper) triangular matrices is a lower (upper) triangular matrix.

In fact, it is easy to see that the inverse of a nonsingular lower triangular matrix L is also lower triangular. This is because the fact that L is nonsingular means that it can be row-reduced to the identity matrix, and following the method described in Section 2.6, it should be clear from the form of L that the elementary row operations required to accomplish this will transform the identity matrix I into lower triangular form. Of course, this also applies equally to upper triangular matrices.

Now, if we apply a type β transformation e_β to the identity matrix I, then what we have is a diagonal matrix of the form

$$E_\beta = e_\beta(I) = \begin{bmatrix} 1 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & & \vdots \\ 0 & \cdots & k & \cdots & 0 \\ \vdots & & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & 1 \end{bmatrix}.$$

And if we apply a type γ transformation to the identity matrix by adding a scalar multiple of row i to row j where $j > i$, then we obtain a lower triangular matrix of the form

$$E_\gamma = e_\gamma(I) = \begin{bmatrix} 1 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & & \vdots & & \vdots \\ 0 & \cdots & 1 & \cdots & 0 & \cdots & 0 \\ \vdots & & \vdots & \ddots & \vdots & & \vdots \\ 0 & \cdots & k & \cdots & 1 & \cdots & 0 \\ \vdots & & \vdots & & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & 0 & \cdots & 1 \end{bmatrix}.$$

A moments thought should make it clear that multiplying a diagonal matrix times a lower triangular matrix in either order always results in a lower triangular matrix. So *if* we can reduce a matrix to row echelon form *without requiring any row interchanges*, then the row echelon form U will be a product $U = E_r \cdots E_1 A$ of elementary matrices acting on A, and the product $E_r \cdots E_1$ will be lower triangular. (An elementary matrix that results from a type α transformation is not lower triangular.) Since U is by definition upper triangular, we see that writing $A = E_1^{-1} \cdots E_r^{-1} U$ shows that A has been written as the product of a lower triangular matrix $E_1^{-1} \cdots E_r^{-1} = (E_r \cdots E_1)^{-1}$ times an upper triangular matrix U.

Actually, the type β transformations are not needed at all. If we start from the first row of the matrix A and subtract multiples of this row from the remaining rows (i.e., use a_{11} as a pivot), then we can put A into the form that has all 0's in the first column below a_{11}. Next, we use a_{22} as a pivot to make all of the entries in the second column below a_{22} equal to zero. Continuing this procedure, we eventually arrive at an (upper triangular) row echelon form for A by using only type γ transformations. But a product of type γ elementary matrices is just a lower triangular matrix that has all 1's on the diagonal. Such a lower triangular matrix is called **special lower triangular** (or sometimes **unit lower triangular**). Therefore we may write $A = LU$ where L is special lower triangular (the product of elementary matrices) and U is upper triangular (the row echelon form of A). This is called the ***LU* factorization** of A. Note that if any pivot turns out to be zero, then we would have to interchange rows to continue the process, and this approach fails. We will treat this case a little later.

In summary, *if we can reduce a matrix A to row echelon form without having to interchange any rows, then A has an LU factorization.*

Example 2.13. Let us find the LU factorization of

$$A = \begin{bmatrix} 1 & -3 & 2 \\ -3 & 7 & -5 \\ 2 & -5 & 8 \end{bmatrix}.$$

We have the sequence

$$\begin{aligned} A = \begin{bmatrix} 1 & -3 & 2 \\ -3 & 7 & -5 \\ 2 & -5 & 8 \end{bmatrix} &\xrightarrow[A_3-2A_1]{A_2+3A_1} \begin{bmatrix} 1 & -3 & 2 \\ 0 & -2 & 1 \\ 0 & 1 & 4 \end{bmatrix} \\ &\xrightarrow{A_3+(1/2)A_2} \begin{bmatrix} 1 & -3 & 2 \\ 0 & -2 & 1 \\ 0 & 0 & 9/2 \end{bmatrix} = U. \end{aligned}$$

The corresponding elementary matrices are (in order)

$$E_1 = \begin{bmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad E_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -2 & 0 & 1 \end{bmatrix} \qquad E_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1/2 & 1 \end{bmatrix}$$

and these have inverses

$$E_1^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ -3 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad E_2^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 2 & 0 & 1 \end{bmatrix} \qquad E_3^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -1/2 & 1 \end{bmatrix}.$$

Then $E_3E_2E_1A = U$ so we should have $E_1^{-1}E_2^{-1}E_3^{-1}U = LU = A$. And indeed we do see that

$$\begin{aligned} E_1^{-1}E_2^{-1}E_3^{-1}U &= \begin{bmatrix} 1 & 0 & 0 \\ -3 & 1 & 0 \\ 2 & -1/2 & 1 \end{bmatrix} \begin{bmatrix} 1 & -3 & 2 \\ 0 & -2 & 1 \\ 0 & 0 & 9/2 \end{bmatrix} \\ &= \begin{bmatrix} 1 & -3 & 2 \\ -3 & 7 & -5 \\ 2 & -5 & 8 \end{bmatrix} = A \end{aligned}$$

In other words, we have factored A into the form $A = LU$ as claimed.

Next, we can get even more clever and save some work by making the following observation based on the previous example. We started with the $(1, 1)$ entry of A as the pivot, and *subtracted* a scalar multiple of the first row from each of the rows below. Let k_i be the scalar multiple required to get a 0 in the $(i, 1)$ position. The corresponding *inverse* elementary matrix is then just the identity matrix with k_i in the $(i, 1)$ position. Since we do this once for each row, the product of all the inverse elementary matrices has as its first column below its $(1, 1)$ entry just the scalars $k_2, \dots, k_n$.

Now we go to the second row of A and subtract scalar multiples of a_{22} from each of the rows below. Again, the corresponding inverse elementary matrix is the identity matrix but with its $(i, 2)$ entry just the scalar that was required to get a 0 into the $(i, 2)$ position of A, and the product of these inverse elementary matrices has the required scalars as its second column below the $(2, 2)$ entry.

Continuing this procedure, we see that the required L matrix is just the identity matrix with columns below the main diagonal that are made up of the scalar multiples that were required to put the original matrix into row echelon form. This gives us a quick way to write down L without computing the individual elementary matrices.

Example 2.14. Let us find the LU factorization of the matrix

$$A = \begin{bmatrix} 3 & 1 & 3 & -4 \\ 6 & 4 & 8 & -10 \\ 3 & 2 & 5 & -1 \\ -9 & 5 & -2 & -4 \end{bmatrix}.$$

To reduce this to row echelon form we have the sequence

$$A = \begin{bmatrix} 3 & 1 & 3 & -4 \\ 6 & 4 & 8 & -10 \\ 3 & 2 & 5 & -1 \\ -9 & 5 & -2 & -4 \end{bmatrix} \xrightarrow{\substack{A_2 - 2A_1 \\ A_3 - A_1 \\ A_4 - (-3)A_1}} \begin{bmatrix} 3 & 1 & 3 & -4 \\ 0 & 2 & 2 & -2 \\ 0 & 1 & 2 & 3 \\ 0 & 8 & 7 & -16 \end{bmatrix}$$

$$\xrightarrow{\substack{A_3 - (1/2)A_2 \\ A_4 - 4A_2}} \begin{bmatrix} 3 & 1 & 3 & -4 \\ 0 & 2 & 2 & -2 \\ 0 & 0 & 1 & 4 \\ 0 & 0 & -1 & 4 \end{bmatrix}$$

$$\xrightarrow{A_4 - (-1)A_3} \begin{bmatrix} 3 & 1 & 3 & -4 \\ 0 & 2 & 2 & -2 \\ 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & -4 \end{bmatrix} = U.$$

For column 1 we have the scalar multiples $2, 1$ and -3; for column 2 we have $1/2$ and 4; and for column 3 we have -1. So the matrix L in this case is

$$L = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ 1 & 1/2 & 1 & 0 \\ -3 & 4 & -1 & 1 \end{bmatrix}$$

and

$$\begin{aligned} LU &= \begin{bmatrix} 1 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ 1 & 1/2 & 1 & 0 \\ -3 & 4 & -1 & 1 \end{bmatrix} \begin{bmatrix} 3 & 1 & 3 & -4 \\ 0 & 2 & 2 & -2 \\ 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & -4 \end{bmatrix} \\ &= \begin{bmatrix} 3 & 1 & 3 & -4 \\ 6 & 4 & 8 & -10 \\ 3 & 2 & 5 & -1 \\ -9 & 5 & -2 & -4 \end{bmatrix} = A. \end{aligned}$$

Before giving an example of how the LU factorization can be used to solve systems of linear equations, let us make some remarks.

First of all, it is straightforward to generalize our definition of LU factorization to include matrices that are not square. All we have to do is require that the U matrix be in row echelon form. The rest of the process remains unchanged.

Second, we still have to take care of the situation where some rows of A must be interchanged in order to achieve the row echelon form. We will address this shortly.

And third, while in general the row echelon form of a matrix is not unique, if A is nonsingular, then in fact the LU factorization is unique. To see this, suppose we have two different factorizations $A = LU$ and $A = L'U'$ where U and U' are two row echelon forms of A. We first observe that L and L' (or any special lower triangular matrix for that matter) are necessarily invertible. This is because the diagonal entries of L are all equal to 1, so by subtracting suitable multiples of each row from the rows below it, we can row reduce L to the identity matrix (which is nonsingular).

By assumption, A is nonsingular too, so both $U = L^{-1}A$ and $U' = L'^{-1}A$ must be nonsingular (by Theorem 2.20, Corollary 3). Then using $LU = L'U'$ we may write

$$L'^{-1}(LU)U^{-1} = L'^{-1}(L'U')U^{-1}$$

or

$$(L'^{-1}L)(UU^{-1}) = (L'^{-1}L')(U'U^{-1})$$

which then implies that $L'^{-1}L = U'U^{-1}$. But the left side of this equation is a product of lower triangular matrices and hence is also lower triangular, while the right side is similarly upper triangular. Since the only matrix that is both special lower triangular and upper triangular is the identity matrix (think about it), it follows that ${L'}^{-1}L = I$ and $U'U^{-1} = I$ so that $L = L'$ and $U = U'$ and the factorization is unique. Thus we have proved the following theorem.

Theorem 2.24. *If a nonsingular matrix A has an LU factorization, then the matrices L and U are unique.*

Now let's see how the LU factorization can be used to solve a system of inhomogeneous linear equations. First some more terminology. Recall that we solve an inhomogeneous system $AX = Y$ by reducing to row echelon form. For example, in the case of three equations and three unknowns we might end up with something like

$$\begin{aligned} 2x - 3y + z &= 5 \\ 4y - z &= 2 \\ 3z &= 6. \end{aligned}$$

To solve this, we start at the bottom and find $z = 2$. We then go up one row and use this value for z to find $4y = z + 2 = 4$ so that $y = 1$. Now go up another row and use the previous two results to find $2x = 5+3y-z = 6$ so that $x = 3$. This procedure of starting at the bottom of the row echelon form of an inhomogeneous system of linear equations and working your way up to find the solution to the system is called **back substitution**.

Note that the row echelon form of the equations in the example above can be written as $UX = Y$ where U is upper triangular. Now suppose we have a system that is something like

$$\begin{aligned} x &= 2 \\ 2x - y &= 1 \\ 3x + y - 4z &= 3. \end{aligned}$$

This can be written in the form $LX = Y$ where L is lower triangular. To solve this we start at the top and work our way down. This is referred to as **forward substitution**.

So, if we have the system $AX = Y$, we write A in its factored form $A = LU$ so that our system is $LUX = Y$. If we now define the vector $Z = UX$, then we have the system $LZ = Y$ which can be solved by forward substitution. Then given this solution Z, we can solve $UX = Z$ for X by back substitution.

Example 2.15. Let us solve the system $AX = Y$ where

$$A = \begin{bmatrix} 2 & 1 & 3 \\ 4 & -1 & 3 \\ -2 & 5 & 5 \end{bmatrix} \qquad \text{and} \qquad Y = \begin{bmatrix} 1 \\ -4 \\ 9 \end{bmatrix}.$$

We have the sequence

$$A = \begin{bmatrix} 2 & 1 & 3 \\ 4 & -1 & 3 \\ -2 & 5 & 5 \end{bmatrix} \xrightarrow[A_3-(-1)A_1]{A_2-2A_1} \begin{bmatrix} 2 & 1 & 3 \\ 0 & -3 & -3 \\ 0 & 6 & 8 \end{bmatrix}$$

$$\xrightarrow{A_3-(-2)A_2} \begin{bmatrix} 2 & 1 & 3 \\ 0 & -3 & -3 \\ 0 & 0 & 2 \end{bmatrix}$$

so that

$$A = \begin{bmatrix} 2 & 1 & 3 \\ 4 & -1 & 3 \\ -2 & 5 & 5 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ -1 & -2 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 & 3 \\ 0 & -3 & -3 \\ 0 & 0 & 2 \end{bmatrix} = LU.$$

We first solve $LZ = Y$ for Z as follows. Writing $Z = (z_1, z_2, z_3)$, this system is

$$\begin{aligned} z_1 \qquad\qquad &= 1 \\ 2z_1 + z_2 \qquad &= -4 \\ -z_1 - 2z_2 + z_3 &= 9 \end{aligned}$$

which is easily solved by forward substitution to find $z_1 = 1$, $z_2 = -4 - 2z_1 = -6$ and $z_3 = 9 + z_1 + 2z_2 = -2$.

Now we solve $UX = Z$ for X using back substitution. Write $X = (x_1, x_2, x_3)$ so we have the system

$$\begin{aligned} 2x_1 + x_2 + 3x_3 &= 1 \\ -3x_2 - 3x_3 &= -6 \\ 2x_3 &= -2 \end{aligned}$$

which yields $x_3 = -1$, $-3x_2 = -6 + 3x_3$ or $x_2 = 3$, and $2x_1 = 1 - x_2 - 3x_3$ or $x_1 = 1/2$. In other words, our solution is

$$X = \begin{bmatrix} 1/2 \\ 3 \\ -1 \end{bmatrix}.$$

The last item we still have to address with respect to the LU factorization is what to do if we need to use type α transformations (i.e., interchanging rows) to get our system into row echelon form. This can occur if adding a multiple of one row to another results in a zero pivot (i.e., diagonal entry). Let's take a look at the elementary matrices that arise from a row interchange.

For definiteness, consider a system of four equations. The general result will be quite obvious. If we switch say rows 2 and 3 then we have

$$I = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \xrightarrow[I_3 \to I_2]{I_2 \to I_3} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Note that there is always a single 1 in each row and column, and that the columns are just the standard basis vectors $\{e_i\}$ for $\mathbb{R}^4$, although their order has been changed. (Don't confuse this notation for the basis vectors with the similar notation for an elementary row operation.) We say that the ordered set of basis vectors $\{e_1, e_3, e_2, e_4\}$ is a **permutation** of the ordered set $\{e_1, e_2, e_3, e_4\}$. In other words, a permutation us just a rearrangement of a set that was originally in some standard order. We will denote such an elementary matrix by the letter P and refer to it as a **permutation matrix**.

If we have two permutation matrices, for example

$$P_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \qquad \text{and} \qquad P_2 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

then we can form the product

$$P = P_1 P_2 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

which is also another permutation matrix. Since the ith row of P^T is the same as the ith column of P, and each is just one of the standard basis vectors e_k, we see that the (i,j)th entry of the product $P^T P$ is given by (see Example 1.13)

$$(P^T P)_{ij} = \sum_k p^T_{ik} p_{kj} = \sum_k p_{ki} p_{kj} = \langle e_r, e_s \rangle = \delta_{rs}$$

where e_r is the basis vector that is the ith column of P, and e_s is the basis vector that is the jth column of P. In other words, $P^T P = I$ and we have proved

Theorem 2.25. *For any permutation matrix P we have $P^{-1} = P^T$.*

Now, if we attempt the LU factorization of a matrix A and find at some point that we obtain a zero as pivot, then we need to interchange two rows in order to proceed and the factorization fails. However, with hindsight we could have performed that interchange first, and then proceeded with the reduction.

Let us denote the combination (or product) of all required interchanges (or permutations) by P. While the original matrix A does not have an LU factorization because of the required interchanges, the permuted matrix PA does have such an LU factorization by its construction. We therefore have $PA = LU$ or $A = P^{-1}LU = P^T LU$. This is called the $\boldsymbol{P^T LU}$ **factorization** (or **permuted** $\boldsymbol{LU}$ **factorization**) of A.

Note also that while the interchanges performed to achieve the LU factorization are not unique — in general you have a choice of which rows to switch — once that choice has been made and the LU factorization exists, it is unique. We therefore have the following result.

Theorem 2.26. *Every square matrix $A \in M_n(\mathcal{F})$ has a $P^T LU$ factorization.*

One of the main advantages of the $P^T LU$ factorization is its efficiency. Since each column of L consists of the scalars that were needed to obtain zeros in the corresponding positions of A (as illustrated in Example 2.14), this means that we can construct L as we perform the type γ operations. Similarly, if we need to perform a type α interchange, we apply it to the identity matrix I. In addition, all subsequent interchanges get applied to the previous result and this gives us the matrix P. However, we must also apply the interchange to the corresponding *column* of L. This is because had the interchange been done *before* the type γ operation, only that same column of L would be different. And any subsequent interchanges must also be done to *all previous columns* because it is as if the interchanges were performed before we started with the type γ operations. Our next example illustrates the technique.

Example 2.16. Let us find the $P^T LU$ factorization of the matrix

$$A = \begin{bmatrix} 1 & 2 & -1 & 0 \\ 2 & 4 & -2 & -1 \\ -3 & -5 & 6 & 1 \\ -1 & 2 & 8 & -2 \end{bmatrix}.$$

Performing the type γ operations on the first column we have

$$A \to \begin{bmatrix} 1 & 2 & -1 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 1 & 3 & 1 \\ 0 & 4 & 7 & -2 \end{bmatrix} \quad L = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ -3 & 0 & 1 & 0 \\ -1 & 0 & 0 & 1 \end{bmatrix} \quad P = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Since the $(2,2)$ entry of A is now 0, we interchange rows 2 and 3 to obtain

$$A \to \begin{bmatrix} 1 & 2 & -1 & 0 \\ 0 & 1 & 3 & 1 \\ 0 & 0 & 0 & -1 \\ 0 & 4 & 7 & -2 \end{bmatrix} \quad L = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -3 & 1 & 0 & 0 \\ 2 & 0 & 1 & 0 \\ -1 & 0 & 0 & 1 \end{bmatrix} \quad P = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Note that the interchange was only applied to the first column of L.

We now use the $(2,2)$ entry of A as pivot to obtain

$$A \to \begin{bmatrix} 1 & 2 & -1 & 0 \\ 0 & 1 & 3 & 1 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & -5 & -6 \end{bmatrix} \quad L = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -3 & 1 & 0 & 0 \\ 2 & 0 & 1 & 0 \\ -1 & 4 & 0 & 1 \end{bmatrix} \quad P = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Next we have to perform another interchange to get a nonzero entry in the $(3,3)$ position and we are done:

$$U = \begin{bmatrix} 1 & 2 & -1 & 0 \\ 0 & 1 & 3 & 1 \\ 0 & 0 & -5 & -6 \\ 0 & 0 & 0 & -1 \end{bmatrix} \quad L = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -3 & 1 & 0 & 0 \\ -1 & 4 & 1 & 0 \\ 2 & 0 & 0 & 1 \end{bmatrix} \quad P = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{bmatrix}.$$

Observe that this last interchange had to be applied to all previous columns of L. This is again because had the interchange been done at the start, the resulting L would have been different.

Finally, we can check that

$$\begin{aligned} PA &= \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 2 & -1 & 0 \\ 2 & 4 & -2 & -1 \\ -3 & -5 & 6 & 1 \\ -1 & 2 & 8 & -2 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 2 & -1 & 0 \\ -3 & -5 & 6 & 1 \\ -1 & 2 & 8 & -2 \\ 2 & 4 & -2 & -1 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 & 0 & 0 \\ -3 & 1 & 0 & 0 \\ -1 & 4 & 1 & 0 \\ 2 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & -1 & 0 \\ 0 & 1 & 3 & 1 \\ 0 & 0 & -5 & -6 \\ 0 & 0 & 0 & -1 \end{bmatrix} = LU. \end{aligned}$$

Exercises

1. Find the LU factorization of the following matrices.

(a) $\begin{bmatrix} 1 & 2 \\ -3 & -1 \end{bmatrix}$ (b) $\begin{bmatrix} 2 & -4 \\ 3 & 1 \end{bmatrix}$ (c) $\begin{bmatrix} 1 & 3 \\ -1 & 0 \end{bmatrix}$ (d) $\begin{bmatrix} 1 & 3 \\ 3 & 1 \end{bmatrix}$

(e) $\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 8 & 7 & 9 \end{bmatrix}$ (f) $\begin{bmatrix} 2 & 2 & -1 \\ 4 & 0 & 4 \\ 3 & 4 & 4 \end{bmatrix}$ (g) $\begin{bmatrix} 1 & 0 & -1 \\ 2 & 3 & 2 \\ -3 & 1 & 0 \end{bmatrix}$

(h) $\begin{bmatrix} 1 & 2 & 3 & -1 \\ 2 & 6 & 3 & 0 \\ 0 & 6 & -6 & 7 \\ -1 & -2 & -9 & 0 \end{bmatrix}$ (i) $\begin{bmatrix} 2 & 2 & 2 & 1 \\ -2 & 4 & -1 & 2 \\ 4 & 4 & 7 & 3 \\ 6 & 9 & 5 & 8 \end{bmatrix}$

(j) $\begin{bmatrix} 1 & 1 & -2 & 3 \\ -1 & 2 & 3 & 0 \\ -2 & 1 & 1 & -2 \\ 3 & 0 & -1 & 5 \end{bmatrix}$ (k) $\begin{bmatrix} 2 & 1 & 3 & 1 \\ 1 & 4 & 0 & 1 \\ 3 & 0 & 2 & 2 \\ 1 & 1 & 2 & 2 \end{bmatrix}$

2. Using the generalized definition mentioned in the text to find the LU factorization of the following matrices.

(a) $\begin{bmatrix} 1 & 0 & 1 & -2 \\ 0 & 3 & 3 & 1 \\ 0 & 0 & 0 & 5 \end{bmatrix}$ (b) $\begin{bmatrix} 1 & 2 & 0 & -1 & 1 \\ -2 & -7 & 3 & 8 & -2 \\ 1 & 1 & 3 & 5 & 2 \\ 0 & 3 & -3 & -6 & 0 \end{bmatrix}$

3. Find a $P^T LU$ factorization for each of the following matrices.

(a) $\begin{bmatrix} 0 & 1 & 4 \\ -1 & 2 & 1 \\ 1 & 3 & 3 \end{bmatrix}$ (b) $\begin{bmatrix} 0 & 0 & 1 & 2 \\ -1 & 1 & 3 & 2 \\ 0 & 2 & 1 & 1 \\ 1 & 1 & -1 & 0 \end{bmatrix}$ (c) $\begin{bmatrix} 0 & -1 & 1 & 3 \\ -1 & 1 & 1 & 2 \\ 0 & 1 & -1 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}$

4. For each of the following systems of equations, find a $P^T LU$ factorization for the matrix of coefficients, and then use forward and back substitution to solve the system.

(a) $\begin{aligned} 4x - 4y + 2z &= 1 \\ -3x + 3y + z &= 3 \\ -3x + y - 2z &= -5 \end{aligned}$ (b) $\begin{aligned} y - z + w &= 0 \\ y + z &= 1 \\ x - y + z - 3w &= 2 \\ x + 2y - z + w &= 4 \end{aligned}$

(c) $\begin{aligned} x - y + 2z + w &= 0 \\ -x + y - 3z &= 1 \\ x - y + 4z - 3w &= 2 \\ x + 2y - z + w &= 4 \end{aligned}$

5. (a) Let U be an upper triangular square matrix with no zeros on the diagonal. Show that we can always write $U = DV$ where D is a diagonal matrix and V is special upper triangular. If $A = LU = LDV$, then this is called an ***LDV* factorization** of A.
   (b) Find the LDV factorization of each of the following matrices.

   (i) $\begin{bmatrix} 2 & 1 & 1 \\ 4 & 5 & 2 \\ 2 & -2 & 0 \end{bmatrix}$ (ii) $\begin{bmatrix} 2 & -4 & 0 \\ 3 & -1 & 4 \\ -1 & 2 & 2 \end{bmatrix}$

2.9 The QR Factorization*

Another very useful matrix factorization is based on the Gram-Schmidt process (see the corollary to Theorem 1.21). We will show that if $A \in M_{m\times n}(\mathbb{R})$ is such that all $n \le m$ columns are linearly independent, then it is possible to write $A = QR$ where Q has orthonormal columns and R is a nonsingular upper triangular matrix.

To begin with, the assumption that all n columns of A are linearly independent means that they form a basis for $\mathbb{R}^n$. Let us denote these columns by $\{u_1, \dots, u_n\}$, and let $\{w_1, \dots, w_n\}$ be the corresponding orthonormal basis constructed by the Gram-Schmidt process. According to the Gram-Schmidt algorithm, this means that for each $k = 1, \dots, n$ we know that the linear span of $\{u_1, \dots, u_k\}$ is equal to the linear span of $\{w_1, \dots, w_k\}$.

By definition of linear span, we can then write each (column) vector u_i as a linear combination of the (column) vectors $w_1, \dots, w_i$. In particular, there are coefficients $r_{ij} \in \mathbb{R}$ so that

$$\begin{aligned}
u_1 &= w_1 r_{11} \\
u_2 &= w_1 r_{12} + w_2 r_{22} \\
&\ \vdots \\
u_n &= w_1 r_{1n} + w_2 r_{2n} + \cdots + w_n r_{nn}.
\end{aligned}$$

Since u_i is the ith column of A, let us use a shorthand notation and write a matrix as a function of its columns. Then these equations can be written in matrix form as

$$\begin{aligned}
A &= \begin{bmatrix} u_1 & u_2 & \cdots & u_n \end{bmatrix} \\
&= \begin{bmatrix} w_1 & w_2 & \cdots & w_n \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ 0 & r_{22} & \cdots & r_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & r_{nn} \end{bmatrix} := QR
\end{aligned}$$

By construction, the matrix Q has orthonormal columns (the w_i as determined by the Gram-Schmidt process). Furthermore, it is also a consequence of the Gram-Schmidt process that all of the diagonal entries r_{ii} are nonzero. For if $r_{ii} = 0$, then $u_i = w_1 r_{1i} + \cdots + w_{i-1} r_{i-1\, i}$ which implies that u_i is a linear combination of $u_1, \dots, u_{i-1}$ (since each w_k is a linear combination of $u_1, \dots, u_k$). Since this contradicts the fact that the u_i are all linearly independent, it must be that all of the r_{ii} are nonzero.

Lastly, it follows from the fact that R is upper triangular and none of its diagonal entries are zero, that $\operatorname{rank}(R) = n$ so that R is nonsingular. This is the ***QR* factorization**, and we have now proved the following.

Theorem 2.27. *If $A \in M_{m\times n}(\mathbb{R})$ has $n \leq m$ linearly independent columns, then we can write $A = QR$ where $Q \in M_{m\times n}(\mathbb{R})$ has n linearly independent columns, and $R \in M_n(\mathbb{R})$ is a nonsingular upper triangular matrix.*

We also point out that the fact that Q has orthonormal columns means that $Q^T Q = I$. Using this, we multiply $A = QR$ from the left by Q^T to obtain $Q^T A = R$, and this gives us a practical method of finding R given A and Q.

Let us remark however, that as we will see in a later chapter, a *square* matrix A such that $A^T A = I$ is said to be **orthogonal**. An orthogonal matrix has the property that its rows and columns each form an orthonormal set, and $A^T = A^{-1}$. However, in our discussion of the QR factorization, Q is not orthogonal unless it also happens to be square.

Example 2.17. Let us find the QR factorization of the matrix

$$A = \begin{bmatrix} 3 & -1 & 2 \\ 0 & 0 & 9 \\ 4 & 7 & 11 \end{bmatrix}.$$

Noting that the columns of A are the vectors u_i used in Example 1.15, we have the orthonormal basis

$$w_1 = \begin{bmatrix} 3/5 \\ 0 \\ 4/5 \end{bmatrix} \qquad w_2 = \begin{bmatrix} -4/5 \\ 0 \\ 3/5 \end{bmatrix} \qquad w_3 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

so that

$$Q = \begin{bmatrix} 3/5 & -4/5 & 0 \\ 0 & 0 & 1 \\ 4/5 & 3/5 & 0 \end{bmatrix}.$$

To find R we use the fact that $Q^T Q = I$ to write $R = Q^T A$ so that

$$\begin{aligned} R = Q^T A &= \begin{bmatrix} 3/5 & 0 & 4/5 \\ -4/5 & 0 & 3/5 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 3 & -1 & 2 \\ 0 & 0 & 9 \\ 4 & 7 & 11 \end{bmatrix} \\ &= \begin{bmatrix} 5 & 5 & 10 \\ 0 & 5 & 5 \\ 0 & 0 & 9 \end{bmatrix}. \end{aligned}$$

We easily verify that

$$QR = \begin{bmatrix} 3/5 & -4/5 & 0 \\ 0 & 0 & 1 \\ 4/5 & 3/5 & 0 \end{bmatrix} \begin{bmatrix} 5 & 5 & 10 \\ 0 & 5 & 5 \\ 0 & 0 & 9 \end{bmatrix} = \begin{bmatrix} 3 & -1 & 2 \\ 0 & 0 & 9 \\ 4 & 7 & 11 \end{bmatrix} = A.$$

Exercises

1. For each of the following matrices, use the Gram-Schmidt process to find an orthonormal basis for the column space, and find the QR factorization of the matrix.

$$\text{(a)} \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix} \qquad \text{(b)} \begin{bmatrix} 2 & 1 & -1 \\ 0 & 1 & 3 \\ -1 & -1 & 1 \end{bmatrix} \qquad \text{(c)} \begin{bmatrix} 0 & 0 & 2 \\ 0 & 4 & 1 \\ -1 & 0 & 1 \end{bmatrix}$$

$$\text{(d)} \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 2 \\ -1 & 1 & 0 \\ 1 & 5 & 1 \end{bmatrix} \qquad \text{(e)} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 1 & 0 \\ 1 & 1 & 2 & 1 \\ 1 & 0 & 1 & 1 \end{bmatrix}$$

2. (a) Prove that an upper triangular matrix U is orthogonal if and only if it is a diagonal matrix.
   (b) Prove that the QR factorization of a matrix is unique if all of the diagonal entries of R are positive.

3. If we have a system of linear equations $AX = Y$ where $A = QR$, then $RX = Q^TY$. Since R is upper triangular, it is easy to solve for X using back substitution. Use this approach to solve the following systems of equations.

$$\text{(a)} \begin{bmatrix} 1 & 2 \\ -1 & 3 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} -1 \\ 2 \end{bmatrix} \qquad \text{(b)} \begin{bmatrix} 2 & 1 & -1 \\ 1 & 0 & 2 \\ 2 & -1 & 3 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix}$$

$$\text{(c)} \begin{bmatrix} 1 & 1 & 0 \\ -1 & 0 & 1 \\ 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

Chapter 3

Determinants

Suppose we want to solve the system of equations

$$\begin{aligned} ax + by &= f \\ cx + dy &= g \end{aligned}$$

where $a, b, c, d, f, g \in \mathcal{F}$. It is easily verified that if we reduce the augmented matrix to reduced row-echelon form we obtain

$$\begin{bmatrix} 1 & 0 & (fd - gb)/\Delta \\ 0 & 1 & (ag - cf)/\Delta \end{bmatrix}$$

where $\Delta = ad - cb$. We must therefore have $\Delta \neq 0$ if a solution is to exist for every choice of f and g. If $A \in M_2(\mathcal{F})$ is the matrix of coefficients of our system, we call the number Δ the **determinant** of A, and write this as $\det A$. While it is possible to proceed from this point and define the determinant of larger matrices by induction, we prefer to take another more useful approach in developing the general theory.

It is unfortunate that many authors nowadays seem to feel that determinants are no longer as important as they once were, and as a consequence they are treated only briefly if at all. While it may be true that in and of themselves determinants aren't as important, it is our strong opinion that by developing the theory of determinants in a more sophisticated manner, we introduce the student to some very powerful tools that are of immense use in many other current fields of mathematical and physical research that utilize the techniques of differential geometry.

While the techniques and notation that we are about to introduce can seem too abstract and intimidating to the beginning student, it really doesn't take long to gain familiarity and proficiency with the method, and it quickly becomes second nature well worth the effort it takes to learn.

3.1 The Levi-Civita Symbol

In order to ease into the notation we will use, we begin with an elementary treatment of the vector cross product. This will give us a very useful computational tool that is of importance in and of itself. While the reader is probably already familiar with the cross product, we will still go through its development from scratch just for the sake of completeness.

To begin with, consider two vectors $\vec{A}$ and $\vec{B}$ in $\mathbb{R}^3$ (with Cartesian coordinates). There are two ways to define their **vector product** (or **cross product**) $\vec{A} \times \vec{B}$. The first way is to *define* $\vec{A} \times \vec{B}$ as that vector with norm given by

$$\|\vec{A} \times \vec{B}\| = \|\vec{A}\| \, \|\vec{B}\| \sin\theta$$

where θ is the angle between $\vec{A}$ and $\vec{B}$, and whose direction is such that the triple $(\vec{A}, \vec{B}, \vec{A} \times \vec{B})$ has the same "orientation" as the standard basis vectors $(\hat{x}, \hat{y}, \hat{z})$. This is commonly referred to as "the right hand rule." In other words, if you rotate $\vec{A}$ into $\vec{B}$ through the smallest angle between them with your right hand as if you were using a screwdriver, then the screwdriver points in the direction of $\vec{A} \times \vec{B}$. Note that by definition, $\vec{A} \times \vec{B}$ is perpendicular to the plane spanned by $\vec{A}$ and $\vec{B}$.

The second way to define $\vec{A} \times \vec{B}$ is in terms of its vector components. I will start from this definition and show that it is in fact equivalent to the first definition. So, we *define* $\vec{A} \times \vec{B}$ to be the vector $\vec{C}$ with components

$$\begin{aligned}
C_x &= (\vec{A} \times \vec{B})_x = A_y B_z - A_z B_y \\
C_y &= (\vec{A} \times \vec{B})_y = A_z B_x - A_x B_z \\
C_z &= (\vec{A} \times \vec{B})_z = A_x B_y - A_y B_x
\end{aligned}$$

Before proceeding, note that instead of labeling components by (x, y, z) it will be very convenient for us to use (x_1, x_2, x_3). This is standard practice, and it will greatly facilitate many equations throughout the remainder of this text. (Besides, it's what we have been doing ever since Example **??**.) Using this notation, the above equations are written

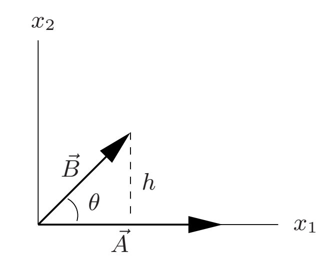
$$\begin{aligned}
C_1 &= (\vec{A} \times \vec{B})_1 = A_2 B_3 - A_3 B_2 \\
C_2 &= (\vec{A} \times \vec{B})_2 = A_3 B_1 - A_1 B_3 \\
C_3 &= (\vec{A} \times \vec{B})_3 = A_1 B_2 - A_2 B_1
\end{aligned}$$

We now see that each equation can be obtained from the previous by cyclically permuting the subscripts $1 \to 2 \to 3 \to 1$.

Using these equations, it is easy to multiply out components and verify that $\vec{A} \cdot \vec{C} = A_1 C_1 + A_2 C_2 + A_3 C_3 = 0$, and similarly $\vec{B} \cdot \vec{C} = 0$. This shows that $\vec{A} \times \vec{B}$ is perpendicular to both $\vec{A}$ and $\vec{B}$, in agreement with our first definition.

Next, there are two ways to show that $\|\vec{A} \times \vec{B}\|$ is also the same as in the first definition. The easy way is to note that any two vectors $\vec{A}$ and $\vec{B}$ in $\mathbb{R}^3$

(both based at the same origin) define a plane. So we choose our coordinate axes so that $\vec{A}$ lies along the x_1-axis as shown below.



Then $\vec{A}$ and $\vec{B}$ have components $\vec{A} = (A_1, 0, 0)$ and $\vec{B} = (B_1, B_2, 0)$ so that

$$\begin{aligned}(\vec{A}\times\vec{B})_1 &= A_2B_3 - A_3B_2 = 0\\ (\vec{A}\times\vec{B})_2 &= A_3B_1 - A_1B_3 = 0\\ (\vec{A}\times\vec{B})_3 &= A_1B_2 - A_2B_1 = A_1B_2\end{aligned}$$

and therefore $\vec{C} = \vec{A}\times\vec{B} = (0, 0, A_1B_2)$. But $A_1 = \|\vec{A}\|$ and $B_2 = h = \|\vec{B}\|\sin\theta$ so that $\|\vec{C}\|^2 = \sum_{i=1}^3 {C_i}^2 = (A_1B_2)^2 = (\|\vec{A}\|\,\|\vec{B}\|\sin\theta)^2$ and therefore

$$\|\vec{A}\times\vec{B}\| = \|\vec{A}\|\,\|\vec{B}\|\sin\theta.$$

Since both the length of a vector and the angle between two vectors is independent of the orientation of the coordinate axes, this result holds for arbitrary $\vec{A}$ and $\vec{B}$ (see also the discussion after Lemma 1.3). Therefore $\|\vec{A}\times\vec{B}\|$ is the same as in our first definition.

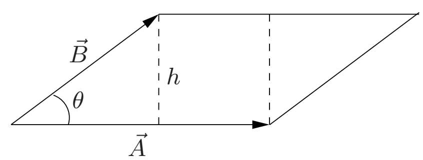
The second way to see this is with a very unenlightening brute force calculation:

$$\begin{aligned}\|\vec{A}\times\vec{B}\|^2 &= (\vec{A}\times\vec{B})\cdot(\vec{A}\times\vec{B}) = {(\vec{A}\times\vec{B})_1}^2 + {(\vec{A}\times\vec{B})_2}^2 + {(\vec{A}\times\vec{B})_3}^2\\
&= (A_2B_3 - A_3B_2)^2 + (A_3B_1 - A_1B_3)^2 + (A_1B_2 - A_2B_1)^2\\
&= {A_2}^2{B_3}^2 + {A_3}^2{B_2}^2 + {A_3}^2{B_1}^2 + {A_1}^2{B_3}^2 + {A_1}^2{B_2}^2 + {A_2}^2{B_1}^2\\
&\qquad - 2(A_2B_3A_3B_2 + A_3B_1A_1B_3 + A_1B_2A_2B_1)\\
&= ({A_2}^2 + {A_3}^2){B_1}^2 + ({A_1}^2 + {A_3}^2){B_2}^2 + ({A_1}^2 + {A_2}^2){B_3}^2\\
&\qquad - 2(A_2B_2A_3B_3 + A_1B_1A_3B_3 + A_1B_1A_2B_2)\\
&= (\text{add and subtract terms})\\
&= ({A_1}^2 + {A_2}^2 + {A_3}^2){B_1}^2 + ({A_1}^2 + {A_2}^2 + {A_3}^2){B_2}^2\\
&\quad + ({A_1}^2 + {A_2}^2 + {A_3}^2){B_3}^2 - ({A_1}^2{B_1}^2 + {A_2}^2{B_2}^2 + {A_3}^2{B_3}^2)\\
&\qquad - 2(A_2B_2A_3B_3 + A_1B_1A_3B_3 + A_1B_1A_2B_2)\end{aligned}$$

$$\begin{aligned}
&= ({A_1}^2 + {A_2}^2 + {A_3}^2)({B_1}^2 + {B_2}^2 + {B_3}^2) - (A_1 B_1 + A_2 B_2 + A_3 B_3)^2 \\
&= \|\vec{A}\|^2\,\|\vec{B}\|^2 - (\vec{A}\cdot\vec{B})^2 = \|\vec{A}\|^2\,\|\vec{B}\|^2 - \|\vec{A}\|^2\,\|\vec{B}\|^2\cos^2\theta \\
&= \|\vec{A}\|^2\,\|\vec{B}\|^2(1-\cos^2\theta) = \|\vec{A}\|^2\,\|\vec{B}\|^2\sin^2\theta
\end{aligned}$$

so again we have $\|\vec{A}\times\vec{B}\| = \|\vec{A}\|\,\|\vec{B}\|\sin\theta$.

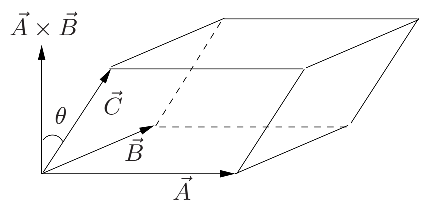
To see the geometrical meaning of the vector product, first take a look at the parallelogram with sides defined by $\vec{A}$ and $\vec{B}$.



In the figure, the height h is equal to $B\sin\theta$ (where $A = \|\vec{A}\|$ and similarly for B), and the area of the parallelogram is equal to the area of the two triangles plus the area of the rectangle:

$$\begin{aligned}
\text{area} &= 2\cdot\frac{1}{2}(B\cos\theta)h + (A - B\cos\theta)h \\
&= Ah = AB\sin\theta = \|\vec{A}\times\vec{B}\|.
\end{aligned}$$

Now suppose we have a third vector $\vec{C}$ that is not coplanar with $\vec{A}$ and $\vec{B}$, and consider the parallelepiped defined by the three vectors as shown below.



The volume of this parallelepiped is given by the area of the base times the height, and hence is equal to

$$\text{volume} = \|\vec{A}\times\vec{B}\|\,\|\vec{C}\|\cos\theta = (\vec{A}\times\vec{B})\cdot\vec{C}.$$

So we see that the so-called **scalar triple product** $(\vec{A}\times\vec{B})\cdot\vec{C}$ represents the volume spanned by the three vectors. In Sections 8.5 and 8.6 we will generalize this result to $\mathbb{R}^n$.

Most of our discussion so far should be familiar to most readers. Now we turn to a formalism that is probably not so familiar. While this is a text on linear algebra and not vector analysis, our formulation of determinants will use a generalization of the permutation symbol that we now introduce. Just keep in mind that the long term benefits of what we are about to do far outweigh the effort required to learn it.

While the concept of permutation should be fairly intuitive, let us make some rather informal definitions. If we have a *set* of n numbers $\{a_1, a_2, \dots, a_n\}$ then, as mentioned in Section 1.3, these n numbers can be arranged into $n!$ *ordered* collections $(a_{i_1}, a_{i_2}, \dots, a_{i_n})$ where $(i_1, i_2, \dots, i_n)$ is just the set $(1, 2, \dots, n)$ arranged in any one of the $n!$ possible orderings. Such an arrangement is called a **permutation** of the set $\{a_1, a_2, \dots, a_n\}$. If we have a set S of n numbers, then we denote the set of all permutations of these numbers by S_n. This is called the **permutation group** of **order** n. Because there are $n!$ rearrangements (i.e., distinct orderings) of a set of n numbers (this can really be any n objects), the permutation group of order n consists of $n!$ elements. It is conventional to denote an element of S_n (i.e., a particular permutation) by Greek letters such as σ, τ, θ etc.

Now, it is fairly obvious intuitively that any permutation can be achieved by a suitable number of interchanges of pairs of elements. Each interchange of a pair is called a **transposition**. (The formal proof of this assertion is, however, more difficult than you might think.) For example, let the ordered set $(1, 2, 3, 4)$ be permuted to the ordered set $(4, 2, 1, 3)$. This can be accomplished as a sequence of transpositions as follows:

$$(1, 2, 3, 4) \xrightarrow{1 \leftrightarrow 4} (4, 2, 3, 1) \xrightarrow{1 \leftrightarrow 3} (4, 2, 1, 3).$$

It is also easy enough to find a different sequence that yields the same final result, and hence the sequence of transpositions resulting in a given permutation is by no means unique. However, it is a fact (also not easy to prove formally) that whatever sequence you choose, the number of transpositions is either always an even number or always an odd number. In particular, if a permutation σ consists of m transpositions, then we define the **sign** of the permutation by

$$\operatorname{sgn}\sigma = (-1)^m.$$

Because of this, it makes sense to talk about a permutation as being either **even** (if m is even) or **odd** (if m is odd).

Now that we have a feeling for what it means to talk about an even or an odd permutation, let us define the **Levi-Civita symbol** ε_{ijk} (also frequently referred to as the **permutation symbol**) by

$$\varepsilon_{ijk} = \begin{cases} 1 & \text{if } (i,j,k) \text{ is an } \textit{even} \text{ permutation of } (1,2,3) \\ -1 & \text{if } (i,j,k) \text{ is an } \textit{odd} \text{ permutation of } (1,2,3) \\ 0 & \text{if } (i,j,k) \text{ is not a permutation of } (1,2,3) \end{cases} .$$

In other words,

$$\varepsilon_{123} = -\varepsilon_{132} = \varepsilon_{312} = -\varepsilon_{321} = \varepsilon_{231} = -\varepsilon_{213} = 1$$

and $\varepsilon_{ijk} = 0$ if there are any repeated indices. We also say that ε_{ijk} is **antisymmetric** in all three indices, meaning that it changes sign upon interchanging any two indices. For a given order (i, j, k) the resulting number ε_{ijk} is also called the **sign** of the permutation.

Before delving further into some of the properties of the Levi-Civita symbol, let's take a brief look at how it is used. Given two vectors $\vec{A}$ and $\vec{B}$, we can let $i = 1$ and form the double sum $\sum_{j,k=1}^{3} \varepsilon_{1jk} A_j B_k$. Since $\varepsilon_{ijk} = 0$ if any two indices are repeated, the only possible values for j and k are 2 and 3. Then

$$\sum_{j,k=1}^{3} \varepsilon_{1jk} A_j B_k = \varepsilon_{123} A_2 B_3 + \varepsilon_{132} A_3 B_2 = A_2 B_3 - A_3 B_2 = (\vec{A} \times \vec{B})_1.$$

But the components of the cross product are cyclic permutations of each other, and ε_{ijk} doesn't change sign under cyclic permutations, so we have the important general result

$$(\vec{A} \times \vec{B})_i = \sum_{j,k=1}^{3} \varepsilon_{ijk} A_j B_k. \tag{3.1}$$

(A cyclic permutation is one of the form $1 \to 2 \to 3 \to 1$ or $x \to y \to z \to x$.)

Now, in order to handle various vector identities, we need to prove some other properties of the Levi-Civita symbol. The first identity to prove is this:

$$\sum_{i,j,k=1}^{3} \varepsilon_{ijk} \varepsilon_{ijk} = 3! = 6. \tag{3.2}$$

But this is actually easy, because (i, j, k) must all be different, and there are 3! ways to order $(1, 2, 3)$. In other words, there are 3! permutations of $\{1, 2, 3\}$. For every case where all three indices are different, whether ε_{ijk} is $+1$ or -1, we always have $(\varepsilon_{ijk})^2 = +1$, and therefore summing over the 3! possibilities yields the desired result.

Before continuing, it will be extremely convenient to introduce a notational shorthand device, called the **Einstein summation convention**. According to this convention, repeated indices are automatically summed over. (The range of summation is always clear from the context.) Furthermore, in order to make it easier to distinguish just which indices are to be summed, we require that one be a subscript and one a superscript. In this notation, a vector would be written as $\vec{A} = A^i \hat{x}_i = A^i e_i$. In those relatively rare cases where an index might be repeated but is not to be summed over, we will explicitly state that there is to be no sum.

From a 'bookkeeping' standpoint, it is important to keep the placement of any free (i.e., unsummed over) indices the same on both sides of an equation. For example, we would always write something like $A_{ij} B^{jk} = C_i{}^k$ and not $A_{ij} B^{jk} = C_{ik}$. In particular, the ith component of the cross product is written

$$(\vec{A} \times \vec{B})_i = \varepsilon_{ijk} A^j B^k. \tag{3.3}$$

For our present purposes, raising and lowering an index is purely a notational convenience. And in order to maintain the proper index placement, we will frequently move an index up or down as necessary. In a later chapter, when we discuss tensors in general, we will see that there is a technical difference between upper and lower indices, but even then, in the particular case of Cartesian coordinates there is no difference. (This is because the "metric" on $\mathbb{R}^3$ with Cartesian coordinates is just δ_{ij}.) While this may seem quite confusing at first, with a little practice it becomes second nature and results in vastly simplified calculations.

Using this convention, equation (3.2) is simply written $\varepsilon_{ijk}\varepsilon^{ijk} = 6$. This also applies to the Kronecker delta, so that we have expressions like $A^i\delta_i^j = \sum_{i=1}^3 A^i\delta_i^j = A^j$ (where δ_i^j is numerically the same as δ_{ij}). An inhomogeneous system of linear equations would be written as simply $a^i{}_j x^j = y^i$, and the dot product as

$$\vec{A}\cdot\vec{B} = A_iB^i = A^iB_i. \tag{3.4}$$

Note also that indices that are summed over are "dummy indices" meaning, for example, that $A_iB^i = A_kB^k$. This is simply another way of writing $\sum_{i=1}^3 A_iB_i = A_1B_1 + A_2B_2 + A_3B_3 = \sum_{k=1}^3 A_kB_k$.

As we have said, the Levi-Civita symbol greatly simplifies many calculations dealing with vectors. Let's look at some examples.

Example 3.1. Let us take a look at the scalar triple product. We have

$$\begin{aligned}
\vec{A}\cdot(\vec{B}\times\vec{C}) &= A_i(\vec{B}\times\vec{C})^i = A_i\varepsilon^{ijk}B_jC_k \\
&= B_j\varepsilon^{jki}C_kA_i \quad (\text{because } \varepsilon^{ijk} = -\varepsilon^{jik} = +\varepsilon^{jki}) \\
&= B_j(\vec{C}\times\vec{A})^j \\
&= \vec{B}\cdot(\vec{C}\times\vec{A}).
\end{aligned}$$

Note also that this formalism automatically takes into account the anti-symmetry of the cross product:

$$(\vec{C}\times\vec{A})_i = \varepsilon_{ijk}C^jA^k = -\varepsilon_{ikj}C^jA^k = -\varepsilon_{ikj}A^kC^j = -(\vec{A}\times\vec{C})_i.$$

It doesn't get any easier than this.

Of course, this formalism works equally well with vector calculus equations involving the gradient ∇. This is the vector defined by

$$\nabla = \hat{x}\frac{\partial}{\partial x} + \hat{y}\frac{\partial}{\partial y} + \hat{z}\frac{\partial}{\partial z} = \hat{x}_1\frac{\partial}{\partial x_1} + \hat{x}_2\frac{\partial}{\partial x_2} + \hat{x}_3\frac{\partial}{\partial x_3} = e^i\frac{\partial}{\partial x_i}.$$

In fact, it will also be convenient to simplify our notation further by defining $\nabla_i = \partial/\partial x_i = \partial_i$, so that $\nabla = e^i\partial_i$.

Example 3.2. Let us prove the well-known identity $\nabla \cdot (\nabla \times \vec{A}) = 0$. We have

$$\nabla \cdot (\nabla \times \vec{A}) = \nabla_i (\nabla \times \vec{A})^i = \partial_i (\varepsilon^{ijk} \partial_j A_k) = \varepsilon^{ijk} \partial_i \partial_j A_k.$$

But now notice that ε^{ijk} is antisymmetric in i and j (so that $\varepsilon^{ijk} = -\varepsilon^{jik}$), while the product $\partial_i \partial_j$ is symmetric in i and j (because we assume that the order of differentiation can be interchanged so that $\partial_i \partial_j = \partial_j \partial_i$). Then

$$\varepsilon^{ijk} \partial_i \partial_j = -\varepsilon^{jik} \partial_i \partial_j = -\varepsilon^{jik} \partial_j \partial_i = -\varepsilon^{ijk} \partial_i \partial_j$$

where the last step follows because i and j are dummy indices, and we can therefore relabel them. But then $\varepsilon^{ijk} \partial_i \partial_j = 0$ and we have proved our identity.

The last step in the previous example is actually a special case of a general result. To see this, suppose that we have an object $A^{ij\cdots}$ that is labeled by two or more indices, and suppose that it is antisymmetric in two of those indices (say i, j). This means that $A^{ij\cdots} = -A^{ji\cdots}$. Now suppose that we have another object $S_{ij\cdots}$ that is symmetric in i and j, so that $S_{ij\cdots} = S_{ji\cdots}$. If we multiply A times S and sum over the indices i and j, then using the symmetry and antisymmetry properties of S and A we have

$$\begin{aligned} A^{ij\cdots} S_{ij\cdots} &= -A^{ji\cdots} S_{ij\cdots} && \text{by the antisymmetry of } A \\ &= -A^{ji\cdots} S_{ji\cdots} && \text{by the symmetry of S} \\ &= -A^{ij\cdots} S_{ij\cdots} && \text{by relabeling the dummy indices } i \text{ and } j \end{aligned}$$

and therefore we have the general result

$$A^{ij\cdots} S_{ij\cdots} = 0.$$

It is also worth pointing out that the indices i and j need not be the first pair of indices, nor do they need to be adjacent. For example, we still have $A^{\cdots i \cdots j \cdots} S_{\cdots i \cdots j \cdots} = 0$.

Now suppose that we have an arbitrary object T^{ij} without any particular symmetry properties. Then we can turn this into an antisymmetric object $T^{[ij]}$ by a process called **antisymmetrization** as follows:

$$T^{ij} \to T^{[ij]} := \frac{1}{2!}(T^{ij} - T^{ji}).$$

In other words, we add up all possible permutations of the indices, with the sign of each permutation being either $+1$ (for an even permutation) or -1 (for an odd permutation), and then divide this sum by the total number of permutations, which in this case is $2!$. If we have something of the form T^{ijk} then we would have

$$T^{ijk} \to T^{[ijk]} := \frac{1}{3!}(T^{ijk} - T^{ikj} + T^{kij} - T^{kji} + T^{jki} - T^{jik})$$

where we alternate signs with each transposition. The generalization to an arbitrary number of indices should be clear. Note also that we could antisymmetrize only over a subset of the indices if required.

It is also important to note that it is impossible to have a nonzero antisymmetric object with more indices than the dimension of the space we are working in. This is simply because at least one index will necessarily be repeated. For example, if we are in $\mathbb{R}^3$, then anything of the form T^{ijkl} must have at least one index repeated because each index can only range between 1, 2 and 3.

Now, why did we go through all of this? Well, first recall that we can write the Kronecker delta in any of the equivalent forms $\delta_{ij} = \delta^i_j = \delta^j_i$. Then we can construct quantities like

$$\delta^{[1}_i \delta^{2]}_j = \frac{1}{2!}\left(\delta^1_i \delta^2_j - \delta^2_i \delta^1_j\right) = \delta^1_{[i}\delta^2_{j]}$$

and

$$\delta^{[1}_i \delta_j{}^2 \delta^{3]}_k = \frac{1}{3!}\left(\delta^1_i\delta^2_j\delta^3_k - \delta^1_i\delta^3_j\delta^2_k + \delta^3_i\delta^1_j\delta^2_k - \delta^3_i\delta^2_j\delta^1_k + \delta^2_i\delta^3_j\delta^1_k - \delta^2_i\delta^1_j\delta^3_k\right).$$

In particular, we now want to show that

$$\varepsilon_{ijk} = 3!\, \delta^{[1}_i \delta_j{}^2 \delta^{3]}_k. \tag{3.5}$$

Clearly, if $i = 1, j = 2$ and $k = 3$ we have

$$\begin{aligned} 3!\, \delta^{[1}_1 \delta_2{}^2 \delta^{3]}_3 &= 3!\, \frac{1}{3!}\left(\delta^1_1\delta^2_2\delta^3_3 - \delta^1_1\delta^3_2\delta^2_3 + \delta^3_1\delta^1_2\delta^2_3 - \delta^3_1\delta^2_2\delta^1_3 + \delta^2_1\delta^3_2\delta^1_3 - \delta^2_1\delta^1_2\delta^3_3\right) \\ &= 1 - 0 + 0 - 0 + 0 - 0 = 1 = \varepsilon_{123} \end{aligned}$$

so equation (3.5) is correct in this particular case. But now we make the crucial observation that both sides of equation (3.5) are antisymmetric in (i, j, k), and hence the equation must hold for all values of (i, j, k). This is because any permutation of (i, j, k) results in the same change of sign on both sides, and both sides also equal 0 if any two indices are repeated. Therefore equation (3.5) is true in general.

To derive what is probably the most useful identity involving the Levi-Civita symbol, we begin with the fact that $\varepsilon^{123} = 1$. Multiplying the left side of equation (3.5) by 1 in this form yields

$$\varepsilon_{ijk}\, \varepsilon^{123} = 3!\, \delta^{[1}_i \delta_j{}^2 \delta^{3]}_k.$$

But now we again make the observation that both sides are antisymmetric in $(1, 2, 3)$, and hence both sides are equal for all values of the upper indices, and we have the fundamental result

$$\varepsilon_{ijk}\, \varepsilon^{nlm} = 3!\, \delta^{[n}_i \delta_j{}^l \delta^{m]}_k. \tag{3.6}$$

We now set $n = k$ and sum over k. (This process of setting two indices equal to each other and summing is called **contraction**.) Using the fact that

$$\delta_k^k = \sum_{i=1}^{3} \delta_k^k = 3$$

along with terms such as $\delta_i^k \delta_k^m = \delta_i^m$ we find

$$\begin{aligned}
\varepsilon_{ijk}\, \varepsilon^{klm} &= 3!\, \delta_i^{[k} \delta_j^{\,l} \delta_k^{m]} \\
&= \delta_i^k \delta_j^l \delta_k^m - \delta_i^k \delta_j^m \delta_k^l + \delta_i^m \delta_j^k \delta_k^l - \delta_i^m \delta_j^l \delta_k^k + \delta_i^l \delta_j^m \delta_k^k - \delta_i^l \delta_j^k \delta_k^m \\
&= \delta_i^m \delta_j^l - \delta_i^l \delta_j^m + \delta_i^m \delta_j^l - 3\delta_i^m \delta_j^l + 3\delta_i^l \delta_j^m - \delta_i^l \delta_j^m \\
&= \delta_i^l \delta_j^m - \delta_i^m \delta_j^l .
\end{aligned}$$

In other words, we have the extremely useful result

$$\varepsilon_{ijk}\, \varepsilon^{klm} = \delta_i^l \delta_j^m - \delta_i^m \delta_j^l . \tag{3.7}$$

This result is so useful that it should definitely be memorized.

Example 3.3. Let us derive the well-known triple vector product known as the "*bac* − *cab*" rule. We simply compute using equation (3.7):

$$\begin{aligned}
[\vec{A} \times (\vec{B} \times \vec{C})]_i &= \varepsilon_{ijk} A^j (\vec{B} \times \vec{C})^k = \varepsilon_{ijk} \varepsilon^{klm} A^j B_l C_m \\
&= (\delta_i^l \delta_j^m - \delta_i^m \delta_j^l) A^j B_l C_m = A^m B_i C_m - A^j B_j C_i \\
&= B_i (\vec{A} \cdot \vec{C}) - C_i (\vec{A} \cdot \vec{B})
\end{aligned}$$

and therefore

$$\vec{A} \times (\vec{B} \times \vec{C}) = \vec{B}(\vec{A} \cdot \vec{C}) - \vec{C}(\vec{A} \cdot \vec{B}).$$

We also point out that some of the sums in this derivation can be done in more than one way. For example, we have either $\delta_i^l \delta_j^m A^j B_l C_m = A^m B_i C_m = B_i(\vec{A} \cdot \vec{C})$ or $\delta_i^l \delta_j^m A^j B_l C_m = A^j B_i C_j = B_i(\vec{A} \cdot \vec{C})$, but the end result is always the same. Note also that at every step along the way, the only index that isn't repeated (and hence summed over) is i.

Example 3.4. Equation (3.7) is just as useful in vector calculus calculations. Here is an example to illustrate the technique.

$$\begin{aligned}
[\nabla \times (\nabla \times \vec{A})]_i &= \varepsilon_{ijk} \partial^j (\nabla \times \vec{A})^k = \varepsilon_{ijk} \varepsilon^{klm} \partial^j \partial_l A_m \\
&= (\delta_i^l \delta_j^m - \delta_i^m \delta_j^l) \partial^j \partial_l A_m = \partial^j \partial_i A_j - \partial^j \partial_j A_i \\
&= \partial_i (\nabla \cdot \vec{A}) - \nabla^2 A_i
\end{aligned}$$

and hence we have the identity

$$\nabla \times (\nabla \times \vec{A}) = \nabla(\nabla \cdot \vec{A}) - \nabla^2 \vec{A}$$

which is very useful in discussing the theory of electromagnetic waves.

Exercises

1. Using the Levi-Civita symbol, prove the following vector identities in $\mathbb{R}^3$ equipped with a Cartesian coordinate system (where the vectors are actually vector fields where necessary, f is differentiable, and $\nabla^i = \partial_i = \partial/\partial x^i$):
   (a) $\vec{A} \times (\vec{B} \times \vec{C}) = (\vec{A} \cdot \vec{C})\vec{B} - (\vec{A} \cdot \vec{B})\vec{C}$
   (b) $(\vec{A} \times \vec{B}) \cdot (\vec{C} \times \vec{D}) = (\vec{A} \cdot \vec{C})(\vec{B} \cdot \vec{D}) - (\vec{A} \cdot \vec{D})(\vec{B} \cdot \vec{C})$
   (c) $\nabla \times \nabla f = 0$
   (d) $\nabla \cdot (\nabla \times \vec{A}) = 0$
   (e) $\nabla \cdot (\vec{A} \times \vec{B}) = \vec{B} \cdot (\nabla \times \vec{A}) - \vec{A} \cdot (\nabla \times \vec{B})$
   (f) $\nabla \times (\vec{A} \times \vec{B}) = \vec{A}(\nabla \cdot \vec{B}) - \vec{B}(\nabla \cdot \vec{A}) + (\vec{B} \cdot \nabla)\vec{A} - (\vec{A} \cdot \nabla)\vec{B}$

2. Using the divergence theorem ($\int_V \nabla \cdot \vec{A}\, d^3x = \int_S \vec{A} \cdot \vec{n}\, da$), prove

$$\int_V \nabla \times \vec{A}\, d^3x = \int_S \vec{n} \times \vec{A}\, da.$$

   [*Hint*: Let $\vec{C}$ be a *constant* vector and show

$$\vec{C} \cdot \int_V \nabla \times \vec{A}\, d^3x = \int_S (\vec{n} \times \vec{A}) \cdot \vec{C}\, da = \vec{C} \cdot \int_S \vec{n} \times \vec{A}\, da.]$$

3. Let $A_{i_1 \cdots i_r}$ be an antisymmetric object, and suppose $T^{i_1 \cdots i_r \cdots}{}_{\cdots}$ is an arbitrary object. Show

$$A_{i_1 \cdots i_r} T^{i_1 \cdots i_r \cdots}{}_{\cdots} = A_{i_1 \cdots i_r} T^{[i_1 \cdots i_r] \cdots}{}_{\cdots}.$$

3.2 Definitions and Elementary Properties

In treating vectors in $\mathbb{R}^3$, we used the permutation symbol ε_{ijk} defined in the last section. We are now ready to apply the same techniques to the theory of determinants. The idea is that we want to define a mapping from a matrix $A \in M_n(\mathcal{F})$ to $\mathcal{F}$ in a way that has certain algebraic properties. Since a matrix in $M_n(\mathcal{F})$ has components a_{ij} with i and j ranging from 1 to n, we are going to need a higher dimensional version of the Levi-Civita symbol already introduced. The obvious extension to n dimensions is the following.

We define

$$\varepsilon^{i_1\cdots i_n} = \begin{cases} 1 & \text{if } i_1,\dots,i_n \text{ is an \textit{even} permutation of } 1,\dots,n \\ -1 & \text{if } i_1,\dots,i_n \text{ is an \textit{odd} permutation of } 1,\dots,n \\ 0 & \text{if } i_1,\dots,i_n \text{ is not a permutation of } 1,\dots,n \end{cases}.$$

Again, there is no practical difference between $\varepsilon^{i_1\cdots i_n}$ and $\varepsilon_{i_1\cdots i_n}$. Using this, we define the **determinant** of $A = (a_{ij}) \in M_n(\mathcal{F})$ to be the number

$$\det A = \varepsilon^{i_1\cdots i_n} a_{1i_1} a_{2i_2} \cdots a_{ni_n}. \tag{3.8}$$

Look carefully at what this expression consists of. Since $\varepsilon^{i_1\cdots i_n}$ vanishes unless $(i_1,\dots,i_n)$ are all distinct, and there are $n!$ such distinct orderings, we see that $\det A$ consists of $n!$ terms in the sum, where each term is a product of n factors a_{ij}, and where each term consists precisely of one factor from each row and each column of A. In other words, $\det A$ is a sum of terms where each term is a product of one element from each row and each column, and the sum is over all such possibilities.

The determinant is frequently written as

$$\det A = \begin{vmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn} \end{vmatrix}.$$

The determinant of an $n \times n$ matrix is said to be of **order** n. Note also that the determinant is only defined for square matrices.

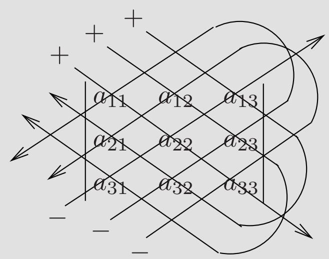
Example 3.5. Leaving the easier 2×2 case to you to verify, we will work out the 3×3 case and show that it gives the same result that you probably learned in a more elementary course. So, for $A = (a_{ij}) \in M_3(\mathcal{F})$ we have

$$\begin{aligned} \det A &= \varepsilon^{ijk} a_{1i} a_{2j} a_{3k} \\ &= \varepsilon^{123} a_{11}a_{22}a_{33} + \varepsilon^{132} a_{11}a_{23}a_{32} + \varepsilon^{312} a_{13}a_{21}a_{32} \\ &\qquad + \varepsilon^{321} a_{13}a_{22}a_{31} + \varepsilon^{231} a_{12}a_{23}a_{31} + \varepsilon^{213} a_{12}a_{21}a_{33} \\ &= a_{11}a_{22}a_{33} - a_{11}a_{23}a_{32} + a_{13}a_{21}a_{32} \\ &\qquad - a_{13}a_{22}a_{31} + a_{12}a_{23}a_{31} - a_{12}a_{21}a_{33} \end{aligned}$$

You may recognize this in either of the mnemonic forms (sometimes called **Sarrus's rule**)

$$\begin{array}{c|ccc|cc} & a_{11} & a_{12} & a_{13} & a_{11} & a_{12} \\ & a_{21} & a_{22} & a_{23} & a_{21} & a_{22} \\ & a_{31} & a_{32} & a_{33} & a_{31} & a_{32} \\ - & - & - & + & + & + \end{array}$$

or



Here, we are to add together all products of terms connected by a $(+)$ line, and subtract all of the products connected by a $(-)$ line. We will see in a later section that this 3×3 determinant may be expanded as a sum of three 2×2 determinants.

Example 3.6. Let $A = (a_{ij})$ be a diagonal matrix, i.e., $a_{ij} = 0$ if $i \neq j$. Then

$$\begin{aligned} \det A &= \varepsilon^{i_1 \cdots i_n} a_{1i_1} \cdots a_{ni_n} = \varepsilon^{1 \cdots n} a_{11} \cdots a_{nn} \\ &= a_{11} \cdots a_{nn} = \prod_{i=1}^{n} a_{ii} \end{aligned}$$

so that

$$\begin{vmatrix} a_{11} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & a_{nn} \end{vmatrix} = \prod_{i=1}^{n} a_{ii}.$$

In particular, we see that $\det I = 1$.

We now prove a number of useful properties of determinants. These are all very straightforward applications of the definition (3.8) once you have become comfortable with the notation. In fact, in our opinion, this approach to determinants affords the simplest way in which to arrive at these results, and is far less confusing than the usual inductive proofs.

Theorem 3.1. *For any $A \in M_n(\mathcal{F})$ we have*

$$\det A = \det A^T.$$

Proof. This is simply an immediate consequence of our definition of determinant. We saw that $\det A$ is a sum of all possible products of one element from each row and each column, and no product can contain more than one term from a given column because the corresponding ε symbol would vanish. This means that an equivalent way of writing all $n!$ such products is (note the order of subscripts is reversed)

$$\det A = \varepsilon^{i_1\cdots i_n} a_{i_1 1} \cdots a_{i_n n}.$$

But $a_{ij} = a^T{}_{ji}$ so this is just

$$\det A = \varepsilon^{i_1\cdots i_n} a_{i_1 1} \cdots a_{i_n n} = \varepsilon^{i_1\cdots i_n} a^T{}_{1 i_1} \cdots a^T{}_{n i_n} = \det A^T. \quad \blacksquare$$

In order to help us gain some additional practice manipulating these quantities, we prove this theorem again based on another result which we will find very useful in its own right. We start from the definition $\det A = \varepsilon^{i_1\cdots i_n} a_{1 i_1} \cdots a_{n i_n}$. Again using $\varepsilon_{1\cdots n} = 1$ we have

$$\varepsilon_{1\cdots n} \det A = \varepsilon^{i_1\cdots i_n} a_{1 i_1} \cdots a_{n i_n}. \tag{3.9}$$

By definition of the permutation symbol, the left side of this equation is antisymmetric in $(1, \ldots, n)$. But so is the right side because, taking a_{1i_1} and a_{2i_2} as an example, we see that

$$\begin{aligned} \varepsilon^{i_1 i_2\cdots i_n} a_{1 i_1} a_{2 i_2} \cdots a_{n i_n} &= \varepsilon^{i_1 i_2\cdots i_n} a_{2 i_2} a_{1 i_1} \cdots a_{n i_n} \\ &= -\varepsilon^{i_2 i_1\cdots i_n} a_{2 i_2} a_{1 i_1} \cdots a_{n i_n} \\ &= -\varepsilon^{i_1 i_2\cdots i_n} a_{2 i_1} a_{1 i_2} \cdots a_{n i_n} \end{aligned}$$

where the last line follows by a relabeling of the dummy indices i_1 and i_2.

So, by a now familiar argument, both sides of equation (3.9) must be true for any values of the indices $(1, \ldots, n)$ and we have the extremely useful result

$$\varepsilon_{j_1\cdots j_n} \det A = \varepsilon^{i_1\cdots i_n} a_{j_1 i_1} \cdots a_{j_n i_n}. \tag{3.10}$$

This equation will turn out to be very helpful in many proofs that would otherwise be considerably more difficult.

Let us now use equation (3.10) to prove Theorem 3.1. We begin with the analogous result to equation (3.2). This is

$$\varepsilon^{i_1\cdots i_n} \varepsilon_{i_1\cdots i_n} = n!. \tag{3.11}$$

Using this, we multiply equation (3.10) by $\varepsilon^{j_1\cdots j_n}$ to yield

$$n! \det A = \varepsilon^{j_1\cdots j_n} \varepsilon^{i_1\cdots i_n} a_{j_1 i_1} \cdots a_{j_n i_n}.$$

On the other hand, by definition of $\det A^T$ we have

$$\det A^T = \varepsilon^{i_1\cdots i_n} a^T{}_{1 i_1} \cdots a^T{}_{n i_n} = \varepsilon^{i_1\cdots i_n} a_{i_1 1} \cdots a_{i_n n}.$$

Multiplying the left side of this equation by $1 = \varepsilon_{1\cdots n}$ and again using the antisymmetry of both sides in $(1, \ldots, n)$ yields

$$\varepsilon_{j_1\cdots j_n} \det A^T = \varepsilon^{i_1\cdots i_n} a_{i_1 j_1} \cdots a_{j_n i_n}.$$

(This also follows by applying equation (3.10) to A^T directly.)

Now multiply this last equation by $\varepsilon^{j_1\cdots j_n}$ to obtain

$$n!\det A^T = \varepsilon^{i_1\cdots i_n}\varepsilon^{j_1\cdots j_n}a_{i_1j_1}\cdots a_{j_ni_n}.$$

Relabeling the dummy indices i and j we have

$$n!\det A^T = \varepsilon^{j_1\cdots j_n}\varepsilon^{i_1\cdots i_n}a_{j_1i_1}\cdots a_{i_nj_n}$$

which is exactly the same as the above expression for $n!\det A$, and we have again proved Theorem 3.1.

Looking at the definition $\det A = \varepsilon^{i_1\cdots i_n}a_{1i_1}\cdots a_{ni_n}$, we see that we can view the determinant as a function of the rows of A: $\det A = \det(A_1,\dots,A_n)$. Since each row is actually a vector in $\mathcal{F}^n$, we can replace A_1 (for example) by any linear combination of two vectors in $\mathcal{F}^n$ so that $A_1 = rB_1 + sC_1$ where $r,s\in\mathcal{F}$ and $B_1,C_1\in\mathcal{F}^n$. Let $B=(b_{ij})$ be the matrix with rows $B_i = A_i$ for $i=2,\dots,n$, and let $C=(c_{ij})$ be the matrix with rows $C_i=A_i$ for $i=2,\dots,n$. Then

$$\begin{aligned}\det A &= \det(A_1,A_2,\dots,A_n) = \det(rB_1+sC_1,A_2,\dots,A_n)\\ &= \varepsilon^{i_1\cdots i_n}(rb_{1i_1}+sc_{1i_1})a_{2i_2}\cdots a_{ni_n}\\ &= r\varepsilon^{i_1\cdots i_n}b_{1i_1}a_{2i_2}\cdots a_{ni_n} + s\varepsilon^{i_1\cdots i_n}c_{1i_1}a_{2i_2}\cdots a_{ni_n}\\ &= r\det B + s\det C.\end{aligned}$$

Since this argument clearly could have been applied to any of the rows of A, we have proved the following theorem.

Theorem 3.2. *Let $A\in M_n(\mathcal{F})$ have row vectors $A_1,\dots,A_n$ and assume that for some $i=1,\dots,n$ we have*

$$A_i = rB_i + sC_i$$

where $B_i,C_i\in\mathcal{F}^n$ and $r,s\in\mathcal{F}$. Let $B\in M_n(\mathcal{F})$ have rows $A_1,\dots,A_{i-1},B_i,A_{i+1},\dots,A_n$ and $C\in M_n(\mathcal{F})$ have rows $A_1,\dots,A_{i-1},C_i,A_{i+1},\dots,A_n$. Then

$$\det A = r\det B + s\det C.$$

In fact, a simple induction argument gives the following result.

Corollary 1. *Let $A\in M_n(\mathcal{F})$ have row vectors $A_1,\dots,A_n$ and suppose that for some $i=1,\dots,n$ we have*

$$A_i = \sum_{j=1}^k r_jB_j$$

where $B_j \in \mathcal{F}^n$ for $j = 1, \ldots, k$ and each $r_i \in \mathcal{F}$. Then

$$\begin{aligned}\det A &= \det(A_1, \ldots, A_{i-1}, \sum_{j=1}^{k} r_j B_j, A_{i+1}, \ldots, A_n) \\ &= \sum_{j=1}^{k} \det(A_1, \ldots, A_{i-1}, B_i, A_{i+1}, \ldots, A_n).\end{aligned}$$

Corollary 2. *If any row of $A \in M_n(\mathcal{F})$ is zero, then $\det A = 0$.*

Proof. If any row of A is zero, then clearly $\det A = \varepsilon^{i_1 \cdots i_n} a_{1i_1} \cdots a_{ni_n} = 0$ because each product in the sum of products contains an element from each row. This result also follows from the theorem by letting $r = s = 0$. ∎

Corollary 3. *If $A \in M_n(\mathcal{F})$ and $r \in \mathcal{F}$, then $\det(rA) = r^n \det A$.*

Proof. Since $rA = (ra_{ij})$ we have

$$\begin{aligned}\det(rA) &= \varepsilon^{i_1 \cdots i_n} (ra_{1i_1}) \cdots (ra_{ni_n}) \\ &= r^n \varepsilon^{i_1 \cdots i_n} a_{1i_1} \cdots a_{ni_n} \\ &= r^n \det A.\end{aligned}$$ ∎

Let us go back and restate an earlier result (equation (3.10)) for emphasis, and also look at two of its immmediate consequences.

Theorem 3.3. *If $A \in M_n(\mathcal{F})$, then*

$$\varepsilon_{j_1 \cdots j_n} \det A = \varepsilon^{i_1 \cdots i_n} a_{j_1 i_1} \cdots a_{j_n i_n}.$$

Corollary 1. *If $B \in M_n(\mathcal{F})$ is obtained from $A \in M_n(\mathcal{F})$ by interchanging two rows of A, the $\det B = -\det A$.*

Proof. This is really just what the theorem says in words. (See the discussion between equations (3.9) and (3.10).) For example, let B result from interchanging rows 1 and 2 of A. Then

$$\det B = \varepsilon^{i_1 i_2 \cdots i_n} b_{1i_1} b_{2i_2} \cdots b_{ni_n} = \varepsilon^{i_1 i_2 \cdots i_n} a_{2i_1} a_{1i_2} \cdots a_{ni_n}$$

$$
\begin{aligned}
&= \varepsilon^{i_1 i_2 \cdots i_n} a_{1i_2} a_{2i_1} \cdots a_{ni_n} = -\varepsilon^{i_2 i_1 \cdots i_n} a_{1i_2} a_{2i_1} \cdots a_{ni_n} \\
&= -\varepsilon^{i_1 i_2 \cdots i_n} a_{1i_1} a_{2i_2} \cdots a_{ni_n} \\
&= -\det A = \varepsilon_{213\cdots n} \det A.
\end{aligned}
$$

where again the next to last line follows by relabeling. ∎

Corollary 2. *If $A \in M_n(\mathcal{F})$ has two identical rows, then* $\det A = 0$.

Proof. If B is the matrix obtained by interchanging two identical rows of A, then by the previous corollary we have

$$\det A = \det B = -\det A$$

and therefore $\det A = 0$. ∎

Here is another way to view Theorem 3.3 and its corollaries. As we did earlier, we view $\det A$ as a function of the rows of A. Then the corollaries state that $\det A = 0$ if any two rows are the same, and $\det A$ changes sign if two nonzero rows are interchanged. In other words, we have

$$\det(A_{j_1}, \dots, A_{j_n}) = \varepsilon_{j_1 \cdots j_n} \det A.$$

If it isn't immediately obvious to you that this is true, then note that for $(j_1, \dots, j_n) = (1, \dots, n)$ it's just an identity. So by the antisymmetry of both sides, it must be true for all $j_1, \dots, j_n$.

3.3 Additional Properties of Determinants

It is important to realize that because $\det A^T = \det A$, these last two theorems and their corollaries apply to columns as well as to rows. Furthermore, these results now allow us easily see what happens to the determinant of a matrix A when we apply elementary row (or column) operations to A. In fact, if you think for a moment, the answer should be obvious. For a type α transformation (i.e., interchanging two rows), we have just seen that $\det A$ changes sign (Theorem 3.3, Corollary 1). For a type β transformation (i.e., multiply a single row by a nonzero scalar), we can let $r = k$, $s = 0$ and $B_i = A_i$ in Theorem 3.2 to see that $\det A \to k \det A$. And for a type γ transformation (i.e., add a multiple of one row to another) we have (for $A_i \to A_i + kA_j$ and using Theorems 3.2 and 3.3, Corollary 2)

$$
\begin{aligned}
\det(A_1, \dots, A_i + kA_j, \dots, A_n) &= \det A + k \det(A_1, \dots, A_j, \dots, A_j, \dots, A_n) \\
&= \det A + 0 = \det A.
\end{aligned}
$$

Summarizing these results, we have the following theorem.

Theorem 3.4. *Suppose $A \in M_n(\mathcal{F})$ and let $B \in M_n(\mathcal{F})$ be row equivalent to A.*

(i) If B results from the interchange of two rows of A, then $\det B = -\det A$.

(ii) If B results from multiplying any row (or column) of A by a scalar k, then $\det B = k \det A$.

(iii) If B results from adding a multiple of one row of A to another row, then $\det B = \det A$.

Corollary. *If R is the reduced row-echelon form of a matrix A, then* $\det R = 0$ *if and only if* $\det A = 0$.

Proof. This follows from Theorem 3.4 since A and R are row-equivalent. ∎

Besides the very easy to handle diagonal matrices, another type of matrix that is easy to deal with are the triangular matrices. To be precise, a matrix $A \in M_n(\mathcal{F})$ is said to be **upper-triangular** if $a_{ij} = 0$ for $i > j$, and A is said to be **lower-triangular** if $a_{ij} = 0$ for $i < j$. Thus a matrix is upper-triangular if it is of the form

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ 0 & a_{22} & a_{23} & \cdots & a_{2n} \\ 0 & 0 & a_{33} & \cdots & a_{3n} \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & a_{nn} \end{bmatrix}$$

and lower-triangular if it is of the form

$$\begin{bmatrix} a_{11} & 0 & 0 & \cdots & 0 \\ a_{21} & a_{22} & 0 & \cdots & 0 \\ a_{31} & a_{32} & a_{33} & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \cdots & a_{nn} \end{bmatrix}.$$

We will use the term **triangular** to mean either upper- or lower-triangular.

Theorem 3.5. *If $A \in M_n(\mathcal{F})$ is a triangular matrix, then*

$$\det A = \prod_{i=1}^{n} a_{ii}.$$

Proof. If A is lower-triangular, then A is of the form shown above. Now look carefully at the definition $\det A = \varepsilon^{i_1 \cdots i_n} a_{1i_1} \cdots a_{ni_n}$. Since A is lower-triangular we have $a_{ij} = 0$ for $i < j$. But then we must have $i_1 = 1$ or else $a_{1i_1} = 0$. Now consider a_{2i_2}. Since $i_1 = 1$ and $a_{2i_2} = 0$ if $2 < i_2$, we must have $i_2 = 2$. Next, $i_1 = 1$ and $i_2 = 2$ means that $i_3 = 3$ or else $a_{3i_3} = 0$. Continuing in this way we see that the only nonzero term in the sum is when $i_j = j$ for each $j = 1, \ldots, n$ and hence

$$\det A = \varepsilon^{12\cdots n} a_{11} \cdots a_{nn} = \prod_{i=1}^{n} a_{ii}.$$

If A is an upper-triangular matrix, then the theorem follows from Theorem 3.1. ∎

An obvious corollary is the following (which was also shown directly in Example 3.6).

Corollary. *If $A \in M_n(\mathcal{F})$ is diagonal, then* $\det A = \prod_{i=1}^{n} a_{ii}$.

Another fundamental result that relates the rank of a matrix to its determinant is this.

Theorem 3.6. *A matrix $A \in M_n(\mathcal{F})$ is singular if and only if* $\det A = 0$.

Proof. Let R be the reduced row-echelon form of A. If A is singular, then by definition $\operatorname{rank}(A) < n$ so that by Theorem 2.6 there must be at least one zero row in the matrix R. Hence $\det R = 0$ by Theorem 3.2, Corollary 2, and therefore $\det A = 0$ by the corollary to Theorem 3.4.

Conversely, we want to show that $\det A = 0$ implies $\operatorname{rank}(A) < n$. To do this we prove the contrapositive, i.e., we show $\operatorname{rank}(A) = n$ implies $\det A \neq 0$. So, assume that $\operatorname{rank}(A) = n$. Then, by Theorem 2.7, we must have $R = I_n$ so that $\det R = 1$. Hence $\det A \neq 0$ by the corollary to Theorem 3.4. In other words, if $\det A = 0$ it follows that $\operatorname{rank}(A) < n$. ∎

Before turning to the method of evaluating determinants known as "expansion by cofactors," there is one last very fundamental result that we should prove.

Theorem 3.7. *If $A, B \in M_n(\mathcal{F})$, then*

$$\det(AB) = (\det A)(\det B).$$

Proof. If either A or B is singular (i.e., their rank is less than n) then so is AB (by the corollary to Theorem 2.17). But then (by Theorem 3.6) either $\det A = 0$ or $\det B = 0$, and also $\det(AB) = 0$ so the theorem is true in this case.

Now assume that both A and B are nonsingular, and let $C = AB$. Then $C_i = (AB)_i = \sum_k a_{ik} B_k$ for each $i = 1, \ldots, n$ (see Section 2.5) so that from Theorem 3.2, Corollary 1 we see that

$$\begin{aligned}
\det C &= \det(C_1, \ldots, C_n) \\
&= \det\left(\sum_{j_1} a_{1j_1} B_{j_1}, \ldots, \sum_{j_n} a_{nj_n} B_{j_n}\right) \\
&= \sum_{j_1} \cdots \sum_{j_n} a_{1j_1} \cdots a_{nj_n} \det(B_{j_1}, \ldots, B_{j_n}).
\end{aligned}$$

But $\det(B_{j_1}, \ldots, B_{j_n}) = \varepsilon_{j_1 \cdots j_n} \det B$ (see the discussion following Theorem 3.3) so we have

$$\begin{aligned}
\det C &= \sum_{j_1} \cdots \sum_{j_n} a_{1j_1} \cdots a_{nj_n} \varepsilon^{j_1 \cdots j_n} \det B \\
&= (\det A)(\det B).
\end{aligned}$$

∎

Corollary. *If $A \in M_n(\mathcal{F})$ is nonsingular, then*

$$\det A^{-1} = (\det A)^{-1}.$$

Proof. If A is nonsingular, then A^{-1} exists, and hence by the theorem we have

$$1 = \det I = \det(AA^{-1}) = (\det A)(\det A^{-1})$$

and therefore

$$\det A^{-1} = (\det A)^{-1}.$$

∎

Exercises

1. Compute the determinants of the following matrices directly from the definition:

   (a) $\begin{bmatrix} 1 & 2 & 3 \\ 4 & -2 & 3 \\ 2 & 5 & -1 \end{bmatrix}$ (b) $\begin{bmatrix} 2 & 0 & 1 \\ 3 & 2 & -3 \\ -1 & -3 & 5 \end{bmatrix}$

2. Consider the following real matrix:

   $$A = \begin{bmatrix} 2 & 1 & 9 & 1 \\ 4 & 3 & -1 & 2 \\ 1 & 4 & 3 & 2 \\ 3 & 2 & 1 & 4 \end{bmatrix}$$

   Evaluate $\det A$ by reducing A to upper-triangular form and using Theorem 3.4.

3. Using the definition, show that

$$\begin{vmatrix} a_1 & 0 & 0 & 0 \\ b_1 & b_2 & b_3 & b_4 \\ c_1 & c_2 & c_3 & c_4 \\ d_1 & d_2 & d_3 & d_4 \end{vmatrix} = a_1 \begin{vmatrix} b_2 & b_3 & b_4 \\ c_2 & c_3 & c_4 \\ d_2 & d_3 & d_4 \end{vmatrix}.$$

4. Evaluate the determinant of the following matrix:

$$\begin{bmatrix} 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & \cdots & 1 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 1 & 0 & \cdots & 0 & 0 \end{bmatrix}$$

5. If A is nonsingular and $A^{-1} = A^T$, show $\det A = \pm 1$ (such a matrix A is said to be **orthogonal**).

6. Consider a complex matrix $U \in M_n(\mathbb{C})$.
   (a) If $U^* = (u_{ij}{}^*)$, show $\det U^* = (\det U)^*$.
   (b) Let $U^\dagger = U^{*T}$ (this is called the **adjoint** or **conjugate transpose** of U, and is not to be confused with the classical adjoint introduced in the next section). Suppose U is such that $U^\dagger U = UU^\dagger = I$ (such a U is said to be **unitary**). Show that we may write $\det U = e^{i\phi}$ for some real ϕ.

7. (a) If A is a real $n \times n$ matrix and k is a positive odd integer, show that $A^k = I_n$ implies that $\det A = 1$.
   (b) If $A^n = 0$ for some positive integer n, show that $\det A = 0$. (A matrix for which $A^n = 0$ is said to be **nilpotent**.)

8. If the **anticommutator** $[A, B]_+ = AB + BA = 0$, show that A and/or B in $M_n(\mathcal{F})$ must be singular if n is odd. What can you say if n is even?

9. Suppose C is a 3×3 matrix that can be expressed as the product of a 3×2 matrix A and a 2×3 matrix B. Show that $\det C = 0$. Generalize this result to $n \times n$ matrices.

10. Recall that A is **symmetric** if $A^T = A$. If A is symmetric, show that

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

11. Recall that a matrix A is said to be **antisymmetric** if $A^T = -A$, i.e., $a^T_{ij} = -a_{ji}$. If A is an antisymmetric square matrix of odd size, prove that $\det A = 0$.

12. (a) Recall (see Exercise 2.5.7) that if $A \in M_n(\mathcal{F})$, then $\operatorname{tr} A = \sum_i a_{ii}$. If A is a 2×2 matrix, prove that $\det(I + A) = 1 + \det A$ if and only if $\operatorname{tr} A = 0$. Is this true for any size square matrix?

(b) If $|a_{ij}| \ll 1$, show that $\det(I + A) \cong 1 + \operatorname{tr} A$.

13. Two matrices A and A' are said to be **similar** if there exists a nonsingular matrix P such that $A' = PAP^{-1}$. The operation of transforming A into A' in this manner is called a **similarity transformation**.
    (a) Show this defines an equivalence relation on the set of all matrices.
    (b) Show that the determinant is invariant under a similarity transformation.
    (c) Show that the trace (Exercise 2.5.7) is also invariant.

14. Consider the matrices
$$A = \begin{bmatrix} 2 & 0 & -1 \\ 3 & 0 & 2 \\ 4 & -3 & 7 \end{bmatrix} \qquad B = \begin{bmatrix} 3 & 2 & -4 \\ 1 & 0 & -2 \\ -2 & 3 & 3 \end{bmatrix}$$
    (a) Evaluate $\det A$ and $\det B$.
    (b) Find AB and BA.
    (c) Evaluate $\det AB$ and $\det BA$.

15. Show that
$$\begin{vmatrix} a_1 & b_1 + xa_1 & c_1 + yb_1 + za_1 \\ a_2 & b_2 + xa_2 & c_2 + yb_2 + za_2 \\ a_3 & b_3 + xa_3 & c_3 + yb_3 + za_3 \end{vmatrix} = \begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}.$$

16. Find all values of x for which each of the following determinants is zero:

(a) $\begin{vmatrix} x-1 & 1 & 1 \\ 0 & x-4 & 1 \\ 0 & 0 & x-2 \end{vmatrix}$ (b) $\begin{vmatrix} 1 & x & x \\ x & 1 & x \\ x & x & 1 \end{vmatrix}$

(c) $\begin{vmatrix} 1 & x & x^2 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \end{vmatrix}$

17. Given a matrix A, the matrix that remains after any rows and/or columns of A have been deleted is called a **submatrix** of A, and the determinant of a square submatrix is called a **subdeterminant**. Show that the rank of a matrix A is the size of the largest nonvanishing subdeterminant of A. [*Hint*: Think about Theorem 2.6, Corollary 2 of Theorem 3.2, and Theorem 3.4.]

18. Show that the following determinant is zero:
$$\begin{vmatrix} a^2 & (a+1)^2 & (a+2)^2 & (a+3)^2 \\ b^2 & (b+1)^2 & (b+2)^2 & (b+3)^2 \\ c^2 & (c+1)^2 & (c+2)^2 & (c+3)^2 \\ d^2 & (d+1)^2 & (d+2)^2 & (d+3)^2 \end{vmatrix}.$$
[*Hint*: You need not actually evaluate it.]

19. Show that
$$\begin{vmatrix} 1 & 6 & 11 & 16 & 21 \\ 2 & 7 & 12 & 17 & 22 \\ 3 & 8 & 13 & 18 & 23 \\ 4 & 9 & 14 & 19 & 24 \\ 5 & 10 & 15 & 20 & 25 \end{vmatrix} = 0.$$

20. (a) If E is an elementary matrix, show (without using Theorem 3.1) that $\det E^T = \det E$.
    (b) Use Theorem 2.23 to show that $\det A^T = \det A$ for any $A \in M_n(\mathcal{F})$.

21. Use the material of this section to give a proof (independent of Chapter 2) that the product of nonsingular matrices is nonsingular.

3.4 Expansion by Cofactors

We now turn to the method of evaluating determinants known as **expansion by cofactors** (or sometimes **expansion by minors** or the **Laplace expansion formula**). This is really more of a theoretical tool than it is practical. The easiest way in general to evaluate determinants is probably to row reduce to triangular form and use Theorem 3.5.

Example 3.7. Consider the matrix A given by
$$A = \begin{bmatrix} 2 & -1 & 3 \\ 1 & 2 & -1 \\ -3 & 0 & 2 \end{bmatrix}.$$
Then we have
$$\begin{aligned}
\det A &= \begin{vmatrix} 2 & -1 & 3 \\ 1 & 2 & -1 \\ -3 & 0 & 2 \end{vmatrix} \\
&= 2 \begin{vmatrix} 1 & -1/2 & 3/2 \\ 1 & 2 & -1 \\ -3 & 0 & 2 \end{vmatrix} \begin{matrix} \leftarrow (1/2)A_1 \\ \\ \\ \end{matrix} \\
&= 2 \begin{vmatrix} 1 & -1/2 & 3/2 \\ 0 & 5/2 & -5/2 \\ 0 & -3/2 & 13/2 \end{vmatrix} \begin{matrix} \\ \leftarrow -A_1 + A_2 \\ \leftarrow \ 3A_1 + A_3 \end{matrix} \\
&= 2 \begin{vmatrix} 1 & -1/2 & 3/2 \\ 0 & 5/2 & -5/2 \\ 0 & 0 & 5 \end{vmatrix} \begin{matrix} \\ \\ \leftarrow (3/5)A_2 + A_3 \end{matrix} \\
&= (2)(1)(5/2)(5) = 25.
\end{aligned}$$

The reader should verify this result by the direct calculation of $\det A$.

However, the cofactor expansion approach gives us a useful formula for the inverse of a matrix as we shall see. The idea is to reduce a determinant of order n to a sum of n determinants of order $n-1$, and then reduce each of these to determinants of order $n-2$ and so forth down to determinants of order 3 or even 2. We will do this by simply taking a careful look at the definition of determinant. By way of notation, if $A \in M_n(\mathcal{F})$, we let $A_{rs} \in M_{n-1}(\mathcal{F})$ be the matrix obtained from A by deleting the rth row and sth column. The matrix A_{rs} is called the rsth **minor matrix** of A, and $\det A_{rs}$ is called the rsth **minor** of A. And the number $a'_{rs} = (-1)^{r+s} \det A_{rs}$ is called the rsth **cofactor** of A.

Example 3.8. Let $A \in M_3(\mathbb{R})$ be defined by

$$A = \begin{bmatrix} 2 & -1 & 5 \\ 0 & 3 & 4 \\ 1 & 2 & -3 \end{bmatrix}.$$

Then the $(2,3)$ minor matrix is constructed by deleting the 2nd row and 3rd column and is given by

$$A_{23} = \begin{bmatrix} 2 & -1 \\ 1 & 2 \end{bmatrix}.$$

The $(2,3)$ minor is then $\det A_{23} = 4 - (-1) = 5$, and the $(2,3)$ cofactor is $(-1)^{2+3} \det A_{23} = -5$.

Theorem 3.8. *Let $A \in M_n(\mathcal{F})$. Then for any $r = 1, \ldots, n$ we have*

$$\det A = \sum_{s=1}^{n} (-1)^{r+s} a_{rs} \det A_{rs} = \sum_{s=1}^{n} a_{rs} a'_{rs}.$$

Proof. Start from $\det A = \varepsilon^{i_1 \cdots i_r \cdots i_n} a_{1i_1} \cdots a_{ri_r} \cdots a_{ni_n}$. Explicitly writing out the sum over the particular index $i_r = 1, \ldots, n$ we have

$$\begin{aligned} \det A = \varepsilon^{i_1 \cdots 1 \cdots i_n} a_{1i_1} \cdots a_{r1} \cdots a_{ni_n} &+ \varepsilon^{i_1 \cdots 2 \cdots i_n} a_{1i_1} \cdots a_{r2} \cdots a_{ni_n} \\ &+ \cdots + \varepsilon^{i_1 \cdots n \cdots i_n} a_{1i_n} \cdots a_{rn} \cdots a_{ni_n} \end{aligned}$$

Note that the scalars a_{rj} (i.e., $a_{r1}, a_{r2}, \ldots, a_{rn}$) are not summed over in any way, and can be brought outside each ε factor.

In order to help with bookkeeping, we introduce the so-called "magic hat" $\widehat{\ }$ that makes things disappear. For example, $12\widehat{3}4 = 124$. Using this as a placeholder in our previous expression for $\det A$, we factor out each a_{rj} and write

$$\det A = a_{r1}\varepsilon^{i_1\cdots 1\cdots i_n} a_{1i_1} \cdots \widehat{a_{r1}} \cdots a_{ni_n} + a_{r2}\varepsilon^{i_1\cdots 2\cdots i_n} a_{1i_1} \cdots \widehat{a_{r2}} \cdots a_{ni_n} + \cdots + a_{rn}\varepsilon^{i_1\cdots n\cdots i_n} a_{1i_n} \cdots \widehat{a_{rn}} \cdots a_{ni_n} \tag{3.12}$$

Now look at the first ε term in equation (3.12) and write it as

$$\varepsilon^{i_1\cdots i_{r-1}1 i_{r+1}\cdots i_n} = (-1)^{r-1}\varepsilon^{1 i_1\cdots i_{r-1}\widehat{i_r} i_{r+1}\cdots i_n}.$$

Because $(-1)^2 = 1$ we can write $(-1)^{r-1} = (-1)^{r-1}(-1)^2 = (-1)^{r+1}$ and the first term becomes

$$(-1)^{r+1} a_{r1}\varepsilon^{1 i_1\cdots i_{r-1}\widehat{i_r} i_{r+1}\cdots i_n} a_{1i_1} \cdots \widehat{a_{r1}} \cdots a_{ni_n}.$$

Now, because of the 1 in the ε superscripts, none of the rest of the i_j's can take the value 1. Then a moments thought should make it clear that (except for the $(-1)^{r+1}a_{r1}$ factor) this last expression is just the determinant of the matrix that results by deleting the rth row and 1st column of A.

$$\begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{r-1,1} & a_{r-1,2} & \cdots & a_{r-1,n} \\ a_{r1} & a_{r2} & \cdots & a_{rn} \\ a_{r+1,1} & a_{r+1,2} & \cdots & a_{r+1,n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix} \xrightarrow[\text{and 1st column}]{\text{delete } r\text{th row}} \begin{vmatrix} a_{12} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{r-1,2} & \cdots & a_{r-1,n} \\ a_{r+1,2} & \cdots & a_{r+1,n} \\ \vdots & & \vdots \\ a_{n2} & \cdots & a_{nn} \end{vmatrix}$$

Indeed, it's just the sum of all possible products of one element from each row and column as before, with the appropriate sign. To see that the sign is correct, note that the first term in equation (3.8) is $(i_1, i_2, \ldots, i_n) = (1, 2, \ldots, n)$ so that the contribution to the determinant is $+a_{11}a_{22}\cdots a_{nn}$ which is the diagonal term. Then what we have here with the rth row and 1st column deleted also has the diagonal term $a_{12}a_{23}\cdots a_{r-1,r}a_{r+1,r+1}\cdots a_{nn}$. (Note that in the determinant on the right above, the entry in the $(r-1)$th row and $(r-1)$th column is $a_{r-1,r}$ and the entry in the rth row and rth column is $a_{r+1,r+1}$.) Furthermore, the column indices on this diagonal term are $(2\,3\cdots r\; r+1\cdots n) = (i_1 i_2 \cdots i_{r-1} i_{r+1}\cdots i_n)$ so the ε term becomes

$$\varepsilon^{1 i_1 i_2\cdots i_{r-1} i_{r+1}\cdots i_n} = \varepsilon^{123\cdots r\; r+1\cdots n} = +1\,.$$

Thus the entire first term becomes simply $(-1)^{r+1}a_{r1}\det A_{r1}$.

Now look at the second term in equation (3.12) and use $(-1)^4 = 1$ to obtain

$$\varepsilon^{i_1\cdots i_{r-1}2\, i_{r+1}\cdots i_n} = (-1)^{r-2}\varepsilon^{i_1 2\, i_3\cdots\widehat{i_r}\cdots i_n} = (-1)^{r+2}\varepsilon^{i_1 2\, i_3\cdots\widehat{i_r}\cdots i_n}$$

and therefore

$$\begin{aligned} a_{r2}\varepsilon^{i_1\cdots 2\cdots i_n} a_{1i_1}\cdots\widehat{a_{r2}}\cdots a_{ni_n} &= (-1)^{r+2} a_{r2}\varepsilon^{i_1 2\, i_3\cdots\widehat{i_r}\cdots i_n} a_{1i_1}\cdots\widehat{a_{r2}}\cdots a_{ni_n} \\ &= (-1)^{r+2} a_{r2}\det A_{r2}. \end{aligned}$$

We continue this procedure until the last term which is (note that here the n superscript is moved to the right because every term with $j > r$ has to go to the right to keep the correct increasing order in the ε superscripts)

$$\begin{aligned} a_{rn}\varepsilon^{i_1\cdots i_{r-1} n i_{r+1}\cdots i_n} & a_{1i_1}\cdots\widehat{a_{rn}}\cdots a_{ni_n} \\ &= (-1)^{n-r} a_{rn}\varepsilon^{i_1\cdots\widehat{i_r}\cdots i_n n} a_{1i_1}\cdots\widehat{a_{rn}}\cdots a_{ni_n} \\ &= (-1)^{r+n} a_{rn}\det A_{rn} \end{aligned}$$

where we also used the fact that $(-1)^{n-r} = 1/(-1)^{n-r} = (-1)^{r-n} = (-1)^{r+n}$.

Putting all of this together we have the desired result

$$\det A = \sum_{s=1}^{n} (-1)^{r+s} a_{rs}\det A_{rs} = \sum_{s=1}^{n} a_{rs}a'_{rs}. \qquad \blacksquare$$

This theorem gives $\det A$ as a sum of cofactors of the rth row. If we apply the theorem to the matrix A^T and then use the fact that $\det A^T = \det A$, we have the following corollary.

Corollary 1. *Using the same notation as in the theorem, for any $s = 1, \ldots, n$ we have*

$$\det A = \sum_{r=1}^{n} a_{rs}a'_{rs}.$$

Note that in this corollary the sum is over the row index, and hence this is an expansion in terms of the sth *column*, whereas in Theorem 3.8 the expansion is in terms of the rth *row*.

Corollary 2. *Using the same notation as in the theorem we have*

$$\sum_{s=1}^{n} a_{ks}a'_{rs} = 0 \quad \textit{if } k \neq r$$

$$\sum_{r=1}^{n} a_{rk}a'_{rs} = 0 \quad \textit{if } k \neq s$$

Proof. Given $A \in M_n(\mathcal{F})$, define $B \in M_n(\mathcal{F})$ by $B_i = A_i$ for $i \neq r$ and $B_r = A_k$ (where $k \neq r$). In other words, replace the rth row of A by the kth row of A

to obtain B. Since B now has two identical rows it follows that $\det B = 0$. Next, observe that $B_{rs} = A_{rs}$ since both matrices delete the rth row (so that the remaining rows are identical), and hence $b'_{rs} = a'_{rs}$ for each $s = 1, \dots, n$. Therefore, since $B_r = A_k$ we have

$$0 = \det B = \sum_{s=1}^{n} b_{rs} b'_{rs} = \sum_{s=1}^{n} b_{rs} a'_{rs} = \sum_{s=1}^{n} a_{ks} a'_{rs}.$$

Similarly, the other result follows by replacing the sth column of A by the kth column so that $b_{rs} = a_{rk}$ and then using Corollary 1. ∎

As we will shortly see, these corollaries will give us a general equation for the inverse of a matrix.

Example 3.9. As in Example 3.8, let $A \in M_3(\mathbb{R})$ be defined by

$$A = \begin{bmatrix} 2 & -1 & 5 \\ 0 & 3 & 4 \\ 1 & 2 & -3 \end{bmatrix}.$$

To evaluate $\det A$ we expand by the second row:

$$\begin{aligned} \det A &= a_{21}a'_{21} + a_{22}a'_{22} + a_{23}a'_{23} \\ &= 0 + (-1)^{2+2}(3) \begin{vmatrix} 2 & 5 \\ 1 & -3 \end{vmatrix} + (-1)^{2+3}(4) \begin{vmatrix} 2 & -1 \\ 1 & 2 \end{vmatrix} \\ &= 3(-6-5) - 4(4-(-1)) = -53. \end{aligned}$$

The reader should repeat this calculation using other rows and columns to see that they all yield the same result.

Example 3.10. Let us evaluate $\det A$ where

$$A = \begin{bmatrix} 5 & 4 & 2 & 1 \\ 2 & 3 & 1 & -2 \\ -5 & -7 & -3 & 9 \\ 1 & -2 & -1 & 4 \end{bmatrix}.$$

Since type γ transformations don't change the determinant (see Theorem 3.4), we do the following sequence of elementary row transformations: (i) $A_1 \to A_1 - 2A_2$ (ii) $A_3 \to A_3 + 3A_2$ (iii) $A_4 \to A_4 + A_2$ and this gives us the matrix

$$B = \begin{bmatrix} 1 & -2 & 0 & 5 \\ 2 & 3 & 1 & -2 \\ 1 & 2 & 0 & 3 \\ 3 & 1 & 0 & 2 \end{bmatrix}$$

with $\det B = \det A$.

Now expand by cofactors of the third column (since there is only one term):

$$\det A = (-1)^{2+3}(1)\begin{vmatrix} 1 & -2 & 5 \\ 1 & 2 & 3 \\ 3 & 1 & 2 \end{vmatrix} = 38$$

where you can either evaluate this 3×3 determinant directly or reduce it to a sum of 2×2 determinants.

We are now in a position to prove a general formula for the inverse of a matrix. Combining Theorem 3.8 and its corollaries, we obtain (for $k, r, s = 1, \ldots, n$)

$$\sum_{s=1}^{n} a_{ks}a'_{rs} = \delta_{kr} \det A \tag{3.13a}$$

$$\sum_{r=1}^{n} a_{rk}a'_{rs} = \delta_{ks} \det A. \tag{3.13b}$$

Since each $a'_{ij} \in \mathcal{F}$, we may use the them to form a new matrix $(a'_{ij}) \in M_n(\mathcal{F})$. The transpose of this new matrix is called the **adjoint** of A (or sometimes the **classical adjoint** to distinguish it from another type of adjoint to be discussed later) and is denoted by $\operatorname{adj} A$. In other words,

$$\operatorname{adj} A = (a'_{ij})^T.$$

Noting that the (i, j)th entry of I_n is $I_{ij} = \delta_{ij}$, it is now easy to prove the following.

Theorem 3.9. *For any $A \in M_n(\mathcal{F})$ we have $A(\operatorname{adj} A) = (\det A)I = (\operatorname{adj} A)A$. In particular, if A is nonsingular, then*

$$A^{-1} = \frac{\operatorname{adj} A}{\det A}.$$

Proof. Using $(\operatorname{adj} A)_{sr} = a'_{rs}$, we may write equation (3.13a) in matrix form as

$$A(\operatorname{adj} A) = (\det A)I$$

and equation (3.13b) as

$$(\operatorname{adj} A)A = (\det A)I.$$

Therefore, if A is nonsingular, we have $\det A \neq 0$ (Theorem 3.6) and hence

$$\frac{A(\operatorname{adj} A)}{\det A} = I = \frac{(\operatorname{adj} A)A}{\det A}.$$

Thus the uniqueness of the inverse (Theorem 2.20, Corollary 1) implies

$$A^{-1} = \frac{(\operatorname{adj} A)}{\det A}. \quad ∎$$

We remark that it is important to realize that the equations

$$A(\operatorname{adj} A) = (\det A)I$$

and

$$(\operatorname{adj} A)A = (\det A)I$$

are valid even if A is singular.

Example 3.11. Let us use this method to find the inverse of the matrix

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

used in Example 2.12. Leaving the details to the reader, we evaluate the cofactors using the formula $a'_{rs} = (-1)^{r+s} \det A_{rs}$ to obtain $a'_{11} = -2$, $a'_{12} = -4$, $a'_{13} = -6$, $a'_{21} = -1$, $a'_{22} = -2$, $a'_{23} = 3$, $a'_{31} = -7$, $a'_{32} = -2$, and $a'_{33} = -3$. Hence we find

$$\operatorname{adj} A = \begin{bmatrix} -2 & -1 & -7 \\ -4 & -2 & -2 \\ -6 & 3 & -3 \end{bmatrix}.$$

To evaluate $\det A$, we may either calculate directly or by minors to obtain $\det A = -12$. Alternatively, from equation (3.13a) we have

$$\begin{aligned} (\det A)I = A(\operatorname{adj} A) &= \begin{bmatrix} -1 & 2 & 1 \\ 0 & 3 & -2 \\ 2 & -1 & 0 \end{bmatrix} \begin{bmatrix} -2 & -1 & -7 \\ -4 & -2 & -2 \\ -6 & 3 & -3 \end{bmatrix} \\ &= \begin{bmatrix} -12 & 0 & 0 \\ 0 & -12 & 0 \\ 0 & 0 & -12 \end{bmatrix} \\ &= -12 \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \end{aligned}$$

so that we again find that $\det A = -12$. In any case, we see that

$$A^{-1} = \frac{\operatorname{adj} A}{-12} = \begin{bmatrix} 1/6 & 1/12 & 7/12 \\ 1/3 & 1/6 & 1/6 \\ 1/2 & -1/4 & 1/4 \end{bmatrix}$$

which agrees with Example 2.12 as it should.

If the reader thinks about Theorem 2.6, Corollary 2 of Theorem 3.2, and Theorem 3.4 (or has already worked Exercise 3.3.17), our next theorem should come as no real surprise. By way of more terminology, given a matrix A, the matrix that remains after any rows and/or columns have been deleted is called a **submatrix** of A.

Theorem 3.10. *Let A be a matrix in $M_{m\times n}(\mathcal{F})$, and let k be the largest integer such that some submatrix $B \in M_k(\mathcal{F})$ of A has a nonzero determinant. Then* $\operatorname{rank}(A) = k$.

Proof. Since B is a $k\times k$ submatrix of A with $\det B \neq 0$, it follows from Theorem 3.6 that B is nonsingular and hence $\operatorname{rank}(B) = k$. This means that the k rows of B are linearly independent, and hence the k rows of A that contain the rows of B must also be linearly independent. Therefore $\operatorname{rank}(A) = \operatorname{rr}(A) \geq k$. By definition of k, there can be no $r \times r$ submatrix of A with nonzero determinant if $r > k$. We will now show that if $\operatorname{rank}(A) = r$, then there necessarily exists an $r \times r$ submatrix with nonzero determinant. This will prove that $\operatorname{rank}(A) = k$.

If $\operatorname{rank}(A) = r$, let A' be the matrix with r linearly independent rows $A_{i_1}, \dots, A_{i_r}$. Clearly $\operatorname{rank}(A') = r$ also. But by definition of rank, we can also choose r linearly independent columns of A'. This results in a nonsingular matrix A'' of size r, and hence $\det A'' \neq 0$ by Theorem 3.6. ∎

Exercises

1. Verify the result of Example 3.7 by direct calculation.

2. Verify the 3×3 determinant in Example 3.10.

3. Verify the terms a'_{ij} in Example 3.11.

4. Evaluate the following determinants by expanding by minors of either rows or columns:

(a) $\begin{vmatrix} 2 & -1 & 5 \\ 0 & 3 & 4 \\ 1 & 2 & -3 \end{vmatrix}$ (b) $\begin{vmatrix} 2 & 5 & 5 & 3 \\ 7 & -8 & 2 & 3 \\ 1 & -1 & 4 & -2 \\ -3 & 9 & -1 & 3 \end{vmatrix}$

(c) $\begin{vmatrix} 3 & 2 & 2 & 3 \\ 1 & -4 & 2 & 1 \\ 4 & 5 & -1 & 0 \\ -1 & -4 & 2 & 7 \end{vmatrix}$ (d) $\begin{vmatrix} 3 & 1 & 0 & 4 & 2 & 1 \\ 2 & 0 & 1 & 0 & 5 & 1 \\ 0 & 4 & -1 & 1 & -1 & 2 \\ 0 & 0 & 0 & 2 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 1 & 0 & 1 \end{vmatrix}$

5. Let $A \in M_n(\mathcal{F})$ be a matrix with 0's down the main diagonal and 1's elsewhere. Show that $\det A = n - 1$ if n is odd, and $\det A = 1 - n$ if n is even.

6. (a) Show that the determinant of the matrix
$$\begin{bmatrix} 1 & a & a^2 \\ 1 & b & b^2 \\ 1 & c & c^2 \end{bmatrix}$$
is $(c - a)(c - b)(b - a)$.

   (b) Define the function $f(x) = Ae^{i\alpha x} + Be^{i\beta x} + Ce^{i\gamma x}$ and assume that A, B, C and α, β, γ are all nonzero constants. If $f(x) = 0$ for all $x \in \mathbb{R}$, show that $\alpha = \beta = \gamma$.

7. Consider the matrix $V_n \in M_n(\mathcal{F})$ defined by
$$\begin{bmatrix} 1 & x_1 & {x_1}^2 & \cdots & {x_1}^{n-1} \\ 1 & x_2 & {x_2}^2 & \cdots & {x_2}^{n-1} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_n & {x_n}^2 & \cdots & {x_n}^{n-1} \end{bmatrix}.$$
Prove that
$$\det V_n = \prod_{i<j} (x_j - x_i)$$
where the product is over all pairs i and j satisfying $1 \le i, j \le n$. This matrix is called the **Vandermonde matrix** of order n. [*Hint*: This should be done by induction on n. The idea is to show that
$$\det V_n = (x_2 - x_1)(x_3 - x_1) \cdots (x_{n-1} - x_1)(x_n - x_1) \det V_{n-1}.$$
Perform elementary column operations on V_n to obtain a new matrix ${V'}_n$ with a 1 in the $(1, 1)$ position and 0's in every other position of the first row. Now factor out the appropriate term from each of the other rows.]

8. The obvious method for deciding if two quadratic polynomials have a common root involves the quadratic formula, and hence taking square roots. This exercise investigates an alternative "root free" approach. (We assume that the reader knows that x_0 is a root of the polynomial $p(x)$ if and only if $p(x_0) = 0$.)

   (a) Show that
$$\begin{aligned} \det A &= \begin{vmatrix} 1 & -(x_1 + x_2) & x_1 x_2 & 0 \\ 0 & 1 & -(x_1 + x_2) & x_1 x_2 \\ 1 & -(y_1 + y_2) & y_1 y_2 & 0 \\ 0 & 1 & -(y_1 + y_2) & y_1 y_2 \end{vmatrix} \\ &= (x_1 - y_1)(x_1 - y_2)(x_2 - y_1)(x_2 - y_2). \end{aligned}$$

(b) Using this result, show that the polynomials

$$a_0x^2 + a_1x + a_2 \qquad (a_0 \neq 0)$$

$$b_0x^2 + b_1x + b_2 \qquad (b_0 \neq 0)$$

have a common root if and only if

$$\begin{vmatrix} a_0 & a_1 & a_2 & 0 \\ 0 & a_0 & a_1 & a_2 \\ b_0 & b_1 & b_2 & 0 \\ 0 & b_0 & b_1 & b_2 \end{vmatrix} = 0.$$

[*Hint*: Note that if x_1 and x_2 are the roots of the first polynomial, then

$$(x - x_1)(x - x_2) = x^2 + (a_1/a_0)x + a_2/a_0$$

and similarly for the second polynomial.]

9. Show that

$$\begin{vmatrix} x & 0 & 0 & 0 & \cdots & 0 & a_0 \\ -1 & x & 0 & 0 & \cdots & 0 & a_1 \\ 0 & -1 & x & 0 & \cdots & 0 & a_2 \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & -1 & x + a_{n+1} \end{vmatrix} = x^n + a_{n-1}x^{n-1} + \cdots + a_0.$$

Explain why this shows that given any polynomial $p(x)$ of degree n, there exists a matrix $A \in M_n(\mathcal{F})$ such that $\det(xI - A) = p(x)$.

10. Consider the following real matrix:

$$A = \begin{bmatrix} a & b & c & d \\ b & -a & d & -c \\ c & -d & -a & b \\ d & c & -b & -a \end{bmatrix}.$$

Show that $\det A = 0$ implies $a = b = c = d = 0$. [*Hint*: Find AA^T and use Theorems 3.1 and 3.7.]

11. Let u, v and w be three vectors in $\mathbb{R}^3$ with the standard inner product, and consider the determinant $G(u, v, w)$ (the **Gramian** of $\{u, v, w\}$) defined by

$$G(u, v, w) = \begin{vmatrix} \langle u, u\rangle & \langle u, v\rangle & \langle u, w\rangle \\ \langle v, u\rangle & \langle v, v\rangle & \langle v, w\rangle \\ \langle w, u\rangle & \langle w, v\rangle & \langle w, w\rangle \end{vmatrix}.$$

Show $G(u, v, w) = 0$ if and only if $\{u, v, w\}$ are linearly dependent. As we shall see in Chapter 8, $G(u, v, w)$ represents the volume of the parallelepiped in $\mathbb{R}^3$ defined by $\{u, v, w\}$.)

12. Find the inverse (if it exists) of the following matrices:

(a) $\begin{bmatrix} 1 & -1 & 2 \\ 1 & 2 & 0 \\ 4 & 1 & 3 \end{bmatrix}$ (b) $\begin{bmatrix} 1 & 3 & 2 \\ 2 & 1 & 3 \\ 3 & 2 & 1 \end{bmatrix}$ (c) $\begin{bmatrix} -2 & 2 & 3 \\ 4 & 3 & -6 \\ 1 & -1 & 2 \end{bmatrix}$

(d) $\begin{bmatrix} 8 & 2 & 5 \\ -7 & 3 & -4 \\ 9 & -6 & 4 \end{bmatrix}$ (e) $\begin{bmatrix} 1 & 2 & 3 & 1 \\ 1 & 3 & 3 & 2 \\ 2 & 4 & 3 & 3 \\ 1 & 1 & 1 & 1 \end{bmatrix}$ (f) $\begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 1 \\ 3 & 4 & 1 & 2 \\ 4 & 1 & 2 & 3 \end{bmatrix}$

13. Find the inverse of
$$\begin{bmatrix} \cos\theta & -\sin\theta \\ -\sin\theta & -\cos\theta \end{bmatrix}.$$

14. Suppose that an n-square matrix A is nilpotent (i.e., $A^k = 0$ for some integer $k > 0$). Prove that $I_n + A$ is nonsingular, and find its inverse. [*Hint*: Note that $(I + A)(I - A) = I - A^2$ etc.]

15. Let $P \in M_n(\mathcal{F})$ be such that $P^2 = P$. If $\lambda \neq 1$, prove that $I_n - \lambda P$ is invertible, and that
$$(I_n - \lambda P)^{-1} = I_n + \frac{\lambda}{1-\lambda} P.$$

16. If $A = (a_{ij})$ is a symmetric matrix, show $(a'_{ij}) = (\operatorname{adj} A)^T$ is also symmetric.

17. If $a, b, c \in \mathbb{R}$, find the inverse of
$$\begin{bmatrix} 1 & a & b \\ -a & 1 & c \\ -b & -c & 1 \end{bmatrix}.$$

18. (a) Using $A^{-1} = (\operatorname{adj} A)/\det A$, show that the inverse of an upper (lower) triangular matrix is upper (lower) triangular.
    (b) If $a \neq 0$, find the inverse of
$$\begin{bmatrix} a & b & c & d \\ 0 & a & b & c \\ 0 & 0 & a & b \\ 0 & 0 & 0 & a \end{bmatrix}.$$

19. Let $A \in M_n(\mathbb{R})$ have all integer entries. Show the following are equivalent:
    (a) $\det A = \pm 1$.
    (b) All entries of A^{-1} are integers.

20. For each of the following matrices A, find the value(s) of x for which the **characteristic matrix** $xI - A$ is invertible.

(a) $\begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}$ (b) $\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$

(c) $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$ (d) $\begin{bmatrix} 0 & 1 & 2 \\ 0 & 1 & 3 \\ 0 & 0 & -1 \end{bmatrix}$

21. Let $A \in M_n(\mathcal{F})$ have exactly one nonzero entry in each row and column. Show that A is invertible, and that its inverse is of the same form.

22. If $A \in M_n(\mathcal{F})$, show $\det(\operatorname{adj} A) = (\det A)^{n-1}$.

23. Show that A is nonsingular if and only if $\operatorname{adj} A$ is nonsingular.

24. Using determinants, find the rank of each of the following matrices:

(a) $\begin{bmatrix} 1 & 2 & 3 & 4 \\ -1 & 2 & 1 & 0 \end{bmatrix}$ (b) $\begin{bmatrix} -1 & 0 & 1 & 2 \\ 1 & 1 & 3 & 0 \\ -1 & 2 & 4 & 1 \end{bmatrix}$

3.5 Determinants and Linear Equations

Suppose that we have a system of n equations in n unknowns which we write in the usual way as

$$\sum_{j=1}^{n} a_{ij} x_j = y_i, \qquad i = 1, \ldots, n.$$

We assume that $A = (a_{ij}) \in M_n(\mathcal{F})$ is nonsingular. In matrix form, this system may be written as $AX = Y$ as we saw earlier. Since A is nonsingular, A^{-1} exists (Theorem 2.20, Corollary 1) and $\det A \neq 0$ (Theorem 3.6). Therefore the solution to $AX = Y$ is given by

$$X = A^{-1}Y = \frac{(\operatorname{adj} A)Y}{\det A}.$$

But $\operatorname{adj} A = (a'_{ij})^T$ so that

$$x_j = \sum_{i=1}^{n} \frac{(\operatorname{adj} A)_{ji} y_i}{\det A} = \sum_{i=1}^{n} \frac{a'_{ij} y_i}{\det A}.$$

From Corollary 1 of Theorem 3.8, we see that $\sum_i y_i a'_{ij}$ is just the expansion by minors of the jth column of the matrix C whose columns are given by $C^i = A^i$ for $i \neq j$ and $C^j = Y$. We are thus led to the following result, called **Cramer's rule**.

Theorem 3.11. *If $A = (a_{ij}) \in M_n(\mathcal{F})$ is nonsingular, then the system of linear equations*

$$\sum_{j=1}^{n} a_{ij} x_j = y_i, \qquad i = 1, \ldots, n$$

has the unique solution

$$x_j = \frac{1}{\det A} \det(A^1, \ldots, A^{j-1}, Y, A^{j+1}, \ldots, A^n).$$

Proof. This theorem was actually proved in the preceding discussion, where uniqueness follows from Theorem 2.13. However, it is instructive to give a more direct proof as follows. We write our system as $\sum A^i x_i = Y$ and simply compute using Corollary 1 of Theorem 3.2 and Corollary 2 of Theorem 3.3:

$$\begin{aligned}
\det(A^1, \ldots, A^{j-1}, Y, A^{j+1}, \ldots, A^n)& \\
&= \det(A^1, \ldots, A^{j-1}, \sum_i A^i x_i, A^{j+1}, \ldots, A^n) \\
&= \sum_i x_i \det(A^1, \ldots, A^{j-1}, A^i, A^{j+1}, \ldots, A^n) \\
&= x_j \det(A^1, \ldots, A^{j-1}, A^j, A^{j+1}, \ldots, A^n) \\
&= x_j \det A.
\end{aligned}$$

∎

Corollary. *A homogeneous system of equations*

$$\sum_{j=1}^{n} a_{ij} x_j = 0, \qquad i = 1, \ldots, n$$

has a nontrivial solution if and only if $\det A = 0$.

Proof. We see from Cramer's rule that if $\det A \neq 0$, then the solution of the homogeneous system is just the zero vector (by Corollary 2 of Theorem 3.2 as applied to columns instead of rows). This shows that the if the system has a nontrivial solution, then $\det A = 0$.

On the other hand, if $\det A = 0$ then the columns of A must be linearly dependent (Theorem 3.6). But the system $\sum_j a_{ij} x_j = 0$ may be written as $\sum_j A^j x_j = 0$ where A^j is the jth column of A. Hence the linear dependence of the A^j shows that the x_j may be chosen such that they are not all zero, and therefore a nontrivial solution exists. (We remark that this corollary also follows directly from Theorems 2.9 and 3.6.) ∎

Example 3.12. Let us solve the system

$$\begin{aligned} 5x + 2y + z &= 3 \\ 2x - y + 2z &= 7 \\ x + 5y - z &= 6 \end{aligned}$$

We see that $A = (a_{ij})$ is nonsingular since

$$\begin{vmatrix} 5 & 2 & 1 \\ 2 & -1 & 2 \\ 1 & 5 & -1 \end{vmatrix} = -26 \neq 0.$$

We then have

$$x = \frac{-1}{26}\begin{vmatrix} 3 & 2 & 1 \\ 7 & -1 & 2 \\ 6 & 5 & -1 \end{vmatrix} = (-1/26)(52) = -2$$

$$y = \frac{-1}{26}\begin{vmatrix} 5 & 3 & 1 \\ 2 & 7 & 2 \\ 1 & 6 & -1 \end{vmatrix} = (-1/26)(-78) = 3$$

$$z = \frac{-1}{26}\begin{vmatrix} 5 & 2 & 3 \\ 2 & -1 & 7 \\ 1 & 5 & 6 \end{vmatrix} = (-1/26)(-182) = 7.$$

Exercises

1. Using Cramers rule, find a solution (if it exists) of the following systems of equations:

(a) $\begin{aligned} 3x + y - z &= 0 \\ x - y + 3z &= 1 \\ 2x + 2y + z &= 7 \end{aligned}$

(b) $\begin{aligned} 2x + y + 2z &= 0 \\ 3x - 2y + z &= 1 \\ -x + 2y + 2z &= -7 \end{aligned}$

(c) $\begin{aligned} 2x - 3y + z &= 10 \\ -x + 3y + 2z &= -2 \\ 4x + 4y + 5z &= 4 \end{aligned}$

(d) $\begin{aligned} x + 2y - 3z + t &= -9 \\ 2x + y + 2z - t &= 3 \\ -x + y + 2z - t &= 0 \\ 3x + 4y + z + 4t &= 3 \end{aligned}$

2. By calculating the inverse of the matrix of coefficients, solve the following systems:

(a) $\begin{aligned} 2x - 3y + z &= a \\ x + 2y + 3z &= b \\ 3x - y + 2z &= c \end{aligned}$

(b) $\begin{aligned} x + 2y + 4z &= a \\ -x + 3y - 2z &= b \\ 2x - y + z &= c \end{aligned}$

(c)
$$\begin{aligned} 2x + y + 2z - 3t &= a \\ 3x + 2y + 3z - 5t &= b \\ 2x + 2y + z - t &= c \\ 5x + 5y + 2z - 2t &= d \end{aligned}$$

(d)
$$\begin{aligned} 6x + y + 4z - 3t &= a \\ 2x - y &= b \\ x + y + z &= c \\ -3x - y - 2z + t &= d \end{aligned}$$

3. If $\det A \neq 0$ and $AB = AC$, show that $B = C$.

4. Find, if possible, a 2×2 matrix X that satisfies each of the given equations:

   (a) $\begin{bmatrix} 2 & 3 \\ 1 & 2 \end{bmatrix} X \begin{bmatrix} 3 & 4 \\ 2 & 3 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$

   (b) $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} X \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 3 & 2 \end{bmatrix}$

5. Consider the system

$$\begin{aligned} ax + by &= \alpha + \beta t \\ cx + dy &= \gamma + \delta t \end{aligned}$$

   where t is a parameter, $\beta^2 + \delta^2 \neq 0$ and

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} \neq 0.$$

   Show that the set of solutions as t varies is a straight line in the direction of the vector

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} \begin{bmatrix} \beta \\ \delta \end{bmatrix}.$$

6. Let A, B, C and D be 2×2 matrices, and let R and S be vectors (i.e., 2×1 matrices). Show that the system

$$\begin{aligned} AX + BY &= R \\ CX + DY &= S \end{aligned}$$

   can always be solved for vectors X and Y if

$$\begin{vmatrix} a_{11} & a_{12} & b_{11} & b_{12} \\ a_{21} & a_{22} & b_{21} & b_{22} \\ c_{11} & c_{12} & d_{11} & d_{12} \\ c_{21} & c_{22} & d_{21} & d_{22} \end{vmatrix} \neq 0.$$

Chapter 4

Linear Transformations and Matrices

In Section 2.1 we defined matrices by systems of linear equations, and in Section 2.5 we showed that the set of all matrices over a field $\mathcal{F}$ may be endowed with certain algebraic properties such as addition and multiplication. In this chapter we present another approach to defining matrices, and we will see that it also leads to the same algebraic behavior as well as yielding important new properties.

4.1 Linear Transformations

Recall that vector space homomorphisms were defined in Section 1.3. We now repeat that definition using some new terminology. In particular, a mapping $T : U \to V$ of two vector spaces over the same field $\mathcal{F}$ is called a **linear transformation** if it has the following properties for all $x, y \in U$ and $a \in \mathcal{F}$:

(LT1) $T(x + y) = T(x) + T(y)$
(LT2) $T(ax) = aT(x)$.

Letting $a = 0$ and -1 shows

$$T(0) = 0$$

and

$$T(-x) = -T(x).$$

We also see that

$$T(x - y) = T(x + (-y)) = T(x) + T(-y) = T(x) - T(y).$$

It should also be clear that by induction we have, for any finite sum,

$$T\Big(\sum a_i x_i\Big) = \sum T(a_i x_i) = \sum a_i T(x_i)$$

for any vectors $x_i \in U$ and scalars $a_i \in \mathcal{F}$.

Example 4.1. Let $T : \mathbb{R}^3 \to \mathbb{R}^2$ be the "projection" mapping defined for any $u = (x, y, z) \in \mathbb{R}^3$ by

$$T(u) = T(x, y, z) = (x, y, 0).$$

Then if $v = (x', y', z')$ we have

$$\begin{aligned} T(u + v) &= T(x + x', y + y', z + z') \\ &= (x + x', y + y', 0) \\ &= (x, y, 0) + (x', y', 0) \\ &= T(u) + T(v) \end{aligned}$$

and

$$T(au) = T(ax, ay, az) = (ax, ay, 0) = a(x, y, 0) = aT(u).$$

Hence T is a linear transformation.

Example 4.2. Let $P \in M_n(\mathcal{F})$ be a fixed invertible matrix. We define a mapping $S : M_n(\mathcal{F}) \to M_n(\mathcal{F})$ by $S(A) = P^{-1}AP$. It is easy to see that this defines a linear transformation since

$$S(\alpha A + B) = P^{-1}(\alpha A + B)P = \alpha P^{-1}AP + P^{-1}BP = \alpha S(A) + S(B).$$

Example 4.3. Let V be a real inner product space, and let W be any subspace of V. By Theorem 1.22 we have $V = W \oplus W^\perp$, and hence by Theorem 1.12, any $v \in V$ has a unique decomposition $v = x + y$ where $x \in W$ and $y \in W^\perp$. Now define the mapping $T : V \to W$ by $T(v) = x$. Then

$$T(v_1 + v_2) = x_1 + x_2 = T(v_1) + T(v_2)$$

and

$$T(av) = ax = aT(v)$$

so that T is a linear transformation. This mapping is called the **orthogonal projection** of V onto W.

Example 4.4. Let $A \in M_n(\mathcal{F})$ be a fixed matrix. Then the mapping T_A defined by $T_A(X) = AX$ for $X \in \mathcal{F}^n$ is clearly a linear transformation from $\mathcal{F}^n \to \mathcal{F}^n$.

Let $T : U \to V$ be a linear transformation, and let $\{e_i\}$ be a basis for U. Then for any $x \in U$ we have $x = \sum x_i e_i$, and hence

$$T(x) = T\Big(\sum x_i e_i\Big) = \sum x_i T(e_i).$$

Therefore, if we know all of the $T(e_i)$, then we know $T(x)$ for any $x \in U$. In other words, *a linear transformation is determined by specifying its values on a basis*. Our first theorem formalizes this fundamental observation.

Theorem 4.1. *Let U and V be finite-dimensional vector spaces over $\mathcal{F}$, and let $\{e_1, \ldots, e_n\}$ be a basis for U. If $v_1, \ldots, v_n$ are any n arbitrary vectors in V, then there exists a unique linear transformation $T : U \to V$ such that $T(e_i) = v_i$ for each $i = 1, \ldots, n$.*

Proof. For any $x \in U$ we have $x = \sum_{i=1}^n x_i e_i$ for some unique set of scalars x_i (Theorem 1.4, Corollary 2). We define the mapping T by

$$T(x) = \sum_{i=1}^n x_i v_i$$

for any $x \in U$. Since the x_i are unique, this mapping is well-defined (see Exercise 4.1.1). Letting $x = e_i$ in the definition of T and noting that for any $i = 1, \ldots, n$ we have $e_i = \sum_j \delta_{ij} e_j$, it follows that

$$T(e_i) = \sum_{j=1}^n \delta_{ij} v_j = v_i.$$

We show that T so defined is a linear transformation.

If $x = \sum x_i e_i$ and $y = \sum y_i e_i$, then $x + y = \sum (x_i + y_i) e_i$, and hence

$$T(x+y) = \sum (x_i + y_i) v_i = \sum x_i v_i + \sum y_i v_i = T(x) + T(y).$$

Also, if $c \in \mathcal{F}$ then $cx = \sum (cx_i) e_i$, and thus

$$T(cx) = \sum (cx_i) v_i = c \sum x_i v_i = cT(x)$$

which shows that T is indeed a linear transformation.

Now suppose that $T' : U \to V$ is any other linear transformation defined by $T'(e_i) = v_i$. Then for any $x \in U$ we have

$$\begin{aligned} T'(x) &= T'\Big(\sum x_i e_i\Big) = \sum x_i T'(e_i) = \sum x_i v_i \\ &= \sum x_i T(e_i) = T\Big(\sum x_i e_i\Big) = T(x) \end{aligned}$$

and hence $T'(x) = T(x)$ for all $x \in U$. This means that $T' = T$ which thus proves uniqueness. ∎

Example 4.5. Let $T : \mathcal{F}^m \to \mathcal{F}^n$ be a linear transformation, and let $\{e_1, \ldots, e_m\}$ be the standard basis for $\mathcal{F}^m$. We may uniquely define T by specifying any m vectors $v_1, \ldots, v_m$ in $\mathcal{F}^n$. In other words, we define T by the requirement $T(e_i) = v_i$ for each $i = 1, \ldots, m$. Since for any $x \in \mathcal{F}^m$ we have $x = \sum_{i=1}^m x_i e_i$, the linearity of T means

$$T(x) = \sum_{i=1}^m x_i v_i.$$

In terms of the standard basis $\{f_1, \ldots, f_n\}$ for $\mathcal{F}^n$, each v_i has components $\{v_{1i}, \ldots, v_{ni}\}$ defined by $v_i = \sum_{j=1}^n f_j v_{ji}$. (Note the order of indices in this equation.) Now define the matrix $A = (a_{ij}) \in M_{n \times m}(\mathcal{F})$ with column vectors given by $A^i = v_i \in \mathcal{F}^n$. In other words (remember these are columns),

$$A^i = (a_{1i}, \ldots, a_{ni}) = (v_{1i}, \ldots, v_{ni}) = v_i.$$

Writing out $T(x)$ we have

$$T(x) = \sum_{i=1}^m x_i v_i = x_1 \begin{bmatrix} v_{11} \\ \vdots \\ v_{n1} \end{bmatrix} + \cdots + x_m \begin{bmatrix} v_{1m} \\ \vdots \\ v_{nm} \end{bmatrix} = \begin{bmatrix} v_{11}x_1 + \cdots + v_{1m}x_m \\ \vdots \\ v_{n1}x_1 + \cdots + v_{nm}x_m \end{bmatrix}$$

and therefore, in terms of the matrix A, our transformation takes the form

$$T(x) = \begin{bmatrix} v_{11} & \cdots & v_{1m} \\ \vdots & & \vdots \\ v_{n1} & \cdots & v_{nm} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_m \end{bmatrix}.$$

We have therefore constructed an explicit matrix representation of the transformation T. We shall have much more to say about such matrix representations shortly.

Given vector spaces U and V, we claim that the set of all linear transformations from U to V can itself be made into a vector space. To accomplish this we proceed as follows. If U and V are vector spaces over $\mathcal{F}$ and $f, g : U \to V$ are mappings, we naturally define

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

and

$$(cf)(x) = cf(x)$$

for $x \in U$ and $c \in \mathcal{F}$. In addition, if $h : V \to W$ (where W is another vector space over $\mathcal{F}$), then we may define the composite mapping $h \circ g : U \to W$ in the usual way by

$$(h \circ g)(x) = h(g(x)).$$

Theorem 4.2. *Let U, V and W be vector spaces over $\mathcal{F}$, let $c \in \mathcal{F}$ be any scalar, and let $f, g : U \to V$ and $h : V \to W$ be linear transformations. Then the mappings $f + g$, cf, and $h \circ g$ are all linear transformations.*

Proof. First, we see that for $x, y \in U$ and $c \in \mathcal{F}$ we have

$$\begin{aligned}(f+g)(x+y) &= f(x+y) + g(x+y)\\ &= f(x) + f(y) + g(x) + g(y)\\ &= (f+g)(x) + (f+g)(y)\end{aligned}$$

and

$$\begin{aligned}(f+g)(cx) &= f(cx) + g(cx) = cf(x) + cg(x) = c[f(x) + g(x)]\\ &= c(f+g)(x)\end{aligned}$$

and hence $f + g$ is a linear transformation. The proof that cf is a linear transformation is left to the reader (Exercise 4.1.3). Finally, we see that

$$\begin{aligned}(h \circ g)(x+y) &= h(g(x+y)) = h(g(x) + g(y)) = h(g(x)) + h(g(y))\\ &= (h \circ g)(x) + (h \circ g)(y)\end{aligned}$$

and

$$(h \circ g)(cx) = h(g(cx)) = h(cg(x)) = ch(g(x)) = c(h \circ g)(x)$$

so that $h \circ g$ is also a linear transformation. ∎

We define the **zero mapping** $0 : U \to V$ by $0x = 0$ for all $x \in U$. Since

$$0(x+y) = 0 = 0x + 0y$$

and

$$0(cx) = 0 = c(0x)$$

it follows that the zero mapping is a linear transformation. Next, given a mapping $f : U \to V$, we define its **negative** $-f : U \to V$ by $(-f)(x) = -f(x)$ for all $x \in U$. If f is a linear transformation, then $-f$ is also linear because cf is linear for any $c \in \mathcal{F}$ and $-f = (-1)f$. Lastly, we note that

$$\begin{aligned}[f + (-f)](x) &= f(x) + (-f)(x) = f(x) + [-f(x)] = f(x) + f(-x)\\ &= f(x - x) = f(0)\\ &= 0\end{aligned}$$

for all $x \in U$ so that $f + (-f) = (-f) + f = 0$ for all linear transformations f.

With all of this algebra out of the way, we are now in a position to easily prove our claim.

Theorem 4.3. *Let U and V be vector spaces over $\mathcal{F}$. Then the set $L(U,V)$ of all linear transformations of U to V with addition and scalar multiplication defined as above is a linear vector space over $\mathcal{F}$.*

Proof. We leave it to the reader to show that the set of all such linear transformations obeys the properties (VS1)–(VS8) given in Section 1.2 (see Exercise 4.1.4). ∎

We denote the vector space defined in Theorem 4.3 by $L(U,V)$. (Some authors denote this space by $\operatorname{Hom}(U,V)$ since a linear transformation is just a vector space homomorphism). The space $L(U,V)$ is often called the space of **linear transformations** (or **mappings**). In the particular case that U and V are finite-dimensional, we have the following important result.

Theorem 4.4. *Let $\dim U = m$ and $\dim V = n$. Then*

$$\dim L(U,V) = (\dim U)(\dim V) = mn.$$

Proof. We prove the theorem by exhibiting a basis for $L(U,V)$ that contains mn elements. Let $\{e_1,\dots,e_m\}$ be a basis for U, and let $\{\bar{e}_1,\dots,\bar{e}_n\}$ be a basis for V. Define the mn linear transformations $E^i{}_j \in L(U,V)$ by

$$E^i{}_j(e_k) = \delta^i{}_k \bar{e}_j$$

where $i,k = 1,\dots,m$ and $j = 1,\dots,n$. Theorem 4.1 guarantees that the mappings $E^i{}_j$ are unique. To show that $\{E^i{}_j\}$ is a basis, we must show that it is linearly independent and spans $L(U,V)$.

If

$$\sum_{i=1}^m \sum_{j=1}^n a^j{}_i E^i{}_j = 0$$

for some set of scalars $a^j{}_i$, then for any e_k we have

$$0 = \sum_{i,j} a^j{}_i E^i{}_j(e_k) = \sum_{i,j} a^j{}_i \delta^i{}_k \bar{e}_j = \sum_j a^j{}_k \bar{e}_j.$$

But the $\bar{e}_j$ are a basis and hence linearly independent, and thus we must have $a^j{}_k = 0$ for every $j = 1,\dots,n$ and $k = 1,\dots,m$. This shows that the $E^i{}_j$ are linearly independent.

Now suppose $f \in L(U,V)$ and let $x \in U$. Then $x = \sum x^i e_i$ and

$$f(x) = f\left(\sum x^i e_i\right) = \sum x^i f(e_i).$$

Since $f(e_i) \in V$, we must have $f(e_i) = \sum_j c^j{}_i \bar{e}_j$ for some set of scalars $c^j{}_i$, and hence

$$f(e_i) = \sum_j c^j{}_i \bar{e}_j = \sum_{j,k} c^j{}_k \delta^k{}_i \bar{e}_j = \sum_{j,k} c^j{}_k E^k{}_j(e_i).$$

But this means that $f = \sum_{j,k} c^j{}_k E^k{}_j$ (Theorem 4.1), and therefore $\{E^k{}_j\}$ spans $L(U, V)$. ∎

Suppose we have a linear mapping $\phi : V \to \mathcal{F}$ of a vector space V to the field of scalars. By definition, this means that

$$\phi(ax + by) = a\phi(x) + b\phi(y)$$

for every $x, y \in V$ and $a, b \in \mathcal{F}$. The mapping ϕ is called a **linear functional** on V.

Example 4.6. Consider the space $M_n(\mathcal{F})$ of n-square matrices over $\mathcal{F}$. Since the trace of any $A = (a_{ij}) \in M_n(\mathcal{F})$ is defined by

$$\operatorname{tr} A = \sum_{i=1}^{n} a_{ii}$$

(see Exercise 2.5.7), it is easy to show that tr defines a linear functional on $M_n(\mathcal{F})$ (Exercise 4.1.5).

Example 4.7. Let $C[a, b]$ denote the space of all real-valued continuous functions defined on the interval $[a, b]$ (see Exercise 1.2.6). We may define a linear functional L on $C[a, b]$ by

$$L(f) = \int_a^b f(x)\, dx$$

for every $f \in C[a, b]$. It is also left to the reader (Exercise 4.1.5) to show that this does indeed define a linear functional on $C[a, b]$.

Let V be a vector space over $\mathcal{F}$. Since $\mathcal{F}$ is also a vector space over itself, we may consider the space $L(V, \mathcal{F})$. This vector space is the set of all linear functionals on V, and is called the **dual space** of V (or the **space of linear functionals** on V). The dual space is generally denoted by V^*, and its elements are frequently denoted by Greek letters. From the proof of Theorem 4.4, we see that if $\{e_i\}$ is a basis for V, then V^* has a unique basis $\{\omega^j\}$ defined by

$$\omega^j(e_i) = \delta^j{}_i.$$

The basis $\{\omega^j\}$ is referred to as the **dual basis** to the basis $\{e_i\}$. We also see that Theorem 4.4 shows $\dim V^* = \dim V$.

(Let us point out that we make no real distinction between subscripts and superscripts. For our purposes, we use whichever is more convenient from a notational standpoint. However, in tensor analysis and differential geometry, subscripts and superscripts are used precisely to distinguish between a vector space and its dual. We shall follow this convention in Chapter 8.)

Example 4.8. Consider the space $V = \mathcal{F}^n$ of all n-tuples of scalars. If we write any $x \in V$ as a column vector, then V^* is just the space of row vectors. This is because if $\phi \in V^*$ we have

$$\phi(x) = \phi\Big(\sum x_i e_i\Big) = \sum x_i \phi(e_i)$$

where the e_i are the standard (column) basis vectors for $V = \mathcal{F}^n$. Thus, since $\phi(e_i) \in \mathcal{F}$, we see that every $\phi(x)$ is the product of some scalar $\phi(e_i)$ times the scalar x_i, summed over $i = 1, \dots, n$. If we write $\phi(e_i) = a_i$, it then follows that we may write

$$\phi(x) = \phi(x_1, \dots, x_n) = (a_1, \dots, a_n) \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \tag{*}$$

or simply $\phi(x) = \sum a_i x_i$. This expression is in fact the origin of the term "linear form."

Since any row vector in $\mathcal{F}^n$ can be expressed in terms of the basis vectors $\omega^1 = (1, 0, \dots, 0), \dots, \omega^n = (0, 0, \dots, 1)$, we see from (*) that the ω^j do indeed form the basis dual to $\{e_i\}$ since they clearly have the property that $\omega^j(e_i) = \delta^j{}_i$. In other words, the row vector ω^j is just the transpose of the corresponding column vector e_j.

Since U^* is a vector space, the reader may wonder whether or not we may form the space $U^{**} = (U^*)^*$. The answer is "yes," and the space U^{**} is called the **double dual** (or **second dual**) of U. In fact, for finite-dimensional vector spaces, it is essentially true that $U^{**} = U$ (in the sense that U and U^{**} are isomorphic). However, we prefer to postpone our discussion of these matters until a later chapter when we can treat all of this material in the detail that it warrants.

Exercises

1. Verify that the mapping T of Theorem 4.1 is well-defined.

2. Repeat Example 4.5, except now let the matrix $A = (a_{ij})$ have *row* vectors $A_i = v_i \in \mathcal{F}^n$. What is the matrix representation of the operation $T(x)$?

3. Show that cf is a linear transformation in the proof of Theorem 4.2.

4. Prove Theorem 4.3.

5. (a) Show that the function tr defines a linear functional on $M_n(\mathcal{F})$ (see Example 4.6).
   (b) Show that the mapping L defined in Example 4.7 defines a linear functional.

6. Explain whether or not each of the following mappings f is linear:

   (a) $f : \mathbb{R}^2 \to \mathbb{R}$ defined by $f(x, y) = |x - y|$.
   (b) $f : \mathbb{R}^2 \to \mathbb{R}$ defined by $f(x, y) = xy$.
   (c) $f : \mathbb{R}^2 \to \mathbb{R}^2$ defined by $f(x, y) = (x + y, x)$.
   (d) $f : \mathbb{R}^2 \to \mathbb{R}^2$ defined by $f(x, y) = (\sin x, y)$.
   (e) $f : \mathbb{R}^2 \to \mathbb{R}^3$ defined by $f(x, y) = (x + 1, 2y, x + y)$.
   (f) $f : \mathbb{R}^3 \to \mathbb{R}$ defined by $f(x, y, z) = 2x - 3y + 4z$.
   (g) $f : \mathbb{R}^3 \to \mathbb{R}^2$ defined by $f(x, y, z) = (|x| , 0)$.
   (h) $f : \mathbb{R}^3 \to \mathbb{R}^3$ defined by $f(x, y, z) = (1, -x, y + z)$.

7. Let $T : U \to V$ be a bijective linear transformation. Define T^{-1} and show that it is also a linear transformation.

8. Let $T : U \to V$ be a linear transformation, and suppose that we have the set of vectors $u_1, \dots, u_n \in U$ with the property that $T(u_1), \dots, T(u_n) \in V$ is linearly independent. Show that $\{u_1, \dots, u_n\}$ is linearly independent.

9. Let $B \in M_n(\mathcal{F})$ be arbitrary. Show that the mapping $T : M_n(\mathcal{F}) \to M_n(\mathcal{F})$ defined by $T(A) = [A, B]_+ = AB + BA$ is linear. Is the same true for the mapping $T'(A) = [A, B] = AB - BA$?

10. Let $T : \mathcal{F}^2 \to \mathcal{F}^2$ be the linear transformation defined by the system

$$\begin{aligned} y_1 &= -3x_1 + x_2 \\ y_2 &= x_1 - x_2 \end{aligned}$$

and let S be the linear transformation defined by the system

$$\begin{aligned} y_1 &= x_1 + x_2 \\ y_2 &= x_1 \end{aligned}$$

Find a system of equations that defines each of the following linear transformations:

(a) $2T$ (b) $T - S$ (c) ST (d) TS
(e) T^2 (f) $T^2 + 2S$

11. Does there exist a linear transformation $T : \mathbb{R}^3 \to \mathbb{R}^2$ with the property that $T(1, -1, 1) = (1, 0)$ and $T(1, 1, 1) = (0, 1)$?

12. Suppose $u_1 = (1, -1)$, $u_2 = (2, -1)$, $u_3 = (-3, 2)$ and $v_1 = (1, 0)$, $v_2 = (0, 1)$, $v_3 = (1, 1)$. Does there exist a linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$ with the property that $Tu_i = v_i$ for each $i = 1, 2,$ and 3?

13. Find $T(x, y, z)$ if $T : \mathbb{R}^3 \to \mathbb{R}$ is the linear transformation defined by $T(1, 1, 1) = 3$, $T(0, 1, -2) = 1$ and $T(0, 0, 1) = -2$.

14. Let V be the set of all complex numbers considered as a vector space over the real field. Find a mapping $T : V \to V$ that is a linear transformation on V, but is not a linear transformation on the space $\mathbb{C}^1$ (i.e., the set of complex numbers considered as a complex vector space).

15. If V is finite-dimensional and $x_1, x_2 \in V$ with $x_1 \neq x_2$, prove there exists a linear functional $f \in V^*$ such that $f(x_1) \neq f(x_2)$.

4.2 Further Properties of Linear Transformations

Suppose $T \in L(U, V)$ where U and V are finite-dimensional over $\mathcal{F}$. We define the **image** of T to be the set

$$\operatorname{Im} T = \{T(x) \in V : x \in U\}$$

and the **kernel** of T to be the set

$$\operatorname{Ker} T = \{x \in U : T(x) = 0\}.$$

(Many authors call $\operatorname{Im} T$ the **range** of T, but we use this term to mean the space V in which T takes its values. Note also that this definition of the kernel is almost the same as we used for the null space $\ker A$ of a matrix A in Section 2.5. In fact, after we have defined the matrix representation of a linear transformation in the next section, we will see that these definitions are exactly the same.)

Since $T(0) = 0 \in V$, we see that $0 \in \operatorname{Im} T$, and hence $\operatorname{Im} T \neq \varnothing$. Now suppose $x', y' \in \operatorname{Im} T$. Then there exist $x, y \in U$ such that $T(x) = x'$ and $T(y) = y'$. Then for any $a, b \in \mathcal{F}$ we have

$$ax' + by' = aT(x) + bT(y) = T(ax + by) \in \operatorname{Im} T$$

(since $ax + by \in U$), and thus $\operatorname{Im} T$ is a subspace of V. Similarly, we see that $0 \in \operatorname{Ker} T$, and if $x, y \in \operatorname{Ker} T$ then

$$T(ax + by) = aT(x) + bT(y) = 0$$

so that $\operatorname{Ker} T$ is also a subspace of U. $\operatorname{Ker} T$ is frequently called the **null space** of T.

We now restate Theorem 1.5 in our current terminology.

Theorem 4.5. *A linear transformation $T \in L(U, V)$ is an isomorphism if and only if* $\operatorname{Ker} T = \{0\}$.

For example, the projection mapping T defined in Example 4.1 is not an isomorphism because $T(0,0,z) = (0,0,0)$ for all $(0,0,z) \in \mathbb{R}^3$. In fact, if x_0 and y_0 are fixed, then we have $T(x_0, y_0, z) = (x_0, y_0, 0)$ independently of z.

If $T \in L(U, V)$, we define the **rank** of T to be the number

$$\operatorname{rank} T = \dim(\operatorname{Im} T)$$

and the **nullity** of T to be the number

$$\operatorname{nul} T = \dim(\operatorname{Ker} T).$$

We will shortly show that this definition of rank is essentially the same as our previous definition of the rank of a matrix. The relationship between $\operatorname{rank} T$ and $\operatorname{nul} T$ is given in the following important result, sometimes called the **rank theorem** (or the **dimension theorem**).

Theorem 4.6. *If U and V are finite-dimensional over $\mathcal{F}$ and $T \in L(U, V)$, then*

$$\operatorname{rank} T + \operatorname{nul} T = \dim U$$

or, alternatively,

$$\dim(\operatorname{Im} T) + \dim(\operatorname{Ker} T) = \dim U.$$

Proof. Let $\{u_1, \dots, u_n\}$be a basis for U and suppose $\operatorname{Ker} T = \{0\}$. Then for any $x \in U$ we have

$$T(x) = T\Big(\sum x_i u_i\Big) = \sum x_i T(u_i)$$

for some set of scalars x_i, and therefore $\{T(u_i)\}$ spans $\operatorname{Im} T$. If $\sum c_i T(u_i) = 0$, then

$$0 = \sum c_i T(u_i) = \sum T(c_i u_i) = T\Big(\sum c_i u_i\Big)$$

which implies $\sum c_i u_i = 0$ (since $\operatorname{Ker} T = \{0\}$). But the u_i are linearly independent so that we must have $c_i = 0$ for every i, and hence $\{T(u_i)\}$ is linearly independent. Since $\operatorname{nul} T = \dim(\operatorname{Ker} T) = 0$ and $\operatorname{rank} T = \dim(\operatorname{Im} T) = n = \dim U$, we see that $\operatorname{rank} T + \operatorname{nul} T = \dim U$.

Now suppose that $\operatorname{Ker} T \neq \{0\}$, and let $\{w_1, \dots, w_k\}$ be a basis for $\operatorname{Ker} T$. By Theorem 1.10, we may extend this to a basis $\{w_1, \dots, w_n\}$ for U. Since $T(w_i) = 0$ for each $i = 1, \dots, k$ it follows that the vectors $T(w_{k+1}), \dots, T(w_n)$ span $\operatorname{Im} T$. If

$$\sum_{j=k+1}^{n} c_j T(w_j) = 0$$

for some set of scalars c_j, then

$$0 = \sum_{j=k+1}^{n} c_j T(w_j) = \sum_{j=k+1}^{n} T(c_j w_j) = T\left(\sum_{j=k+1}^{n} c_j w_j\right)$$

so that $\sum_{j=k+1}^{n} c_j w_j \in \operatorname{Ker} T$. This means that

$$\sum_{j=k+1}^{n} c_j w_j = \sum_{j=1}^{k} a_j w_j$$

for some set of scalars a_j. But this is just

$$\sum_{j=1}^{k} a_j w_j - \sum_{j=k+1}^{n} c_j w_j = 0$$

and hence

$$a_1 = \cdots = a_k = c_{k+1} = \cdots = c_n = 0$$

since the w_j are linearly independent. Therefore $T(w_{k+1}), \dots, T(w_n)$ are linearly independent and thus form a basis for $\operatorname{Im} T$. We have therefore shown that

$$\dim U = k + (n - k) = \dim(\operatorname{Ker} T) + \dim(\operatorname{Im} T) = \operatorname{nul} T + \operatorname{rank} T. \qquad \blacksquare$$

An extremely important special case of the space $L(U, V)$ is the space $L(V, V)$ of all linear transformations of V into itself. This space is frequently written as $L(V)$, and its elements are usually called **linear operators** on V, or simply **operators**. We will have much more to say about operators in a later chapter.

A linear transformation $T \in L(U, V)$ is said to be **invertible** if there exists a linear transformation $T^{-1} \in L(V, U)$ such that $TT^{-1} = T^{-1}T = I$. (Note that technically TT^{-1} is the identity on V and $T^{-1}T$ is the identity on U, but the meaning of this statement should be clear.) This is exactly the same definition we had in Section 2.6 for matrices. The unique mapping T^{-1} is called the **inverse** of T.

We now show that if a linear transformation is invertible, then it has the properties that we would expect.

Theorem 4.7. *A linear transformation $T \in L(U, V)$ is invertible if and only if it is a bijection (i.e., one-to-one and onto).*

Proof. First suppose that T is invertible. If $T(x_1) = T(x_2)$ for $x_1, x_2 \in U$, then the fact that $T^{-1}T = I$ implies

$$x_1 = T^{-1}T(x_1) = T^{-1}T(x_2) = x_2$$

and hence T is injective. If $y \in V$, then using $TT^{-1} = I$ we have

$$y = I(y) = (TT^{-1})y = T(T^{-1}(y))$$

so that $y = T(x)$ where $x = T^{-1}(y)$. This shows that T is also surjective, and hence a bijection.

Conversely, let T be a bijection. We must define a linear transformation $T^{-1} \in L(V, U)$ with the desired properties. Let $y \in V$ be arbitrary. Since T is surjective, there exists a vector $x \in U$ such that $T(x) = y$. The vector x is unique because T is injective. We may therefore define a mapping $T^{-1} : V \to U$ by the rule $T^{-1}(y) = x$ where $y = T(x)$. To show that T^{-1} is linear, let $y_1, y_2 \in V$ be arbitrary and choose $x_1, x_2 \in U$ such that $T(x_1) = y_1$ and $T(x_2) = y_2$. Using the linearity of T we then see that

$$T(x_1 + x_2) = T(x_1) + T(x_2) = y_1 + y_2$$

and hence

$$T^{-1}(y_1 + y_2) = x_1 + x_2.$$

But then

$$T^{-1}(y_1 + y_2) = x_1 + x_2 = T^{-1}(y_1) + T^{-1}(y_2).$$

Similarly, if $T(x) = y$ and $a \in \mathcal{F}$, then $T(ax) = aT(x) = ay$ so that

$$T^{-1}(ay) = ax = aT^{-1}(y).$$

We have thus shown that $T^{-1} \in L(V, U)$. Finally, we note that for any $y \in V$ and $x \in U$ such that $T(x) = y$ we have

$$TT^{-1}(y) = T(x) = y$$

and

$$T^{-1}T(x) = T^{-1}(y) = x$$

so that $TT^{-1} = T^{-1}T = I$. ∎

A linear transformation $T \in L(U, V)$ is said to be **nonsingular** if $\operatorname{Ker} T = \{0\}$. In other words, T is nonsingular if it is one-to-one (Theorem 4.5). As we might expect, T is said to be **singular** if it is not nonsingular. In other words, T is singular if $\operatorname{Ker} T \neq \{0\}$.

Now suppose U and V are both finite-dimensional and $\dim U = \dim V$. If $\operatorname{Ker} T = \{0\}$, then $\operatorname{nul} T = 0$ and the rank theorem (Theorem 4.6) shows that $\dim U = \dim(\operatorname{Im} T)$. In other words, we must have $\operatorname{Im} T = V$, and hence T is surjective. Conversely, if T is surjective then we are forced to conclude that $\operatorname{nul} T = 0$, and thus T is also injective. Hence a linear transformation between two finite-dimensional vector spaces of the same dimension is one-to-one if and only if it is onto. Combining this discussion with Theorem 4.7, we obtain the following result and its obvious corollary.

Theorem 4.8. *Let U and V be finite-dimensional vector spaces with $\dim U = \dim V$. Then the following statements are equivalent for any linear transformation $T \in L(U, V)$:*

(i) T is invertible.

(ii) T is nonsingular.

(iii) T is surjective.

Corollary. *A linear operator $T \in L(V)$ on a finite-dimensional vector space is invertible if and only if it is nonsingular.*

Example 4.9. Let $V = \mathcal{F}^n$ so that any $x \in V$ may be written in terms of components as $x = (x_1, \dots, x_n)$. Given any matrix $A = (a_{ij}) \in M_{m \times n}(\mathcal{F})$, we define a linear transformation $T : \mathcal{F}^n \to \mathcal{F}^m$ by $T(x) = y$ which is again given in component form by

$$y_i = \sum_{j=1}^{n} a_{ij} x_j, \qquad i = 1, \dots, m.$$

We claim that T is one-to-one if and only if the homogeneous system

$$\sum_{j=1}^{n} a_{ij} x_j = 0, \qquad i = 1, \dots, m$$

has only the trivial solution. (Note that if T is one-to-one, this is the same as requiring that the solution of the nonhomogeneous system be unique. It also follows from Corollary 5 of Theorem 2.20 that if T is one-to-one, then A is nonsingular.)

First let T be one-to-one. Clearly $T(0) = 0$, and if $v = (v_1, \dots, v_n)$ is a solution of the homogeneous system, then $T(v) = 0$. But if T is one-to-one, then $v = 0$ is the only solution. Conversely, let the homogeneous system have only the trivial solution. If $T(u) = T(v)$, then

$$0 = T(u) - T(v) = T(u - v)$$

which implies that $u - v = 0$ or $u = v$.

Example 4.10. Let $T \in L(\mathbb{R}^2)$ be defined by

$$T(x, y) = (y, 2x - y).$$

If $T(x, y) = (0, 0)$, then we must have $x = y = 0$, and hence $\operatorname{Ker} T = \{0\}$. By the corollary to Theorem 4.8, T is invertible, and we now show how to find T^{-1}.

Suppose we write $(x', y') = T(x, y) = (y, 2x - y)$. Then $y = x'$ and $2x - y = y'$ so that solving for x and y in terms of x' and y' we obtain $x = (x' + y')/2$ and $y = x'$. We therefore see that

$$T^{-1}(x', y') = (x'/2 + y'/2, x').$$

Note this also shows that T is surjective since for any $(x', y') \in \mathbb{R}^2$ we found a point $(x, y) = (x'/2 + y'/2, x')$ such that $T(x, y) = (x', y')$.

Our next example shows the importance of finite-dimensionality in Theorem 4.8.

Example 4.11. Let $V = \mathcal{F}[x]$, the (infinite-dimensional) space of all polynomials over $\mathcal{F}$ (see Example 1.2). For any $v \in V$ with $v = \sum_{i=0}^{n} a_i x^i$ we define $T \in L(V)$ by

$$T(v) = \sum_{i=0}^{n} a_i x^{i+1}$$

(this is just a "multiplication by x" operation). We leave it to the reader to show that T is linear and nonsingular (see Exercise 4.2.1). However, it is clear that T can not be surjective (for example, T takes scalars into polynomials of degree 1), so it can not be invertible. However, it is nevertheless possible to find a left inverse ${T_L}^{-1}$ for T. To see this, we let ${T_L}^{-1}$ be the operation of subtracting the constant term and then dividing by x:

$${T_L}^{-1}(v) = \sum_{i=1}^{n} a_i x^{i-1}.$$

We again leave it to the reader (Exercise 4.2.1) to show that this is a linear transformation, and that ${T_L}^{-1}T = I$ while $T{T_L}^{-1} \neq I$.

Example 4.12. While the operation T defined above is an example of a nonsingular linear transformation that is not surjective, we can also give an example of a linear transformation on $\mathcal{F}[x]$ that is surjective but not nonsingular. To see this, consider the operation $D = d/dx$ that takes the derivative of every polynomial in $\mathcal{F}[x]$. It is easy to see that D is a linear transformation, but D can not possibly be nonsingular since the derivative of any constant polynomial $p(x) = c$ is zero. Note though, that the image of D is all of $\mathcal{F}[x]$, and it is in fact possible to find a right inverse of D. Indeed, if we let ${D_R}^{-1}(f) = \int_0^x f(t)\,dt$ be the (indefinite) integral operator, then

$${D_R}^{-1}\left(\sum_{i=0}^{n} a_i x^i\right) = \sum_{i=0}^{n} \frac{a_i x^{i+1}}{i+1}$$

and hence $D{D_R}^{-1} = I$. However, it is obvious that ${D_R}^{-1}D \neq I$ because ${D_R}^{-1}D$ applied to a constant polynomial yields zero.

Exercises

1. (a) Verify that the mapping T in Example 4.9 is linear.
   (b) Verify that the mapping T in Example 4.10 is linear.

(c) Verify that the mapping T in Example 4.11 is linear and nonsingular.
(d) Verify that $T{T_L}^{-1} \neq I$ in Example 4.11.

2. Find a linear transformation $T : \mathbb{R}^3 \to \mathbb{R}^4$ whose image is generated by the vectors $(1, 2, 0, -4)$ and $(2, 0, -1, -3)$.

3. For each of the following linear transformations T, find the dimension and a basis for $\operatorname{Im} T$ and $\operatorname{Ker} T$:
   (a) $T : \mathbb{R}^3 \to \mathbb{R}^3$ defined by
   $$T(x, y, z) = (x + 2y - z, y + z, x + y - 2z).$$
   (b) $T : \mathbb{R}^4 \to \mathbb{R}^3$ defined by
   $$T(x, y, z, t) = (x - y + z + t, x + 2z - t, x + y + 3z - 3t).$$

4. Consider the space $M_2(\mathbb{R})$ of real 2×2 matrices, and define the matrix
   $$B = \begin{bmatrix} 1 & 2 \\ 0 & 3 \end{bmatrix}.$$
   Find the dimension and exhibit a specific basis for the kernel of the linear transformation $T : M_2(\mathbb{R}) \to M_2(\mathbb{R})$ defined by $T(A) = AB - BA = [A, B]$.

5. Show that a linear transformation is nonsingular if and only if it takes linearly independent sets into linearly independent sets.

6. Consider the operator $T : \mathbb{R}^3 \to \mathbb{R}^3$ defined by
   $$T(x, y, z) = (2x, 4x - y, 2x + 3y - z).$$
   (a) Show that T is invertible.
   (b) Find a formula for T^{-1} .

7. Let E be a **projection** (or **idempotent**) operator on a space V, i.e., $E^2 = E$ on V. Define $U = \operatorname{Im} E$ and $W = \operatorname{Ker} E$. Show that:
   (a) $E(u) = u$ for every $u \in U$.
   (b) If $E \neq I$, then E is singular.
   (c) $V = U \oplus W$.

8. If $S : U \to V$ and $T : V \to U$ are nonsingular linear transformations, show that ST is nonsingular. What can be said if S and/or T is singular?

9. Let $S : U \to V$ and $T : V \to W$ be linear transformations.
   (a) Show that $TS : U \to W$ is linear.
   (b) Show that $\operatorname{rank}(TS) \leq \operatorname{rank} T$ and $\operatorname{rank}(TS) \leq \operatorname{rank} S$, i.e., $\operatorname{rank}(TS) \leq \min\{\operatorname{rank} T, \operatorname{rank} S\}$.

10. If $S, T \in L(V)$ and S is nonsingular, show that $\operatorname{rank}(ST) = \operatorname{rank}(TS) = \operatorname{rank} T$.

11. If $S, T \in L(U, V)$, show that $\operatorname{rank}(S + T) \leq \operatorname{rank} S + \operatorname{rank} T$. Give an example of two nonzero linear transformations $S, T \in L(U, V)$ such that $\operatorname{rank}(S + T) = \operatorname{rank} S + \operatorname{rank} T$.

12. Suppose that $V = U \oplus W$ and consider the linear operators E_1 and E_2 on V defined by $E_1(v) = u$ and $E_2(v) = w$ where $u \in U$, $w \in W$ and $v = u + w$. Show that:
    (a) E_1 and E_2 are projection operators on V (see Exercise 7 above).
    (b) $E_1 + E_2 = I$.
    (c) $E_1 E_2 = 0 = E_2 E_1$.
    (d) $V = \operatorname{Im} E_1 \oplus \operatorname{Im} E_2$.

13. Prove that the nonsingular elements in $L(V)$ form a group.

14. Recall that an operator $T \in L(V)$ is said to be **nilpotent** if $T^n = 0$ for some positive integer n. Suppose that T is nilpotent and $T(x) = \alpha x$ for some nonzero $x \in V$ and some $\alpha \in \mathcal{F}$. Show that $\alpha = 0$.

15. If $\dim V = 1$, show that $L(V)$ is isomorphic to $\mathcal{F}$.

16. Let $V = \mathbb{C}^3$ have the standard basis $\{e_i\}$, and let $T \in L(V)$ be defined by $T(e_1) = (1, 0, i)$, $T(e_2) = (0, 1, 1)$ and $T(e_3) = (i, 1, 0)$. Is T invertible?

17. Let V be finite-dimensional, and suppose $T \in L(V)$ has the property that $\operatorname{rank}(T^2) = \operatorname{rank} T$. Show that $(\operatorname{Im} T) \cap (\operatorname{Ker} T) = \{0\}$.

4.3 Matrix Representations

By now it should be apparent that there seems to be a definite similarity between Theorems 4.6 and 2.10. This is indeed the case, but to formulate this relationship precisely, we must first describe the representation of a linear transformation by matrices.

Consider a linear transformation $T \in L(U, V)$, and let U and V have bases $\{u_1, \dots, u_n\}$ and $\{v_1, \dots, v_m\}$ respectively. Since $T(u_i) \in V$, it follows from Corollary 2 of Theorem 1.4 that there exists a unique set of scalars $a_{1i}, \dots, a_{mi}$ such that

$$T(u_i) = \sum_{j=1}^{m} v_j a_{ji} \tag{4.1}$$

for each $i = 1, \dots, n$. Thus, the linear transformation T leads in a natural way to a matrix (a_{ij}) defined with respect to the given bases. On the other hand, if we are given a matrix (a_{ij}), then $\sum_{j=1}^{m} v_j a_{ji}$ is a vector in V for each $i = 1, \dots, n$. Hence, by Theorem 4.1, there exists a unique linear transformation T defined by $T(u_i) = \sum_{j=1}^{m} v_j a_{ji}$.

Now let x be any vector in U. Then $x = \sum_{i=1}^n x_i u_i$ so that

$$T(x) = T\left(\sum_{i=1}^n x_i u_i\right) = \sum_{i=1}^n x_i T(u_i) = \sum_{i=1}^n \sum_{j=1}^m v_j a_{ji} x_i .$$

But $T(x) \in V$ so we may write

$$y = T(x) = \sum_{j=1}^m y_j v_j .$$

Since $\{v_i\}$ is a basis for V, comparing these last two equations shows that

$$y_j = \sum_{i=1}^n a_{ji} x_i$$

for each $j = 1, \ldots, m$. The reader should note which index is summed over in this expression for y_j.

If we write out both of the systems $T(u_i) = \sum_{j=1}^m v_j a_{ji}$ and $y_j = \sum_{i=1}^n a_{ji} x_i$, we have

$$\begin{aligned} T(u_1) &= a_{11}v_1 + \cdots + a_{m1}v_m \\ &\vdots \\ T(u_n) &= a_{1n}v_1 + \cdots + a_{mn}v_m \end{aligned} \tag{4.2}$$

and

$$\begin{aligned} y_1 &= a_{11}x_1 + \cdots + a_{1n}x_n \\ &\vdots \\ y_m &= a_{m1}x_1 + \cdots + a_{mn}x_n \end{aligned} \tag{4.3}$$

We thus see that the matrix of coefficients in (4.2) is the transpose of the matrix of coefficients in (4.3). We shall call the $m \times n$ matrix of coefficients in equations (4.3) the **matrix representation** of the linear transformation T, and we say that T is **represented** by the matrix $A = (a_{ij})$ with respect to the given (ordered) bases $\{u_i\}$ and $\{v_i\}$.

We will sometimes use the notation $[A]$ to denote the matrix corresponding to an operator $A \in L(U, V)$. This will avoid the confusion that may arise when the same letter is used to denote both the transformation and its representation matrix. In addition, if the particular bases chosen are important, then we will write the matrix representation of the above transformation as $[A]_u^v$, and if $A \in L(V)$, then we write simply $[A]_v$.

In order to make these definitions somewhat more transparent, let us make the following observation. If $x \in U$ has coordinates $(x_1, \ldots, x_n)$ relative to a basis for U, and $y \in V$ has coordinates $(y_1, \ldots, y_m)$ relative to a basis for V, then the expression $y = A(x)$ may be written in matrix form as $Y = [A]X$ where both X and Y are column vectors. In other words, $[A]X$ is the coordinate vector corresponding to the result of the transformation A acting on the vector x. An equivalent way of writing this that emphasizes the bases involved is

$$[y]_v = [A(x)]_v = [A]_u^v [x]_u .$$

If $\{v_j\}$ is a basis for V, then we may clearly write

$$v_i = \sum_j v_j \delta_{ji}$$

where the δ_{ji} are now to be interpreted as the components of v_i *with respect to the basis* $\{v_j\}$. In other words, v_1 has components $(1, 0, \dots, 0)$, v_2 has components $(0, 1, \dots, 0)$ and so forth. Hence, writing out $[A(u_1)]_v = \sum_{j=1}^m v_j a_{j1}$, we see that

$$[A(u_1)]_v = \begin{bmatrix} a_{11} \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ a_{21} \\ \vdots \\ 0 \end{bmatrix} + \cdots + \begin{bmatrix} 0 \\ 0 \\ \vdots \\ a_{m1} \end{bmatrix} = \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{bmatrix}$$

so that $[A(u_1)]_v$ is just the first column of $[A]_u^v$. Similarly, it is easy to see that in general, $[A(u_i)]_v$ is the ith column of $[A]_u^v$. In other words, *the matrix representation* $[A]_u^v$ *of a linear transformation* $A \in L(U, V)$ *has columns that are nothing more than the images under* A *of the basis vectors of* U.

We summarize this very important discussion as a theorem for easy reference.

Theorem 4.9. *Let* U *and* V *have bases* $\{u_1, \dots, u_n\}$ *and* $\{v_1, \dots, v_m\}$ *respectively. Then for any* $A \in L(U, V)$ *the vector*

$$[A(u_i)]_v = \sum_{j=1}^m v_j a_{ji}$$

is the i*th column of the matrix* $[A]_u^v = (a_{ji})$ *that represents* A *relative to the given bases.*

Example 4.13. Let V have a basis $\{v_1, v_2, v_3\}$, and let $A \in L(V)$ be defined by

$$\begin{aligned} A(v_1) &= 3v_1 \phantom{{}- 2v_2} + v_3 \\ A(v_2) &= v_1 - 2v_2 - v_3 \\ A(v_3) &= v_2 + v_3 \end{aligned}$$

Then the representation of A (relative to this basis) is

$$[A]_v = \begin{bmatrix} 3 & 1 & 0 \\ 0 & -2 & 1 \\ 1 & -1 & 1 \end{bmatrix}.$$

The reader may be wondering why we wrote $A(u_i) = \sum_j v_j a_{ji}$ rather than $A(u_i) = \sum_j a_{ij} v_j$. The reason is that we want the matrix corresponding to a combination of linear transformations to be the product of the individual matrix representations taken in the same order. (The argument that follows is based on what we learned in Chapter 2 about matrix multiplication, even though technically we have not yet defined this operation within the framework of our current discussion. In fact, our present formulation can be taken as the *definition* of matrix multiplication.)

To see what this means, suppose $A, B \in L(V)$. If we had written (note the order of subscripts) $A(v_i) = \sum_j a_{ij} v_j$ and $B(v_i) = \sum_j b_{ij} v_j$, then we would have found that

$$\begin{aligned}(AB)(v_i) = A(B(v_i)) &= A\biggl(\sum_j b_{ij} v_j\biggr) = \sum_j b_{ij} A(v_j) \\ &= \sum_{j,k} b_{ij} a_{jk} v_k = \sum_k c_{ik} v_k\end{aligned}$$

where $c_{ik} = \sum_j b_{ij} a_{jk}$. As a matrix product, we would then have $[C] = [B][A]$. However, if we write (as we did) $A(v_i) = \sum_j v_j a_{ji}$ and $B(v_i) = \sum_j v_j b_{ji}$, then we obtain

$$\begin{aligned}(AB)(v_i) = A(B(v_i)) &= A\biggl(\sum_j v_j b_{ji}\biggr) = \sum_j A(v_j) b_{ji} \\ &= \sum_{j,k} v_k a_{kj} b_{ji} = \sum_k v_k c_{ki}\end{aligned}$$

where now $c_{ki} = \sum_j a_{kj} b_{ji}$. Since the matrix notation for this is $[C] = [A][B]$, we see that the order of the matrix representation of transformations is preserved as desired. We have therefore proven the following result.

Theorem 4.10. *For any operators $A, B \in L(V)$ we have $[AB] = [A][B]$.*

From equation (4.3) above, we see that any nonhomogeneous system of m linear equations in n unknowns defines an $m \times n$ matrix (a_{ij}). According to our discussion, this matrix should also define a linear transformation in a consistent manner.

Example 4.14. Consider the space $\mathbb{R}^2$ with the standard basis

$$e_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \qquad e_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

so that any $X \in \mathbb{R}^2$ may be written as

$$X = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = x_1 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

Suppose we have the system of equations

$$\begin{aligned} y_1 &= 2x_1 - x_2 \\ y_2 &= x_1 + 3x_2 \end{aligned}$$

which we may write in matrix form as $[A]X = Y$ where

$$[A] = \begin{bmatrix} 2 & -1 \\ 1 & 3 \end{bmatrix}.$$

Hence we have a linear transformation $A(x) = [A]X$. In particular,

$$A(e_1) = \begin{bmatrix} 2 & -1 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \end{bmatrix} = 2e_1 + e_2$$

$$A(e_2) = \begin{bmatrix} 2 & -1 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} -1 \\ 3 \end{bmatrix} = -e_1 + 3e_2.$$

We now see that letting the ith column of $[A]$ be $A(e_i)$, we arrive back at the original form $[A]$ that represents the linear transformation $A(e_1) = 2e_1 + e_2$ and $A(e_2) = -e_1 + 3e_2$.

Example 4.15. Consider the space $V = \mathbb{R}^2$ with basis vectors $v_1 = (1, 1)$ and $v_2 = (-1, 0)$. Let T be the linear operator on $\mathbb{R}^2$ defined by

$$T(x, y) = (4x - 2y, 2x + y).$$

To find the matrix of T relative to the given basis, all we do is compute the effect of T on each basis vector:

$$\begin{aligned} T(v_1) &= T(1, 1) = (2, 3) = 3v_1 + v_2 \\ T(v_2) &= T(-1, 0) = (-4, -2) = -2v_1 + 2v_2. \end{aligned}$$

Since the matrix of T has columns given by the image of each basis vector, we must have

$$[T] = \begin{bmatrix} 3 & -2 \\ 1 & 2 \end{bmatrix}.$$

Theorem 4.11. *Let U and V be vector spaces over $\mathcal{F}$ with bases $\{u_1, \dots, u_n\}$ and $\{v_1, \dots, v_m\}$ respectively. Suppose $A \in L(U, V)$ and let $[A]$ be the matrix representation of A with respect to the given bases. Then the mapping $\phi : A \to [A]$ is an isomorphism of $L(U, V)$ onto the vector space $M_{m\times n}(\mathcal{F})$ of all $m \times n$ matrices over $\mathcal{F}$.*

Proof. Part of this was proved in the discussion above, but for ease of reference, we repeat it here. Given any $(a_{ij}) \in M_{m\times n}(\mathcal{F})$, we define the linear transformation $A \in L(U, V)$ by

$$A(u_i) = \sum_{j=1}^{m} v_j a_{ji}$$

for each $i = 1, \ldots, n$. According to Theorem 4.1, the transformation A is uniquely defined and is in $L(U, V)$. By definition, $[A] = (a_{ij})$, and hence ϕ is surjective. On the other hand, given any $A \in L(U, V)$, it follows from Corollary 2 of Theorem 1.4 that for each $i = 1, \ldots, n$ there exists a unique set of scalars $a_{1i}, \ldots, a_{mi} \in \mathcal{F}$ such that $A(u_i) = \sum_{j=1}^{m} v_j a_{ji}$. Therefore, any $A \in L(U, V)$ has led to a unique matrix $(a_{ij}) \in M_{m\times n}(\mathcal{F})$. Combined with the previous result that ϕ is surjective, this shows that ϕ is injective and hence a bijection. Another way to see this is to note that if we also have $B \in L(U, V)$ with $[B] = [A]$, then

$$(B - A)(u_i) = B(u_i) - A(u_i) = \sum_{j=1}^{m} v_j(b_{ji} - a_{ji}) = 0.$$

Since $B - A$ is linear (Theorem 4.3), it follows that $(B - A)x = 0$ for all $x \in U$, and hence $B = A$ so that ϕ is one-to-one.

Finally, to show that ϕ is an isomorphism we must show that it is also a vector space homomorphism (i.e., a linear transformation). But this is easy if we simply observe that

$$(A + B)(u_i) = A(u_i) + B(u_i) = \sum_j v_j a_{ji} + \sum_j v_j b_{ji} = \sum_j v_j(a_{ji} + b_{ji})$$

and, for any $c \in \mathcal{F}$,

$$(cA)(u_i) = c(A(u_i)) = c\Bigg(\sum_j v_j a_{ji}\Bigg) = \sum_j v_j(ca_{ji}).$$

Therefore we have shown that

$$[A + B] = [A] + [B]$$

and

$$[cA] = c[A]$$

so that ϕ is a homomorphism. ∎

It may be worth recalling that the space $M_{m\times n}(\mathcal{F})$ is clearly of dimension mn since, for example, we have

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} = a\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + b\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} + c\begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} + d\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}.$$

Therefore Theorem 4.11 provides another proof that $\dim L(U, V) = mn$.

We now return to the relationship between Theorems 4.6 and 2.10. In particular, we would like to know how the rank of a linear transformation is related to the rank of a matrix. The answer was essentially given in Theorem 4.9.

Theorem 4.12. *If $A \in L(U, V)$ is represented by $[A] = (a_{ji}) \in M_{m\times n}(\mathcal{F})$, then* $\operatorname{rank} A = \operatorname{rank}[A]$.

Proof. Recall that $\operatorname{rank} A = \dim(\operatorname{Im} A)$ and $\operatorname{rank}[A] = \operatorname{cr}[A]$. For any $x \in U$ we have

$$A(x) = A\Big(\sum x_i u_i\Big) = \sum x_i A(u_i)$$

so that the $A(u_i)$ span $\operatorname{Im} A$. But $[A(u_i)]$ is just the ith column of $[A]$, and hence the $[A(u_i)]$ also span the column space of $[A]$. Therefore the number of linearly independent columns of $[A]$ is the same as the number of linearly independent vectors in the image of A (see Exercise 4.3.1). This means that $\operatorname{rank} A = \operatorname{cr}[A] = \operatorname{rank}[A]$. ∎

Recall from Section 2.5 that the kernel of a matrix is the subspace $\ker(A) = \{X \in \mathbb{R}^n : AX = 0\}$. In other words, $\ker(A)$ is just the matrix version of the kernel of the linear transformation represented by A. Then Theorem 4.12 shows that the rank theorem (Theorem 4.6) can be directly stated in terms of matrices, and gives us another way to view Theorem 2.18.

Corollary. *Let $A \in M_{m\times n}(\mathcal{F})$. Then*

$$\operatorname{rank} A + \dim(\ker A) = n.$$

Suppose we have a system of n linear equations in n unknowns written in matrix form as $[A]X = Y$ where $[A]$ is the matrix representation of the corresponding linear transformation $A \in L(V)$, and $\dim V = n$. If we are to solve this for a unique X, then $[A]$ must be of rank n (Theorem 2.13). Hence $\operatorname{rank} A = n$ also so that $\operatorname{nul} A = \dim(\operatorname{Ker} A) = 0$ by Theorem 4.6. But this means that $\operatorname{Ker} A = \{0\}$ and thus A is nonsingular. Note also that Theorem 2.10 now says that the dimension of the solution space is zero (which it must be for the solution to be unique) which agrees with $\operatorname{Ker} A = \{0\}$.

All of this merely shows the various interrelationships between the matrix nomenclature and the concept of a linear transformation that should be expected in view of Theorem 4.11. Our discussion is summarized by the following useful characterization.

Theorem 4.13. *A linear transformation $A \in L(V)$ is nonsingular if and only if* $\det[A] \neq 0$.

Proof. Let $\dim V = n$. If A is nonsingular, then $\operatorname{Ker} A = \{0\}$ so that $\operatorname{nul} A = 0$ and hence $\operatorname{rank}[A] = \operatorname{rank} A = n$ (Theorem 4.6). Therefore $[A]^{-1}$ exists (Theorem 2.20) so that $\det[A] \neq 0$ (Theorem 3.6). The converse follows by an exact reversal of the argument. ∎

Exercises

1. Suppose $A \in L(U, V)$ and let $\{u_i\}$, $\{v_i\}$ be bases for U and V respectively. Show directly that $\{A(u_i)\}$ is linearly independent if and only if the columns of $[A]$ are also linearly independent.

2. Let V be the space of all real polynomials of degree less than or equal to 3. In other words, elements of V are of the form $f(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3$ where each $a_i \in \mathbb{R}$.

   (a) Show the derivative mapping $D = d/dx$ is an element of $L(V)$.
   (b) Find the matrix of D relative to the ordered basis $\{f_i\}$ for V defined by $f_i(x) = x^{i-1}$.

3. Let $T : \mathbb{R}^3 \to \mathbb{R}^2$ be defined by $T(x, y, z) = (x + y, 2z - x)$.

   (a) Find the matrix of T relative to the standard bases for $\mathbb{R}^3$ and $\mathbb{R}^2$.
   (b) Find the matrix of T relative to the basis $\{\alpha_i\}$ for $\mathbb{R}^3$ and $\{\beta_i\}$ for $\mathbb{R}^2$ where $\alpha_1 = (1, 0, -1)$, $\alpha_2 = (1, 1, 1)$, $\alpha_3 = (1, 0, 0)$, $\beta_1 = (0, 1)$ and $\beta_2 = (1, 0)$.

4. Relative to the standard basis, let $T \in L(\mathbb{R}^3)$ have the matrix representation
$$\begin{bmatrix} 1 & 2 & 1 \\ 0 & 1 & 1 \\ -1 & 3 & 4 \end{bmatrix}.$$
   Find a basis for $\operatorname{Im} T$ and $\operatorname{Ker} T$.

5. Let $T \in L(\mathbb{R}^3)$ be defined by $T(x, y, z) = (3x + z, -2x + y, -x + 2y + 4z)$.

   (a) Find the matrix of T relative to the standard basis for $\mathbb{R}^3$.
   (b) Find the matrix of T relative to the basis $\{\alpha_i\}$ given by $\alpha_1 = (1, 0, 1)$, $\alpha_2 = (-1, 2, 1)$ and $\alpha_3 = (2, 1, 1)$.
   (c) Show that T is invertible, and give a formula for T^{-1} similar to that given in part (a) for T.

6. Let $T : \mathcal{F}^n \to \mathcal{F}^m$ be the linear transformation defined by
$$T(x_1, \ldots, x_n) = \left(\sum_{i=1}^n a_{1i} x_i, \ldots, \sum_{i=1}^n a_{mi} x_i \right).$$

(a) Show that the matrix of T relative to the standard bases of $\mathcal{F}^n$ and $\mathcal{F}^m$ is given by

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}.$$

(b) Find the matrix representation of $T : \mathbb{R}^4 \to \mathbb{R}^2$ defined by

$$T(x, y, z, t) = (3x - 4y + 2z - 5t, 5x + 7y - z - 2t)$$

relative to the standard bases of $\mathbb{R}^n$.

7. Let $\dim U = m$, $\dim V = n$, and suppose that $T \in L(U, V)$ has rank r. Prove there exists a basis for U and a basis for V relative to which the matrix of T takes the form

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

[*Hint*: Show that $\operatorname{Ker} T$ has a basis $\{w_1, \ldots, w_{m-r}\}$, and then extend this to a basis $\{u_1, \ldots, u_r, w_1, \ldots, w_{m-r}\}$ for U. Define $v_i = T(u_i)$, and show this is a basis for $\operatorname{Im} T$. Now extend this to a basis for V.]

8. Let $\{e_i\}$ be the standard basis for $\mathbb{R}^3$, and let $\{f_i\}$ be the standard basis for $\mathbb{R}^2$.

(a) Define $T : \mathbb{R}^3 \to \mathbb{R}^2$ by $T(e_1) = f_2$, $T(e_2) = f_1$ and $T(e_3) = f_1 + f_2$. Write down the matrix $[T]_e^f$.
(b) Define $S : \mathbb{R}^2 \to \mathbb{R}^3$ by $S(f_1) = (1, 2, 3)$ and $S(f_2) = (2, -1, 4)$. Write down $[S]_f^e$.
(c) Find $ST(e_i)$ for each $i = 1, 2, 3$ and write down the matrix $[ST]_e$ of the linear operator $ST : \mathbb{R}^3 \to \mathbb{R}^3$. Verify that $[ST] = [S][T]$.

9. Suppose $T \in L(V)$ and let W be a subspace of V. We say that W is **invariant under** T (or T**-invariant**) if $T(W) \subset W$. If $\dim W = m$, show that T has a "block matrix" representation of the form

$$\begin{bmatrix} A & B \\ 0 & C \end{bmatrix}$$

where A is an $m \times m$ matrix.

10. Let $T \in L(V)$, and suppose that $V = U \oplus W$ where both U and W are T-invariant (see problem 9). If $\dim U = m$ and $\dim W = n$, show that T has a matrix representation of the form

$$\begin{bmatrix} A & 0 \\ 0 & C \end{bmatrix}$$

where A is an $m \times m$ matrix and C is an $n \times n$ matrix.

11. Show that $A \in L(V)$ is nonsingular implies $[A^{-1}] = [A]^{-1}$.

4.4 Change of Basis

Suppose we have a linear operator $A \in L(V)$. Then, given a basis for V, we can write down the corresponding matrix $[A]$. If we change to a new basis for V, then we will have a new representation for A. We now investigate the relationship between the matrix representations of A in each of these bases.

Given a vector space V, let us consider two arbitrary bases $\{e_1, \dots, e_n\}$ and $\{\bar{e}_1, \dots, \bar{e}_n\}$ for V. Then any vector $x \in V$ may be written as either $x = \sum x_i e_i$ or as $x = \sum \bar{x}_i \bar{e}_i$. (It is important to realize that vectors and linear transformations exist independently of the coordinate system used to describe them, and their components may vary from one coordinate system to another.) Since each $\bar{e}_i$ is a vector in V, we may write its components in terms of the basis $\{e_i\}$. In other words, we define the **transition matrix** $[P] = (p_{ij}) \in M_n(\mathcal{F})$ by

$$\bar{e}_i = \sum_{j=1}^n e_j p_{ji}$$

for each $i = 1, \dots, n$. There is nothing mysterious here; p_{ji} is merely the jth component of the vector $\bar{e}_i$ with respect to the basis $\{e_1, \dots, e_n\}$. The matrix $[P]$ must be unique for the given bases according to Corollary 2 of Theorem 1.4.

Note that $[P]$ defines a linear transformation $P \in L(V)$ by $P(e_i) = \bar{e}_i$. Since $\{P(e_i)\} = \{\bar{e}_i\}$ spans $\operatorname{Im} P$ and the $\bar{e}_i$ are linearly independent, it follows that $\operatorname{rank}(P) = n$ so that P is nonsingular and hence P^{-1} exists. By Theorem 4.11, we have $I = [I] = [PP^{-1}] = [P][P^{-1}]$ and hence we conclude that $[P^{-1}] = [P]^{-1}$. (However, it is also quite simple to show directly that if a linear operator A is nonsingular, then $[A^{-1}] = [A]^{-1}$. See Exercise 4.3.11).

Let us emphasize an earlier remark. From Theorem 4.9 we know that $[\bar{e}_i] = [P(e_i)]$ is just the ith column vector of $[P]$. Since relative to the basis $\{e_i\}$ we have $e_1 = (1, 0, \dots, 0)$, $e_2 = (0, 1, \dots, 0)$ and so on, it follows that the ith column of $[P]$ represents the components of $\bar{e}_i$ relative to the basis $\{e_i\}$. In other words, the matrix entry p_{ji} is the jth component of the ith basis vector $\bar{e}_i$ relative to the basis $\{e_i\}$. Written out in full, this statement is

$$\begin{aligned}
\bar{e}_i = \sum_{j=1}^n e_j p_{ji} &= e_1 p_{1i} + e_2 p_{2i} + \cdots + e_n p_{ni} \\
&= \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} p_{1i} + \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix} p_{2i} + \cdots + \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix} p_{ni} = \begin{bmatrix} p_{1i} \\ p_{2i} \\ \vdots \\ p_{ni} \end{bmatrix}
\end{aligned}$$

which is just the ith column of $[P]$.

The transition matrix enables us to easily relate the components of any $x \in V$ between the two coordinate systems. To see this, we observe that

$$x = \sum_i x_i e_i = \sum_j \bar{x}_j \bar{e}_j = \sum_{i,j} \bar{x}_j e_i p_{ij} = \sum_{i,j} p_{ij} \bar{x}_j e_i$$

and hence the uniqueness of the expansion implies $x_i = \sum_j p_{ij}\bar{x}_j$ so that

$$\bar{x}_j = \sum_i p^{-1}{}_{ji} x_i .$$

This discussion proves the following theorem.

Theorem 4.14. *Let $[P]$ be the transition matrix from a basis $\{e_i\}$ to a basis $\{\bar{e}_i\}$ for a space V. Then for any $x \in V$ we have*

$$[x]_{\bar{e}} = [P]^{-1}[x]_e$$

which we sometimes write simply as $\overline{X} = P^{-1}X$.

From now on we will omit the brackets on matrix representations unless they are needed for clarity. Thus we will usually write both a linear transformation $A \in L(U, V)$ and its representation $[A] \in M_{m\times n}(\mathcal{F})$ as simply A. Furthermore, to avoid possible ambiguity, we will sometimes denote a linear transformation by T, and its corresponding matrix representation by $A = (a_{ij})$.

Using the above results, it is now an easy matter for us to relate the representation of a linear operator $A \in L(V)$ in one basis to its representation in another basis. If $A(e_i) = \sum_j e_j a_{ji}$ and $A(\bar{e}_i) = \sum_j \bar{e}_j \bar{a}_{ji}$, then on the one hand we have

$$A(\bar{e}_i) = \sum_j \bar{e}_j \bar{a}_{ji} = \sum_{j,k} e_k p_{kj} \bar{a}_{ji}$$

while on the other hand,

$$A(\bar{e}_i) = A\biggl(\sum_j e_j p_{ji}\biggr) = \sum_j A(e_j) p_{ji} = \sum_{j,k} e_k a_{kj} p_{ji} .$$

Therefore, since $\{e_k\}$ is a basis for V, we may equate each component in these two equations to obtain $\sum_j p_{kj}\bar{a}_{ji} = \sum_j a_{kj} p_{ji}$ or

$$\bar{a}_{ri} = \sum_{j,k} p^{-1}{}_{rk} a_{kj} p_{ji} .$$

In matrix notation this is just (omitting the brackets on P)

$$[A]_{\bar{e}} = P^{-1}[A]_e P$$

which we will usually write in the form $\bar{A} = P^{-1}AP$ for simplicity.

If $A, B \in M_n(\mathcal{F})$, then B is said to be **similar** to A if there exists a nonsingular matrix S such that $B = S^{-1}AS$, in which case A and B are said to be related by a **similarity transformation**. We leave it to the reader to show that this defines an equivalence relation on $M_n(\mathcal{F})$ (see Exercise 4.4.1).

We have shown that the matrix representations of an operator with respect to two different bases are related by a similarity transformation. So a reasonable question is if we have a representation in one basis, does every other matrix similar to the first represent the operator in some other basis? Here is the answer.

Theorem 4.15. *If $T \in L(V)$ is represented by $A \in M_n(\mathcal{F})$ relative to the basis $\{e_i\}$, then a matrix $\bar{A} \in M_n(\mathcal{F})$ represents T relative to some basis $\{\bar{e}_i\}$ if and only if $\bar{A}$ is similar to A. If this is the case, then*

$$\bar{A} = P^{-1}AP$$

where P is the transition matrix from the basis $\{e_i\}$ to the basis $\{\bar{e}_i\}$.

Proof. The discussion above showed that if A and $\bar{A}$ represent T in two different bases, then $\bar{A} = P^{-1}AP$ where P is the transition matrix from $\{e_i\}$ to $\{\bar{e}_i\}$.

On the other hand, suppose that T is represented by $A = (a_{ij})$ in the basis $\{e_i\}$, and assume that $\bar{A}$ is similar to A. Then $\bar{A} = P^{-1}AP$ for some nonsingular matrix $P = (p_{ij})$. We define a new basis $\{\bar{e}_i\}$ for V by

$$\bar{e}_i = P(e_i) = \sum_j e_j p_{ji}$$

(where we use the same symbol for both the operator P and its matrix representation). Then

$$T(\bar{e}_i) = T\left(\sum_j e_j p_{ji}\right) = \sum_j T(e_j)p_{ji} = \sum_{j,k} e_k a_{kj} p_{ji}$$

while on the other hand, if T is represented by some matrix $C = (c_{ij})$ in the basis $\{\bar{e}_i\}$, then

$$T(\bar{e}_i) = \sum_j \bar{e}_j c_{ji} = \sum_{j,k} e_k p_{kj} c_{ji}.$$

Equating the coefficients of e_k in both of these expressions yields

$$\sum_j a_{kj} p_{ji} = \sum_j p_{kj} c_{ji}$$

so that

$$c_{ri} = \sum_{j,k} p^{-1}{}_{rk} a_{kj} p_{ji}$$

and hence

$$C = P^{-1}AP = \bar{A}.$$

Therefore $\bar{A}$ represents T in the basis $\{\bar{e}_i\}$. ∎

Example 4.16. Consider the linear transformation $T \in L(\mathbb{R}^3)$ defined by

$$T(x,y,z) = \begin{bmatrix} 9x+y \\ 9y \\ 7z \end{bmatrix}.$$

Let $\{e_i\}$ be the standard basis for $\mathbb{R}^3$, and let $\{\bar{e}_i\}$ be the basis defined by

$$\bar{e}_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \qquad \bar{e}_2 = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} \qquad \bar{e}_3 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}.$$

Let us first find the representation $\bar{A} = [T]_{\bar{e}}$ directly from the definition $T(\bar{e}_i) = \sum_{j=1}^3 \bar{e}_j \bar{a}_{ji}$. We will go through two ways of doing this to help clarify the various concepts involved.

We have $T(\bar{e}_1) = T(1,0,1) = (9,0,7)$. Then we write $(9,0,7) = a(1,0,1) + b(1,0,-1) + c(0,1,1)$ and solve for a, b, c to obtain $T(\bar{e}_1) = 8\bar{e}_1 + \bar{e}_2$. Similarly, we find $T(\bar{e}_2) = T(1,0,-1) = (9,0,-7) = \bar{e}_1 + 8\bar{e}_2$ and $T(\bar{e}_3) = T(0,1,1) = (1,9,7) = (-1/2)\bar{e}_1 + (3/2)\bar{e}_2 + 9\bar{e}_3$. This shows that the representation $[T]_{\bar{e}}$ is given by

$$\bar{A} = [T]_{\bar{e}} = \begin{bmatrix} 8 & 1 & -1/2 \\ 1 & 8 & 3/2 \\ 0 & 0 & 9 \end{bmatrix}.$$

Another way is to use the fact that everything is simple with respect to the standard basis for $\mathbb{R}^3$. We see that $T(e_1) = T(1,0,0) = (9,0,0) = 9e_1, T(e_2) = T(0,1,0) = (1,9,0) = e_1 + 9e_2$ and $T(e_3) = T(0,0,1) = (0,0,7) = 7e_3$. Note that this shows

$$A = [T]_e = \begin{bmatrix} 9 & 1 & 0 \\ 0 & 9 & 0 \\ 0 & 0 & 7 \end{bmatrix}$$

which we will need below when we use the transition matrix to find $\bar{A}$.

It is easy to see that $\bar{e}_1 = e_1 + e_3, \bar{e}_2 = e_1 - e_3$ and $\bar{e}_3 = e_2 + e_3$, so inverting these equations we have $e_1 = (1/2)(\bar{e}_1 + \bar{e}_2), e_3 = (1/2)(\bar{e}_1 - \bar{e}_2)$ and $e_2 = \bar{e}_3 - e_3 = -(1/2)(\bar{e}_1 - \bar{e}_2) + \bar{e}_3$. Then using the linearity of T we have

$$\begin{aligned} T(\bar{e}_1) &= T(e_1 + e_3) = T(e_1) + T(e_3) = 9e_1 + 7e_3 \\ &= (9/2)(\bar{e}_1 + \bar{e}_2) + (7/2)(\bar{e}_1 - \bar{e}_2) \\ &= 8\bar{e}_1 + \bar{e}_2 \end{aligned}$$

$$\begin{aligned} T(\bar{e}_2) &= T(e_1 - e_3) = T(e_1) - T(e_3) = 9e_1 - 7e_3 \\ &= (9/2)(\bar{e}_1 + \bar{e}_2) - (7/2)(\bar{e}_1 - \bar{e}_2) \\ &= \bar{e}_1 + 8\bar{e}_2 \end{aligned}$$

$$\begin{aligned} T(\bar{e}_3) &= T(e_2 + e_3) = T(e_2) + T(e_3) = e_1 + 9e_2 + 7e_3 \\ &= (1/2)(\bar{e}_1 + \bar{e}_2) - (9/2)(\bar{e}_1 - \bar{e}_2) + 9\bar{e}_3 + (7/2)(\bar{e}_1 - \bar{e}_2) \\ &= -(1/2)\bar{e}_1 + (3/2)\bar{e}_2 + 9\bar{e}_3 \end{aligned}$$

and, as expected, this gives the same result as we had above for $[T]_{\bar{e}}$.

Now we will use the transition matrix P to find $\bar{A} = [T]_{\bar{e}}$. The matrix P is defined by $\bar{e}_i = Pe_i = \sum_{j=1}^{3} e_j p_{ji}$ is just the ith column of P, so we immediately have

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

There are a number of ways to find P^{-1}. We could use the row reduction method described in Section 2.6, we could use Theorem 3.9, or we could use the fact that the inverse matrix is defined by $e_i = P^{-1}\bar{e}_i$ and use the expressions we found above for each e_i in terms of the $\bar{e}_i$'s. This last approach is the easiest for us and we can just write down the result

$$P^{-1} = \frac{1}{2}\begin{bmatrix} 1 & -1 & 1 \\ 1 & 1 & -1 \\ 0 & 2 & 0 \end{bmatrix}.$$

We now see that

$$\begin{aligned} [T]_{\bar{e}} = P^{-1}[T]_e P &= \frac{1}{2}\begin{bmatrix} 1 & -1 & 1 \\ 1 & 1 & -1 \\ 0 & 2 & 0 \end{bmatrix}\begin{bmatrix} 9 & 1 & 0 \\ 0 & 9 & 0 \\ 0 & 0 & 7 \end{bmatrix}\begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & -1 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 8 & 1 & -1/2 \\ 1 & 8 & 3/2 \\ 0 & 0 & 9 \end{bmatrix} \end{aligned}$$

which agrees with our previous approaches.

Also realize that a vector $X = (x, y, z) \in \mathbb{R}^3$ has components x, y, z only with respect to the standard basis $\{e_i\}$ for $\mathbb{R}^3$. In other words

$$X = \begin{bmatrix} x \\ y \\ z \end{bmatrix} = x\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + y\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + z\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = xe_1 + ye_2 + ze_3.$$

But with respect to the basis $\{\bar{e}_i\}$ we have

$$\begin{aligned} \overline{X} = P^{-1}X &= \frac{1}{2}\begin{bmatrix} 1 & -1 & 1 \\ 1 & 1 & -1 \\ 0 & 2 & 0 \end{bmatrix}\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \frac{1}{2}\begin{bmatrix} x - y + z \\ x + y - z \\ 2y \end{bmatrix} \\ &= \frac{1}{2}(x - y + z)\bar{e}_1 + \frac{1}{2}(x + y - z)\bar{e}_2 + y\bar{e}_3 \\ &= \bar{x}\bar{e}_1 + \bar{y}\bar{e}_2 + \bar{z}\bar{e}_3. \end{aligned}$$

Note that by Theorem 3.7 and its corollary we have

$$\det \bar{A} = \det(P^{-1}AP) = (\det P^{-1})(\det A)(\det P) = \det A$$

and hence all matrices which represent a linear operator T have the same determinant. Another way of stating this is to say that the determinant is **invariant** under a similarity transformation. We thus define the **determinant of a linear operator** $T \in L(V)$ as $\det A$, where A is any matrix representing T.

Another important quantity associated with a matrix $A \in M_n(\mathcal{F})$ is the sum $\sum_{i=1}^n a_{ii}$ of its diagonal elements. This sum is called the **trace**, and is denoted by $\operatorname{tr} A$ (see Exercise 2.5.7). A simple but useful result is the following.

Theorem 4.16. *If $A, B \in M_n(\mathcal{F})$, then $\operatorname{tr}(AB) = \operatorname{tr}(BA)$.*

Proof. We simply compute

$$\begin{aligned} \operatorname{tr}(AB) &= \sum_i (AB)_{ii} = \sum_{i,j} a_{ij}b_{ji} = \sum_j \sum_i b_{ji}a_{ij} = \sum_j (BA)_{jj} \\ &= \operatorname{tr}(BA). \end{aligned}$$ ∎

From this theorem it is easy to show that the trace is also invariant under a similarity transformation (see Exercise 3.3.13). Because of this, it also makes sense to speak of the **trace of a linear operator**.

Example 4.17. Consider the space $V = \mathbb{R}^2$ with its standard basis $e_1 = (1, 0)$ and $e_2 = (0, 1)$, and let $\bar{e}_1 = (1, 2)$, $\bar{e}_2 = (3, -1)$ be another basis. We then see that

$$\begin{aligned} \bar{e}_1 &= \ \ e_1 + 2e_2 \\ \bar{e}_2 &= 3e_1 - \ \ e_2 \end{aligned}$$

and consequently the transition matrix P from $\{e_i\}$ to $\{\bar{e}_i\}$ and its inverse P^{-1} are given by

$$P = \begin{bmatrix} 1 & 3 \\ 2 & -1 \end{bmatrix} \qquad \text{and} \qquad P^{-1} = \begin{bmatrix} 1/7 & 3/7 \\ 2/7 & -1/7 \end{bmatrix}.$$

Note that P^{-1} may be found either using Theorem 3.9, or by solving for $\{e_i\}$ in terms of $\{\bar{e}_i\}$ to obtain

$$\begin{aligned} e_1 &= (1/7)\bar{e}_1 + (2/7)\bar{e}_2 \\ e_2 &= (3/7)\bar{e}_1 - (1/7)\bar{e}_2 \end{aligned}$$

Now let T be the operator defined by

$$\begin{aligned} T(e_1) &= \ \ (20/7)e_1 - \ \ (2/7)e_2 \\ T(e_2) &= (-3/7)e_1 + (15/7)e_2 \end{aligned}$$

so that relative to the basis $\{e_i\}$ we have

$$A = \begin{bmatrix} 20/7 & -3/7 \\ -2/7 & 15/7 \end{bmatrix}.$$

We thus find that

$$\bar{A} = P^{-1}AP = \begin{bmatrix} 1/7 & 3/7 \\ 2/7 & -1/7 \end{bmatrix} \begin{bmatrix} 20/7 & -3/7 \\ -2/7 & 15/7 \end{bmatrix} \begin{bmatrix} 1 & 3 \\ 2 & -1 \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}.$$

Alternatively, we have

$$\begin{aligned} T(\bar{e}_1) &= T(e_1 + 2e_2) = T(e_1) + 2T(e_2) = 2e_1 + 4e_2 = 2\bar{e}_1 \\ T(\bar{e}_2) &= T(3e_1 - e_2) = 3T(e_1) - T(e_2) = 9e_1 - 3e_2 = 3\bar{e}_2 \end{aligned}$$

so that again we find

$$\bar{A} = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}$$

We now see that

$$\operatorname{tr} A = 20/7 + 15/7 = 5 = \operatorname{tr} \bar{A}$$

and also

$$\det A = 6 = \det \bar{A}$$

as they should.

You should also verify that the determinant and trace are invariant in Example 4.16.

We point out that in this example, $\bar{A}$ turns out to be a diagonal matrix. In this case the basis $\{\bar{e}_i\}$ is said to **diagonalize** the operator T. While it is certainly *not* true that there always exists a basis in which an operator takes on a diagonal representation, in the next chapter we will study the conditions under which we can in fact find such a basis of eigenvectors.

Let us make one related additional comment about our last example. While it is true that (algebraically speaking) a linear operator is completely determined once its effect on a basis is known, there is no real geometric interpretation of this when the matrix representation of an operator is of the same form as A in Example 4.17. However, if the representation is diagonal as it is with $\bar{A}$, then in this basis the operator represents a magnification factor in each direction. In other words, we see that $\bar{A}$ represents a multiplication of any vector in the $\bar{e}_1$ direction by 2, and a multiplication of any vector in the $\bar{e}_2$ direction by 3. This is the physical interpretation that we will attach to eigenvalues as studied in the next chapter.

Exercises

1. Show that the set of similar matrices defines an equivalence relation on $M_n(\mathcal{F})$.

2. Let $\{e_i\}$ be the standard basis for $\mathbb{R}^3$, and consider the basis $f_1 = (1,1,1)$, $f_2 = (1,1,0)$ and $f_3 = (1,0,0)$.
   (a) Find the transition matrix P from $\{e_i\}$ to $\{f_i\}$.
   (b) Find the transition matrix Q from $\{f_i\}$ to $\{e_i\}$.
   (c) Verify that $Q = P^{-1}$.
   (d) Show that $[v]_f = P^{-1}[v]_e$ for any $v \in \mathbb{R}^3$.
   (e) Define $T \in L(\mathbb{R}^3)$ by $T(x,y,z) = (2y+z, x-4y, 3x)$. Show that $[T]_f = P^{-1}[T]_e P$.

3. Let $\{e_1, e_2\}$ be a basis for V, and define $T \in L(V)$ by $T(e_1) = 3e_1 - 2e_2$ and $T(e_2) = e_1 + 4e_2$. Define the basis $\{f_i\}$ for V by $f_1 = e_1 + e_2$ and $f_2 = 2e_1 + 3e_2$. Find $[T]_f$.

4. Consider the field $\mathbb{C}$ as a vector space over $\mathbb{R}$, and define the linear "conjugation operator" $T \in L(\mathbb{C})$ by $T(z) = z^*$ for each $z \in \mathbb{C}$.
   (a) Find the matrix of T relative to the basis $\{e_j\} = \{1, i\}$.
   (b) Find the matrix of T relative to the basis $\{f_j\} = \{1+i, 1+2i\}$.
   (c) Find the transition matrices P and Q that go from $\{e_j\}$ to $\{f_j\}$ and from $\{f_j\}$ to $\{e_j\}$ respectively.
   (d) Verify that $Q = P^{-1}$.
   (e) Show that $[T]_f = P^{-1}[T]_e P$.
   (f) Verify that $\operatorname{tr}[T]_f = \operatorname{tr}[T]_e$ and $\det[T]_f = \det[T]_e$.

5. Let $\{e_i\}$, $\{f_i\}$ and $\{g_i\}$ be bases for V, and let P and Q be the transition matrices from $\{e_i\}$ to $\{f_i\}$ and from $\{f_i\}$ to $\{g_i\}$ respectively. Show that QP is the transition matrix from $\{e_i\}$ to $\{g_i\}$.

6. Let A be a 2×2 matrix such that only A is similar to itself. Show that A has the form
$$\begin{bmatrix} a & 0 \\ 0 & a \end{bmatrix}.$$

7. Show that similar matrices have the same rank.

8. (a) Let $\{e_i\}$ be the standard basis for $\mathbb{R}^n$, and let $\{f_i\}$ be any other orthonormal basis (relative to the standard inner product). Show that the transition matrix P from $\{e_i\}$ to $\{f_i\}$ is **orthogonal**, i.e., $P^T = P^{-1}$.
   (b) Let $T \in L(\mathbb{R}^3)$ have the following matrix relative to the standard basis:
$$\begin{bmatrix} 2 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 3 \end{bmatrix}.$$

Find the matrix of T relative to the basis $f_1 = (2/3, 2/3, -1/3)$, $f_2 = (1/3, 2/3, -2/3)$ and $f_3 = (2/3, -1/3, 2/3)$.

9. Let $T \in L(\mathbb{R}^2)$ have the following matrix relative to the standard basis $\{e_i\}$ for $\mathbb{R}^2$:
$$[T]_e = \begin{bmatrix} a & b \\ c & d \end{bmatrix}.$$
   (a) Suppose there exist two linearly independent vectors f_1 and f_2 in $\mathbb{R}^2$ with the property that $T(f_1) = \lambda_1 f_1$ and $T(f_2) = \lambda_2 f_2$ (where $\lambda_i \in \mathbb{R}$). If P is the transition matrix from the basis $\{e_i\}$ to the basis $\{f_i\}$, show
$$[T]_f = P^{-1}[T]_e P = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}.$$
   (b) Prove there exists a nonzero vector $x \in \mathbb{R}^2$ with the property that $T(x) = x$ if and only if
$$\begin{vmatrix} a-1 & b \\ c & d-1 \end{vmatrix} = 0.$$
   (c) Prove there exists a one-dimensional T-invariant subspace of $\mathbb{R}^2$ if and only if
$$\begin{vmatrix} a-\lambda & b \\ c & d-\lambda \end{vmatrix} = 0$$
   for some scalar λ. (Recall that a subspace W is T-invariant if $T(W) \subset W$.)
   (d) Let $T \in L(\mathbb{C}^2)$ be represented by the matrix
$$\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$
   where $\theta \in \mathbb{R}$. Show that there exist two one-dimensional (complex) T-invariant subspaces, and hence show that T and the matrix
$$\begin{bmatrix} e^{i\theta} & 0 \\ 0 & e^{-i\theta} \end{bmatrix}$$
   are similar over the complex field.

10. Let $V = \mathbb{R}^2$ have basis vectors $e_1 = (1, 1)$ and $e_2 = (1, -1)$. Suppose we define another basis for V by $\bar{e}_1 = (2, 4)$ and $\bar{e}_2 = (3, 1)$. Define the transition operator $P \in L(V)$ as usual by $\bar{e}_i = Pe_i$. Write down the matrix $[P]_e^{\bar{e}}$.

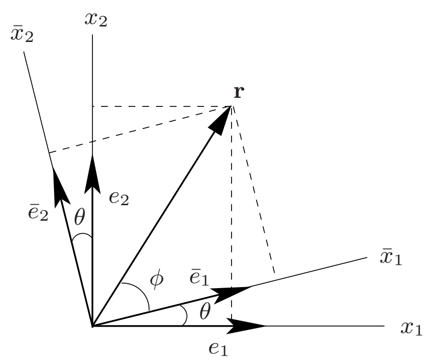
11. Let U have bases $\{u_i\}$ and $\{\bar{u}_i\}$ and let V have bases $\{v_i\}$ and $\{\bar{v}_i\}$. Define the transition operators $P \in L(U)$ and $Q \in L(V)$ by $\bar{u}_i = Pu_i$ and $\bar{v}_i = Qv_i$. If $T \in L(U, V)$, express $[T]_u^v$ in terms of $[T]_{\bar{u}}^{\bar{v}}$.

12. Show that the transition matrix defined by the Gram-Schmidt process is upper-triangular with strictly positive determinant.

4.5 Orthogonal Transformations

In this last section of the chapter we will take a brief look at one particular kind of linear transformation that is of great importance in physics and engineering. For simplicity of notation, we will let $x = x_1$ and $y = x_2$, and follow the summation convention used in Chapter 3.

To begin with, let $\mathbf{r}$ be a vector in $\mathbb{R}^2$ and consider a counterclockwise rotation of the x_1x_2-plane about the x_3-axis as shown below.



The vectors e_i and $\bar{e}_i$ are the usual orthonormal basis vectors with $\|e_i\| = \|\bar{e}_i\| = 1$. From the geometry of the diagram we see that

$$\begin{aligned}\bar{e}_1 &= (\cos\theta)e_1 + (\sin\theta)e_2 \\ \bar{e}_2 &= -(\sin\theta)e_1 + (\cos\theta)e_2\end{aligned}$$

so that $\bar{e}_i = P(e_i) = e_j p^j{}_i$ and the transition matrix $(p^j{}_i)$ is given by

$$(p^j{}_i) = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}. \tag{4.4}$$

Since $\mathbf{r} = x^j e_j = \bar{x}^i \bar{e}_i = \bar{x}^i e_j p^j{}_i = (p^j{}_i \bar{x}^i)e_j$ we see that $x^j = p^j{}_i \bar{x}^i$ or $\bar{x}^i = (p^{-1})^i{}_j x^j$ as shown in the discussion prior to Theorem 4.14.

You can easily compute the matrix P^{-1}, but it is better to make the general observation that rotating the coordinate system doesn't change the length of $\mathbf{r}$. So using $\|\mathbf{r}\|^2 = x^i x_i = \bar{x}^j \bar{x}_j$ together with $x^i = p^i{}_j \bar{x}^j$ this becomes

$$x^i x_i = p^i{}_j \bar{x}^j p_i{}^k \bar{x}_k = (p^T)^k{}_i p^i{}_j \bar{x}^j \bar{x}_k := \bar{x}^j \bar{x}_j$$

so that we must have

$$(p^T)^k{}_i p^i{}_j = \delta^k_j .$$

In matrix notation this is just $P^T P = I$ which implies that $P^T = P^{-1}$. This is the definition of an **orthogonal transformation** (or **orthogonal matrix**). In other words, a matrix $A \in M_n(\mathcal{F})$ is said to be **orthogonal** if and only if $A^T = A^{-1}$.

As an important consequence of this definition, note that if A is orthogonal, then

$$1 = \det I = \det(AA^{-1}) = \det(AA^T) = (\det A)(\det A^T) = (\det A)^2$$

and hence

$$\det A = \pm 1. \tag{4.5}$$

Going back to our example rotation, we therefore have

$$P^{-1} = P^T = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}$$

so that $\bar{x}^i = (p^{-1})^i{}_j x^j = (p^T)^i{}_j x^j$ or

$$\begin{aligned} \bar{x}_1 &= (\cos\theta)x_1 + (\sin\theta)x_2 \\ \bar{x}_2 &= -(\sin\theta)x_1 + (\cos\theta)x_2 \end{aligned}$$

To check these results, we first verify that $P^{-1} = P^T$:

$$P^T P = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I.$$

Next, from the diagram we see that

$$\begin{aligned} x_1 = r\cos(\theta+\phi) &= r\cos\theta\cos\phi - r\sin\theta\sin\phi \\ &= (\cos\theta)\bar{x}_1 - (\sin\theta)\bar{x}_2 \end{aligned}$$

$$\begin{aligned} x_2 = r\sin(\theta+\phi) &= r\sin\theta\cos\phi + r\cos\theta\sin\phi \\ &= (\sin\theta)\bar{x}_1 + (\cos\theta)\bar{x}_2 \end{aligned}$$

In matrix form this is

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \end{bmatrix} \tag{4.6}$$

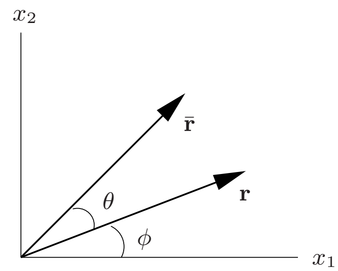
or, alternatively,

$$\begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \tag{4.7}$$

which is the same as we saw above using $(p^T)^i{}_j$.

To be completely precise, the rotation that we have just described is properly called a **passive transformation** because it left the vector alone and rotated the coordinate system. An alternative approach is to leave the coordinate system alone and rotate the vector itself. This is called an **active transformation**. One must be very careful when reading the literature to be aware of just which type of rotation is under consideration. Let's compare the two types of rotation.

With an active transformation we have the following situation:



Here the vector $\mathbf{r}$ is rotated by θ to give the vector $\bar{\mathbf{r}}$ where, of course, $\|\mathbf{r}\| = \|\bar{\mathbf{r}}\|$. In the *passive* case we defined the transition matrix P by $\bar{e}_i = P(e_i)$. Now, in the *active* case we define a linear transformation T by $\bar{\mathbf{r}} = T(\mathbf{r})$. From the diagram, the components of $\bar{\mathbf{r}}$ are given by

$$\begin{aligned}\bar{x}_1 = r\cos(\theta+\phi) &= r\cos\theta\cos\phi - r\sin\theta\sin\phi \\ &= (\cos\theta)x_1 - (\sin\theta)x_2\end{aligned}$$

$$\begin{aligned}\bar{x}_2 = r\sin(\theta+\phi) &= r\sin\theta\cos\phi + r\cos\theta\sin\phi \\ &= (\sin\theta)x_1 + (\cos\theta)x_2\end{aligned}$$

or

$$\begin{bmatrix}\bar{x}_1\\ \bar{x}_2\end{bmatrix} = \begin{bmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{bmatrix}\begin{bmatrix}x_1\\ x_2\end{bmatrix}. \tag{4.8}$$

From equation (4.3) we see that this matrix is just the matrix representation of T.

Another way to write this is

$$(\bar{x}_1, \bar{x}_2) = T(x_1, x_2) = ((\cos\theta)x_1 - (\sin\theta)x_2, (\sin\theta)x_1 + (\cos\theta)x_2).$$

Then the first column of $[T]$ is

$$T(e_1) = T(1,0) = (\cos\theta, \sin\theta)$$

and the second column is

$$T(e_2) = T(0,1) = (-\sin\theta, \cos\theta)$$

so that

$$[T] = \begin{bmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{bmatrix}.$$

as in equation (4.8).

Carefully compare the matrix in equation (4.8) with that in equation (4.7). The matrix in equation (4.8) is obtained from the matrix in equation (4.7) by letting $\theta \to -\theta$. This is the effective difference between active and passive rotations. If a passive transformation rotates the coordinate system *counter-clockwise* by an angle θ, then the corresponding active transformation rotates the vector by the same angle but in the *clockwise* direction.

An interesting application of orthogonal transformations is the following. It seems intuitively obvious that the most general displacement of a rigid body (for instance, a rock) is a combination of a translation plus a rotation. While it is possible to prove this by looking at the eigenvalues of an orthogonal transformation (see, e.g., Goldstein [20]), we will approach the problem from a different point of view.

In order to give a mathematical description to the physical displacement of a rigid body, we want to look at all mathematical transformations that leave the distance between two points unchanged. Such a transformation is called an **isometry** (or **rigid motion**). We begin with a careful definition of isometry, reviewing some of what was discussed in Section 1.5.

Let V be a (finite-dimensional) real vector space with a positive definite inner product. This means that $\langle x, x\rangle = 0$ if and only if $x = 0$. Given the associated norm on V defined by $\|x\| = \langle x, x\rangle^{1/2}$, we may define a distance function, also called a **metric** on V, by

$$d(x, y) := \|x - y\| .$$

Note that a general property of the norm is

$$\begin{aligned} \|x - y\|^2 &= \langle x - y, x - y\rangle^2 \\ &= \langle x, x\rangle - \langle x, y\rangle - \langle y, x\rangle + \langle y, y\rangle \\ &= \|x\|^2 - 2\langle x, y\rangle + \|y\|^2 \end{aligned} \tag{4.9}$$

A function $f : V \to V$ is called an **isometry** (or **rigid motion**) if

$$\|f(x) - f(y)\| = \|x - y\| \tag{4.10}$$

or, equivalently,

$$d(f(x), f(y)) = d(x, y). \tag{4.11}$$

Let us consider what happens in the particular case that $f(0) = 0$. In this case we have first of all

$$\|f(x)\|^2 = \|f(x) - 0\|^2 = \|f(x) - f(0)\|^2 = \|x - 0\|^2 = \|x\|^2 \tag{4.12}$$

and thus isometries that leave the origin fixed also preserve the norm (or length) of vectors in V. It is also true that such isometries preserve inner products. To see this, first note that on the one hand, from equation (4.9) we have

$$\|f(x) - f(y)\|^2 = \|f(x)\|^2 - 2\langle f(x), f(y)\rangle + \|f(y)\|^2$$

while on the other hand, by definition of isometry, we also have $\|f(x) - f(y)\|^2 = \|x - y\|^2$. Therefore, equating the above equation with equation (4.9) and using equation (4.12) shows that

$$\langle f(x), f(y)\rangle = \langle x, y\rangle. \tag{4.13}$$

Again, this only applies to the special case where $f(0) = 0$, i.e., to those situations where the isometry leaves the origin fixed.

For example, it should be clear that the motion of a solid physical object through space is an isometry, because by definition of solid it follows that the distance between any two points in the object doesn't change. It is also obvious on physical grounds that the composition $F \circ G$ of two isometries F and G is another isometry. Mathematically this follows by simply noting that applying equation (4.11) twice we have

$$d(F(G(x)), F(G(y))) = d(G(x), G(y)) = d(x, y).$$

As a specific example, consider translations. A function $g : V \to V$ is called a **translation** if there exists a vector $v_0 \in V$ such that $g(x) = x + v_0$. We will sometimes denote the translation by g_{v_0} if we wish to emphasize the displacement vector v_0. It is easy to see that $g(x)$ is an isometry because

$$\|g(x) - g(y)\| = \|(x + v_0) - (y + v_0)\| = \|x - y\|.$$

Recall that a mapping $f : V \to V$ is said to be **linear** if $f(x + ay) = f(x) + af(y)$ for all $x, y \in V$ and $a \in \mathbb{R}$. If T is a linear transformation on V that preserves inner products, i.e., with the property that

$$\langle T(x), T(y)\rangle = \langle x, y\rangle$$

then we also say that T is an **orthogonal transformation**. (We will see below that this agrees with our previous definition.) It is easy to show that an orthogonal transformation is an isometry. Indeed, we first note that

$$\|T(x)\|^2 = \langle T(x), T(x)\rangle = \langle x, x\rangle = \|x\|^2$$

so that $\|T(x)\| = \|x\|$ and T preserves norms. But then the fact that T is linear shows that

$$\|T(x) - T(y)\| = \|T(x - y)\| = \|x - y\|$$

and thus T is an isometry.

Now let F be any isometry on V, and define the function $T : V \to V$ by

$$T(x) = F(x) - F(0).$$

Since T is just the composition of F with translation by $-F(0)$ (i.e., $T = g_{-F(0)} \circ F$) it is also an isometry. In addition, it has the property that $T(0) = 0$.

We first claim that T is in fact a linear transformation on V. This means we need to show that $T(x + ay) = T(x) + aT(y)$ where $a \in \mathbb{R}$. The proof is straightforward. Using our previous results we have

$$\begin{aligned}
&\|T(x+ay) - [T(x) + aT(y)]\|^2 \\
&\qquad = \|[T(x + ay) - T(x)] - aT(y)\|^2 \\
&\qquad = \|T(x + ay) - T(x)\|^2 - 2a\langle T(x + ay) - T(x), T(y)\rangle + a^2 \|T(y)\|^2
\end{aligned}$$

$$\begin{aligned}
&= \|(x + ay) - x\|^2 - 2a\langle T(x + ay), T(y)\rangle + 2a\langle T(x), T(y)\rangle + a^2 \|y\|^2 \\
&= a^2 \|y\|^2 - 2a\langle x + ay, y\rangle + 2a\langle x, y\rangle + a^2 \|y\|^2 \\
&= 2a^2 \|y\|^2 - 2a\langle x, y\rangle - 2a^2 \|y\|^2 + 2a\langle x, y\rangle \\
&= 0.
\end{aligned}$$

Since the inner product is positive definite, it follows that the vector inside the norm on the left hand side must equal zero, and hence $T(x+ay) = T(x)+aT(y)$ as claimed.

In summary, we have shown that T is a linear transformation and, additionally, it is an isometry with the property that $T(0) = 0$ so it also preserves inner products. This is just the definition of an orthogonal transformation. But now note that $F(x) = T(x) + F(0)$ so that F is in fact the composition of the orthogonal transformation T with a translation by $F(0)$. In other words, we have proved that in fact *any isometry is the composition of an orthogonal transformation followed by a translation*. This result is known as **Chasles' theorem**.

Furthermore, this decomposition is unique. To see this, suppose we have two such decompositions

$$F(x) = T_1(x) + v_1 = T_2(x) + v_2$$

where T_1 and T_2 are orthogonal transformations and $v_1, v_2 \in V$ are the translation vectors. Since $T_i(0) = 0$, letting $x = 0$ shows that $v_1 = v_2$ so the translation is unique. But then this leaves $T_1(x) = T_2(x)$ for all $x \in V$ so that $T_1 = T_2$ also.

Finally, let us show what our definition of orthogonal transformation means in terms of matrix representations. Let $T \in L(U, V)$ where U has the basis $\{e_1, \dots, e_n\}$ and V has the basis $\{\bar{e}_1, \dots, \bar{e}_m\}$. Then the matrix representation $A = (a^i{}_j)$ of T is defined by

$$T(e_i) = \sum_{j=1}^{m} \bar{e}_j a^j{}_i \qquad \text{for } i = 1, \dots, n. \tag{4.14}$$

If T is orthogonal, then by definition this means that

$$\langle T(e_i), T(e_j)\rangle = \langle e_i, e_j\rangle \tag{4.15}$$

and therefore this equation becomes (using the summation convention)

$$\langle \bar{e}_k a^k{}_i, \bar{e}_l a^l{}_j\rangle = a^k{}_i a^l{}_j \langle \bar{e}_k, \bar{e}_l\rangle = \langle e_i, e_j\rangle.$$

Now, *if both $\{e_i\}$ and $\{\bar{e}_j\}$ are orthonormal bases*, then the left hand side of equation (4.15) becomes (note this is just the scalar product $A^i \cdot A^j$)

$$a^k{}_i a^l{}_j \delta_{kl} = a^k{}_i a_{kj} = (a^T)_i{}^k a_{kj} \tag{4.16}$$

and this must equal δ_{ij} (the right hand side of equation (4.15)). In other words, $A^T A = I$ so that (at least in the finite-dimensional case) $A^T = A^{-1}$, which is what we took earlier as the *definition* of an orthogonal transformation (or matrix). Note this means that we also have $AA^T = I$.

We emphasize that this result depended on the fact that the bases were orthonormal. Let us see what this implies for the matrix A. With respect to the basis $\{\bar{e}_i\}$, we have that $\bar{e}_1$ has components $(1, 0, 0, \dots, 0)$, $\bar{e}_2$ has components $(0, 1, 0, \dots, 0)$, and so forth down to $\bar{e}_m = (0, 0, 0, \dots, 1)$. Keeping in mind that our vectors are really columns, we see from equation (4.14) that T takes the ith basis vector e_i of U into the ith column of the matrix $A = (a^j{}_i)$. But then equation (4.15) shows that these columns are orthonormal. Furthermore, writing out $AA^T = I$ in terms of components we have

$$a_{ij}(a^T)^j{}_k = a_{ij}a_k{}^j = \delta_{ik}$$

which is the row version of equation (4.16) and shows that the rows of A are also orthonormal (i.e., this is just $A_i \cdot A_k = \delta_{ik}$). In other words, an orthogonal matrix has rows and columns that both form orthonormal sets of vectors.

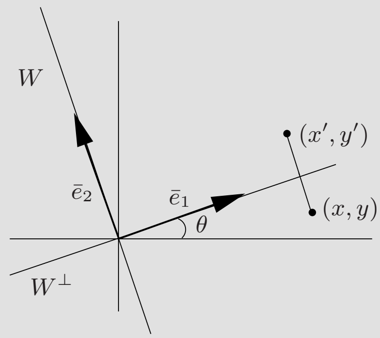
Now, we have defined an orthogonal transformation (matrix) A as one for which $A^T = A^{-1}$, and we saw that rotations in $\mathbb{R}^2$ are examples of this. An interesting and important question is whether or not all orthogonal transformations are in fact rotations. Well, consider the linear transformation $T \in L(\mathbb{R}^2)$ defined by $T(x, y) = (-x, y)$. With respect to the standard basis, T has the matrix representation

$$A = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$$

and it is easy to see that $A^{-1} = A^T = A$ so that A is orthogonal. This is clearly not a rotation (there is no angle θ in equation (4.4) that satisfies this), and whereas taking the determinant of equation (4.4) yields $+1$, here we have $\det A = -1$. This type of orthogonal transformation is called a **reflection**.

To be precise, first recall from Theorem 1.22 that if W is a subspace of a finite-dimensional inner product space V, then we can always write $V = W \oplus W^\perp$. If W is a one-dimensional subspace of $\mathbb{R}^2$, then a linear operator $T \in L(\mathbb{R}^2)$ is said to be a **reflection about** $W^\perp$ if $T(w) = w$ for all $w \in W^\perp$ and $T(w) = -w$ for all $w \in W$. In the example we just gave, the transformation T is a reflection about the subspace of $\mathbb{R}^2$ that is the y-axis.

Example 4.18. Let W be the subspace of $\mathbb{R}^2$ that is a line passing through the origin as shown. Let $\bar{e}_1$ be a basis for $W^\perp$ and $\bar{e}_2$ be a basis for W.



Relative to the standard basis for $\mathbb{R}^2$ we have

$$\bar{e}_1 = (\cos\theta, \sin\theta) = (\cos\theta)e_1 + (\sin\theta)e_2$$

and

$$\bar{e}_2 = (-\sin\theta, \cos\theta) = (-\sin\theta)e_1 + (\cos\theta)e_2.$$

If T is the reflection about $W^\perp$, then $T(\bar{e}_1) = \bar{e}_1$ and $T(\bar{e}_2) = -\bar{e}_2$ so the matrix representation of T with respect to the basis $\{\bar{e}_i\}$ is

$$[T]_{\bar{e}} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

The transition matrix from $\{e_i\}$ to $\{\bar{e}_i\}$ is

$$P = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

which is orthogonal (it's just a rotation of the coordinates) so that

$$P^{-1} = P^T = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}.$$

Then the matrix representation of T with respect to the standard basis for $\mathbb{R}^2$ is

$$\begin{aligned} [T]_e = P[T]_{\bar{e}}P^{-1} &= \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \\ &= \begin{bmatrix} \cos^2\theta - \sin^2\theta & 2\sin\theta\cos\theta \\ 2\sin\theta\cos\theta & \sin^2\theta - \cos^2\theta \end{bmatrix} \\ &= \begin{bmatrix} \cos 2\theta & \sin 2\theta \\ \sin 2\theta & -\cos 2\theta \end{bmatrix} \qquad (4.17) \end{aligned}$$

Now look at what happens to a point $\mathbf{r} = (x, y)$ in the plane. Let $\tan\alpha = y/x$ so this point makes an angle α with respect to the x-axis. Then with respect to the basis $\{\bar{e}_1, \bar{e}_2\}$ this point has coordinates

$$\mathbf{r} = (\bar{x}, \bar{y}) = (r\cos(\theta - \alpha), -r\sin(\theta - \alpha))$$

The action of T on $(\bar{x}, \bar{y})$ is the point

$$T(\bar{x}, \bar{y}) = (\bar{x}, -\bar{y}) = (r\cos(\theta - \alpha), r\sin(\theta - \alpha))$$

which is the point (x', y') shown in the figure above. Thus we see that T indeed represents a reflection about the line through the origin.

We now claim that any orthogonal transformation T on $\mathbb{R}^2$ is either a rotation (so that $\det T = +1$) or a reflection about some line through the origin (so that $\det T = -1$). To prove this, we first show that any orthogonal transformation on a finite-dimensional space V takes an orthonormal basis into another orthonormal basis. But this is easy since by definition T is nonsingular so that (by the rank theorem) $\operatorname{rank} T = \dim(\operatorname{Im} T) = \dim V$. Therefore, if $\{e_1, \dots, e_n\}$ is an orthonormal basis for V, then $\{T(e_1), \dots, T(e_n)\}$ must be linearly independent, and it is also orthonormal since

$$\begin{aligned} \langle T(e_i), T(e_j)\rangle &= \langle e_k a^k{}_i, e_l a^l{}_j\rangle = a^k{}_i a^l{}_j \langle e_k, e_l\rangle = a^k{}_i a^l{}_j \delta_{kl} = a^k{}_i a_{kj} \\ &= (a^T)_{jk} a^k{}_i = \delta_{ij} \end{aligned}$$

where $A = (a_{ij})$ is the matrix representation of T. (We will prove this again in a slightly different manner in the next chapter.)

Going back to $\mathbb{R}^2$, that fact that $T(e_1)$ and $T(e_2)$ are unit vectors (they are orthonormal) means that there is an angle θ such that the vector $T(e_1)$ has components $(\cos\theta, \sin\theta)$ (just look at the figure at the beginning of this section). If we think of this as defining a line through the origin with slope $\sin\theta/\cos\theta$, then $T(e_2)$ must lie on the line perpendicular to this and through the origin, so it has slope $-\cos\theta/\sin\theta$. Therefore $T(e_2)$ must have components that are either $(-\sin\theta, \cos\theta)$ or $(\sin\theta, -\cos\theta)$.

In the first case we have

$$A = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

so that $\det A = \det T = +1$. Referring to equation (4.4) and the figure above it, it is clear that this A represents a rotation in the plane.

In the second case we have

$$A = \begin{bmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{bmatrix}$$

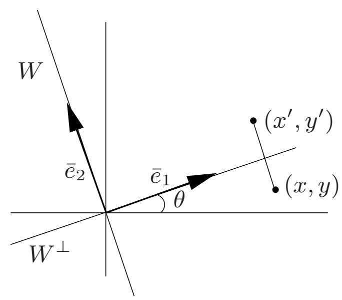
so that $\det A = \det T = -1$. Comparing this with equation (4.17), we see that this represents a reflection about the line making an angle $\theta/2$ with respect to the x-axis.

Putting this all together we see that indeed every orthogonal transformation of $\mathbb{R}^2$ is either a rotation or a reflection as claimed. In other words, we have proved

Theorem 4.17. *Let $T \in L(\mathbb{R}^2)$ be orthogonal. Then T is either a rotation or a reflection, and in fact is a rotation if and only if* $\det T = +1$ *and a reflection if and only if* $\det T = -1$.

Exercises

1. Referring to the figure below,



the reflection T about $W^\perp$ is defined by its matrix with respect to the basis $\{\bar{\mathbf{e}}_i\}$ as

$$[T]_{\bar{e}} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

as we saw in Example 4.18. The (*orthogonal*) transition matrix from the standard basis $\{\mathbf{e}_i\}$ for $\mathbb{R}^2$ to the basis $\{\bar{\mathbf{e}}_i\}$ for $W^\perp \oplus W$ is given by

$$P = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

and therefore the matrix of T with respect to the standard basis $\{\mathbf{e}_i\}$ is given by

$$\begin{aligned} [T]_e = P[T]_{\bar{e}}P^{-1} &= \begin{bmatrix} \cos^2\theta - \sin^2\theta & 2\sin\theta\cos\theta \\ 2\sin\theta\cos\theta & \sin^2\theta - \cos^2\theta \end{bmatrix} \\ &= \begin{bmatrix} \cos 2\theta & \sin 2\theta \\ \sin 2\theta & -\cos 2\theta \end{bmatrix}. \end{aligned}$$

With respect to $\{\mathbf{e}_i\}$, the point (x, y) has coordinates

$$\begin{bmatrix} x \\ y \end{bmatrix}_e = \begin{bmatrix} r\cos\alpha \\ r\sin\alpha \end{bmatrix}$$

where α is defined by $\tan\alpha = y/x$. And with respect to $\{\bar{\mathbf{e}}_i\}$ this same point has coordinates

$$\begin{bmatrix} x \\ y \end{bmatrix}_{\bar{e}} = \begin{bmatrix} \bar{x} \\ \bar{y} \end{bmatrix} = \begin{bmatrix} r\cos(\theta - \alpha) \\ -r\sin(\theta - \alpha) \end{bmatrix}.$$

(a) Show that

$$\begin{bmatrix} \bar{x} \\ \bar{y} \end{bmatrix} = \begin{bmatrix} x \\ y \end{bmatrix}_{\bar{e}} = P^{-1} \begin{bmatrix} x \\ y \end{bmatrix}_e .$$

(b) What are the coordinates of the reflected point (x', y') with respect to the bases $\{\bar{\mathbf{e}}_i\}$ and $\{\mathbf{e}_i\}$?

(c) Show that

$$\begin{bmatrix} x' \\ y' \end{bmatrix}_e = P \begin{bmatrix} x' \\ y' \end{bmatrix}_{\bar{e}} .$$

2. Let A, B and C be linear operators on $\mathbb{R}^2$ with the following matrices relative to the standard basis $\{e_i\}$:

$$[A]_e = \begin{bmatrix} 4 & 6 \\ -2 & -3 \end{bmatrix} \qquad [B]_e = \begin{bmatrix} 1/2 & -\sqrt{3}/2 \\ \sqrt{3}/2 & 1/2 \end{bmatrix}$$

$$[C]_e = \begin{bmatrix} 7 & 3 \\ -10 & -4 \end{bmatrix}.$$

(a) If $f_1 = (2, -1)$ and $f_2 = (3, -2)$, show that $A(f_1) = f_1$ and $A(f_2) = 0$.
(b) Find $[A]_f$.
(c) What is the geometric effect of A?
(d) Show that B is a rotation about the origin of the xy-plane, and find the angle of rotation.
(e) If $f_1 = (1, -2)$ and $f_2 = (3, -5)$, find $C(f_1)$ and $C(f_2)$.
(f) Find $[C]_f$.
(g) What is the geometric effect of C?

Chapter 5

Eigenvalues and Eigenvectors

The concept of an eigenvector is one of the most important applications of linear algebra in all of mathematics, physics and engineering. For example, all of modern quantum theory is based on the existence of eigenvectors and eigenvalues of Hermitian operators. In this chapter we will take a careful look at when such eigenvectors exist and how to find them. However, since eigenvalues are the roots of polynomials, we first need to review some elementary properties of polynomials.

5.1 Polynomials

It is not our intention in this book to treat the general theory of polynomials. However, a brief look at some of their basic properties is useful in examining the existence of eigenvalues. The reader should feel free to skim this section and only read the statements of the theorems if desired.

We assume that the reader has a basic familiarity with polynomials, and knows what it means to add and multiply them. Let us denote the set of all polynomials in the indeterminant (or variable) x over the field $\mathcal{F}$ by $\mathcal{F}[x]$. You should know that if $p \in \mathcal{F}[x]$ is of the form $p = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n$ (where $a_n \neq 0$), then the integer n is called the **degree** of p and is denoted by $\deg p$.

The first basic result we need is the following formal statement of the process of long division that you should have learned in high school. Because the proof is by induction and not very enlightening, we omit it. The polynomials q and r defined in the theorem are called the **quotient** and **remainder** respectively. Simply put, dividing a polynomial f by a polynomial g gives a quotient q and remainder r.

Theorem 5.1 (**Division Algorithm**). *Given $f, g \in \mathcal{F}[x]$ with $g \neq 0$, there exist unique polynomials $q, r \in \mathcal{F}[x]$ such that*

$$f = qg + r$$

where either $r = 0$ or $\deg r < \deg g$.

Example 5.1. Consider the polynomials

$$\begin{aligned} f &= 2x^4 + x^2 - x + 1 \\ g &= 2x - 1. \end{aligned}$$

Define the polynomial

$$f_1 = f - x^3 g = x^3 + x^2 - x + 1.$$

Now let

$$f_2 = f_1 - (1/2)x^2 g = (3/2)x^2 - x + 1.$$

Again, we let

$$f_3 = f_2 - (3/4)xg = (-1/4)x + 1$$

so that

$$f_4 = f_3 + (1/8)g = 7/8.$$

Since $\deg(7/8) < \deg g$, we are finished with the division. Combining the above polynomials we see that

$$f = [x^3 + (1/2)x^2 + (3/4)x - (1/8)]g + f_4$$

and therefore

$$\begin{aligned} q &= x^3 + (1/2)x^2 + (3/4)x - (1/8) \\ r &= 7/8. \end{aligned}$$

This may also be written out in a more familiar form as

$$
\begin{array}{r|llllll}
\multicolumn{1}{r}{} & & x^3 & +\,(1/2)x^2 & +\,(3/4)x & -\,(1/8) & \\ \cline{2-7}
2x-1 & 2x^4 & + & x^2\,- & x\,+ & 1 & \\
\multicolumn{1}{r}{} & 2x^4\,- & x^3 & & & & \\ \cline{2-7}
\multicolumn{1}{r}{} & & x^3\,+ & x^2\,- & x\,+ & 1 & \\
\multicolumn{1}{r}{} & & x^3 - (1/2)x^2 & & & & \\ \cline{3-7}
\multicolumn{1}{r}{} & & & (3/2)x^2\,- & x\,+ & 1 & \\
\multicolumn{1}{r}{} & & & (3/2)x^2 - (3/4)x & & & \\ \cline{4-7}
\multicolumn{1}{r}{} & & & & -\,(1/4)x\,+ & 1 & \\
\multicolumn{1}{r}{} & & & & -\,(1/4)x + (1/8) & & \\ \cline{5-7}
\multicolumn{1}{r}{} & & & & & 7/8 &
\end{array}
$$

It should be noted that at each step in the division, we eliminated the highest remaining power of f by subtracting the appropriate multiple of g.

Since our goal is to be able to find and discuss the roots of polynomials, the utility of Theorem 5.1 lies in our next two results. But first we need some additional terminology.

If $f(x)$ is a polynomial in $\mathcal{F}[x]$, then $c \in \mathcal{F}$ is said to be a **zero** (or **root**) of f if $f(c) = 0$. We shall also sometimes say that c is a **solution** of the polynomial equation $f(x) = 0$. If $f, g \in \mathcal{F}[x]$ and $g \neq 0$, then we say that f is **divisible** by g (or g **divides** f) over $\mathcal{F}$ if $f = qg$ for some $q \in \mathcal{F}[x]$. In other words, f is divisible by g if the remainder in the division of f by g is zero. In this case we also say that g is a **factor** of f (over $\mathcal{F}$).

Theorem 5.2 (**Remainder Theorem**). *Suppose $f \in \mathcal{F}[x]$ and $c \in \mathcal{F}$. Then the remainder in the division of f by $x - c$ is $f(c)$. In other words,*

$$f(x) = (x - c)q(x) + f(c).$$

Proof. We see from the division algorithm that $f = (x - c)q + r$ where either $r = 0$ or $\deg r < \deg(x - c) = 1$, and hence either $r = 0$ or $\deg r = 0$ (in which case $r \in \mathcal{F}$). In either case, we may substitute c for x to obtain

$$f(c) = (c - c)q(c) + r = r.$$ ∎

Corollary (**Factor Theorem**). *If $f \in \mathcal{F}[x]$ and $c \in \mathcal{F}$, then $x - c$ is a factor of f if and only if $f(c) = 0$.*

Proof. Rephrasing the statement of the corollary as $f = q(x - c)$ if and only if $f(c) = 0$, it is clear that this follows directly from the theorem. ∎

Example 5.2. If we divide $f = x^3 - 5x^2 + 7x$ by $g = x - 2$, we obtain $q = x^2 - 3x + 1$ and $r = 2$. It is also easy see that $f(2) = 8 - 5(4) + 7(2) = 2$ as it should according to Theorem 5.2.

Since the only fields we are dealing with are $\mathbb{R}$ and $\mathbb{C}$, let us see just what the difference is between them concerning the roots of polynomials. By way of terminology, a field $\mathcal{F}$ is said to be **algebraically closed** if every polynomial $f \in \mathcal{F}[x]$ with $\deg f > 0$ has at least one zero (or root) in $\mathcal{F}$.

Our next theorem is called the **Fundamental Theorem of Algebra**. This is the result on which much of the remainder of this text is based, because it states that any polynomial over $\mathbb{C}$ always has a root. In particular, we will always be able to find eigenvalues over a complex space, and this will also allow us to put any complex matrix into triangular form. Unfortunately, the proof of this result is based either on the theory of complex variables or on the compactness of metric spaces, and in either case is far beyond the scope of this book.

Theorem 5.3 (**Fundamental Theorem of Algebra**). *The complex number field $\mathbb{C}$ is algebraically closed.*

As another bit of terminology, a polynomial $p = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n$ is said to be **monic** if $a_n = 1$. The fundamental theorem of algebra together with the factor theorem now gives us the result we need.

Theorem 5.4. *Let $\mathcal{F}$ be an algebraically closed field. Then every monic polynomial $f \in \mathcal{F}[x]$ can be factored into the form*

$$f = \prod_{i=1}^{n}(x - a_i)$$

where each $a_i \in \mathcal{F}$.

Proof. Let $f \in \mathcal{F}[x]$ be of degree $n \geq 1$. Since $\mathcal{F}$ is algebraically closed there exists $a_1 \in \mathcal{F}$ such that $f(a_1) = 0$, and hence by the factor theorem,

$$f = (x - a_1)q_1$$

where $q_1 \in \mathcal{F}[x]$ and $\deg q_1 = n - 1$. (This is a consequence of the general fact that if $\deg p = m$ and $\deg q = n$, then $\deg pq = m + n$. Just look at the largest power of x in the product $pq = (a_0 + a_1x + a_2x^2 + \cdots + a_mx^m)(b_0 + b_1x + b_2x^2 + \cdots + b_nx^n)$.)

Now, by the algebraic closure of $\mathcal{F}$ there exists $a_2 \in \mathcal{F}$ such that $q_1(a_2) = 0$, and therefore

$$q_1 = (x - a_2)q_2$$

where $\deg q_2 = n - 2$. It is clear that we can continue this process a total of n times, finally arriving at

$$f = c(x - a_1)(x - a_2)\cdots(x - a_n)$$

where $c \in \mathcal{F}$ is nonzero. In particular, $c = 1$ if q_{n-1} is monic. ∎

Observe that Theorem 5.4 shows that any polynomial of degree n over an algebraically closed field has exactly n roots, but it doesn't require that these roots be distinct, and in general they are not.

Note also that while Theorem 5.3 shows that the field $\mathbb{C}$ is algebraically closed, it is not true that $\mathbb{R}$ is algebraically closed. This should be obvious because any quadratic equation of the form $ax^2 + bx + c = 0$ has solutions given by the quadratic formula

$$\frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

and if $b^2 - 4ac < 0$, then there is no solution for x in the real number system. However, in the case of $\mathbb{R}[x]$, we do have the following result.

Theorem 5.5. *Suppose $f = a_0 + a_1 x + \cdots + a_n x^n \in \mathbb{R}[x]$. If $\alpha \in \mathbb{C}$ is a root of f, then so is α^*. Furthermore, if $\alpha \neq \alpha^*$, then $(x - \alpha)(x - \alpha^*)$ is a factor of f.*

Proof. If $\alpha \in \mathbb{C}$ is a root of f, then $a_0 + a_1\alpha + \cdots + a_n\alpha^n = 0$. Taking the complex conjugate of this equation and remembering that each $a_i \in \mathbb{R}$, we obtain $a_0 + a_1\alpha^* + \cdots + a_n{\alpha^*}^n = 0$ so that α^* is also a root of f. The second part of the theorem now follows directly from the factor theorem: If α is a root of f we can write $f = (x - \alpha)g$ so that $0 = f(\alpha^*) = (\alpha^* - \alpha)g(\alpha^*)$. Then $g(\alpha^*) = 0$ so that $g = (x - \alpha^*)h$ and hence $f = (x - \alpha)(x - \alpha^*)h$. Thus $(x - \alpha)(x - \alpha^*)$ is a factor of f. ∎

Exercises

1. Use the division algorithm (i.e., long division) to find the quotient and remainder when $f = 2x^4 - x^3 + x - 1 \in \mathbb{R}[x]$ is divided by $g = 3x^3 - x^2 + 3 \in \mathbb{R}[x]$.

2. Find the remainder when $ix^9 + 3x^7 + x^6 - 2ix + 1 \in \mathbb{C}[x]$ is divided by $x + i \in \mathbb{C}[x]$.

3. Factor the following polynomials into their prime factors in both $\mathbb{R}[x]$ and $\mathbb{Q}[x]$:
   (a) $2x^3 - x^2 + x + 1$.
   (b) $3x^3 + 2x^2 - 4x + 1$.
   (c) $x^6 + 1$.
   (d) $x^4 + 16$.

4. Find the greatest common divisor of the following pairs of polynomials over $\mathbb{R}[x]$. Express your result in the form defined in Theorem **??**.
   (a) $4x^3 + 2x^2 - 2x - 1$ and $2x^3 - x^2 + x + 1$.
   (b) $x^3 - x + 1$ and $2x^4 + x^2 + x - 5$.
   (c) $x^4 + 3x^2 + 2$ and $x^5 - x$.
   (d) $x^3 + x^2 - 2x - 2$ and $x^4 - 2x^3 + 3x^2 - 6x$.

5. Use the remainder theorem to find the remainder when $2x^5 - 3x^3 + 2x + 1 \in \mathbb{R}[x]$ is divided by:

(a) $x - 2 \in \mathbb{R}[x]$.
(b) $x + 3 \in \mathbb{R}[x]$.

6. (a) Is $x - 3$ a factor of $3x^3 - 9x^2 - 7x + 21$ over $\mathbb{Q}[x]$?
(b) Is $x + 2$ a factor of $x^3 + 8x^2 + 6x - 8$ over $\mathbb{R}[x]$?
(c) For which $k \in \mathbb{Q}$ is $x - 1$ a factor of $x^3 + 2x^2 + x + k$ over $\mathbb{Q}[x]$?
(d) For which $k \in \mathbb{C}$ is $x + i$ a factor of $ix^9 + 3x^7 + x^6 - 2ix + k$ over $\mathbb{C}[x]$?

7. Find the greatest common divisor and least common multiple of the following pairs of polynomials:
(a) $(x-1)(x+2)^2$ and $(x+2)(x-4)$.
(b) $(x-2)^2(x-3)^4(x-i)$ and $(x-1)(x-2)(x-3)^3$.
(c) $(x^2+1)(x^2-1)$ and $(x+i)^3(x^3-1)$.

8. Let $V_n \subset \mathcal{F}[x]$ denote the set of all polynomials of degree $\le n$, and let $a_0, a_1, \ldots, a_n \in \mathcal{F}$ be distinct.
(a) Show V_n is a vector space over $\mathcal{F}$ with basis $\{1, x, x^2, \ldots, x^n\}$, and hence that $\dim V_n = n + 1$.
(b) For each $i = 0, \ldots, n$ define the mapping $T_i : V_n \to \mathcal{F}$ by $T_i(f) = f(a_i)$. Show that the T_i are linear functionals on V_n, i.e., that $T_i \in {V_n}^*$.
(c) For each $k = 0, \ldots, n$ define the polynomial

$$\begin{aligned} p_k(x) &= \frac{(x-a_0)\cdots(x-a_{k-1})(x-a_{k+1})\cdots(x-a_n)}{(a_k-a_0)\cdots(a_k-a_{k-1})(a_k-a_{k+1})\cdots(a_k-a_n)} \\ &= \prod_{i\ne k}\left(\frac{x-a_i}{a_k-a_i}\right) \in V_n. \end{aligned}$$

Show that $T_i(p_j) = \delta_{ij}$.
(d) Show $p_0, \ldots, p_n$ forms a basis for V_n, and hence that any $f \in V_n$ may be written as

$$f = \sum_{i=0}^{n} f(a_i) p_i .$$

(e) Now let $b_0, b_1, \ldots, b_n \in \mathcal{F}$ be arbitrary, and define $f = \sum b_i p_i$. Show $f(a_j) = b_j$ for $0 \le j \le n$. Thus there exists a polynomial of degree $\le n$ that takes on given values at $n + 1$ distinct points.
(f) Now assume that $f, g \in \mathcal{F}[x]$ are of degree $\le n$ and satisfy $f(a_j) = b_j = g(a_j)$ for $0 \le j \le n$. Prove $f = g$, and hence that the polynomial defined in part (e) is unique. This is called the **Lagrange interpolation formula**.

5.2 Eigenvalues and Eigenvectors

We begin with some basic definitions. If $T \in L(V)$, then an element $\lambda \in \mathcal{F}$ is called an **eigenvalue** (also called a **characteristic value** or **characteristic**

root) of T if there exists a nonzero vector $v \in V$ such that $T(v) = \lambda v$. In this case, we call the vector v an **eigenvector** (or **characteristic vector**) belonging to the eigenvalue λ. Note that while an eigenvector is nonzero by definition, an eigenvalue may very well be zero.

Throughout the remainder of this chapter we will frequently leave off the parentheses around vector operands. In other words, we sometimes write Tv rather than $T(v)$. This simply serves to keep our notation as uncluttered as possible.

If T has an eigenvalue λ, then $Tv = \lambda v$ or $(T - \lambda)v = 0$. But then $v \in \operatorname{Ker}(T - \lambda 1)$ and $v \neq 0$ so that $T - \lambda 1$ is singular. (Note that here we use 1 as the identity operator, because λ itself is just a scalar. When we deal with matrices, then instead of writing $\lambda 1$ we will write λI.) Conversely, if $T - \lambda 1$ is singular, then there exists $v \neq 0$ such that $(T - \lambda 1)v = 0$ so that $Tv = \lambda v$. This proves the following.

Theorem 5.6. *A linear operator $T \in L(V)$ has eigenvalue $\lambda \in \mathcal{F}$ if and only if $\lambda 1 - T$ is singular.*

Note, in particular, that 0 is an eigenvalue of T if and only if T is singular. And, as a side remark, observe that there is no essential difference between writing $(T - \lambda)v = 0$ and $(\lambda - T)v = 0$. One or the other of these forms may be more preferable depending on just what calculation we are doing, and we will freely go back and forth between them.

In an exactly analogous manner, we say that an element $\lambda \in \mathcal{F}$ is an **eigenvalue** of a matrix $A \in M_n(\mathcal{F})$ if there exists a nonzero (column) vector $v \in \mathcal{F}^n$ such that $Av = \lambda v$, and we call v an **eigenvector** of A belonging to the eigenvalue λ. Given a basis $\{e_i\}$ for $\mathcal{F}^n$, we can write this matrix eigenvalue equation in terms of components as $a^i{}_j v^j = \lambda v^i$ or, written out as

$$\sum_{j=1}^{n} a_{ij} v_j = \lambda v_i, \qquad i = 1, \dots, n.$$

Now suppose $T \in L(V)$ and $v \in V$. If $\{e_1, \dots, e_n\}$ is a basis for V, then $v = \sum v_i e_i$ and hence

$$T(v) = T\Big(\sum_i v_i e_i\Big) = \sum_i v_i T(e_i) = \sum_{i,j} e_j a_{ji} v_i$$

where $A = (a_{ij})$ is the matrix representation of T relative to the basis $\{e_i\}$. Using this result, we see that if $T(v) = \lambda v$, then

$$\sum_{i,j} e_j a_{ji} v_i = \lambda \sum_j v_j e_j$$

and hence equating components shows that $\sum_i a_{ji} v_i = \lambda v_j$. We thus see that (as expected) the isomorphism between $L(V)$ and $M_n(\mathcal{F})$ (see Theorem 4.11) shows that λ is an eigenvalue of the linear transformation T if and only if λ is also an eigenvalue of the corresponding matrix representation A. Using the notation of Chapter 4, we can say that $T(v) = \lambda v$ if and only if $[T]_e[v]_e = \lambda[v]_e$.

It is important to point out that eigenvectors are only specified up to an overall constant. This is because if $Tv = \lambda v$, then for any $c \in \mathcal{F}$ we have $T(cv) = cT(v) = c\lambda v = \lambda(cv)$ so that cv is also an eigenvector with eigenvalue λ. Because of this, we are always free to normalize the eigenvectors to any desired value.

Example 5.3. Let us find all of the eigenvectors and associated eigenvalues of the matrix

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

This means that we must find a vector $v = (x, y)$ such that $Av = \lambda v$. In matrix notation, this equation takes the form

$$\begin{bmatrix} 1 & 2 \\ 3 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \lambda \begin{bmatrix} x \\ y \end{bmatrix}$$

and the equation $(T - \lambda I)v = 0$ becomes

$$\begin{bmatrix} 1-\lambda & 2 \\ 3 & 2-\lambda \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = 0\,.$$

This is equivalent to the system

$$\begin{aligned} (1-\lambda)x + 2y &= 0 \\ 3x + (2-\lambda)y &= 0 \end{aligned} \tag{5.1}$$

Since this homogeneous system of equations has a nontrivial solution if and only if the determinant of the coefficient matrix is zero (Corollary to Theorem 3.11), we must have

$$\begin{vmatrix} 1-\lambda & 2 \\ 3 & 2-\lambda \end{vmatrix} = \lambda^2 - 3\lambda - 4 = (\lambda - 4)(\lambda + 1) = 0\,.$$

We thus see that the eigenvalues are $\lambda = 4$ and $\lambda = -1$. (The roots of this polynomial are found either by inspection, or by applying the elementary quadratic formula.)

Substituting $\lambda = 4$ into equations (5.1) yields

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

or $y = (3/2)x$. This means that every eigenvector corresponding to the eigenvalue $\lambda = 4$ has the form $v = (x, 3x/2)$. In other words, every multiple of the vector $v = (2, 3)$ is also an eigenvector with eigenvalue equal to 4. If we substitute $\lambda = -1$ in equations (5.1), then we similarly find $y = -x$, and hence every multiple of the vector $v = (1, -1)$ is an eigenvector with eigenvalue equal to -1. (Note that both of equations (5.1) give the same information. This is not surprising because the determinant of the coefficients vanishes so we know that the rows are linearly dependent, and hence each supplies the same information.)

Recall from the factor theorem (Corollary to Theorem 5.2) that if $c \in \mathcal{F}$ is a root of $f \in \mathcal{F}[x]$, then $(x - c)$ divides f. If c is such that $(x - c)^m$ divides f but no higher power of $x - c$ divides f, then we say that c is a root of **multiplicity** m. In counting the number of roots that a polynomial has, we shall always count a root of multiplicity m as m roots. A root of multiplicity 1 is frequently called a **simple** root.

Theorem 5.7. *If $v_1, \dots, v_r$ are eigenvectors belonging to the distinct eigenvalues $\lambda_1, \dots, \lambda_r$ of $T \in L(V)$, then the set $\{v_1, \dots, v_r\}$ is linearly independent.*

Proof. We will prove the contrapositive of the statement in the theorem. That is, we assume that the set $\{v_1, \dots, v_r\}$ is linearly *dependent* and show that this leads to a contradiction.

If $\{v_1, \dots, v_r\}$ is linearly dependent, let v_{k+1} be the first vector that is dependent on the previous ones. In other words, $\{v_1, \dots, v_k\}$ is linearly independent, but there exist scalars $c_1, \dots, c_k$ not all equal to 0 such that

$$v_{k+1} = c_1 v_1 + \cdots c_k v_k.$$

Act on this with T and use $Tv_i = \lambda_i v_i$ to obtain

$$\lambda_{k+1} v_{k+1} = c_1 \lambda_1 v_1 + \cdots c_k \lambda_k v_k.$$

On the other hand, we can multiply the first equation by λ_{k+1} and subtract this from the second equation to yield

$$0 = c_1(\lambda_1 - \lambda_{k+1})v_1 + \cdots + c_k(\lambda_k - \lambda_{k+1})v_k.$$

But $v_1, \dots, v_k$ is linearly independent which implies that

$$c_1(\lambda_1 - \lambda_{k+1}) = \cdots = c_k(\lambda_k - \lambda_{k+1})v_k = 0.$$

Since the λ_i are all distinct by hypothesis, this shows that $c_1 = \cdots = c_k = 0$ and hence $v_{k+1} = c_1 v_1 + \cdots + c_k v_k = 0$. But this is impossible since eigenvectors are nonzero by definition. Therefore the set $\{v_1, \dots, v_r\}$ must be linearly *independent.* ∎

Corollary 1. *Suppose* $T \in L(V)$ *and* $\dim V = n$. *Then* T *can have at most* n *distinct eigenvalues in* $\mathcal{F}$.

Proof. Since $\dim V = n$, there can be at most n independent vectors in V. Since n distinct eigenvalues result in n independent eigenvectors, this corollary follows directly from Theorem 5.7. ∎

Corollary 2. *Suppose* $T \in L(V)$ *and* $\dim V = n$. *If* T *has* n *distinct eigenvalues, then there exists a basis for* V *which consists of eigenvectors of* T.

Proof. If T has n distinct eigenvalues, then (by Theorem 5.7) T must have n linearly independent eigenvectors. But n is the number of elements in any basis for V, and hence these n linearly independent eigenvectors in fact form a basis for V. ∎

It should be remarked that one eigenvalue can belong to more than one linearly independent eigenvector. In fact, if $T \in L(V)$ and λ is an eigenvalue of T, then the set

$$V_\lambda := \{v \in V : Tv = \lambda v\}$$

of all eigenvectors of T belonging to λ is a subspace of V called the **eigenspace** of λ. It is also easy to see that $V_\lambda = \operatorname{Ker}(\lambda 1 - T)$ (see Exercise 5.2.1).

Exercises

1. (a) If $T \in L(V)$ and λ is an eigenvalue of T, show that the set V_λ of all eigenvectors of T belonging to λ is a T-invariant subspace of V (i.e., a subspace with the property that $T(v) \in V_\lambda$ for all $v \in V_\lambda$).
   (b) Show $V_\lambda = \operatorname{Ker}(\lambda 1 - T)$.

2. An operator $T \in L(V)$ with the property that $T^n = 0$ for some $n \in \mathbb{Z}^+$ is said to be **nilpotent**. Show that the only eigenvalue of a nilpotent operator is 0.

3. If $S, T \in L(V)$, show that ST and TS have the same eigenvalues. [*Hint*: First use Theorems 4.13 and 5.6 to show that 0 is an eigenvalue of ST if and only if 0 is an eigenvalue of TS. Now assume $\lambda \neq 0$, and let $ST(v) = \lambda v$. Show that Tv is an eigenvector of TS.]

4. (a) Consider the rotation operator $R(\alpha) \in L(\mathbb{R}^2)$ defined in Example **??**. Does $R(\alpha)$ have any eigenvectors? Explain.
   (b) Repeat part (a) but now consider rotations in $\mathbb{R}^3$.

5. For each of the following matrices, find all eigenvalues and linearly independent eigenvectors:

(a) $\begin{bmatrix} 2 & 2 \\ 1 & 3 \end{bmatrix}$ (b) $\begin{bmatrix} 4 & 2 \\ 3 & 3 \end{bmatrix}$ (c) $\begin{bmatrix} 5 & -1 \\ 1 & 3 \end{bmatrix}$

6. Consider the spaces $D[\mathbb{R}]$ and $F[\mathbb{R}]$ defined in Exercise 1.2.6, and let $d : D[\mathbb{R}] \to F[\mathbb{R}]$ be the usual derivative operator.
   (a) Show the eigenfunctions (i.e., eigenvectors) of d are of the form $\exp(\lambda x)$ where λ is the corresponding eigenvalue.
   (b) Suppose $\lambda_1, \ldots, \lambda_r \in \mathbb{R}$ are distinct. Show the set

$$S = \{\exp(\lambda_1 x), \ldots, \exp(\lambda_r x)\}$$

   is linearly independent. [*Hint*: Consider the linear span of S.]

7. Suppose $T \in L(V)$ is invertible. Show that λ is an eigenvalue of T if and only if $\lambda \neq 0$ and λ^{-1} is an eigenvalue of T^{-1}.

8. Suppose $T \in L(V)$ and $\dim V = n$. If T has n linearly independent eigenvectors, what can you say about the matrix representation of T?

9. Let V be a two-dimensional space over $\mathbb{R}$, and let $\{e_1, e_2\}$ be a basis for V. Find the eigenvalues and eigenvectors of the operator $T \in L(V)$ defined by:
   (a) $Te_1 = e_1 + e_2$ $\quad Te_2 = e_1 - e_2$.
   (b) $Te_1 = 5e_1 + 6e_2$ $\quad Te_2 = -7e_2$.
   (c) $Te_1 = e_1 + 2e_2$ $\quad Te_2 = 3e_1 + 6e_2$.

10. Suppose $A \in M_n(\mathbb{C})$ and define $R_i = \sum_{j=1}^n |a_{ij}|$ and $P_i = R_i - |a_{ii}|$.
   (a) Show that if $Ax = 0$ for some nonzero $x = (x_1, \ldots, x_n)$, then for any r such that $x_r \neq 0$ we have

$$|a_{rr}|\,|x_r| = \left| \sum_{j \neq r} a_{rj} x_j \right| .$$

   (b) Show that part (a) implies that for some r we have $|a_{rr}| \leq P_r$.
   (c) Prove that if $|a_{ii}| > P_i$ for all $i = 1, \ldots, n$ then all eigenvalues of A are nonzero (or, equivalently, that $\det A \neq 0$).

11. (a) Suppose $A \in M_n(\mathbb{C})$ and let λ be an eigenvalue of A. Using the previous exercise prove **Gershgorin's Theorem**: $|\lambda - a_{rr}| \leq P_r$ for some r with $1 \leq r \leq n$.
   (b) Use this result to show that every eigenvalue λ of the matrix

$$A = \begin{bmatrix} 4 & 1 & 1 & 0 & 1 \\ 1 & 3 & 1 & 0 & 0 \\ 1 & 2 & 3 & 1 & 0 \\ 1 & 1 & -1 & 4 & 0 \\ 1 & 1 & 1 & 1 & 5 \end{bmatrix}$$

   satisfies $1 \leq |\lambda| \leq 9$.

5.3 Characteristic Polynomials

So far our discussion has dealt only theoretically with the existence of eigenvalues of an operator $T \in L(V)$. From a practical standpoint (as we saw in Example 5.3), it is much more convenient to deal with the matrix representation of an operator. Recall that the definition of an eigenvalue $\lambda \in \mathcal{F}$ and eigenvector $v = \sum v_i e_i$ of a matrix $A = (a_{ij}) \in M_n(\mathcal{F})$ is given in terms of components by $\sum_j a_{ij} v_j = \lambda v_i$ for each $i = 1, \ldots, n$. This may be written in the form

$$\sum_{j=1}^{n} a_{ij} v_j = \lambda \sum_{j=1}^{n} \delta_{ij} v_j$$

or, alternatively, as

$$\sum_{j=1}^{n} (\lambda \delta_{ij} - a_{ij}) v_j = 0.$$

In matrix notation, this is

$$(\lambda I - A)v = 0.$$

By the corollary to Theorem 3.11, this set of homogeneous equations has a nontrivial solution if and only if $\det(\lambda I - A) = 0$.

Another way to see this is to note that by Theorem 5.6, λ is an eigenvalue of the operator $T \in L(V)$ if and only if $\lambda 1 - T$ is singular. But according to Theorem 4.13, this means $\det(\lambda 1 - T) = 0$ (recall that the determinant of a linear transformation T is defined to be the determinant of any matrix representation of T). In other words, λ is an eigenvalue of T if and only if $\det(\lambda 1 - T) = 0$. This proves the following important result.

Theorem 5.8. *Suppose $T \in L(V)$ and $\lambda \in \mathcal{F}$. Then λ is an eigenvalue of T if and only if* $\det(\lambda 1 - T) = 0$.

It is also worth pointing out that there is no real difference between the statements $\det(\lambda 1 - T) = 0$ and $\det(T - \lambda 1) = 0$, and we will use whichever one is most appropriate for what we are doing at the time.

Let $[T]$ be a matrix representation of T. The matrix $xI - [T]$ is called the **characteristic matrix** of $[T]$, and the expression $\det(x1 - T) = 0$ is called the **characteristic** (or **secular**) **equation** of T. The determinant $\det(x1 - T)$ is frequently denoted by $\Delta_T(x)$. Writing out the determinant in a particular basis, we see that $\det(x1 - T)$ is of the form

$$\Delta_T(x) = \begin{vmatrix} x - a_{11} & -a_{12} & \cdots & -a_{1n} \\ -a_{21} & x - a_{22} & \cdots & -a_{2n} \\ \vdots & \vdots & & \vdots \\ -a_{n1} & -a_{n2} & \cdots & x - a_{nn} \end{vmatrix}$$

where $A = (a_{ij})$ is the matrix representation of T in the chosen basis. Since the expansion of a determinant contains exactly one element from each row and each column, we see that (see Exercise 5.3.1)

$$\begin{aligned}\det(x1 - T) = (x - a_{11})(x - a_{22}) \cdots (x - a_{nn}) \\ &+ \text{terms containing } n-1 \text{ factors of the form } x - a_{ii} \\ &+ \cdots + \text{terms with no factors containing } x \\ &= x^n - (\operatorname{tr} A)x^{n-1} + \text{terms of lower degree in } x + (-1)^n \det A.\end{aligned}$$

This monic polynomial is called the **characteristic polynomial** of T.

From the discussion following Theorem 4.15, we see that if $A' = P^{-1}AP$ is similar to A, then

$$\det(xI - A') = \det(xI - P^{-1}AP) = \det[P^{-1}(xI - A)P] = \det(xI - A)$$

(since $\det P^{-1} = (\det P)^{-1}$). We thus see that similar matrices have the same characteristic polynomial (the converse of this statement is *not* true), and hence also the same eigenvalues. Therefore the eigenvalues (*not* eigenvectors) of an operator $T \in L(V)$ do not depend on the basis chosen for V. Note also that according to Section 4.4 (or Exercise 3.3.13), we may as well write $\operatorname{tr} T$ and $\det T$ (rather than $\operatorname{tr} A$ and $\det A$) since these are independent of the particular basis chosen.

Using this terminology, we may rephrase Theorem 5.8 in the following form which we state as a corollary.

Corollary. *A scalar $\lambda \in \mathcal{F}$ is an eigenvalue of $T \in L(V)$ if and only if λ is a root of the characteristic polynomial $\Delta_T(x)$.*

Since the characteristic polynomial is of degree n in x, it follows from Theorem 5.4 that if we are in an algebraically closed field (such as $\mathbb{C}$), then there must exist n roots. In this case, the characteristic polynomial may be factored into the form

$$\det(x1 - T) = (x - \lambda_1)(x - \lambda_2) \cdots (x - \lambda_n)$$

where the eigenvalues λ_i are not necessarily distinct. Expanding this expression we have

$$\det(x1 - T) = x^n - \left(\sum_{i=1}^{n} \lambda_i\right) x^{n-1} + \cdots + (-1)^n \lambda_1 \lambda_2 \cdots \lambda_n.$$

Comparing this with the above general expression for the characteristic polynomial, we see that

$$\operatorname{tr} T = \sum_{i=1}^{n} \lambda_i$$

and

$$\det T = \prod_{i=1}^{n} \lambda_i .$$

It should be remembered that this result only applies to an algebraically closed field (or to any other field $\mathcal{F}$ as long as all n roots of the characteristic polynomial lie in $\mathcal{F}$).

Example 5.4. Let us find the eigenvalues and eigenvectors of the matrix

$$A = \begin{bmatrix} 1 & 4 \\ 2 & 3 \end{bmatrix}.$$

The characteristic polynomial of A is given by

$$\Delta_A(x) = \begin{vmatrix} x-1 & -4 \\ -2 & x-3 \end{vmatrix} = x^2 - 4x - 5 = (x-5)(x+1)$$

and hence the eigenvalues of A are $\lambda = 5, -1$. To find the eigenvectors corresponding to each eigenvalue, we must solve $Av = \lambda v$ or $(\lambda I - A)v = 0$. Written out for $\lambda = 5$ this is

$$\begin{bmatrix} 4 & -4 \\ -2 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

We must therefore solve the set of homogeneous linear equations

$$\begin{aligned} 4x - 4y &= 0 \\ -2x + 2y &= 0 \end{aligned}$$

which is clearly equivalent to the single equation $x - y = 0$, or $x = y$. This means that every eigenvector corresponding to the eigenvalue $\lambda = 5$ is a multiple of the vector $(1, 1)$, and thus the corresponding eigenspace is one-dimensional.

For $\lambda = -1$ we have

$$\begin{bmatrix} -2 & -4 \\ -2 & -4 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

and the equation to be solved is (since both are the same) $-2x - 4y = 0$. The solution is thus $-x = 2y$ so the eigenvector is a multiple of $(2, -1)$.

We now note that

$$\operatorname{tr} A = 1 + 3 = 4 = \sum_{i=1}^{2} \lambda_i$$

and

$$\det A = 3 - 8 = -5 = \prod_{i=1}^{2} \lambda_i .$$

It is also easy to see these relationships hold for the matrix given in Example 5.3.

It is worth remarking that the existence of eigenvalues of a given operator (or matrix) depends on the particular field we are working with. For example, the matrix

$$A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$$

has characteristic polynomial $x^2 + 1$ which has no real roots, but does have the complex roots $\pm i$. In other words, A has no eigenvalues in $\mathbb{R}$, but does have the eigenvalues $\pm i$ in $\mathbb{C}$ (see Exercise 5.3.3).

Let us now take a careful look at what happens if a space V has a basis of eigenvectors of an operator T. Suppose that $T \in L(V)$ with $\dim V = n$. If V has a basis $\{v_1, \dots, v_n\}$ that consists entirely of eigenvectors of T, then the matrix representation of T in this basis is defined by

$$T(v_i) = \sum_{j=1}^{n} v_j a_{ji} = \lambda_i v_i = \sum_{j=1}^{n} \delta_{ji} \lambda_j v_j$$

and therefore $a_{ji} = \delta_{ji}\lambda_j$. In other words, T is represented by a diagonal matrix in a basis of eigenvectors. Conversely, if T is represented by a diagonal matrix $a_{ji} = \delta_{ji}\lambda_j$ relative to some basis $\{v_i\}$, then reversing the argument shows that each v_i is an eigenvector of T. This proves the following theorem.

Theorem 5.9. *A linear operator $T \in L(V)$ can be represented by a diagonal matrix if and only if V has a basis consisting of eigenvectors of T. If this is the case, then the diagonal elements of the matrix representation are precisely the eigenvalues of T. (Note however, that the eigenvalues need not necessarily be distinct.)*

If $T \in L(V)$ is represented in some basis $\{e_i\}$ by a matrix A, and in the basis of eigenvectors $\{v_i\}$ by a diagonal matrix D, then Theorem 4.15 tells us that A and D must be similar matrices. This proves the following version of Theorem 5.9, which we state as a corollary.

Corollary 1. *A matrix $A \in M_n(\mathcal{F})$ is similar to a diagonal matrix D if and only if A has n linearly independent eigenvectors.*

Corollary 2. *A linear operator $T \in L(V)$ can be represented by a diagonal matrix if T has $n = \dim V$ distinct eigenvalues.*

Proof. This follows from Corollary 2 of Theorem 5.7. ∎

Note that the existence of $n = \dim V$ distinct eigenvalues of $T \in L(V)$ is a sufficient but not necessary condition for T to have a diagonal representation. For example, the identity operator has the usual diagonal representation, but its only eigenvalues are $\lambda = 1$. In general, if any eigenvalue has multiplicity greater than 1, then there will be fewer distinct eigenvalues than the dimension of V. However, in this case we *may* be able to choose an appropriate linear combination of eigenvectors in each eigenspace so the matrix of T will still be diagonal. We shall have more to say about this in Section 5.6.

We say that a matrix A is **diagonalizable** if it is similar to a diagonal matrix D. If P is a nonsingular matrix such that $D = P^{-1}AP$, then we say that P **diagonalizes** A. It should be noted that if λ is an eigenvalue of a matrix A with eigenvector v (i.e., $Av = \lambda v$), then for any nonsingular matrix P we have

$$(P^{-1}AP)(P^{-1}v) = P^{-1}Av = P^{-1}\lambda v = \lambda(P^{-1}v).$$

In other words, $P^{-1}v$ is an eigenvector of $P^{-1}AP$. Similarly, we say that $T \in L(V)$ is **diagonalizable** if there exists a basis for V that consists entirely of eigenvectors of T.

While all of this sounds well and good, the reader might wonder exactly how the transition matrix P is to be constructed. Actually, the method has already been given in Section 4.4. If $T \in L(V)$ and A is the matrix representation of T in a basis $\{e_i\}$, then P is defined to be the transformation that takes the basis $\{e_i\}$ into the basis $\{v_i\}$ of eigenvectors. In other words, $v_i = Pe_i = \sum_j e_j p_{ji}$. This means that the ith column of (p_{ji}) is just the ith eigenvector of A. The fact that P must be nonsingular coincides with the requirement that T (or A) have n linearly independent eigenvectors v_i.

Example 5.5. Referring to Example 5.4, we found the eigenvectors $v_1 = (1, 1)$ and $v_2 = (2, -1)$ belonging to the matrix

$$A = \begin{bmatrix} 1 & 4 \\ 2 & 3 \end{bmatrix}.$$

Then P and P^{-1} are given by

$$P = \begin{bmatrix} 1 & 2 \\ 1 & -1 \end{bmatrix}$$

and

$$P^{-1} = \frac{\operatorname{adj} P}{\det P} = \begin{bmatrix} 1/3 & 2/3 \\ 1/3 & -1/3 \end{bmatrix}$$

and therefore

$$D = P^{-1}AP = \begin{bmatrix} 1/3 & 2/3 \\ 1/3 & -1/3 \end{bmatrix} \begin{bmatrix} 1 & 4 \\ 2 & 3 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} 5 & 0 \\ 0 & -1 \end{bmatrix}.$$

We see that D is a diagonal matrix, and that the diagonal elements are just the eigenvalues of A. Note also that

$$\begin{aligned} D(P^{-1}v_1) &= \begin{bmatrix} 5 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} 1/3 & 2/3 \\ 1/3 & -1/3 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} 5 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = 5 \begin{bmatrix} 1 \\ 0 \end{bmatrix} \\ &= \lambda_1(P^{-1}v_1) \end{aligned}$$

with a similar result holding for $P^{-1}v_2$. Observe that since $P^{-1}v_i = e_i$, this is just $De_i = \lambda_i e_i$.

Example 5.6. Let us show that the matrix

$$A = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}$$

is not diagonalizable. The characteristic equation is $\Delta_A(x) = (x-1)^2 = 0$, and hence there are two identical roots $\lambda = 1$. If there existed an eigenvector $v = (x, y)$, it would have to satisfy the equation $(\lambda I - A)v = 0$ or

$$\begin{bmatrix} 0 & -2 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Since this yields $-2y = 0$, the eigenvectors must be of the form $(x, 0)$, and hence it is impossible to find two linearly independent such eigenvectors.

Exercises

1. Suppose $T \in L(V)$ has matrix representation $A = (a_{ij})$, and $\dim V = n$. Prove

$$\begin{aligned} &\det(x1 - T) \\ &\qquad = x^n - (\operatorname{tr} A)x^{n-1} + \text{terms of lower degree in } x + (-1)^n \det A. \end{aligned}$$

   [*Hint*: Use the definition of determinant.]

2. If $T \in L(V)$ and $\Delta_T(x)$ is a product of distinct linear factors, prove that T is diagonalizable.

3. Consider the following matrices:

$$A = \begin{bmatrix} 2 & -1 \\ 1 & 4 \end{bmatrix} \qquad B = \begin{bmatrix} 3 & -1 \\ 13 & -3 \end{bmatrix}$$

(a) Find all eigenvalues and linearly independent eigenvectors over $\mathbb{R}$.
(b) Find all eigenvalues and linearly independent eigenvectors over $\mathbb{C}$.

4. For each of the following matrices, find all eigenvalues, a basis for each eigenspace, and determine whether or not the matrix is diagonalizable:

(a) $\begin{bmatrix} 1 & -3 & 3 \\ 3 & -5 & 3 \\ 6 & -6 & 4 \end{bmatrix}$ (b) $\begin{bmatrix} -3 & 1 & -1 \\ -7 & 5 & -1 \\ -6 & 6 & -2 \end{bmatrix}$

5. Consider the operator $T \in L(\mathbb{R}^3)$ defined by

$$T(x, y, z) = (2x + y, y - z, 2y + 4z).$$

Find all eigenvalues and a basis for each eigenspace.

6. Let $A = (a_{ij})$ be a triangular matrix, and assume that all of the diagonal entries of A are distinct. Is A diagonalizable? Explain.

7. Suppose $A \in M_3(\mathbb{R})$. Show that A can not be a zero of the polynomial $f = x^2 + 1$.

8. If $A \in M_n(\mathcal{F})$, show that A and A^T have the same eigenvalues.

9. Suppose A is a block triangular matrix with square matrices A_{ii} on the diagonal. Show that the characteristic polynomial of A is the product of the characteristic polynomials of the A_{ii}.

10. For each of the following matrices A, find a nonsingular matrix P (if it exists) such that $P^{-1}AP$ is diagonal:

(a) $\begin{bmatrix} 3 & 1 & 1 \\ 2 & 4 & 2 \\ 1 & 1 & 3 \end{bmatrix}$ (b) $\begin{bmatrix} 1 & 2 & 2 \\ 1 & 2 & -1 \\ -1 & 1 & 4 \end{bmatrix}$ (c) $\begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

11. Consider the following real matrix:

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}.$$

Find necessary and sufficient conditions on a, b, c and d so that A is diagonalizable.

12. Let A be an **idempotent** matrix (i.e., $A^2 = A$) of rank r. Show that A is similar to the matrix

$$B = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}.$$

13. Let V be the space of all real polynomials $f \in \mathbb{R}[x]$ of degree at most 2, and define $T \in L(V)$ by $Tf = f + f' + xf'$ where f' denotes the usual derivative with respect to x.

(a) Write down the most obvious basis $\{e_1, e_2, e_3\}$ for V you can think of, and then write down $[T]_e$.

(b) Find all eigenvalues of T, and then find a nonsingular matrix P such that $P^{-1}[T]_e P$ is diagonal.

14. Prove that any real symmetric 2×2 matrix is diagonalizable.

5.4 Block Matrices

Before proceeding, there is another definition that will greatly facilitate our description of invariant subspaces. (See Exercise 4.3.9.) In particular, suppose we are given a matrix $A = (a_{ij}) \in M_{m\times n}(\mathcal{F})$. Then, by partitioning the rows and columns of A in some manner, we obtain what is called a **block matrix**. To illustrate, suppose $A \in M_{3\times 5}(\mathbb{R})$ is given by

$$A = \begin{bmatrix} 7 & 5 & 5 & 4 & -1 \\ 2 & 1 & -3 & 0 & 5 \\ 0 & 8 & 2 & 1 & -9 \end{bmatrix}.$$

Then we may partition A into blocks to obtain (for example) the matrix

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$$

where

$$A_{11} = \begin{bmatrix} 7 & 5 & 5 \end{bmatrix} \qquad A_{12} = \begin{bmatrix} 4 & -1 \end{bmatrix}$$

$$A_{21} = \begin{bmatrix} 2 & 1 & -3 \\ 0 & 8 & 2 \end{bmatrix} \qquad A_{22} = \begin{bmatrix} 0 & 5 \\ 1 & -9 \end{bmatrix}.$$

(Do not confuse these A_{ij} with minor matrices or the entry a_{ij} of A.)

If A and B are block matrices that are partitioned into the same number of blocks such that each of the corresponding blocks is of the same size, then it is clear that (in an obvious notation)

$$A + B = \begin{bmatrix} A_{11} + B_{11} & \cdots & A_{1n} + B_{1n} \\ \vdots & & \vdots \\ A_{m1} + B_{m1} & \cdots & A_{mn} + B_{mn} \end{bmatrix}.$$

In addition, if C and D are block matrices such that the number of columns in each C_{ij} is equal to the number of rows in each D_{jk}, then the product of C and D is also a block matrix CD where $(CD)_{ik} = \sum_j C_{ij} D_{jk}$. Thus block matrices are multiplied as if each block were just a single element of each matrix in the product. In other words, each $(CD)_{ik}$ is a matrix that is the sum of a product of matrices. The proof of this fact is an exercise in matrix multiplication, and is left to the reader (see Exercise 5.4.1).

Theorem 5.10. *If $A \in M_n(\mathcal{F})$ is a block triangular matrix of the form*

$$\begin{bmatrix} A_{11} & A_{12} & A_{13} & \cdots & A_{1k} \\ 0 & A_{22} & A_{23} & \cdots & A_{2k} \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & A_{kk} \end{bmatrix}$$

where each A_{ii} is a square matrix and the 0's are zero matrices of appropriate size, then

$$\det A = \prod_{i=1}^{k} \det A_{ii}.$$

Proof. What is probably the simplest proof of this theorem is outlined in Exercise 5.4.3. However, the proof that follows serves as a good illustration of the meaning of the terms in the definition of the determinant. It is actually a lot easier to understand than it is to explain.

We first note that only the diagonal matrices are required to be square matrices. Because each A_{ii} is square, we can simply prove the theorem for the case $k = 2$, and the general case will then follow by induction. (For example, consider the case where $k = 3$:

$$\left[\begin{array}{c|cc} A_{11} & A_{12} & A_{13} \\ \hline 0 & A_{22} & A_{23} \\ 0 & 0 & A_{33} \end{array}\right]$$

Because each A_{ii} is square, this is a 2×2 block triangular matrix where the two diagonal matrices are square.)

We thus let $A = (a_{ij}) \in M_n(\mathcal{F})$ be of the form

$$A = \begin{bmatrix} B & C \\ 0 & D \end{bmatrix}$$

where $B = (b_{ij}) \in M_r(\mathcal{F})$, $D = (d_{ij}) \in M_s(\mathcal{F})$, $C = (c_{ij}) \in M_{r\times s}(\mathcal{F})$ and $r + s = n$. Note that for $1 \le i, j \le r$ we have $a_{ij} = b_{ij}$, for $1 \le i, j \le s$ we have $a_{i+r,j+r} = d_{ij}$, and if $i > r$ and $j \le r$ then $a_{ij} = 0$.

From the definition of determinant we have

$$\det A = \varepsilon^{i_1 \cdots i_r i_{r+1} \cdots i_{r+s}} a_{1i_1} \cdots a_{ri_r} a_{r+1\, i_{r+1}} \cdots a_{r+s\, i_{r+s}}.$$

Now observe that for rows $r + j$ where $r + 1 \le r + j \le r + s$ we must have the column index $i_{r+j} > r$ or else $a_{r+j\, i_{r+j}} = 0$, and for this range of indices we are summing over the D block. But if the indices $i_{r+1}, \dots, i_{r+s}$ only cover the range $r + 1, \dots r + s$ then the rest of the indices $i_1, \dots, i_r$ can only take the values

$1, \ldots, r$ and for this range of columns the elements $a_{1i_1}, \ldots, a_{ri_r}$ are precisely the elements of B.

For a *fixed* set of indices $i_1, \ldots, i_r$ each of which takes a value between 1 and r, the ε symbol still sums over all possible $i_{r+1}, \ldots, i_{r+s}$ where each index ranges between $r+1$ and $r+s$, with the correct sign of the permutation of those indices. In other words, for each set $i_1, \ldots, i_r$ the rest of the sum gives $\det D$. Then this factor of $\det D$ just multiplies each of the sums over all possible combinations of $i_1, \ldots, i_r$ where each index now ranges between 1 and r, again with the proper sign of the permutation. In other words, it just gives the term $\det B$, and the final result is indeed $\det A = (\det B)(\det D)$. ∎

Example 5.7. Consider the matrix

$$A = \begin{bmatrix} 1 & -1 & 2 & 3 \\ 2 & 2 & 0 & 2 \\ 4 & 1 & -1 & -1 \\ 1 & 2 & 3 & 0 \end{bmatrix}.$$

Subtract multiples of row 1 from rows 2, 3 and 4 to obtain the matrix

$$\begin{bmatrix} 1 & -1 & 2 & 3 \\ 0 & 4 & -4 & -4 \\ 0 & 5 & -9 & -13 \\ 0 & 3 & 1 & -3 \end{bmatrix}.$$

Now subtract 5/4 times row 2 from row 3, and 3/4 times row 2 from row 4. This yields the matrix

$$B = \left[\begin{array}{cc|cc} 1 & -1 & 2 & 3 \\ 0 & 4 & -4 & -4 \\ \hline 0 & 0 & -4 & -8 \\ 0 & 0 & 4 & 0 \end{array}\right]$$

with $\det B = \det A$ (by Theorem 3.4). Since B is in block triangular form we have

$$\det A = \det B = \begin{vmatrix} 1 & -1 \\ 0 & 4 \end{vmatrix} \begin{vmatrix} -4 & -8 \\ 4 & 0 \end{vmatrix} = 4(32) = 128.$$

Exercises

1. Prove the multiplication formula given in the text (just prior to Theorem 5.10) for block matrices.

2. Suppose $A \in M_n(\mathcal{F})$, $D \in M_m(\mathcal{F})$, $U \in M_{n\times m}(\mathcal{F})$ and $V \in M_{m\times n}(\mathcal{F})$,

and consider the $(n+m) \times (n+m)$ matrix

$$M = \begin{bmatrix} A & U \\ V & D \end{bmatrix}.$$

If A^{-1} exists, show that

$$\begin{bmatrix} A^{-1} & 0 \\ -VA^{-1} & I_m \end{bmatrix} \begin{bmatrix} A & U \\ V & D \end{bmatrix} = \begin{bmatrix} I_n & A^{-1}U \\ 0 & -VA^{-1}U + D \end{bmatrix}$$

and hence that

$$\begin{vmatrix} A & U \\ V & D \end{vmatrix} = (\det A)\det(D - VA^{-1}U).$$

3. Let A be a block triangular matrix of the form

$$A = \begin{bmatrix} B & C \\ 0 & D \end{bmatrix}$$

where B and D are square matrices. Prove that $\det A = (\det B)(\det D)$ by using elementary row operations on A to create a block triangular matrix

$$\tilde{A} = \begin{bmatrix} \tilde{B} & \tilde{C} \\ 0 & \tilde{D} \end{bmatrix}$$

where $\tilde{B}$ and $\tilde{D}$ are upper-triangular.

4. Show that
$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^T = \begin{bmatrix} A^T & C^T \\ B^T & D^T \end{bmatrix}.$$

5.5 Invariant Subspaces

Suppose $T \in L(V)$ and let W be a subspace of V. Then W is said to be **invariant under** T (or simply T**-invariant**) if $T(w) \in W$ for every $w \in W$. For example, if $V = \mathbb{R}^3$ then the xy-plane is invariant under the linear transformation that rotates every vector in $\mathbb{R}^3$ about the z-axis. As another example, note that if $v \in V$ is an eigenvector of T, then $T(v) = \lambda v$ for some $\lambda \in \mathcal{F}$, and hence v generates a one-dimensional subspace of V that is invariant under T (this is *not* necessarily the same as the eigenspace of λ).

Another way to describe the invariance of W under T is to say that $T(W) \subset W$. Then clearly $T^2(W) = T(T(W)) \subset W$, and in general $T^n(W) \subset W$ for every $n \in \mathbb{Z}^+$. Since W is a subspace of V, this means $f(T)(W) \subset W$ for any $f(x) \in \mathcal{F}[x]$. In other words, if W is invariant under T, then W is also invariant under any polynomial in T (over the same field as W).

If $W \subset V$ is T-invariant, we may focus our attention on the effect of T on W alone. To do this, we define the **restriction** of T to W as that operator

$T|W : W \to W$ defined by $(T|W)(w) = T(w)$ for every $w \in W$. In other words, the restriction is an operator $T|W$ that acts only on the subspace W, and gives the same result as the full operator T gives when it acts on those vectors in V that happen to be in W. We will frequently write T_W instead of $T|W$.

Now suppose $T \in L(V)$ and let $W \subset V$ be a T-invariant subspace. Furthermore let $\{v_1, \dots, v_n\}$ be a basis for V, where the first $m < n$ vectors form a basis for W. If $A = (a_{ij})$ is the matrix representation of T relative to this basis for V, then a little thought should convince you that A must be of the block matrix form

$$A = \begin{bmatrix} B & C \\ 0 & D \end{bmatrix}$$

where $a_{ij} = 0$ for $j \leq m$ and $i > m$. This is because $T(w) \in W$ and any $w \in W$ has components $(w_1, \dots, w_m, 0, \dots, 0)$ relative to the above basis for V. It should also be reasonably clear that B is just the matrix representation of T_W. The formal proof of this fact is given in our next theorem.

Theorem 5.11. *Let W be a subspace of V and suppose $T \in L(V)$. Then W is T-invariant if and only if T can be represented in the block matrix form*

$$A = \begin{bmatrix} B & C \\ 0 & D \end{bmatrix}$$

where B is a matrix representation of T_W.

Proof. First suppose that W is T-invariant. Choose a basis $\{v_1, \dots, v_m\}$ for W, and extend this to a basis $\{v_1, \dots, v_m, v_{m+1}, \dots, v_n\}$ for V (see Theorem 1.10). Then, since $T(v_i) \in W$ for each $i = 1, \dots, m$ there exist scalars b_{ij} such that

$$T_W(v_i) = T(v_i) = v_1 b_{1i} + \cdots + v_m b_{mi}$$

for each $i = 1, \dots, m$. In addition, since $T(v_i) \in V$ for each $i = m+1, \dots, n$ there also exist scalars c_{ij} and d_{ij} such that

$$T(v_i) = v_1 c_{1i} + \cdots + v_m c_{mi} + v_{m+1} d_{m+1,i} + \cdots + v_n d_{ni}$$

for each $i = m+1, \dots, n$.

Because T takes the ith basis vector into the ith column of the matrix representation of T (Theorem 4.9), we see that this representation is given by an $n \times n$ matrix A of the form

$$A = \begin{bmatrix} b_{11} & \cdots & b_{1m} & c_{1\,m+1} & \cdots & c_{1n} \\ b_{21} & \cdots & b_{2m} & c_{2\,m+1} & \cdots & c_{2n} \\ \vdots & & \vdots & \vdots & & \vdots \\ b_{m1} & \cdots & b_{mm} & c_{m\,m+1} & \cdots & c_{mn} \\ 0 & & 0 & d_{m+1\,m+1} & \cdots & d_{m+1\,n} \\ \vdots & & \vdots & \vdots & & \vdots \\ 0 & \cdots & 0 & d_{n\,m+1} & \cdots & d_{nn} \end{bmatrix}$$

or, in block matrix form as

$$A = \begin{bmatrix} B & C \\ 0 & D \end{bmatrix}$$

where B is an $m \times m$ matrix that represents T_W, C is an $m \times (n-m)$ matrix, and D is an $(n-m) \times (n-m)$ matrix.

Conversely, if A has the stated form and $\{v_1, \dots, v_n\}$ is a basis for V, then the subspace W of V defined by vectors of the form

$$w = \sum_{i=1}^{m} \alpha_i v_i$$

where each $\alpha_i \in \mathcal{F}$ will be invariant under T. Indeed, for each $i = 1, \dots, m$ we have

$$T(v_i) = \sum_{j=1}^{n} v_j a_{ji} = v_1 b_{1i} + \cdots + v_m b_{mi} \in W$$

and hence $T(w) = \sum_{i=1}^{m} \alpha_i T(v_i) \in W$. ∎

Corollary. *Suppose $T \in L(V)$ and W is a T-invariant subspace of V. Then the characteristic polynomial of T_W divides the characteristic polynomial of T.*

Proof. This is Exercise 5.5.2. ∎

Recall from Theorem 1.18 that the orthogonal complement $W^\perp$ of a set $W \subset V$ is a subspace of V. If W is a subspace of V and both W and $W^\perp$ are T-invariant then, since $V = W \oplus W^\perp$ (Theorem 1.22), a little more thought should convince you that the matrix representation of T will now be of the block diagonal form

$$A = \begin{bmatrix} B & 0 \\ 0 & D \end{bmatrix}.$$

We now proceed to discuss a variation of Theorem 5.11 in which we take into account the case where V can be decomposed into a direct sum of subspaces.

Let us assume that $V = W_1 \oplus \cdots \oplus W_r$ where each W_i is a T-invariant subspace of V. Then we define the restriction of T to W_i to be the operator $T_i = T_{W_i} = T|W_i$. In other words, $T_i(w_i) = T(w_i) \in W_i$ for any $w_i \in W_i$. Given any $v \in V$ we have $v = v_1 + \cdots + v_r$ where $v_i \in W_i$ for each $i = 1, \dots, r$ and hence

$$T(v) = \sum_{i=1}^{r} T(v_i) = \sum_{i=1}^{r} T_i(v_i).$$

This shows that T is completely determined by the effect of each T_i on W_i. In this case we call T the **direct sum** of the T_i and we write

$$T = T_1 \oplus \cdots \oplus T_r.$$

We also say that T is **reducible** (or **decomposable**) into the operators T_i, and the spaces W_i are said to **reduce** T, or to form a T**-invariant direct sum decomposition** of V. In other words, T is reducible if there exists a basis for V such that $V = W_1 \oplus \cdots \oplus W_r$ and each W_i is T-invariant.

For each $i = 1, \dots, r$ we let $\mathcal{B}_i = \{w_{(i)1}, \dots, w_{(i)n_i}\}$ be a basis for W_i so that $\mathcal{B} = \bigcup_{i=1}^r \mathcal{B}_i$ is a basis for $V = W_1 \oplus \cdots \oplus W_r$ (Theorem 1.15). Using a somewhat cluttered notation, we let $A_i = (a^{(i)}{}_{kj})$ be the matrix representation of T_i with respect to the basis $\mathcal{B}_i$ (where k and j label the rows and columns respectively of the matrix A_i). Therefore we see that

$$T(w_{(i)j}) = T_i(w_{(i)j}) = \sum_{k=1}^{n_i} w_{(i)k} a^{(i)}{}_{kj}$$

where $i = 1, \dots, r$ and $j = 1, \dots, n_i$. If A is the matrix representation of T with respect to the basis $\mathcal{B} = \{w_{(1)1}, \dots, w_{(1)n_1}, \dots, w_{(r)1}, \dots, w_{(r)n_r}\}$ for V, then since the ith column of A is just the image of the ith basis vector under T, we see that A must be of the block diagonal form

$$\begin{bmatrix} A_1 & 0 & \cdots & 0 \\ 0 & A_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & A_r \end{bmatrix}.$$

If this is not immediately clear, then a minute's thought should help, keeping in mind that each A_i is an $n_i \times n_i$ matrix, and A is an $n \times n$ matrix where $n = \sum_{i=1}^r n_i$. It is also helpful to think of the elements of $\mathcal{B}$ as being numbered from 1 to n rather than by the confusing double subscripts (also refer to the proof of Theorem 5.11).

The matrix A is called the **direct sum** of the matrices $A_1, \dots, A_r$ and we write

$$A = A_1 \oplus \cdots \oplus A_r.$$

In this case we also say that the matrix A is **reducible**. Thus a representation $[T]$ of T is reducible if there exists a basis for V in which $[T]$ is block diagonal. (Some authors say that a representation is **reducible** if there exists a basis for V in which the matrix of T is triangular. In this case, if there exists a basis for V in which the matrix is block diagonal, then the representation is said to be **completely reducible**. We shall not follow this convention.) This discussion proves the following theorem.

Theorem 5.12. *Suppose $T \in L(V)$ and assume $V = W_1 \oplus \cdots \oplus W_r$ where each W_i is T-invariant. If A_i is the matrix representation of $T_i = T|W_i$, then the matrix representation of T is given by the matrix $A = A_1 \oplus \cdots \oplus A_r$.*

Corollary. *Suppose $T \in L(V)$ and $V = W_1 \oplus \cdots \oplus W_r$ where each W_i is T-invariant. If $\Delta_T(x)$ is the characteristic polynomial for T and $\Delta_i(x)$ is the characteristic polynomial for $T_i = T|W_i$, then $\Delta_T(x) = \Delta_1(x) \cdots \Delta_r(x)$.*

Proof. See Exercise 5.5.3. ∎

Example 5.8. Referring to Example 1.8, consider the space $V = \mathbb{R}^3$. We write $V = W_1 \oplus W_2$ where $W_1 = \mathbb{R}^2$ (the xy-plane) and $W_2 = \mathbb{R}^1$ (the z-axis). Note that W_1 has basis vectors $w_{(1)1} = (1, 0, 0)$ and $w_{(1)2} = (0, 1, 0)$, and W_2 has basis vector $w_{(2)1} = (0, 0, 1)$.

Now let $T \in L(V)$ be the linear operator that rotates any $v \in V$ counterclockwise by an angle θ about the z-axis. Then clearly both W_1 and W_2 are T invariant. Letting $\{e_i\}$ be the standard basis for $\mathbb{R}^3$, we have $T_i = T|W_i$ and consequently (see Section 4.5),

$$\begin{aligned}
T_1(e_1) &= T(e_1) = (\cos\theta)e_1 + (\sin\theta)e_2 \\
T_1(e_2) &= T(e_2) = -(\sin\theta)e_1 + (\cos\theta)e_2 \\
T_2(e_3) &= T(e_3) = e_3
\end{aligned}$$

Thus $V = W_1 \oplus W_2$ is a T-invariant direct sum decomposition of V, and T is the direct sum of T_1 and T_2. It should be clear that the matrix representation of T is given by

$$\left[\begin{array}{rr|r} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ \hline 0 & 0 & 1 \end{array}\right]$$

which is just the direct sum of the matrix representations of T_1 and T_2.

Exercises

1. Suppose $V = W_1 \oplus W_2$ and let $T_1 : W_1 \to V$ and $T_2 : W_2 \to V$ be linear. Show that $T = T_1 \oplus T_2$ is linear.
2. Prove the corollary to Theorem 5.11.
3. Prove the corollary to Theorem 5.12.
4. A **group** $(G, \diamond)$ is a nonempty set G together with a binary operation called **multiplication** and denoted by $\diamond$ that obeys the following axioms:

   (G1) $a, b \in G$ implies $a \diamond b \in G$ (closure);
   (G2) $a, b, c \in G$ implies $(a \diamond b) \diamond c = a \diamond (b \diamond c)$ (associativity);
   (G3) There exists $e \in G$ such that $a \diamond e = e \diamond a = a$ for all $a \in G$ (identity);

(G4) For each $a \in G$, there exists $a^{-1} \in G$ such that $a \diamond a^{-1} = a^{-1} \diamond a = e$ (inverse).

(As a side remark, a group is said to be **abelian** if it also has the property that

(G5) $a \diamond b = b \diamond a$ for all $a, b \in G$ (commutativity).

In the case of abelian groups, the group multiplication operation is frequently denoted by $+$ and called **addition**. Also, there is no standard notation for the group multiplication symbol, and our choice of $\diamond$ is completely arbitrary.)

If the number of elements in G is finite, then G is said to be a **finite** group.

We will simplify the notation by leaving out the group multiplication symbol and assuming that it is understood for the particular group under discussion.

Let V be a finite-dimensional inner product space over $\mathbb{C}$, and let G be a finite group. If for each $g \in G$ there is a linear operator $U(g) \in L(V)$ such that

$$U(g_1)U(g_2) = U(g_1 g_2)$$

then the collection $U(G) = \{U(g)\}$ is said to form a **representation** of G. If W is a subspace of V with the property that $U(g)(W) \subset W$ for *all* $g \in G$, then we say W is $U(G)$-**invariant** (or simply **invariant**). Furthermore, we say that the representation $U(G)$ is **irreducible** if there is no nontrivial $U(G)$-invariant subspace (i.e., the only invariant subspaces are $\{0\}$ and V itself).

(a) Prove **Schur's lemma 1**: Let $U(G)$ be an irreducible representation of G on V. If $A \in L(V)$ is such that $AU(g) = U(g)A$ for all $g \in G$, then $A = \lambda 1$ where $\lambda \in \mathbb{C}$. [*Hint*: Let λ be an eigenvalue of A with corresponding eigenspace V_λ. Show that V_λ is $U(G)$-invariant.]

(b) If $S \in L(V)$ is nonsingular, show that $U'(G) = S^{-1}U(G)S$ is also a representation of G on V. (Two representations of G related by such a similarity transformation are said to be **equivalent**.)

(c) Prove **Schur's lemma 2**: Let $U(G)$ and $U'(G)$ be two irreducible representations of G on V and V' respectively, and suppose $A \in L(V', V)$ is such that $AU'(g) = U(g)A$ for all $g \in G$. Then either $A = 0$, or else A is an isomorphism of V' onto V so that A^{-1} exists and $U(G)$ is equivalent to $U'(G)$. [*Hint*: Show that $\operatorname{Im} A$ is invariant under $U(G)$, and that $\operatorname{Ker} A$ is invariant under $U'(G)$.]

5. Relative to the standard basis for $\mathbb{R}^2$, let $T \in L(\mathbb{R}^2)$ be represented by

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

(a) Prove that the only T-invariant subspaces of $\mathbb{R}^2$ are $\{0\}$ and $\mathbb{R}^2$ itself.

(b) Suppose $U \in L(\mathbb{C}^2)$ is also represented by A. Show that there exist one-dimensional U-invariant subspaces.

6. Find all invariant subspaces over $\mathbb{R}$ of the operator represented by

$$A = \begin{bmatrix} 2 & -5 \\ 1 & -2 \end{bmatrix}.$$

5.6 More on Diagonalization

If an operator $T \in L(V)$ is diagonalizable, then in a (suitably numbered) basis of eigenvectors, its matrix representation A will take the form

$$A = \begin{bmatrix} \lambda_1 I_{m_1} & 0 & 0 & \cdots & 0 & 0 \\ 0 & \lambda_2 I_{m_2} & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & \lambda_r I_{m_r} \end{bmatrix}$$

where each λ_i is repeated m_i times and I_{m_i} is the $m_i \times m_i$ identity matrix. Note that $m_1 + \cdots + m_r = \dim V = n$. Thus the characteristic polynomial for T has the form

$$\Delta_T(x) = \det(xI - A) = (x - \lambda_1)^{m_1} \cdots (x - \lambda_r)^{m_r}$$

which is a product of (possibly repeated) *linear* factors. However, we stress that just because the characteristic polynomial factors into a product of linear terms does not mean that the operator is diagonalizable. We now investigate the conditions that determine just when an operator will be diagonalizable.

Let us assume that T is diagonalizable, and hence that the characteristic polynomial factors into linear terms (so that $\sum_{i=1}^r m_i = \dim V = n$). For each distinct eigenvalue λ_i, we have seen that the corresponding eigenspace V_{λ_i} is just $\operatorname{Ker}(T - \lambda_i 1)$. Relative to a basis of eigenvectors, the matrix $[T - \lambda_i 1]$ is diagonal with precisely m_i zeros along its main diagonal (just look at the matrix A shown above and subtract off $\lambda_i I$). From Theorem 4.12 we know that the rank of a linear transformation is the same as the rank of its matrix representation, and hence $\operatorname{rank}(T - \lambda_i 1)$ is just the number of remaining nonzero rows in $[T - \lambda_i 1]$ which is $\dim V - m_i$ (see Theorem 2.6). But from the rank theorem (Theorem 4.6) we then see that

$$\begin{aligned} \dim V_{\lambda_i} &= \dim \operatorname{Ker}(T - \lambda_i 1) = \dim V - \operatorname{rank}(T - \lambda_i 1) = n - (n - m_i) \\ &= m_i. \end{aligned}$$

In other words, if T is diagonalizable, then the dimension of each eigenspace V_{λ_i} is just the multiplicity of the eigenvalue λ_i. Let us clarify this in terms of some common terminology. In so doing, we will also repeat this conclusion from a slightly different viewpoint.

Given a linear operator $T \in L(V)$, what we have called the multiplicity of an eigenvalue λ is the largest positive integer m such that $(x - \lambda)^m$ divides the characteristic polynomial $\Delta_T(x)$. This is properly called the **algebraic multiplicity** of λ, in contrast to the **geometric multiplicity** which is the number of linearly independent eigenvectors belonging to that eigenvalue. In other words, the geometric multiplicity of λ is the dimension of V_λ. In general, we will use the word "multiplicity" to mean the algebraic multiplicity. The set of all eigenvalues of a linear operator $T \in L(V)$ is called the **spectrum** of T. If some eigenvalue in the spectrum of T is of algebraic multiplicity > 1, then the spectrum is said to be **degenerate**.

What we have just shown then is that if T is diagonalizable, then the algebraic and geometric multiplicities of each eigenvalue must be the same. In fact, we can arrive at this conclusion from a slightly different viewpoint that also illustrates much of what we have already covered and, in addition, proves the converse. First we prove a preliminary result.

If $T \in L(V)$ has an eigenvalue λ of algebraic multiplicity m, then it is not hard for us to show that the dimension of the eigenspace V_λ must be less than or equal to m. Note that since every element of V_λ is an eigenvector of T with eigenvalue λ, the space V_λ must be a T-invariant subspace of V. Furthermore, every basis for V_λ will obviously consist of eigenvectors corresponding to λ.

Theorem 5.13. *Let $T \in L(V)$ have eigenvalue λ. Then the geometric multiplicity of λ is always less than or equal to its algebraic multiplicity. In other words, if λ has algebraic multiplicity m, then $\dim V_\lambda \le m$.*

Proof. Suppose $\dim V_\lambda = r$ and let $\{v_1, \dots, v_r\}$ be a basis for V_λ. By Theorem 1.10, we extend this to a basis $\{v_1, \dots, v_n\}$ for V. Relative to this basis, T must have the matrix representation (see Theorem 5.11)

$$\begin{bmatrix} \lambda I_r & C \\ 0 & D \end{bmatrix}.$$

Applying Theorem 5.10 and the fact that the determinant of a diagonal matrix is just the product of its (diagonal) elements, we see that the characteristic polynomial $\Delta_T(x)$ of T is given by

$$\begin{aligned} \Delta_T(x) &= \begin{vmatrix} (x-\lambda)I_r & -C \\ 0 & xI_{n-r} - D \end{vmatrix} \\ &= \det[(x-\lambda)I_r] \det(xI_{n-r} - D) \\ &= (x-\lambda)^r \det(xI_{n-r} - D) \end{aligned}$$

which shows that $(x-\lambda)^r$ divides $\Delta_T(x)$. Since by definition m is the largest positive integer such that $(x-\lambda)^m \mid \Delta_T(x)$, it follows that $r \le m$. ∎

Note that a special case of this theorem arises when an eigenvalue is of (algebraic) multiplicity 1. In this case, it then follows that the geometric and algebraic multiplicities are necessarily equal. We now proceed to show just when this will be true in general. Recall that any polynomial over an algebraically closed field will factor into linear terms (Theorem 5.4).

Theorem 5.14. *Assume that $T \in L(V)$ has a characteristic polynomial that factors into (not necessarily distinct) linear terms. Let T have distinct eigenvalues $\lambda_1, \ldots, \lambda_r$ with (algebraic) multiplicities $m_1, \ldots, m_r$ respectively, and let $\dim V_{\lambda_i} = d_i$. Then T is diagonalizable if and only if $m_i = d_i$ for each $i = 1, \ldots, r$.*

Proof. Let $\dim V = n$. We note that since the characteristic polynomial of T is of degree n and factors into linear terms, it follows that $m_1 + \cdots + m_r = n$. We first assume that T is diagonalizable. By definition, this means that V has a basis consisting of n linearly independent eigenvectors of T. Since each of these basis eigenvectors must belong to at least one of the eigenspaces V_{λ_i}, it follows that $V = V_{\lambda_1} + \cdots + V_{\lambda_r}$ and consequently $n \le d_1 + \cdots + d_r$. From Theorem 5.13 we know that $d_i \le m_i$ for each $i = 1, \ldots, r$ and hence

$$n \le d_1 + \cdots + d_r \le m_1 + \cdots + m_r = n$$

which implies $d_1 + \cdots + d_r = m_1 + \cdots + m_r$ or

$$(m_1 - d_1) + \cdots + (m_r - d_r) = 0.$$

But each term in this equation is nonnegative (by Theorem 5.13), and hence we must have $m_i = d_i$ for each i.

Conversely, suppose $d_i = m_i$ for each $i = 1, \ldots, r$. For each i, we know that any basis for V_{λ_i} consists of linearly independent eigenvectors corresponding to the eigenvalue λ_i, while by Theorem 5.7, we know that eigenvectors corresponding to distinct eigenvalues are linearly independent. Therefore the union $\mathcal{B}$ of the bases of $\{V_{\lambda_i}\}$ forms a linearly independent set of $d_1+\cdots+d_r = m_1+\cdots+m_r$ vectors. But $m_1 + \cdots + m_r = n = \dim V$, and hence $\mathcal{B}$ forms a basis for V. Since this shows that V has a basis of eigenvectors of T, it follows by definition that T must be diagonalizable. ∎

Corollary 1. *An operator $T \in L(V)$ is diagonalizable if and only if*

$$V = W_1 \oplus \cdots \oplus W_r$$

where $W_1, \ldots, W_r$ are the eigenspaces corresponding to the distinct eigenvalues of T.

Proof. This is Exercise 5.6.1. ∎

Using Theorem 4.6, we see that the geometric multiplicity of an eigenvalue λ is given by

$$\dim V_\lambda = \dim(\operatorname{Ker}(T - \lambda 1)) = \operatorname{nul}(T - \lambda 1) = \dim V - \operatorname{rank}(T - \lambda 1).$$

This observation together with Theorem 5.14 proves the next corollary.

Corollary 2. *An operator $T \in L(V)$ whose characteristic polynomial factors into linear terms is diagonalizable if and only if the algebraic multiplicity of λ is equal to* $\dim V - \operatorname{rank}(T - \lambda 1)$ *for each eigenvalue λ.*

Example 5.9. Consider the operator $T \in L(\mathbb{R}^3)$ defined by

$$T(x, y, z) = (9x + y, 9y, 7z).$$

Relative to the standard basis for $\mathbb{R}^3$, the matrix representation of T is given by

$$A = \begin{bmatrix} 9 & 1 & 0 \\ 0 & 9 & 0 \\ 0 & 0 & 7 \end{bmatrix}$$

and hence the characteristic polynomial is

$$\Delta_A(x) = \det(A - \lambda I) = (9 - \lambda)^2(7 - \lambda)$$

which is a product of linear factors. However,

$$A - 9I = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -2 \end{bmatrix}$$

which clearly has rank equal to 2, and hence $\operatorname{nul}(A - 9I) = 3 - 2 = 1$ which is not the same as the algebraic multiplicity of $\lambda = 9$ (which is 2). Thus T is not diagonalizable.

Example 5.10. Consider the operator on $\mathbb{R}^3$ defined by the following matrix:

$$A = \begin{bmatrix} 5 & -6 & -6 \\ -1 & 4 & 2 \\ 3 & -6 & -4 \end{bmatrix}.$$

In order to avoid factoring a cubic polynomial, we compute the characteristic polynomial $\Delta_A(x) = \det(xI - A)$ by applying Theorem 3.4 as follows (the reader should be able to see exactly what elementary row operations were performed in each step).

$$\begin{aligned}
\begin{vmatrix} x-5 & 6 & 6 \\ 1 & x-4 & -2 \\ -3 & 6 & x+4 \end{vmatrix} &= \begin{vmatrix} x-2 & 0 & -x+2 \\ 1 & x-4 & -2 \\ -3 & 6 & x+4 \end{vmatrix} \\
&= (x-2)\begin{vmatrix} 1 & 0 & -1 \\ 1 & x-4 & -2 \\ -3 & 6 & x+4 \end{vmatrix} \\
&= (x-2)\begin{vmatrix} 1 & 0 & -1 \\ 0 & x-4 & -1 \\ 0 & 6 & x+1 \end{vmatrix} \\
&= (x-2)\begin{vmatrix} x-4 & -1 \\ 6 & x+1 \end{vmatrix} \\
&= (x-2)^2(x-1).
\end{aligned}$$

We now see that A has eigenvalue $\lambda_1 = 1$ with (algebraic) multiplicity 1, and eigenvalue $\lambda_2 = 2$ with (algebraic) multiplicity 2. From Theorem 5.13 we know that the algebraic and geometric multiplicities of λ_1 are necessarily the same and equal to 1, so we need only consider λ_2. Observing that

$$A - 2I = \begin{bmatrix} 3 & -6 & -6 \\ -1 & 2 & 2 \\ 3 & -6 & -6 \end{bmatrix}$$

it is obvious that $\operatorname{rank}(A-2I) = 1$, and hence $\operatorname{nul}(A-2I) = 3-1 = 2$. This shows that A is indeed diagonalizable.

Let us now construct bases for the eigenspaces $W_i = V_{\lambda_i}$. This means we seek vectors $v = (x,y,z) \in \mathbb{R}^3$ such that $(A - \lambda_i I)v = 0$. This is easily solved by the usual row reduction techniques as follows. For $\lambda_1 = 1$ we have

$$\begin{aligned}
A - I = \begin{bmatrix} 4 & -6 & -6 \\ -1 & 3 & 2 \\ 3 & -6 & -5 \end{bmatrix} &\to \begin{bmatrix} 1 & 0 & -1 \\ -1 & 3 & 2 \\ 3 & -6 & -5 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & -1 \\ 0 & 3 & 1 \\ 0 & -6 & -2 \end{bmatrix} \\
&\to \begin{bmatrix} 1 & 0 & -1 \\ 0 & 3 & 1 \\ 0 & 0 & 0 \end{bmatrix}
\end{aligned}$$

which has the solutions $x = z$ and $y = -z/3 = -x/3$. Therefore W_1 is spanned by the single eigenvector $v_1 = (3, -1, 3)$. As to $\lambda_2 = 2$, we proceed in a similar manner to obtain

$$A - 2I = \begin{bmatrix} 3 & -6 & -6 \\ -1 & 2 & 2 \\ 3 & -6 & -6 \end{bmatrix} \to \begin{bmatrix} 1 & -2 & -2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

which implies that any vector (x, y, z) with $x = 2y + 2z$ will work. For example, we can let $x = 0$ and $y = 1$ to obtain $z = -1$, and hence one basis vector for W_2 is given by $v_2 = (0, 1, -1)$. If we let $x = 1$ and $y = 0$, then we have $z = 1/2$ so that another independent basis vector for W_2 is given by $v_3 = (2, 0, 1)$.

In terms of these eigenvectors, the transformation matrix P that diagonalizes A is given by

$$P = \begin{bmatrix} 3 & 0 & 2 \\ -1 & 1 & 0 \\ 3 & -1 & 1 \end{bmatrix}$$

and we leave it to the reader to verify that $AP = PD$ (i.e., $P^{-1}AP = D$) where D is the diagonal matrix with diagonal elements $d_{11} = 1$ and $d_{22} = d_{33} = 2$.

Exercises

1. Prove Corollary 1 of Theorem 5.14.

2. Show that two similar matrices A and B have the same eigenvalues, and these eigenvalues have the same geometric multiplicities.

3. Let $\lambda_1, \dots, \lambda_r \in \mathcal{F}$ be distinct, and let $D \in M_n(\mathcal{F})$ be diagonal with a characteristic polynomial of the form

$$\Delta_D(x) = (x - \lambda_1)^{d_1} \cdots (x - \lambda_r)^{d_r}.$$

   Let V be the space of all $n \times n$ matrices B that commute with D, i.e., the set of all B such that $BD = DB$. Prove $\dim V = {d_1}^2 + \cdots + {d_r}^2$.

4. Relative to the standard basis, let $T \in L(\mathbb{R}^4)$ be represented by

$$A = \begin{bmatrix} 0 & 0 & 0 & 0 \\ a & 0 & 0 & 0 \\ 0 & b & 0 & 0 \\ 0 & 0 & c & 0 \end{bmatrix}.$$

   Find conditions on a, b and c such that T is diagonalizable.

5. Determine whether or not each of the following matrices is diagonalizable. If it is, find a nonsingular matrix P and a diagonal matrix D such that $P^{-1}AP = D$.

   (a) $\begin{bmatrix} 1 & 1 & 0 \\ 0 & 2 & 2 \\ 0 & 0 & 3 \end{bmatrix}$ (b) $\begin{bmatrix} 7 & -4 & 0 \\ 8 & -5 & 0 \\ 6 & -6 & 3 \end{bmatrix}$ (c) $\begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & -1 \\ 0 & 1 & 1 \end{bmatrix}$

   (d) $\begin{bmatrix} -1 & -3 & -9 \\ 0 & 5 & 18 \\ 0 & -2 & -7 \end{bmatrix}$ (e) $\begin{bmatrix} 3 & -1 & -2 \\ 2 & 0 & -2 \\ 2 & -1 & -1 \end{bmatrix}$ (f) $\begin{bmatrix} -1 & 1 & 0 \\ 0 & 5 & 0 \\ 4 & -2 & 5 \end{bmatrix}$

(g) $\begin{bmatrix} 3 & 1 & 1 \\ 2 & 4 & 2 \\ -1 & -1 & 1 \end{bmatrix}$

6. Determine whether or not each of the following operators $T \in L(\mathbb{R}^3)$ is diagonalizable. If it is, find an eigenvector basis for $\mathbb{R}^3$ such that $[T]$ is diagonal.
   (a) $T(x, y, z) = (-y, x, 3z)$.
   (b) $T(x, y, z) = (8x + 2y - 2z, 3x + 3y - z, 24x + 8y - 6z)$.
   (c) $T(x, y, z) = (4x + z, 2x + 3y + 2z, x + 4z)$.
   (d) $T(x, y, z) = (-2y - 3z, x + 3y + 3z, z)$.

7. Suppose a matrix A is diagonalizable. Prove that A^m is diagonalizable for any positive integer m.

8. Summarize several of our results by proving the following theorem:

   Let V be finite-dimensional, suppose $T \in L(V)$ has distinct eigenvalues $\lambda_1, \ldots, \lambda_r$, and let $W_i = \operatorname{Ker}(T - \lambda_i 1)$. Then the following are equivalent:
   (a) T is diagonalizable.
   (b) $\Delta_T(x) = (x - \lambda_1)^{m_1} \cdots (x - \lambda_r)^{m_r}$ and W_i is of dimension m_i for each $i = 1, \ldots, r$.
   (c) $\dim W_1 + \cdots + \dim W_r = \dim V$.

9. Let V_3 be the space of real polynomials of degree at most 3, and let f' and f'' denote the first and second derivatives of $f \in V$. Define $T \in L(V_3)$ by $T(f) = f' + f''$. Decide whether or not T is diagonalizable and, if it is, find a basis for V_3 such that $[T]$ is diagonal.

10. (a) Let V_2 be the space of real polynomials of degree at most 2, and define $T \in L(V_2)$ by $T(ax^2 + bx + c) = cx^2 + bx + a$. Decide whether or not T is diagonalizable and, if it is, find a basis for V_2 such that $[T]$ is diagonal.
    (b) Repeat part (a) with $T = (x + 1)(d/dx)$. (See Exercise 5.3.13.)

5.7 Diagonalizing Normal Matrices

This section presents a simple and direct method of treating two important results: the triangular form for complex matrices and the diagonalization of normal matrices. To begin with, suppose we have a matrix $A \in M_n(\mathbb{C})$. We define the **adjoint** (or **Hermitian adjoint**) of A to be the matrix $A^\dagger = A^{*T}$. In other words, the adjoint of A is its complex conjugate transpose. From Theorem 2.15(iv), it is easy to see that

$$(AB)^\dagger = B^\dagger A^\dagger.$$

If it so happens that $A^\dagger = A$, then A is said to be a **Hermitian** matrix.

If a matrix $U \in M_n(\mathbb{C})$ has the property that $U^\dagger = U^{-1}$, then we say U is **unitary**. Thus a matrix U is unitary if $UU^\dagger = U^\dagger U = I$. (Note that by Theorem 2.20, it is only necessary to require either $UU^\dagger = I$ or $U^\dagger U = I$. However, the full definition is necessary in the case of infinite-dimensional spaces.) We also see that the product of two unitary matrices U and V is unitary since $(UV)^\dagger UV = V^\dagger U^\dagger UV = V^\dagger IV = V^\dagger V = I$. If a matrix $N \in M_n(\mathbb{C})$ has the property that it commutes with its adjoint, i.e., $NN^\dagger = N^\dagger N$, then N is said to be a **normal** matrix. Note that Hermitian and unitary matrices are automatically normal.

Example 5.11. Consider the matrix $A \in M_2(\mathbb{C})$ given by

$$A = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -1 \\ i & i \end{bmatrix}.$$

Then the adjoint of A is given by

$$A^\dagger = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -i \\ -1 & -i \end{bmatrix}$$

and we leave it to the reader to verify that $AA^\dagger = A^\dagger A = I$, and hence show that A is unitary.

A convenient property of the adjoint is this. If $A \in M_n(\mathbb{C})$ and $x, y \in \mathbb{C}^n$, then $Ax \in \mathbb{C}^n$ also, so we may use the standard inner product on $\mathbb{C}^n$ (see Example 1.9) to write (using $A^\dagger = A^{*T}$)

$$\begin{aligned} \langle Ax, y\rangle &= \sum_{i=1}^n (Ax)_i^* y_i = \sum_{i=1}^n a_{ij}^* x_j^* y_i = \sum_{i=1}^n x_j^* a_{ji}^\dagger y_i \\ &= \langle x, A^\dagger y\rangle. \end{aligned} \tag{5.2}$$

In the particular case of a unitary matrix, we see that

$$\langle Ux, Uy\rangle = \langle x, U^\dagger U y\rangle = \langle x, y\rangle$$

so that unitary transformations also preserve the angle between two vectors (and hence maintains orthogonality as well). Choosing $y = x$ we also see that

$$\|Ux\|^2 = \langle Ux, Ux\rangle = \langle x, U^\dagger Ux\rangle = \langle x, Ix\rangle = \langle x, x\rangle = \|x\|^2$$

so that unitary transformations preserve lengths of vectors, i.e., they are really just rotations in $\mathbb{C}^n$.

It is well worth pointing out that in the case of a real matrix $A \in M_n(\mathcal{F})$, instead of the adjoint $A^\dagger$ we have the transpose A^T and equation (5.2) becomes

$$\langle Ax, y\rangle = \langle x, A^T y\rangle$$

or equivalently

$$\langle A^T x, y\rangle = \langle x, Ay\rangle. \tag{5.3}$$

We will use this below when we prove that a real symmetric matrix has all real eigenvalues.

Note that since $U \in M_n(\mathbb{C})$, the rows U_i and columns U^i of U are just vectors in $\mathbb{C}^n$. This means we can take their inner product relative to the standard inner product on $\mathbb{C}^n$. Writing out the relation $UU^\dagger = I$ in terms of components, we have

$$(UU^\dagger)_{ij} = \sum_{k=1}^n u_{ik}u^\dagger_{kj} = \sum_{k=1}^n u_{ik}u^*_{jk} = \sum_{k=1}^n u^*_{jk}u_{ik} = \langle U_j, U_i\rangle = \delta_{ij}$$

and from $U^\dagger U = I$ we see that

$$(U^\dagger U)_{ij} = \sum_{k=1}^n u^\dagger_{ik}u_{kj} = \sum_{k=1}^n u^*_{ki}u_{kj} = \langle U^i, U^j\rangle = \delta_{ij}.$$

In other words, a matrix is unitary if and only if its rows (or columns) each form an orthonormal set. Note we have shown that if the rows (columns) of $U \in M_n(\mathbb{C})$ form an orthonormal set, then so do the columns (rows), and either of these is a sufficient condition for U to be unitary. For example, the reader can easily verify that the matrix A in Example 5.11 satisfies these conditions.

It is also worth pointing out that Hermitian and unitary matrices have important analogues over the real number system. If $A \in M_n(\mathbb{R})$ is Hermitian, then $A = A^\dagger = A^T$, and we say A is **symmetric**. If $U \in M_n(\mathbb{R})$ is unitary, then $U^{-1} = U^\dagger = U^T$, and we say U is **orthogonal**. Repeating the above calculations over $\mathbb{R}$, it is easy to see that a real matrix is orthogonal if and only if its rows (or columns) form an orthonormal set.

Let us summarize what we have shown so far in this section.

Theorem 5.15. *The following conditions on a matrix $U \in M_n(\mathbb{C})$ are equivalent:*
(i) U is unitary.
(ii) The rows U_i of U form an orthonormal set.
(iii) The columns U^i of U form an orthonormal set.

Note that the equivalence of (ii) and (iii) in this theorem means that the rows of U form an orthonormal set if and only if the columns of U form an orthonormal set. But the rows of U are just the columns of U^T, and hence U is unitary if and only if U^T is unitary.

Corollary. *The following conditions on a matrix $A \in M_n(\mathbb{R})$ are equivalent:*
(i) A is orthogonal.

(ii) The rows A_i of A form an orthonormal set.
(iii) The columns A^i of A form an orthonormal set.

Our next theorem details several useful properties of orthogonal and unitary matrices.

Theorem 5.16. *(i) If A is an orthogonal matrix, then* $\det A = \pm 1$.
(ii) If U is a unitary matrix, then $|\det U| = 1$*. Alternatively,* $\det U = e^{i\phi}$ *for some real number ϕ.*

Proof. (i) We have $AA^T = I$, and hence (from Theorems 3.7 and 3.1)

$$1 = \det I = \det(AA^T) = (\det A)(\det A^T) = (\det A)^2$$

so that $\det A = \pm 1$.
(ii) If $UU^\dagger = I$ then, as above, we have

$$\begin{aligned} 1 = \det I = \det(UU^\dagger) &= (\det U)(\det U^\dagger) = (\det U)(\det U^T)^* \\ &= (\det U)(\det U)^* = |\det U|^2 . \end{aligned}$$

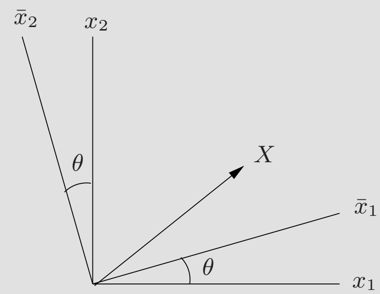
Since the absolute value is defined to be positive, this shows $|\det U| = 1$ and hence $\det U = e^{i\phi}$ for some real ϕ. ∎

Example 5.12. Let us take another look at rotations in $\mathbb{R}^2$ as shown, for example, in the figure below (see Section 4.5). Recall that if we have two bases $\{e_i\}$ and $\{\bar{e}_i\}$, then they are related by a transition matrix $A = (a_{ij})$ defined by $\bar{e}_i = \sum_j e_j a_{ji}$. In addition, if $X = \sum x^i e_i = \sum \bar{x}^i \bar{e}_i$, then $x^i = \sum_j a_{ij}\bar{x}^j$. If both $\{e_i\}$ and $\{\bar{e}_i\}$ are orthonormal bases, then

$$\langle e_i, \bar{e}_j \rangle = \left\langle e_i, \sum_k e_k a_{kj} \right\rangle = \sum_k a_{kj} \langle e_i, e_k \rangle = \sum_k a_{kj}\delta_{ik} = a_{ij}.$$

Using the usual dot product on $\mathbb{R}^2$ as our inner product (see Section 1.5, Lemma 1.3) and referring to the figure below, we see that the elements a_{ij} are given by (also see Section A.6 for the trigonometric identities)

$$\begin{aligned} a_{11} &= e_1 \cdot \bar{e}_1 = |e_1|\,|\bar{e}_1| \cos\theta = \cos\theta \\ a_{12} &= e_1 \cdot \bar{e}_2 = |e_1|\,|\bar{e}_2| \cos(\pi/2 + \theta) = -\sin\theta \\ a_{21} &= e_2 \cdot \bar{e}_1 = |e_2|\,|\bar{e}_1| \cos(\pi/2 - \theta) = \sin\theta \\ a_{22} &= e_2 \cdot \bar{e}_2 = |e_2|\,|\bar{e}_2| \cos\theta = \cos\theta \end{aligned}$$



Thus the matrix A is given by

$$(a_{ij}) = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}.$$

We leave it to the reader to compute directly and show $A^T A = AA^T = I$ and $\det A = +1$.

Example 5.13. Referring to Example 5.12, we can show that any (real) 2×2 orthogonal matrix with $\det A = +1$ has the form

$$(a_{ij}) = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

for some $\theta \in \mathbb{R}$. To see this, suppose A has the form

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

where $a, b, c, d \in \mathbb{R}$. Since A is orthogonal, its rows form an orthonormal set, and hence we have

$$a^2 + b^2 = 1, \quad c^2 + d^2 = 1, \quad ac + bd = 0, \quad ad - bc = 1$$

where the last equation follows from $\det A = 1$.

If $a = 0$, then the first of these equations yields $b = \pm 1$, the third then yields $d = 0$, and the last yields $-c = 1/b = \pm 1$ which is equivalent to $c = -b$. In other words, if $a = 0$, then A has either of the forms

$$\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \quad \text{or} \quad \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}.$$

The first of these is of the required form if we choose $\theta = -90^\circ = -\pi/2$, and the second is of the required form if we choose $\theta = +90^\circ = +\pi/2$.

Now suppose that $a \neq 0$. From the third equation we have $c = -bd/a$, and substituting this into the second equation, we find $(a^2 + b^2)d^2 = a^2$. Using the first equation, this becomes $a^2 = d^2$ or $a = \pm d$. If $a = -d$, then the third equation yields $b = c$, and hence the last equation yields $-a^2 - b^2 = 1$ which is impossible. Therefore $a = +d$, the third equation then yields $c = -b$, and we are left with

$$\begin{bmatrix} a & -c \\ c & a \end{bmatrix}$$

Since $\det A = a^2 + c^2 = 1$, there exists a real number θ such that $a = \cos\theta$ and $c = \sin\theta$ which gives us the desired form for A.

One of the most important and useful properties of matrices over $\mathbb{C}$ is that they can always be put into triangular form by an appropriate transformation. To show this, it will be helpful to recall from Section 2.5 that if A and B are two matrices for which the product AB is defined, then the ith row of AB is given by $(AB)_i = A_i B$ and the ith column of AB is given by $(AB)^i = AB^i$.

Theorem 5.17 (**Schur Canonical Form**). *If $A \in M_n(\mathbb{C})$, then there exists a unitary matrix $U \in M_n(\mathbb{C})$ such that $U^\dagger AU$ is upper-triangular. Furthermore, the diagonal entries of $U^\dagger AU$ are just the eigenvalues of A.*

Proof. The proof is by induction. If $n = 1$ there is nothing to prove, so we assume the theorem holds for any square matrix of size $n - 1 \geq 1$, and suppose A is of size n. Since we are dealing with the algebraically closed field $\mathbb{C}$, we know that A has n (not necessarily distinct) eigenvalues (see Section 5.3). Let λ be one of these eigenvalues, and denote the corresponding eigenvector by $\widetilde{U}^1$. By Theorem 1.10 we extend $\widetilde{U}^1$ to a basis for $\mathbb{C}^n$, and by the Gram-Schmidt process (Theorem 1.21) we assume this basis is orthonormal. From our discussion above, we see that this basis may be used as the columns of a unitary matrix $\widetilde{U}$ with $\widetilde{U}^1$ as its first column. We then see that

$$\begin{aligned}(\widetilde{U}^\dagger A\widetilde{U})^1 &= \widetilde{U}^\dagger(A\widetilde{U})^1 = \widetilde{U}^\dagger(A\widetilde{U}^1) = \widetilde{U}^\dagger(\lambda\widetilde{U}^1) = \lambda(\widetilde{U}^\dagger\widetilde{U}^1) \\ &= \lambda(\widetilde{U}^\dagger\widetilde{U})^1 = \lambda I^1\end{aligned}$$

and hence $\widetilde{U}^\dagger A\widetilde{U}$ has the form

$$\widetilde{U}^\dagger A\widetilde{U} = \begin{bmatrix} \lambda & * \cdots * \\ 0 & \\ \vdots & \boxed{\quad B \quad} \\ 0 & \end{bmatrix}$$

where $B \in M_{n-1}(\mathbb{C})$ and the *'s are (in general) nonzero scalars.

By our induction hypothesis, we may choose a unitary matrix $W \in M_{n-1}(\mathbb{C})$ such that $W^\dagger BW$ is upper-triangular. Let $V \in M_n(\mathbb{C})$ be a unitary matrix of the form

$$V = \begin{bmatrix} 1 & 0 \cdots 0 \\ 0 & \\ \vdots & \boxed{\quad W \quad} \\ 0 & \end{bmatrix}$$

and define the unitary matrix $U = \widetilde{U}V \in M_n(\mathbb{C})$. Then

$$U^\dagger AU = (\widetilde{U}V)^\dagger A(\widetilde{U}V) = V^\dagger(\widetilde{U}^\dagger A\widetilde{U})V$$

is upper-triangular since (in an obvious shorthand notation)

$$\begin{aligned} V^\dagger(\widetilde{U}^\dagger A\widetilde{U})V &= \begin{bmatrix} 1 & 0 \\ 0 & W^\dagger \end{bmatrix} \begin{bmatrix} \lambda & * \\ 0 & B \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & W \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & W^\dagger \end{bmatrix} \begin{bmatrix} \lambda & * \\ 0 & BW \end{bmatrix} \\ &= \begin{bmatrix} \lambda & * \\ 0 & W^\dagger BW \end{bmatrix} \end{aligned}$$

and $W^\dagger BW$ is upper-triangular by the induction hypothesis.

It is easy to see (using Theorem 3.5) that the roots of $\det(\lambda I - U^\dagger AU)$ are just the diagonal entries of $U^\dagger AU$ because $\lambda I - U^\dagger AU$ is of the upper triangular form

$$\begin{bmatrix} \lambda - (U^\dagger AU)_{11} & * & & * \\ 0 & \lambda - (U^\dagger AU)_{22} & & * \\ \vdots & \vdots & \ddots & * \\ 0 & 0 & & \lambda - (U^\dagger AU)_{nn} \end{bmatrix}$$

where the *'s are just some in general nonzero entries. But

$$\det(\lambda I - U^\dagger AU) = \det[U^\dagger(\lambda I - A)U] = \det(\lambda I - A)$$

so that A and $U^\dagger AU$ have the same eigenvalues. ∎

Corollary. *If $A \in M_n(\mathbb{R})$ has all its eigenvalues in $\mathbb{R}$, then the matrix U defined in Theorem 5.17 may be chosen to have all real entries.*

Proof. If $\lambda \in \mathbb{R}$ is an eigenvalue of A, then $A - \lambda I$ is a real matrix with determinant $\det(A - \lambda I) = 0$, and therefore the homogeneous system of equations $(A - \lambda I)X = 0$ has a real solution. Defining $\widetilde{U}^1 = X$, we may now proceed as in Theorem 5.17. The details are left to the reader (see Exercise 5.7.1). ∎

We say that two matrices A, $B \in M_n(\mathbb{C})$ are **unitarily similar** if there exists a unitary matrix U such that $B = U^\dagger AU = U^{-1}AU$. Since this defines an equivalence relation on the set of all matrices in $M_n(\mathbb{C})$, it is also common to say that A and B are **unitarily equivalent**.

We leave it to the reader to show that if A and B are unitarily similar and A is normal, then B is also normal (see Exercise 5.7.2). In particular, suppose U is unitary and N is such that $U^\dagger NU = D$ is diagonal. Since any diagonal matrix is automatically normal, it follows that N must be normal also. In other words, any matrix unitarily similar to a diagonal matrix is normal. We now show that the converse is also true, i.e., that any normal matrix is unitarily similar to a diagonal matrix. This extremely important result is the basis for many physical applications in both classical and quantum physics.

To see this, suppose N is normal, and let $U^\dagger NU = D$ be the Schur canonical form of N. Then D is both upper-triangular and normal (since it is unitarily similar to a normal matrix). We claim that the only such matrices are diagonal. For, consider the $(1,1)$ elements of $DD^\dagger$ and $D^\dagger D$. From what we showed above, we have

$$(DD^\dagger)_{11} = \langle D_1, D_1\rangle = |d_{11}|^2 + |d_{12}|^2 + \cdots + |d_{1n}|^2$$

and

$$(D^\dagger D)_{11} = \langle D^1, D^1\rangle = |d_{11}|^2 + |d_{21}|^2 + \cdots + |d_{n1}|^2 .$$

But D is upper-triangular so that $d_{21} = \cdots = d_{n1} = 0$. By normality we must have $(DD^\dagger)_{11} = (D^\dagger D)_{11}$, and therefore $d_{12} = \cdots = d_{1n} = 0$ also. In other words, with the possible exception of the $(1,1)$ entry, all entries in the first row and column of D must be zero. In the same manner, we see that

$$(DD^\dagger)_{22} = \langle D_2, D_2\rangle = |d_{21}|^2 + |d_{22}|^2 + \cdots + |d_{2n}|^2$$

and

$$(D^\dagger D)_{22} = \langle D^2, D^2\rangle = |d_{12}|^2 + |d_{22}|^2 + \cdots + |d_{n2}|^2 .$$

Since the fact that D is upper-triangular means $d_{32} = \cdots = d_{n2} = 0$ and we just showed that $d_{21} = d_{12} = 0$, it again follows by normality that $d_{23} = \cdots = d_{2n} = 0$. Thus all entries in the second row and column with the possible exception of the $(2,2)$ entry must be zero.

Continuing this procedure, it is clear that D must be diagonal as claimed. In other words, *an upper-triangular normal matrix is necessarily diagonal.* This discussion proves the following very important theorem.

Theorem 5.18. *A matrix $N \in M_n(\mathbb{C})$ is normal if and only if there exists a unitary matrix U such that $U^\dagger NU$ is diagonal.*

Corollary. *If $A = (a_{ij}) \in M_n(\mathbb{R})$ is symmetric, then its eigenvalues are real and there exists an orthogonal matrix S such that $S^T AS$ is diagonal.*

Proof. If the eigenvalues are real, then the rest of this corollary follows from the corollary to Theorem 5.17 and the real analogue of the proof of Theorem 5.18.

Now suppose $A = A^T$ so that $a_{ij} = a_{ji}$. If λ is an eigenvalue of A, then there exists a (nonzero and not necessarily real) vector $x \in \mathbb{C}^n$ such that $Ax = \lambda x$ and hence

$$\langle x, Ax\rangle = \lambda\langle x, x\rangle = \lambda \|x\|^2 .$$

On the other hand, using equation (5.3) we see that

$$\langle x, Ax\rangle = \langle A^T x, x\rangle = \langle x, A^T x\rangle^* = \langle x, Ax\rangle^* = \lambda^*\langle x, x\rangle^* = \lambda^* \|x\|^2 .$$

Subtracting these last two equations yields $(\lambda - \lambda^*) \|x\|^2 = 0$ and hence $\lambda = \lambda^*$ since $\|x\| \neq 0$ by definition. ∎

Let us make some observations. Note that any basis relative to which a normal matrix N is diagonal is by definition a basis of eigenvectors. The unitary transition matrix U that diagonalizes N has columns that are precisely these eigenvectors, and since the columns of a unitary matrix are orthonormal, it follows that the eigenvector basis is in fact orthonormal. Of course, the analogous result holds for a real symmetric matrix also.

In fact, in the next chapter we will show directly that the eigenvectors belonging to distinct eigenvalues of a normal operator are orthogonal.

Exercises

1. Finish the proof of the corollary to Theorem 5.17.

2. Show that if $A, B \in M_n(\mathcal{F})$ are unitarily similar and A is normal, then B is also normal.

3. Suppose A, $B \in M_n(\mathbb{C})$ commute (i.e., $AB = BA$).
   (a) Prove there exists a unitary matrix U such that $U^\dagger AU$ and $U^\dagger BU$ are both upper-triangular. [*Hint*: Let $V_\lambda \subset \mathbb{C}^n$ be the eigenspace of B corresponding to the eigenvalue λ. Show V_λ is invariant under A, and hence show that A and B have a common eigenvector $\widetilde{U}^1$. Now proceed as in the proof of Theorem 5.17.]
   (b) Show that if A and B are also normal, then there exists a unitary matrix U such that $U^\dagger AU$ and $U^\dagger BU$ are diagonal.

4. Can every matrix $A \in M_n(\mathcal{F})$ be written as a product of two unitary matrices? Explain.

5. (a) Prove that if H is Hermitian, then $\det H$ is real.
   (b) Is it the case that every square matrix A can be written as the product of finitely many Hermitian matrices? Explain.

6. A matrix M is **skew-Hermitian** if $M^\dagger = -M$.
   (a) Show skew-Hermitian matrices are normal.
   (b) Show any square matrix A can be written as a sum of a skew-Hermitian matrix and a Hermitian matrix.

7. Describe all diagonal unitary matrices. Prove any $n \times n$ diagonal matrix can be written as a finite sum of unitary diagonal matrices. [*Hint*: Do the cases $n = 1$ and $n = 2$ to get the idea.]

8. Using the previous exercise, show any $n \times n$ normal matrix can be written as the sum of finitely many unitary matrices.

9. If A is unitary, does this imply $\det A^k = 1$ for some integer k? What if A is a real, unitary matrix (i.e., orthogonal)?

10. (a) Is an $n \times n$ matrix A that is similar (but not necessarily *unitarily* similar) to a Hermitian matrix necessarily Hermitian?
    (b) If A is similar to a normal matrix, is A necessarily normal?

11. If N is normal and $Nx = \lambda x$, prove $N^\dagger x = \lambda^* x$. [*Hint*: First treat the case where N is diagonal.]

12. Does the fact that A is similar to a diagonal matrix imply A is normal?

13. Discuss the following conjecture: If N_1 and N_2 are normal, then $N_1 + N_2$ is normal if and only if $N_1 N_2^\dagger = N_2^\dagger N_1$.

14. (a) If $A \in M_n(\mathbb{R})$ is nonzero and skew-symmetric, show A can not have any real eigenvalues.
    (b) What can you say about the eigenvalues of such a matrix?
    (c) What can you say about the rank of A?

5.8 The Singular Value Decomposition*

Throughout this section it will be very convenient for us to view a matrix as a function of its rows or columns. For example, consider a matrix $C \in M_{m\times n}(\mathcal{F})$ given by

$$C = \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ c_{21} & c_{22} & \cdots & c_{2n} \\ \vdots & \vdots & & \vdots \\ c_{m1} & c_{m2} & \cdots & c_{mn} \end{bmatrix}.$$

This matrix has columns

$$C^i = \begin{bmatrix} c_{1i} \\ c_{2i} \\ \vdots \\ c_{mi} \end{bmatrix}$$

and we can write C as a row vector where each entry is a column:

$$C = \begin{bmatrix} C^1 & C^2 & \cdots & C^n \end{bmatrix}.$$

We also have the transpose

$$C^T = \begin{bmatrix} c_{11} & c_{21} & \cdots & c_{m1} \\ c_{12} & c_{22} & \cdots & c_{m2} \\ \vdots & \vdots & & \vdots \\ c_{1n} & c_{2n} & \cdots & c_{mn} \end{bmatrix}$$

which we can write as a column vector with entries that are rows, each of which is the transpose of the corresponding column of C:

$$C^T = \begin{bmatrix} (C^1)^T \\ (C^2)^T \\ \vdots \\ (C^n)^T \end{bmatrix}.$$

With this notation out of the way, let us turn to the problem at hand. We know that not every matrix can be diagonalized, but we have seen (Theorem 5.18) that a normal matrix $A \in M_n(\mathbb{C})$ can be diagonalized by finding its eigenvalues and eigenvectors. But if A is not square, then it can't even have any eigenvectors (because it takes a vector of one dimension and maps it into a vector of a different dimension). However, if $A \in M_{m\times n}(\mathcal{F})$, then $A^\dagger A$ is a Hermitan $n \times n$ matrix (i.e., $(A^\dagger A)^\dagger = A^\dagger A$ so it's normal) and *it* can therefore be diagonalized. Note that $A : \mathcal{F}^n \to \mathcal{F}^m$ and $A^\dagger A : \mathcal{F}^n \to \mathcal{F}^n$.

What can we say about the eigenvalues of $A^\dagger A$? Well, if $A^\dagger Ax = \lambda x$ then

$$\langle x, A^\dagger Ax\rangle = \langle Ax, Ax\rangle = \|Ax\|^2 \geq 0$$

so that

$$0 \leq \langle x, A^\dagger Ax\rangle = \langle x, \lambda x\rangle = \lambda\langle x, x\rangle = \lambda \|x\|^2 .$$

Since $\|x\|^2 \geq 0$, it follows that $\lambda \geq 0$, and hence the eigenvalues of $A^\dagger A$ are not only real, but are in fact also nonnegative.

Because of this, we can take the square roots of the eigenvalues λ_i, and we define the **singular values** of A to be the numbers $\sigma_i = \sqrt{\lambda_i}$. (Some authors define the singular values to be the square root of the nonzero eigenvalues only.) It is conventional to list the singular values in decreasing order $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n$ where repeated roots are listed separately. The orthonormal eigenvectors of $A^\dagger A$ are called the **singular vectors**, and they form an orthonormal basis for $\mathcal{F}^n$.

Let the orthonormal eigenvectors of $A^\dagger A$ be denoted by $\{v_1, \ldots, v_n\}$ so that $A^\dagger Av_i = \lambda_i v_i$. Then

$$\langle Av_i, Av_j\rangle = \langle v_i, A^\dagger Av_j\rangle = \langle v_i, \lambda_j v_j\rangle = \lambda_j\langle v_i, v_j\rangle = \lambda_j\delta_{ij}$$

so the vectors $Av_i \in \mathcal{F}^m$ are orthogonal and $\|Av_i\| = \langle Av_i, Av_i\rangle^{1/2} = \sqrt{\lambda_i} = \sigma_i$. If we let $\lambda_1, \ldots, \lambda_r$ be the nonzero eigenvalues, then for $i = 1, \ldots, r$ we may normalize the Av_i by defining the vectors $u_i \in \mathcal{F}^m$ as

$$u_i = \frac{1}{\sigma_i}Av_i, \qquad i = 1, \ldots, r$$

where now $\langle u_i, u_j \rangle = \delta_{ij}$.

If $r < m$ then the $\{u_i\}$ will not be a basis for $\mathcal{F}^m$, and in this case we assume that they are somehow extended to form a complete orthonormal basis for $\mathcal{F}^m$. However this is accomplished, we then define the matrix $U \in M_m(\mathcal{F})$ to have u_i as its ith column. By Theorem 5.15 we know that U is unitary. We also use the orthonormal eigenvectors $v_i \in \mathcal{F}^n$ as the columns of another unitary matrix $V \in M_n(\mathcal{F})$.

By construction, we have $Av_i = \sigma_i u_i$ for $i = 1, \ldots, r$ and $Av_i = 0$ for $i = r+1, \ldots, n$. Let us write the matrix AV as a function of its columns. Recalling that the ith column of AV is given by $AV^i = Av_i$ we have

$$\begin{aligned} AV &= [\,(AV)^1\ (AV)^2\ \cdots\ (AV)^n\,] = [\,Av_1\ Av_2\ \cdots\ Av_n\,] \\ &= [\,\sigma_1 u_1\ \sigma_2 u_2\ \cdots\ \sigma_r u_r\ 0\ \cdots\ 0\,]. \end{aligned}$$

Defining the block diagonal matrix $\Sigma \in M_{m\times n}(\mathcal{F})$ by

$$\Sigma = \left[\begin{array}{ccc|c} \sigma_1 & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & \sigma_r & 0 \\ \hline 0 & \cdots & 0 & 0 \end{array}\right] \tag{5.4}$$

where the 0's are of the appropriate size, we may write the right side of the equation for AV as (remember the u_i are column vectors)

$$[\,\sigma_1 u_1\ \sigma_2 u_2\ \cdots\ \sigma_r u_r\ 0\ \cdots\ 0\,] = [\,u_1\ u_2\ \cdots\ u_m\,]\left[\begin{array}{ccc|c} \sigma_1 & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & \sigma_r & 0 \\ \hline 0 & \cdots & 0 & 0 \end{array}\right]$$

or simply $AV = U\Sigma$. Finally, using the fact that V is unitary (so $V^{-1} = V^\dagger$), we have proved the following important theorem, called the **singular value decomposition** (abbreviated SVD).

Theorem 5.19. *Let $A \in M_{m\times n}(\mathcal{F})$ have singular values $\sigma_1, \ldots, \sigma_r > 0$ and $\sigma_{r+1}, \ldots, \sigma_n = 0$. Then there exists a unitary matrix $U \in M_m(\mathcal{F})$, a unitary matrix $V \in M_n(\mathcal{F})$ and a matrix $\Sigma \in M_{m\times n}(\mathcal{F})$ of the form shown in equation* (5.4) *such that*

$$A = U\Sigma V^\dagger.$$

Corollary. *Let $A \in M_{m\times n}(\mathbb{R})$ have singular values $\sigma_1, \ldots, \sigma_r > 0$ and $\sigma_{r+1}, \ldots, \sigma_n = 0$. Then there exists an orthogonal matrix $U \in M_m(\mathbb{R})$, an orthogonal matrix $V \in M_n(\mathbb{R})$ and a matrix $\Sigma \in M_{m\times n}(\mathbb{R})$ of the form shown in equation* (5.4) *such that*

$$A = U\Sigma V^T.$$

Proof. This follows exactly as in the theorem by using the corollaries to Theorems 5.15 and 5.18. ∎

Example 5.14. Let us find the SVD of the matrix

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

In this case $A \in M_{m\times n}(\mathbb{R})$, so we use the corollary to Theorem 5.19.

We have

$$A^T A = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

To find the eigenvalues of $A^T A$ we solve

$$\begin{aligned} \det(A^T A - \lambda I) &= \begin{vmatrix} 1-\lambda & 1 & 0 \\ 1 & 1-\lambda & 0 \\ 0 & 0 & 1-\lambda \end{vmatrix} = (1-\lambda)[(1-\lambda)^2 - 1] \\ &= (\lambda-2)(\lambda-1)\lambda = 0 \end{aligned}$$

and therefore $\lambda_1 = 2, \lambda_2 = 1$ and $\lambda_3 = 0$.

Now we find the corresponding eigenvectors $\mathbf{x}$ by solving $(A^T A - \lambda_i I)\mathbf{x} = 0$. For $\lambda_1 = 2$ we have

$$\begin{bmatrix} -1 & 1 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

or $x = y, z = 0$ so that the normalized singular vector is

$$v_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}.$$

For $\lambda_2 = 1$ we have

$$\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

or $x = y = 0$ with no restriction on z, so the normalized singular vector is

$$v_2 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

Finally, for $\lambda_3 = 0$ we have

$$\begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

so that $x = -y$ and $z = 0$ and the normalized singular vector is

$$v_3 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}.$$

Note that $\{v_1, v_2, v_3\}$ are orthonormal. The singular values are $\sigma_1 = \sqrt{2}, \sigma_2 = 1$ and $\sigma_3 = 0$ and the matrices $V \in M_3(\mathbb{R})$ and $\Sigma \in M_{2\times 3}(\mathbb{R})$ are given by

$$V = \begin{bmatrix} 1/\sqrt{2} & 0 & 1/\sqrt{2} \\ 1/\sqrt{2} & 0 & -1/\sqrt{2} \\ 0 & 1 & 0 \end{bmatrix} \qquad \Sigma = \begin{bmatrix} \sqrt{2} & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}.$$

To find the matrix $U \in M_2(\mathbb{R})$ we compute

$$u_1 = \frac{1}{\sigma_1} A v_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$u_2 = \frac{1}{\sigma_2} A v_2 = \frac{1}{1} \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

which are already an orthonormal basis for $\mathbb{R}^2$, and hence

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

Putting this all together we have the SVD

$$\begin{aligned} A = U\Sigma V^T &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \sqrt{2} & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} & 0 \\ 0 & 0 & 1 \\ 1/\sqrt{2} & -1/\sqrt{2} & 0 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \end{aligned}$$

Example 5.15. Let us find the SVD of the matrix

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

We form

$$A^T A = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

and find the eigenvalues from

$$\begin{vmatrix} 2-\lambda & 1 \\ 1 & 2-\lambda \end{vmatrix} = \lambda^2 - 4\lambda + 3 = (\lambda-1)(\lambda-3) = 0$$

so that $\lambda_1 = 3$ and $\lambda_2 = 1$. Therefore the singular values of A are $\sigma_1 = \sqrt{3}$ and $\sigma_2 = 1$.

We next find the corresponding eigenvectors from $(A^TA - \lambda_i I)\mathbf{x} = 0$. For $\lambda_1 = 3$ we solve

$$\begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

so that $x = y$ and the normalized singular vector is

$$v_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

For $\lambda_2 = 1$ we have

$$\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

so that $x = -y$ and

$$v_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

This gives us the matrices

$$V = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \qquad \Sigma = \begin{bmatrix} \sqrt{3} & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}.$$

To find U we first calculate

$$u_1 = \frac{1}{\sigma_1} A v_1 = \frac{1}{\sqrt{6}} \begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \frac{1}{\sqrt{6}} \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix}$$

$$u_2 = \frac{1}{\sigma_2} A v_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}.$$

However, in this case we need another linearly independent vector u_3 so that we have a basis for $\mathbb{R}^3$. The easiest way is to simply note that by inspection the standard basis vector e_3 of $\mathbb{R}^3$ is fairly obviously independent of u_1 and u_2, so we need only apply the Gram-Schmidt process to orthogonalize the three vectors. Since u_1 and u_2 are already orthonormal to each other, we subtract off the component of e_3 in the direction of both of these and then normalize the result. In other words, using the standard inner product on $\mathbb{R}^3$ we have

$$\begin{aligned} e_3 - \langle u_1, e_3\rangle u_1 - \langle u_2, e_3\rangle u_2 &= e_3 - \frac{1}{\sqrt{6}} u_1 - \frac{(-1)}{\sqrt{2}} u_2 \\ &= \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} - \frac{1}{6}\begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix} + \frac{1}{2}\begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix} = \frac{1}{3}\begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix} \end{aligned}$$

so normalizing this we obtain

$$u_3 = \frac{1}{\sqrt{3}}\begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}.$$

The matrix U is now given by

$$U = \begin{bmatrix} 2/\sqrt{6} & 0 & -1/\sqrt{3} \\ 1/\sqrt{6} & 1/\sqrt{2} & 1/\sqrt{3} \\ 1/\sqrt{6} & -1/\sqrt{2} & 1/\sqrt{3} \end{bmatrix}$$

and the SVD is

$$\begin{aligned} A = U\Sigma V^T &= \frac{1}{\sqrt{2}}\begin{bmatrix} 2/\sqrt{6} & 0 & -1/\sqrt{3} \\ 1/\sqrt{6} & 1/\sqrt{2} & 1/\sqrt{3} \\ 1/\sqrt{6} & -1/\sqrt{2} & 1/\sqrt{3} \end{bmatrix}\begin{bmatrix} \sqrt{3} & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}. \end{aligned}$$

One very interesting consequence of the SVD is that it helps visualize the geometric effect of a linear transformation. (We will have more to say about geometry and the determinant of a linear transformation in Section 8.7.) First we need some basic results dealing with the rank of a matrix.

Theorem 5.20. *Let $A \in M_{m\times n}(\mathcal{F})$ and let $P \in M_m(\mathcal{F})$ and $Q \in M_n(\mathcal{F})$ be nonsingular. Then* $\operatorname{rank}(PA) = \operatorname{rank}(A)$ *and* $\operatorname{rank}(AQ) = \operatorname{rank}(A)$.

Proof. If P is nonsingular, then from the corollary to Theorem 2.17 we have

$$\operatorname{rank}(PA) \le \min\{\operatorname{rank} A, \operatorname{rank} P\} = \operatorname{rank}(A).$$

But P^{-1} is also nonsingular so the same corollary shows that

$$\operatorname{rank}(A) = \operatorname{rank}(P^{-1}(PA)) \le \operatorname{rank}(PA) \le \operatorname{rank}(A)$$

and this implies that $\operatorname{rank}(PA) = \operatorname{rank}(A)$.

An exactly analogous argument using $A = (AQ)Q^{-1}$ shows that $\operatorname{rank}(AQ) = \operatorname{rank}(A)$. ∎

As an important application of this theorem, suppose we have a linear transformation T with matrix representation $A = [T]_e$ with respect to some basis $\{e_i\}$. If $\bar{A} = [T]_{\bar{e}}$ is its representation with respect to another basis $\{\bar{e}_i\}$, then we know that $\bar{A} = P^{-1}AP$ where P is the nonsingular transition matrix from the basis $\{e_i\}$ to the basis $\{\bar{e}_i\}$ (Theorem 4.15). Then from Theorem 5.20 we see that

$$\operatorname{rank}(P^{-1}AP) = \operatorname{rank}(AP) = \operatorname{rank} A$$

and hence the rank of a linear transformation is independent of its representation (which is to be expected).

In particular, if A is diagonalizable, then its diagonal form consists of its eigenvalues along the diagonal, and hence the rank of A is just the number of nonzero eigenvalues (which is the number of nonzero rows in what is the row echelon form). This proves the next theorem.

Theorem 5.21. *Let $A \in M_n(\mathcal{F})$ be diagonalizable. Then* $\operatorname{rank}(A)$ *is just the number of nonzero eigenvalues of A.*

Before proving our next theorem, let us point out that given vectors $X, Y \in \mathbb{C}^n$, their inner product can be written in the equivalent forms

$$\langle X, Y \rangle = \sum_{i=1}^{n} x_i^* y_i = X^{*T} Y = X^\dagger Y$$

where $X^{*T} Y = X^\dagger Y$ is the matrix product of a $1 \times n$ matrix X^{*T} with an $n \times 1$ matrix Y. In particular, we have

$$\|X\|^2 = \langle X, X \rangle = X^\dagger X.$$

Theorem 5.22. *Let $A \in M_{m \times n}(\mathcal{F})$. Then* $\operatorname{rank}(A^\dagger A) = \operatorname{rank}(A)$.

Proof. Since both $A \in M_{m\times n}(\mathcal{F})$ and $A^\dagger A \in M_n(\mathcal{F})$ have the same number n of columns, the rank theorem for matrices (Theorem 2.18) shows that

$$\operatorname{rank} A + \dim(\ker A) = n = \operatorname{rank}(A^\dagger A) + \dim(\ker(A^\dagger A)).$$

Because of this, we need only show that $\dim(\ker A) = \dim(\ker(A^\dagger A))$.

So, if $X \in \ker A$, then $AX = 0$ so that $(A^\dagger A)X = A^\dagger(AX) = 0$ and therefore $X \in \ker(A^\dagger A)$, i.e., $\ker A \subset \ker(A^\dagger A)$. On the other hand, suppose

$X \in \ker(A^\dagger A)$ so that $(A^\dagger A)X = 0$. Since X, AX and $(A^\dagger A)X$ are all vectors in $\mathcal{F}^n$, we can use the standard inner product on $\mathbb{C}^n$ to write

$$0 = \langle X, (A^\dagger A)X\rangle = X^\dagger(A^\dagger A)X = (AX)^\dagger(AX) = \|AX\|^2 .$$

Therefore, the fact that the norm is positive definite (property (N1) in Theorem 1.17) implies $AX = 0$ so that $X \in \ker A$, i.e., $\ker(A^\dagger A) \subset \ker A$. Therefore $\ker A = \ker(A^\dagger A)$ and it follows that $\dim(\ker A) = \dim(\ker(A^\dagger A))$ as required. ∎

The following corollary is immediate.

Corollary. *Let* $A \in M_{m\times n}(\mathbb{R})$. *Then* $\operatorname{rank}(A^T A) = \operatorname{rank}(A)$.

We are now ready to prove a number of propertiesof the SVD in the special case of a matrix $A \in M_{m\times n}(\mathbb{R})$. Recall that the row space of A is denoted by $\operatorname{row}(A)$, and the row rank by $\operatorname{rr}(A)$; the column space by $\operatorname{col}(A)$ with rank $\operatorname{cr}(A)$; and the kernel (or null space) of A by $\ker(A)$.

Theorem 5.23. *Let* $A \in M_{m\times n}(\mathbb{R})$ *have nonzero singular values* $\sigma_1, \dots, \sigma_r$ *and singular value decomposition* $U\Sigma V^T$. *In other words,* $A^T A$ *has eigenvalues* $\sigma_1^2 \ge \cdots \ge \sigma_n^2$ *and corresponding orthonormal eigenvectors* $v_1, \dots, v_n$*; and for* $i = 1, \dots, r$ *we have the orthonormal vectors* $u_i = (1/\sigma_i)Av_i$ *which (if* $r < m$*) are extended to form an orthonormal basis* $\{u_1, \dots, u_m\}$ *for* $\mathbb{R}^m$. *Then*

(i) $\operatorname{rank}(A) = r$.
(ii) $\{u_1, \dots, u_r\}$ *is an orthonormal basis for* $\operatorname{col}(A)$.
(iii) $\{u_{r+1}, \dots, u_m\}$ *is an orthonormal basis for* $\ker(A^T)$.
(iv) $\{v_{r+1}, \dots, v_n\}$ *is an orthonormal basis for* $\ker(A)$.
(v) $\{v_1, \dots, v_r\}$ *is an orthonormal basis for* $\operatorname{row}(A)$.

Proof. (i) It follows from Theorem 5.21 that $\operatorname{rank}(A^T A) = r$, and from Theorem 5.22 that $\operatorname{rank}(A) = \operatorname{rank}(A^T A)$.

Alternatively, since U and V are orthogonal (and hence nonsingular), we can use Theorem 5.20 directly to see that $\operatorname{rank}(A) = \operatorname{rank}(U\Sigma V^T) = \operatorname{rank}(U\Sigma) = \operatorname{rank}(\Sigma) = r$.

(ii) As we have seen in Example 2.10, if $A \in M_{m\times n}(\mathbb{R})$ and $X \in \mathbb{R}^n$, then $AX = \sum_{i=1}^n A^i x_i$ is a linear combination of the columns of A. In particular, each Av_i is a linear combination of the columns of A, so that for each $i = 1, \dots, r$ we see that $u_i = (1/\sigma_i)Av_i$ is a linear combination of the columns of A and hence is in the column space of A. Since $\{u_1, \dots, u_r\}$ are orthonormal, they are linearly independent (Theorem 1.19), and there are $r = \operatorname{rank}(A) = \operatorname{cr}(A)$ of them. Hence they form a basis for $\operatorname{col}(A)$.

(iii) By Theorem 1.22 we can write $\mathbb{R}^m = \operatorname{col}(A) \oplus (\operatorname{col}(A))^\perp$. But $\{u_1, \dots, u_m\}$ is an orthonormal basis for $\mathbb{R}^m$ while by part (ii) $\{u_1, \dots, u_r\}$ is a basis for col(A). Therefore it must be that $\{u_{r+1}, \dots, u_m\}$ is a basis for $(\mathrm{col}(A))^\perp$, which is just $\ker(A^T)$ by Theorem 2.19.

(iv) By construction of the SVD, we know that $Av_i = 0$ for $i = r+1, \dots, n$. Therefore $v_{r+1}, \dots, v_n$ are all in $\ker(A)$. Furthermore, these $n-r$ vectors are linearly independent because they are orthonormal. By the rank theorem for matrices (Theorem 2.18) we know that $\dim(\ker A) = n - \mathrm{rank}(A) = n - r$, and hence $\{v_{r+1}, \dots, v_n\}$ must form an orthonormal basis for $\ker(A)$.

(v) By Theorem 1.22 again we can write $\mathbb{R}^n = \ker(A) \oplus (\ker(A))^\perp$ where $\{v_1, \dots, v_n\}$ is an orthonormal basis for $\mathbb{R}^n$ and $\{v_{r+1}, \dots, v_n\}$ is an orthonormal basis for $\ker(A)$. But then $\{v_1, \dots, v_r\}$ are all in $(\ker(A))^\perp$ which is of dimension $n - \dim(\ker(A)) = n - (n-r) = r$. Since $\{v_1, \dots, v_r\}$ are linearly independent (they are orthonormal), they must be a basis for $(\ker(A))^\perp$, which is the same as row(A) by Theorem 2.19 and the corollary to Theorem 1.22. ∎

Now suppose we have a *unit vector* $X \in \mathbb{R}^n$, and let us look at the effect of $A \in M_{m\times n}(\mathbb{R})$ acting on X. Write the SVD of A in the usual manner as $A = U\Sigma V^T$ where r is the number of nonzero singular values of A. Expressing U as a function of its columns and V^T as a function of its rows, we have

$$\begin{aligned} AX &= U\Sigma V^T X \\ &= \begin{bmatrix} u_1 & \cdots & u_m \end{bmatrix} \left[\begin{array}{ccc|c} \sigma_1 & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & \sigma_r & 0 \\ \hline 0 & \cdots & 0 & 0 \end{array}\right] \begin{bmatrix} v_1^T \\ \vdots \\ v_n^T \end{bmatrix} X. \end{aligned}$$

But each v_i^T is a (row) vector in $\mathbb{R}^n$ and X is a column vector in $\mathbb{R}^n$, so

$$V^T X = \begin{bmatrix} v_1^T \\ \vdots \\ v_n^T \end{bmatrix} X = \begin{bmatrix} v_1^T X \\ \vdots \\ v_n^T X \end{bmatrix}$$

where now each $v_i^T X$ is just a scalar. In fact, since $\|X\| = 1$ and V is orthogonal, we see that

$$\left\|V^T X\right\|^2 = \langle V^T X, V^T X\rangle = \langle X, VV^T X\rangle = \langle X, X\rangle = 1$$

so that $V^T X$ is also a unit vector. Written out this says

$$(v_1^T X)^2 + \cdots + (v_n^T X)^2 = 1. \tag{5.5}$$

Next we have

$$\Sigma V^T X = \left[\begin{array}{ccc|c} \sigma_1 & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & \sigma_r & 0 \\ \hline 0 & \cdots & 0 & 0 \end{array}\right] \begin{bmatrix} v_1^T X \\ \vdots \\ v_r^T X \\ v_{r+1}^T X \\ \vdots \\ v_n^T X \end{bmatrix} = \begin{bmatrix} \sigma_1 v_1^T X \\ \vdots \\ \sigma_r v_r^T X \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

where again each $\sigma_i v_i^T X$ is a scalar. Finally, we have

$$AX = U\Sigma V^T X = \begin{bmatrix} u_1 & \cdots & u_r & u_{r+1} & \cdots & u_m \end{bmatrix} \begin{bmatrix} \sigma_1 v_1^T X \\ \vdots \\ \sigma_r v_r^T X \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

or

$$AX = (\sigma_1 v_1^T X)u_1 + \cdots + (\sigma_r v_r^T X)u_r. \tag{5.6}$$

Now we are ready to see how the linear transformation A acts on the unit sphere in $\mathbb{R}^n$.

Theorem 5.24. *Let $A \in M_{m\times n}(\mathbb{R})$ have singular value decomposition $U\Sigma V^T$ as in the corollary to Theorem 5.19. Then $\{AX : X \in \mathbb{R}^n, \|X\| = 1\}$ is either the surface of an ellipsoid in $\mathbb{R}^m$ if* $\operatorname{rank}(A) = r = n$*, or a solid ellipsoid in $\mathbb{R}^m$ if* $\operatorname{rank}(A) = r < n$.

Proof. Observe that $A : \mathbb{R}^n \to \mathbb{R}^m$, and that $\operatorname{rank}(A)$ is the dimension of the image of A. By Theorem 5.23(i), we know that $\operatorname{rank}(A) = r$ is the number of nonzero singular values of A.

First, if $\operatorname{rank}(A) = r = n$, then necessarily $n \le m$. And from equation (5.6) we see that

$$AX = (\sigma_1 v_1^T X)u_1 + \cdots + (\sigma_n v_n^T X)u_n := y_1 u_1 + \cdots + y_n u_n \in \mathbb{R}^m$$

where we have defined the scalars $y_i = \sigma_i v_i^T X$. Since $n \le m$, let us define the vector $Y \in \mathbb{R}^m$ by

$$Y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

Then we can write this last equation in the form $AX = UY$, and using the fact that U is orthogonal we see that

$$\begin{aligned} \|AX\|^2 = \|UY\|^2 = \|Y\|^2 &= {y_1}^2 + \cdots + {y_n}^2 \\ &= (\sigma_1 v_1^T X)^2 + \cdots + (\sigma_n v_n^T X)^2. \end{aligned}$$

But then using equation (5.5) we have

$$\frac{{y_1}^2}{{\sigma_1}^2} + \cdots + \frac{{y_n}^2}{{\sigma_n}^2} = (v_1^T X)^2 + \cdots + (v_n^T X)^2 = 1$$

which is the equation of an ellipsoid in $\mathbb{R}^m$. In other words, A takes a vector X on the unit sphere in $\mathbb{R}^n$ and maps it to the vector UY in $\mathbb{R}^m$, where the vectors UY describe an ellipsoid (i.e., a surface in $\mathbb{R}^m$).

Now suppose that $\operatorname{rank}(A) = r < n$. From equation (5.6) again we now have

$$AX = y_1 u_1 + \cdots + y_r u_r$$

so we define

$$Y = \begin{bmatrix} y_1 \\ \vdots \\ y_r \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

and $AX = UY$ with $\|AX\|^2 = \|Y\|^2 = {y_1}^2 + \cdots + {y_r}^2$ as in the first case. But now, since $r < n$ we have

$$\frac{{y_1}^2}{{\sigma_1}^2} + \cdots + \frac{{y_r}^2}{{\sigma_r}^2} = (v_1^T X)^2 + \cdots + (v_r^T X)^2 \le (v_1^T X)^2 + \cdots + (v_n^T X)^2 = 1.$$

In other words,

$$\frac{{y_1}^2}{{\sigma_1}^2} + \cdots + \frac{{y_r}^2}{{\sigma_r}^2} \le 1$$

which is the equation of a solid ellipsoid in $\mathbb{R}^m$. ∎

Exercises

1. Find the SVD of each of the following matrices.

(a) $\begin{bmatrix} 3 \\ 4 \end{bmatrix}$ (b) $\begin{bmatrix} 3 & 4 \end{bmatrix}$ (c) $\begin{bmatrix} 0 & 0 \\ 0 & 3 \\ -2 & 0 \end{bmatrix}$ (d) $\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ -2 & 2 \end{bmatrix}$

(e) $\begin{bmatrix} 2 & 0 & 1 \\ 0 & 2 & 0 \end{bmatrix}$ (f) $\begin{bmatrix} 2 & 1 & 0 & -1 \\ 0 & -1 & 1 & 1 \end{bmatrix}$ (g) $\begin{bmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix}$

2. Prove that if $A \in M_n(\mathcal{F})$ is nonsingular, then the singular values of A^{-1} are the inverses of the singular values of A.

Chapter 6

Operators and Diagonalization

6.1 The Adjoint Operator

Let V be finite-dimensional over $\mathbb{C}$, and assume that V has an inner product $\langle \cdot\,, \cdot \rangle$ defined on it. Thus for any $X, Y \in V$ we have $\langle X, Y \rangle \in \mathbb{C}$. For example, with respect to the standard basis $\{e_i\}$ for $\mathbb{C}^n$ (which is the same as the standard basis for $\mathbb{R}^n$), we have $X = \sum x^i e_i$ and hence

$$\begin{aligned}\langle X, Y \rangle &= \Big\langle \sum_i x^i e_i, \sum_j y^j e_j \Big\rangle = \sum_{i,j} x^{i*} y^j \langle e_i, e_j \rangle = \sum_{i,j} x^{i*} y^j \delta_{ij} \\ &= \sum_i x^{i*} y^i = X^{*T} Y.\end{aligned}$$

In particular, for any $T \in L(V)$ and $X \in V$ we have the vector $TX \in V$, and hence it is meaningful to write expressions of the form $\langle TX, Y \rangle$ and $\langle X, TY \rangle$.

Since we are dealing with finite-dimensional vector spaces, the Gram-Schmidt process (Theorem 1.21) guarantees we can always work with an orthonormal basis. Hence, let us consider a complex inner product space V with basis $\{e_i\}$ such that $\langle e_i, e_j \rangle = \delta_{ij}$. Then we see that for any $u = \sum u^j e_j \in V$ we have

$$\langle e_i, u \rangle = \Big\langle e_i, \sum_j u^j e_j \Big\rangle = \sum_j u^j \langle e_i, e_j \rangle = \sum_j u^j \delta_{ij} = u^i$$

and therefore

$$u = \sum_i \langle e_i, u \rangle e_i.$$

Now consider the vector Te_j. Applying this last equation to the vector $u = Te_j$ we have

$$Te_j = \sum_i \langle e_i, Te_j \rangle e_i.$$

But this is precisely the definition of the matrix $A = (a_{ij})$ that represents T relative to the basis $\{e_i\}$. In other words, this extremely important result shows that the **matrix elements** a_{ij} **of the operator** $T \in L(V)$ are given by

$$a_{ij} = \langle e_i, Te_j \rangle.$$

It is important to note however, that this definition depended on the use of an orthonormal basis for V. To see the self-consistency of this definition, we go back to our original definition of (a_{ij}) as $Te_j = \sum_k e_k a_{kj}$. Taking the scalar product of both sides of this equation with e_i yields (using the orthonormality of the e_i)

$$\langle e_i, Te_j \rangle = \left\langle e_i, \sum_k e_k a_{kj} \right\rangle = \sum_k a_{kj} \langle e_i, e_k \rangle = \sum_k a_{kj} \delta_{ik} = a_{ij}.$$

Recall from Section 4.1 that the dual vector space $V^* = L(V, \mathcal{F}) : V \to \mathcal{F}$ is defined to be the space of linear functionals on V. In other words, if $\phi \in V^*$, then for every u, $v \in V$ and a, $b \in \mathcal{F}$ we have

$$\phi(au + bv) = a\phi(u) + b\phi(v) \in \mathcal{F}.$$

We now prove a very important result that is the basis for the definition of the operator adjoint.

Theorem 6.1. *Let V be a finite-dimensional inner product space over $\mathbb{C}$. Then, given any linear functional L on V, there exists a unique $u \in V$ such that $Lv = \langle u, v \rangle$ for all $v \in V$.*

Proof. Let $\{e_i\}$ be an orthonormal basis for V and define $u = \sum_i (Le_i)^* e_i$. Now define the linear functional L_u on V by $L_u v = \langle u, v \rangle$ for every $v \in V$. Then, in particular, we have

$$L_u e_i = \langle u, e_i \rangle = \left\langle \sum_j (Le_j)^* e_j, e_i \right\rangle = \sum_j Le_j \langle e_j, e_i \rangle = \sum_j Le_j \delta_{ji} = Le_i.$$

Since L and L_u agree on a basis for V, they must agree on any $v \in V$, and hence $L = L_u = \langle u, \cdot \rangle$.

As to the uniqueness of the vector u, suppose $u' \in V$ has the property that $Lv = \langle u', v \rangle$ for every $v \in V$. Then $Lv = \langle u, v \rangle = \langle u', v \rangle$ so that $\langle u - u', v \rangle = 0$. Since v was arbitrary we may choose $v = u - u'$. Then $\langle u - u', u - u' \rangle = 0$ which implies (since the inner product is positive definite) $u - u' = 0$ or $u = u'$. ∎

The importance of finite-dimensionality in this theorem is shown by the following example.

Example 6.1. Let $V = \mathbb{R}[x]$ be the (infinite-dimensional) space of all polynomials over $\mathbb{R}$, and define an inner product on V by

$$\langle f, g\rangle = \int_0^1 f(x)g(x)\,dx$$

for every $f, g \in V$. We will give an example of a linear functional L on V for which there does not exist a polynomial $h \in V$ with the property that $Lf = \langle h, f\rangle$ for all $f \in V$.

To show this, define the nonzero linear functional L by

$$Lf = f(0).$$

(L is nonzero since, e.g., $L(a+x) = a$.) Now suppose there exists a polynomial $h \in V$ such that $Lf = f(0) = \langle h, f\rangle$ for every $f \in V$. Then, in particular, we have

$$L(xf) = 0f(0) = 0 = \langle h, xf\rangle$$

for every $f \in V$. Choosing $f = xh$ we see that

$$0 = \langle h, x^2h\rangle = \int_0^1 x^2h^2\,dx.$$

Since the integrand is strictly positive, this forces h to be the zero polynomial. Thus we are left with $Lf = \langle h, f\rangle = \langle 0, f\rangle = 0$ for every $f \in V$, and hence $L = 0$. But this contradicts the fact that $L \neq 0$, and hence no such polynomial h can exist.

Note the fact that V is infinite-dimensional is required when we choose $f = xh$. The reason for this is that if V consisted of all polynomials of degree less than or equal to some positive integer N, then $f = xh$ could have degree greater than N.

Now consider an operator $T \in L(V)$, and let u be an arbitrary element of V. Then the mapping $L_u : V \to \mathbb{C}$ defined by $L_u v = \langle u, Tv\rangle$ for every $v \in V$ is a linear functional on V. Applying Theorem 6.1, we see there exists a unique $u' \in V$ such that $\langle u, Tv\rangle = L_u v = \langle u', v\rangle$ for every $v \in V$. We now define the mapping $T^\dagger : V \to V$ by $T^\dagger u = u'$. In other words, we define the **adjoint** $T^\dagger$ of an operator $T \in L(V)$ by

$$\langle T^\dagger u, v\rangle = \langle u, Tv\rangle$$

for all $u, v \in V$. The mapping $T^\dagger$ is unique because u' is unique for a given u. Thus, if $\widetilde{T}^\dagger u = u' = T^\dagger u$, then $(\widetilde{T}^\dagger - T^\dagger)u = 0$ for every $u \in V$, and hence $\widetilde{T}^\dagger - T^\dagger = 0$ or $\widetilde{T}^\dagger = T^\dagger$.

Note further that

$$\langle Tu, v\rangle = \langle v, Tu\rangle^* = \langle T^\dagger v, u\rangle^* = \langle u, T^\dagger v\rangle.$$

However, it follows from the definition that $\langle u, T^\dagger v\rangle = \langle T^{\dagger\dagger} u, v\rangle$. Therefore the uniqueness of the adjoint implies $T^{\dagger\dagger} = T$.

Let us show the map $T^\dagger$ is linear. For all $u_1, u_2, v \in V$ and $a, b \in \mathbb{C}$ we have

$$\begin{aligned}\langle T^\dagger(au_1 + bu_2), v\rangle &= \langle au_1 + bu_2, Tv\rangle \\ &= a^*\langle u_1, Tv\rangle + b^*\langle u_2, Tv\rangle \\ &= a^*\langle T^\dagger u_1, v\rangle + b^*\langle T^\dagger u_2, v\rangle \\ &= \langle aT^\dagger u_1, v\rangle + \langle bT^\dagger u_2, v\rangle \\ &= \langle aT^\dagger u_1 + bT^\dagger u_2, v\rangle.\end{aligned}$$

Since this is true for every $v \in V$, we must have

$$T^\dagger(au_1 + bu_2) = aT^\dagger u_1 + bT^\dagger u_2.$$

Thus $T^\dagger$ is linear and $T^\dagger \in L(V)$.

If $\{e_i\}$ is an orthonormal basis for V, then the matrix elements of T are given by $a_{ij} = \langle e_i, Te_j\rangle$. Similarly, the matrix elements b_{ij} of $T^\dagger$ are related to those of T because

$$b_{ij} = \langle e_i, T^\dagger e_j\rangle = \langle Te_i, e_j\rangle = \langle e_j, Te_i\rangle^* = a_{ji}^*.$$

In other words, if A is the matrix representation of T relative to the orthonormal basis $\{e_i\}$, then A^{*T} is the matrix representation of $T^\dagger$. This explains the symbol and terminology used for the Hermitian adjoint. Note that if V is a real vector space, then the matrix representation of $T^\dagger$ is simply A^T, and we may denote the corresponding operator by T^T.

We summarize this discussion in the following theorem, which is valid only in finite-dimensional vector spaces. (It is also worth pointing out that $T^\dagger$ depends on the particular inner product defined on V.)

Theorem 6.2. *Let T be a linear operator on a finite-dimensional complex inner product space V. Then there exists a unique linear operator $T^\dagger$ on V defined by $\langle T^\dagger u, v\rangle = \langle u, Tv\rangle$ for all $u, v \in V$. Furthermore, if A is the matrix representation of T relative to an orthonormal basis $\{e_i\}$, then $A^\dagger = A^{*T}$ is the matrix representation of $T^\dagger$ relative to this same basis. If V is a real space, then the matrix representation of $T^\dagger$ is simply A^T.*

Example 6.2. Let us give an example that shows the importance of finite-dimensionality in defining an adjoint operator. Consider the space $V = \mathbb{R}[x]$ of all polynomials over $\mathbb{R}$, and let the inner product be as in Example 6.1. Define the differentiation operator $D \in L(V)$ by $Df = df/dx$. We show there exists no adjoint operator $D^\dagger$ that satisfies $\langle Df, g\rangle = \langle f, D^\dagger g\rangle$.

Using $\langle Df, g\rangle = \langle f, D^\dagger g\rangle$, we integrate by parts to obtain

$$\begin{aligned}\langle f, D^\dagger g\rangle = \langle Df, g\rangle &= \int_0^1 (Df)g\,dx = \int_0^1 [D(fg) - fDg]\,dx \\ &= (fg)(1) - (fg)(0) - \langle f, Dg\rangle.\end{aligned}$$

Rearranging, this general result may be written as

$$\langle f, (D + D^\dagger)g\rangle = (fg)(1) - (fg)(0).$$

We now let $f = x^2(1-x)^2 p$ for any $p \in V$. Then $f(1) = f(0) = 0$ so we are left with

$$\begin{aligned}0 = \langle f, (D + D^\dagger)g\rangle &= \int_0^1 x^2(1-x^2)p(D + D^\dagger)g\,dx \\ &= \langle x^2(1-x)^2(D + D^\dagger)g, p\rangle.\end{aligned}$$

Since this is true for every $p \in V$, it follows that $x^2(1-x)^2(D + D^\dagger)g = 0$. But $x^2(1-x)^2 > 0$ except at the endpoints, and hence we must have $(D + D^\dagger)g = 0$ for all $g \in V$, and thus $D + D^\dagger = 0$. However, the above general result then yields

$$0 = \langle f, (D + D^\dagger)g\rangle = (fg)(1) - (fg)(0)$$

which is certainly not true for every $f, g \in V$. Hence $D^\dagger$ must not exist.

We leave it to the reader to find where the infinite-dimensionality of $V = \mathbb{R}[x]$ enters into this example.

While this example shows that not every operator on an infinite-dimensional space has an adjoint, there are in fact some operators on some infinite-dimensional spaces that do indeed have an adjoint. A particular example of this is given in Exercise 6.2.3. In fact, the famous Riesz representation theorem asserts that any continuous linear functional on a Hilbert space does indeed have an adjoint. While this fact should be well known to anyone who has studied quantum mechanics, its development is beyond the scope of this text.

An operator $T \in L(V)$ is said to be **Hermitian** (or **self-adjoint**) if $T^\dagger = T$. The elementary properties of the adjoint operator $T^\dagger$ are given in the following theorem. Note that if V is a real vector space, then the properties of the matrix representing an adjoint operator simply reduce to those of the transpose. Hence, a real Hermitian operator is represented by a (real) symmetric matrix.

Theorem 6.3. *Suppose $S, T \in L(V)$ and $c \in \mathbb{C}$. Then*
(i) $(S + T)^\dagger = S^\dagger + T^\dagger$.
(ii) $(cT)^\dagger = c^* T^\dagger$.
(iii) $(ST)^\dagger = T^\dagger S^\dagger$.

(iv) $T^{\dagger\dagger} = (T^\dagger)^\dagger = T$.
(v) $I^\dagger = I$ *and* $0^\dagger = 0$.
(vi) $(T^\dagger)^{-1} = (T^{-1})^\dagger$.

Proof. Let $u, v \in V$ be arbitrary. Then, from the definitions, we have

(i) $$\begin{aligned}\langle (S+T)^\dagger u, v\rangle &= \langle u, (S+T)v\rangle = \langle u, Sv + Tv\rangle = \langle u, Sv\rangle + \langle u, Tv\rangle \\ &= \langle S^\dagger u, v\rangle + \langle T^\dagger u, v\rangle = \langle (S^\dagger + T^\dagger)u, v\rangle\ .\end{aligned}$$

(ii) $\langle (cT)^\dagger u, v\rangle = \langle u, cTv\rangle = c\langle u, Tv\rangle = c\langle T^\dagger u, v\rangle = \langle c^* T^\dagger u, v\rangle$.

(iii) $$\begin{aligned}\langle (ST)^\dagger u, v\rangle &= \langle u, (ST)v\rangle = \langle u, S(Tv)\rangle = \langle S^\dagger u, Tv\rangle \\ &= \langle T^\dagger (S^\dagger u), v\rangle = \langle (T^\dagger S^\dagger)u, v\rangle\ .\end{aligned}$$

(iv) This was shown in the discussion preceding Theorem 6.2.

(v) $\langle Iu, v\rangle = \langle u, v\rangle = \langle u, Iv\rangle = \langle I^\dagger u, v\rangle$.
$\langle 0u, v\rangle = \langle 0, v\rangle = 0 = \langle u, 0v\rangle = \langle 0^\dagger u, v\rangle$.

(vi) $I = I^\dagger = (TT^{-1})^\dagger = (T^{-1})^\dagger T^\dagger$ so that $(T^{-1})^\dagger = (T^\dagger)^{-1}$.

The proof is completed by noting that the adjoint and inverse operators are unique. ∎

Corollary. *If $T \in L(V)$ is nonsingular, then so is $T^\dagger$.*

Proof. This follows from Theorems 6.3(vi) and 4.8. ∎

We now group together several other useful properties of operators for easy reference.

Theorem 6.4. *(i) Let V be an inner product space over either $\mathbb{R}$ or $\mathbb{C}$, let $T \in L(V)$, and suppose $\langle u, Tv\rangle = 0$ for all $u, v \in V$. Then $T = 0$.*

(ii) Let V be an inner product space over $\mathbb{C}$, let $T \in L(V)$, and suppose $\langle u, Tu\rangle = 0$ for all $u \in V$. Then $T = 0$.

(iii) Let V be a real inner product space, let $T \in L(V)$ be Hermitian, and suppose $\langle u, Tu\rangle = 0$ for all $u \in V$. Then $T = 0$.

Proof. (i) Let $u = Tv$. Then, by definition of the inner product, we see that $\langle Tv, Tv\rangle = 0$ implies $Tv = 0$ for all $v \in V$ which implies $T = 0$.

(ii) For any $u, v \in V$ we have (by hypothesis)

$$\begin{aligned}0 &= \langle u+v, T(u+v)\rangle \\ &= \langle u, Tu\rangle + \langle u, Tv\rangle + \langle v, Tu\rangle + \langle v, Tv\rangle\end{aligned}$$

$$\begin{aligned} &= 0 + \langle u, Tv\rangle + \langle v, Tu\rangle + 0 \\ &= \langle u, Tv\rangle + \langle v, Tu\rangle \qquad (*) \end{aligned}$$

Since v is arbitrary, we may replace it with iv to obtain

$$0 = i\langle u, Tv\rangle - i\langle v, Tu\rangle.$$

Dividing this by i and adding to (*) results in $0 = \langle u, Tv\rangle$ for any $u, v \in V$. By (i) this implies $T = 0$.

(iii) For any $u, v \in V$ we have $\langle u + v, T(u + v)\rangle = 0$ which also yields (*). Therefore, using (*), the fact that $T^\dagger = T$, and the fact that V is real, we obtain

$$\begin{aligned} 0 &= \langle T^\dagger u, v\rangle + \langle v, Tu\rangle = \langle Tu, v\rangle + \langle v, Tu\rangle = \langle v, Tu\rangle + \langle v, Tu\rangle \\ &= 2\langle v, Tu\rangle. \end{aligned}$$

Since this holds for any $u, v \in V$ we have $T = 0$ by (i). (Note that in this particular case, $T^\dagger = T^T$.) ∎

Exercises

1. Suppose $S, T \in L(V)$.
   (a) If S and T are Hermitian, show ST and TS are Hermitian if and only if $[S, T] = ST - TS = 0$.
   (b) If T is Hermitian, show $S^\dagger TS$ is Hermitian for all S.
   (c) If S is nonsingular and $S^\dagger TS$ is Hermitian, show T is Hermitian.

2. Consider $V = M_n(\mathbb{C})$ with the inner product $\langle A, B\rangle = \operatorname{tr}(B^\dagger A)$. For each $M \in V$, define the operator $T_M \in L(V)$ by $T_M(A) = MA$. Show $(T_M)^\dagger = T_{M^\dagger}$.

3. Consider the space $V = \mathbb{C}[x]$. If $f = \sum a_i x^i \in V$, we define the complex conjugate of f to be the polynomial $f^* = \sum a_i^* x^i \in V$. In other words, if $t \in \mathbb{R}$, then $f^*(t) = (f(t))^*$. We define an inner product on V by

   $$\langle f, g\rangle = \int_0^1 f(t)^* g(t)\, dt.$$

   For each $f \in V$, define the operator $T_f \in L(V)$ by $T_f(g) = fg$. Show $(T_f)^\dagger = T_{f^*}$.

4. Let V be the space of all real polynomials of degree ≤ 3, and define an inner product on V by

   $$\langle f, g\rangle = \int_0^1 f(x)g(x)\, dx.$$

   For any $t \in \mathbb{R}$, find a polynomial $h_t \in V$ such that $\langle h_t, f\rangle = f(t)$ for all $f \in V$.

5. If V is as in the previous exercise and D is the usual differentiation operator on V, find $D^\dagger$.

6. Let $V = \mathbb{C}^2$ with the standard inner product.
   (a) Define $T \in L(V)$ by $Te_1 = (1, -2)$, $Te_2 = (i, -1)$. If $v = (z_1, z_2)$, find $T^\dagger v$.
   (b) Define $T \in L(V)$ by $Te_1 = (1 + i, 2)$, $Te_2 = (i, i)$. Find the matrix representation of $T^\dagger$ relative to the usual basis for V. Is it true that $[T, T^\dagger] = 0$?

7. Let V be a finite-dimensional inner product space and suppose $T \in L(V)$. Show $\operatorname{Im} T^\dagger = (\operatorname{Ker} T)^\perp$.

8. Let V be a finite-dimensional inner product space, and suppose $E \in L(V)$ is idempotent, i.e., $E^2 = E$. Prove $E^\dagger = E$ if and only if $[E, E^\dagger] = 0$.

9. For each of the following inner product spaces V and $L \in V^*$, find a vector $u \in V$ such that $Lv = \langle u, v\rangle$ for all $v \in V$:
   (a) $V = \mathbb{R}^3$ and $L(x, y, z) = x - 2y + 4z$.
   (b) $V = \mathbb{C}^2$ and $L(z_1, z_2) = z_1 - z_2$.
   (c) V is the space of all real polynomials of degree ≤ 2 with inner product as in Exercise 6.2.4, and $Lf = f(0) + Df(1)$. (Here D is the usual differentiation operator.)

10. (a) Let $V = \mathbb{R}^2$ and define $T \in L(V)$ by $T(x, y) = (2x + y, x - 3y)$. Find $T^\dagger(3, 5)$.
   (b) Let $V = \mathbb{C}^2$ and define $T \in L(V)$ by $T(z_1, z_2) = (2z_1 + iz_2, (1 - i)z_1)$. Find $T^\dagger(3 - i, 1 + i2)$.
   (c) Let V be as in Exercise 6.2.9(c), and define $T \in L(V)$ by $Tf = 3f + Df$. Find $T^\dagger f$ where $f = 3x^2 - x + 4$.

6.2 Normal Operators

Let V be a complex inner product space with the induced norm. Another important class of operators $U \in L(V)$ is that for which $\|Uv\| = \|v\|$ for all $v \in V$. Such operators are called **isometric** because they preserve the length of the vector v. Furthermore, for any $v, w \in V$ we see that

$$\|Uv - Uw\| = \|U(v - w)\| = \|v - w\|$$

so that U preserves distances as well. This is sometimes described by saying that U is an **isometry**. (Note that in Section 4.5 when we discussed isometries, there was no requirement that the function be linear.)

If we write out the norm as an inner product and assume that the adjoint operator exists, we see that an isometric operator satisfies

$$\langle v, v\rangle = \langle Uv, Uv\rangle = \langle v, (U^\dagger U)v\rangle$$

and hence $\langle v, (U^\dagger U - 1)v\rangle = 0$ for any $v \in V$. But then from Theorem 6.4(ii)) it follows that

$$U^\dagger U = 1.$$

In fact, this is sometimes taken as the definition of an isometric operator. Note this applies equally well to an infinite-dimensional space.

If V is finite-dimensional, then (by Theorems 2.20 and 4.11) it follows that $U^\dagger = U^{-1}$, and hence

$$U^\dagger U = UU^\dagger = 1.$$

Any operator that satisfies either $U^\dagger U = UU^\dagger = 1$ or $U^\dagger = U^{-1}$ is said to be **unitary**. It is clear that a unitary operator is necessarily isometric. If V is simply a real space, then unitary operators are called **orthogonal**.

Because of the importance of isometric and unitary operators in both mathematics and physics, it is worth arriving at both of these definitions from a slightly different viewpoint that also aids in our understanding of these operators.

Let V be a complex vector space with an inner product defined on it. We say that an operator U is **unitary** if $\|Uv\| = \|v\|$ for all $v \in V$ and, in addition, it is a mapping of V *onto* itself. Since $\|Uv\| = \|v\|$, we see that $Uv = 0$ if and only if $v = 0$, and hence $\operatorname{Ker} U = \{0\}$. Therefore U is one-to-one and U^{-1} exists (Theorem 4.5). Since U is surjective, the inverse is defined on all of V also. Note there has been no mention of finite-dimensionality. This was avoided by requiring that the mapping be surjective.

Starting from $\|Uv\| = \|v\|$, we may write $\langle v, (U^\dagger U)v\rangle = \langle v, v\rangle$. As we did in the proof of Theorem 6.4, if we first substitute $v = v_1 + v_2$ and then $v = v_1 + iv_2$, divide the second of these equations by i, add to the first, and use $\langle v_i, (U^\dagger U)v_i\rangle = \langle v_i, v_i\rangle$, we find $\langle v_1, (U^\dagger U)v_2\rangle = \langle v_1, v_2\rangle$ or $\langle v_1, (U^\dagger U - 1)v_2\rangle = 0$. Since this holds for all $v_1, v_2 \in V$, it follows from Theorem 6.4(i) that $U^\dagger U = 1$. If we now multiply this equation from the left by U we have $UU^\dagger U = U$, and hence $(UU^\dagger)(Uv) = Uv$ for all $v \in V$. But as v varies over all of V, so does Uv since U is surjective. We then define $v' = Uv$ so that $(UU^\dagger)v' = v'$ for all $v' \in V$. This shows $U^\dagger U = 1$ implies $UU^\dagger = 1$. What we have just done then, is show that a surjective norm-preserving operator U has the property that $U^\dagger U = UU^\dagger = 1$. It is important to emphasize that this approach is equally valid in infinite-dimensional spaces.

We now define an **isometric** operator Ω to be an operator defined on all of V with the property that $\|\Omega v\| = \|v\|$ for all $v \in V$. This differs from a unitary operator in that we do not require that Ω also be surjective. Again, the requirement that Ω preserve the norm tells us that Ω has an inverse (since it must be one-to-one), but this inverse is not necessarily defined on the whole of V. For example, let $\{e_i\}$ be an orthonormal basis for V, and define the "shift operator" Ω by

$$\Omega(e_i) = e_{i+1}.$$

This Ω is clearly defined on all of V, but the image of Ω is not all of V since it does not include the vector e_1. Therefore Ω^{-1} is not defined on e_1.

Exactly as we did for unitary operators, we can show $\Omega^\dagger\Omega = 1$ for an isometric operator Ω. If V happens to be finite-dimensional, then obviously $\Omega\Omega^\dagger = 1$. Thus, on a finite-dimensional space, an isometric operator is also unitary.

Finally, let us show an interesting relationship between the inverse Ω^{-1} of an isometric operator and its adjoint $\Omega^\dagger$. From $\Omega^\dagger\Omega = 1$, we may write $\Omega^\dagger(\Omega v) = v$ for every $v \in V$. If we define $\Omega v = v'$, then for every $v' \in \operatorname{Im}\Omega$ we have $v = \Omega^{-1}v'$, and hence

$$\Omega^\dagger v' = \Omega^{-1}v' \quad \text{for } v' \in \operatorname{Im}\Omega.$$

On the other hand, if $w' \in (\operatorname{Im}\Omega)^\perp$, then automatically $\langle w', \Omega v\rangle = 0$ for every $v \in V$. Therefore this may be written as $\langle \Omega^\dagger w', v\rangle = 0$ for every $v \in V$, and hence (choose $v = \Omega^\dagger w'$)

$$\Omega^\dagger w' = 0 \quad \text{for } w' \in (\operatorname{Im}\Omega)^\perp.$$

In other words, we have

$$\Omega^\dagger = \begin{cases} \Omega^{-1} & \text{on } \operatorname{Im}\Omega \\ 0 & \text{on } (\operatorname{Im}\Omega)^\perp \end{cases}.$$

For instance, using our earlier example of the shift operator, we see that $\langle e_1, e_i\rangle = 0$ for $i \neq 1$, and hence $e_1 \in (\operatorname{Im}\Omega)^\perp$. Therefore $\Omega^\dagger(e_1) = 0$ so we clearly can not have $\Omega\Omega^\dagger = 1$.

Our next theorem contains the operator versions of what was shown for matrices in Section 5.7.

Theorem 6.5. *Let V be a complex finite-dimensional inner product space. Then the following conditions on an operator $U \in L(V)$ are equivalent:*
(i) $U^\dagger = U^{-1}$.
(ii) $\langle Uv, Uw\rangle = \langle v, w\rangle$ for all $v, w \in V$.
(iii) $\|Uv\| = \|v\|$.

Proof. (i) $\Rightarrow$ (ii): $\langle Uv, Uw\rangle = \langle v, (U^\dagger U)w\rangle = \langle v, Iw\rangle = \langle v, w\rangle$.
(ii) $\Rightarrow$ (iii): $\|Uv\| = \langle Uv, Uv\rangle^{1/2} = \langle v, v\rangle^{1/2} = \|v\|$.
(iii) $\Rightarrow$ (i): $\langle v, (U^\dagger U)v\rangle = \langle Uv, Uv\rangle = \langle v, v\rangle = \langle v, Iv\rangle$, and therefore $\langle v, (U^\dagger U - I)v\rangle = 0$. Hence (by Theorem 6.4(ii)) we must have $U^\dagger U = I$, and therefore $U^\dagger = U^{-1}$ (since V is finite-dimensional). ∎

From part (iii) of this theorem we see U preserves the length of any vector. In particular, U preserves the length of a unit vector, hence the designation "unitary." Note also that if v and w are orthogonal, then $\langle v, w\rangle = 0$ and hence $\langle Uv, Uw\rangle = \langle v, w\rangle = 0$. Thus U maintains orthogonality as well.

Condition (ii) of this theorem is sometimes described by saying that a unitary transformation **preserves inner products**. In general, we say that a linear transformation (i.e., a vector space homomorphism) T of an inner product space V *onto* an inner product space W (over the same field) is an **inner product space isomorphism** of V onto W if it also preserves inner products. Therefore, one may define a unitary operator as an inner product space isomorphism.

It is also worth commenting on the case of unitary operators defined on a real vector space. Since in this case the adjoint reduces to the transpose, we have $U^\dagger = U^T = U^{-1}$. If V is a real vector space, then an operator $T = L(V)$ that satisfies $T^T = T^{-1}$ is said to be an **orthogonal** transformation. It should be clear that Theorem 6.5 also applies to real vector spaces if we replace the adjoint by the transpose.

From Theorem 6.2 we see that a complex matrix A represents a unitary operator relative to an orthonormal basis if and only if $A^\dagger = A^{-1}$. We therefore say that a complex matrix A is a **unitary matrix** if $A^\dagger = A^{-1}$. In the special case that A is a real matrix with the property that $A^T = A^{-1}$, then we say A is an **orthogonal matrix**. These definitions agree with what was discussed in Section 5.7.

Example 6.3. Suppose $V = \mathbb{R}^n$ and $X \in V$. In terms of an orthonormal basis $\{e_i\}$ for V we may write $X = \sum_i x^i e_i$. Now suppose we are given another orthonormal basis $\{\bar{e}_i\}$ related to the first basis by $\bar{e}_i = A(e_i) = \sum_j e_j a_{ji}$ for some real matrix (a_{ij}). Relative to this new basis we have $A(X) = \overline{X} = \sum_i \bar{x}^i \bar{e}_i$ where $x^i = \sum_j a_{ij}\bar{x}^j$ (see Section 4.4). Then

$$\begin{aligned}
\|X\|^2 &= \left\langle \sum_i x^i e_i, \sum_j x^j e_j \right\rangle = \sum_{i,j} x^i x^j \langle e_i, e_j\rangle = \sum_{i,j} x^i x^j \delta_{ij} \\
&= \sum_i (x^i)^2 = \sum_{i,j,k} a_{ij} a_{ik} \bar{x}^j \bar{x}^k = \sum_{i,j,k} a^T_{ji} a_{ik} \bar{x}^j \bar{x}^k \\
&= \sum_{j,k} (A^T A)_{jk} \bar{x}^j \bar{x}^k .
\end{aligned}$$

If A is orthogonal, then $A^T = A^{-1}$ so that $(A^T A)_{jk} = \delta_{jk}$ and we are left with

$$\|X\|^2 = \sum_i (x^i)^2 = \sum_j (\bar{x}^j)^2 = \left\|\overline{X}\right\|^2$$

so the length of X is unchanged under an orthogonal transformation. An equivalent way to see this is to assume that A simply represents a rotation so the length of a vector remains unchanged by definition. This then forces A to be an orthogonal transformation (see Exercise 6.2.1).

Another way to think of orthogonal transformations is the following. We saw in Section 1.5 that the angle θ between two vectors $X, Y \in \mathbb{R}^n$ is defined by

$$\cos\theta = \frac{\langle X, Y\rangle}{\|X\|\,\|Y\|}.$$

Under the orthogonal transformation A, we then have $\overline{X} = A(X)$ and also

$$\cos\bar{\theta} = \frac{\langle \overline{X}, \overline{Y} \rangle}{\|\overline{X}\| \, \|\overline{Y}\|}.$$

But $\|\overline{X}\| = \|X\|$ and $\|\overline{Y}\| = \|Y\|$, and in addition,

$$\begin{aligned}\langle X, Y \rangle &= \Big\langle \sum_i x^i e_i, \sum_j y^j e_j \Big\rangle = \sum_i x^i y^i = \sum_{i,j,k} a_{ij}\bar{x}^j a_{ik}\bar{y}^k \\ &= \sum_{j,k} \delta_{jk}\bar{x}^j \bar{y}^j = \sum_j \bar{x}^j \bar{y}^j \\ &= \langle \overline{X}, \overline{Y} \rangle\end{aligned}$$

so that $\theta = \bar{\theta}$ (this also follows from the real vector space version of Theorem 6.5). Therefore an orthogonal transformation also preserves the angle between two vectors, and hence is nothing more than a rotation in $\mathbb{R}^n$.

Since a unitary operator is an inner product space isomorphism, our next theorem should come as no surprise.

Theorem 6.6. *Let V be finite-dimensional over $\mathbb{C}$ (resp. $\mathbb{R}$). A linear transformation $U \in L(V)$ is unitary (resp. orthogonal) if and only if it takes an orthonormal basis for V into an orthonormal basis for V.*

Proof. We consider the case where V is complex, leaving the real case to the reader. Let $\{e_i\}$ be an orthonormal basis for V, and assume U is unitary. Then from Theorem 6.5(ii) we have

$$\langle Ue_i, Ue_j \rangle = \langle e_i, e_j \rangle = \delta_{ij}$$

so that $\{Ue_i\}$ is also an orthonormal set. But any orthonormal set is linearly independent (Theorem 1.19), and hence $\{Ue_i\}$ forms a basis for V (since there are as many of the Ue_i as there are e_i).

Conversely, suppose both $\{e_i\}$ and $\{Ue_i\}$ are orthonormal bases for V and let $v, w \in V$ be arbitrary. Then

$$\begin{aligned}\langle v, w \rangle &= \Big\langle \sum_i v^i e_i, \sum_j w^j e_j \Big\rangle = \sum_{i,j} v^{i*} w^j \langle e_i, e_j \rangle = \sum_{i,j} v^{i*} w^j \delta_{ij} \\ &= \sum_i v^{i*} w^i.\end{aligned}$$

However, we also have

$$\langle Uv, Uw \rangle = \Big\langle U\Big(\sum_i v^i e_i\Big), U\Big(\sum_j w^j e_j\Big)\Big\rangle = \sum_{i,j} v^{i*} w^j \langle Ue_i, Ue_j \rangle$$

$$= \sum_{i,j} v^{i*} w^j \delta_{ij} = \sum_i v^{i*} w^i = \langle v, w \rangle.$$

This shows that U is unitary (Theorem 6.5). ∎

Recall also that Ue_i is the ith column of $[U]$. Therefore, if $\{Ue_i\}$ is an orthonormal set, then the columns of $[U]$ are orthonormal and hence (by Theorem 5.15) $[U]$ must be a unitary matrix.

Corollary. *Let V and W be finite-dimensional inner product spaces over $\mathbb{C}$. Then there exists an inner product space isomorphism of V onto W if and only if* $\dim V = \dim W$.

Proof. Clearly $\dim V = \dim W$ if V and W are isomorphic. On the other hand, let $\{e_1, \ldots, e_n\}$ be an orthonormal basis for V, and let $\{\bar{e}_1, \ldots, \bar{e}_n\}$ be an orthonormal basis for W. (These bases exist by Theorem 1.21.) We define the (surjective) linear transformation U by the requirement $Ue_i = \bar{e}_i$. U is unique by Theorem 4.1. Since $\langle Ue_i, Ue_j \rangle = \langle \bar{e}_i, \bar{e}_j \rangle = \delta_{ij} = \langle e_i, e_j \rangle$, the proof of Theorem 6.6 shows that U preserves inner products. In particular, we see that $\|Uv\| = \|v\|$ for every $v \in V$, and hence $\operatorname{Ker} U = \{0\}$ (by property (N1) of Theorem 1.17). Thus U is also one-to-one (Theorem 4.5). ∎

Let us now take a look at some rather basic properties of the eigenvalues and eigenvectors of the operators we have been discussing. To simplify our terminology, we remark that a complex inner product space is also called a **unitary space**, while a real inner product space is sometimes called a **Euclidean space**. If H is an operator such that $H^\dagger = -H$, then H is said to be **anti-Hermitian** (or **skew-Hermitian**). Furthermore, if P is an operator such that $P = S^\dagger S$ for some operator S, then we say that P is **positive** (or **positive semidefinite** or **nonnegative**). If S also happens to be nonsingular (and hence P is also nonsingular), then we say that P is **positive definite**. Note that a positive operator is necessarily Hermitian since $(S^\dagger S)^\dagger = S^\dagger S$. The reason P is called positive is shown in part (iv) of the following theorem.

Theorem 6.7. *(i) The eigenvalues of a Hermitian operator are real.*

(ii) The eigenvalues of an isometry (and hence also of a unitary transformation) have absolute value one.

(iii) The eigenvalues of an anti-Hermitian operator are pure imaginary.

(iv) A positive (positive definite) operator has eigenvalues that are real and nonnegative (positive).

Proof. (i) If H is Hermitian, $v \neq 0$, and $Hv = \lambda v$, we have

$$\begin{aligned} \lambda \langle v, v \rangle &= \langle v, \lambda v \rangle = \langle v, Hv \rangle = \langle H^\dagger v, v \rangle = \langle Hv, v \rangle \\ &= \langle \lambda v, v \rangle = \lambda^* \langle v, v \rangle. \end{aligned}$$

But $\langle v, v\rangle \neq 0$, and hence $\lambda = \lambda^*$.

(ii) If Ω is an isometry, $v \neq 0$, and $\Omega v = \lambda v$, then we have (using Theorem 1.17)

$$\|v\| = \|\Omega v\| = \|\lambda v\| = |\lambda|\,\|v\|\,.$$

But $v \neq 0$, and hence $|\lambda| = 1$.

(iii) If $H^\dagger = -H$, $v \neq 0$, and $Hv = \lambda v$, then

$$\begin{aligned}\lambda\langle v, v\rangle &= \langle v, \lambda v\rangle = \langle v, Hv\rangle = \langle H^\dagger v, v\rangle = \langle -Hv, v\rangle = \langle -\lambda v, v\rangle \\ &= -\lambda^*\langle v, v\rangle.\end{aligned}$$

But $\langle v, v\rangle \neq 0$, and hence $\lambda = -\lambda^*$. This shows that λ is pure imaginary.

(iv) Let $P = S^\dagger S$ be a positive definite operator. If $v \neq 0$, then the fact that S is nonsingular means $Sv \neq 0$, and hence $\langle Sv, Sv\rangle = \|Sv\|^2 > 0$. Then, for $Pv = (S^\dagger S)v = \lambda v$, we see that

$$\lambda\langle v, v\rangle = \langle v, \lambda v\rangle = \langle v, Pv\rangle = \langle v, (S^\dagger S)v\rangle = \langle Sv, Sv\rangle.$$

But $\langle v, v\rangle = \|v\|^2 > 0$ also, and therefore we must have $\lambda > 0$.

If P is positive, then S is singular and the only difference is that now for $v \neq 0$ we have $\langle Sv, Sv\rangle = \|Sv\|^2 \geq 0$ which implies that $\lambda \geq 0$. ∎

We say that an operator N is **normal** if $N^\dagger N = NN^\dagger$. Note this implies that for any $v \in V$ we have

$$\begin{aligned}\|Nv\|^2 &= \langle Nv, Nv\rangle = \langle (N^\dagger N)v, v\rangle = \langle (NN^\dagger)v, v\rangle = \langle N^\dagger v, N^\dagger v\rangle \\ &= \left\|N^\dagger v\right\|^2.\end{aligned}$$

Now let λ be a complex number. It is easy to see that if N is normal then so is $N - \lambda 1$ since (from Theorem 6.3)

$$\begin{aligned}(N-\lambda 1)^\dagger(N-\lambda 1) &= (N^\dagger - \lambda^* 1)(N - \lambda 1) = N^\dagger N - \lambda N^\dagger - \lambda^* N + \lambda^*\lambda 1 \\ &= (N - \lambda 1)(N^\dagger - \lambda^* 1) = (N - \lambda 1)(N - \lambda 1)^\dagger.\end{aligned}$$

Using $N - \lambda 1$ instead of N in the previous result we obtain

$$\|Nv - \lambda v\|^2 = \left\|N^\dagger v - \lambda^* v\right\|^2.$$

Since the norm is positive definite, this equation proves the next theorem.

Theorem 6.8. *Let N be a normal operator and let λ be an eigenvalue of N. Then $Nv = \lambda v$ if and only if $N^\dagger v = \lambda^* v$.*

In words, if v is an eigenvector of a normal operator N with eigenvalue λ, then v is also an eigenvector of $N^\dagger$ with eigenvalue λ^*. (Note it is always true that if λ is an eigenvalue of an operator T, then λ^* will be an eigenvalue of $T^\dagger$. However, the eigenvectors will in general be different. See Exercise 6.2.9.)

Corollary. *If N is normal and $Nv = 0$ for some $v \in V$, then $N^\dagger v = 0$.*

Proof. This follows from Theorem 6.8 by taking $\lambda = \lambda^* = 0$. Alternatively, using $N^\dagger N = NN^\dagger$ along with the fact that $Nv = 0$, we see that

$$\langle N^\dagger v, N^\dagger v\rangle = \langle v, (NN^\dagger)v\rangle = \langle v, (N^\dagger N)v\rangle = 0.$$

Since the inner product is positive definite, this requires that $N^\dagger v = 0$. ∎

As was the case with matrices, if $H^\dagger = H$, then $H^\dagger H = HH = HH^\dagger$ so that any Hermitian operator is normal. Furthermore, if U is unitary, then $U^\dagger U = UU^\dagger$ $(= 1)$ so that U is also normal. A Hermitian operator T defined on a *real* inner product space is said to be **symmetric**. This is equivalent to requiring that with respect to an orthonormal basis, the matrix elements a_{ij} of T are given by

$$a_{ij} = \langle e_i, Te_j\rangle = \langle Te_i, e_j\rangle = \langle e_j, Te_i\rangle = a_{ji}.$$

Therefore a symmetric operator is represented by a real symmetric matrix. It is also true that **antisymmetric** operators (i.e., $T^T = -T$) and anti-Hermitian operators ($H^\dagger = -H$) are normal.

Theorem 6.9. *(i) Eigenvectors belonging to* distinct *eigenvalues of a Hermitian operator are orthogonal.*

(ii) Eigenvectors belonging to distinct *eigenvalues of an isometric operator are orthogonal. Hence the eigenvectors belonging to distinct eigenvalues of a unitary operator are orthogonal.*

(iii) Eigenvectors belonging to distinct *eigenvalues of a normal operator are orthogonal.*

Proof. As we remarked above, Hermitian and unitary operators are special cases of normal operators. Thus parts (i) and (ii) follow from part (iii). However, it is instructive to give independent proofs of parts (i) and (ii).

Assume T is an operator on a unitary space, and $Tv_i = \lambda_i v_i$ for $i = 1, 2$ with $\lambda_1 \neq \lambda_2$. We may then also assume without loss of generality that $\lambda_1 \neq 0$.

(i) If $T = T^\dagger$, then (using Theorem 6.7(i))

$$\begin{aligned}\lambda_2\langle v_1, v_2\rangle &= \langle v_1, \lambda_2 v_2\rangle = \langle v_1, Tv_2\rangle = \langle T^\dagger v_1, v_2\rangle = \langle Tv_1, v_2\rangle \\ &= \langle \lambda_1 v_1, v_2\rangle = \lambda_1^*\langle v_1, v_2\rangle = \lambda_1\langle v_1, v_2\rangle.\end{aligned}$$

But $\lambda_1 \neq \lambda_2$, and hence $\langle v_1, v_2\rangle = 0$.

(ii) If T is isometric, then $T^\dagger T = 1$ and we have

$$\langle v_1, v_2\rangle = \langle v_1, (T^\dagger T)v_2\rangle = \langle Tv_1, Tv_2\rangle = \lambda_1^*\lambda_2\langle v_1, v_2\rangle.$$

But by Theorem 6.7(ii) we have $|\lambda_1|^2 = \lambda_1^*\lambda_1 = 1$, and thus $\lambda_1^* = 1/\lambda_1$. Therefore, multiplying the above equation by λ_1, we see that $\lambda_1\langle v_1, v_2\rangle = \lambda_2\langle v_1, v_2\rangle$ and hence, since $\lambda_1 \neq \lambda_2$, this shows that $\langle v_1, v_2\rangle = 0$.

(iii) If T is normal, then

$$\langle v_1, Tv_2\rangle = \lambda_2\langle v_1, v_2\rangle$$

while on the other hand, using Theorem 6.8 we have

$$\langle v_1, Tv_2\rangle = \langle T^\dagger v_1, v_2\rangle = \langle \lambda_1^* v_1, v_2\rangle = \lambda_1\langle v_1, v_2\rangle.$$

Therefore $\langle v_1, v_2\rangle = 0$ since $\lambda_1 \neq \lambda_2$. ∎

Theorem 6.10. *(i) Let T be an operator on a unitary space V, and let W be a T-invariant subspace of V. Then $W^\perp$ is invariant under $T^\dagger$.*

(ii) Let U be a unitary operator on a unitary space V, and let W be a U-invariant subspace of V. Then $W^\perp$ is also invariant under U.

Proof. (i) For any $v \in W$ we have $Tv \in W$ since W is T-invariant. Let $w \in W^\perp$ be arbitrary. We must show that $T^\dagger w \in W^\perp$. But this is easy because

$$\langle T^\dagger w, v\rangle = \langle w, Tv\rangle = 0$$

by definition of $W^\perp$. Thus $T^\dagger w \in W^\perp$ so that $W^\perp$ is invariant under $T^\dagger$.

(ii) The fact that U is unitary means $U^{-1} = U^\dagger$ exists, and hence U is nonsingular (so $\operatorname{Ker} U = \{0\}$). In other words, for any $v' \in W$ there exists $v \in W$ such that $Uv = v'$ (by Theorem 4.6). Now let $w \in W^\perp$ be arbitrary. Then

$$\langle Uw, v'\rangle = \langle Uw, Uv\rangle = \langle w, (U^\dagger U)v\rangle = \langle w, v\rangle = 0$$

by definition of $W^\perp$. Thus $Uw \in W^\perp$ so that $W^\perp$ is invariant under U. ∎

Repeating the proof of part (ii) of this theorem in the case of an orthogonal operator on a Euclidean space (i.e., a real inner product space), we obtain the following corollary.

Corollary. *Let T be an orthogonal operator on a finite-dimensional Euclidean space V, and let W be a T-invariant subspace of V. Then $W^\perp$ is also T-invariant.*

Recall from the discussion in Section 5.6 that the algebraic multiplicity of a given eigenvalue is the number of times the eigenvalue is repeated as a root of the characteristic polynomial. We also defined the geometric multiplicity as the number of linearly independent eigenvectors corresponding to this eigenvalue (i.e., the dimension of its eigenspace).

Theorem 6.11. *Let H be a Hermitian operator on a finite-dimensional unitary space V. Then the algebraic multiplicity of any eigenvalue λ of H is equal to its geometric multiplicity.*

Proof. Let $V_\lambda = \{v \in V : Hv = \lambda v\}$ be the eigenspace corresponding to the eigenvalue λ. Clearly V_λ is invariant under H since $Hv = \lambda v \in V_\lambda$ for every $v \in V_\lambda$. By Theorem 6.10(i), we then have that $V_\lambda^\perp$ is also invariant under $H^\dagger = H$. Furthermore, by Theorem 1.22 we see $V = V_\lambda \oplus V_\lambda^\perp$. Applying Theorem 5.12 we may write $H = H_1 \oplus H_2$ where $H_1 = H|V_\lambda$ and $H_2 = H|V_\lambda^\perp$.

Let A be the matrix representation of H, and let A_i be the matrix representation of H_i $(i = 1, 2)$. By Theorem 5.12, we also have $A = A_1 \oplus A_2$. Using Theorem 5.10, it then follows that the characteristic polynomial of A is given by

$$\det(xI - A) = \det(xI - A_1)\det(xI - A_2).$$

Now, H_1 is a Hermitian operator on the finite-dimensional space V_λ with only the single eigenvalue λ. Therefore λ is the only root of $\det(xI - A_1) = 0$, and hence it must occur with an algebraic multiplicity equal to the dimension of V_λ (since this is just the size of the matrix A_1). In other words, if $\dim V_\lambda = m$, then $\det(xI - A_1) = (x - \lambda)^m$. On the other hand, λ is not an eigenvalue of A_2 by definition, and hence $\det(\lambda I - A_2) \neq 0$. This means that $\det(xI - A)$ contains $(x - \lambda)$ as a factor exactly m times. ∎

Corollary. *Any Hermitian operator H on a finite-dimensional unitary space V is diagonalizable.*

Proof. Since V is a unitary space, the characteristic polynomial of H will factor into (not necessarily distinct) linear terms. The conclusion then follows from Theorems 6.11 and 5.14. ∎

In fact, from Theorem 5.18 we know that any normal matrix is unitarily similar to a diagonal matrix. This means that given any normal operator $T \in L(V)$, there is an orthonormal basis for V that consists of eigenvectors of T.

Exercises

1. Let $V = \mathbb{R}^n$ with the standard inner product, and suppose the length of any $X \in V$ remains unchanged under $A \in L(V)$. Show that A must be an orthogonal transformation.

2. Let V be the space of all continuous complex-valued functions defined on $[0, 2\pi]$, and define an inner product on V by

$$\langle f, g\rangle = \frac{1}{2\pi}\int_0^{2\pi} f(x)^* g(x)\, dx.$$

Suppose there exists $h \in V$ such that $|h(x)| = 1$ for all $x \in [0, 2\pi]$, and define $T_h \in L(V)$ by $T_h f = hf$. Prove that T is unitary.

3. Let W be a finite-dimensional subspace of an inner product space V, and recall that $V = W \oplus W^\perp$ (see Exercise 1.6.11). Define $U \in L(V)$ by

$$U(w_1 + w_2) = w_1 - w_2$$

where $w_1 \in W$ and $w_2 \in W^\perp$.

(a) Prove that U is a Hermitian operator.
(b) Let $V = \mathbb{R}^3$ have the standard inner product, and let $W \subset V$ be spanned by the vector $(1, 0, 1)$. Find the matrix of U relative to the standard basis for V.

4. Let V be a finite-dimensional inner product space. An operator $\Omega \in L(V)$ is said to be a **partial isometry** if there exists a subspace W of V such that $\|\Omega w\| = \|w\|$ for all $w \in W$, and $\|\Omega w\| = 0$ for all $w \in W^\perp$. Let Ω be a partial isometry and suppose $\{w_1, \dots, w_k\}$ is an orthonormal basis for W.

(a) Show $\langle \Omega u, \Omega v \rangle = \langle u, v \rangle$ for all $u, v \in W$. [*Hint*: Use Exercise 1.5.7.]
(b) Show $\{\Omega w_1, \dots, \Omega w_k\}$ is an orthonormal basis for $\operatorname{Im} \Omega$.
(c) Show there exists an orthonormal basis $\{v_i\}$ for V such that the first k columns of $[\Omega]_v$ form an orthonormal set, and the remaining columns are zero.
(d) Let $\{u_1, \dots, u_r\}$ be an orthonormal basis for $(\operatorname{Im} \Omega)^\perp$. Show that $\{\Omega w_1, \dots, \Omega w_k, u_1, \dots, u_r\}$ is an orthonormal basis for V.
(e) Suppose $T \in L(V)$ satisfies $T(\Omega w_i) = w_i$ (for $1 \le i \le k$) and $T u_i = 0$ (for $1 \le i \le r$). Show T is well-defined, and that $T = \Omega^\dagger$.
(f) Show $\Omega^\dagger$ is a partial isometry.

5. Let V be a complex inner product space, and suppose $H \in L(V)$ is Hermitian. Show:

(a) $\|v + iHv\| = \|v - iHv\|$ for all $v \in V$.
(b) $u + iHu = v + iHv$ if and only if $u = v$.
(c) $1 + iH$ and $1 - iH$ are nonsingular.
(d) If V is finite-dimensional, then $U = (1 - iH)(1 + iH)^{-1}$ is a unitary operator. (U is called the **Cayley transform** of H. This result is also true in an infinite-dimensional Hilbert space but the proof is considerably more difficult.)

6. Let V be a unitary space and suppose $T \in L(V)$. Define $T_+ = (1/2)(T + T^\dagger)$ and $T_- = (1/2i)(T - T^\dagger)$.

(a) Show T_+ and T_- are Hermitian, and that $T = T_+ + iT_-$.
(b) If T'_+ and T'_- are Hermitian operators such that $T = T'_+ + iT'_-$, show $T'_+ = T_+$ and $T'_- = T_-$.
(c) Prove T is normal if and only if $[T_+, T_-] = 0$.

7. Let V be a finite-dimensional inner product space, and suppose $T \in L(V)$ is both positive and unitary. Prove $T = 1$.

8. Let $H \in M_n(\mathbb{C})$ be Hermitian. Then for any nonzero $x \in \mathbb{C}^n$ we define the **Rayleigh quotient** to be the number

$$R(x) = \frac{\langle x, Hx\rangle}{\|x\|^2}.$$

Prove $\max\{R(x) : x \neq 0\}$ is the largest eigenvalue of H, and that $\min\{R(x) : x \neq 0\}$ is the smallest eigenvalue of H.

9. If V is finite-dimensional and $T \in L(V)$ has eigenvalue λ, show $T^\dagger$ has eigenvalue λ^*.

6.3 More on Orthogonal Transformations

In this section we will take a more detailed look at the structure of orthogonal transformations on a real inner product space V (i.e., a finite-dimensional Euclidean space). Our goal is to show that there is a basis for V such that these operators can be written as the direct sum of two-dimensional rotations plus at most a single reflection. This result will be used in Section 8.8 when we discuss oriented bases for vector spaces.

Let us first make some careful definitions. (The reader might want to first review Section 4.5.) A linear operator $T \in L(V)$ on a finite-dimensional inner product space V is called a **rotation** if T is either the identity on V or else there exists a two-dimensional subspace W of V with basis $\{e_1, e_2\}$ such that for some real number θ we have

$$\begin{aligned} T(e_1) &= (\cos\theta)e_1 + (\sin\theta)e_2 \\ T(e_2) &= (-\sin\theta)e_1 + (\cos\theta)e_2 \end{aligned}$$

and $T(v) = v$ for all $v \in W^\perp$. The operator T is said to be a **rotation of** W **about** $W^\perp$.

If there exists a one-dimensional subspace W of V such that $T(w) = -w$ for all $w \in W$ and $T(v) = v$ for all $v \in W^\perp$, then T is said to be a **reflection of** V **about** $W^\perp$. Since, as have seen earlier, orthogonal transformations preserve inner products, we have the following slight generalization of Theorem 4.17.

Theorem 6.12. *Let $T \in L(V)$ be orthogonal on a two-dimensional inner product space V. Then T is either a rotation or a reflection, and in fact is a rotation if and only if $\det T = +1$ and a reflection if and only if $\det T = -1$.*

Now let T be any linear operator on a finite-dimensional real space V, and let A be the matrix representation of T with respect to some basis for V. Since

V is real, there is no guarantee that A has an eigenvalue in $\mathbb{R}$, but we can use A to *define* another operator T' on $\mathbb{C}^n$ by $T'(x) = Ax$ for all $x \in \mathbb{C}^n$. Now T' *does* have an eigenvalue $\lambda \in \mathbb{C}$ and corresponding eigenvector $v \in \mathbb{C}^n$. Let us write both of these in terms of their real and imaginary parts as $\lambda = \lambda_{\text{re}} + i\lambda_{\text{im}}$ and $v = v_{\text{re}} + iv_{\text{im}}$ where $\lambda_{\text{re}}, \lambda_{\text{im}} \in \mathbb{R}$ and $v_{\text{re}}, v_{\text{im}} \in \mathbb{R}^n$.

The eigenvalue equation for T' is $T'(v) = Av = \lambda(v)$, and in terms of real and imaginary components this becomes

$$A(v_{\text{re}} + iv_{\text{im}}) = (\lambda_{\text{re}} + i\lambda_{\text{im}})(v_{\text{re}} + iv_{\text{im}})$$

or

$$Av_{\text{re}} + iAv_{\text{im}} = (\lambda_{\text{re}}v_{\text{re}} - \lambda_{\text{im}}v_{\text{im}}) + i(\lambda_{\text{re}}v_{\text{im}} + \lambda_{\text{im}}v_{\text{re}}).$$

Equating real and imaginary parts we have the equations

$$Av_{\text{re}} = \lambda_{\text{re}}v_{\text{re}} - \lambda_{\text{im}}v_{\text{im}}$$

and

$$Av_{\text{im}} = \lambda_{\text{re}}v_{\text{im}} + \lambda_{\text{im}}v_{\text{re}}$$

which shows that both Av_{re} and Av_{im} are in the *real* space spanned by $\{v_{\text{re}}, v_{\text{im}}\}$. Since eigenvectors are nonzero by definition, we can't have both v_{re} and v_{im} equal to zero, so we define the nonzero subspace $W \subset \mathbb{R}^n$ to be the linear span of v_{re} and v_{im}. Thus we have $1 \le \dim W \le 2$.

By construction, W is invariant under A and we have proved the following.

Theorem 6.13. *Let V be a finite-dimensional real vector space, and let $T \in L(V)$. Then there exists a T-invariant subspace W of V with $1 \le \dim W \le 2$.*

We now apply this result to the particular case of orthogonal operators, and show that V can be written as a direct sum of 1- or 2-dimensional T-invariant subspaces. In other words, we show that T is reducible, and the W_i form a T-invariant direct sum decomposition of V (see Section 5.5).

Theorem 6.14. *Let T be an orthogonal operator on a finite-dimensional Euclidean space V. Then $V = W_1 \oplus \cdots \oplus W_r$ where each W_i is T-invariant, $1 \le \dim W_i \le 2$ and vectors belonging to distinct subspaces W_i and W_j are orthogonal (i.e., the subspaces W_i are pairwise orthogonal).*

Proof. The proof is by induction on $\dim V$. If $\dim V = 1$ there is nothing to prove, so let $\dim V = n > 1$ and assume that the theorem holds for $\dim V < n$. By Theorem 6.13 there exists a T-invariant subspace W_1 such that $1 \le \dim W_1 \le 2$. If $W_1 = V$ then we are done. Otherwise, by Theorem 1.22 we know that $V = W_1 \oplus W_1^\perp$ where $W_1^\perp$ is nonzero. Furthermore, $W_1^\perp$ is also

T-invariant (by the corollary to Theorem 6.10), and the restriction $T_{W_1^\perp}$ of T to $W_1^\perp$ is also clearly orthogonal.

Since $\dim W_1^\perp < n$, we may apply our induction hypothesis to the operator $T_{W_1^\perp}$ to conclude that $W_1^\perp = W_2 \oplus \cdots \oplus W_r$ where each W_i is T-invariant, and the subspaces W_i are pairwise orthogonal. But then $V = W_1 \oplus W_1^\perp = W_1 \oplus \cdots \oplus W_r$. ∎

Now we can combine Theorems 6.12 and 6.14 for our main result.

Theorem 6.15. *Let T be an orthogonal operator on a finite-dimensional Euclidean space $V = W_1 \oplus \cdots \oplus W_r$ where the W_i are pairwise orthogonal T-invariant subspaces of V, and each W_i is of dimension either 1 or 2.*

(i) If $\det T = +1$, then there are an even number of W_i for which T_{W_i} is a reflection, and if $\det T = -1$, then there are an odd number of W_i for which T_{W_i} is a reflection.

(ii) The decomposition $V = W_1 \oplus \cdots \oplus W_r$ can be made so that if $\det T = +1$ then there are no W_i's such that T_{W_i} is a reflection, and if $\det T = -1$, then there is precisely one W_i such that T_{W_i} is a reflection. Furthermore, if T_{W_k} is a reflection, then $\dim W_k = 1$.

Proof. (i) Suppose that T_{W_i} is a reflection for m of the W_i's. By Theorem 5.12, the matrix representation of T is the direct sum of the matrix representations of T_{W_i}, and hence by Theorems 5.10 and 6.12 we have

$$\det T = (\det T_{W_1}) \cdots (\det T_{W_r}) = (-1)^m.$$

This proves part (i).

(ii) Let $\widetilde{W} = \{v \in V : T(v) = -v\}$. Then $\widetilde{W}$ is T-invariant as is $W := \widetilde{W}^\perp$. Note that $V = W \oplus \widetilde{W}$. Applying Theorem 6.14 to T_W, we may write $W = W_1 \oplus \cdots \oplus W_s$ where these W_i's are pairwise orthogonal and each has dimension 1 or 2. Because $\widetilde{W}$ contains all vectors in V that are reflected by T, and W is the orthogonal complement to $\widetilde{W}$, it follows that for each $i = 1, \ldots, s$ it must be that T_{W_i} is a rotation. Formally, we see that if $w \in W_i$ is such that $T(w) = -w$, then $w \in W_i \cap \widetilde{W} \subset W \cap \widetilde{W} = \{0\}$.

Next, we assume that $\widetilde{W} \neq 0$ or we are done. So, choose an orthonormal basis $\{e_j\}$ for $\widetilde{W}$. We write this basis as a pairwise disjoint union $\tilde{e}_1 \cup \cdots \cup \tilde{e}_q$ where (by definition) each $\tilde{e}_1, \ldots, \tilde{e}_{q-1}$ consists of exactly two of the e_j's, and $\tilde{e}_q$ also consists of two of the e_j's if $\dim \widetilde{W}$ is an even number, and only a single e_j if $\dim \widetilde{W}$ is an odd number. For each $j = 1, \ldots, q$ define W_{s+j} to be the span of $\tilde{e}_j$. Then

$$\begin{aligned} V &= W \oplus \widetilde{W} = W_1 \oplus \cdots \oplus W_s \oplus \widetilde{W} \\ &= W_1 \oplus \cdots \oplus W_s \oplus W_{s+1} \oplus \cdots \oplus W_{s+q} \end{aligned}$$

where all of the W_i's are pairwise orthogonal.

Since each $\tilde{e}_j$ consists of either 1 or 2 of the basis vectors e_i for $\widetilde{W}$ (so that $T(e_i) = -e_i$), we see that if $\tilde{e}_j$ consists of 2 basis vectors, then the matrix representation $[T_{W s+j}]_{\tilde{e}_j}$ is given by

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

and hence $\det T_{W s+j} = +1$. Then (by Theorem 6.12) T_{W_i} is a rotation for all $1 \leq i < s+q$. If $\tilde{e}_q$ consists of a single vector e_j, then $\dim W_{s+q} = 1$ and $\det T_{W_{s+q}} = -1$ so that (by Theorem 6.12 again) $T_{W_{s+q}}$ represents a reflection. Finally, letting $s+q = r$ we see that our decomposition satisfies the requirements of the theorem. ∎

It's not hard to see what is really going on here. Recall that a rotation in $\mathbb{R}^2$ looks like

$$\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}.$$

If we have more than a single reflection, then we can group them into pairs by choosing a rotation with $\theta = \pi$. This then leaves us with at most a single unpaired reflection.

As we saw in Section 5.5, the decomposition of V described in Theorem 6.15 shows that we may write the orthogonal operator T in the form

$$T = T_1 \oplus \cdots \oplus T_r$$

where $T_i = T_{W_i}$ and each T_i represents a rotation for all but at most one of the T_i's which would be a reflection.

Exercises

1. Define the matrices

$$A = \begin{bmatrix} \frac{1}{2} & \frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & -\frac{1}{2} \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

   (a) Show that A represents a reflection.
   (b) Find the axis in $\mathbb{R}^3$ about which A reflects.
   (c) Show that both AB and BA represent rotations.

2. Show that the composite of two rotations on $\mathbb{R}^3$ is another rotation on $\mathbb{R}^3$.

3. Consider the matrices

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{bmatrix} \qquad B = \begin{bmatrix} \cos\phi & -\sin\phi & 0 \\ \sin\phi & \cos\phi & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

   where $\theta, \phi \in \mathbb{R}$.

(a) Show that both A and B represent rotations.
(b) Show that the product AB is a rotation.
(c) Find the axis of rotation for AB.

4. Prove that it is impossible for an orthogonal operator to be both a rotation and a reflection.

5. Prove that the composite of two reflections is a rotation on any two- or three-dimensional Euclidean space.

6. Let V be a two-dimensional Euclidean space, and let $x \neq y \in V$ be unit vectors. Show there exists a unique rotation T such that $T(x) = y$.

6.4 Projections

In this section we introduce the concept of projection operators and show how they may be related to direct sum decompositions where each of the subspaces in the direct sum is invariant under some linear operator.

Suppose that U and W are subspaces of a vector space V with the property that $V = U \oplus W$. Then every $v \in V$ has a unique representation of the form $v = u + w$ where $u \in U$ and $w \in W$ (Theorem 1.14). We now define the mapping $E : V \to V$ by $Ev = u$. Note that E is well-defined since the direct sum decomposition is unique. Moreover, given any other $v' \in V = U \oplus W$ with $v' = u' + w'$, we know that $v + v' = (u + u') + (w + w')$ and $kv = ku + kw$, and hence it is easy to see that E is in fact linear because

$$E(v + v') = u + u' = Ev + Ev'$$

and

$$E(kv) = ku = k(Ev).$$

The linear operator $E \in L(V)$ is called the **projection of V on U in the direction of W**. Furthermore, since any $u \in U \subset V$ may be written in the form $u = u + 0$, we also see that $Eu = u$ and therefore

$$E^2 v = E(Ev) = Eu = u = Ev.$$

In other words, a projection operator E has the property that $E^2 = E$. By way of terminology, any operator $T \in L(V)$ with the property that $T^2 = T$ is said to be **idempotent**.

On the other hand, given a vector space V, suppose we have an operator $E \in L(V)$ with the property that $E^2 = E$. We claim that $V = \operatorname{Im} E \oplus \operatorname{Ker} E$. Indeed, first note that if $u \in \operatorname{Im} E$, then by definition this means there exists $v \in V$ with the property that $Ev = u$. It therefore follows that

$$Eu = E(Ev) = E^2 v = Ev = u$$

and thus $Eu = u$ for any $u \in \operatorname{Im} E$. Conversely, the equation $Eu = u$ obviously says that $u \in \operatorname{Im} E$, and hence we see that $u \in \operatorname{Im} E$ if and only if $Eu = u$.

Next, note that given any $v \in V$ we may clearly write

$$v = Ev + v - Ev = Ev + (1 - E)v$$

where by definition, $Ev \in \operatorname{Im} E$. Since

$$E[(1 - E)v] = (E - E^2)v = (E - E)v = 0$$

we see that $(1 - E)v \in \operatorname{Ker} E$, and hence $V = \operatorname{Im} E + \operatorname{Ker} E$. We claim that this sum is in fact direct.

To see this, let $w \in \operatorname{Im} E \cap \operatorname{Ker} E$. Since $w \in \operatorname{Im} E$ and $E^2 = E$, we have seen that $Ew = w$, while the fact that $w \in \operatorname{Ker} E$ means that $Ew = 0$. Therefore $w = 0$ so that $\operatorname{Im} E \cap \operatorname{Ker} E = \{0\}$, and hence

$$V = \operatorname{Im} E \oplus \operatorname{Ker} E.$$

Since we have now shown that any $v \in V$ may be written in the unique form $v = u + w$ with $u \in \operatorname{Im} E$ and $w \in \operatorname{Ker} E$, it follows that $Ev = Eu + Ew = u + 0 = u$ so that E is the projection of V on $\operatorname{Im} E$ in the direction of $\operatorname{Ker} E$.

It is also of use to note that

$$\operatorname{Ker} E = \operatorname{Im}(1 - E)$$

and

$$\operatorname{Ker}(1 - E) = \operatorname{Im} E.$$

To see this, suppose $w \in \operatorname{Ker} E$. Then

$$w = Ew + (1 - E)w = (1 - E)w$$

which implies $w \in \operatorname{Im}(1 - E)$, and hence $\operatorname{Ker} E \subset \operatorname{Im}(1 - E)$. On the other hand, if $w \in \operatorname{Im}(1 - E)$ then there exists $w' \in V$ such that $w = (1 - E)w'$ and hence

$$Ew = (E - E^2)w' = (E - E)w' = 0$$

so that $w \in \operatorname{Ker} E$. This shows that $\operatorname{Im}(1 - E) \subset \operatorname{Ker} E$, and therefore $\operatorname{Ker} E = \operatorname{Im}(1 - E)$. The similar proof that $\operatorname{Ker}(1 - E) = \operatorname{Im} E$ is left as an exercise for the reader (Exercise 6.4.1).

Theorem 6.16. *Let V be a vector space with $\dim V = n$, and suppose $E \in L(V)$ has rank $k = \dim(\operatorname{Im} E)$. Then E is idempotent (i.e., $E^2 = E$) if and only if any one of the following statements is true:*

(i) If $v \in \operatorname{Im} E$, then $Ev = v$.

(ii) $V = \operatorname{Im} E \oplus \operatorname{Ker} E$ and E is the projection of V on $\operatorname{Im} E$ in the direction of $\operatorname{Ker} E$.

(iii) $\operatorname{Im} E = \operatorname{Ker}(1 - E)$ and $\operatorname{Ker} E = \operatorname{Im}(1 - E)$.

(iv) It is possible to choose a basis for V such that $[E] = I_k \oplus 0_{n-k}$.

Proof. Suppose $E^2 = E$. In view of the above discussion, all that remains is to prove part (iv). Applying part (ii), we let $\{e_1, \dots, e_k\}$ be a basis for $\operatorname{Im} E$ and $\{e_{k+1}, \dots, e_n\}$ be a basis for $\operatorname{Ker} E$. By part (i), we know that $Ee_i = e_i$ for $i = 1, \dots, k$ and, by definition of $\operatorname{Ker} E$, we have $Ee_i = 0$ for $i = k+1, \dots, n$. But then $[E]$ has the desired form since the ith column of $[E]$ is just Ee_i.

Conversely, suppose (i) is true and $v \in V$ is arbitrary. Then $E^2v = E(Ev) = Ev$ implies $E^2 = E$. Now suppose (ii) is true and $v \in V$. Then $v = u + w$ where $u \in \operatorname{Im} E$ and $w \in \operatorname{Ker} E$. Therefore $Ev = Eu + Ew = Eu = u$ (by definition of projection) and $E^2v = E^2u = Eu = u$ so that $E^2v = Ev$ for all $v \in V$, and hence $E^2 = E$. If (iii) holds and $v \in V$, then $Ev \in \operatorname{Im} E = \operatorname{Ker}(1 - E)$ so that $0 = (1 - E)Ev = Ev - E^2v$ and hence $E^2v = Ev$ again. Similarly, $(1 - E)v \in \operatorname{Im}(1 - E) = \operatorname{Ker} E$ so that $0 = E(1 - E)v = Ev - E^2v$ and hence $E^2v = Ev$. In either case, we have $E^2 = E$. Finally, from the form of $[E]$ given in (iv), it is obvious that $E^2 = E$. ∎

Note that part (ii) of this theorem is really a particular example of the rank theorem.

It is also worth making the following observation. If we are given a vector space V and a subspace $W \subset V$, then there may be many subspaces $U \subset V$ with the property that $V = U \oplus W$. For example, the space $\mathbb{R}^3$ is not necessarily represented by the usual orthogonal Cartesian coordinate system. Rather, it may be viewed as consisting of a line plus any (oblique) plane not containing the given line. However, in the particular case that $V = W \oplus W^\perp$, then $W^\perp$ is uniquely specified by W (see Section 1.6). In this case, the projection $E \in L(V)$ defined by $Ev = w$ with $w \in W$ is called the **orthogonal projection** of V on W. In other words, E is an orthogonal projection if $(\operatorname{Im} E)^\perp = \operatorname{Ker} E$. By the corollary to Theorem 1.22, this is equivalent to the requirement that $(\operatorname{Ker} E)^\perp = \operatorname{Im} E$.

It is not hard to generalize these results to the direct sum of more than two subspaces. Indeed, suppose we have a vector space V such that $V = W_1 \oplus \cdots \oplus W_r$. Since any $v \in V$ has the unique representation as $v = w_1 + \cdots + w_r$ with $w_i \in W_i$, we may define for each $j = 1, \dots, r$ the operator $E_j \in L(V)$ by $E_jv = w_j$. That each E_j is in fact linear is easily shown exactly as above for the simpler case. It should also be clear that $\operatorname{Im} E_j = W_j$ (see Exercise 6.4.2). If we write

$$w_j = 0 + \cdots + 0 + w_j + 0 + \cdots + 0$$

as the unique expression for $w_j \in W_j \subset V$, then we see that $E_jw_j = w_j$, and hence for any $v \in V$ we have

$$E_j{}^2v = E_j(E_jv) = E_jw_j = w_j = E_jv$$

so that $E_j{}^2 = E_j$.

The representation of each w_j as E_jv is very useful because we may write any $v \in V$ as

$$v = w_1 + \cdots + w_r = E_1v + \cdots + E_rv = (E_1 + \cdots + E_r)v$$

and thus we see that $E_1 + \cdots + E_r = 1$. Furthermore, since the image of E_j is W_j, it follows that if $E_j v = 0$ then $w_j = 0$, and hence

$$\operatorname{Ker} E_j = W_1 \oplus \cdots W_{j-1} \oplus W_{j+1} \oplus \cdots \oplus W_r.$$

We then see that for any $j = 1, \ldots, r$ we have $V = \operatorname{Im} E_j \oplus \operatorname{Ker} E_j$ exactly as before. It is also easy to see that $E_i E_j = 0$ if $i \neq j$ because $\operatorname{Im} E_j = W_j \subset \operatorname{Ker} E_i$.

Theorem 6.17. *Let V be a vector space, and suppose $V = W_1 \oplus \cdots \oplus W_r$. Then for each $j = 1, \ldots, r$ there exists a linear operator $E_j \in L(V)$ with the following properties:*
(i) $1 = E_1 + \cdots + E_r$.
(ii) $E_i E_j = 0$ *if* $i \neq j$.
(iii) ${E_j}^2 = E_j$.
(iv) $\operatorname{Im} E_j = W_j$.
Conversely, if $\{E_1, \ldots, E_r\}$ are linear operators on V that obey properties (i) and (ii), then each E_j is idempotent and $V = W_1 \oplus \cdots \oplus W_r$ where $W_j = \operatorname{Im} E_j$.

Proof. In view of the previous discussion, we only need to prove the converse statement. From (i) and (ii) we see that

$$E_j = E_j 1 = E_j(E_1 + \cdots + E_r) = {E_j}^2 + \sum_{i \neq j} E_j E_i = {E_j}^2$$

which shows that each E_j is idempotent. Next, property (i) shows us that for any $v \in V$ we have

$$v = 1v = E_1 v + \cdots + E_r v$$

and hence $V = W_1 + \cdots + W_r$ where we have defined $W_j = \operatorname{Im} E_j$. Now suppose $0 = w_1 + \cdots + w_r$ where each $w_j \in W_j$. If we can show this implies $w_1 = \cdots = w_r = 0$, then any $v \in V$ will have a unique representation $v = v_1 + \cdots + v_r$ with $v_i \in W_i$. This is because if

$$v = v_1 + \cdots + v_r = v_1' + \cdots + v_r'$$

then

$$(v_1 - v_1') + \cdots + (v_r - v_r') = 0$$

would imply $v_i - v_i' = 0$ for each i, and thus $v_i = v_i'$. Hence it will follow that $V = W_1 \oplus \cdots \oplus W_r$ (Theorem 1.14).

Since $w_1 + \cdots + w_r = 0$, it is obvious that $E_j(w_1 + \cdots + w_r) = 0$. However, note that $E_j w_i = 0$ if $i \neq j$ (because $w_i \in \operatorname{Im} E_i$ and $E_j E_i = 0$), while $E_j w_j = w_j$ (since $w_j = E_j w'$ for some $w' \in V$ and hence $E_j w_j = {E_j}^2 w' = E_j w' = w_j$). This shows that $w_1 = \cdots = w_r = 0$ as desired. ∎

We now turn our attention to invariant direct sum decompositions, referring to Section 5.5 for notation. We saw in Corollary 1 of Theorem 5.14 that a diagonalizable operator $T \in L(V)$ leads to a direct sum decomposition of V in terms of the eigenspaces of T. However, Theorem 6.17 shows us that such a decomposition should lead to a collection of projections on these eigenspaces. Our next theorem elaborates on this observation in detail. Before stating and proving this result however, let us take another look at a matrix that has been diagonalized.

We observe that a diagonal matrix of the form

$$D = \begin{bmatrix} \lambda_1 I_{m_1} & 0 & 0 & \cdots & 0 & 0 \\ 0 & \lambda_2 I_{m_2} & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & \lambda_r I_{m_r} \end{bmatrix}$$

can also be written as

$$D = \lambda_1 \begin{bmatrix} I_{m_1} & & & \\ & 0_{m_2} & & \\ & & \ddots & \\ & & & 0_{m_r} \end{bmatrix} + \lambda_2 \begin{bmatrix} 0_{m_1} & & & \\ & I_{m_2} & & \\ & & \ddots & \\ & & & 0_{m_r} \end{bmatrix}$$
$$+ \cdots + \lambda_r \begin{bmatrix} 0_{m_1} & & & \\ & 0_{m_2} & & \\ & & \ddots & \\ & & & I_{m_r} \end{bmatrix}.$$

If we define E_i to be the matrix obtained from D by setting $\lambda_i = 1$ and $\lambda_j = 0$ for each $j \neq i$ (i.e., the ith matrix in the above expression), then this may be written in the simple form

$$D = \lambda_1 E_1 + \lambda_2 E_2 + \cdots + \lambda_r E_r$$

where clearly

$$I = E_1 + E_2 + \cdots + E_r.$$

Furthermore, it is easy to see that the matrices E_i have the property that

$$E_i E_j = 0 \text{ if } i \neq j$$

and

$$E_i{}^2 = E_i \neq 0.$$

With these observations in mind, we now prove this result in general.

Theorem 6.18. *If $T \in L(V)$ is a diagonalizable operator with distinct eigenvalues $\lambda_1, \dots, \lambda_r$ then there exist linear operators $E_1, \dots, E_r$ in $L(V)$ such that:*

(i) $1 = E_1 + \cdots + E_r$.

(ii) $E_i E_j = 0$ *if* $i \neq j$.

(iii) $T = \lambda_1 E_1 + \cdots + \lambda_r E_r$.

(iv) ${E_j}^2 = E_j$.

(v) $\operatorname{Im} E_j = W_j$ *where* $W_j = \operatorname{Ker}(T - \lambda_j 1)$ *is the eigenspace corresponding to* λ_j.

Conversely, if there exist distinct scalars $\lambda_1, \dots, \lambda_r$ and distinct nonzero linear operators $E_1, \dots, E_r$ satisfying properties (i), (ii) and (iii), then properties (iv) and (v) are also satisfied, and T is diagonalizable with $\lambda_1, \dots, \lambda_r$ as its distinct eigenvalues.

Proof. First assume T is diagonalizable with distinct eigenvalues $\lambda_1, \dots, \lambda_r$ and let $W_1, \dots, W_r$ be the corresponding eigenspaces. By Corollary 1 of Theorem 5.14 we know $V = W_1 \oplus \cdots \oplus W_r$. Then Theorem 6.17 shows the existence of the projection operators $E_1, \dots, E_r$ satisfying properties (i), (ii), (iv) and (v). As to property (iii), we see (by property (i)) that for any $v \in V$ we have $v = E_1 v + \cdots + E_r v$. Since $E_j v \in W_j$, we know from the definition of eigenspace that $T(E_j v) = \lambda_j (E_j v)$, and therefore

$$\begin{aligned} Tv &= T(E_1 v) + \cdots + T(E_r v) \\ &= \lambda_1 (E_1 v) + \cdots + \lambda_r (E_r v) \\ &= (\lambda_1 E_1 + \cdots + \lambda_r E_r) v \end{aligned}$$

which verifies property (iii).

Now suppose we are given a linear operator $T \in L(V)$ together with distinct scalars $\lambda_1, \dots, \lambda_r$ and (nonzero) linear operators $E_1, \dots, E_r$ that obey properties (i), (ii) and (iii). Multiplying (i) by E_i and using (ii) proves (iv). Now multiply (iii) from the right by E_i and use property (ii) to obtain $TE_i = \lambda_i E_i$ or $(T - \lambda_i 1)E_i = 0$. If $w_i \in \operatorname{Im} E_i$ is arbitrary, then $w_i = E_i w_i'$ for some $w_i' \in V$ and hence $(T - \lambda_i 1) w_i = (T - \lambda_i 1) E_i w_i' = 0$ which shows that $w_i \in \operatorname{Ker}(T - \lambda_i 1)$. Since $E_i \neq 0$, this shows the existence of a nonzero vector $w_i \in \operatorname{Ker}(T - \lambda_i 1)$ with the property that $Tw_i = \lambda_i w_i$. This proves that each λ_i is an eigenvalue of T. We claim there are no other eigenvalues of T other than $\{\lambda_i\}$.

To see this, let α be any scalar and assume that $(T - \alpha 1)v = 0$ for some nonzero $v \in V$. Using properties (i) and (iii), we see that

$$T - \alpha 1 = (\lambda_1 - \alpha) E_1 + \cdots + (\lambda_r - \alpha) E_r$$

and hence letting both sides of this equation act on v yields

$$0 = (\lambda_1 - \alpha) E_1 v + \cdots + (\lambda_r - \alpha) E_r v.$$

Multiplying this last equation from the left by E_i and using properties (ii) and (iv), we then see that $(\lambda_i - \alpha) E_i v = 0$ for every $i = 1, \dots, r$. Since $v \neq 0$ may be written as $v = E_1v + \cdots + E_rv$, it must be true that $E_jv \neq 0$ for some j, and hence in this case we have $\lambda_j - \alpha = 0$ or $\alpha = \lambda_j$.

We must still show that T is diagonalizable, and that $\operatorname{Im} E_i = \operatorname{Ker}(T - \lambda_i 1)$. It was shown in the previous paragraph that any nonzero $w_j \in \operatorname{Im} E_j$ satisfies $Tw_j = \lambda_j w_j$, and hence any nonzero vector in the image of any E_i is an eigenvector of E_i. Note this says that $\operatorname{Im} E_i \subset \operatorname{Ker}(T - \lambda_i 1)$. Using property (i), we see that any $w \in V$ may be written as $w = E_1w + \cdots + E_rw$ which shows that V is spanned by eigenvectors of T. But this is just what we mean when we say that T is diagonalizable.

Finally, suppose $w_i \in \operatorname{Ker}(T - \lambda_i 1)$ is arbitrary. Then $(T - \lambda_i 1)w_i = 0$ and hence (exactly as we showed above)

$$0 = (\lambda_1 - \lambda_i)E_1w_i + \cdots + (\lambda_r - \lambda_i)E_rw_i.$$

Thus for each $j = 1, \ldots, r$ we have

$$0 = (\lambda_j - \lambda_i)E_jw_i$$

which implies $E_jw_i = 0$ for $j \neq i$. Since $w_i = E_1w_i + \cdots + E_rw_i$ while $E_jw_i = 0$ for $j \neq i$, we conclude that $w_i = E_iw_i$ which shows that $w_i \in \operatorname{Im} E_i$. In other words, we have also shown that $\operatorname{Ker}(T - \lambda_i 1) \subset \operatorname{Im} E_i$. Together with our earlier result, this proves $\operatorname{Im} E_i = \operatorname{Ker}(T - \lambda_i 1)$. ∎

Exercises

1. (a) Let E be an idempotent linear operator. Show $\operatorname{Ker}(1 - E) = \operatorname{Im} E$.
   (b) If $E^2 = E$, show $(1 - E)^2 = 1 - E$.

2. Let $V = W_1 \oplus \cdots \oplus W_r$ and suppose $v = w_1 + \cdots + w_r \in V$. For each $j = 1, \ldots, r$ we define the operator E_j on V by $E_jv = w_j$.
   (a) Show $E_j \in L(V)$.
   (b) Show $\operatorname{Im} E_j = W_j$.

3. Let $E_1, \ldots, E_r$ and $W_1, \ldots, W_r$ be as defined in Theorem 6.17, and suppose $T \in L(V)$.
   (a) If $TE_i = E_iT$ for every E_i, prove every $W_j = \operatorname{Im} E_j$ is T-invariant.
   (b) If every W_j is T-invariant, prove $TE_i = E_iT$ for every E_i. [*Hint*: Let $v \in V$ be arbitrary. Show that property (i) of Theorem 6.17 implies $T(E_iv) = w_i$ for some $w_i \in W_i = \operatorname{Im} E_i$. Now show that $E_j(TE_i)v = (E_iw_i)\delta_{ij}$, and hence that $E_j(Tv) = T(E_jv)$.]

4. Prove property (v) in Theorem 6.18 holds for the matrices E_i given prior to the theorem.

5. Let W be a finite-dimensional subspace of an inner product space V.
   (a) Show there exists precisely one orthogonal projection on W.
   (b) Let E be the orthogonal projection on W. Show that for any $v \in V$ we have $\|v - Ev\| \leq \|v - w\|$ for every $w \in W$. In other words, show Ev is the unique element of W that is "closest" to v.

6.5 The Spectral Theorem

We now turn to another major topic of this chapter, the so-called spectral theorem. This important result is actually nothing more than another way of looking at Theorem 5.18. We begin with a simple version that is easy to understand and visualize if the reader will refer back to the discussion prior to Theorem 6.18.

Theorem 6.19. *Suppose $A \in M_n(\mathbb{C})$ is a diagonalizable matrix with distinct eigenvalues $\lambda_1, \dots, \lambda_r$. Then A can be written in the form*

$$A = \lambda_1 E_1 + \cdots + \lambda_r E_r$$

where the E_i are $n \times n$ matrices with the following properties:
(i) Each E_i is idempotent (i.e., $E_i^2 = E_i$).
(ii) $E_i E_j = 0$ for $i \neq j$.
(iii) $E_1 + \cdots + E_r = I$.
(iv) $AE_i = E_i A$ for every E_i.

Proof. Since A is diagonalizable by assumption, let $D = P^{-1}AP$ be the diagonal form of A for some nonsingular matrix P (whose columns are just the eigenvectors of A). Remember that the diagonal elements of D are just the eigenvalues λ_i of A. Let P_i be the $n \times n$ diagonal matrix with diagonal element 1 wherever a λ_i occurs in D, and 0's everywhere else. It should be clear that the collection $\{P_i\}$ obeys properties (i)–(iii), and that

$$P^{-1}AP = D = \lambda_1 P_1 + \cdots + \lambda_r P_r.$$

If we now define $E_i = PP_iP^{-1}$, then we have

$$A = PDP^{-1} = \lambda_1 E_1 + \cdots + \lambda_r E_r$$

where the E_i also obey properties (i)–(iii) by virtue of the fact that the P_i do. Using (i) and (ii) in this last equation we find

$$AE_i = (\lambda_1 E_1 + \cdots + \lambda_r E_r)E_i = \lambda_i E_i$$

and similarly it follows that $E_i A = \lambda_i E_i$ so that each E_i commutes with A, i.e., $E_i A = AE_i$. ∎

By way of terminology, the collection of eigenvalues $\lambda_1, \dots, \lambda_r$ is called the **spectrum** of A, the sum $E_1 + \cdots + E_r = I$ is called the **resolution of the identity** induced by A, and the expression $A = \lambda_1 E_1 + \cdots + \lambda_r E_r$ is called the **spectral decomposition** of A. These definitions also apply to arbitrary normal operators as in Theorem 6.21 below.

Corollary. *Let A be diagonalizable with spectral decomposition as in Theorem 6.19. If $f(x) \in \mathbb{C}[x]$ is any polynomial, then*

$$f(A) = f(\lambda_1)E_1 + \cdots + f(\lambda_r)E_r.$$

Proof. Using properties (i)–(iii) in Theorem 6.19, it is easy to see that for any $m > 0$ we have

$$A^m = (\lambda_1)^m E_1 + \cdots + (\lambda_r)^m E_r.$$

The result for arbitrary polynomials now follows easily from this result. ∎

It is easy to see how to construct the spectral decomposition. Let D be the diagonal form of A. Disregarding multiplicities, we can write D as

$$D = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} = U^\dagger A U$$

where U is the unitary matrix whose columns are the eigenvectors of A. If we let v_i be the ith eigenvector and write $U = [v_1 \cdots v_n]$ as a function of its columns, then what we have is

$$\begin{aligned} A = UDU^\dagger &= \begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix} \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} \begin{bmatrix} v_1^\dagger \\ \vdots \\ v_n^\dagger \end{bmatrix} \\ &= \begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix} \begin{bmatrix} \lambda_1 v_1^\dagger \\ \vdots \\ \lambda_n v_n^\dagger \end{bmatrix} \\ &= \lambda_1 v_1 v_1^\dagger + \cdots + \lambda_n v_n v_n^\dagger . \end{aligned}$$

Note that each v_i is a column vector, and each $v_i^\dagger$ is a row vector, and hence each term $v_i v_i^\dagger$ is an $n \times n$ matrix.

Before turning to our proof of the spectral theorem, we first prove a simple but useful characterization of orthogonal projections.

Theorem 6.20. *Let V be an inner product space and suppose $E \in L(V)$. Then E is an orthogonal projection if and only if $E^2 = E = E^\dagger$.*

Proof. We first assume that E is an orthogonal projection. By definition this means $E^2 = E$, and hence we must show that $E^\dagger = E$. From Theorem 6.16

we know that $V = \operatorname{Im} E \oplus \operatorname{Ker} E = \operatorname{Im} E \oplus (\operatorname{Im} E)^\perp$. Suppose $v,\ w \in V$ are arbitrary. Then we may write $v = v_1 + v_2$ and $w = w_1 + w_2$ where $v_1,\ w_1 \in \operatorname{Im} E$ and $v_2,\ w_2 \in (\operatorname{Im} E)^\perp$. Therefore

$$\langle v, Ew\rangle = \langle v_1 + v_2, w_1\rangle = \langle v_1, w_1\rangle + \langle v_2, w_1\rangle = \langle v_1, w_1\rangle$$

and

$$\langle v, E^\dagger w\rangle = \langle Ev, w\rangle = \langle v_1, w_1 + w_2\rangle = \langle v_1, w_1\rangle + \langle v_1, w_2\rangle = \langle v_1, w_1\rangle.$$

In other words, $\langle v, (E - E^\dagger)w\rangle = 0$ for all $v,\ w \in V$, and hence $E = E^\dagger$ (by Theorem 6.4(i)).

On the other hand, if $E^2 = E = E^\dagger$, then we know from Theorem 6.16 that E is a projection of V on $\operatorname{Im} E$ in the direction of $\operatorname{Ker} E$, i.e., $V = \operatorname{Im} E \oplus \operatorname{Ker} E$. Therefore, we need only show that $\operatorname{Im} E$ and $\operatorname{Ker} E$ are orthogonal subspaces. To show this, let $w \in \operatorname{Im} E$ and $w' \in \operatorname{Ker} E$ be arbitrary. Then $Ew = w$ and $Ew' = 0$ so that

$$\langle w', w\rangle = \langle w', Ew\rangle = \langle E^\dagger w', w\rangle = \langle Ew', w\rangle = 0. \qquad \blacksquare$$

We are now in a position to prove the spectral theorem for normal operators. In order to distinguish projection operators from their matrix representations in this theorem, we denote the operators by π_i and the corresponding matrices by E_i.

Theorem 6.21 (**Spectral Theorem for Normal Operators**). *Let V be a finite-dimensional unitary space, and let N be a normal operator on V with distinct eigenvalues $\lambda_1, \dots, \lambda_r$. Then*

(i) $N = \lambda_1\pi_1 + \cdots + \lambda_r\pi_r$ where each π_i is the orthogonal projection of V onto a subspace $W_i = \operatorname{Im} \pi_i$.

(ii) $\pi_i\pi_j = 0$ for $i \neq j$.

(iii) $\pi_1 + \cdots + \pi_r = 1$.

(iv) $V = W_1 \oplus \cdots \oplus W_r$ where the subspaces W_i are mutually orthogonal.

(v) $W_j = \operatorname{Im} \pi_j = \operatorname{Ker}(N - \lambda_j 1)$ is the eigenspace corresponding to λ_j.

Proof. Choose any orthonormal basis $\{e_i\}$ for V, and let A be the matrix representation of N relative to this basis. As discussed following Theorem 5.6, the normal matrix A has the same eigenvalues as the normal operator N. By Theorem 5.18 we know that A is diagonalizable, and hence applying Theorem 6.19 we may write

$$A = \lambda_1 E_1 + \cdots + \lambda_r E_r$$

where $E_i^2 = E_i$, $E_iE_j = 0$ if $i \neq j$, and $E_1 + \cdots + E_r = I$. Furthermore, A is diagonalized by a unitary matrix P, and as we saw in the proof of Theorem 6.19, $E_i = PP_iP^\dagger$ where each P_i is a real diagonal matrix. Since each P_i is clearly Hermitian, this implies $E_i^\dagger = E_i$, and hence each E_i is an orthogonal projection (Theorem 6.20).

Now define $\pi_i \in L(V)$ as that operator whose matrix representation relative to the basis $\{e_i\}$ is just E_i. From the isomorphism between linear transformations and their representations (Theorem 4.11), it should be clear that

$$\begin{aligned} N &= \lambda_1\pi_1 + \cdots + \lambda_r\pi_r \\ \pi_i^\dagger &= \pi_i \\ \pi_i^2 &= \pi_i \\ \pi_i\pi_j &= 0 \text{ for } i \neq j \\ \pi_1 + \cdots + \pi_r &= 1. \end{aligned}$$

Since $\pi_i^2 = \pi_i = \pi_i^\dagger$, Theorem 6.20 tells us that each π_i is an orthogonal projection of V on the subspace $W_i = \operatorname{Im}\pi_i$. Since $\pi_1 + \cdots + \pi_r = 1$, we see that for any $v \in V$ we have $v = \pi_1 v + \cdots + \pi_r v$ so that $V = W_1 + \cdots + W_r$. To show this sum is direct suppose, for example, that

$$w_1 \in W_1 \cap (W_2 + \cdots + W_r).$$

This means that $w_1 = w_2 + \cdots + w_r$ where $w_i \in W_i$ for each $i = 1, \ldots, r$. Since $w_i \in W_i = \operatorname{Im}\pi_i$, it follows that there exists $v_i \in V$ such that $\pi_i v_i = w_i$ for each i. Then

$$w_i = \pi_i v_i = \pi_i^2 v_i = \pi_i w_i$$

and if $i \neq j$, then $\pi_i\pi_j = 0$ implies

$$\pi_i w_j = (\pi_i\pi_j)v_j = 0.$$

Applying π_1 to $w_1 = w_2 + \cdots + w_r$, we obtain $w_1 = \pi_1 w_1 = 0$. Hence we have shown that $W_1 \cap (W_2 + \cdots + W_r) = \{0\}$. Since this argument can clearly be applied to any of the W_i, we have proved that $V = W_1 \oplus \cdots \oplus W_r$.

Next we note that for each i, π_i is the *orthogonal* projection of V on $W_i = \operatorname{Im}\pi_i$ in the direction of $W_i^\perp = \operatorname{Ker}\pi_i$, so that $V = W_i \oplus W_i^\perp$. Therefore, since $V = W_1 \oplus \cdots \oplus W_r$, it follows that for each $j \neq i$ we must have $W_j \subset W_i^\perp$, and hence the subspaces W_i must be mutually orthogonal. Finally, the fact that $W_j = \operatorname{Ker}(N - \lambda_j 1)$ was proved in Theorem 6.18. ∎

The observant reader will have noticed the striking similarity between the spectral theorem and Theorem 6.18. In fact, part of Theorem 6.21 is essentially a corollary of Theorem 6.18. This is because a normal operator is diagonalizable, and hence satisfies the hypotheses of Theorem 6.18. However, note that in the present case we have used the existence of an inner product in our proof, whereas in Section 6.4, no such structure was assumed to exist.

Theorem 6.22. *Let $\sum_{j=1}^r \lambda_j E_j$ be the spectral decomposition of a normal operator N on a finite-dimensional unitary space. Then for each $i = 1, \ldots, r$ there exists a polynomial $f_i(x) \in \mathbb{C}[x]$ such that $f_i(\lambda_j) = \delta_{ij}$ and $f_i(N) = E_i$.*

Proof. For each $i = 1, \ldots, r$ we must find a polynomial $f_i(x) \in \mathbb{C}[x]$ with the property that $f_i(\lambda_j) = \delta_{ij}$. It should be obvious that the polynomials $f_i(x)$ defined by

$$f_i(x) = \prod_{j \neq i} \frac{x - \lambda_j}{\lambda_i - \lambda_j}$$

have this property. From the corollary to Theorem 6.19 we have $p(N) = \sum_j p(\lambda_j) E_j$ for any $p(x) \in \mathbb{C}[x]$, and hence

$$f_i(N) = \sum_j f_i(\lambda_j) E_j = \sum_j \delta_{ij} E_j = E_i$$

as required. ∎

Exercises

1. Let T be an operator on a finite-dimensional unitary space. Prove T is unitary if and only if T is normal and $|\lambda| = 1$ for every eigenvalue λ of T.

2. Let H be a normal operator on a finite-dimensional unitary space. Prove that H is Hermitian if and only if every eigenvalue of H is real.

3. Let $V_n \subset \mathcal{F}[x]$ denote the set of all polynomials of degree $\leq n$, and let $a_0, a_1, \ldots, a_n \in \mathcal{F}$ be distinct.
   (a) Show V_n is a vector space over $\mathcal{F}$ with basis $\{1, x, x^2, \ldots, x^n\}$, and hence that $\dim V_n = n + 1$.
   (b) For each $i = 0, \ldots, n$ define the mapping $T_i : V_n \to \mathcal{F}$ by $T_i(f) = f(a_i)$. Show that the T_i are linear functionals on V_n, i.e., that $T_i \in {V_n}^*$.
   (c) For each $k = 0, \ldots, n$ define the polynomial
   $$\begin{aligned} p_k(x) &= \frac{(x - a_0) \cdots (x - a_{k-1})(x - a_{k+1}) \cdots (x - a_n)}{(a_k - a_0) \cdots (a_k - a_{k-1})(a_k - a_{k+1}) \cdots (a_k - a_n)} \\ &= \prod_{i \neq k} \left(\frac{x - a_i}{a_k - a_i} \right) \in V_n. \end{aligned}$$
   Show that $T_i(p_j) = \delta_{ij}$.
   (d) Show $p_0, \ldots, p_n$ forms a basis for V_n, and hence that any $f \in V_n$ may be written as
   $$f = \sum_{i=0}^{n} f(a_i) p_i.$$
   (e) Now let $b_0, b_1, \ldots, b_n \in \mathcal{F}$ be arbitrary, and define $f = \sum b_i p_i$. Show $f(a_j) = b_j$ for $0 \leq j \leq n$. Thus there exists a polynomial of degree $\leq n$ that takes on given values at $n + 1$ distinct points.

(f) Now assume that $f, g \in \mathcal{F}[x]$ are of degree $\leq n$ and satisfy $f(a_j) = b_j = g(a_j)$ for $0 \leq j \leq n$. Prove $f = g$, and hence that the polynomial defined in part (e) is unique. This is called the **Lagrange interpolation formula**.

(g) Let N be an operator on a finite-dimensional unitary space. Prove that N is normal if and only if $N^\dagger = g(N)$ for some polynomial g. [*Hint*: If N is normal with eigenvalues $\lambda_1, \ldots, \lambda_r$, show the existence of a polynomial g such that $g(\lambda_i) = \lambda_i^*$ for each i.]

6.6 Positive Operators

Before proving the main result of this section (the polar decomposition theorem), let us briefly discuss functions of a linear transformation. We have already seen an example of such a function. If A is a normal operator with spectral decomposition $A = \sum \lambda_i E_i$, then we saw that the linear transformation $p(A)$ was given by $p(A) = \sum p(\lambda_i) E_i$ where $p(x)$ is any polynomial in $\mathbb{C}[x]$ (Corollary to Theorem 6.19).

In order to generalize this notion, let N be a normal operator on a unitary space, and hence N has spectral decomposition $\sum \lambda_i E_i$. If f is an arbitrary complex-valued function (defined at least at each of the λ_i), we define a linear transformation $f(N)$ by

$$f(N) = \sum_i f(\lambda_i) E_i.$$

What we are particularly interested in is the function $f(x) = \sqrt{x}$ defined for all real $x \geq 0$ as the positive square root of x.

Recall (see Section 6.2) that we defined a positive operator P by the requirement that $P = S^\dagger S$ for some operator S. It is then clear that $P^\dagger = P$, and hence P is Hermitian (and therefore also normal). From Theorem 6.7(iv), the eigenvalues of $P = \sum \lambda_j E_j$ are real and non-negative, and we can define $\sqrt{P}$ by

$$\sqrt{P} = \sum_j \sqrt{\lambda_j} E_j$$

where each $\lambda_j \geq 0$.

Using the properties of the E_j, it is easy to see that $(\sqrt{P}\,)^2 = P$. Furthermore, since E_j is an orthogonal projection, it follows that $E_j^\dagger = E_j$ (Theorem 6.20), and therefore $(\sqrt{P}\,)^\dagger = \sqrt{P}$ so that $\sqrt{P}$ is Hermitian. Note that since $P = S^\dagger S$ we have

$$\langle Pv, v\rangle = \langle (S^\dagger S)v, v\rangle = \langle Sv, Sv\rangle = \|Sv\|^2 \geq 0.$$

Using $\sum E_j = 1$, let us write $v = \sum E_j v = \sum v_j$ where the nonzero v_j are mutually orthogonal (either from Theorem 6.21(iv) or by direct calculation since $\langle v_i, v_j\rangle = \langle E_i v, E_j v\rangle = \langle v, E_i^\dagger E_j v\rangle = \langle v, E_i E_j v\rangle = 0$ for $i \neq j$). Then

$$\sqrt{P}(v) = \sum_j \sqrt{\lambda_j} E_j v = \sum_j \sqrt{\lambda_j} v_j$$

and hence we also have (using $\langle v_j, v_k\rangle = 0$ if $j \neq k$)

$$\begin{aligned}\langle\sqrt{P}(v), v\rangle &= \left\langle \sum_j \sqrt{\lambda_j}\, v_j, \sum_k v_k \right\rangle = \sum_{j,k}\sqrt{\lambda_j}\langle v_j, v_k\rangle = \sum_j \sqrt{\lambda_j}\langle v_j, v_j\rangle \\ &= \sum_j \sqrt{\lambda_j}\,\|v_j\|^2 \geq 0.\end{aligned}$$

In summary, we have shown that $\sqrt{P}$ satisfies

(PT1) $(\sqrt{P}\,)^2 = P$
(PT2) $(\sqrt{P}\,)^\dagger = \sqrt{P}$
(PT3) $\langle\sqrt{P}(v), v\rangle \geq 0$

and it is natural to ask about the uniqueness of any operator satisfying these three properties. For example, if we let $T = \sum \pm\sqrt{\lambda_j}E_j$, then we still have $T^2 = \sum \lambda_j E_j = P$ regardless of the sign chosen for each term. Let us denote the fact that $\sqrt{P}$ satisfies properties (PT2) and (PT3) above by the expression $\sqrt{P} \geq 0$. In other words, by the statement $A \geq 0$ we mean $A^\dagger = A$ and $\langle Av, v\rangle \geq 0$ for every $v \in V$ (i.e., A is a positive Hermitian operator).

We now claim that if $P = T^2$ and $T \geq 0$, then $T = \sqrt{P}$. To prove this, we first note that $T \geq 0$ implies $T^\dagger = T$ (property (PT2)), and hence T must also be normal. Now let $\sum \mu_i F_i$ be the spectral decomposition of T. Then

$$\sum_i (\mu_i)^2 F_i = T^2 = P = \sum_j \lambda_j E_j .$$

If $v_i \neq 0$ is an eigenvector of T corresponding to μ_i, then property (PT3) tells us (using the fact that each μ_i is real because T is Hermitian)

$$0 \leq \langle Tv_i, v_i\rangle = \langle \mu_i v_i, v_i\rangle = \mu_i \|v_i\|^2 .$$

But $\|v_i\| > 0$, and hence $\mu_i \geq 0$. In other words, any operator $T \geq 0$ has nonnegative eigenvalues.

Since each μ_i is distinct and nonnegative, so is each $(\mu_i)^2$, and hence each $(\mu_i)^2$ must be equal to some λ_j. Therefore the corresponding F_i and E_j must be equal (by Theorem 6.21(v)). By suitably numbering the eigenvalues, we may write $(\mu_i)^2 = \lambda_i$, and thus $\mu_i = \sqrt{\lambda_i}$. This shows that

$$T = \sum_i \mu_i F_i = \sum_i \sqrt{\lambda_i}\, E_i = \sqrt{P}$$

as claimed.

We summarize this discussion in the next result which gives us three equivalent definitions of a **positive transformation**.

Theorem 6.23. *Let P be an operator on a unitary space V. Then the following conditions are equivalent:*
(i) $P = T^2$ for some unique Hermitian operator $T \geq 0$.
(ii) $P = S^\dagger S$ for some operator S.
(iii) $P^\dagger = P$ and $\langle Pv, v\rangle \geq 0$ for every $v \in V$.

Proof. (i) $\Rightarrow$ (ii): If $P = T^2$ and $T^\dagger = T$, then $P = TT = T^\dagger T$.

(ii) $\Rightarrow$ (iii): If $P = S^\dagger S$, then $P^\dagger = P$ and $\langle Pv, v\rangle = \langle S^\dagger Sv, v\rangle = \langle Sv, Sv\rangle = \|Sv\|^2 \geq 0$.

(iii) $\Rightarrow$ (i): Note that property (iii) is just our statement that $P \geq 0$. Since $P^\dagger = P$, we see that P is normal, and hence we may write $P = \sum \lambda_j E_j$. Defining $T = \sum \sqrt{\lambda_j} E_j$, we have $T^\dagger = T$ (since every E_j is Hermitian), and the preceding discussion shows that $T \geq 0$ is the unique operator with the property that $P = T^2$. ∎

We remark that in the particular case that P is positive definite, then $P = S^\dagger S$ where S is nonsingular. This means that P is also nonsingular.

Finally, we are in a position to prove the last result of this section, the so-called **polar decomposition** (or **factorization**) of an operator. While we state and prove this theorem in terms of matrices, it should be obvious by now that it applies just as well to operators. Also note that this theorem is quite similar to the SVD discussed in Section 5.8 and can in fact be proved as a direct consequence of Theorem 5.19.

Theorem 6.24 (Polar Decomposition). *If $A \in M_n(\mathbb{C})$, then there exist unique positive Hermitian matrices H_1, $H_2 \in M_n(\mathbb{C})$ and (not necessarily unique) unitary matrices U_1, $U_2 \in M_n(\mathbb{C})$ such that $A = U_1 H_1 = H_2 U_2$. Moreover, $H_1 = (A^\dagger A)^{1/2}$ and $H_2 = (AA^\dagger)^{1/2}$. In addition, the matrices U_1 and U_2 are uniquely determined if and only if A is nonsingular.*

Proof. Let $(\lambda_1)^2, \dots, (\lambda_n)^2$ be the eigenvalues of the positive Hermitian matrix $A^\dagger A$, and assume the λ_i are numbered so that $\lambda_i > 0$ for $i = 1, \dots, r$ and $\lambda_i = 0$ for $i = r+1, \dots, n$ (see Theorem 6.7(iv)). (Note that if A is nonsingular, then $A^\dagger A$ is positive definite and hence $r = n$.) Applying Theorem 5.18, we let $\{v_1, \dots, v_n\}$ be the corresponding orthonormal eigenvectors of $A^\dagger A$. For each $i = 1, \dots, r$ we define the vectors $w_i = Av_i/\lambda_i$. Then

$$\begin{aligned}\langle w_i, w_j\rangle &= \langle Av_i/\lambda_i, Av_j/\lambda_j\rangle = \langle v_i, A^\dagger A v_j\rangle/\lambda_i\lambda_j \\ &= \langle v_i, v_j\rangle(\lambda_j)^2/\lambda_i\lambda_j = \delta_{ij}(\lambda_j)^2/\lambda_i\lambda_j\end{aligned}$$

so that $w_1, \dots, w_r$ are also orthonormal. We now extend these to an orthonormal basis $\{w_1, \dots, w_n\}$ for $\mathbb{C}^n$. If we define the columns of the matrices V, $W \in$

$M_n(\mathbb{C})$ by $V^i = v_i$ and $W^i = w_i$, then V and W will be unitary by Theorem 5.15.

Defining the Hermitian matrix $D \in M_n(\mathbb{C})$ by

$$D = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$$

it is easy to see that the equations $Av_i = \lambda_i w_i$ may be written in matrix form as $AV = WD$. Using the fact that V and W are unitary, we define $U_1 = WV^\dagger$ and $H_1 = VDV^\dagger$ to obtain

$$A = WDV^\dagger = (WV^\dagger)(VDV^\dagger) = U_1 H_1.$$

Since $\det(\lambda I - VDV^\dagger) = \det(\lambda I - D)$, we see that H_1 and D have the same nonnegative eigenvalues, and hence H_1 is a positive Hermitian matrix. We can now apply this result to the matrix $A^\dagger$ to write $A^\dagger = \widetilde{U}_1 \widetilde{H}_1$ or $A = \widetilde{H}_1^\dagger \widetilde{U}_1^\dagger = \widetilde{H}_1 \widetilde{U}_1^\dagger$. If we define $H_2 = \widetilde{H}_1$ and $U_2 = \widetilde{U}_1^\dagger$, then we obtain $A = H_2 U_2$ as desired.

We now observe that using $A = U_1 H_1$ we may write

$$A^\dagger A = H_1 U_1^\dagger U_1 H_1 = (H_1)^2$$

and similarly

$$AA^\dagger = H_2 U_2 U_2^\dagger H_2 = (H_2)^2$$

so that H_1 and H_2 are unique even if A is singular. Since U_1 and U_2 are unitary, they are necessarily nonsingular, and hence H_1 and H_2 are nonsingular if $A = U_1 H_1 = H_2 U_2$ is nonsingular. In this case, $U_1 = AH_1^{-1}$ and $U_2 = H_2^{-1} A$ will also be unique. On the other hand, suppose A is singular. Then $r \neq n$ and $w_r, \ldots, w_n$ are not unique. This means $U_1 = WV^\dagger$ (and similarly U_2) is not unique. In other words, if U_1 and U_2 are unique, then A must be nonsingular. ∎

The reason this is called the "polar decomposition" is because of its analogy to complex numbers. Recall that a complex number z can be written in polar form as $z = |z| e^{i\theta} = re^{i\theta}$. Here, the unitary matrix is the analogue of $e^{i\theta}$ which gives a unit length "direction," and the Hermitian matrix plays the role of a distance.

Exercises

1. Let V be a unitary space and let $E \in L(V)$ be an orthogonal projection.
   (a) Show directly that E is a positive transformation.
   (b) Show $\|Ev\| \leq \|v\|$ for all $v \in V$.

2. Prove that if A and B are commuting positive transformations, then AB is also positive.

3. This exercise is related to Exercise 5.5.4. Prove that any representation $\{D(a) : a \in G\}$ of a finite group G is equivalent to a unitary representation as follows:

(a) Consider the matrix $X = \sum_{a\in G} D^\dagger(a)D(a)$. Show X is Hermitian and positive definite, and hence that $X = S^2$ for some Hermitian S.
(b) Show $D(a)^\dagger XD(a) = X$.
(c) Show $U(a) = SD(a)S^{-1}$ is a unitary representation.

6.7 The Matrix Exponential Series*

We now use Theorem 6.19 to prove a very useful result, namely, that any unitary matrix U can be written in the form e^{iH} for some Hermitian matrix H. Before proving this however, we must first discuss some of the theory of sequences and series of matrices. In particular, we must define just what is meant by expressions of the form e^{iH}. If the reader already knows something about sequences and series of numbers, then the rest of this section should present no difficulty. However, if not, then the stated results may be taken on faith.

Let $\{S_r\}$ be a sequence of complex matrices where each $S_r \in M_n(\mathbb{C})$ has entries $s_{ij}^{(r)}$. We say that $\{S_r\}$ **converges** to the **limit** $S = (s_{ij}) \in M_n(\mathbb{C})$ if each of the n^2 sequences $\{s_{ij}^{(r)}\}$ converges to a limit s_{ij}. We then write $S_r \to S$ or $\lim_{r\to\infty} S_r = S$ (or even simply $\lim S_r = S$). In other words, a sequence $\{S_r\}$ of matrices converges if and only if every entry of S_r forms a convergent sequence.

Similarly, an infinite series of matrices

$$\sum_{r=1}^{\infty} A_r$$

where $A_r = (a_{ij}^{(r)})$ is said to be **convergent** to the **sum** $S = (s_{ij})$ if the sequence of partial sums

$$S_m = \sum_{r=1}^{m} A_r$$

converges to S. Another way to say this is that the series $\sum A_r$ converges to S if and only if each of the n^2 series $\sum a_{ij}^{(r)}$ converges to s_{ij} for each $i, j = 1, \ldots, n$. We adhere to the convention of leaving off the limits in a series if they are infinite.

Our next theorem proves several intuitively obvious properties of sequences and series of matrices.

Theorem 6.25. *(i) Let $\{S_r\}$ be a convergent sequence of $n \times n$ matrices with limit S, and let P be any $n \times n$ matrix. Then $PS_r \to PS$ and $S_rP \to SP$.*

(ii) If $S_r \to S$ and P is nonsingular, then $P^{-1}S_rP \to P^{-1}SP$.

(iii) If $\sum A_r$ converges to A and P is nonsingular, then $\sum P^{-1}A_rP$ converges to $P^{-1}AP$.

Proof. (i) Since $S_r \to S$, we have $\lim s_{ij}^{(r)} = s_{ij}$ for all i, $j = 1, \ldots, n$. Therefore

$$\lim(PS_r)_{ij} = \lim\left(\sum_k p_{ik} s_{kj}^{(r)}\right) = \sum_k p_{ik} \lim s_{kj}^{(r)} = \sum_k p_{ik} s_{kj} = (PS)_{ij}.$$

Since this holds for all i, $j = 1, \ldots, n$ we must have $PS_r \to PS$. It should be obvious that we also have $S_r P \to SP$.

(ii) As in part (i), we have

$$\begin{aligned}
\lim(P^{-1} S_r P)_{ij} &= \lim\left(\sum_{k,m} p_{ik}^{-1} s_{km}^{(r)} p_{mj}\right) \\
&= \sum_{k,m} p_{ik}^{-1} p_{mj} \lim s_{km}^{(r)} \\
&= \sum_{k,m} p_{ik}^{-1} p_{mj} s_{km} \\
&= (P^{-1} SP)_{ij}.
\end{aligned}$$

Note that we may use part (i) to formally write this as

$$\lim(P^{-1} S_r P) = P^{-1} \lim(S_r P) = P^{-1} SP.$$

(iii) If we write the mth partial sum as

$$S_m = \sum_{r=1}^m P^{-1} A_r P = P^{-1}\left(\sum_{r=1}^m A_r\right) P$$

then we have

$$\begin{aligned}
\lim_{m \to \infty} (S_m)_{ij} &= \sum_{k,l} \lim\left[p_{ik}^{-1}\left(\sum_{r=1}^m a_{kl}^{(r)}\right) p_{lj}\right] \\
&= \sum_{k,l} p_{ik}^{-1} p_{lj} \lim \sum_{r=1}^m a_{kl}^{(r)} \\
&= \sum_{k,l} p_{ik}^{-1} p_{lj} a_{kl} \\
&= (P^{-1} AP)_{ij}.
\end{aligned}$$

∎

Theorem 6.26. *For any $A = (a_{ij}) \in M_n(\mathbb{C})$ the following series converges:*

$$\sum_{k=0}^{\infty} \frac{A^k}{k!} = I + A + \frac{A^2}{2!} + \cdots + \frac{A^r}{r!} + \cdots .$$

Proof. Choose a positive real number $M > \max\{n, |a_{ij}|\}$ where the max is taken over all i, $j = 1, \ldots, n$. Then $|a_{ij}| < M$ and $n < M < M^2$. Now consider the term $A^2 = (b_{ij}) = \left(\sum_k a_{ik}a_{kj}\right)$. We have (by Theorem 1.17, property (N3))

$$|b_{ij}| \leq \sum_{k=1}^{n} |a_{ik}|\,|a_{kj}| < \sum_{k=1}^{n} M^2 = nM^2 < M^4.$$

Proceeding by induction, suppose that for $A^r = (c_{ij})$, it has been shown that $|c_{ij}| < M^{2r}$. Then $A^{r+1} = (d_{ij})$ where

$$|d_{ij}| \leq \sum_{k=1}^{n} |a_{ik}|\,|c_{kj}| < nMM^{2r} = nM^{2r+1} < M^{2(r+1)}.$$

This proves $A^r = (a_{ij}^{(r)})$ has the property that $\left|a_{ij}^{(r)}\right| < M^{2r}$ for every $r \geq 1$.

Now, for each of the n^2 terms i, $j = 1, \ldots, n$ we have

$$\sum_{r=0}^{\infty} \frac{\left|a_{ij}^{(r)}\right|}{r!} < \sum_{r=0}^{\infty} \frac{M^{2r}}{r!} = \exp(M^2)$$

so that each of these n^2 series (i.e., for each i, $j = 1, \ldots, n$) must converge by the comparison test. Hence the series $I + A + A^2/2! + \cdots$ must converge (since a series that converges absolutely must converge). ∎

We call the series in Theorem 6.26 the **matrix exponential series**, and denote its sum by $e^A = \exp A$. In general, the series for e^A is extremely difficult, if not impossible, to evaluate. However, there are important exceptions.

Example 6.4. Let A be the diagonal matrix

$$A = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}.$$

Then it is easy to see that

$$A^r = \begin{bmatrix} (\lambda_1)^r & 0 & \cdots & 0 \\ 0 & (\lambda_2)^r & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & (\lambda_n)^r \end{bmatrix}.$$

and hence

$$\exp A = I + A + \frac{A^2}{2!} + \cdots = \begin{bmatrix} e^{\lambda_1} & 0 & \cdots & 0 \\ 0 & e^{\lambda_2} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & e^{\lambda_n} \end{bmatrix}.$$

Example 6.5. Consider the 2×2 matrix

$$J = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$$

and let

$$A = \theta J = \begin{bmatrix} 0 & -\theta \\ \theta & 0 \end{bmatrix}$$

where $\theta \in \mathbb{R}$. Noting $J^2 = -I$, we see that $A^2 = -\theta^2 I$, $A^3 = -\theta^3 J$, $A^4 = \theta^4 I$, $A^5 = \theta^5 J$, $A^6 = -\theta^6 I$, and so forth. From elementary calculus we know

$$\sin\theta = \theta - \frac{\theta^3}{3!} + \frac{\theta^5}{5!} - \cdots$$

and

$$\cos\theta = 1 - \frac{\theta^2}{2!} + \frac{\theta^4}{4!} - \cdots$$

and hence

$$\begin{aligned} e^A &= I + A + A^2/2! + \cdots \\ &= I + \theta J - \frac{\theta^2}{2!} I - \frac{\theta^3}{3!} J + \frac{\theta^4}{4!} I + \frac{\theta^5}{5!} J - \frac{\theta^6}{6!} I + \cdots \\ &= I\left(1 - \frac{\theta^2}{2!} + \frac{\theta^4}{4!} - \cdots\right) + J\left(\theta - \frac{\theta^3}{3!} + \frac{\theta^5}{5!} - \cdots\right) \\ &= (\cos\theta) I + (\sin\theta) J. \end{aligned}$$

In other words, using the explicit forms of I and J we see that

$$e^A = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

so that $e^{\theta J}$ represents a rotation in $\mathbb{R}^2$ by an angle θ.

Theorem 6.27. *Let $A \in M_n(\mathbb{C})$ be diagonalizable, and let $\lambda_1, \dots, \lambda_r$ be the distinct eigenvalues of A. Then the matrix power series*

$$\sum_{s=0}^{\infty} a_s A^s$$

converges if and only if the series

$$\sum_{s=0}^{\infty} a_s (\lambda_i)^s$$

converges for each $i = 1, \ldots, r$.

Proof. Since A is diagonalizable, choose a nonsingular matrix P such that $D = P^{-1}AP$ is diagonal. It is then easy to see that for every $s \geq 1$ we have

$$a_s D^s = a_s P^{-1} A^s P = P^{-1} a_s A^s P$$

where the n diagonal entries of D^s are just the numbers $(\lambda_i)^s$. By Theorem 6.25(iii), we know that $\sum a_s A^s$ converges if and only if $\sum a_s D^s$ converges. But by definition of series convergence, $\sum a_s D^s$ converges if and only if $\sum_s a_s(\lambda_i)^s$ converges for every $i = 1, \ldots, r$. ∎

Theorem 6.28. *Let* $f(x) = a_0 + a_1 x + a_2 x^2 + \cdots$ *be any power series with coefficients in* $\mathbb{C}$, *and let* $A \in M_n(\mathbb{C})$ *be diagonalizable with spectral decomposition* $A = \lambda_1 E_1 + \cdots + \lambda_r E_r$. *Then, if the series*

$$f(A) = a_0 I + a_1 A + a_2 A^2 + \cdots$$

converges, its sum is

$$f(A) = f(\lambda_1)E_1 + \cdots + f(\lambda_r)E_r.$$

Proof. As in the proof of Theorem 6.19, let the diagonal form of A be

$$D = P^{-1}AP = \lambda_1 P_1 + \cdots + \lambda_r P_r$$

so that $E_i = PP_iP^{-1}$. Now note that

$$\begin{aligned} P^{-1}f(A)P &= a_0 P^{-1}P + a_1 P^{-1}AP + a_2 P^{-1}APP^{-1}AP + \cdots \\ &= a_0 I + a_1 D + a_2 D^2 + \cdots \\ &= f(P^{-1}AP) \\ &= f(D). \end{aligned}$$

Using properties (i)–(iii) of Theorem 6.19 applied to the P_i, it is easy to see that $D^k = (\lambda_1)^k P_1 + \cdots + (\lambda_r)^k P_r$ and hence

$$f(D) = f(\lambda_1)P_1 + \cdots + f(\lambda_r)P_r.$$

Then if $f(A) = \sum a_k A^k$ converges, so does $\sum P^{-1} a_k A^k P = P^{-1} f(A) P = f(D)$ (Theorem 6.25(iii)), and we have

$$f(A) = f(PDP^{-1}) = Pf(D)P^{-1} = f(\lambda_1)E_1 + \cdots + f(\lambda_r)E_r.$$ ∎

Example 6.6. Consider the exponential series e^A where A is diagonalizable. Then, if $\lambda_1, \dots, \lambda_r$ are the distinct eigenvalues of A, we have the spectral decomposition $A = \lambda_1 E_1 + \cdots + \lambda_r E_r$. Using $f(A) = e^A$, Theorem 6.28 yields

$$e^A = e^{\lambda_1} E_1 + \cdots + e^{\lambda_r} E_r$$

in agreement with Example 6.4.

We can now prove our earlier assertion that a unitary matrix U can be written in the form e^{iH} for some Hermitian matrix H.

Theorem 6.29. *Every unitary matrix U can be written in the form e^{iH} for some Hermitian matrix H. Conversely, if H is Hermitian, then e^{iH} is unitary.*

Proof. By Theorem 6.7(ii), the distinct eigenvalues of U may be written in the form $e^{i\lambda_1}, \dots, e^{i\lambda_k}$ where each λ_i is real. Since U is also normal, it follows from Theorem 5.18 that there exists a unitary matrix P such that $P^\dagger UP = P^{-1}UP$ is diagonal. In fact

$$P^{-1}UP = e^{i\lambda_1} P_1 + \cdots + e^{i\lambda_k} P_k$$

where the P_i are the idempotent matrices used in the proof of Theorem 6.19.

From Example 6.4 we see that the matrix $e^{i\lambda_1} P_1 + \cdots + e^{i\lambda_k} P_k$ is just e^{iD} where

$$D = \lambda_1 P_1 + \cdots + \lambda_k P_k$$

is a diagonal matrix with the λ_i as diagonal entries. Therefore, using Theorem 6.25(iii) we see that

$$U = Pe^{iD}P^{-1} = e^{iPDP^{-1}} = e^{iH}$$

where $H = PDP^{-1}$. Since D is a real diagonal matrix it is clearly Hermitian, and since P is unitary (so that $P^{-1} = P^\dagger$), it follows that $H^\dagger = (PDP^\dagger)^\dagger = PDP^\dagger = H$ so that H is Hermitian also.

Conversely, suppose H is Hermitian with distinct real eigenvalues $\lambda_1, \dots, \lambda_k$. Since H is also normal, there exists a unitary matrix P that diagonalizes H. Then as above, we may write this diagonal matrix as

$$P^{-1}HP = \lambda_1 P_1 + \cdots + \lambda_k P_k$$

so that (from Example 6.4 again)

$$P^{-1}e^{iH}P = e^{iP^{-1}HP} = e^{i\lambda_1} P_1 + \cdots + e^{i\lambda_k} P_k.$$

Using the properties of the P_i, it is easy to see that the right hand side of this equation is diagonal and unitary. Indeed, using

$$(e^{i\lambda_1}P_1 + \cdots + e^{i\lambda_k}P_k)^\dagger = e^{-i\lambda_1}P_1 + \cdots + e^{-i\lambda_k}P_k$$

we have

$$(e^{i\lambda_1}P_1 + \cdots + e^{i\lambda_k}P_k)^\dagger(e^{i\lambda_1}P_1 + \cdots + e^{i\lambda_k}P_k) = I$$

and

$$(e^{i\lambda_1}P_1 + \cdots + e^{i\lambda_k}P_k)(e^{i\lambda_1}P_1 + \cdots + e^{i\lambda_k}P_k)^\dagger = I.$$

Therefore the left hand side must also be unitary, and hence (using $P^{-1} = P^\dagger$)

$$\begin{aligned} I &= (P^{-1}e^{iH}P)^\dagger(P^{-1}e^{iH}P) \\ &= P^\dagger(e^{iH})^\dagger PP^{-1}e^{iH}P \\ &= P^\dagger(e^{iH})^\dagger e^{iH}P \end{aligned}$$

so that $PP^{-1} = I = (e^{iH})^\dagger e^{iH}$. Similarly we see that $e^{iH}(e^{iH})^\dagger = I$, and thus e^{iH} is unitary. ∎

While this theorem is also true in infinite dimensions (i.e., in a Hilbert space), its proof is considerably more difficult. The reader is referred to the books listed in the bibliography for this generalization.

Given a *constant* matrix A, we now wish to show that

$$\frac{de^{tA}}{dt} = Ae^{tA}. \tag{6.1}$$

To see this, we first define the derivative of a matrix $M = M(t)$ to be that matrix whose elements are just the derivatives of the corresponding elements of M. In other words, if $M(t) = (m_{ij}(t))$, then $(dM/dt)_{ij} = dm_{ij}/dt$. Now note that (with $M(t) = tA$)

$$e^{tA} = I + tA + \frac{(tA)^2}{2!} + \frac{(tA)^3}{3!} + \cdots$$

and hence (since the a_{ij} are constant) taking the derivative with respect to t yields the desired result:

$$\begin{aligned} \frac{de^{tA}}{dt} &= 0 + A + tAA + \frac{(tA)^2A}{2!} + \cdots \\ &= A\left[I + tA + \frac{(tA)^2}{2!} + \cdots\right] \\ &= Ae^{tA}. \end{aligned}$$

Next, given two matrices A and B (of compatible sizes), we recall their **commutator** is the matrix $[A, B] = AB - BA = -[B, A]$. If $[A, B] = 0$, then $AB = BA$ and we say that A and B **commute**. Now consider the function

$f(x) = e^{xA}Be^{-xA}$. Leaving it to the reader to verify that the product rule for derivatives also holds for matrices, we obtain (note that $Ae^{xA} = e^{xA}A$)

$$\begin{aligned}
\frac{df}{dx} &= Ae^{xA}Be^{-xA} - e^{xA}Be^{-xA}A = Af - fA = [A, f] \\
\frac{d^2f}{dx^2} &= [A, df/dx] = [A, [A, f]] \\
&\vdots
\end{aligned}$$

Expanding $f(x)$ in a Taylor series about $x = 0$, we find (using $f(0) = B$)

$$\begin{aligned}
f(x) &= f(0) + \left(\frac{df}{dx}\right)_0 x + \left(\frac{d^2f}{dx^2}\right)_0 \frac{x^2}{2!} + \cdots \\
&= B + [A, B]x + [A, [A, B]]\frac{x^2}{2!} + \cdots .
\end{aligned}$$

Setting $x = 1$, we finally obtain

$$e^A Be^{-A} = B + [A, B] + \frac{1}{2!}[A, [A, B]] + \frac{1}{3!}[A, [A, [A, B]]] + \cdots \tag{6.2}$$

Note that setting $B = I$ shows that $e^A e^{-A} = I$ as we would hope.

In the particular case that both A and B commute with their commutator $[A, B]$, then we find from equation (6.2) that $e^A Be^{-A} = B + [A, B]$ and hence $e^A B = Be^A + [A, B]e^A$ or

$$[e^A, B] = [A, B]e^A. \tag{6.3}$$

Example 6.7. We now show that if A and B are two matrices *that both commute with their commutator* $[A, B]$, then

$$e^A e^B = \exp\left\{A + B + \frac{1}{2}[A, B]\right\}. \tag{6.4}$$

(This is sometimes referred to as **Weyl's formula**.)

To prove this, we start with the function $f(x) = e^{xA}e^{xB}e^{-x(A+B)}$. Then

$$\begin{aligned}
\frac{df}{dx} &= e^{xA}Ae^{xB}e^{-x(A+B)} + e^{xA}e^{xB}Be^{-x(A+B)} - e^{xA}e^{xB}(A + B)e^{-x(A+B)} \\
&= e^{xA}Ae^{xB}e^{-x(A+B)} - e^{xA}e^{xB}Ae^{-x(A+B)} \\
&= e^{xA}[A, e^{xB}]e^{-x(A+B)} \qquad (6.5)
\end{aligned}$$

As a special case, note $[A, B] = 0$ implies $df/dx = 0$ so that f is independent of x. Since $f(0) = I$, it follows that we may choose $x = 1$ to obtain $e^A e^B e^{-(A+B)} = I$ or $e^A e^B = e^{A+B}$ (as long as $[A, B] = 0$).

From equation (6.3) we have (replacing A by xB and B by A) $[A, e^{xB}] = x[A,B]e^{xB}$. Using this along with the fact that A commutes with the commutator $[A,B]$ (so that $e^{xA}[A,B] = [A,B]e^{xA}$), we have

$$\frac{df}{dx} = xe^{xA}[A,B]e^{xB}e^{-x(A+B)} = x[A,B]f.$$

Since A and B are independent of x, we may formally integrate this from 0 to x to obtain

$$\ln \frac{f(x)}{f(0)} = [A,B]\frac{x^2}{2}.$$

Using $f(0) = I$, this is $f(x) = \exp\{[A,B]x^2/2\}$ so that setting $x = 1$ we find

$$e^A e^B e^{-(A+B)} = \exp\left\{\frac{1}{2}[A,B]\right\}.$$

Finally, multiplying this equation from the right by e^{A+B} and using the fact that$[[A,B]/2, A+B] = 0$ yields equation (6.4).

Exercises

1. (a) Let N be a normal operator on a finite-dimensional unitary space. Prove
$$\det e^N = e^{\operatorname{tr} N}.$$
   (b) Prove this holds for any $N \in M_n(\mathbb{C})$. [*Hint*: Use Theorem 5.17.]

2. If the limit of a sequence of unitary operators exists, is it also unitary? Why?

3. Let T be a unitary operator. Show the sequence $\{T^n : n = 0, 1, 2, \dots\}$ contains a subsequence $\{T^{n_k} : k = 0, 1, 2, \dots\}$ that converges to a unitary operator. [*Hint*: You will need the fact that the unit disk in $\mathbb{C}^2$ is compact.]

Chapter 7

Linear Forms

7.1 Bilinear Forms

Let U and V be vector spaces over $\mathcal{F}$. We say that a mapping $f : U \times V \to \mathcal{F}$ is **bilinear** if it has the following properties for all $u_1, u_2 \in U$, for all $v_1, v_2 \in V$ and all $a, b \in \mathcal{F}$:

(BM1) $f(au_1 + bu_2, v_1) = af(u_1, v_1) + bf(u_2, v_1)$.
(BM2) $f(u_1, av_1 + bv_2) = af(u_1, v_1) + bf(u_1, v_2)$.

In other words, f is bilinear if for each $v \in V$ the mapping $u \mapsto f(u, v)$ is linear, and if for each $u \in U$ the mapping $v \mapsto f(u, v)$ is linear. In the particular case that $V = U$, then the bilinear map $f : V \times V \to \mathcal{F}$ is called a **bilinear form** on V. Rather than write expressions like $f(u, v)$, we will sometimes write the bilinear map as $\langle u, v\rangle$ if there is no need to refer to the mapping f explicitly. While this notation is used to denote several different operations, the context generally makes it clear exactly what is meant.

We say the bilinear map $f : U \times V \to \mathcal{F}$ is **nondegenerate** if $f(u, v) = 0$ for all $v \in V$ implies $u = 0$, and $f(u, v) = 0$ for all $u \in U$ implies $v = 0$.

Example 7.1. Suppose $A = (a_{ij}) \in M_n(\mathcal{F})$. Then we may interpret A as a bilinear form on $\mathcal{F}^n$ as follows. In terms of the standard basis $\{e_i\}$ for $\mathcal{F}^n$, any $X \in \mathcal{F}^n$ may be written as $X = \sum x_i e_i$, and hence for all $X, Y \in \mathcal{F}^n$ we define the bilinear form f_A by

$$f_A(X, Y) = \sum_{i,j} a_{ij} x_i y_j = X^T AY.$$

Here the row vector X^T is the transpose of the column vector X, and the expression $X^T AY$ is just the usual matrix product. It should be easy for the reader to verify that f_A is actually a bilinear form on $\mathcal{F}^n$.

Example 7.2. Recall from Section 4.1 that the vector space $V^* = L(V, \mathcal{F}) : V \to \mathcal{F}$ is defined to be the space of linear functionals on V. In other words, if $\phi \in V^*$, then for every $u, v \in V$ and $a, b \in \mathcal{F}$ we have

$$\phi(au + bv) = a\phi(u) + b\phi(v) \in \mathcal{F}.$$

The space V^* is called the **dual space** of V. If V is finite-dimensional, then viewing $\mathcal{F}$ as a one-dimensional vector space (over $\mathcal{F}$), it follows from Theorem 4.4 that $\dim V^* = \dim V$. In particular, given a basis $\{e_i\}$ for V, the proof of Theorem 4.4 showed that a unique basis $\{\omega^i\}$ for V^* is defined by the requirement that

$$\omega^i(e_j) = \delta^i_j$$

where we now again use superscripts to denote basis vectors in the dual space. We refer to the basis $\{\omega^i\}$ for V^* as the basis **dual** to the basis $\{e_i\}$ for V. Elements of V^* are usually referred to as **1-forms**, and are commonly denoted by Greek letters such as β, ϕ, θ and so forth. Similarly, we often refer to the ω^i as **basis 1-forms**.

Suppose $\alpha, \beta \in V^*$. Since α and β are linear, we may define a bilinear form $f : V \times V \to \mathcal{F}$ by

$$f(u, v) = \alpha(u)\beta(v)$$

for all $u, v \in V$. This form is usually denoted by $\alpha \otimes \beta$ and is called the **tensor product** of α and β. In other words, the tensor product of two elements α, $\beta \in V^*$ is defined for all $u, v \in V$ by

$$(\alpha \otimes \beta)(u, v) = \alpha(u)\beta(v).$$

We may also define the bilinear form $g : V \times V \to \mathcal{F}$ by

$$g(u, v) = \alpha(u)\beta(v) - \alpha(v)\beta(u).$$

We leave it to the reader to show that this is indeed a bilinear form. The mapping g is usually denoted by $\alpha \wedge \beta$, and is called the **wedge product** or the **antisymmetric tensor product** of α and β. In other words

$$(\alpha \wedge \beta)(u, v) = \alpha(u)\beta(v) - \alpha(v)\beta(u).$$

Note that $\alpha \wedge \beta$ is just $\alpha \otimes \beta - \beta \otimes \alpha$. We will have much more to say about these mappings in Chapter 8.

Generalizing Example 7.1 leads to the following theorem.

Theorem 7.1. *Given a bilinear map $f : \mathcal{F}^m \times \mathcal{F}^n \to \mathcal{F}$, there exists a unique matrix $A \in M_{m\times n}(\mathcal{F})$ such that $f = f_A$. In other words, there exists a unique matrix A such that $f(X, Y) = X^T AY$ for all $X \in \mathcal{F}^m$ and $Y \in \mathcal{F}^n$.*

Proof. In terms of the standard bases for $\mathcal{F}^m$ and $\mathcal{F}^n$, we have the column vectors $X = \sum_{i=1}^m x_i e_i \in \mathcal{F}^m$ and $Y = \sum_{j=1}^n y_j e_j \in \mathcal{F}^n$. Using the bilinearity of f we then have

$$f(X,Y) = f\left(\sum_i x_i e_i, \sum_j y_j e_j\right) = \sum_{i,j} x_i y_j f(e_i, e_j).$$

If we define $a_{ij} = f(e_i, e_j)$, then we see our expression becomes

$$f(X,Y) = \sum_{i,j} x_i a_{ij} y_j = X^T A Y.$$

To prove the uniqueness of the matrix A, suppose there exists a matrix A' such that $f = f_{A'}$. Then for all $X \in \mathcal{F}^m$ and $Y \in \mathcal{F}^n$ we have

$$f(X,Y) = X^T AY = X^T A' Y$$

and hence $X^T(A - A')Y = 0$. Now let $C = A - A'$ so that

$$X^T C Y = \sum_{i,j} c_{ij} x_i y_j = 0$$

for all $X \in \mathcal{F}^m$ and $Y \in \mathcal{F}^n$. In particular, choosing $X = e_i$ and $Y = e_j$, we find that $c_{ij} = 0$ for every i and j. Thus $C = 0$ so that $A = A'$. ∎

The matrix A defined in this theorem is said to **represent** the bilinear map f relative to the standard bases for $\mathcal{F}^m$ and $\mathcal{F}^n$. It thus appears that f is represented by the mn elements $a_{ij} = f(e_i, e_j)$. It is extremely important to realize that the elements a_{ij} are *defined* by the expression $f(e_i, e_j)$ and, conversely, given a matrix $A = (a_{ij})$, we *define* the expression $f(e_i, e_j)$ by requiring $f(e_i, e_j) = a_{ij}$. In other words, to say we are given a bilinear map $f : \mathcal{F}^m \times \mathcal{F}^n \to \mathcal{F}$ *means* that we are given values of $f(e_i, e_j)$ for each i and j. Then, given these values, we can evaluate expressions of the form $f(X,Y) = \sum_{i,j} x_i y_j f(e_i, e_j)$. Conversely, if we are given each of the $f(e_i, e_j)$, then we have defined a bilinear map on $\mathcal{F}^m \times \mathcal{F}^n$.

We denote the set of all bilinear maps on U and V by $\mathcal{B}(U \times V, \mathcal{F})$ and the set of all bilinear forms as simply $\mathcal{B}(V) = \mathcal{B}(V \times V, \mathcal{F})$. It is easy to make $\mathcal{B}(U \times V, \mathcal{F})$ into a vector space over $\mathcal{F}$. To do so, we simply define

$$(af + bg)(u, v) = af(u, v) + bg(u, v)$$

for any f, $g \in \mathcal{B}(U \times V, \mathcal{F})$ and a, $b \in \mathcal{F}$. The reader should have no trouble showing that $af + bg$ is itself a bilinear mapping.

It is left to the reader (see Exercise 7.1.1) to show that the association $A \mapsto f_A$ defined in Theorem 7.1 is actually an isomorphism between $M_{m\times n}(\mathcal{F})$ and $\mathcal{B}(F^m \times F^n, \mathcal{F})$. More generally, it should be clear that Theorem 7.1 applies equally well to any pair of finite-dimensional vector spaces U and V, and from now on we shall treat it as such.

Theorem 7.2. *Let V be finite-dimensional over $\mathcal{F}$, and let V^* have basis $\{\omega^i\}$. Define the elements $f^{ij} \in \mathcal{B}(V)$ by*

$$f^{ij}(u,v) = \omega^i(u)\omega^j(v)$$

for all u, $v \in V$. Then $\{f^{ij}\}$ forms a basis for $\mathcal{B}(V)$ which thus has dimension $(\dim V)^2$.

Proof. Let $\{e_i\}$ be the basis for V dual to the $\{\omega^i\}$ basis for V^*, and define $a_{ij} = f(e_i, e_j)$. Given any $f \in \mathcal{B}(V)$, we claim that $f = \sum_{i,j} a_{ij} f^{ij}$. To prove this, it suffices to show that $f(e_r, e_s) = \left(\sum_{i,j} a_{ij} f^{ij}\right)(e_r, e_s)$ for all r and s.

We first note that

$$\begin{aligned} \left(\sum_{i,j} a_{ij} f^{ij}\right)(e_r, e_s) &= \sum_{i,j} a_{ij}\omega^i(e_r)\omega^j(e_s) \\ &= \sum_{i,j} a_{ij}\delta^i_r \delta^j_s = a_{rs} \\ &= f(e_r, e_s). \end{aligned}$$

Since f is bilinear, it follows from this that $f(u,v) = \left(\sum_{i,j} a_{ij} f^{ij}\right)(u,v)$ for all u, $v \in V$ so that $f = \sum_{i,j} a_{ij} f^{ij}$. Hence $\{f^{ij}\}$ spans $\mathcal{B}(V)$.

Now suppose $\sum_{i,j} a_{ij} f^{ij} = 0$ (note this 0 is actually an element of $\mathcal{B}(V)$). Applying this to (e_r, e_s) and using the above result, we see that

$$0 = \left(\sum_{i,j} a_{ij} f^{ij}\right)(e_r, e_s) = a_{rs}.$$

Therefore $\{f^{ij}\}$ is linearly independent and hence forms a basis for $\mathcal{B}(V)$. ∎

It should be mentioned in passing that the functions f^{ij} defined in Theorem 7.2 can be written as the tensor product $\omega^i \otimes \omega^j : V \times V \to \mathcal{F}$ (see Example 7.2). Thus the set of bilinear forms $\omega^i \otimes \omega^j$ forms a basis for the space $V^* \otimes V^*$ which is called the **tensor product** of the two spaces V^*. This remark is not meant to be a complete treatment by any means, and we will return to these ideas in Chapter 8.

We also note that if $\{e_i\}$ is a basis for V and $\dim V = n$, then the matrix A of any $f \in \mathcal{B}(V)$ has elements $a_{ij} = f(e_i, e_j)$, and hence $A = (a_{ij})$ has n^2 independent elements. Thus, $\dim \mathcal{B}(V) = n^2$ as we saw above.

Theorem 7.3. *Let P be the transition matrix from a basis $\{e_i\}$ for V to a new basis $\{\bar{e}_i\}$. If A is the matrix of $f \in \mathcal{B}(V)$ relative to $\{e_i\}$, then $\bar{A} = P^T AP$ is the matrix of f relative to the basis $\{\bar{e}_i\}$.*

Proof. Let $X, Y \in V$ be arbitrary. In Section 4.4 we showed that the transition matrix $P = (p_{ij})$ defined by $\bar{e}_i = P(e_i) = \sum_j e_j p_{ji}$ also transforms the components of $X = \sum_i x_i e_i = \sum_j \bar{x}_j \bar{e}_j$ as $x_i = \sum_j p_{ij} \bar{x}_j$. In matrix notation, this may be written as $[X]_e = P[X]_{\bar{e}}$ (see Theorem 4.14), and hence $[X]_e^T = [X]_{\bar{e}}^T P^T$. From Theorem 7.1 we then have

$$f(X, Y) = [X]_e^T A[Y]_e = [X]_{\bar{e}}^T [P]^T A[P][Y]_{\bar{e}} = [X]_{\bar{e}}^T \bar{A}[Y]_{\bar{e}}.$$

Since X and Y are arbitrary, this shows that $\bar{A} = P^T AP$ is the unique representation of f in the new basis $\{\bar{e}_i\}$. ∎

Just as the transition matrix led to the definition of a similarity transformation, we now say that a matrix B is **congruent** to a matrix A if there exists a nonsingular matrix P such that $B = P^T AP$. It is easy to see that congruent matrices have the same rank. Indeed, by Theorem 2.16 we know that $\operatorname{rank}(P) = \operatorname{rank}(P^T)$. Then using Theorem 5.20 it follows that

$$\operatorname{rank}(B) = \operatorname{rank}(P^T AP) = \operatorname{rank}(AP) = \operatorname{rank}(A).$$

We are therefore justified in defining $\operatorname{rank}(f)$, the **rank** of a bilinear form f on V, to be the rank of any matrix representation of f. We leave it to the reader to show that f is nondegenerate if and only if $\operatorname{rank}(f) = \dim V$ (see Exercise 7.1.3).

Exercises

1. Show the association $A \mapsto f_A$ defined in Theorem 7.1 is an isomorphism between $M_{m \times m}(\mathcal{F})$ and $\mathcal{B}(F^m \times F^n, \mathcal{F})$.

2. Let $V = M_{m \times n}(\mathcal{F})$ and suppose $A \in M_m(\mathcal{F})$ is fixed. Then for any X, $Y \in V$ we define the mapping $f_A : V \times V \to \mathcal{F}$ by $f_A(X, Y) = \operatorname{tr}(X^T AY)$. Show this defines a bilinear form on V.

3. Prove that a bilinear form f on V is nondegenerate if and only if $\operatorname{rank}(f) = \dim V$.

4. (a) Let $V = \mathbb{R}^3$ and define $f \in \mathcal{B}(V)$ by

$$f(X, Y) = 3x_1y_1 - 2x_1y_2 + 5x_2y_1 + 7x_2y_2 - 8x_2y_3 + 4x_3y_2 - x_3y_3.$$

   Write out $f(X, Y)$ as a matrix product $X^T AY$.

   (b) Suppose $A \in M_n(\mathcal{F})$ and let $f(X, Y) = X^T AY$ for $X, Y \in \mathcal{F}^n$. Show $f \in \mathcal{B}(\mathcal{F}^n)$.

5. Let $V = \mathbb{R}^2$ and define $f \in \mathcal{B}(V)$ by

$$f(X, Y) = 2x_1y_1 - 3x_1y_2 + x_2y_2.$$

   (a) Find the matrix representation A of f relative to the basis $v_1 = (1, 0)$, $v_2 = (1, 1)$.

(b) Find the matrix representation $\bar{A}$ of f relative to the basis $\bar{v}_1 = (2, 1)$, $\bar{v}_2 = (1, -1)$.
(c) Find the transition matrix P from the basis $\{v_i\}$ to the basis $\{\bar{v}_i\}$ and verify that $\bar{A} = P^T AP$.

6. Let $V = M_n(\mathbb{C})$, and for all A, $B \in V$ define
$$f(A, B) = n \operatorname{tr}(AB) - (\operatorname{tr} A)(\operatorname{tr} B).$$
(a) Show this defines a bilinear form on V.
(b) Let $U \subset V$ be the subspace of traceless matrices. Show that f is degenerate, but that $f_U = f|U$ is nondegenerate.
(c) Let $W \subset V$ be the subspace of all traceless skew-Hermitian matrices A (i.e., $\operatorname{tr} A = 0$ and $A^\dagger = A^{*T} = -A$). Show $f_W = f|W$ is negative definite, i.e., that $f_W(A, A) < 0$ for all nonzero $A \in W$.
(d) Let $\widetilde{V} \subset V$ be the set of all matrices $A \in V$ with the property that $f(A, B) = 0$ for all $B \in V$. Show $\widetilde{V}$ is a subspace of V. Give an explicit description of $\widetilde{V}$ and find its dimension.

7.2 Symmetric and Antisymmetric Bilinear Forms

An extremely important type of bilinear form is one for which $f(u, u) = 0$ for all $u \in V$. Such forms are said to be **alternating**. If f is alternating, then for every u, $v \in V$ we have
$$\begin{aligned} 0 &= f(u + v, u + v) \\ &= f(u, u) + f(u, v) + f(v, u) + f(v, v) \\ &= f(u, v) + f(v, u) \end{aligned}$$
and hence
$$f(u, v) = -f(v, u).$$
A bilinear form that satisfies this condition is called **antisymmetric** (or **skew-symmetric**).

It is also worth pointing out the simple fact that the diagonal matrix elements of any representation of an alternating (or antisymmetric) bilinear form will necessarily be zero. This is because the diagonal elements are given by $a_{ii} = f(e_i, e_i) = 0$.

Theorem 7.4. *Let $f \in \mathcal{B}(V)$ be alternating. Then there exists a basis for V in which the matrix A of f takes the block diagonal form*
$$A = M \oplus \cdots \oplus M \oplus 0 \oplus \cdots \oplus 0$$
where 0 is the 1×1 matrix (0), and
$$M = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}.$$

Moreover, the number of blocks consisting of the matrix M is just $(1/2)\operatorname{rank}(f)$.

Proof. We first note the theorem is clearly true if $f = 0$. Next we note that if $\dim V = 1$, then any vector $v_i \in V$ is of the form $v_i = a_i u$ for some basis vector u and scalar a_i. Therefore, for any $v_1, v_2 \in V$ we have

$$f(v_1, v_2) = f(a_1 u, a_2 u) = a_1 a_2 f(u, u) = 0$$

so that again $f = 0$. We now assume $f \neq 0$ and $\dim V > 1$, and proceed by induction on $\dim V$. In other words, we assume the theorem is true for $\dim V < n$, and proceed to show it is also true for $\dim V = n$.

Since $\dim V > 1$ and $f \neq 0$, there exist nonzero vectors $u_1, u_2 \in V$ such that $f(u_1, u_2) \neq 0$. Moreover, we can always multiply u_1 by the appropriate scalar so that

$$f(u_1, u_2) = 1 = -f(u_2, u_1).$$

It is also true that u_1 and u_2 must be linearly independent because if $u_2 = ku_1$, then $f(u_1, u_2) = f(u_1, ku_1) = kf(u_1, u_1) = 0$.

We now define the two-dimensional subspace $U \subset V$ spanned by the vectors $\{u_1, u_2\}$. By definition, the matrix $(a_{ij}) \in M_2(\mathcal{F})$ of f restricted to U is given by $a_{ij} = f(u_i, u_j)$, and hence it is easy to see that (a_{ij}) is given by the matrix M defined in the statement of the theorem.

Since any $u \in U$ is of the form $u = au_1 + bu_2$, we see that

$$f(u, u_1) = af(u_1, u_1) + bf(u_2, u_1) = -b$$

and

$$f(u, u_2) = af(u_1, u_2) + bf(u_2, u_2) = a.$$

Now define the set

$$W = \{w \in V : f(w, u) = 0 \text{ for every } u \in U\}.$$

We claim that $V = U \oplus W$ (compare this with Theorem 1.22).

To show $U \cap W = \{0\}$, we assume $v \in U \cap W$. Then $v \in U$ has the form $v = \alpha u_1 + \beta u_2$ for some scalars α and β. But $v \in W$ so that $0 = f(v, u_1) = -\beta$ and $0 = f(v, u_2) = \alpha$, and hence $v = 0$. We now show $V = U + W$.

Let $v \in V$ be arbitrary, and define the vectors

$$\begin{aligned} u &= f(v, u_2)u_1 - f(v, u_1)u_2 \in U \\ w &= v - u. \end{aligned}$$

If we can show $w \in W$, then we will have shown that $v = u + w \in U + W$ as desired. But this is easy to do since we have

$$\begin{aligned} f(u, u_1) &= f(v, u_2)f(u_1, u_1) - f(v, u_1)f(u_2, u_1) = f(v, u_1) \\ f(u, u_2) &= f(v, u_2)f(u_1, u_2) - f(v, u_1)f(u_2, u_2) = f(v, u_2) \end{aligned}$$

and therefore we find that

$$\begin{aligned} f(w,u_1) &= f(v-u,u_1) = f(v,u_1) - f(u,u_1) = 0 \\ f(w,u_2) &= f(v-u,u_2) = f(v,u_2) - f(u,u_2) = 0. \end{aligned}$$

These equations show that $f(w,u) = 0$ for every $u \in U$, and thus $w \in W$. This completes the proof that $V = U \oplus W$, and hence it follows that $\dim W = \dim V - \dim U = n - 2 < n$.

Next we note that the restriction of f to W is just an alternating bilinear form on W and therefore, by our induction hypothesis, there exists a basis $\{u_3,\dots,u_n\}$ for W such that the matrix of f restricted to W has the desired form. But the matrix of V is the direct sum of the matrices of U and W, where the matrix of U was shown above to be M. Therefore $\{u_1,u_2,\dots,u_n\}$ is a basis for V in which the matrix of f has the desired form.

Finally, it should be clear that the rows of the matrix of f that are made up of the portion $M \oplus \cdots \oplus M$ are necessarily linearly independent (by definition of direct sum and the fact that the rows of M are independent). Since each M contains two rows, we see that $\operatorname{rank}(f) = \operatorname{rr}(f)$ is precisely twice the number of M matrices in the direct sum. ∎

Corollary 1. *Any nonzero alternating bilinear form must have even rank.*

Proof. Since the number of M blocks in the matrix of f is $(1/2)\operatorname{rank}(f)$, it follows that $\operatorname{rank}(f)$ must be an even number. ∎

Corollary 2. *If there exists a nondegenerate, alternating form on V, then $\dim V$ is even.*

Proof. This is Exercise 7.2.7. ∎

If $f \in \mathcal{B}(V)$ is alternating, then the matrix elements a_{ij} representing f relative to any basis $\{e_i\}$ for V are given by

$$a_{ij} = f(e_i,e_j) = -f(e_j,e_i) = -a_{ji}.$$

Any matrix $A = (a_{ij}) \in M_n(\mathcal{F})$ with the property that $a_{ij} = -a_{ji}$ (i.e., $A = -A^T$) is said to be **antisymmetric**. If we are given any element a_{ij} of an antisymmetric matrix, then we automatically know a_{ji}. Because of this, we say a_{ij} and a_{ji} are not **independent**. Since the diagonal elements of any such antisymmetric matrix must be zero, this means the maximum number of independent elements in A is given by $(n^2 - n)/2$. Therefore the subspace of $\mathcal{B}(V)$ consisting of nondegenerate alternating bilinear forms is of dimension $n(n-1)/2$.

Another extremely important class of bilinear forms on V is that for which $f(u, v) = f(v, u)$ for all u, $v \in V$. In this case we say that f is **symmetric**, and we have the matrix representation

$$a_{ij} = f(e_i, e_j) = f(e_j, e_i) = a_{ji}.$$

As expected, any matrix $A = (a_{ij})$ with the property that $a_{ij} = a_{ji}$ (i.e., $A = A^T$) is said to be **symmetric**. In this case, the number of independent elements of A is $[(n^2 - n)/2] + n = (n^2 + n)/2$, and hence the subspace of $\mathcal{B}(V)$ consisting of symmetric bilinear forms has dimension $n(n+1)/2$.

It is also easy to prove generally that a matrix $A \in M_n(\mathcal{F})$ represents a symmetric bilinear form on V if and only if A is a symmetric matrix. Indeed, if f is a symmetric bilinear form, then for all X, $Y \in V$ we have

$$X^T AY = f(X, Y) = f(Y, X) = Y^T AX.$$

But $X^T AY$ is just a 1×1 matrix, and hence $(X^T AY)^T = X^T AY$. Therefore (using Theorem 2.15) we have

$$Y^T AX = X^T AY = (X^T AY)^T = Y^T A^T X^{TT} = Y^T A^T X.$$

Since X and Y are arbitrary, this implies $A = A^T$. Conversely, suppose A is a symmetric matrix. Then for all X, $Y \in V$ we have

$$X^T AY = (X^T AY)^T = Y^T A^T X^{TT} = Y^T AX$$

so that A represents a symmetric bilinear form. The analogous result holds for antisymmetric bilinear forms as well (see Exercise 7.2.2).

Note that adding the dimensions of the symmetric and antisymmetric subspaces of $\mathcal{B}(V)$ we find

$$n(n-1)/2 + n(n+1)/2 = n^2 = \dim \mathcal{B}(V).$$

This should not be surprising since, for an arbitrary bilinear form $f \in \mathcal{B}(V)$ and any X, $Y \in V$, we can always write

$$f(X, Y) = (1/2)[f(X, Y) + f(Y, X)] + (1/2)[f(X, Y) - f(Y, X)].$$

In other words, any bilinear form can always be written as the sum of a symmetric and an antisymmetric bilinear form.

There is another particular type of form that is worth distinguishing. In particular, let V be finite-dimensional over $\mathcal{F}$, and let $f = \langle \cdot\, , \cdot \rangle$ be a symmetric bilinear form on V. We define the mapping $q : V \to \mathcal{F}$ by

$$q(X) = f(X, X) = \langle X, X \rangle$$

for every $X \in V$. The mapping q is called the **quadratic form associated** with the symmetric bilinear form f. It is clear that (by definition) q is represented by a symmetric matrix A, and hence it may be written in the alternative forms

$$q(X) = X^T AX = \sum_{i,j} a_{ij} x_i x_j = \sum_i a_{ii}(x_i)^2 + 2 \sum_{i<j} a_{ij} x_i x_j .$$

This expression for q in terms of the variables x_i is called the **quadratic polynomial** corresponding to the symmetric matrix A. In the case where A happens to be a diagonal matrix, then $a_{ij} = 0$ for $i \neq j$ and we are left with the simple form $q(X) = a_{11}(x_1)^2 + \cdots + a_{nn}(x_n)^2$. In other words, the quadratic polynomial corresponding to a diagonal matrix contains no "cross product" terms.

While we will show in the next section that every quadratic form has a diagonal representation, let us first look at a special case.

Example 7.3. Consider the *real* quadratic polynomial on $\mathbb{R}^n$ defined by

$$q(Y) = \sum_{i,j} b_{ij} y_i y_j$$

(where $b_{ij} = b_{ji}$ as usual for a quadratic form). If it happens that $b_{11} = 0$ but, for example, that $b_{12} \neq 0$, then we make the substitutions

$$\begin{aligned} y_1 &= x_1 + x_2 \\ y_2 &= x_1 - x_2 \\ y_i &= x_i \qquad\qquad \text{for } i = 3, \dots, n. \end{aligned}$$

A little algebra (which you should check) then shows that $q(Y)$ takes the form

$$q(Y) = \sum_{i,j} c_{ij} x_i x_j$$

where now $c_{11} \neq 0$. This means we can focus our attention on the case $q(X) = \sum_{i,j} a_{ij} x_i x_j$ where it is assumed that $a_{11} \neq 0$.

Thus, given the real quadratic form $q(X) = \sum_{i,j} a_{ij} x_i x_j$ where $a_{11} \neq 0$, let us make the substitutions

$$\begin{aligned} x_1 &= y_1 - (1/a_{11})[a_{12} y_2 + \cdots + a_{1n} y_n] \\ x_i &= y_i \quad \text{for each } i = 2, \dots, n. \end{aligned}$$

Some more algebra shows that $q(X)$ now takes the form

$$q(x_1, \dots, x_n) = a_{11}(y_1)^2 + q'(y_2, \dots, y_n)$$

where q' is a new quadratic polynomial. Continuing this process, we eventually arrive at a new set of variables in which q has a diagonal representation. This is called **completing the square**.

Given any quadratic form q, it is possible to fully recover the values of f from those of q. To show this, let u, $v \in V$ be arbitrary. Then

$$q(u + v) = \langle u + v, u + v \rangle$$

$$\begin{aligned}&= \langle u,u\rangle + \langle u,v\rangle + \langle v,u\rangle + \langle v,v\rangle \\ &= q(u) + 2f(u,v) + q(v)\end{aligned}$$

and therefore

$$f(u,v) = (1/2)[q(u+v) - q(u) - q(v)].$$

This equation is called the **polar form** of f.

Exercises

1. (a) Show that if f is a nondegenerate, antisymmetric bilinear form on V, then $n = \dim V$ is even.
   (b) Show there exists a basis for V in which the matrix of f takes the block matrix form

   $$\begin{bmatrix} 0 & D \\ -D & 0 \end{bmatrix}$$

   where D is the $(n/2)\times(n/2)$ matrix

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

2. Show that a matrix $A \in M_n(\mathcal{F})$ represents an antisymmetric bilinear form on V if and only if A is antisymmetric.

3. Reduce each of the following quadratic forms to diagonal form:
   (a) $q(x,y,z) = 2x^2 - 8xy + y^2 - 16xz + 14yz + 5z^2$.
   (b) $q(x,y,z) = x^2 - xz + y^2$.
   (c) $q(x,y,z) = xy + y^2 + 4xz + z^2$.
   (d) $q(x,y,z) = xy + yz$.

4. (a) Find all antisymmetric bilinear forms on $\mathbb{R}^3$.
   (b) Find a basis for the space of all antisymmetric bilinear forms on $\mathbb{R}^n$.

5. Let V be finite-dimensional over $\mathbb{C}$. Prove:
   (a) The equation

   $$(Ef)(u,v) = (1/2)[f(u,v) - f(v,u)]$$

   for every $f \in \mathcal{B}(V)$ defines a linear operator E on $\mathcal{B}(V)$.
   (b) E is a projection, i.e., $E^2 = E$.
   (c) If $T \in L(V)$, the equation

   $$(T^\dagger f)(u,v) = f(Tu,Tv)$$

   defines a linear operator $T^\dagger$ on $\mathcal{B}(V)$.

(d) $ET^\dagger = T^\dagger E$ for all $T \in \mathcal{B}(V)$.

6. Let V be finite-dimensional over $\mathbb{C}$, and suppose f, $g \in \mathcal{B}(V)$ are antisymmetric. Show there exists an invertible $T \in L(V)$ such that $f(Tu, Tv) = g(u, v)$ for all u, $v \in V$ if and only if f and g have the same rank.

7. Prove Corollary 2 of Theorem 7.4.

7.3 Diagonalization of Symmetric Bilinear Forms

We now turn to the diagonalization of quadratic forms. One obvious way to accomplish this in the case of a real quadratic form is to treat the matrix representation A of the quadratic form as the matrix representation of a linear operator. Then from the corollary to Theorem 5.18 we know that there exists an orthogonal matrix S that diagonalizes A so that $S^{-1}AS = S^TAS$ is diagonal. Hence from Theorem 7.3 it follows that this will represent the quadratic form in the new basis.

However, there is an important distinction between bilinear forms and linear operators. In order to diagonalize a linear operator we must find its eigenvalues and eigenvectors. This is possible because a linear operator is a mapping from a vector space V into itself, so the equation $Tv = \lambda v$ makes sense. But a bilinear form is a mapping from $V \times V \to \mathcal{F}$, and the concept of an eigenvalue has no meaning in this case.

We can avoid this issue because it is also possible to diagonalize a symmetric bilinear form using a nonorthogonal transformation. It is this approach that we follow in the proof of our next theorem. After the proof we will give an example that should clarify the algorithm that was described. Note that there is no requirement in this theorem that the bilinear form be real, and in fact it applies to a vector space over any field.

Theorem 7.5. *Let f be a symmetric bilinear form on a finite-dimensional space V. Then there exists a basis $\{e_i\}$ for V in which f is represented by a diagonal matrix. Alternatively, if f is represented by a (symmetric) matrix A in one basis, then there exists a nonsingular transition matrix P to the basis $\{e_i\}$ such that P^TAP is diagonal.*

Proof. Let the (symmetric) matrix representation of f be $A = (a_{ij}) \in M_n(\mathcal{F})$, and first assume $a_{11} \neq 0$. For each $i = 2, \ldots, n$ we multiply the ith row of A by a_{11}, and then add $-a_{i1}$ times the first row to this new ith row. In other words, this combination of two elementary row operations results in $A_i \to a_{11}A_i - a_{i1}A_1$. Following this procedure for each $i = 2, \ldots, n$ yields the first column of A in the form $A^1 = (a_{11}, 0, \ldots, 0)$ (remember this is a column vector, not a row vector). We now want to put the first row of A into the same form. However, this is easy because A is symmetric. We thus perform exactly the same operations (in the same sequence), but on columns instead of rows, resulting in $A^i \to a_{11}A^i - a_{i1}A^1$. Therefore the first row is also transformed into the form $A_1 = (a_{11}, 0, \ldots, 0)$. In other words, this sequence of operations results in the transformed A having the block matrix form

$$\begin{bmatrix} a_{11} & 0 \\ 0 & B \end{bmatrix}$$

where B is a matrix of size less than that of A. We can also write this in the form $(a_{11}) \oplus B$.

Now look carefully at what we did in the particular case of $i = 2$. Let us denote the multiplication operation by the elementary matrix E_m, and the addition operation by E_a (see Section 2.7). Then what was done in performing the row operations was simply to carry out the multiplication $(E_a E_m)A$. Next, because A is symmetric, we carried out exactly the same operations but applied to the columns instead of the rows. As we saw at the end of Section 2.7, this is equivalent to the multiplication $A(E_m^T E_a^T)$. In other words, for $i = 2$ we effectively carried out the multiplication

$$E_a E_m A E_m^T E_a^T.$$

For each succeeding value of i we then carried out this same procedure, and the final net effect on A was simply a multiplication of the form

$$E_s \cdots E_1 A E_1^T \cdots E_s^T$$

which resulted in the block matrix $(a_{11}) \oplus B$ shown above. Furthermore, note that if we let $S = E_1^T \cdots E_s^T = (E_s \cdots E_1)^T$, then $(a_{11}) \oplus B = S^T AS$ must be symmetric since $(S^T AS)^T = S^T A^T S = S^T AS$. This means that in fact the matrix B must also be symmetric.

We can now repeat this procedure on the matrix B and, by induction, we eventually arrive at a diagonal representation of A given by

$$D = E_r \cdots E_1 A E_1^T \cdots E_r^T$$

for some set of elementary row transformations E_i. But from Theorems 7.3 and 7.5, we know that $D = P^T AP$, and therefore P^T is given by the product $e_r \cdots e_1(I) = E_r \cdots E_1$ of elementary row operations applied to the identity matrix exactly as they were applied to A. It should be emphasized that we were able to arrive at this conclusion only because A is symmetric, thereby allowing each column operation to be the transpose of the corresponding row operation. Note however, that while the order of the row and column operations performed is important within their own group, the associativity of the matrix product allows the column operations (as a group) to be performed independently of the row operations.

We still must take into account the case where $a_{11} = 0$. If $a_{11} = 0$ but $a_{ii} \neq 0$ for some $i > 1$, then we can bring a_{ii} into the first diagonal position by interchanging the ith row and column with the first row and column respectively.

We then follow the procedure given above. If $a_{ii} = 0$ for every $i = 1, \dots, n$ then we can pick any $a_{ij} \neq 0$ and apply the operations $A_i \to A_i + A_j$ and $A^i \to A^i + A^j$. This puts $2a_{ij} \neq 0$ into the ith diagonal position, and allows us to proceed as in the previous case (which then goes into the first case treated). ∎

Example 7.4. Let us find the transition matrix P such that $D = P^T AP$ is diagonal, with A given by

$$\begin{bmatrix} 1 & -3 & 2 \\ -3 & 7 & -5 \\ 2 & -5 & 8 \end{bmatrix}.$$

We begin by forming the matrix $(A|I)$:

$$\left[\begin{array}{rrr|rrr} 1 & -3 & 2 & 1 & 0 & 0 \\ -3 & 7 & -5 & 0 & 1 & 0 \\ 2 & -5 & 8 & 0 & 0 & 1 \end{array}\right].$$

Now carry out the following sequence of elementary row operations to both A and I, and identical column operations to A only:

$$\begin{array}{r} \\ A_2 + 3A_1 \to \\ A_3 - 2A_1 \to \end{array} \left[\begin{array}{rrr|rrr} 1 & -3 & 2 & 1 & 0 & 0 \\ 0 & -2 & 1 & 3 & 1 & 0 \\ 0 & 1 & 4 & -2 & 0 & 1 \end{array}\right]$$

$$\begin{array}{c} \left[\begin{array}{rrr|rrr} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & -2 & 1 & 3 & 1 & 0 \\ 0 & 1 & 4 & -2 & 0 & 1 \end{array}\right] \\ \uparrow \quad \uparrow \\ A^2 + 3A^1 \quad A^3 - 2A^1 \end{array}$$

$$\begin{array}{r} \\ \\ 2A_3 + A_2 \to \end{array} \left[\begin{array}{rrr|rrr} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & -2 & 1 & 3 & 1 & 0 \\ 0 & 0 & 9 & -1 & 1 & 2 \end{array}\right]$$

$$\begin{array}{c} \left[\begin{array}{rrr|rrr} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & -2 & 0 & 3 & 1 & 0 \\ 0 & 0 & 18 & -1 & 1 & 2 \end{array}\right] \\ \uparrow \\ 2A^3 + A^2 \end{array}$$

We have thus diagonalized A, and the final form of the matrix $(A|I)$ is just $(D|P^T)$. Note also that

$$P^T AP = \begin{bmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ -1 & 1 & 2 \end{bmatrix} \begin{bmatrix} 1 & -3 & 2 \\ -3 & 7 & -5 \\ 2 & -5 & 8 \end{bmatrix} \begin{bmatrix} 1 & 3 & -1 \\ 0 & 1 & 1 \\ 0 & 0 & 2 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 18 \end{bmatrix} = D.$$

Since Theorem 7.5 tells us that every symmetric bilinear form has a diagonal representation, it follows that the associated quadratic form $q(X)$ has the **diagonal representation**

$$q(X) = X^T AX = a_{11}(x_1)^2 + \cdots + a_{nn}(x_n)^2$$

where A is the diagonal matrix representing the (symmetric) bilinear form.

Let us stop for a minute and look at what we have (or have not) actually done. If we treat the matrix representation of a real symmetric bilinear form as a real symmetric matrix A and diagonalize it by finding its eigenvectors, then we what we have done is construct an *orthogonal* matrix P whose columns are the eigenvectors of A, and is such that $P^{-1}AP = P^T AP$ is a diagonal matrix with diagonal entries that are just the eigenvalues of A. In other words, not only has the transformation $P^T AP$ diagonalized the bilinear form, the same transformation $P^{-1}AP$ has diagonalized the linear operator whose matrix representation was given by A with respect to some basis.

However, if we perform the *nonorthogonal* diagonalization as described in the proof of Theorem 7.5, then the resulting matrix $P^T AP$ has diagonal entries that are *not* the eigenvalues of A. Furthermore, since the matrix P is not orthogonal in general, the matrix $P^T AP$ is *not* the representation of any *linear operator* with respect to any basis. But given this P (which is nonsingular), we could construct the matrix $P^{-1}AP$ which would represent the linear operator in a new basis with transition matrix P, but this representation would not be diagonal.

Let us now specialize this discussion somewhat and consider only real symmetric bilinear forms. We begin by noting that in general, the diagonal representation of a symmetric bilinear form f has positive, negative, and zero entries. We can always renumber the basis vectors so that the positive entries appear first, followed by the negative entries and then the zero entries. It is in fact true, as we now show, that any other diagonal representation of f has the same number of positive and negative entries. If there are P positive entries and N negative entries, then the difference $S = P - N$ is called the **signature** of f.

Theorem 7.6. *Let $f \in \mathcal{B}(V)$ be a real symmetric bilinear form. Then every diagonal representation of f has the same number of positive and negative entries.*

Proof. Let $\{e_1, \ldots, e_n\}$ be the basis for V in which the matrix of f is diagonal (see Theorem 7.5). By suitably numbering the e_i, we may assume that the first

P entries are positive and the next N entries are negative (also note there could be $n-P-N$ zero entries). Now let $\{e'_1, \dots, e'_n\}$ be another basis for V in which the matrix of f is also diagonal. Again, assume the first P' entries are positive and the next N' entries are negative. Since the rank of f is just the rank of any matrix representation of f, and since the rank of a matrix is just the dimension of its row (or column) space, it is clear that $\operatorname{rank}(f) = P + N = P' + N'$. Because of this, we need only show that $P = P'$.

Let U be the linear span of the P vectors $\{e_1, \dots, e_P\}$, let W be the linear span of $\{e'_{P'+1}, \dots, e'_n\}$, and note $\dim U = P$ and $\dim W = n - P'$. Then for all nonzero vectors $u \in U$ and $w \in W$, we have $f(u,u) > 0$ and $f(w,w) \leq 0$ (this inequality is $\leq$ and not $<$ because if $P' + N' \neq n$, then the last of the basis vectors that span W will define a diagonal element in the matrix of f that is 0). Hence it follows that $U \cap W = \{0\}$, and therefore (by Theorem 1.11)

$$\begin{aligned} \dim(U+W) &= \dim U + \dim W - \dim(U \cap W) = P + n - P' - 0 \\ &= P - P' + n. \end{aligned}$$

Since U and W are subspaces of V, it follows that $\dim(U+W) \leq \dim V = n$, and therefore $P - P' + n \leq n$. This shows $P \leq P'$. Had we let U be the span of $\{e'_1, \dots, e'_{P'}\}$ and W be the span of $\{e_{P+1}, \dots, e_n\}$, we would have found that $P' \leq P$. Therefore $P = P'$ as claimed. ∎

While Theorem 7.5 showed that any quadratic form has a diagonal representation, the important special case of a real quadratic form allows an even simpler representation. This corollary is known as **Sylvester's theorem** (or the **law of inertia** or the **principal axis theorem**).

Corollary. *Let f be a real symmetric bilinear form. Then f has a unique diagonal representation of the form*

$$\begin{bmatrix} I_r & & \\ & -I_s & \\ & & 0_t \end{bmatrix}$$

where I_r and I_s are the $r \times r$ and $s \times s$ identity matrices, and 0_t is the $t \times t$ zero matrix. In particular, the associated quadratic form q has a representation of the form

$$q(x_1, \dots, x_n) = (x_1)^2 + \cdots + (x_r)^2 - (x_{r+1})^2 - \cdots - (x_{r+s})^2.$$

Proof. Let f be represented by a (real) symmetric $n \times n$ matrix A. By Theorem 7.6, there exists a nonsingular matrix P_1 such that $D = P_1^T A P_1 = (d_{ij})$ is a diagonal representation of f with a unique number r of positive entries followed by a unique number s of negative entries. We let $t = n - r - s$ be the unique number of zero entries in D.

Now let P_2 be the diagonal matrix with diagonal entries

$$(P_2)_{ii} = \begin{cases} 1/\sqrt{d_{ii}} & \text{for } i = 1, \dots, r \\ 1/\sqrt{-d_{ii}} & \text{for } i = r+1, \dots, r+s \\ 1 & \text{for } i = r+s+1, \dots, n \end{cases} .$$

Since P_2 is diagonal, it is obvious that $(P_2)^T = P_2$. We leave it to the reader to multiply out the matrices and show that

$$P_2^T D P_2 = P_2^T P_1^T A P_1 P_2 = (P_1 P_2)^T A (P_1 P_2)$$

is a congruence transformation that takes A into the desired form. ∎

We say that a real symmetric bilinear form $f \in \mathcal{B}(V)$ is **nonnegative** (or **positive semidefinite**) if

$$q(X) = X^T A X = \sum_{i,j} a_{ij} x_i x_j = f(X, X) \geq 0$$

for all $X \in V$, and we say that f is **positive definite** if $q(X) > 0$ for all nonzero $X \in V$. In particular, from Theorem 7.6 we see that f is nonnegative if and only if the signature $S = \operatorname{rank}(f) \leq \dim V$, and f will be positive definite if and only if $S = \dim V$.

Example 7.5. The quadratic form $(x_1)^2 - 4x_1 x_2 + 5(x_2)^2$ is positive definite because it can be written in the form

$$(x_1 - 2x_2)^2 + (x_2)^2$$

which is nonnegative for all real values of x_1 and x_2, and is zero only if $x_1 = x_2 = 0$.

The quadratic form $(x_1)^2 + (x_2)^2 + 2(x_3)^2 - 2x_1 x_3 - 2x_2 x_3$ can be written in the form

$$(x_1 - x_3)^2 + (x_2 - x_3)^2.$$

Since this is nonnegative for all real values of x_1, x_2 and x_3 but is zero for nonzero values (e.g., $x_1 = x_2 = x_3 \neq 0$), this quadratic form is nonnegative but not positive definite.

Exercises

1. Determine the rank and signature of the following real quadratic forms:
   (a) $x^2 + 2xy + y^2$.
   (b) $x^2 + xy + 2xz + 2y^2 + 4yz + 2z^2$.

2. Find the transition matrix P such that $P^T AP$ is diagonal where A is given by:

(a) $\begin{bmatrix} 1 & 2 & -3 \\ 2 & 5 & -4 \\ -3 & -4 & 8 \end{bmatrix}$ (b) $\begin{bmatrix} 0 & 1 & 1 \\ 1 & -2 & 2 \\ 1 & 2 & -1 \end{bmatrix}$

(c) $\begin{bmatrix} 1 & 1 & -2 & -3 \\ 1 & 2 & -5 & -1 \\ -2 & -5 & 6 & 9 \\ -3 & -1 & 9 & 11 \end{bmatrix}$

3. Let f be the symmetric bilinear form associated with the real quadratic form $q(x, y) = ax^2 + bxy + cy^2$. Show:
   (a) f is nondegenerate if and only if $b^2 - 4ac \neq 0$.
   (b) f is positive definite if and only if $a > 0$ and $b^2 - 4ac < 0$.

4. If A is a real, symmetric, positive definite matrix, show there exists a nonsingular matrix P such that $A = P^T P$.

The remaining exercises are all related.

5. Let V be finite-dimensional over $\mathbb{C}$, let S be the subspace of all symmetric bilinear forms on V, and let Q be the set of all quadratic forms on V.
   (a) Show Q is a subspace of all functions from V to $\mathbb{C}$.
   (b) Suppose $T \in L(V)$ and $q \in Q$. Show the equation $(T^\dagger q)(v) = q(Tv)$ defines a quadratic form $T^\dagger q$ on V.
   (c) Show the function $T^\dagger$ is a linear operator on Q, and show $T^\dagger$ is invertible if and only if T is invertible.

6. (a) Let q be the quadratic form on $\mathbb{R}^2$ defined by $q(x, y) = ax^2+2bxy+cy^2$ (where $a \neq 0$). Find an invertible $T \in L(\mathbb{R}^2)$ such that
$$(T^\dagger q)(x, y) = ax^2 + (c - b^2/a)y^2.$$
   [*Hint*: Complete the square to find T^{-1} (and hence T).]
   (b) Let q be the quadratic form on $\mathbb{R}^2$ defined by $q(x, y) = 2bxy$. Find an invertible $T \in L(\mathbb{R}^2)$ such that
$$(T^\dagger q)(x, y) = 2bx^2 - 2by^2.$$
   (c) Let q be the quadratic form on $\mathbb{R}^3$ defined by $q(x, y, z) = xy+2xz+z^2$. Find an invertible $T \in L(\mathbb{R}^3)$ such that
$$(T^\dagger q)(x, y, z) = x^2 - y^2 + z^2.$$

7. Suppose $A \in M_n(\mathbb{R})$ is symmetric, and define a quadratic form q on $\mathbb{R}^n$ by
$$q(X) = \sum_{i,j=1}^{n} a_{ij} x_i x_j.$$

Show there exists $T \in L(\mathbb{R}^n)$ such that

$$(T^\dagger q)(X) = \sum_{i=1}^{n} c_i(x_i)^2$$

where each c_i is either 0 or ± 1.

7.4 Hermitian Forms*

Let us now briefly consider how some of the results of the previous sections carry over to the case of bilinear forms over the complex number field.

We say that a mapping $f : V \times V \to \mathbb{C}$ is a **Hermitian form** on V if for all $u_1, u_2, v \in V$ and $a, b \in \mathbb{C}$ we have

(HF1) $f(au_1 + bu_2, v) = a^* f(u_1, v) + b^* f(u_2, v)$.
(HF2) $f(u_1, v) = f(v, u_1)^*$.

(We should point out that many authors define a Hermitian form by requiring that the scalars a and b on the right hand side of property (HF1) not be the complex conjugates as we have defined it. In this case, the scalars on the right hand side of property (HF3) below will be the complex conjugates of what we have shown.) As was the case for the Hermitian inner product (see Section 1.5), we see that

$$\begin{aligned} f(u, av_1 + bv_2) &= f(av_1 + bv_2, u)^* = [a^* f(v_1, u) + b^* f(v_2, u)]^* \\ &= af(v_1, u)^* + bf(v_2, u)^* = af(u, v_1) + bf(u, v_2) \end{aligned}$$

which we state as

(HF3) $f(u, av_1 + bv_2) = af(u, v_1) + bf(u, v_2)$.

Since $f(u, u) = f(u, u)^*$ it follows that $f(u, u) \in \mathbb{R}$ for all $u \in V$.

Along with a Hermitian form f is the **associated Hermitian quadratic form** $q : V \to \mathbb{R}$ defined by $q(u) = f(u, u)$ for all $u \in V$. A little algebra (Exercise 7.4.1) shows that f may be obtained from q by the **polar form** expression of f which is

$$f(u, v) = (1/4)[q(u + v) - q(u - v)] - (i/4)[q(u + iv) - q(u - iv)].$$

We also say that f is **nonnegative semidefinite** if $q(u) = f(u, u) \geq 0$ for all $u \in V$, and **positive definite** if $q(u) = f(u, u) > 0$ for all nonzero $u \in V$. For example, the usual Hermitian inner product on $\mathbb{C}^n$ is a positive definite form since for every nonzero $X = (x^1, \ldots, x^n) \in \mathbb{C}^n$ we have

$$q(X) = f(X, X) = \langle X, X \rangle = \sum_{i=1}^{n} (x^i)^* x^i = \sum_{i=1}^{n} \left|x^i\right|^2 > 0.$$

As we defined it in Section 5.7, we say that a matrix $H = (h_{ij}) \in M_n(\mathbb{C})$ is **Hermitian** if $h_{ij} = h_{ji}^*$. In other words, H is Hermitian if $H = H^{*T} := H^\dagger$. Note also that for any scalar k we have $k^\dagger = k^*$. Furthermore, using Theorem 2.15(iv), we see that

$$(AB)^\dagger = (AB)^{*T} = (A^*B^*)^T = B^\dagger A^\dagger.$$

By induction, this obviously extends to any finite product of matrices. It is also clear that

$$A^{\dagger\dagger} = A.$$

Example 7.6. Let H be a Hermitian matrix. We show that $f(X,Y) = X^\dagger HY$ defines a Hermitian form on $\mathbb{C}^n$.

Let X_1, X_2, $Y \in \mathbb{C}^n$ be arbitrary, and let a, $b \in \mathbb{C}$. Then (using Theorem 2.15(i))

$$\begin{aligned} f(aX_1 + bX_2, Y) &= (aX_1 + bX_2)^\dagger HY \\ &= (a^*X_1^\dagger + b^*X_2^\dagger)HY \\ &= a^*X_1^\dagger HY + b^*X_2^\dagger HY \\ &= a^*f(X_1,Y) + b^*f(X_2,Y) \end{aligned}$$

which shows that $f(X,Y)$ satisfies property (HF1) of a Hermitian form. Now, since $X^\dagger HY$ is a (complex) scalar we have $(X^\dagger HY)^T = X^\dagger HY$, and therefore

$$f(X,Y)^* = (X^\dagger HY)^* = (X^\dagger HY)^\dagger = Y^\dagger HX = f(Y,X)$$

where we used the fact that $H^\dagger = H$. Thus $f(X,Y)$ satisfies property (HF2), and hence defines a Hermitian form on $\mathbb{C}^n$.

It is probably worth pointing out that $X^\dagger HY$ will not be a Hermitian form if the alternative definition mentioned above is used. In this case, one must use $f(X,Y) = X^T HY^*$ (see Exercise 7.4.2).

Now let V have basis $\{e_i\}$, and let f be a Hermitian form on V. Then for any $X = \sum x_i e_i$ and $Y = \sum y_i e_i$ in V, we see that

$$f(X,Y) = f\left(\sum_i x_i e_i, \sum_j y_j e_j\right) = \sum_{i,j} x_i^* y_j f(e_i, e_j).$$

Just as we did in Theorem 7.1, we define the matrix elements h_{ij} representing a Hermitian form f by $h_{ij} = f(e_i, e_j)$. Note that since $f(e_i, e_j) = f(e_j, e_i)^*$, we see the diagonal elements of $H = (h_{ij})$ must be real. Using this definition for the matrix elements of f we then have

$$f(X,Y) = \sum_{i,j} x_i^* h_{ij} y^j = X^\dagger HY.$$

Following the proof of Theorem 7.1, this shows that any Hermitian form f has a unique representation in terms of the Hermitian matrix H. If we want to make explicit the basis referred to in this expression, we write $f(X,Y) = [X]_e^\dagger H[Y]_e$ where it is understood that the elements h_{ij} are defined with respect to the basis $\{e_i\}$.

Finally, let us prove the complex analogues of Theorems 7.3 and 7.6.

Theorem 7.7. *Let f be a Hermitian form on V, and let P be the transition matrix from a basis $\{e_i\}$ for V to a new basis $\{\bar{e}_i\}$. If H is the matrix of f with respect to the basis $\{e_i\}$ for V, then $\overline{H} = P^\dagger HP$ is the matrix of f relative to the new basis $\{\bar{e}_i\}$.*

Proof. We saw in the proof of Theorem 7.3 that for any $X \in V$ we have $[X]_e = P[X]_{\bar{e}}$, and hence $[X]_e^\dagger = [X]_{\bar{e}}^\dagger P^\dagger$. Therefore, for any $X, Y \in V$ we see that

$$f(X,Y) = [X]_e^\dagger H[Y]_e = [X]_{\bar{e}}^\dagger P^\dagger HP[Y]_{\bar{e}} = [X]_{\bar{e}}^\dagger \overline{H}[Y]_{\bar{e}}$$

where $\overline{H} = P^\dagger HP$ is the (unique) matrix of f relative to the basis $\{\bar{e}_i\}$. ∎

Theorem 7.8. *Let f be a Hermitian form on V. Then there exists a basis for V in which the matrix of f is diagonal, and every other diagonal representation of f has the same number of positive and negative entries.*

Proof. Using the fact that $f(u,u)$ is real for all $u \in V$ along with the appropriate polar form of f, it should be easy for the reader to follow the proofs of Theorems 7.5 and 7.6 and complete the proof of this theorem (see Exercise 7.4.3). ∎

We note that because of this result, our earlier definition for the signature of a bilinear form applies equally well to Hermitian forms.

Exercises

1. Let f be a Hermitian form on V and q the associated quadratic form. Verify the polar form

$$f(u,v) = (1/4)[q(u+v) - q(u-v)] - (i/4)[q(u+iv) - q(u-iv)].$$

2. Verify the statement made at the end of Example 7.6.

3. Prove Theorem 7.8.

4. Show that the algorithm described in Section 7.3 applies to Hermitian matrices if we allow multiplication by complex numbers and, instead of multiplying by E^T on the right, we multiply by E^{*T}.

5. For each of the following Hermitian matrices H, use the results of the previous exercise to find a nonsingular matrix P such that $P^T HP$ is diagonal:

(a) $\begin{bmatrix} 1 & i \\ -i & 2 \end{bmatrix}$ (b) $\begin{bmatrix} 1 & 2+3i \\ 2-3i & -1 \end{bmatrix}$

(c) $\begin{bmatrix} 1 & i & 2+i \\ -i & 2 & 1-i \\ 2-i & 1+i & 2 \end{bmatrix}$ (d) $\begin{bmatrix} 1 & 1+i & 2i \\ 1-i & 4 & 2-3i \\ -2i & 2+3i & 7 \end{bmatrix}$

7.5 Simultaneous Diagonalization*

We now want to investigate how we might simultaneously diagonalize two real quadratic forms in the case where at least one of them is positive definite. This is directly applicable to problems such as finding the normal modes of a system of coupled oscillators, as we will show. In fact, we will look at several ways to go about this simultaneous diagonalization.

For our first approach, we start with the forms $X^T AX$ and $X^T BX$ where both A and B are real symmetric matrices and A is positive definite. Then, by the corollary to Theorem 5.18, we diagonalize A by an orthogonal transformation P so that $P^{-1}AP = P^T AP = D$ where D is the diagonal matrix $\operatorname{diag}(\lambda_1, \dots, \lambda_n)$ consisting of the eigenvalues of A. Under this change of basis, $X \to \overline{X} = P^{-1}X$ or $X = P\overline{X}$ (Theorem 4.14), and hence

$$X^T AX = \overline{X}^T P^T AP\overline{X} = \overline{X}^T D\overline{X} = \lambda_1(\bar{x}_1)^2 + \cdots + \lambda_n(\bar{x}_n)^2.$$

Since A is positive definite, each $\lambda_i > 0$ and we may define the nonorthogonal matrix $Q = \operatorname{diag}(1/\sqrt{\lambda_1}, \dots, 1/\sqrt{\lambda_n}) = Q^T$ so that $Q^T DQ = I$, $\overline{X} \to \widetilde{X} = Q^{-1}\overline{X}$ or $\overline{X} = Q\widetilde{X}$ and

$$X^T AX = \overline{X}^T D\overline{X} = \widetilde{X}^T Q^T DQ\widetilde{X} = \widetilde{X}^T I\widetilde{X} = (\tilde{x}_1)^2 + \cdots + (\tilde{x}_n)^2.$$

(This is the same result as Sylvester's theorem.)

Now, what happens to B under these basis changes? We just compute:

$$X^T BX = \overline{X}^T P^T BP\overline{X} = \widetilde{X}^T Q^T P^T BPQ\widetilde{X} = \widetilde{X}^T \widetilde{B}\widetilde{X}$$

where $\widetilde{B} = (PQ)^T B(PQ)$ is symmetric because B is symmetric. We can therefore diagonalize $\widetilde{B}$ by an orthogonal transformation R so that $\widetilde{B} \to R^{-1}\widetilde{B}R = R^T\widetilde{B}R = \widetilde{D} = \operatorname{diag}(\mu_1, \dots, \mu_n)$ where the μ_i are the eigenvalues of $\widetilde{B}$. Under the transformation R we have $Y = R^{-1}\widetilde{X}$ or $\widetilde{X} = RY$ and

$$X^T BX = \widetilde{X}^T \widetilde{B}\widetilde{X} = Y^T R^T \widetilde{B}RY = Y^T \widetilde{D}Y = \mu_1(y_1)^2 + \cdots + \mu_n(y_n)^2.$$

Finally, we have to look at what this last transformation R does to A. But this is easy because R is orthogonal and A is already diagonal so that

$$X^T AX = \widetilde{X}^T I\widetilde{X} = Y^T R^T IRY = Y^T Y = (y_1)^2 + \cdots + (y_n)^2.$$

In other words, the nonorthogonal transformation $X = P\overline{X} = PQ\widetilde{X} = (PQR)Y$ diagonalizes both A and B.

Now recall that the μ_i were the eigenvalues of $\widetilde{B} = (PQ)^T B(PQ)$, and hence are determined by solving the secular equation $\det(\widetilde{B} - \mu I) = 0$. But using $(PQ)^T A(PQ) = Q^T DQ = I$ we can write this as

$$\begin{aligned} 0 = \det(\widetilde{B} - \mu I) &= \det[(PQ)^T B(PQ) - \mu(PQ)^T A(PQ)] \\ &= \det[(PQ)^T (B - \mu A)(PQ)] \\ &= [\det(PQ)]^2 \det(B - \mu A). \end{aligned}$$

But PQ is nonsingular so $\det(PQ) \neq 0$ and we are left with

$$\det(B - \mu A) = 0 \tag{7.1}$$

as the defining equation for the μ_i.

The second way we can simultaneously diagonalize A and B is to first perform a nonorthogonal diagonalization of A by using elementary row transformations as in the proof of Theorem 7.5. Since A is positive definite, Sylvester's theorem shows that there is a nonsingular matrix P with $P^T AP = I$ and where $\overline{X} = P^{-1}X$ or $X = P\overline{X}$ such that

$$X^T AX = \overline{X}^T P^T AP\overline{X} = \overline{X}^T I\overline{X} = (\bar{x}_1)^2 + \cdots + (\bar{x}_n)^2.$$

Note that even though this is the same result as in the first (orthogonal) approach, the diagonalizing matrices are different.

Using this P, we see that $\overline{B} = P^T BP$ is also symmetric and hence can be diagonalized by an orthogonal matrix Q where $Y = Q^{-1}\overline{X}$ or $\overline{X} = QY$. Then $Q^{-1}\overline{B}Q = Q^T\overline{B}Q = \overline{D} = \operatorname{diag}(\mu_1, \ldots, \mu_n)$ where the μ_i are the eigenvalues of $\overline{B}$. Under these transformations we have

$$\begin{aligned} X^T AX &= \overline{X}^T P^T AP\overline{X} = \overline{X}^T I\overline{X} = Y^T Q^T IQY = Y^T Y \\ &= (y_1)^2 + \cdots + (y_n)^2 \end{aligned}$$

and

$$\begin{aligned} X^T BX &= \overline{X}^T P^T BP\overline{X} = Y^T Q^T P^T BPQY = Y^T Q^T \overline{B}QY = Y^T \overline{D}Y \\ &= \mu_1(y_1)^2 + \cdots + \mu_n(y_n)^2 \end{aligned}$$

so that the transformation $X \to Y = (PQ)^{-1}X$ diagonalizes both A and B.

To find the eigenvalues μ_i we must solve $0 = \det(\overline{B} - \mu I) = \det(P^T BP - \mu I) = \det(P^T BP - \mu P^T AP) = \det[P^T(B - \mu I)P] = (\det P)^2 \det(B - \mu A)$ so that again we have

$$\det(B - \mu A) = 0.$$

Either way, we have proved most of the next theorem.

Theorem 7.9. *Let $X^T AX$ and $X^T BX$ be two real quadratic forms on an n-dimensional Euclidean space V, and assume that $X^T AX$ is positive definite. Then there exists a nonsingular matrix P such that the transformation $X = PY$ reduces $X^T AX$ to the form*

$$X^T AX = Y^T Y = (y_1)^2 + \cdots + (y_n)^2$$

and $X^T BX$ to the form

$$X^T BX = Y^T DY = \mu_1(y_1)^2 + \cdots + \mu_n(y_n)^2$$

where $\mu_1, \dots, \mu_n$ are roots of the equation

$$\det(B - \mu A) = 0.$$

Moreover, the μ_i are real and positive if and only if $X^T BX$ is positive definite.

Proof. In view of the above discussion, all that remains is to prove the last statement of the theorem. Since B is a real symmetric matrix, there exists an orthogonal matrix S that brings it into the form

$$S^T BS = \operatorname{diag}(\lambda_1, \dots, \lambda_n) = \widetilde{D}$$

where the λ_i are the eigenvalues of B. Writing $X = SY$, we see that

$$X^T BX = Y^T S^T BSY = Y^T \widetilde{D} Y = \lambda_1(y_1)^2 + \cdots + \lambda_n(y_n)^2$$

and thus $X^T BX$ is positive definite if and only if $Y^T \widetilde{D} Y$ is positive definite, i.e., if and only if every $\lambda_i > 0$. Since we saw above that

$$(PQ)^T B(PQ) = \operatorname{diag}(\mu_1, \dots, \mu_n) = \overline{D}$$

it follows from Theorem 7.6 that the number of positive λ_i must equal the number of positive μ_i. Therefore $X^T BX$ is positive definite if and only if every $\mu_i > 0$. ∎

Example 7.7. Let us show how Theorem 7.9 can be of help in classical mechanics. This rather long example requires a knowledge of both the Lagrange equations of motion and Taylor series expansions. The details of the physics are given in any modern text on classical mechanics. Our purpose is simply to demonstrate the usefulness of this theorem.

We will first solve the general problem of coupled oscillators undergoing small oscillations using the standard techniques for systems of differential equations. After that, we will show how the same solution can be easily written down based on the results of Theorem 7.9.

Consider the small oscillations of a conservative system of N particles about a point of stable equilibrium. We assume the position $\mathbf{r}_i$ of the ith particle is a function of n generalized coordinates q_i, and not explicitly on the time t. Thus we write $\mathbf{r}_i = \mathbf{r}_i(q_1, \dots, q_n)$, and

$$\dot{\mathbf{r}}_i = \frac{d\mathbf{r}_i}{dt} = \sum_{j=1}^{n} \frac{\partial \mathbf{r}_i}{\partial q_j} \dot{q}_j \qquad i = 1, \dots, N$$

where we denote the derivative with respect to time by a dot.

Since the velocity v_i of the ith particle is given by $\|\dot{\mathbf{r}}_i\|$, the kinetic energy T of the ith particle is $(1/2)m_i {v_i}^2 = (1/2)m_i \dot{\mathbf{r}}_i \cdot \dot{\mathbf{r}}_i$, and hence the kinetic energy of the system of N particles is given by

$$T = \frac{1}{2} \sum_{i=1}^{N} m_i \dot{\mathbf{r}}_i \cdot \dot{\mathbf{r}}_i := \frac{1}{2} \sum_{j,k=1}^{n} M_{jk} \dot{q}_j \dot{q}_k$$

where

$$M_{jk} = \sum_{i=1}^{N} m_i \frac{\partial \mathbf{r}_i}{\partial q_j} \cdot \frac{\partial \mathbf{r}_i}{\partial q_k} = M_{kj}$$

is a function of the coordinates q.

We assume that the equilibrium position of each q_i is at $q_i = 0$ and expand in a Taylor series:

$$M_{ij}(q_1, \dots, q_n) = M_{ij}(0) + \sum_{k=1}^{n} \left(\frac{\partial M_{ij}}{\partial q_k} \right)_0 q_k + \cdots .$$

Since we are considering only small oscillations, we will work to second order in the coordinates. Because T has two factors of $\dot{q}$ in it already, we need keep only the first (constant) term in this expansion. Then denoting the *constant* $M_{ij}(0)$ by $m_{ij} = m_{ji}$ we have

$$T = \frac{1}{2} \sum_{i,j=1}^{n} m_{ij} \dot{q}_i \dot{q}_j \tag{7.2}$$

so that T is a quadratic form in the $\dot{q}_i$'s. Moreover, since $T > 0$ we must have that in fact T is a positive definite quadratic form.

Let the potential energy of the system be $V = V(q_1, \dots, q_n)$. Expanding V in a Taylor series expansion about the equilibrium point (the minimum of potential energy), we have

$$V(q_1, \dots, q_n) = V(0) + \sum_{i=1}^{n} \left(\frac{\partial V}{\partial q_i} \right)_0 q_i + \frac{1}{2} \sum_{i,j=1}^{n} \left(\frac{\partial^2 V}{\partial q_i \partial q_j} \right)_0 q_i q_j + \cdots .$$

Since we are expanding about the minimum we must have $(\partial V / \partial q_i)_0 = 0$ for every i. Furthermore, we may shift the zero of potential and assume $V(0) = 0$

because this has no effect on the force on each particle. To second order we may therefore write the potential as the quadratic form

$$V = \frac{1}{2} \sum_{i,j=1}^{n} v_{ij} q_i q_j \tag{7.3}$$

where the $v_{ij} = v_{ji}$ are *constants*. Furthermore, each $v_{ij} > 0$ because we are expanding about a minimum, and hence V is also a positive definite quadratic form in the q_i's.

The Lagrange equations of motion are

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}_i}\right) - \frac{\partial L}{\partial q_i} = 0 \tag{7.4}$$

where $L = T - V$ is called the **Lagrangian**. Written out, it is

$$L = T - V = \frac{1}{2} \sum_{j,k=1}^{n} (m_{jk} \dot{q}_j \dot{q}_k - v_{jk} a_j q_k). \tag{7.5}$$

Since T is a function of the $\dot{q}_i$'s and V is a function of the q_i's, the equations of motion take the form

$$\frac{d}{dt}\left(\frac{\partial T}{\partial \dot{q}_i}\right) + \frac{\partial V}{\partial q_i} = 0. \tag{7.6}$$

In particular, using the above expressions for T and V the equations of motion become

$$\sum_{k=1}^{n} (m_{ik} \ddot{q}_k + v_{ik} q_k) = 0. \tag{7.7}$$

This is a set of n coupled second order differential equations. As we stated above, let's first look at the standard approach to solving this system, because it's a good application of several results we have covered so far in this book. (We follow the treatment in the excellent text by Fetter and Walecka [12].)

To begin, we change from q_k to a new complex variable z_k where the physical solution is then given by $q_k = \operatorname{Re} z_k$:

$$\sum_{k=1}^{n} (m_{ik} \ddot{z}_k + v_{ik} z_k) = 0. \tag{7.8}$$

We then look for solutions of the form $z_k = z_k^0 e^{i\omega t}$ where all n coordinates oscillate at the same frequency. Such solutions are called **normal modes**. Substituting these into equation (7.8) yields

$$\sum_{k=1}^{n} (v_{ik} - \omega^2 m_{ik}) z_k^0 = 0.$$

In order for this system of linear equations to have a nontrivial solution, we must have

$$\det(v_{ik} - \omega^2 m_{ik}) = 0. \tag{7.9}$$

This characteristic equation is an nth order polynomial in ω^2, and we denote the n roots by ω_s^2, $s = 1, \dots, n$. Each ω_s is called a **normal frequency**, and the corresponding nontrivial solution $z_k^{(s)}$ satisfies

$$\sum_{k=1}^{n} (v_{ik} - \omega_s^2 m_{ik}) z_k^{(s)} = 0. \tag{7.10}$$

We now show that each ω_s^2 is both real and greater than or equal to zero.

To see this, first multiply equation (7.10) by $z_i^{(s)*}$ and sum over i to solve for ω_s^2:

$$\omega_s^2 = \frac{\sum_{ik} z_i^{(s)*} v_{ik} z_k^{(s)}}{\sum_{ik} z_i^{(s)*} m_{ik} z_k^{(s)}}.$$

But v_{ik} and m_{ik} are both real and symmetric, so using the fact that i and k are dummy indices, we see that taking the complex conjugate of this equation shows that $(\omega_s^2)^* = \omega_s^2$ as claimed.

Next, there are two ways to see that $\omega_s^2 \geq 0$. First, we note that if $\omega_s^2 < 0$, then ω_s would contain an imaginary part, and $z_k = z_k^{(s)} e^{i\omega_s t}$ would then either grow or decay exponentially, neither of which leads to a stable equilibrium. Alternatively, simply note that both v_{ik} and m_{ik} are positive definite quadratic forms, and hence $\omega_s^2 \geq 0$.

The next thing we can now show is that because all of the coefficients in equation (7.10) are real, it follows that for each s the ratio $z_k^{(s)}/z_n^{(s)}$ is real for every $k = 1, \dots, n-1$. This is just a consequence of Cramer's rule. To see this, consider a general homogeneous linear system

$$\begin{aligned} a_{11}x_1 + \cdots + a_{1n}x_n &= 0 \\ &\vdots \\ a_{n1}x_1 + \cdots + a_{nn}x_n &= 0 \end{aligned}$$

In order to have a nontrivial solution for $\mathbf{x}$ we must have $\det A = \det(a_{ij}) = 0$. But then $\operatorname{rank} A < n$ so the rows of A are linearly dependent. Let's say that the nth equation is a linear combination of the preceding $n-1$ equations. Then we can disregard it and assume that the remaining $n-1$ equations have at least one nonzero component, say x_n. Dividing by x_n we can write our system as

$$\begin{aligned} a_{11}(x_1/x_n) + \cdots + \quad a_{1\,n-1}(x_{n-1}/x_n) &= \quad -a_{1n} \\ &\vdots \\ a_{n-1\,1}(x_1/x_n) + \cdots + a_{n-1\,n-1}(x_{n-1}/x_n) &= -a_{n-1\,n} \end{aligned}$$

This is now $n-1$ inhomogeneous equations in $n-1$ unknowns, and Cramer's rule lets us solve for the ratios x_k/x_n. Since all of the a_{ij} are real, these ratios are necessarily real by the construction process of Cramer's rule.

So, since $z_k^{(s)}/z_n^{(s)}$ is real for each s, any complex constant can only appear as an overall multiplicative factor independent of k. This means we can write the solution to equation (7.10) in the form

$$z_k^{(s)} = e^{i\phi_s} r_k^{(s)} \qquad k = 1, \ldots, n \tag{7.11}$$

where ϕ_s is real and one component (say $r_n^{(s)}$) can be arbitrarily specified. Using equation (7.11) in equation (7.10) and canceling the common phase $e^{i\phi_s}$ gives us

$$\sum_{k=1}^{n} v_{ik} r_k^{(s)} = \omega_s^2 \sum_{k=1}^{n} m_{ik} r_k^{(s)} \qquad i = 1, \ldots, n. \tag{7.12}$$

We will refer to ω_s as an eigenvalue, and $\mathbf{r}^{(s)}$ as the corresponding eigenvector. Note that this yields the same equation for $\mathbf{r}^{(s)}$ for both $\pm\omega_s$.

Now write equation (7.12) for the tth eigenvalue, multiply it by $r_i^{(s)}$ and sum over i:

$$\sum_{ik} r_i^{(s)} v_{ik} r_k^{(t)} = \omega_t^2 \sum_{ik} r_i^{(s)} m_{ik} r_k^{(t)}$$

Multiply equation (7.12) by $r_i^{(t)}$, sum over i and subtract it from the previous equation to obtain

$$(\omega_t^2 - \omega_s^2) \sum_{ik} r_i^{(t)} m_{ik} r_k^{(s)} = 0.$$

If we assume that $\omega_s^2 \neq \omega_t^2$ for $s \neq t$, then this equation implies

$$\sum_{ik} r_i^{(t)} m_{ik} r_k^{(s)} = 0 \qquad \text{for } s \neq t$$

which is an orthogonality relation between the eigenvectors $\mathbf{r}^{(s)}$ and $\mathbf{r}^{(t)}$. Furthermore, since equation (7.12) only determines $n-1$ real ratios $r_k^{(s)}/r_n^{(s)}$, we are free to multiply all components $z_k^{(s)}$ in equation (7.11) by a common factor that depends only on s (and not k). Then we can choose our solutions to be orthonormal and we have the normalization condition

$$\sum_{ik} r_i^{(t)} m_{ik} r_k^{(s)} = \delta_{st}. \tag{7.13}$$

The solutions in equation (7.11) now become

$$z_k^{(s)} = C^{(s)} e^{i\phi_s} r_k^{(s)} \tag{7.14}$$

where $r_k^{(s)}$ is fixed by the normalization, so $C^{(s)}$ and ϕ_s are the only real parameters that we can still arbitrarily specify. Note also that there is one of these equations for $\omega = +\omega_s$ and one for $\omega = -\omega_s$.

One other remark. We assumed above that $\omega_s^2 \neq \omega_t^2$ for $s \neq t$. If there are in fact repeated roots, then we must apply the Gram-Schmidt process to each eigenspace separately. But be sure to realize that the inner product used must be that defined by equation (7.13).

The general solution to equation (7.8) is a superposition of the solutions (7.14):

$$z_k(t) = \sum_{s=1}^{n} \Big[(z_+^{(s)})_k e^{i\omega_s t} + (z_-^{(s)})_k e^{-i\omega_s t} \Big] \tag{7.15}$$

where

$$(z_\pm^{(s)})_k := C_\pm^{(s)} e^{i\phi_s^\pm} r_k^{(s)} \quad \text{for } k = 1, \dots, n.$$

This solution is a linear combination of all normal modes labeled by s, and the subscript k labels the coordinate under consideration.

Recall that the physical solution is given by $q_k = \operatorname{Re} z_k$. Then redefining the complex number

$$(z_+^{(s)})_k + (z_-^{(s)})_k^* := z_k^{(s)} := C^{(s)} r_k^{(s)} e^{i\phi_s}$$

we have (since $C_\pm^{(s)}$ and $r_k^{(s)}$ are real)

$$\begin{aligned}
q_k(t) = \operatorname{Re} z_k(t) &= \frac{1}{2}(z_k(t) + z_k(t)^*) \\
&= \frac{1}{2} \sum_{s=1}^{n} \Big[(z_+^{(s)})_k e^{i\omega_s t} + (z_-^{(s)})_k e^{-i\omega_s t} + (z_+^{(s)})_k^* e^{-i\omega_s t} + (z_-^{(s)})_k^* e^{i\omega_s t} \Big] \\
&= \frac{1}{2} \sum_{s=1}^{n} \Big[C^{(s)} r_k^{(s)} e^{i\phi_s} e^{i\omega_s t} + C^{(s)} r_k^{(s)} e^{-i\phi_s} e^{-i\omega_s t} \Big] \\
&= \sum_{s=1}^{n} \operatorname{Re}(C^{(s)} r_k^{(s)} e^{i(\omega_s t + \phi_s)})
\end{aligned}$$

and therefore

$$q_k(t) = \sum_{s=1}^{n} C^{(s)} r_k^{(s)} \cos(\omega_s t + \phi_s). \tag{7.16}$$

This is the most general solution to Lagrange's equations for small oscillations about a point of stable equilibrium. Each q_k is the displacement from equilibrium of the kth generalized coordinate, and is given as an expansion in terms of the kth component of the eigenvector $\mathbf{r}^{(s)}$. The coefficients $C^{(s)} \cos(\omega_s t + \phi_s)$ of this expansion are called the **normal coordinates**. We will come back to them shortly.

Let us recast our results using vector and matrix notation. If we define the "potential energy matrix" $V = (v_{ij})$ and the "mass matrix" $M = (m_{ij})$, then the eigenvalue equation (7.9) may be written

$$\det(V - \omega^2 M) = 0 \tag{7.17}$$

and equation (7.12) for the sth eigenvector $r_k^{(s)}$ becomes

$$(V - \omega_s^2 M)\mathbf{r}^{(s)} = 0. \tag{7.18}$$

Using the matrix M we can define an inner product on our solution space, and the orthonormality condition (equation (7.13)) may be written

$$\langle \mathbf{r}^{(s)}, \mathbf{r}^{(t)} \rangle := {\mathbf{r}^{(s)}}^T M \mathbf{r}^{(t)} = \delta_{st}. \tag{7.19}$$

Finally, the general solution (7.16) takes the form

$$\mathbf{q} = \sum_{s=1}^{n} C^{(s)} \mathbf{r}^{(s)} \cos(\omega_s t + \phi_s). \tag{7.20}$$

Other than finding the integration constants $C^{(s)}$ and ϕ_s by specifying the initial conditions of a specific problem, this completes the standard solution to a system of coupled oscillators. We now turn to Theorem 7.9 and study its application to this problem.

Notice that equation (7.17) is just an example of what we denoted by $\det(B - \mu A) = 0$ in Theorem 7.9. If we use the eigenvectors $\mathbf{r}^{(s)}$ defined by equation (7.18) to construct a transition matrix P (i.e., let the sth column of P be $\mathbf{r}^{(s)}$), then equation (7.19) shows that the congruence transformation $P^T MP$ diagonalizes M. (This is the analogue of diagonalizing A in the proof of Theorem 7.9.)

To see this in detail, note that $p_{ij} = i$th entry of the jth column $= r_i^{(j)}$ so that

$$\begin{aligned} (P^T MP)_{ij} &= \sum_{kl} p_{ik}^T m_{kl} p_{lj} = \sum_{kl} p_{ki} m_{kl} p_{lj} \\ &= \sum_{kl} r_k^{(i)} m_{kl} r_l^{(j)} = \delta_{ij} \quad \text{by equation (7.13)} \end{aligned}$$

and therefore

$$P^T MP = I \tag{7.21}$$

where the sth column of P is the eigenvector $\mathbf{r}^{(s)}$ corresponding to the sth normal mode defined by ω_s. In other words, the transition matrix P (sometimes called the **modal matrix**) diagonalizes the mass matrix M.

Again looking back at what we did with two quadratic forms, we should find that P also diagonalizes the potential energy matrix V, and the diagonal elements should be the eigenvalues ω_s^2. That this is indeed the case is straightforward to show:

$$\begin{aligned} (P^T VP)_{ij} &= \sum_{kl} p_{ik}^T v_{kl} p_{lj} = \sum_{kl} p_{ki} v_{kl} p_{lj} \\ &= \sum_{kl} r_k^{(i)} v_{kl} r_l^{(j)} \quad \text{now use equation (7.12)} \\ &= \sum_{kl} r_k^{(i)} \omega_j^2 m_{kl} r_l^{(j)} = \omega_j^2 \sum_{kkl} r_k^{(i)} m_{kl} r_l^{(j)} \\ &= \omega_j^2 \delta_{ij} \quad \text{by equation (7.13).} \end{aligned}$$

In other words,

$$P^T V P = \begin{bmatrix} \omega_1^2 & & 0 \\ & \ddots & \\ 0 & & \omega_n^2 \end{bmatrix} := D_\omega. \tag{7.22}$$

Thus we have shown that, as expected, the transition matrix P simultaneously diagonaizes both of the quadratic forms M and V. We could do this because at least one of them was positive definite.

What else do we know about the transition matrix? It takes us from our original basis (essentially the standard basis on $\mathbb{R}^n$) to the basis of eigenvectors $\{\mathbf{r}^{(s)}\}$. Since the $\mathbf{q}(t)$'s are time dependent linear combinations of the $\mathbf{r}^{(s)}$'s (see equation (7.20)), let us define new coordinates $\mathbf{q}'(t)$ in the usual way by

$$\mathbf{q}'(t) = P^{-1}\mathbf{q}(t) \tag{7.23}$$

or $\mathbf{q}(t) = P\mathbf{q}'(t)$ (see Theorem 4.14). From equation (7.21) we see that $(P^T M)P = I$ which implies that $P^{-1} = P^T M$, and hence

$$\mathbf{q}'(t) = (P^T M)\mathbf{q}(t). \tag{7.24}$$

Since P and M are constant, real matrices, we have simply changed to a new basis that is a linear combination of the original generalized coordinates $\mathbf{q}$.

Now let's see what the Lagrangian looks like in these new coordinates. Writing equation (7.5) in matrix notation we have

$$L = \frac{1}{2}(\dot{\mathbf{q}}^T M \dot{\mathbf{q}} - \mathbf{q}^T V \mathbf{q}). \tag{7.25}$$

We use $\mathbf{q}(t) = P\mathbf{q}'(t)$ to write this in terms of the new coordinates as (remember P is a *constant* matrix)

$$L = \frac{1}{2}(\dot{\mathbf{q}}'^T P^T M P \dot{\mathbf{q}}' - \mathbf{q}'^T P^T V P \mathbf{q}').$$

But from equations (7.21) and (7.22) we know that P diagonalizes both M and V so we are left with the simple form

$$L = \frac{1}{2}(\dot{\mathbf{q}}'^T \dot{\mathbf{q}}' - \mathbf{q}'^T D_\omega \mathbf{q}'). \tag{7.26}$$

In terms of components this is

$$L = \frac{1}{2}\sum_{k=1}^{n}(\dot{q}_k'^2 - \omega_k^2 q_k'^2) \tag{7.27}$$

and we see that the new coordinates have diagonalized the Lagrangian. Furthermore, knowing $\mathbf{q}'$ we know $\mathbf{q} = P\mathbf{q}'$ which are the original generalized coordinates (i.e., the displacements from equilibrium). This means that we can take q'_k, $k = 1, \ldots, n$ as new generalized coordinates, called the **normal coordinates**.

We write the Lagrange equations (7.4) in terms of the q'_i and use equation (7.27) to obtain

$$\ddot{q}'_k + \omega_k^2 q'_k = 0. \tag{7.28}$$

This shows that the diagonalized Lagrangian leads to equations of motion that have become *decoupled*, so our original problem of n *coupled* second order differential equations has now become a set of n independent uncoupled simple harmonic oscillators. Each normal coordinate q'_k oscillates independently with an angular frequency ω_k (the normal mode frequency). Note also that equations (7.27) and (7.28) are very general – they hold for any system undergoing small oscillations about a point of static equilibrium.

The solution to equation (7.28) is

$$q'_k(t) = C^{(k)} \cos(\omega_k t + \phi_k)$$

or

$$\mathbf{q}'(t) = \begin{bmatrix} C^{(1)} \cos(\omega_1 t + \phi_1) \\ \vdots \\ C^{(n)} \cos(\omega_n t + \phi_n) \end{bmatrix}. \tag{7.29}$$

From $\mathbf{q} = P\mathbf{q}'$ (or $q_k = \sum_l p_{kl} q'_l$) and the definition of P (i.e., $p_{kl} = r_k^{(l)}$) we have

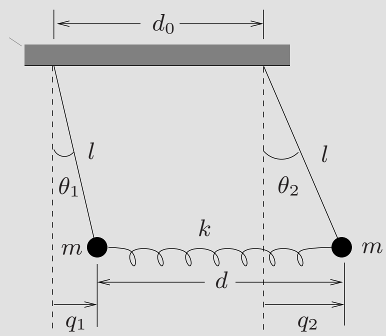
$$q_k(t) = \sum_{l=1}^{n} r_k^{(l)} C^{(l)} \cos(\omega_l t + \phi_l) = \sum_{l=1}^{n} q'_l r_k^{(l)} \tag{7.30}$$

which is the same as equation (7.16) as it should be. Note that the normal coordinates are just the coefficients (or amplitudes) of the eigenvectors $\mathbf{r}^{(s)}$ in the expansion of $\mathbf{q}(t)$.

Summarizing, we start from the Lagrangian (7.25) written in terms of the coordinates $\mathbf{q}$. This consists of two real quadratic forms, and we use them to construct the eigenvalue equation (7.18). We solve for the eigenvalues ω_s using equation (7.17), and then find the eigenvectors $\mathbf{r}^{(s)}$ using equation (7.18) again. These eigenvectors are normalized according to equation (7.19), and these normalized eigenvectors are used to construct the transition matrix P. Defining new coordinates $\mathbf{q}'(t) = P^{-1}\mathbf{q}(t)$ (equation (7.23)), the Lagrangian becomes equation (7.26), and the Lagrange equations of motion (7.28) have the solution (7.29). Finally, converting back to the $\mathbf{q}$'s from the $\mathbf{q}'$'s we have the solutions given in equations (7.30).

Of course, all we really have to do is find the eigenvalues and eigenvectors, because we have solved the problem in full generality and we know that the solution is simply given by equation (7.30).

Example 7.8. Let us work out a specific example. Consider two identical pendulums moving in a common plane as shown below. Assume that at equilibrium the pendulums hang vertically with the spring at its natural length d_0. Each is of mass m and length l, and they are connected by a spring of force constant k. We let their separation at any instant be d, and the generalized coordinates be the horizontal displacements q_i. Note that if $\theta_1 \neq \theta_2$, the two masses do not lie on a horizontal line. Since we are considering small oscillations only, to second order in θ we have $\sin\theta_i \approx \theta_i$ so that $q_i = l\sin\theta_i \approx l\theta_i$ or $\theta_i \approx q_i/l$.



The kinetic energy is easy to find. Each mass has velocity $l\dot{\theta}_i \approx \dot{q}_i$ so that

$$T = \frac{1}{2}m(\dot{q}_1^2 + \dot{q}_2^2).$$

However, the potential energy is somewhat more difficult.

Take the origin of our coordinate system to be at the pivot point of the first mass, with the y-axis pointing down and the x-axis to the right. The total potential energy of the system is due to both gravity and the spring. Defining the gravitational potential energy to be zero when the pendulums are at their lowest point, each pendulum has gravitational potential energy given by

$$V_{\text{grav}} = mgl(1 - \cos\theta_i) \approx mgl\frac{\theta_i^2}{2} = \frac{mg}{2l}q_i^2.$$

(This is because gravitational potential energy is defined by $\mathbf{F} = -\nabla V = mg\hat{y}$. Integrating from l to $y = l\cos\theta_i$ and taking $V(l) = 0$ gives $V = mgl(1 - \cos\theta)$. Then using $\cos\theta \approx 1 - \theta^2/2$ yields the above result.)

The potential energy of the spring is $V_{\text{spring}} = (1/2)k(d - d_0)^2$, so we must find d as function of the q_i's. From the figure, it is fairly clear that to first order in the q_i's we have $d = d_0 + (q_2 - q_1)$, but we can calculate it as follows. The (x, y) coordinates of the first mass are $(l\sin\theta_1, l\cos\theta_1)$ and the second are $(d_0 + l\sin\theta_2, l\cos\theta_2)$. Using the Pythagorean theorem we find

$$\begin{aligned} d^2 &= (d_0 + l\sin\theta_2 - l\sin\theta_1)^2 + (l\cos\theta_2 - l\cos\theta_1)^2 \\ &= d_0^2 + 2l^2[1 - (\sin\theta_1\sin\theta_2 + \cos\theta_1\cos\theta_2)] + 2d_0(l\sin\theta_2 - l\sin\theta_1) \end{aligned}$$

$$\begin{aligned}
&= d_0^2 + 2l^2[1 - \cos(\theta_1 - \theta_2)] + 2d_0(q_2 - q_1) \\
&= d_0^2 + l^2(\theta_1 - \theta_2)^2 + 2d_0(q_2 - q_1) \\
&= (d_0 + q_2 - q_1)^2
\end{aligned}$$

and therefore $d = d_0 + q_2 - q_1$ as expected. The total potential energy of our system is then

$$\begin{aligned}
V = V_{\text{grav}} + V_{\text{spring}} &= \frac{mg}{2l}(q_1^2 + q_2^2) + \frac{1}{2}k(q_2 - q_1)^2 \\
&= \frac{1}{2}\Big(\frac{mg}{l} + k\Big)(q_1^2 + q_2^2) - \frac{1}{2}k(q_1 q_2 + q_2 q_1).
\end{aligned}$$

We now have both T and V in the form of equations (7.2) and (7.3), and we can immediately write down the mass and potential energy matrices

$$M = \begin{bmatrix} m & 0 \\ 0 & m \end{bmatrix} \qquad V = \begin{bmatrix} mg/l + k & -k \\ -k & mg/l + k \end{bmatrix}.$$

Now we solve equation (7.17) for the normal frequencies ω^2.

$$\begin{aligned}
\det(V - \omega^2 M) &= \begin{vmatrix} mg/l + k - m\omega^2 & -k \\ -k & mg/l + k - m\omega^2 \end{vmatrix} \\
&= (mg/l + k - m\omega^2)^2 - k^2 \\
&= m^2(\omega^2)^2 - 2m\omega^2(mg/l + k) + (mg/l + k)^2 - k^2 = 0.
\end{aligned}$$

This is actually very easy to solve for ω^2 using the quadratic formula, and the result is $\omega^2 = (g/l + k/m) \pm k/m$ or

$$\omega_1 = (g/l)^{1/2} \quad \text{and} \quad \omega_2 = (g/l + 2k/m)^{1/2}.$$

The next step is to use these normal frequencies in equation (7.18) to find the eigenvectors $\mathbf{r}^{(s)}$. For ω_1 we have

$$\det(V - \omega_1^2 M)\mathbf{r}^{(1)} = \begin{vmatrix} k & -k \\ -k & k \end{vmatrix} \begin{bmatrix} r_1^{(1)} \\ r_2^{(1)} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

so that $kr_1^{(1)} - kr_2^{(1)} = 0$ or $r_1^{(1)} = r_2^{(1)}$. (Note that since $\det(V - \omega^2 M) = 0$, the rows are linearly dependent, and we need only look at one row for each ω_s^2.) In this mode, both masses have the same amplitude and move together, so the spring plays no role at all.

For the second mode ω_2 we have

$$\det(V - \omega_2^2 M)\mathbf{r}^{(2)} = \begin{vmatrix} -k & -k \\ -k & -k \end{vmatrix} \begin{bmatrix} r_1^{(2)} \\ r_2^{(2)} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

so that $r_1^{(2)} = -r_2^{(2)}$. Now the pendulums move in opposite directions but with equal amplitudes. In this mode the spring plays a definite role, which should be obvious because ω_2 depends on k while ω_1 does not.

Since the eigenvalue equation (7.18) is homogeneous, we are free to multiply each $\mathbf{r}^{(s)}$ by a constant N, and we normalize them according to equation (7.19). This gives

$$\mathbf{r}^{(1)T} M \mathbf{r}^{(1)} = N^2 \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} m & 0 \\ 0 & m \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = N^2(2m) = 1$$

so that $N = 1/\sqrt{2m}$. The vector $\mathbf{r}^{(2)}$ gives the same result so we have the normalized eigenvectors

$$\mathbf{r}^{(1)} = \frac{1}{\sqrt{2m}} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad \text{and} \quad \mathbf{r}^{(2)} = \frac{1}{\sqrt{2m}} \begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

We leave it to the reader to verify that $\mathbf{r}^{(1)T} M \mathbf{r}^{(2)} = \mathbf{r}^{(2)T} M \mathbf{r}^{(1)} = 0$.

The transition (or modal) matrix P is given by

$$P = \frac{1}{\sqrt{2m}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

and we also leave it to you to verify that P diagonalizes both M and V:

$$P^T M P = I \quad \text{and} \quad P^T V P = \begin{bmatrix} g/l & 0 \\ 0 & g/l + 2k/m \end{bmatrix} = D_\omega.$$

Using equation (7.24), the normal coordinates $\mathbf{q}'(t)$ defined by $\mathbf{q}'(t) = P^{-1}\mathbf{q}(t)$ are given by

$$\begin{aligned} \mathbf{q}'(t) = P^T M \mathbf{q}(t) &= \frac{m}{\sqrt{2m}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} q_1(t) \\ q_2(t) \end{bmatrix} \\ &= \sqrt{\frac{m}{2}} \begin{bmatrix} q_1(t) + q_2(t) \\ q_1(t) - q_2(t) \end{bmatrix} \end{aligned}$$

where, from equation (7.29),

$$\mathbf{q}'(t) = \begin{bmatrix} q_1'(t) \\ q_2'(t) \end{bmatrix} = \begin{bmatrix} C^{(1)} \cos(\omega_1 t + \phi_1) \\ C^{(2)} \cos(\omega_2 t + \phi_2) \end{bmatrix}.$$

To determine the four constants $C^{(i)}$ and ϕ_i we must specify the initial conditions for the problem. In other words, we need to specify $\mathbf{q}(0)$ and $\dot{\mathbf{q}}(0)$ so that

$$\mathbf{q}'(0) = \begin{bmatrix} C^{(1)} \cos\phi_1 \\ C^{(2)} \cos\phi_2 \end{bmatrix} = \sqrt{\frac{m}{2}} \begin{bmatrix} q_1(0) + q_2(0) \\ q_1(0) - q_2(0) \end{bmatrix}$$

and

$$\dot{\mathbf{q}}'(0) = \begin{bmatrix} -\omega_1 C^{(1)} \sin\phi_1 \\ -\omega_2 C^{(2)} \sin\phi_2 \end{bmatrix} = \sqrt{\frac{m}{2}} \begin{bmatrix} \dot{q}_1(0) + \dot{q}_2(0) \\ \dot{q}_1(0) - \dot{q}_2(0) \end{bmatrix}.$$

Let us assume that at $t = 0$ the left pendulum is displaced a distance α to the right and released from rest. In other words, $q_1(0) = \alpha$ and $q_2(0) = \dot{q}_1(0) = \dot{q}_2(0) = 0$. The equation for $\dot{\mathbf{q}}'(0)$ yields

$$\dot{\mathbf{q}}'(0) = \begin{bmatrix} -\omega_1 C^{(1)} \sin\phi_1 \\ -\omega_2 C^{(2)} \sin\phi_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

so that $\phi_1 = \phi_2 = 0$. Now we can use the equation for $\mathbf{q}'(0)$ to obtain

$$\begin{bmatrix} C^{(1)} \\ C^{(2)} \end{bmatrix} = \sqrt{\frac{m}{2}} \begin{bmatrix} \alpha \\ \alpha \end{bmatrix}$$

and hence $C^{(1)} = C^{(2)} = \alpha\sqrt{m/2}$.

Using these constants, the normal coordinates are now

$$\mathbf{q}'(t) = \alpha\sqrt{\frac{m}{2}} \begin{bmatrix} \cos\omega_1 t \\ \cos\omega_2 t \end{bmatrix}$$

and the complete time dependent solution for $\mathbf{q}(t)$ is

$$\mathbf{q}(t) = P\mathbf{q}'(t) = \frac{\alpha}{2} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} \cos\omega_1 t \\ \cos\omega_2 t \end{bmatrix}$$

or

$$\mathbf{q}(t) = \begin{bmatrix} q_1(t) \\ q_2(t) \end{bmatrix} = \frac{\alpha}{2} \begin{bmatrix} \cos\omega_1 t + \cos\omega_2 t \\ \cos\omega_1 t - \cos\omega_2 t \end{bmatrix}. \tag{7.31}$$

While this is the solution to our problem, we can put it into a form that makes it much easier to see what is going on.

Start from the trigonometric identities

$$\cos(a \pm b) = \cos a \cos b \mp \sin a \sin b$$

and

$$\sin(a \pm b) = \sin a \cos b \pm \cos a \sin b.$$

Noting that $a = (1/2)(a+b) + (1/2)(a-b)$ and $b = (1/2)(a+b) - (1/2)(a-b)$ we can write

$$\begin{aligned} \cos a \pm \cos b &= \cos\left[\frac{1}{2}(a+b) + \frac{1}{2}(a-b)\right] \pm \cos\left[\frac{1}{2}(a+b) - \frac{1}{2}(a-b)\right] \\ &= \begin{cases} 2\cos\frac{1}{2}(a+b)\cos\frac{1}{2}(a-b) \\ -2\sin\frac{1}{2}(a+b)\sin\frac{1}{2}(a-b) \end{cases} \end{aligned}$$

Applying this to our solution $\mathbf{q}(t)$ we have the final form of our solution

$$\mathbf{q}(t) = \begin{bmatrix} q_1(t) \\ q_2(t) \end{bmatrix} = \begin{bmatrix} \left[\alpha \cos \frac{1}{2}(\omega_2 - \omega_1)t\right] \cos \frac{1}{2}(\omega_2 + \omega_1)t \\ \left[\alpha \sin \frac{1}{2}(\omega_2 - \omega_1)t\right] \sin \frac{1}{2}(\omega_2 + \omega_1)t \end{bmatrix}. \tag{7.32}$$

To understand what this means physically, let us consider the limit of a weak spring, i.e., $k \ll mg/l$. In this case ω_2 is only slightly greater than ω_1, and the sin and cos of $\omega_2 - \omega_1$ vary very slowly compared to the sin and cos of $\omega_2 + \omega_1$. This means that the terms in brackets in equation (7.32) act like very slowly varying amplitudes of the much more rapidly oscillating second terms. This is the phenomena of **beats**.

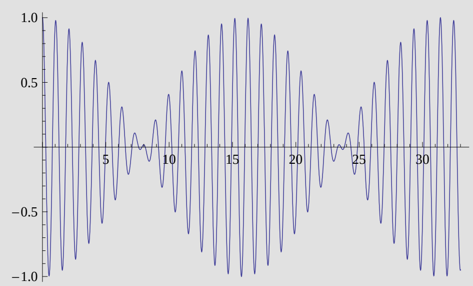


Figure 7.1: A plot of $q_1(t) = [\cos 0.2t] \cos 6t$

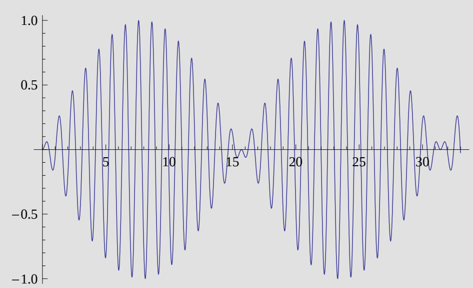


Figure 7.2: A plot of $q_2(t) = [\sin 0.2t] \sin 6t$

The amplitude of the first pendulum goes to 0 after a time $\tau = \pi/(\omega_2 - \omega_1)$ at which time the second pendulum achieves its maximum amplitude. Then the situation reverses, and the oscillations vary periodically. If the motion is strictly periodic, then we have $q_1(0) = q_1(t_p)$ for some time t_p, and therefore from equation (7.31) we must have $2 = \cos \omega_1 t_p + \cos \omega_2 t_p$. But the only way to achieve this is to have both cosines equal to 1, so $\omega_1 t_p = 2\pi n$ and $\omega_2 t_p = 2\pi m$ for some integers n and m. Therefore $\omega_2/\omega_1 = m/n$ which is a rational number. This means that if ω_2/ω_1 is an irrational number, then despite the apparent regularity we see with the beats, the system never returns exactly to its initial configuration.

The last thing we will cover in this section is yet another way to simultaneously diagonalize two real, symmetric bilinear forms. From a practical standpoint this approach won't give us a different method of solving the problem, but it is an interesting application of many of the ideas we have covered in this text.

First recall the distinction between a linear transformation and a bilinear form. As we have pointed out, it makes no sense to talk about the eigenvectors of a bilinear form because it is a mapping from $V \times V \to \mathcal{F}$ and not from $V \to V$. However, we will show that it is possible to combine both bilinear forms into an operator on V that can be diagonalized in the usual manner.

Given a vector space V, the dual space V^* is the space of linear functionals on V. Then given a real, positive definite symmetric bilinear form $A : V \times V \to \mathbb{R}$, there is an associated mapping $\widetilde{A} : V \to V^*$ defined by $\widetilde{A}(v) = A(v, \cdot) \in V^*$ for all $v \in V$. The fact that A is positive definite means $A(v, v) = 0$ if and only if $v = 0$, and therefore $\operatorname{Ker} \widetilde{A} = \{0\}$ so that $\widetilde{A}$ is one-to-one and $\widetilde{A}^{-1}$ exists.

Similarly, given another real, symmetric bilinear form B, we define $\widetilde{B} : V \to V^*$ by $\widetilde{B}(v) = B(v, \cdot)$. Then the composite mapping $\widetilde{A}^{-1} \circ \widetilde{B} : V \to V$ is a linear transformation on V. We will write $\widetilde{A}^{-1}\widetilde{B}$ for simplicity.

$$\begin{tikzcd} V \arrow[r, bend left, "\widetilde{A}"] \arrow[r, bend right, "\widetilde{B}"'] & V^* \arrow[l, bend right=70, "\widetilde{A}^{-1}"'] \end{tikzcd}$$

We now use A to define an inner product on V by $\langle u, v\rangle := A(u, v)$. Using this inner product, we can apply the Gram-Schmidt process to any basis for V and construct a new orthonormal basis $\{e_i\}$ such that

$$\langle e_i, e_j\rangle = A(e_i, e_j) = \delta_{ij}$$

and thus we have diagonalized A to the identity matrix. Relative to this basis, the bilinear form B has matrix elements $B(e_i, e_j) = b_{ij} = b_{ji}$.

Since $\widetilde{A}^{-1}\widetilde{B} : V \to V$, given any $v \in V$ there is some $u \in V$ such that $(\widetilde{A}^{-1}\widetilde{B})v = u$. Acting on both sides of this equation with $\widetilde{A}$ gives $\widetilde{B}v = \widetilde{A}u$ or $B(v, \cdot) = A(u, \cdot)$, i.e., $A(u, w) = B(v, w)$ for all $w \in V$. So given v, we would like to find the corresponding u.

To accomplish this, we use the fact that A and B are bilinear and consider the special case $v = e_i$ and $w = e_j$. Writing $u = \sum_k u_k e_k$ we have

$$A(u, w) = A\Big(\sum_k u_k e_k, e_j\Big) = \sum_k u_k A(e_k, e_j) = \sum_k u_k \delta_{kj} = u_j$$

and
$$B(v,w) = B(e_i,e_j) = b_{ij}$$
so that $A(u,w) = B(v,w)$ implies $u_j = b_{ij} = b_{ji}$. This shows that
$$(\widetilde{A}^{-1}\widetilde{B})e_i = \sum_j e_j b_{ji}$$
and hence the operator $\widetilde{A}^{-1}\widetilde{B}$ has matrix representation (b_{ij}) which is symmetric. Note also that from Section 4.3 we know that if $(\widetilde{A}^{-1}\widetilde{B})x = \bar{x}$, then $\bar{x}_i = \sum_j b_{ij}x_j$.

So, since $\widetilde{A}^{-1}\widetilde{B}$ is a real, symmetric operator it can be diagonalized by an orthogonal transformation, and we let its eigenvectors be denoted by f_i, i.e., $(\widetilde{A}^{-1}\widetilde{B})f_i = \lambda_i f_i$. In addition, by the corollary to Theorem 5.18 we know that the eigenvalues are real.

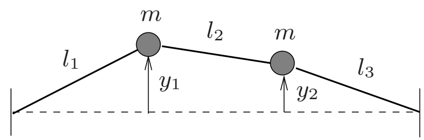
Furthermore, the eigenvectors belonging to distinct eigenvalues are orthogonal even with the inner product we have defined. This follows either from the discussion at the end of Section 5.7 or from Theorem 6.9. Alternatively, we can show directly that this is true because $A(e_i,e_j) = \langle e_i,e_j\rangle = \delta_{ij}$ so that letting $\widetilde{C} = \widetilde{A}^{-1}\widetilde{B}$ for simplicity we have (for any $x,y \in V$)
$$\begin{aligned}\langle \widetilde{C}x,y\rangle &= \sum_{ij}(\widetilde{C}x)_i y_j\langle e_i,e_j\rangle = \sum_{ij}(\widetilde{C}x)_i y_j\delta_{ij} = \sum_i(\widetilde{C}x)_i y_i \\ &= \sum_{ij} b_{ij}x_j y_i = \sum_{ij} x_j b_{ji} y_i = \sum_j x_j(\widetilde{C}y)_j \\ &= \langle x,\widetilde{C}y\rangle.\end{aligned}$$
Therefore, if $\widetilde{C}x = \lambda x$ and $\widetilde{C}y = \mu y$, then on the one hand $\langle\widetilde{C}x,y\rangle = \lambda\langle x,y\rangle$ while on the other hand $\langle x,\widetilde{C}y\rangle = \mu\langle x,y\rangle$. Subtracting these equations and using the above result we have $(\lambda-\mu)\langle x,y\rangle = 0$ so that $\langle x,y\rangle = 0$ if $\lambda \neq \mu$.

What we have then shown is that $(\widetilde{A}^{-1}\widetilde{B})f_i = \lambda_i f_i$ where $\langle f_i,f_j\rangle = A(f_i,f_j) = \delta_{ij}$ so that A is still diagonal with respect to the f_i's. But then $\widetilde{B}f_i = \lambda_i\widetilde{A}f_i$ or $B(f_i,\cdot) = \lambda_i A(f_i,\cdot)$ which means that $B(f_i,f_j) = \lambda_i A(f_i,f_j) = \lambda_i\delta_{ij}$ and B is also diagonal with respect to the f_i's.

Note that the λ_i are solutions to $\det(\widetilde{A}^{-1}\widetilde{B} - \lambda 1) = 0$ or $\det(b_{ij} - \lambda\delta_{ij}) = 0$ where $b_{ij} = B(e_i,e_j)$ and the e_i satisfy $\langle e_i,e_j\rangle = A(e_i,e_j) = \delta_{ij}$ as determined by the Gram-Schmidt process. Finally, since $\delta_{ij} = A(e_i,e_j) = a_{ij}$ we see that the λ_i are solutions of the equation $\det(b_{ij} - \lambda a_{ij}) = 0$ exactly as in our previous two approaches.

Exercises

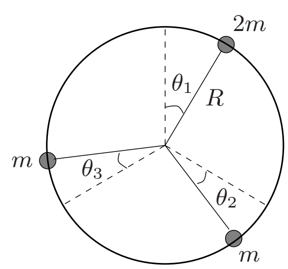
1. Two equal masses m move on a frictionless *horizontal* table. They are held by three identical taut strings of equilibrium length l_0 and *constant* tension τ as shown below.



At equilibrium, everything lies in a straight line between the fixed endpoints. Assume that the masses move in the y-direction only.

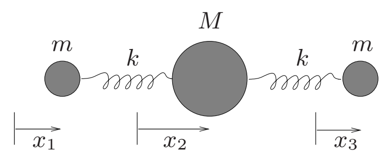
(a) Write down the Lagrangian in the case of small displacements, i.e., $y_i \ll l_0$. [*Hint*: The string tension is the force along the string, and potential energy is equivalent to work done, which is force times distance. Considering only small displacements makes the assumption of constant tension more physically realistic.]
(b) Find the normal frequencies and normal modes of oscillation. Describe the motion.
(c) Find the normalized eigenvectors $\mathbf{r}^{(s)}$ and show that they are orthonormal. Construct the transition matrix P, and show that $P^T V P$ is diagonal with diagonal elements that are precisely the normal frequencies.
(d) Find the normal coordinates $\mathbf{q}' = P^{-1}\mathbf{q} = P^T M\mathbf{q}$.

2. Three beads of masses m, m and $2m$ slide frictionlessly on a hoop of radius R. They are connected by three identical (curved) springs each with force constant k. Let θ_i denote the displacements from equilibrium of each of the three masses as shown in the figure below. (The equilibrium positions are obviously 120° apart. Again, the lines shown inside the hoop are imaginary lines used to label the generalized coordinates of the masses.)



(a) Write down the Lagrangian for the system.
(b) Find the normal frequencies and normalized eigenvectors. (It is easiest to define $\omega_0 = \sqrt{k/m}$ so that $k = m\omega_0^2$. Also, note that there is no small angle approximation in this problem.)
(c) Show that the transition matrix P diagonalizes the potential energy matrix V, and find the normal coordinates $\mathbf{q}' = P^T M\mathbf{q}$.
(d) Suppose at $t = 0$ we have $\dot\theta_1(0) = \dot\theta_2(0) = \dot\theta_3(0) = 0$, and $\theta_2(0) = \theta_3(0) = 0$ while $\theta_1(0) = \alpha$. Find the complete time-dependent solutions for the $\theta_i(t)$'s and describe the resulting motion in reasonable detail.

3. Consider the model of a linear triatomic molecule (such as CO_2) shown below. We have two equal masses m connected to a mass M by equal springs with force constant k. The displacements from equilibrium are the x_i as shown. Assume that the motion takes place in only one dimension.



   (a) Write down the Lagrangian for the system.
   (b) Find the normal frequencies in terms of the constants $\omega_0 = \sqrt{k/m}$ and $\lambda = M/m$. Note that one of the normal frequencies is zero.
   (c) Find the normal modes for the nonzero frequencies and describe them.
   (d) What is the motion of the zero frequency mode? [*Hint*: First look at what the eigenvalue equation says for zero frequency. Now look at what the equation of motion for the normal coordinate says, and then look at the relationship between the normal coordinates q'_k and the generalized coordinates q_k.]

Chapter 8

Multilinear Mappings and Tensors

The concept of a tensor is something that seems to take on a certain mystique to students who haven't yet studied them. However, the basic idea is quite simple. To say it briefly, a tensor is a scalar-valued mapping that takes as its argument several vectors, and is linear in each of them individually. While the notation can get rather cumbersome because of the number of vector spaces involved, that's basically all there is to it. However, to make these definitions as general and useful as we need, we will first have to go back and study dual spaces somewhat more carefully.

We assume that the reader has studied Sections 3.1 and 3.2, and we will use the notation and results from those sections freely. In addition, throughout this chapter we will generally be using the summation convention as discussed in Section 3.1. However, to help ease into the subject, we will also frequently stick with the standard summation symbols.

8.1 Bilinear Functionals and the Dual Space

Recall from the end of Section 4.1 and Example 7.2 that the vector space $V^* = L(V, \mathcal{F}) : V \to \mathcal{F}$ is defined to be the space of linear functionals on V. In other words, if $\phi \in V^*$, then for every u, $v \in V$ and a, $b \in \mathcal{F}$ we have

$$\phi(au + bv) = a\phi(u) + b\phi(v) \in \mathcal{F}.$$

The space V^* is called the **dual space** of V. If V is finite-dimensional, then viewing $\mathcal{F}$ as a one-dimensional vector space (over $\mathcal{F}$), it follows from Theorem 4.4 that $\dim V^* = \dim V$. In particular, given a basis $\{e_i\}$ for V, the proof of Theorem 4.4 showed that a unique basis $\{\omega^i\}$ for V^* is defined by the requirement that

$$\omega^i(e_j) = \delta^i_j$$

where we now use superscripts to denote basis vectors in the dual space. We refer to the basis $\{\omega^i\}$ for V^* as the basis **dual** to the basis $\{e_i\}$ for V. Elements of V^* are usually referred to as 1-**forms**, and are commonly denoted by Greek letters such as β, ϕ, θ and so forth. Similarly, we often refer to the ω^i as **basis** 1-**forms**.

Since applying Theorem 4.4 to the special case of V^* directly may be somewhat confusing, let us briefly go through a slightly different approach to defining a basis for V^*.

Suppose we are given a basis $\{e_1, \dots, e_n\}$ for a finite-dimensional vector space V. Given any set of n scalars ϕ_i, we *define* the linear functionals $\phi \in V^* = L(V, \mathcal{F})$ by $\phi(e_i) = \phi_i$. According to Theorem 4.1, this mapping is unique. In particular, we *define* n linear functionals ω^i by $\omega^i(e_j) = \delta^i_j$. Conversely, given any linear functional $\phi \in V^*$, we *define* the n scalars ϕ_i by $\phi_i = \phi(e_i)$. Then given any $\phi \in V^*$ and any $v = \sum v^i e_i \in V$, we have on the one hand

$$\phi(v) = \phi\Big(\sum v^i e_i\Big) = \sum v^i \phi(e_i) = \sum \phi_i v^i$$

while on the other hand

$$\omega^i(v) = \omega^i\Bigg(\sum_j v^j e_j\Bigg) = \sum_j v^j \omega^i(e_j) = \sum_j v^j \delta^i_j = v^i.$$

Therefore $\phi(v) = \sum_i \phi_i \omega^i(v)$ for any $v \in V$, and we conclude that $\phi = \sum_i \phi_i \omega^i$. This shows that the ω^i span V^*, and we claim that they are in fact a basis for V^*.

To show the ω^i are linearly independent, suppose $\sum_i a_i \omega^i = 0$. We must show that every $a_i = 0$. But for any $j = 1, \dots, n$ we have

$$0 = \sum_i a_i \omega^i(e_j) = \sum_i a_i \delta^i_j = a_j$$

which verifies our claim. This completes the proof that $\{\omega^i\}$ forms a basis for V^*.

There is another common way of denoting the action of V^* on V that is quite similar to the notation used for an inner product. In this approach, the action of the dual basis $\{\omega^i\}$ for V^* on the basis $\{e_i\}$ for V is denoted by writing $\omega^i(e_j)$ as

$$\langle \omega^i, e_j \rangle = \delta^i_j.$$

However, it should be carefully noted that this is *not* an inner product. In particular, the entry on the left inside the bracket is an element of V^*, while the entry on the right is an element of V. Furthermore, from the definition of V^* as a linear vector space, it follows that $\langle \cdot\,, \cdot \rangle$ is linear in both entries. In other words, if $\phi, \theta \in V^*$, and if u, $v \in V$ and a, $b \in \mathcal{F}$, we have

$$\begin{aligned}\langle a\phi + b\theta, u \rangle &= a\langle \phi, u \rangle + b\langle \theta, u \rangle \\ \langle \phi, au + bv \rangle &= a\langle \phi, u \rangle + b\langle \phi, v \rangle.\end{aligned}$$

These relations define what we shall call a **bilinear functional** $\langle \cdot\,, \cdot\rangle : V^* \times V \to \mathcal{F}$ on V^* and V (compare this with definition IP1 of an inner product given in Section 1.5).

We summarize these results as a theorem.

Theorem 8.1. *Let $\{e_1, \dots, e_n\}$ be a basis for V, and let $\{\omega^1, \dots, \omega^n\}$ be the corresponding dual basis for V^* defined by $\omega^i(e_j) = \delta^i_j$. Then any $v \in V$ can be written in the forms*

$$v = \sum_{i=1}^n v^i e_i = \sum_{i=1}^n \omega^i(v) e_i = \sum_{i=1}^n \langle \omega^i, v\rangle e_i$$

and any $\phi \in V^*$ can be written as

$$\phi = \sum_{i=1}^n \phi_i \omega^i = \sum_{i=1}^n \phi(e_i)\omega^i = \sum_{i=1}^n \langle \phi, e_i\rangle \omega^i.$$

This theorem provides us with a simple interpretation of the dual basis. In particular, since we already know that any $v \in V$ has the expansion $v = \sum v^i e_i$ in terms of a basis $\{e_i\}$, we see that $\omega^i(v) = \langle \omega^i, v\rangle = v^i$ is just the ith coordinate of v. In other words, ω^i is just the ith coordinate function on V (relative to the basis $\{e_i\}$).

Let us make another observation. If we write $v = \sum v^i e_i$ and recall that $\phi(e_i) = \phi_i$, then (as we saw above) the linearity of ϕ results in

$$\langle \phi, v\rangle = \phi(v) = \phi\Big(\sum v^i e_i\Big) = \sum v^i \phi(e_i) = \sum \phi_i v^i$$

which looks very much like the standard inner product on $\mathbb{R}^n$. In fact, if V is an inner product space, we shall see that the components of an element $\phi \in V^*$ may be related in a direct way to the components of some vector in V (see Section 8.9).

It is also useful to note that given any nonzero $v \in V$, there exists $\phi \in V^*$ with the property that $\phi(v) \neq 0$. To see this, we use Theorem 1.10 to first extend v to a basis $\{v, v_2, \dots, v_n\}$ for V. Then, according to Theorem 4.1, there exists a unique linear transformation $\phi : V \to \mathcal{F}$ such that $\phi(v) = 1$ and $\phi(v_i) = 0$ for $i = 2, \dots, n$. This ϕ so defined clearly has the desired property. An important consequence of this comes from noting that if v_1, $v_2 \in V$ with $v_1 \neq v_2$, then $v_1 - v_2 \neq 0$, and thus there exists $\phi \in V^*$ such that

$$0 \neq \phi(v_1 - v_2) = \phi(v_1) - \phi(v_2).$$

This proves our next result.

Theorem 8.2. *If V is finite-dimensional and $v_1, v_2 \in V$ with $v_1 \neq v_2$, then there exists $\phi \in V^*$ with the property that $\phi(v_1) \neq \phi(v_2)$.*

Example 8.1. Consider the space $V = \mathbb{R}^2$ consisting of all column vectors of the form

$$v = \begin{bmatrix} v^1 \\ v^2 \end{bmatrix}.$$

Relative to the standard basis we have

$$v = v^1 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + v^2 \begin{bmatrix} 0 \\ 1 \end{bmatrix} = v^1 e_1 + v^2 e_2.$$

If $\phi \in V^*$, then $\phi(v) = \sum \phi_i v^i$, and we may represent ϕ by the row vector $\phi = (\phi_1, \phi_2)$. In particular, if we write the dual basis as $\omega^i = (a_i, b_i)$, then we have

$$1 = \omega^1(e_1) = (a_1, b_1) \begin{bmatrix} 1 \\ 0 \end{bmatrix} = a_1$$

$$0 = \omega^1(e_2) = (a_1, b_1) \begin{bmatrix} 0 \\ 1 \end{bmatrix} = b_1$$

$$0 = \omega^2(e_1) = (a_2, b_2) \begin{bmatrix} 1 \\ 0 \end{bmatrix} = a_2$$

$$1 = \omega^2(e_2) = (a_2, b_2) \begin{bmatrix} 0 \\ 1 \end{bmatrix} = b_2$$

so that $\omega^1 = (1, 0)$ and $\omega^2 = (0, 1)$. Note, for example,

$$\omega^1(v) = (1, 0) \begin{bmatrix} v^1 \\ v^2 \end{bmatrix} = v^1$$

as it should.

Since V^* is a vector space, we can define its dual space V^{**} (the **double dual**) as the space of linear functionals on V^*. We now proceed to show that there is a natural isomorphism between V and V^{**}, and therefore we can generally consider them to be the same space. By a "natural isomorphism" we mean that it is independent of the bases chosen for V and V^{**}.

We first observe that the expression $\langle \cdot\,, u \rangle$ for fixed $u \in V$ defines a linear functional on V^*. (Note that here $\langle \cdot\,, \cdot \rangle$ is a bilinear functional and *not* an inner product.) In other words, we define the functional $f_u : V^* \to \mathcal{F}$ by

$$f_u(\phi) = \langle \phi, u \rangle = \phi(u)$$

for all $\phi \in V^*$. It follows that for all $a, b \in \mathcal{F}$ and $\phi, \omega \in V^*$ we have

$$f_u(a\phi + b\omega) = \langle a\phi + b\omega, u\rangle = a\langle \phi, u\rangle + b\langle \omega, u\rangle = af_u(\phi) + bf_u(\omega)$$

and hence f_u is a linear functional from V^* to $\mathcal{F}$. In other words, $f_u \in L(V^*, \mathcal{F})$ is in the dual space of V^*. As we have already stated, this space is called the **double dual** (or **second dual**) of V, and is denoted by V^{**}.

Recall that $\mathcal{F}$ can be considered to be a 1-dimensional vector space over itself, and from Theorem 4.4 we know that $\dim L(U, V) = (\dim U)(\dim V)$. Hence $\dim V^* = \dim L(V, \mathcal{F}) = \dim V$, and similarly $\dim V^{**} = \dim L(V^*, \mathcal{F}) = \dim V^*$. Therefore we see that $\dim V = \dim V^* = \dim V^{**}$.

Theorem 8.3. *Let V be finite-dimensional over $\mathcal{F}$, and for each $u \in V$ define the function $f_u : V^* \to \mathcal{F}$ by $f_u(\phi) = \phi(u)$ for all $\phi \in V^*$. Then the mapping $f : u \mapsto f_u$ is an isomorphism of V onto V^{**}.*

Proof. We first show that the mapping $f : u \mapsto f_u$ defined above is linear. For any $u, v \in V$ and $a, b \in \mathcal{F}$ we see that

$$\begin{aligned} f_{au+bv}(\phi) &= \langle \phi, au + bv\rangle \\ &= a\langle \phi, u\rangle + b\langle \phi, v\rangle \\ &= af_u(\phi) + bf_v(\phi) \\ &= (af_u + bf_v)(\phi). \end{aligned}$$

Since this holds for all $\phi \in V^*$, it follows that $f_{au+bv} = af_u + bf_v$, and hence the mapping f is indeed linear (so it defines a vector space homomorphism).

Now let $u \in V$ be an arbitrary nonzero vector. By Theorem 8.2 (with $v_1 = u$ and $v_2 = 0$) there exists $\phi \in V^*$ such that $f_u(\phi) = \langle \phi, u\rangle \neq 0$, and hence clearly $f_u \neq 0$. Since it is obviously true that $f_0 = 0$, it follows that $\operatorname{Ker} f = \{0\}$, and thus we have a one-to-one mapping from V into V^{**} (Theorem 4.5).

Finally, since V is finite-dimensional, we know that $\dim V = \dim V^* = \dim V^{**}$, and hence the mapping f must be onto (since it is one-to-one). ∎

The isomorphism $f : u \to f_u$ defined in Theorem 8.3 is called the **natural** (or **evaluation**) **mapping** of V into V^{**}. (We remark without proof that even if V is infinite-dimensional this mapping is linear and injective, but is not surjective.) Because of this isomorphism, we will make the identification $V = V^{**}$ from now on, and hence also view V as the space of linear functionals on V^*. Furthermore, if $\{\omega^i\}$ is a basis for V^*, then the dual basis $\{e_i\}$ for V will be taken to be the basis for V^{**}. In other words, we may write

$$\omega^i(e_j) = e_j(\omega^i) = \delta^i_j$$

so that

$$\phi(v) = v(\phi) = \sum \phi_i v^i.$$

Exercises

1. Find the basis dual to the given basis for each of the following:
   (a) $\mathbb{R}^2$ with basis $e_1 = (2, 1)$, $e_2 = (3, 1)$.
   (b) $\mathbb{R}^3$ with basis $e_1 = (1, -1, 3)$, $e_2 = (0, 1, -1)$, $e_3 = (0, 3, -2)$.

2. Let V be the space of all real polynomials of degree ≤ 1. Define $\omega^1, \omega^2 \in V^*$ by
$$\omega^1(f) = \int_0^1 f(x)\,dx \qquad \text{and} \qquad \omega^2(f) = \int_0^2 f(x)\,dx.$$
   Find a basis $\{e_1, e_2\}$ for V that is dual to $\{\omega^1, \omega^2\}$.

3. Let V be the vector space of all polynomials of degree ≤ 2. Define the linear functionals ω^1, ω^2, $\omega^3 \in V^*$ by
$$\omega^1(f) = \int_0^1 f(x)\,dx, \qquad \omega^2(f) = f'(1), \qquad \omega^3(f) = f(0)$$
   where $f'(x)$ is the usual derivative of $f(x)$. Find the basis $\{e_i\}$ for V which is dual to $\{\omega^i\}$.

4. (a) Let u, $v \in V$ and suppose that $\phi(u) = 0$ implies $\phi(v) = 0$ for all $\phi \in V^*$. Show $v = ku$ for some scalar k.
   (b) Let $\phi, \sigma \in V^*$ and suppose that $\phi(v) = 0$ implies $\sigma(v) = 0$ for all $v \in V$. Show $\sigma = k\phi$ for some scalar k.

5. Let $V = \mathcal{F}[x]$, and for $a \in \mathcal{F}$, define $\phi_a : V \to \mathcal{F}$ by $\phi_a(f) = f(a)$. Show:
   (a) ϕ_a is linear, i.e., $\phi_a \in V^*$.
   (b) If $a \neq b$, then $\phi_a \neq \phi_b$.

6. Let V be finite-dimensional and W a subspace of V. If $\phi \in W^*$, prove ϕ can be extended to a linear functional $\Phi \in V^*$, i.e., $\Phi(w) = \phi(w)$ for all $w \in W$.

8.2 Tensors

Let V be a finite-dimensional vector space over $\mathcal{F}$, and let V^r denote the r-fold Cartesian product $V \times V \times \cdots \times V$. In other words, an element of V^r is an r-tuple $(v_1, \ldots, v_r)$ where each $v_i \in V$. If W is another vector space over $\mathcal{F}$, then a mapping $T : V^r \to W$ is said to be **multilinear** if $T(v_1, \ldots, v_r)$ is linear in each variable. That is, T is multilinear if for each $i = 1, \ldots, r$ we have
$$T(v_1, \ldots, av_i + bv_i', \ldots, v_r) = aT(v_1, \ldots, v_i, \ldots, v_r) + bT(v_1, \ldots, v_i', \ldots, v_r)$$
for all $v_i, v_i' \in V$ and $a, b \in \mathcal{F}$. In the particular case that $W = \mathcal{F}$, the mapping T is variously called an r-**linear form** on V, or a **multilinear form** of **degree**

r on V, or an r-**tensor** on V. The set of all r-tensors on V will be denoted by $\mathcal{T}_r(V)$.

As might be expected, we define addition and scalar multiplication on $\mathcal{T}_r(V)$ by

$$\begin{aligned}(S+T)(v_1,\dots,v_r) &= S(v_1,\dots,v_r)+T(v_1,\dots,v_r)\\ (aT)(v_1,\dots,v_r) &= aT(v_1,\dots,v_r)\end{aligned}$$

for all $S,T\in\mathcal{T}_r(V)$ and $a\in\mathcal{F}$. It should be clear that $S+T$ and aT are both r-tensors. With these operations, $\mathcal{T}_r(V)$ becomes a vector space over $\mathcal{F}$. Note that the particular case of $r=1$ yields $\mathcal{T}_1(V)=V^*$, i.e., the dual space of V, and if $r=2$, then we obtain a bilinear form on V.

Although this definition takes care of most of what we will need in this chapter, it is worth going through a more general (but not really more difficult) definition as follows. The basic idea is that a tensor is a scalar-valued multilinear function with variables in both V and V^*. For example, a tensor could be a function on the space $V^*\times V\times V$. By convention, we will always write all V^* variables before all V variables so that, for example, a tensor on $V\times V^*\times V$ will be replaced by a tensor on $V^*\times V\times V$. (However, not all authors adhere to this convention, so the reader should be very careful when reading the literature.) Note also that by Theorem 8.3, the space of linear functions on V^* is V^{**} which we view as simply V.

Without further ado, we define a **tensor** T on V to be a multilinear map on $V^{*s}\times V^r$:

$$T: V^{*s}\times V^r=\underbrace{V^*\times\cdots\times V^*}_{s\text{ copies}}\times\underbrace{V\times\cdots\times V}_{r\text{ copies}}\to\mathcal{F}$$

where r is called the **covariant order** and s is called the **contravariant order** of T. We shall say that a tensor of covariant order r and contravariant order s is of **type** (or **rank**) $\binom{s}{r}$. Furthermore, as stated above, we will refer to a covariant tensor of rank $\binom{0}{r}$ as simply an r-tensor. (Don't confuse this symbol with the binomial coefficients.)

If we denote the set of all tensors of type $\binom{s}{r}$ by $\mathcal{T}_r^{\,s}(V)$, then defining addition and scalar multiplication exactly as above, we see that $\mathcal{T}_r^{\,s}(V)$ forms a vector space over $\mathcal{F}$. A tensor of type $\binom{0}{0}$ is defined to be a scalar, and hence $\mathcal{T}_0^{\,0}(V)=\mathcal{F}$. A tensor of type $\binom{1}{0}$ is called a **contravariant vector**, and a tensor of type $\binom{0}{1}$ is called a **covariant vector** (or simply a **covector**). In order to distinguish between these types of vectors, we denote the basis vectors for V by a subscript (e.g., e_i), and the basis vectors for V^* by a superscript (e.g., ω^j). Furthermore, we will generally leave off the V and simply write $\mathcal{T}_r$ or $\mathcal{T}_r^{\,s}$.

At this point we are virtually forced to adhere to the Einstein summation convention and sum over repeated indices in any vector or tensor expression where one index is a superscript and one is a subscript. Because of this, we write the vector components with indices in the opposite position from that of the basis vectors. This is why we have been writing $x=\sum_i x^ie_i\in V$ and

$\phi = \sum_j \phi_j \omega^j \in V^*$. Thus we now simply write $x = x^i e_i$ and $\phi = \phi_j \omega^j$ where the summation is to be understood. Generally the limits of the sum will be clear. However, we will revert to the more complete notation if there is any possibility of ambiguity, or those rare circumstances when a repeated index is not to be summed over.

It is also worth emphasizing the trivial fact that the indices summed over are just "dummy indices." In other words, we have $x^i e_i = x^j e_j$ and so on. Throughout this chapter we will be relabelling indices in this manner without further notice, and we will assume that the reader understands what we are doing.

Suppose $T \in \mathcal{T}_r$, and let $\{e_1, \dots, e_n\}$ be a basis for V. Pick any r vectors $v_i \in V$ and expand them in terms of this basis as $v_i = e_j a^j{}_i$ where, as usual, $a^j{}_i \in \mathcal{F}$ is just the jth component of the vector v_i. (Note that here the subscript i on v_i is not a tensor index; the context should make it clear whether we are dealing with a vector or it's components. We also emphasize that the $a^j{}_i$ are not the components of any tensor, and hence the position of indices here is just a matter of notational convenience and consistency with the summation convention.) Using the multilinearity of T we see that

$$T(v_1, \dots, v_r) = T(e_{j_1} a^{j_1}{}_1, \dots, e_{j_r} a^{j_r}{}_r) = a^{j_1}{}_1 \cdots a^{j_r}{}_r T(e_{j_1}, \dots, e_{j_r}).$$

The n^r scalars $T(e_{j_1}, \dots, e_{j_r})$ are called the **components** of T relative to the basis $\{e_i\}$, and are denoted by $T_{j_1 \cdots j_r}$. As we will prove in detail below, this multilinearity essentially shows that a tensor is completely specified once we know its values on a basis (i.e., its components).

This terminology implies that there exists a basis for $\mathcal{T}_r$ such that the $T_{j_1 \cdots j_r}$ are just the components of $T \in \mathcal{T}_r$ with respect to this basis. We now construct this basis, which will prove that $\mathcal{T}_r$ is of dimension n^r. (We will show formally in Section 8.9 that the Kronecker symbols $\delta^i{}_j$ are in fact the components of a tensor, and that these components are the same in any coordinate system. However, for all practical purposes we continue to use the $\delta^i{}_j$ simply as a notational device, and hence we place no importance on the position of the indices, i.e., $\delta^i{}_j = \delta_j{}^i$ etc.)

For each collection $\{i_1, \dots, i_r\}$ (where $1 \le i_k \le n$), we define the tensor $\Omega^{i_1 \cdots i_r}$ (not simply the components of a tensor Ω) to be that element of $\mathcal{T}_r$ whose values on the basis $\{e_i\}$ for V are given by

$$\Omega^{i_1 \cdots i_r}(e_{j_1}, \dots, e_{j_r}) = \delta^{i_1}{}_{j_1} \cdots \delta^{i_r}{}_{j_r}$$

and whose values on an arbitrary collection $\{v_1, \dots, v_r\}$ of vectors are given by multilinearity as

$$\begin{aligned} \Omega^{i_1 \cdots i_r}(v_1, \dots, v_r) &= \Omega^{i_1 \cdots i_r}(e_{j_1} a^{j_1}{}_1, \dots, e_{j_r} a^{j_r}{}_r) \\ &= a^{j_1}{}_1 \cdots a^{j_r}{}_r \Omega^{i_1 \cdots i_r}(e_{j_1}, \dots, e_{j_r}) \\ &= a^{j_1}{}_1 \cdots a^{j_r}{}_r \delta^{i_1}{}_{j_1} \cdots \delta^{i_r}{}_{j_r} \\ &= a^{i_1}{}_1 \cdots a^{i_r}{}_r \end{aligned}$$

That this does indeed define a tensor is guaranteed by this last equation which shows that each $\Omega^{i_1\cdots i_r}$ is in fact linear in each variable (since $v_1 + v_1' = (a^{j_1}{}_1 + a'^{j_1}{}_1)e_{j_1}$ etc.). To prove the n^r tensors $\Omega^{i_1\cdots i_r}$ form a basis for $\mathcal{T}_r$, we must show they are linearly independent and span $\mathcal{T}_r$.

Suppose $\alpha_{i_1\cdots i_r}\Omega^{i_1\cdots i_r} = 0$ where each $\alpha_{i_1\cdots i_r} \in \mathcal{F}$. From the definition of $\Omega^{i_1\cdots i_r}$ we see that applying this to any r-tuple $(e_{j_1}, \ldots, e_{j_r})$ of basis vectors yields $\alpha_{i_1\cdots i_r} = 0$. Since this is true for every such r-tuple, it follows that $\alpha_{i_1\cdots i_r} = 0$ for every r-tuple of indices $(i_1, \ldots, i_r)$, and hence the $\Omega^{i_1\cdots i_r}$ are linearly independent.

Now let $T_{i_1\cdots i_r} = T(e_{i_1}, \ldots, e_{i_r})$ and consider the tensor $T_{i_1\cdots i_r}\Omega^{i_1\cdots i_r} \in \mathcal{T}_r$. Using the definition of $\Omega^{i_1\cdots i_r}$, we see that both $T_{i_1\cdots i_r}\Omega^{i_1\cdots i_r}$ and T yield the same result when applied to any r-tuple $(e_{j_1}, \ldots, e_{j_r})$ of basis vectors, and hence they must be equal as multilinear functions on V^r. This shows that $\{\Omega^{i_1\cdots i_r}\}$ spans $\mathcal{T}_r$.

While we have treated only the space $\mathcal{T}_r$, it is not any more difficult to treat the general space $\mathcal{T}_r^s$. Thus, if $\{e_i\}$ is a basis for V, $\{\omega^j\}$ is a basis for V^* and $T \in \mathcal{T}_r^s$, we define the **components** of T (relative to the given bases) by

$$T^{i_1\cdots i_s}{}_{j_1\cdots j_r} = T(\omega^{i_1}, \ldots, \omega^{i_s}, e_{j_1}, \ldots, e_{j_r}).$$

Defining the n^{r+s} analogous tensors $\Omega^{j_1\cdots j_r}_{i_1\cdots i_s}$, it is easy to mimic the above procedure and hence prove the following result.

Theorem 8.4. *The set $\mathcal{T}_r^s$ of all tensors of type $\binom{s}{r}$ on V forms a vector space of dimension n^{r+s}.*

Proof. This is Exercise 8.2.1. ∎

Since a tensor $T \in \mathcal{T}_r^s$ is a function on $V^{*s} \times V^r$, it would be nice if we could write a basis (e.g., $\Omega^{j_1\cdots j_r}_{i_1\cdots i_s}$) for $\mathcal{T}_r^s$ in terms of the bases $\{e_i\}$ for V and $\{\omega^j\}$ for V^*. We now show this is easy to accomplish by defining a product on $\mathcal{T}_r^s$ called the tensor product. The reader is cautioned not to be intimidated by the notational complexities, since the concepts involved are really quite simple.

Suppose $S \in \mathcal{T}_{r_1}^{s_1}$ and $T \in \mathcal{T}_{r_2}^{s_2}$. Let $u_1, \ldots, u_{r_1}, v_1, \ldots, v_{r_2}$ be vectors in V, and $\alpha^1, \ldots, \alpha^{s_1}, \beta^1, \ldots, \beta^{s_2}$ be covectors in V^*. Note that the product

$$S(\alpha^1, \ldots, \alpha^{s_1}, u_1, \ldots, u_{r_1})T(\beta^1, \ldots, \beta^{s_2}, v_1, \ldots, v_{r_2})$$

is linear in each of its $r_1 + s_1 + r_2 + s_2$ variables. Hence we define the **tensor product** $S \otimes T \in \mathcal{T}_{r_1+r_2}^{s_1+s_2}$ (read "S tensor T") by

$$\begin{aligned}(S \otimes T)(\alpha^1, \ldots, \alpha^{s_1}, \beta^1, \ldots, \beta^{s_2}, u_1, \ldots, u_{r_1}, v_1, \ldots, v_{r_2})\\ = S(\alpha^1, \ldots, \alpha^{s_1}, u_1, \ldots, u_{r_1})T(\beta^1, \ldots, \beta^{s_2}, v_1, \ldots, v_{r_2}).\end{aligned}$$

It is easily shown that the tensor product is both associative and distributive (i.e., bilinear in both factors). In other words, for any scalar $a \in \mathcal{F}$ and tensors R, S and T such that the following formulas make sense, we have

$$\begin{aligned}(R \otimes S) \otimes T &= R \otimes (S \otimes T) \\ R \otimes (S + T) &= R \otimes S + R \otimes T \\ (R + S) \otimes T &= R \otimes T + S \otimes T \\ (aS) \otimes T &= S \otimes (aT) = a(S \otimes T)\end{aligned}$$

(see Exercise 8.2.2). Because of the associativity property (which is a consequence of associativity in $\mathcal{F}$), we will drop the parentheses in expressions such as the top equation and simply write $R \otimes S \otimes T$. This clearly extends to any finite product of tensors. It is important to note, however, that the tensor product is most certainly *not* commutative, i.e., $S \otimes T \neq T \otimes S$.

Now let $\{e_1, \dots, e_n\}$ be a basis for V, and let $\{\omega^j\}$ be its dual basis. We claim that the set $\{\omega^{j_1} \otimes \cdots \otimes \omega^{j_r}\}$ of tensor products, where $1 \leq j_k \leq n$, forms a basis for the space $\mathcal{T}_r$ of covariant tensors. To see this, we note that from the definitions of tensor product and dual space we have

$$\omega^{j_1} \otimes \cdots \otimes \omega^{j_r}(e_{i_1}, \dots, e_{i_r}) = \omega^{j_1}(e_{i_1}) \cdots \omega^{j_r}(e_{i_r}) = \delta^{j_1}{}_{i_1} \cdots \delta^{j_r}{}_{i_r} \tag{8.1}$$

so that $\omega^{j_1} \otimes \cdots \otimes \omega^{j_r}$ and $\Omega^{j_1 \cdots j_r}$ both take the same values on the r-tuples $(e_{i_1}, \dots, e_{i_r})$, and hence they must be equal as multilinear functions on V^r. Since we showed above that $\{\Omega^{j_1 \cdots j_r}\}$ forms a basis for $\mathcal{T}_r$, we have proved that $\{\omega^{j_1} \otimes \cdots \otimes \omega^{j_r}\}$ also forms a basis for $\mathcal{T}_r$.

The method of the previous paragraph is readily extended to the space $\mathcal{T}_r^s$. We must recall however, that we are treating V^{**} and V as the same space. If $\{e_i\}$ is a basis for V, then the dual basis $\{\omega^j\}$ for V^* was defined by $\omega^j(e_i) = \langle \omega^j, e_i \rangle = \delta^j{}_i$. Similarly, given a basis $\{\omega^j\}$ for V^*, we define the basis $\{e_i\}$ for $V^{**} = V$ by $e_i(\omega^j) = \omega^j(e_i) = \delta^j{}_i$. In fact, using tensor products, it is now easy to repeat Theorem 8.4 in its most useful form. Note also that the next theorem shows that a tensor is determined by its values on the bases $\{e_i\}$ and $\{\omega^j\}$.

Theorem 8.5. *Let V have basis $\{e_1, \dots, e_n\}$, and let V^* have the corresponding dual basis $\{\omega^1, \dots, \omega^n\}$. Then a basis for $\mathcal{T}_r^s$ is given by the collection*

$$\{e_{i_1} \otimes \cdots \otimes e_{i_s} \otimes \omega^{j_1} \otimes \cdots \otimes \omega^{j_r}\}$$

where $1 \leq j_1, \dots, j_r, i_1, \dots, i_s \leq n$, and hence $\dim \mathcal{T}_r^s = n^{r+s}$.

Proof. In view of Theorem 8.4, all that is needed is to show that

$$e_{i_1} \otimes \cdots \otimes e_{i_s} \otimes \omega^{j_1} \otimes \cdots \otimes \omega^{j_r} = \Omega_{i_1 \cdots i_s}^{j_1 \cdots j_r}.$$

The details are left to the reader (see Exercise 8.2.1). ∎

Note that what this theorem tells us is that any tensor $T \in \mathcal{T}_r^s(V)$ can be written as

$$T = T^{i_1 \cdots i_s}{}_{j_1 \cdots j_r} e_{i_1} \otimes \cdots \otimes e_{i_s} \otimes \omega^{j_1} \otimes \cdots \otimes \omega^{j_r} .$$

Since the components of a tensor T are defined with respect to a particular basis (and dual basis), we might ask about the relationship between the components of T relative to two different bases. Using the multilinearity of tensors, this is a simple problem to solve.

First, let $\{e_i\}$ be a basis for V and let $\{\omega^j\}$ be its dual basis. If $\{\bar{e}_i\}$ is another basis for V, then there exists a nonsingular transition matrix $A = (a^j{}_i)$ such that

$$\bar{e}_i = e_j a^j{}_i. \tag{8.2}$$

(We emphasize that $a^j{}_i$ is only a matrix, not a tensor. Note also that our definition of the matrix of a linear transformation given in Section 4.3 shows that $a^j{}_i$ is the element of A in the jth row and ith column.) Using $\langle \omega^i, e_j \rangle = \delta^i{}_j$, we have

$$\langle \omega^i, \bar{e}_k \rangle = \langle \omega^i, e_j a^j{}_k \rangle = a^j{}_k \langle \omega^i, e_j \rangle = a^j{}_k \delta^i{}_j = a^i{}_k.$$

Let us denote the inverse of the matrix $A = (a^i{}_j)$ by $A^{-1} = B = (b^i{}_j)$. In other words, $a^i{}_j b^j{}_k = \delta^i{}_k$ and $b^i{}_j a^j{}_k = \delta^i{}_k$. Multiplying both sides of $\langle \omega^i, \bar{e}_k \rangle = a^i{}_k$ by $b^j{}_i$ and summing on i yields

$$\langle b^j{}_i \omega^i, \bar{e}_k \rangle = b^j{}_i a^i{}_k = \delta^j{}_k.$$

But the basis $\{\bar{\omega}^i\}$ dual to $\{\bar{e}_i\}$ also must satisfy $\langle \bar{\omega}^j, \bar{e}_k \rangle = \delta^j{}_k$, and hence comparing this with the previous equation shows that the dual basis vectors transform as

$$\bar{\omega}^j = b^j{}_i \omega^i. \tag{8.3}$$

The reader should compare this carefully with equation (8.2). We say that the dual basis vectors transform **oppositely** (i.e., use the inverse transformation matrix) to the basis vectors.

We now return to the question of the relationship between the components of a tensor in two different bases. For definiteness, we will consider a tensor $T \in \mathcal{T}_1^2$. The analogous result for an arbitrary tensor in $\mathcal{T}_r^s$ will be quite obvious. Let $\{e_i\}$ and $\{\omega^j\}$ be a basis and dual basis for V and V^* respectively. Now consider another pair of bases $\{\bar{e}_i\}$ and $\{\bar{\omega}^j\}$ where $\bar{e}_i = e_j a^j{}_i$ and $\bar{\omega}^i = b^i{}_j \omega^j$. Then we have $T^{ij}{}_k = T(\omega^i, \omega^j, e_k)$ as well as $\overline{T}^{pq}{}_r = T(\bar{\omega}^p, \bar{\omega}^q, \bar{e}_r)$, and therefore

$$\overline{T}^{pq}{}_r = T(\bar{\omega}^p, \bar{\omega}^q, \bar{e}_r) = b^p{}_i b^q{}_j a^k{}_r T(\omega^i, \omega^j, e_k) = b^p{}_i b^q{}_j a^k{}_r T^{ij}{}_k.$$

This is the **classical transformation law** for the components of a tensor of type $\binom{2}{1}$. It should be kept in mind that $(a^i{}_j)$ and $(b^i{}_j)$ are inverse matrices to each other. (In fact, this equation is frequently taken as the *definition* of a tensor (at least in older texts). In other words, according to this approach, any quantity with this transformation property is *defined* to be a tensor.)

In particular, the components x^i of a vector $x = x^i e_i$ transform as

$$\bar{x}^i = b^i{}_j x^j \tag{8.4}$$

while the components α_i of a covector $\alpha = \alpha_i \omega^i$ transform as

$$\bar{\alpha}_i = \alpha_j a^j{}_i. \tag{8.5}$$

We leave it to the reader to verify that these transformation laws lead to the self-consistent formulas $x = x^i e_i = \bar{x}^j \bar{e}_j$ and $\alpha = \alpha_i \omega^i = \bar{\alpha}_j \bar{\omega}^j$ as we should expect (see Exercise 8.2.3).

We point out that these transformation laws are the origin of the terms "contravariant" and "covariant." This is because the components of a vector transform opposite ("contravariant") to the basis vectors e_i, while the components of dual vectors transform the same ("covariant") way as these basis vectors.

It is also worth mentioning that many authors use a prime (or some other method such as a different type of letter) for distinguishing different bases. In other words, if we have a basis $\{e_i\}$ and we wish to transform to another basis which we denote by $\{e_{i'}\}$, then this is accomplished by a transformation matrix $(a^i{}_{j'})$ so that $e_{i'} = e_j a^j{}_{i'}$. In this case, we would write $\omega^{i'} = a^{i'}{}_j \omega^j$ where $(a^{i'}{}_j)$ is the inverse of $(a^i{}_{j'})$. In this notation, the transformation law for the tensor T used above would be written as

$$T^{p'q'}{}_{r'} = b^{p'}{}_i b^{q'}{}_j a^k{}_{r'} T^{ij}{}_k.$$

Note that specifying the components of a tensor with respect to one coordinate system allows the determination of its components with respect to any other coordinate system. Because of this, we shall frequently refer to a tensor by its "generic" components. In other words, we will refer to e.g., $T^{ij}{}_k$, as a "tensor" and not the more accurate description as the "components of the tensor T."

Example 8.2. We know that under a change of basis $e_i \to \bar{e}_i = e_j p^j{}_i$ we also have $x^i \to \bar{x}^i = (p^{-1})^i{}_j x^j$ and $x^i = p^i{}_j \bar{x}^j$. (This is just the requirement that $x = x^i e_i = \bar{x}^i \bar{e}_i$. See Section 4.4.) If x is just the position vector, then the components x^i and $\bar{x}^i$ are just the coordinate functions, and hence we may take their derivatives to write

$$a^i{}_j = p^i{}_j = \frac{\partial x^i}{\partial \bar{x}^j} \qquad \text{and} \qquad b^i{}_j = (p^{-1})^i{}_j = \frac{\partial \bar{x}^i}{\partial x^j}$$

where we have also used the notation of equations (8.2) and (8.3).

Now suppose we have a tensor $T \in \mathcal{T}_1^2(V)$, where V has basis $\{e_i\}$ and corresponding dual basis $\{\omega^i\}$. Then we can write the components with respect to the basis $\{\bar{e}_i\}$ in terms of the components with respect to the basis $\{e_i\}$ by directly applying the definitions:

$$\begin{aligned}\overline{T}^{ij}{}_k &= T(\bar{\omega}^i, \bar{\omega}^j, \bar{e}_k) = T(b^i{}_l \omega^l, b^j{}_m \omega^m, e_n a^n{}_k) \\ &= b^i{}_l b^j{}_m a^n{}_k T(\omega^l, \omega^m, e_n) = b^i{}_l b^j{}_m a^n{}_k T^{lm}{}_n\end{aligned}$$

$$= \frac{\partial \bar{x}^i}{\partial x^l}\frac{\partial \bar{x}^j}{\partial x^m}\frac{\partial x^n}{\partial \bar{x}^k}T^{lm}{}_n\,.$$

This is the classical transformation law for a tensor of type $\binom{2}{1}$.

Example 8.3. For those readers who may have seen a classical treatment of tensors and have had a course in advanced calculus, we will now show how our more modern approach agrees with the classical.

If $\{x^i\}$ is a local coordinate system on a differentiable manifold M, then a (tangent) **vector field** $v(x)$ on M is defined as the derivative operator

$$v = v^i \frac{\partial}{\partial x^i}$$

(where $v^i = v^i(x)$) so that $v(f) = v^i(\partial f/\partial x^i)$ for every smooth function $f : M \to \mathbb{R}$. Since every vector at $x \in M$ can in this manner be written as a linear combination of the $\partial/\partial x^i$, we see that $\{\partial/\partial x^i\}$ forms a basis for the tangent space at x.

We now define the **differential** df of a function by $df(v) = v(f)$ and thus $df(v)$ is just the directional derivative of f in the direction of v. Note that

$$dx^i(v) = v(x^i) = v^j \frac{\partial x^i}{\partial x^j} = v^j \delta^i_j = v^i$$

and hence $df(v) = v^i(\partial f/\partial x^i) = (\partial f/\partial x^i)dx^i(v)$. Since v was arbitrary, we obtain the familiar elementary formula $df = (\partial f/\partial x^i)dx^i$. Furthermore, we see that

$$dx^i\left(\frac{\partial}{\partial x^j}\right) = \frac{\partial x^i}{\partial x^j} = \delta^i_j$$

so that $\{dx^i\}$ forms the basis dual to $\{\partial/\partial x^i\}$.

In summary then, relative to the local coordinate system $\{x^i\}$, we define a basis $\{e_i = \partial/\partial x^i\}$ for a (tangent) space V along with the dual basis $\{\omega^j = dx^j\}$ for the (cotangent) space V^*.

If we now go to a new coordinate system $\{\bar{x}^i\}$ in the same coordinate patch, then from calculus we obtain

$$\frac{\partial}{\partial \bar{x}^i} = \frac{\partial x^j}{\partial \bar{x}^i}\frac{\partial}{\partial x^j}$$

so the expression $\bar{e}_i = e_j a^j{}_i$ implies $a^j{}_i = \partial x^j/\partial \bar{x}^i$. Similarly, we also have

$$d\bar{x}^i = \frac{\partial \bar{x}^i}{\partial x^j}dx^j$$

so that $\bar{\omega}^i = b^i{}_j\omega^j$ implies $b^i{}_j = \partial \bar{x}^i/\partial x^j$. Note that the chain rule from calculus shows us that

$$a^i{}_k b^k{}_j = \frac{\partial x^i}{\partial \bar{x}^k}\frac{\partial \bar{x}^k}{\partial x^j} = \frac{\partial x^i}{\partial x^j} = \delta^i_j$$

and thus $(b^i{}_j)$ is indeed the inverse matrix to $(a^i{}_j)$.

Using these results in the above expression for $\overline{T}^{pq}{}_r$, we see that

$$\overline{T}^{pq}{}_r = \frac{\partial \bar{x}^p}{\partial x^i}\frac{\partial \bar{x}^q}{\partial x^j}\frac{\partial x^k}{\partial \bar{x}^r} T^{ij}{}_k$$

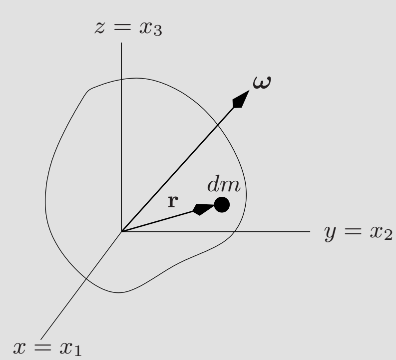
which is just the classical definition of the transformation law for a tensor of type $\binom{2}{1}$.

We also remark that in older texts, a contravariant vector is *defined* to have the same transformation properties as the expression $d\bar{x}^i = (\partial \bar{x}^i/\partial x^j)dx^j$, while a covariant vector is *defined* to have the same transformation properties as the expression $\partial/\partial \bar{x}^i = (\partial x^j/\partial \bar{x}^i)\partial/\partial x^j$.

Finally, let us define a simple classical tensor operation that is frequently quite useful. To begin with, we have seen that the result of operating on a vector $v = v^i e_i \in V$ with a dual vector $\alpha = \alpha_j \omega^j \in V^*$ is just $\langle \alpha, v\rangle = \alpha_j v^i \langle \omega^j, e_i\rangle = \alpha_j v^i \delta^j_i = \alpha_i v^i$. This is sometimes called the **contraction** of α with v. We leave it to the reader to show that the contraction is independent of the particular coordinate system used (see Exercise 8.2.4). (This is just a generalization of the fact that the elementary dot product gives a number (scalar) that is independent of the coordinate system. Geometrically, the length of a vector doesn't depend on the coordinate system, nor does the angle between two vectors.)

If we start with tensors of higher order, then we can perform the same sort of operation. For example, if we have $S \in \mathcal{T}_2^1$ with components $S^i{}_{jk}$ and $T \in \mathcal{T}^2$ with components T^{pq}, then we can form the $\binom{2}{1}$ tensor with components $S^i{}_{jk}T^{jq}$, or a different $\binom{2}{1}$ tensor with components $S^i{}_{jk}T^{pj}$ and so forth. This operation is also called **contraction**. Note that if we start with a $\binom{1}{1}$ tensor T, then we can contract the components of T to obtain the scalar $T^i{}_i$. This is called the **trace** of T.

Example 8.4. Let us give an example of a tensor that occurs in classical mechanics. Consider the rotation of a rigid body about an arbitrary axis of rotation defined by an angular velocity vector $\boldsymbol{\omega}$ as shown in the figure below.



(Don't confuse this $\boldsymbol{\omega}$ with a dual basis vector.) The vector $\mathbf{r}$ points to an arbitary element of mass dm whose velocity is given by $\mathbf{v} = \boldsymbol{\omega} \times \mathbf{r}$. (A particle moving in a circle of radius a moves a distance $ds = a\,d\theta$ where θ is the angle of rotation. Then its speed is $v = ds/dt = a\,\dot{\theta} = a\,\omega$ with a direction tangent to the circle. Defining the angular velocity vector $\boldsymbol{\omega}$ to have magnitude ω and direction given by the right hand rule along the axis of rotation, we see that we may write in general $\mathbf{v} = \boldsymbol{\omega} \times \mathbf{r}$ where $\|\boldsymbol{\omega} \times \mathbf{r}\| = a\,\omega$.)

The kinetic energy of the object is now given by

$$T = \frac{1}{2}\int v^2\,dm = \frac{1}{2}\int \rho(\boldsymbol{\omega} \times \mathbf{r}) \cdot (\boldsymbol{\omega} \times \mathbf{r})\,d^3r$$

where ρ is the density of the object. We rewrite the dot product as follows:

$$\begin{aligned}
(\boldsymbol{\omega} \times \mathbf{r}) \cdot (\boldsymbol{\omega} \times \mathbf{r}) &= (\boldsymbol{\omega} \times \mathbf{r})^i (\boldsymbol{\omega} \times \mathbf{r})_i = \varepsilon^{ijk}\varepsilon_{ilm}\omega_j x_k \omega^l x^m \\
&= (\delta^j_l \delta^k_m - \delta^k_l \delta^j_m)\omega_j x_k \omega^l x^m \\
&= \omega_j x_k \omega^j x^k - \omega_j x_k \omega^k x^j \\
&= \omega^j \omega^k \delta_{jk} r^2 - \omega^j \omega^k x_j x_k \\
&= \omega^j (\delta_{jk} r^2 - x_j x_k)\omega^k .
\end{aligned}$$

If we define the 3×3 symmetric matrix $\mathbf{I} = (I_{jk})$ by

$$I_{jk} = \int \rho(\delta_{jk} r^2 - x_j x_k)\,d^3r$$

then the kinetic energy becomes

$$T = \frac{1}{2}\,\omega^j I_{jk} \omega^k$$

which is often written in the form $T = (1/2)\boldsymbol{\omega} \cdot \mathbf{I} \cdot \boldsymbol{\omega}$. The matrix $\mathbf{I}$ is called the **inertia tensor**. It is a geometric property of the object that depends only on the coordinate system chosen and not on the angular velocity.

To see why $\mathbf{I}$ is called a tensor, first note that the kinetic energy T is just a scalar. (The kinetic energy of an object clearly doesn't depend on the coordinate system chosen.) This means that under rotations of the coordinate system it remains unchanged, i.e., $T(\bar{x}) = T(x)$ where $\mathbf{x} = x^i e_i = \bar{x}^i \bar{e}_i$ is the same point described with respect to two different coordinate systems.

Now $\boldsymbol{\omega}$ is just a normal vector like $\mathbf{x}$, and as we saw in Section 4.5, under rotations $\bar{e}_i = e_j p^j{}_i$ the components of a vector transform as $\bar{\omega}^i = (p^{-1})^i{}_j\,\omega^j$. Since T doesn't depend on the coordinate system we may write

$$\begin{aligned}
\bar{\omega}^i \bar{I}_{ij} \bar{\omega}^j &= (p^{-1})^i{}_k \omega^k \bar{I}_{ij} (p^{-1})^j{}_l \omega^l \\
&= \omega^k \left[(p^{-1})^i{}_k (p^{-1})^j{}_l \bar{I}_{ij}\right] \omega^l \\
&:= \omega^k I_{kl} \omega^l
\end{aligned}$$

and therefore
$$I_{kl} = (p^{-1})^i{}_k (p^{-1})^j{}_l \bar{I}_{ij}.$$
Multiplying both sides of this equation by $p^k{}_r p^l{}_s$ and using $(p^{-1})^i{}_k p^k{}_r = \delta^i_r$ and $(p^{-1})^j{}_l p^l{}_s = \delta^j_s$ we have
$$\bar{I}_{rs} = p^k{}_r p^l{}_s I_{kl}$$
which is the transformation law for a second rank tensor. In other words, in order that the kinetic energy be a scalar, it is necessary that the "matrix" **I** be a tensor.

Finally, taking the derivative of $x^i = p^i{}_j \bar{x}^j$ shows that $p^i{}_j = \partial x^i / \partial \bar{x}^j$ and we may write this in the classical form
$$\bar{I}_{rs} = \frac{\partial x^k}{\partial \bar{x}^r} \frac{\partial x^l}{\partial \bar{x}^s} I_{kl}.$$

Exercises

1. (a) Prove Theorem 8.4.
   (b) Prove Theorem 8.5.

2. Prove the four associative and distributive properties of the tensor product given in the text following Theorem 8.4.

3. Use $v = v^i e_i = \bar{v}^j \bar{e}_j$ and $\alpha = \alpha_i \omega^i = \bar{\alpha}_j \bar{\omega}^j$ to show that equations (8.4) and (8.5) follow from equations (8.2) and (8.3).

4. If $v \in V$ and $\alpha \in V^*$, show that $\langle \alpha, v \rangle$ is independent of the particular basis chosen for V. Generalize this to arbitrary tensors.

5. Let A_i be a covariant vector field (i.e., $A_i = A_i(x)$) with the transformation rule
$$\bar{A}_i = \frac{\partial x^j}{\partial \bar{x}^i} A_j.$$
Show the quantity $\partial_j A_i = \partial A_i / \partial x^j$ does not define a tensor, but that $F_{ij} = \partial_i A_j - \partial_j A_i$ is in fact a second-rank tensor.

8.3 Special Types of Tensors

In order to obtain some of the most useful results concerning tensors, we turn our attention to the space $\mathcal{T}_r$ of covariant tensors on V. We say that a tensor $S \in \mathcal{T}_r$ is **symmetric** if for each pair (i, j) with $1 \le i, j \le r$ and all $v_i \in V$ we have
$$S(v_1, \dots, v_i, \dots, v_j, \dots, v_r) = S(v_1, \dots, v_j, \dots, v_i, \dots, v_r).$$

Similarly, $A \in \mathcal{T}_r$ is said to be **antisymmetric** (or **skew-symmetric** or **alternating**) if

$$A(v_1, \ldots, v_i, \ldots, v_j, \ldots, v_r) = -A(v_1, \ldots, v_j, \ldots, v_i, \ldots, v_r).$$

Note this definition implies that $A(v_1, \ldots, v_r) = 0$ if any two of the v_i are identical. In fact, this was our original definition of an alternating bilinear form. Furthermore, we also see that $A(v_1, \ldots, v_r) = 0$ if any v_i is a linear combination of the rest of the v_j. In particular, this means we must always have $r \leq \dim V$ if we are to have a nonzero antisymmetric tensor of type $\binom{0}{r}$ on V.

It is easy to see that if S_1, $S_2 \in \mathcal{T}_r$ are symmetric, then so is $aS_1 + bS_2$ where $a, b \in \mathcal{F}$. Similarly, $aA_1 + bA_2$ is antisymmetric. Therefore the symmetric tensors form a subspace of $\mathcal{T}_r$ which we denote by $\sum^r(V)$, and the antisymmetric tensors form another subspace of $\mathcal{T}_r$ which is denoted by $\bigwedge^r(V)$ (some authors denote this space by $\bigwedge^r(V^*)$). Elements of $\bigwedge^r(V)$ are generally called **exterior *r*-forms**, or simply ***r*-forms**. These are the r-tensors that are also antisymmetric. According to this terminology, the basis vectors $\{\omega^i\}$ for V^* are referred to as **basis 1-forms**. Note that the only element common to both of these subspaces is the zero tensor. Also note that a 0-form is just an ordinary function.

A particularly important example of an antisymmetric tensor is the determinant function $\det \in \mathcal{T}_n(\mathbb{R}^n)$. Note also that the definition of a symmetric tensor translates into the obvious requirement that (e.g., in the particular case of $\mathcal{T}_2$) $S_{ij} = S_{ji}$, while an antisymmetric tensor obeys $A_{ij} = -A_{ji}$. These definitions can also be extended to include contravariant tensors although we shall have little need to do so.

We point out that even if $S, T \in \sum^r(V)$ are both symmetric, it need not be true that $S \otimes T$ be symmetric (i.e., $S \otimes T \notin \sum^{r+r}(V)$). For example, if $S_{ij} = S_{ji}$ and $T_{pq} = T_{qp}$, it does not necessarily follow that $S_{ij}T_{pq} = S_{ip}T_{jq}$. It is also clear that if A, $B \in \bigwedge^r(V)$, then we do not necessarily have $A \otimes B \in \bigwedge^{r+r}(V)$.

Example 8.5. Suppose $\alpha \in \bigwedge^n(V)$, let $\{e_1, \ldots, e_n\}$ be a basis for V, and for each $i = 1, \ldots, n$ let $v_i = e_j a^j{}_i$ where $a^j{}_i \in \mathcal{F}$. Then, using the multilinearity of α, we may write

$$\alpha(v_1, \ldots, v_n) = a^{j_1}{}_1 \cdots a^{j_n}{}_n \, \alpha(e_{j_1}, \ldots, e_{j_n})$$

where the sums are over all $1 \leq j_k \leq n$. But $\alpha \in \bigwedge^n(V)$ is antisymmetric, and hence $(e_{j_1}, \ldots, e_{j_n})$ must be a permutation of $(e_1, \ldots, e_n)$ in order that the e_{j_k} all be distinct (or else $\alpha(e_{j_1}, \ldots, e_{j_n}) = 0$). In other words, $\alpha(e_{j_1}, \ldots, e_{j_n}) = \varepsilon_{j_1 \cdots j_n} \alpha(e_1, \ldots, e_n)$ and we are left with

$$\begin{aligned} \alpha(v_1, \ldots, v_n) &= \varepsilon_{j_1 \cdots j_n} a^{j_1}{}_1 \cdots a^{j_n}{}_n \alpha(e_1, \ldots, e_n) \\ &= \det(a^i{}_j) \, \alpha(e_1, \ldots, e_n). \end{aligned} \tag{8.6}$$

Let us consider the special case where $\alpha(e_1, \ldots, e_n) = 1$. Note that if $\{\omega^j\}$ is a basis for V^*, then

$$\omega^{j_r}(v_i) = \omega^{j_r}(e_k a^k{}_i) = a^k{}_i \, \omega^{j_r}(e_k) = a^k{}_i \, \delta^{j_r}_k = a^{j_r}{}_i.$$

Using the definition of tensor product, we can therefore write equation (8.6) as

$$\det(a^j{}_i) = \alpha(v_1, \ldots, v_n) = \varepsilon_{j_1 \cdots j_n} \, \omega^{j_1} \otimes \cdots \otimes \omega^{j_n}(v_1, \ldots, v_n)$$

which implies that the determinant function is given by

$$\alpha = \varepsilon_{j_1 \cdots j_n} \, \omega^{j_1} \otimes \cdots \otimes \omega^{j_n}.$$

In other words, if A is a matrix with columns given by $v_1, \ldots, v_n$ then $\det A = \alpha(v_1, \ldots, v_n)$. And in particular, the requirement that $\alpha(e_1, \ldots, e_n) = 1$ is just $\det I = 1$ as it should.

Suppose $A_{i_1 \cdots i_r}$ and $T^{i_1 \cdots i_r}$ (where $r \leq n = \dim V$ and $1 \leq i_k \leq n$) are both antisymmetric tensors, and consider their contraction $A_{i_1 \cdots i_r} T^{i_1 \cdots i_r}$. For any particular set of indices $i_1, \ldots, i_r$ there will be $r!$ different ordered sets $(i_1, \ldots, i_r)$. But by antisymmetry, the values of $A_{i_1 \cdots i_r}$ corresponding to each ordered set will differ only by a sign, and similarly for $T^{i_1 \cdots i_r}$. This means the product of $A_{i_1 \cdots i_r}$ times $T^{i_1 \cdots i_r}$ summed over the $r!$ ordered sets $(i_1, \ldots, i_r)$ is the same as $r!$ times a single product which we choose to be the indices $i_1, \ldots, i_r$ taken in increasing order. In other words, we have

$$A_{i_1 \cdots i_r} T^{i_1 \cdots i_r} = r! A_{|i_1 \cdots i_r|} T^{i_1 \cdots i_r}$$

where $|i_1 \cdots i_r|$ denotes the fact that we are summing over increasing sets of indices only. For example, if we have antisymmetric tensors A_{ijk} and T^{ijk} in $\mathbb{R}^3$, then

$$A_{ijk} T^{ijk} = 3! A_{|ijk|} T^{ijk} = 6 A_{123} T^{123}$$

(where, in this case of course, A_{ijk} and T^{ijk} can only differ by a scalar).

We now want to show that the Levi-Civita *symbol* defined in Section 3.1 is actually a special case of the Levi-Civita *tensor*. Consider the vector space $\mathbb{R}^3$ with the standard orthonormal basis $\{e_1, e_2, e_3\}$. We define the antisymmetric tensor $\varepsilon \in \bigwedge^3(\mathbb{R}^3)$ by the requirement that

$$\varepsilon_{123} = \varepsilon(e_1, e_2, e_3) = +1.$$

Since $\dim \bigwedge^3(\mathbb{R}^3) = 1$, this defines all the components of ε by antisymmetry: $\varepsilon_{213} = -\varepsilon_{231} = \varepsilon_{321} = -1$ etc. In other words, if $\{e_i\}$ is the standard orthonormal basis for $\mathbb{R}^3$, then $\varepsilon_{ijk} := \varepsilon(e_i, e_j, e_k)$ is just the usual Levi-Civita symbol.

If $\{\bar{e}_i = e_j a^j{}_i\}$ is any other *orthonormal* basis for $\mathbb{R}^3$ related to the first basis by an (orthogonal) transition matrix $A = (a^j{}_i)$ with determinant equal to $+1$ (see Equation (4.5)), then we also have

$$\varepsilon(\bar{e}_1, \bar{e}_2, \bar{e}_3) = \varepsilon(e_i, e_j, e_k) a^i{}_1 a^j{}_2 a^k{}_3 = \varepsilon_{ijk} a^i{}_1 a^j{}_2 a^k{}_3 = \det A = +1.$$

So we see that in any orthonormal basis related to the standard basis by an orthogonal transformation with determinant equal to $+1$, the Levi-Civita tensor behaves just like the symbol. (One says that the new basis has the same "orientation" as the original basis. We will discuss orientations later in this chapter.)

The tensor ε is called the **Levi-Civita tensor**. However, we stress that in a non-orthonormal coordinate system, it will not generally be true that $\varepsilon_{123} = +1$. And while we have defined the general ε_{ijk} as the components of a tensor, it is most common to see the **Levi-Civita** (or **permutation**) **symbol** ε_{ijk} *defined* simply as an antisymmetric *symbol* with $\varepsilon_{123} = +1$, and this is how we shall use it. For notational consistency, we also define the permutation symbol ε^{ijk} to have the same values as ε_{ijk}. (Again, for the general tensor, this is only true in an orthonormal cartesian coordinate system.)

As we saw in Section 3.2, this definition can easily be extended to an arbitrary number of dimensions. In other words, we define

$$\varepsilon_{i_1\cdots i_n} = \begin{cases} +1 & \text{if } (i_1, \dots, i_n) \text{ is an even permutation of } (1, 2, \dots, n) \\ -1 & \text{if } (i_1, \dots, i_n) \text{ is an odd permutation of } (1, 2, \dots, n) \\ 0 & \text{otherwise} \end{cases}$$

where

$$\varepsilon_{i_1\cdots i_n}\varepsilon^{i_1\cdots i_n} = n!.$$

8.4 The Exterior Product

We have seen that the tensor product of two elements of $\bigwedge^r(V)$ is not generally another element of $\bigwedge^{r+r}(V)$. However, we can define another product on $\bigwedge^r(V)$ that turns out to be of great use. This product is a broad generalization of the vector cross product in that it applies to an arbitrary number of dimensions and to spaces with a nondegenerate inner product. We adopt the convention of denoting elements of $\bigwedge^r(V)$ by Greek letters such as α, β etc.

Let us introduce some convenient notation for handling multiple indices. Instead of writing the ordered set $(i_1, \dots, i_r)$, we simply write I where the exact range will be clear from the context. Furthermore, we write $\underrightarrow{I}$ to denote the increasing sequence $(i_1 < \cdots < i_r)$. Similarly, we shall also write v_I instead of $(v_{i_1}, \dots, v_{i_r})$.

If $\alpha \in \bigwedge^r(V)$ and $\{e_1, \dots, e_n\}$ is a basis for V, then α is determined by its n^r components

$$a_I = a_{i_1\cdots i_r} = \alpha(e_{i_1}, \dots, e_{i_r}) = \alpha(e_I).$$

For each distinct collection of r indices, the antisymmetry of α tells us that the components will only differ by a sign, and hence we need only consider the components where the indices are taken in increasing order. But the number of ways we can pick r distinct indices from a collection of n possibilities is $n!/r!(n-r)!$ and therefore

$$\dim \bigwedge^r(V) = \frac{n!}{r!(n-r)!}.$$

(This is just the usual binomial coefficient $\binom{n}{r}$. We have n choices for the first index, $n-1$ choices for the second, and on down to $n-r+1$ choices for the rth index. Then the total number of choices is $n(n-1)\cdots(n-r+1) = n!/(n-r)!$. But for each distinct collection of r indices there are $r!$ different orderings, and hence we have over counted each collection by $r!$. Dividing by this factor of $r!$ yields the desired result.)

In particular, note that if $r > n$ then there is necessarily a repeated index in $\alpha(e_{i_1}, \dots, e_{i_r})$ and in this case we have $a_I = 0$ by antisymmetry. This means that there are no nonzero r-forms on an n-dimensional space if $r > n$.

To take full advantage of this notation, we first define the **generalized permutation symbol** ε by

$$\varepsilon^{j_1\cdots j_r}_{i_1\cdots i_r} = \begin{cases} +1 & \text{if } (j_1, \dots, j_r) \text{ is an even permutation of } (i_1, \dots, i_r) \\ -1 & \text{if } (j_1, \dots, j_r) \text{ is an odd permutation of } (i_1, \dots, i_r) \\ \ \ 0 & \text{otherwise} \end{cases} .$$

For example, $\varepsilon^{352}_{235} = +1$, $\varepsilon^{431}_{341} = -1$, $\varepsilon^{142}_{231} = 0$ etc. In particular, if $A = (a^j{}_i)$ is an $n \times n$ matrix, then

$$\det A = \varepsilon^{i_1\cdots i_n}_{1\cdots n} a^1{}_{i_1} \cdots a^n{}_{i_n} = \varepsilon^{1\cdots n}_{i_1\cdots i_n} a^{i_1}{}_1 \cdots a^{i_n}{}_n$$

because

$$\varepsilon^{j_1\cdots j_n}_{1\cdots n} = \varepsilon^{j_1\cdots j_n} = \varepsilon_{j_1\cdots j_n}.$$

Now, for $\alpha \in \bigwedge^r(V)$, $\beta \in \bigwedge^s(V)$ and $v_1, \dots, v_r, v_{r+1}, \dots, v_{r+s} \in V$, we have already defined their tensor product $\alpha \otimes \beta$ by

$$(\alpha \otimes \beta)(v_1, \dots, v_{r+s}) = \alpha(v_1, \dots, v_r)\beta(v_{r+1}, \dots, v_{r+s}).$$

But as we have pointed out, it is not generally true that $\alpha \otimes \beta \in \bigwedge^{r+s}(V)$. What we now wish to do is define a product $\bigwedge^r(V) \times \bigwedge^s(V) \to \bigwedge^{r+s}(V)$ that preserves antisymmetry. Recall the antisymmetrization process we defined in Section 3.1. There we took an object like T_{ij} and formed $T_{[ij]} = (1/2!)(T_{ij} - T_{ji})$. In other words, we added up all permutations of the indices with a coefficient in front of each that is the sign of the permutation, and then divided by the number of permutations. Now we want to do the same to the tensor product.

Consider the special case where $\alpha, \beta \in \bigwedge^1(V)$ are just 1-forms. If we define the product $\alpha \wedge \beta := \alpha \otimes \beta - \beta \otimes \alpha$ which is clearly antisymmetric in α and β, then acting on two vectors $u, v \in V$ we have

$$\begin{aligned} (\alpha \wedge \beta)(u, v) &= (\alpha \otimes \beta)(u, v) - (\beta \otimes \alpha)(u, v) \\ &= \alpha(u)\beta(v) - \beta(u)\alpha(v) \\ &= \alpha(u)\beta(v) - \alpha(v)\beta(u) \end{aligned}$$

which is also clearly antisymmetric in u and v.

Generalizing this result, we define the **exterior** (or **wedge** or **Grassmann**) **product**

$$\wedge : \bigwedge^r(V) \times \bigwedge^s(V) \to \bigwedge^{r+s}(V)$$

as follows. Let $\alpha \in \bigwedge^r(V)$ and $\beta \in \bigwedge^s(V)$. Then $\alpha \wedge \beta \in \bigwedge^{r+s}(V)$ is defined on $(r+s)$-tuples of vectors $v_I = (v_{i_1}, \dots, v_{i_{r+s}})$ by

$$(\alpha \wedge \beta)(v_I) := \sum_{\underrightarrow{J}} \sum_{\underrightarrow{K}} \varepsilon_I^{JK} \alpha(v_J)\beta(v_K). \tag{8.7}$$

Written out in full this is

$$\begin{aligned}(\alpha \wedge \beta)&(v_{i_1}, \dots, v_{i_{r+s}}) \\ &= \sum_{j_1 < \cdots < j_r} \sum_{k_1 < \cdots < k_s} \varepsilon_{i_1 \cdots i_{r+s}}^{j_1 \cdots j_r k_1 \cdots k_s} \alpha(v_{j_1}, \dots, v_{j_r})\beta(v_{k_1}, \dots, v_{k_s}).\end{aligned}$$

Since α has components $\alpha_I = \alpha(e_I) = \alpha(e_1, \dots, e_r)$ we can write the components of the wedge product in terms of the components of α and β separately as

$$(\alpha \wedge \beta)_I := \sum_{\underrightarrow{J}} \sum_{\underrightarrow{K}} \varepsilon_I^{JK} \alpha_J \beta_K.$$

Note also that the wedge product of a 0-form f (i.e., a function) and an r-form α is just $f \wedge \alpha = f\alpha$.

Example 8.6. Suppose $\dim V = 5$ and $\{e_1, \dots, e_5\}$ is a basis for V. If $\alpha \in \bigwedge^2(V)$ and $\beta \in \bigwedge^1(V)$, then

$$\begin{aligned}(\alpha \wedge \beta)(e_5, e_2, e_3) &= \sum_{j_1 < j_2, k} \varepsilon_{523}^{j_1 j_2 k} \alpha(e_{j_1}, e_{j_2})\beta(e_k) \\ &= \varepsilon_{523}^{235}\alpha(e_2, e_3)\beta(e_5) + \varepsilon_{523}^{253}\alpha(e_2, e_5)\beta(e_3) + \varepsilon_{523}^{352}\alpha(e_3, e_5)\beta(e_2) \\ &= \alpha(e_2, e_3)\beta(e_5) - \alpha(e_2, e_5)\beta(e_3) + \alpha(e_3, e_5)\beta(e_2).\end{aligned}$$

Example 8.7. Let us show that the wedge product is not commutative in general. If $\alpha \in \bigwedge^r(V)$ and $\beta \in \bigwedge^s(V)$, we simply compute using components:

$$\begin{aligned}(\alpha \wedge \beta)_I &= \sum_{\underrightarrow{J}} \sum_{\underrightarrow{K}} \varepsilon_I^{JK} \alpha_J \beta_K \\ &= (-1)^{rs} \sum_{\underrightarrow{J}} \sum_{\underrightarrow{K}} \varepsilon_I^{KJ} \beta_K \alpha_J\end{aligned}$$

where going from $JK = j_1 \cdots j_r k_1 \cdots k_s$ to $KJ = k_1 \cdots k_s j_1 \cdots j_r$ requires rs transpositions. But then we have the important general formula

$$\alpha \wedge \beta = (-1)^{rs} \beta \wedge \alpha \tag{8.8}$$

In particular, we see that if either r or s is even, then $\alpha \wedge \beta = \beta \wedge \alpha$, but if both r and s are odd, then $\alpha \wedge \beta = -\beta \wedge \alpha$. Therefore if r is odd we have $\alpha \wedge \alpha = 0$, but if r is even, then $\alpha \wedge \alpha$ is not necessarily zero. In particular, any 1-form α always has the property that $\alpha \wedge \alpha = 0$.

For instance, recall from Example 8.3 that the tangent space of a manifold M has basis $\{\partial/\partial x_i\}$ and corresponding dual basis $\{dx_j\}$. In $\mathbb{R}^3$ this dual basis is $\{dx, dy, dz\}$ and we see that, for example, $dx \wedge dy = -dy \wedge dx$ while $dx \wedge dx = 0$.

Our next theorem is a useful result in many computations. It is simply a contraction of indices in the permutation symbols.

Theorem 8.6. *Let* $I = (i_1, \dots, i_q)$, $J = (j_1, \dots, j_{r+s})$, $K = (k_1, \dots, k_r)$ *and* $L = (l_1, \dots, l_s)$. *Then*

$$\sum_{\underrightarrow{J}} \varepsilon^{IJ}_{1\cdots q+r+s} \varepsilon^{KL}_{J} = \varepsilon^{IKL}_{1\cdots q+r+s}$$

where I, K *and* L *are fixed quantities, and* J *is summed over all increasing subsets* $j_1 < \cdots < j_{r+s}$ *of* $\{1, \dots, q+r+s\}$.

Proof. The only nonvanishing terms on the left hand side can occur when J is a permutation of KL (or else $\varepsilon^{KL}_J = 0$), and of these possible permutations, we only have one in the sum, and that is for the increasing set $\underrightarrow{J}$. If J is an even permutation of KL, then $\varepsilon^{KL}_J = +1$, and $\varepsilon^{IJ}_{1\cdots q+r+s} = \varepsilon^{IKL}_{1\cdots q+r+s}$ since an even number of permutations is required to go from J to KL. If J is an odd permutation of KL, then $\varepsilon^{KL}_J = -1$, and $\varepsilon^{IJ}_{1\cdots q+r+s} = -\varepsilon^{IKL}_{1\cdots q+r+s}$ since an odd number of permutations is required to go from J to KL. The conclusion then follows immediately. ∎

Note that we could have let $J = (j_1, \dots, j_r)$ and left out L entirely in Theorem 8.6. The reason we included L is shown in the next example.

Example 8.8. Let us use Theorem 8.6 to give a simple proof of the associativity of the wedge product. In other words, we want to show that

$$\alpha \wedge (\beta \wedge \gamma) = (\alpha \wedge \beta) \wedge \gamma$$

for any $\alpha \in \bigwedge^q(V)$, $\beta \in \bigwedge^r(V)$ and $\gamma \in \bigwedge^s(V)$. To see this, let $I = (i_1, \dots, i_q)$, $J = (j_1, \dots, j_{r+s})$, $K = (k_1, \dots, k_r)$ and $L = (l_1, \dots, l_s)$. Then we have

$$\begin{aligned}
[\alpha \wedge (\beta \wedge \gamma)](v_1, \dots, v_{q+r+s}) &= \sum_{\underrightarrow{I}, \underrightarrow{J}} \varepsilon^{IJ}_{1\cdots q+r+s} \alpha(v_I)(\beta \wedge \gamma)(v_J) \\
&= \sum_{\underrightarrow{I}, \underrightarrow{J}} \varepsilon^{IJ}_{1\cdots q+r+s} \alpha(v_I) \sum_{\underrightarrow{K}, \underrightarrow{L}} \varepsilon^{KL}_{J} \beta(v_K)\gamma(v_L)
\end{aligned}$$

$$= \sum_{\underset{\rightarrow}{I},\underset{\rightarrow}{K},\underset{\rightarrow}{L}} \varepsilon^{IKL}_{1\cdots q+r+s} \alpha(v_I)\beta(v_K)\gamma(v_L).$$

It is easy to see that had we started with $(\alpha \wedge \beta) \wedge \gamma$, we would have arrived at the same sum. This could also have been done using components only.

As was the case with the tensor product, we simply write $\alpha \wedge \beta \wedge \gamma$ from now on. Note also that a similar calculation can be done for the wedge product of any number of terms.

Our next theorem summarizes some of the most important algebraic properties of the wedge product.

Theorem 8.7. *Suppose $\alpha, \alpha_1, \alpha_2 \in \mathcal{T}_q(V)$, $\beta, \beta_1, \beta_2 \in \mathcal{T}_r(V)$, $\gamma \in \mathcal{T}_s(V)$ and $a \in \mathcal{F}$. Then*

(i) The wedge product is bilinear. That is,

$$\begin{aligned} (\alpha_1 + \alpha_2) \wedge \beta &= \alpha_1 \wedge \beta + \alpha_2 \wedge \beta \\ \alpha \wedge (\beta_1 + \beta_2) &= \alpha \wedge \beta_1 + \alpha \wedge \beta_2 \\ (a\alpha) \wedge \beta &= \alpha \wedge (a\beta) = a(\alpha \wedge \beta) \end{aligned}$$

(ii) $\alpha \wedge \beta = (-1)^{qr} \beta \wedge \alpha$

(iii) The wedge product is associative. That is,

$$\alpha \wedge (\beta \wedge \gamma) = (\alpha \wedge \beta) \wedge \gamma.$$

Proof. Parts (ii) and (iii) were proved in Examples 8.7 and 8.8. Part (i) is left as an exercise (see Exercise 8.4.1). ∎

Example 8.9. If $\alpha_1, \ldots, \alpha_5$ are 1-forms on $\mathbb{R}^5$, let us define

$$\beta = \alpha_1 \wedge \alpha_3 + \alpha_3 \wedge \alpha_5$$

and

$$\gamma = 2\alpha_2 \wedge \alpha_4 \wedge \alpha_5 - \alpha_1 \wedge \alpha_2 \wedge \alpha_4.$$

Using the properties of the wedge product given in Theorem 8.7 we then have

$$\begin{aligned} \beta \wedge \gamma &= (\alpha_1 \wedge \alpha_3 + \alpha_3 \wedge \alpha_5) \wedge (2\alpha_2 \wedge \alpha_4 \wedge \alpha_5 - \alpha_1 \wedge \alpha_2 \wedge \alpha_4) \\ &= 2\alpha_1 \wedge \alpha_3 \wedge \alpha_2 \wedge \alpha_4 \wedge \alpha_5 - \alpha_1 \wedge \alpha_3 \wedge \alpha_1 \wedge \alpha_2 \wedge \alpha_4 \\ &\qquad + 2\alpha_3 \wedge \alpha_5 \wedge \alpha_2 \wedge \alpha_4 \wedge \alpha_5 - \alpha_3 \wedge \alpha_5 \wedge \alpha_1 \wedge \alpha_2 \wedge \alpha_4 \\ &= -2\alpha_1 \wedge \alpha_2 \wedge \alpha_3 \wedge \alpha_4 \wedge \alpha_5 - 0 + 0 - \alpha_1 \wedge \alpha_2 \wedge \alpha_3 \wedge \alpha_4 \wedge \alpha_5 \\ &= -3\alpha_1 \wedge \alpha_2 \wedge \alpha_3 \wedge \alpha_4 \wedge \alpha_5. \end{aligned}$$

In Exercise 8.4.3 you will show that any $\alpha \in \bigwedge^r(V)$ (where $r \leq n = \dim V$) can be written as

$$\alpha = \sum_{\overrightarrow{I}} \alpha(e_I)\omega^I = \sum_{i_1 < \cdots < i_r} \alpha(e_{i_1}, \ldots, e_{i_r})\omega^{i_1} \wedge \cdots \wedge \omega^{i_r}$$
$$= \alpha_{|i_1 \cdots i_n|}\omega^{i_1} \wedge \cdots \wedge \omega^{i_r} .$$

This shows that the collection $\{\omega^{i_1} \wedge \cdots \wedge \omega^{i_r}\}$ where $1 \leq i_k \leq n$ and $i_1 < \cdots < i_r$ forms a basis for $\bigwedge^r(V)$.

Example 8.10. Suppose $\alpha_1, \ldots, \alpha_r \in \bigwedge^1(V)$ and $v_1, \ldots, v_r \in V$. Using Theorem 8.6, it is easy to generalize equation (8.7) to obtain (see Exercise 8.4.2)

$$\begin{aligned}(\alpha_1 \wedge \cdots \wedge \alpha_r)(v_1, \ldots, v_r) &= \sum_{i_1 \cdots i_r} \varepsilon^{i_1 \cdots i_r}_{1 \cdots r} \alpha_1(v_{i_1}) \cdots \alpha_r(v_{i_r}) \\ &= \det(\alpha_i(v_j)).\end{aligned}$$

(Note the sum is not over any increasing indices because each α_i is only a 1-form.)

As a special case, suppose $\{e_i\}$ is a basis for V and $\{\omega^j\}$ is the corresponding dual basis. Then $\omega^j(e_i) = \delta^j_i$ and hence

$$\begin{aligned}\omega^{i_1} \wedge \cdots \wedge \omega^{i_r}(e_{j_1}, \ldots, e_{j_r}) &= \sum_{k_1 \cdots k_r} \varepsilon^{k_1 \cdots k_r}_{j_1 \cdots j_r} \omega^{i_1}(e_{k_1}) \cdots \omega^{i_r}(e_{k_r}) \\ &= \varepsilon^{i_1 \cdots i_r}_{j_1 \cdots j_r}.\end{aligned}$$

In particular, if $\dim V = n$, choosing the indices $(i_1, \ldots, i_n) = (1, \ldots, n) = (j_1, \ldots, j_n)$, we see that

$$\omega^1 \wedge \cdots \wedge \omega^n(e_1, \ldots, e_n) = 1.$$

Example 8.11. Another useful result is the following. Suppose $\dim V = n$, and let $\{\omega^1, \ldots, \omega^n\}$ be a basis for V^*. If $\alpha^1, \ldots, \alpha^n$ are any other 1-forms in $\bigwedge^1(V) = V^*$, then we may expand each α^i in terms of the ω^j as $\alpha^i = a^i{}_j\omega^j$. We then have

$$\begin{aligned}\alpha^1 \wedge \cdots \wedge \alpha^n &= a^1{}_{i_n} \cdots a^n{}_{i_n} \omega^{i_1} \wedge \cdots \wedge \omega^{i_n} \\ &= a^1{}_{i_n} \cdots a^n{}_{i_n} \varepsilon^{i_1 \cdots i_n}_{1 \cdots n} \omega^1 \wedge \cdots \wedge \omega^n \\ &= \det(a^i{}_j)\omega^1 \wedge \cdots \wedge \omega^n\end{aligned}$$

Recalling Example 8.3, if $\{\omega^i = dx^i\}$ is a local basis for a cotangent space

V^* and $\{\alpha^i = dy^i\}$ is any other local basis, then $dy^i = (\partial y^i/\partial x^j)dx^j$ and

$$\det(a^i{}_j) = \det\left(\frac{\partial y^i}{\partial x^j}\right) = \frac{\partial(y^1 \cdots y^n)}{\partial(x^1 \cdots x^n)}$$

is just the usual Jacobian of the transformation. We then have

$$dy^1 \wedge \cdots \wedge dy^n = \frac{\partial(y^1 \cdots y^n)}{\partial(x^1 \cdots x^n)}\, dx^1 \wedge \cdots \wedge dx^n.$$

The reader may recognize $dx^1 \wedge \cdots \wedge dx^n$ as the volume element on $\mathbb{R}^n$, and hence differential forms are a natural way to describe the change of variables in multiple integrals. (For an excellent treatment of tensors and differential forms see [14].)

Theorem 8.8. *If $\alpha^1, \ldots, \alpha^r \in \bigwedge^1(V)$, then $\{\alpha^1, \ldots, \alpha^r\}$ is a linearly dependent set if and only if $\alpha^1 \wedge \cdots \wedge \alpha^r = 0$.*

Proof. If $\{\alpha^1, \ldots, \alpha^r\}$ is linearly dependent, then there exists at least one vector, say α^1, such that $\alpha^1 = \sum_{j \neq 1} a_j \alpha^j$. But then

$$\begin{aligned}
\alpha^1 \wedge \cdots \wedge \alpha^r &= \left(\sum_{j\neq 1} a_j \alpha^j\right) \wedge \alpha^2 \wedge \cdots \wedge \alpha^r \\
&= \sum_{j \neq 1} a_j (\alpha^j \wedge \alpha^2 \wedge \cdots \wedge \alpha^r) \\
&= 0
\end{aligned}$$

since every term in the sum contains a repeated 1-form and hence vanishes.

Conversely, suppose $\alpha^1, \ldots, \alpha^r$ are linearly *independent*. We can then extend them to a basis $\{\alpha^1, \ldots, \alpha^n\}$ for V^* (Theorem 1.10). If $\{e_i\}$ is the corresponding dual basis for V, then $\alpha^1 \wedge \cdots \wedge \alpha^n(e_1, \ldots, e_n) = 1$ which implies $\alpha^1 \wedge \cdots \wedge \alpha^r \neq 0$. Therefore $\{\alpha^1, \ldots, \alpha^r\}$ must be linearly *dependent* if $\alpha^1 \wedge \cdots \wedge \alpha^r = 0$. ∎

Exercises

1. Prove Theorem 8.7(i).

2. Suppose $\alpha_1, \ldots, \alpha_r \in \bigwedge^1(V)$ and $v_1, \ldots, v_r \in V$. Show

$$(\alpha_1 \wedge \cdots \wedge \alpha_r)(v_1, \ldots, v_r) = \det(\alpha_i(v_j)).$$

3. Suppose $\{e_1, \dots, e_n\}$ is a basis for V and $\{\omega^1, \dots, \omega^n\}$ is the corresponding dual basis. If $\alpha \in \bigwedge^r(V)$ (where $r \le n$), show

$$\alpha = \sum_{\overrightarrow{I}} \alpha(e_I)\omega^I = \sum_{i_1<\cdots<i_r} \alpha(e_{i_1}, \dots, e_{i_r})\omega^{i_1} \wedge \cdots \wedge \omega^{i_r}$$

by applying both sides to $(e_{j_1}, \dots, e_{j_r})$.

4. (**Interior Product**) Suppose $\alpha \in \bigwedge^r(V)$ and $v, v_2, \dots, v_r \in V$. We define the $(r-1)$-form $i_v\alpha$ by

$$\begin{array}{ll} i_v\alpha = 0 & \text{if } r = 0. \\ i_v\alpha = \alpha(v) & \text{if } r = 1. \\ i_v\alpha(v_2, \dots, v_r) = \alpha(v, v_2, \dots, v_r) & \text{if } r > 1. \end{array}$$

(a) Prove $i_{u+v} = i_u + i_v$.

(b) If $\alpha \in \bigwedge^r(V)$ and $\beta \in \bigwedge^s(V)$, prove $i_v : \bigwedge^{r+s}(V) \to \bigwedge^{r+s-1}(V)$ is an **anti-derivation**, i.e.,

$$i_v(\alpha \wedge \beta) = (i_v\alpha) \wedge \beta + (-1)^r \alpha \wedge (i_v\beta).$$

(c) If $v = v^i e_i$ and $\alpha = \sum_{\overrightarrow{I}} a_{i_1\cdots i_r}\omega^{i_1} \wedge \cdots \wedge \omega^{i_r}$ where $\{\omega^i\}$ is the basis dual to $\{e_i\}$, show

$$i_v\alpha = \sum_{i_2<\cdots<i_r} b_{i_2\cdots i_r}\omega^{i_2} \wedge \cdots \wedge \omega^{i_r}$$

where

$$b_{i_2\cdots i_r} = \sum_j v^j a_{j i_2\cdots i_r}.$$

(d) If $\alpha = f^1 \wedge \cdots \wedge f^r$ where each f^k is a 1-form, show

$$\begin{aligned} i_v\alpha &= \sum_{k=1}^r (-1)^{k-1} f^k(v) f^1 \wedge \cdots \wedge f^{k-1} \wedge f^{k+1} \wedge \cdots \wedge f^r \\ &= \sum_{k=1}^r (-1)^{k-1} f^k(v) f^1 \wedge \cdots \wedge \widehat{f^k} \wedge \cdots \wedge f^r \end{aligned}$$

where the $\widehat{\ }$ means the term f^k is to be deleted from the expression.

5. Let $V = \mathbb{R}^n$ have the standard basis $\{e_i\}$, and let the corresponding dual basis for V^* be $\{\omega^i\}$.

(a) If $u, v \in V$, show

$$\omega^i \wedge \omega^j(u, v) = \begin{vmatrix} u^i & v^i \\ u^j & v^j \end{vmatrix}$$

and that this is $\pm$ the area of the parallelogram spanned by the projection of u and v onto the $x^i x^j$-plane. What do you think is the significance of the different signs?

(b) Generalize this to $\omega^{i_1} \wedge \cdots \wedge \omega^{i_r}$ where $r \leq n$.

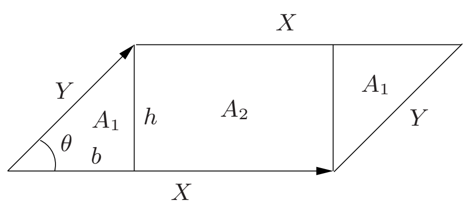
6. Let $I = (i_1, \ldots, i_q)$, $J = (j_1, \ldots, j_p)$, and $K = (k_1, \ldots, k_q)$. Prove the following generalization of equation (3.6):

$$\sum_{\overrightarrow{J}} \varepsilon^{JI}_{1\cdots p+q}\, \varepsilon^{1\cdots p+q}_{JK} = \varepsilon^{I}_{K} = q!\delta^{[i_1}_{k_1} \cdots \delta^{i_q]}_{k_q}.$$

8.5 Volumes in $\mathbb{R}^3$

Instead of starting out with an abstract presentation of volumes, we shall first go through an intuitive elementary discussion beginning with $\mathbb{R}^2$, then going to $\mathbb{R}^3$, and finally generalizing to $\mathbb{R}^n$ in the next section.

First consider a parallelogram in $\mathbb{R}^2$ (with the usual norm) defined by the vectors X and Y as shown.



Note that $h = \|Y\| \sin\theta$ and $b = \|Y\| \cos\theta$, and also that the area of each triangle is given by $A_1 = (1/2)bh$. Then the area of the rectangle is given by $A_2 = (\|X\| - b)h$, and the area of the entire parallelogram is given by

$$A = 2A_1 + A_2 = bh + (\|X\| - b)h = \|X\|\, h = \|X\|\, \|Y\| \sin\theta. \tag{8.9}$$

The reader should recognize this as the magnitude of the elementary "vector cross product" $X \times Y$ of the ordered pair of vectors (X, Y) that is *defined* to have a direction normal to the plane spanned by X and Y, and given by the "right hand rule" (i.e., *out* of the plane in this case).

If we define the usual orthogonal coordinate system with the x-axis parallel to the vector X, then

$$X = (x^1, x^2) = (\|X\|, 0)$$

and

$$Y = (y^1, y^2) = (\|Y\| \cos\theta, \|Y\| \sin\theta)$$

and hence we see that the determinant with columns formed from the vectors X and Y is just

$$\begin{vmatrix} x^1 & y^1 \\ x^2 & y^2 \end{vmatrix} = \begin{vmatrix} \|X\| & \|Y\| \cos\theta \\ 0 & \|Y\| \sin\theta \end{vmatrix} = \|X\|\, \|Y\| \sin\theta = A. \tag{8.10}$$

Notice that if we interchanged the vectors X and Y in the diagram, then the determinant would change sign and the vector $X \times Y$ (which by definition has a direction dependent on the *ordered* pair (X, Y)) would point *into* the page. Thus the area of a parallelogram (which is always positive by definition) defined by two vectors in $\mathbb{R}^2$ is in general given by the absolute value of the determinant in equation (8.10).

In terms of the usual inner product (or "dot product") $\langle \cdot\,, \cdot \rangle$ on $\mathbb{R}^2$, we have $\langle X, X\rangle = \|X\|^2$ and $\langle X, Y\rangle = \langle Y, X\rangle = \|X\|\,\|Y\|\cos\theta$, and hence

$$\begin{aligned} A^2 &= \|X\|^2\,\|Y\|^2 \sin^2\theta \\ &= \|X\|^2\,\|Y\|^2\,(1-\cos^2\theta) \\ &= \|X\|^2\,\|Y\|^2 - \langle X, Y\rangle^2. \end{aligned}$$

Therefore we see that the area is also given by the positive square root of the determinant

$$A^2 = \begin{vmatrix} \langle X, X\rangle & \langle X, Y\rangle \\ \langle Y, X\rangle & \langle Y, Y\rangle \end{vmatrix}. \tag{8.11}$$

It is also worth noting that the inner product may be written in the form $\langle X, Y\rangle = x^1y^1 + x^2y^2$, and thus in terms of matrices we may write

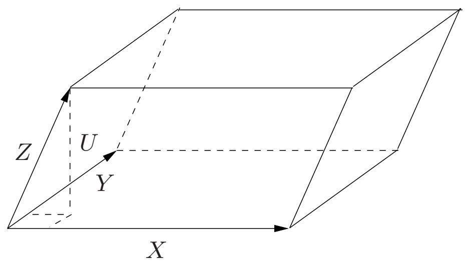
$$\begin{bmatrix} \langle X, X\rangle & \langle X, Y\rangle \\ \langle Y, X\rangle & \langle Y, Y\rangle \end{bmatrix} = \begin{bmatrix} x^1 & x^2 \\ y^1 & y^2 \end{bmatrix} \begin{bmatrix} x^1 & y^1 \\ x^2 & y^2 \end{bmatrix}.$$

Hence taking the determinant of this equation (using Theorems 3.7 and 3.1), we find (at least in $\mathbb{R}^2$) that the determinant in equation (8.11) also implies the area is given by the absolute value of the determinant in equation (8.10).

It is now easy to extend this discussion to a parallelogram in $\mathbb{R}^3$. Indeed, if $X = (x^1, x^2, x^3)$ and $Y = (y^1, y^2, y^3)$ are vectors in $\mathbb{R}^3$, then equation (8.9) is unchanged because any two vectors in $\mathbb{R}^3$ define the plane $\mathbb{R}^2$ spanned by the two vectors. Equation (8.11) also remains unchanged since its derivation did not depend on the specific coordinates of X and Y in $\mathbb{R}^2$. However, the left hand part of equation (8.10) does not apply (although we will see below that the three-dimensional version determines a volume in $\mathbb{R}^3$).

As a final remark on parallelograms, note that if X and Y are linearly dependent, then $aX + bY = 0$ so that $Y = -(a/b)X$, and hence X and Y are co-linear. Therefore θ equals 0 or π so that all equations for the area in terms of $\sin\theta$ are equal to zero. Since X and Y are dependent, this also means that the determinant in equation (8.10) equals zero, and everything is consistent.

We now take a look at volumes in $\mathbb{R}^3$. Consider three linearly independent vectors $X = (x^1, x^2, x^3)$, $Y = (y^1, y^2, y^3)$ and $Z = (z^1, z^2, z^3)$, and consider the parallelepiped with edges defined by these three vectors (in the given order (X, Y, Z)).



We claim that the volume of this parallelepiped is given by both the positive square root of the determinant

$$\begin{vmatrix} \langle X,X\rangle & \langle X,Y\rangle & \langle X,Z\rangle \\ \langle Y,X\rangle & \langle Y,Y\rangle & \langle Y,Z\rangle \\ \langle Z,X\rangle & \langle Z,Y\rangle & \langle Z,Z\rangle \end{vmatrix} \tag{8.12}$$

and the absolute value of the determinant

$$\begin{vmatrix} x^1 & y^1 & z^1 \\ x^2 & y^2 & z^2 \\ x^3 & y^3 & z^3 \end{vmatrix}. \tag{8.13}$$

To see this, first note that the volume of the parallelepiped is given by the product of the area of the base times the height, where the area A of the base is given by equation (8.11) and the height $\|U\|$ is just the projection of Z onto the orthogonal complement in $\mathbb{R}^3$ of the space spanned by X and Y. In other words, if W is the subspace of $V = \mathbb{R}^3$ spanned by X and Y, then (by Theorem 1.22) $V = W^\perp \oplus W$, and hence by Theorem 1.12 we may write

$$Z = U + aX + bY$$

where $U \in W^\perp$ and a, $b \in \mathbb{R}$ are uniquely determined (the uniqueness of a and b actually follows from Theorem 1.3 together with Theorem 1.12).

By definition we have $\langle X,U\rangle = \langle Y,U\rangle = 0$, and therefore

$$\langle X,Z\rangle = a\,\|X\|^2 + b\langle X,Y\rangle \tag{8.14a}$$

$$\langle Y,Z\rangle = a\langle Y,X\rangle + b\,\|Y\|^2 \tag{8.14b}$$

$$\langle U,Z\rangle = \|U\|^2\,. \tag{8.14c}$$

We now wish to solve the first two of these equations for a and b by Cramer's rule (Theorem 3.11). Note the determinant of the matrix of coefficients is just equation (8.11), and hence is just the square of the area A of the base of the parallelepiped.

Applying Cramer's rule we have

$$aA^2 = \begin{vmatrix} \langle X,Z\rangle & \langle X,Y\rangle \\ \langle Y,Z\rangle & \langle Y,Y\rangle \end{vmatrix} = -\begin{vmatrix} \langle X,Y\rangle & \langle X,Z\rangle \\ \langle Y,Y\rangle & \langle Y,Z\rangle \end{vmatrix}$$

$$bA^2 = \begin{vmatrix} \langle X, X\rangle & \langle X, Z\rangle \\ \langle Y, X\rangle & \langle Y, Z\rangle \end{vmatrix}.$$

Denoting the volume by $\mathrm{Vol}(X, Y, Z)$, we now have (using equation (8.14c) together with $U = Z - aX - bY$)

$$\mathrm{Vol}^2(X, Y, Z) = A^2 \|U\|^2 = A^2\langle U, Z\rangle = A^2(\langle Z, Z\rangle - a\langle X, Z\rangle - b\langle Y, Z\rangle)$$

so that substituting the expressions for A^2, aA^2 and bA^2 we find

$$\begin{aligned}\mathrm{Vol}^2(X, Y, Z) = \langle Z, Z\rangle \begin{vmatrix} \langle X, X\rangle & \langle X, Y\rangle \\ \langle Y, X\rangle & \langle Y, Y\rangle \end{vmatrix} &+ \langle X, Z\rangle \begin{vmatrix} \langle X, Y\rangle & \langle X, Z\rangle \\ \langle Y, Y\rangle & \langle Y, Z\rangle \end{vmatrix} \\ &- \langle Y, Z\rangle \begin{vmatrix} \langle X, X\rangle & \langle X, Z\rangle \\ \langle Y, X\rangle & \langle Y, Z\rangle \end{vmatrix}.\end{aligned}$$

Using $\langle X, Y\rangle = \langle Y, X\rangle$ etc., we see this is just the expansion of a determinant by minors of the third row, and hence (using $\det A^T = \det A$)

$$\begin{aligned}\mathrm{Vol}^2(X, Y, Z) &= \begin{vmatrix} \langle X, X\rangle & \langle Y, X\rangle & \langle Z, X\rangle \\ \langle X, Y\rangle & \langle Y, Y\rangle & \langle Z, Y\rangle \\ \langle X, Z\rangle & \langle Y, Z\rangle & \langle Z, Z\rangle \end{vmatrix} \\ &= \begin{vmatrix} x^1 & x^2 & x^3 \\ y^1 & y^2 & y^3 \\ z^1 & z^2 & z^3 \end{vmatrix} \begin{vmatrix} x^1 & y^1 & z^1 \\ x^2 & y^2 & z^2 \\ x^3 & y^3 & z^3 \end{vmatrix} = \begin{vmatrix} x^1 & y^1 & z^1 \\ x^2 & y^2 & z^2 \\ x^3 & y^3 & z^3 \end{vmatrix}^2.\end{aligned}$$

We remark that if the collection $\{X, Y, Z\}$ is linearly dependent, then the volume of the parallelepiped degenerates to zero (since at least one of the parallelograms that form the sides will have zero area). This agrees with the fact that the determinant of equation (8.13) will vanish if two rows are linearly dependent. We also note that the area of the base is given by

$$\|X \times Y\| = \|X\| \, \|Y\| \sin \angle(X, Y)$$

where the direction of the vector $X \times Y$ is up (in this case). Therefore the projection of Z in the direction of $X \times Y$ is just Z dotted into a unit vector in the direction of $X \times Y$, and hence the volume of the parallelepiped is given by the number $Z \cdot (X \times Y)$. This is the so-called **scalar triple product** that should be familiar from elementary courses. We leave it to the reader to show the scalar triple product is given by the determinant of equation (8.13) (see Exercise 8.5.1).

Finally, note that if any two of the vectors X, Y, Z in equation (8.13) are interchanged, then the determinant changes sign even though the volume is unaffected (since it must be positive). This observation will form the basis for the concept of "orientation" to be defined later.

Exercises

1. Show $Z \cdot (X \times Y)$ is given by the determinant in equation (8.13).

2. Find the area of the parallelogram whose vertices are:
   (a) $(0,0)$, $(1,3)$, $(-2,1)$ and $(-1,4)$.
   (b) $(2,4)$, $(4,5)$, $(5,2)$ and $(7,3)$.
   (c) $(-1,3)$, $(1,5)$, $(3,2)$ and $(5,4)$.
   (d) $(0,0,0)$, $(1,-2,2)$, $(3,4,2)$ and $(4,2,4)$.
   (e) $(2,2,1)$, $(3,0,6)$, $(4,1,5)$ and $(1,1,2)$.

3. Find the volume of the parallelepipeds whose adjacent edges are the vectors:
   (a) $(1,1,2)$, $(3,-1,0)$ and $(5,2,-1)$.
   (b) $(1,1,0)$, $(1,0,1)$ and $(0,1,1)$.

4. Prove both algebraically and geometrically that the parallelogram with edges X and Y has the same area as the parallelogram with edges X and $Y + aX$ for any scalar a.

5. Prove both algebraically and geometrically that the volume of the parallelepiped in $\mathbb{R}^3$ with edges X, Y and Z is equal to the volume of the parallelepiped with edges X, Y and $Z + aX + bY$ for any scalars a and b.

6. Show that the parallelepiped in $\mathbb{R}^3$ defined by the three vectors $(2,2,1)$, $(1,-2,2)$ and $(-2,1,2)$ is a cube. Find the volume of this cube.

8.6 Volumes in $\mathbb{R}^n$

Now that we have a feeling for volumes in $\mathbb{R}^3$ expressed as determinants, let us prove the analogous results in $\mathbb{R}^n$. To begin with, we note that parallelograms defined by the vectors X and Y in either $\mathbb{R}^2$ or $\mathbb{R}^3$ contain all points (i.e., vectors) of the form $aX + bY$ for any $a, b \in [0,1]$. Similarly, given three linearly independent vectors $X, Y, Z \in \mathbb{R}^3$, we may define the parallelepiped with these vectors as edges to be that subset of $\mathbb{R}^3$ containing all vectors of the form $aX + bY + cZ$ where $0 \le a, b, c \le 1$. The corners of the parallelepiped are the points $\delta_1 X + \delta_2 Y + \delta_3 Z$ where each δ_i is either 0 or 1.

Generalizing these observations, given any r linearly independent vectors $X_1, \dots, X_r \in \mathbb{R}^n$, we define an ***r*-dimensional parallelepiped** as the set of all vectors of the form $a_1 X_1 + \cdots + a_r X_r$ where $0 \le a_i \le 1$ for each $i = 1, \dots, r$. In $\mathbb{R}^3$, by a **1-volume** we mean a length, a **2-volume** means an area, and a **3-volume** is just the usual volume.

To define the volume of an r-dimensional parallelepiped we proceed by induction on r. In particular, if X is a nonzero vector (i.e., a 1-dimensional parallelepiped) in $\mathbb{R}^n$, we define its 1-volume to be its length $\langle X, X\rangle^{1/2}$. Proceeding, suppose the $(r-1)$-dimensional volume of an $(r-1)$-dimensional parallelepiped has been defined. If we let P_r denote the r-dimensional parallelepiped defined by the r linearly independent vectors $X_1, \ldots, X_r$, then we say the **base** of P_r is the $(r-1)$-dimensional parallelepiped defined by the $r-1$ vectors $X_1, \ldots, X_{r-1}$, and the **height** of P_r is the length of the projection of X_r onto the orthogonal complement in $\mathbb{R}^n$ of the space spanned by $X_1, \ldots, X_{r-1}$. According to our induction hypothesis, the volume of an $(r-1)$-dimensional parallelepiped has already been defined. Therefore we define the ***r*-volume** of P_r to be the product of its height times the $(r-1)$-dimensional volume of its base.

The reader may wonder whether or not the r-volume of an r-dimensional parallelepiped in any way depends on which of the r vectors is singled out for projection. We proceed as if it does not and then, after the next theorem, we shall show that this is indeed the case.

Theorem 8.9. *Let P_r be the r-dimensional parallelepiped defined by the r linearly independent vectors $X_1, \ldots, X_r \in \mathbb{R}^n$. Then the r-volume of P_r is the positive square root of the determinant*

$$\begin{vmatrix} \langle X_1, X_1\rangle & \langle X_1, X_2\rangle & \cdots & \langle X_1, X_r\rangle \\ \langle X_2, X_1\rangle & \langle X_2, X_2\rangle & \cdots & \langle X_2, X_r\rangle \\ \vdots & \vdots & & \vdots \\ \langle X_r, X_1\rangle & \langle X_r, X_2\rangle & \cdots & \langle X_r, X_r\rangle \end{vmatrix}. \tag{8.15}$$

Proof. For the case $r = 1$ we see the theorem is true by the definition of length (or 1-volume) of a vector. Proceeding by induction, we assume the theorem is true for an $(r-1)$-dimensional parallelepiped, and we show that it is also true for an r-dimensional parallelepiped. Hence, let us write

$$A^2 = \mathrm{Vol}^2(P_{r-1}) = \begin{vmatrix} \langle X_1, X_1\rangle & \langle X_1, X_2\rangle & \cdots & \langle X_1, X_{r-1}\rangle \\ \langle X_2, X_1\rangle & \langle X_2, X_2\rangle & \cdots & \langle X_2, X_{r-1}\rangle \\ \vdots & \vdots & & \vdots \\ \langle X_{r-1}, X_1\rangle & \langle X_{r-1}, X_2\rangle & \cdots & \langle X_{r-1}, X_{r-1}\rangle \end{vmatrix}$$

for the volume of the $(r-1)$-dimensional base of P_r. Just as we did in our discussion of volumes in $\mathbb{R}^3$, we write X_r in terms of its projection U onto the orthogonal complement of the space spanned by the $r-1$ vectors $X_1, \ldots, X_r$. This means we can write

$$X_r = U + a_1 X_1 + \cdots + a_{r-1} X_{r-1}$$

where $\langle U, X_i\rangle = 0$ for $i = 1, \ldots, r-1$ and $\langle U, X_r\rangle = \langle U, U\rangle$. We thus have the system of equations

$$\begin{array}{ccccccc} a_1\langle X_1, X_1\rangle & + & a_2\langle X_1, X_2\rangle & + \cdots + & a_{r-1}\langle X_1, X_{r-1}\rangle & = & \langle X_1, X_r\rangle \\ a_1\langle X_2, X_1\rangle & + & a_2\langle X_2, X_2\rangle & + \cdots + & a_{r-1}\langle X_2, X_{r-1}\rangle & = & \langle X_2, X_r\rangle \\ \vdots & & \vdots & & \vdots & & \vdots \\ a_1\langle X_{r-1}, X_1\rangle & + & a_2\langle X_{r-1}, X_2\rangle & + \cdots + & a_{r-1}\langle X_{r-1}, X_{r-1}\rangle & = & \langle X_{r-1}, X_r\rangle \end{array}$$

We write $M_1, \ldots, M_{r-1}$ for the minors of the first $r-1$ elements of the last row in equation (8.15). Solving the above system for the a_i using Cramers rule we obtain

$$\begin{aligned} A^2 a_1 &= (-1)^{r-2} M_1 \\ A^2 a_2 &= (-1)^{r-3} M_2 \\ &\vdots \\ A^2 a_{r-1} &= M_{r-1} \end{aligned}$$

where the factors of $(-1)^{r-k-1}$ in $A^2 a_k$ result from moving the last column of equation (8.15) over to become the kth column of the kth minor matrix.

Using this result, we now have

$$\begin{aligned} A^2 U &= A^2(-a_1 X_1 - a_2 X_2 - \cdots - a_{r-1} X_{r-1} + X_r) \\ &= (-1)^{r-1} M_1 X_1 + (-1)^{r-2} M_2 X_2 + \cdots + (-1) M_{r-1} X_{r-1} + A^2 X_r \end{aligned}$$

and hence, using $\|U\|^2 = \langle U, U\rangle = \langle U, X_r\rangle$, we find that (since $(-1)^{-k} = (-1)^k$)

$$\begin{aligned} A^2 \|U\|^2 &= A^2 \langle U, X_r\rangle \\ &= (-1)^{r-1} M_1 \langle X_r, X_1\rangle + (-1)(-1)^{r-1} M_2 \langle X_r, X_2\rangle + \cdots + A^2 \langle X_r, X_r\rangle \\ &= (-1)^{r-1}[M_1 \langle X_r, X_1\rangle - M_2 \langle X_r, X_2\rangle + \cdots + (-1)^{r-1} A^2 \langle X_r, X_r\rangle]. \end{aligned}$$

Now note that the right hand side of this equation is precisely the expansion of equation (8.15) by minors of the last row, and the left hand side is by definition the square of the r-volume of the r-dimensional parallelepiped P_r. This also shows that the determinant (8.15) is positive. ∎

This result may also be expressed in terms of the matrix $(\langle X_i, X_j\rangle)$ as

$$\mathrm{Vol}(P_r) = [\det(\langle X_i, X_j\rangle)]^{1/2}.$$

The most useful form of this theorem is given in the following corollary.

Corollary. *The n-volume of the n-dimensional parallelepiped in $\mathbb{R}^n$ defined by the vectors $X_1, \ldots, X_n$ where each X_i has coordinates $({x^1}_i, \ldots, {x^n}_i)$ is the absolute value of the determinant of the matrix X given by*

$$X = \begin{bmatrix} {x^1}_1 & {x^1}_2 & \cdots & {x^1}_n \\ {x^2}_1 & {x^2}_2 & \cdots & {x^2}_n \\ \vdots & \vdots & & \vdots \\ {x^n}_1 & {x^n}_2 & \cdots & {x^n}_n \end{bmatrix}.$$

Proof. Note $(\det X)^2 = (\det X)(\det X^T) = \det XX^T$ is just the determinant in equation (8.15) of Theorem 8.9, which is the square of the volume. In other words, $\mathrm{Vol}(P_n) = |\det X|$. ∎

Prior to this theorem, we asked whether or not the r-volume depended on which of the r vectors is singled out for projection. We can now easily show that it does not.

Suppose we have an r-dimensional parallelepiped defined by r linearly independent vectors, and let us label these vectors $X_1, \ldots, X_r$. According to Theorem 8.9, we project X_r onto the space orthogonal to the space spanned by $X_1, \ldots, X_{r-1}$, and this leads to the determinant (8.15). If we wish to project any other vector instead, then we may simply relabel these r vectors to put a different one into position r. In other words, we have made some permutation of the indices in equation (8.15). However, remember that any permutation is a product of transpositions, and hence we need only consider the effect of a single interchange of two indices.

Notice, for example, that the indices 1 and r only occur in rows 1 and r as well as in columns 1 and r. And in general, indices i and j only occur in the ith and jth rows and columns. But we also see that the matrix corresponding to equation (8.15) is symmetric in these indices about the main diagonal, and hence an interchange of the indices i and j has the effect of interchanging *both* rows i and j *as well as* columns i and j in exactly the same manner. Therefore, because we have interchanged the same rows and columns there will be no sign change, and hence the determinant (8.15) remains unchanged. In particular, it always remains positive. It now follows that the volume we have defined is indeed independent of which of the r vectors is singled out to be the height of the parallelepiped.

Now note that according to the above corollary, we know

$$\mathrm{Vol}(P_n) = \mathrm{Vol}(X_1, \ldots, X_n) = |\det X|$$

which is always positive. While our discussion just showed that $\mathrm{Vol}(X_1, \ldots, X_n)$ is independent of any permutation of indices, the actual value of $\det X$ can change sign upon any such permutation. Because of this, we say that the vectors $(X_1, \ldots, X_n)$ are **positively oriented** if $\det X > 0$, and **negatively oriented** if $\det X < 0$. Thus the orientation of a set of vectors depends on the order in which they are written. To take into account the sign of $\det X$, we define the **oriented volume** $\mathrm{Vol}_o(X_1, \ldots, X_n)$ to be $+\mathrm{Vol}(X_1, \ldots, X_n)$ if $\det X \geq 0$, and $-\mathrm{Vol}(X_1, \ldots, X_n)$ if $\det X < 0$. We will return to a careful discussion of orientation in Section 8.8. We also remark that $\det X$ is always nonzero as long as the vectors $(X_1, \ldots, X_n)$ are linearly independent. Thus the above corollary may be expressed in the form

$$\mathrm{Vol}_o(X_1, \ldots, X_n) = \det(X_1, \ldots, X_n)$$

where $\det(X_1, \ldots, X_n)$ means the determinant as a function of the column vectors X_i.

Exercises

1. Find the 3-volume of the three-dimensional parallelepipeds in $\mathbb{R}^4$ defined by the vectors:

   (a) $(2, 1, 0, -1)$, $(3, -1, 5, 2)$ and $(0, 4, -1, 2)$.
   (b) $(1, 1, 0, 0)$, $(0, 2, 2, 0)$ and $(0, 0, 3, 3)$.

2. Find the 2-volume of the parallelogram in $\mathbb{R}^4$ two of whose edges are the vectors $(1, 3, -1, 6)$ and $(-1, 2, 4, 3)$.

3. Prove that if the vectors $X_1, X_2, \dots, X_r$ are mutually orthogonal, the r-volume of the parallelepiped defined by them is equal to the product of their lengths.

4. Prove that r vectors $X_1, X_2, \dots, X_r$ in $\mathbb{R}^n$ are linearly dependent if and only if the determinant (8.15) is equal to zero.

8.7 Linear Transformations and Volumes

One of the most useful applications of Theorem 8.9 and its corollary relates to linear mappings. In fact, this is the approach usually followed in deriving the change of variables formula for multiple integrals. Let $\{e_i\}$ be an orthonormal basis for $\mathbb{R}^n$, and let C_n denote the **unit cube** in $\mathbb{R}^n$. In other words,

$$C_n = \{t_1 e_1 + \cdots + t_n e_n \in \mathbb{R}^n : 0 \le t_i \le 1\}.$$

This is similar to the definition of P_r given previously.

Now let $A : \mathbb{R}^n \to \mathbb{R}^n$ be a linear transformation. Then the matrix of A relative to the basis $\{e_i\}$ is defined by $A(e_i) = e_j a^j{}_i$. Let us write the image of e_i as X_i, so that $X_i = A(e_i) = e_j a^j{}_i$. This means the column vector X_i has components $(a^1{}_i, \dots, a^n{}_i)$. Under the transformation A, the image of C_n becomes

$$A(C_n) = A\Big(\sum t_i e_i\Big) = \sum t_i A(e_i) = \sum t_i X_i$$

(where $0 \le t_i \le 1$) which is just the parallelepiped P_n spanned by the vectors $(X_1, \dots, X_n)$. Therefore the volume of $P_n = A(C_n)$ is given by

$$\operatorname{Vol}(P_n) = \operatorname{Vol}(A(C_n)) = \det(X_1, \dots, X_n) = \left|\det(a^j{}_i)\right|.$$

Recalling that the determinant of a linear transformation is defined to be the determinant of its matrix representation, we have proved the next result.

Theorem 8.10. *Let C_n be the unit cube in $\mathbb{R}^n$ spanned by the orthonormal basis vectors $\{e_i\}$. If $A : \mathbb{R}^n \to \mathbb{R}^n$ is a linear transformation and $P_n = A(C_n)$, then* $\operatorname{Vol}(P_n) = \operatorname{Vol}(A(C_n)) = |\det A|$.

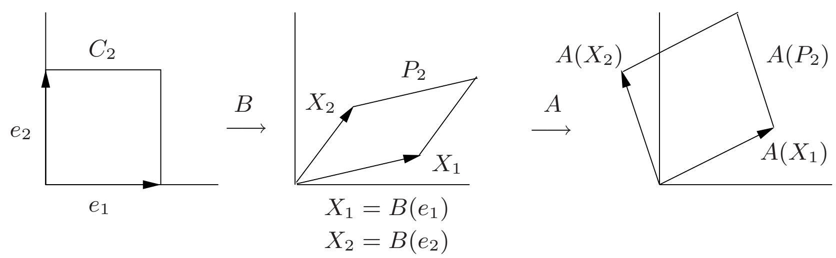
It is quite simple to generalize this result somewhat to include the image of an n-dimensional parallelepiped under a linear transformation A. First, we note that any parallelepiped P_n is just the image of C_n under some linear transformation B. Indeed, if $P_n = \{t_1X_1 + \cdots + t_nX_n : 0 \le t_i \le 1\}$ for some set of vectors X_i, then we may define the transformation B by $B(e_i) = X_i$, and hence $P_n = B(C_n)$. Thus

$$A(P_n) = A(B(C_n)) = (A \circ B)(C_n)$$

and therefore (using Theorem 8.10 along with the fact that the matrix of the composition of two transformations is the matrix product)

$$\begin{aligned}\mathrm{Vol}(A(P_n)) &= \mathrm{Vol}[(A \circ B)(C_n)] = |\det(A \circ B)| = |\det A|\,|\det B| \\ &= |\det A|\,\mathrm{Vol}(P_n).\end{aligned}$$

In other words, $|\det A|$ is a measure of how much the volume of the parallelepiped changes under the linear transformation A. See the figure below for a picture of this in $\mathbb{R}^2$.



We summarize this discussion as a corollary to Theorem 8.10.

Corollary. *Suppose P_n is an n-dimensional parallelepiped in $\mathbb{R}^n$, and let $A : \mathbb{R}^n \to \mathbb{R}^n$ be a linear transformation. Then* $\mathrm{Vol}(A(P_n)) = |\det A|\,\mathrm{Vol}(P_n)$.

Now that we have an intuitive grasp of these concepts, let us look at this material from the point of view of exterior algebra. This more sophisticated approach is of great use in e.g., the theory of integration. We begin by showing that a linear transformation $T \in L(U, V)$ induces a linear transformation $T^* \in L(V^*, U^*)$ in a natural way.

Let U and V be real vector spaces. Given a linear transformation $T \in L(U, V)$ we define the mapping $T^* \in L(V^*, U^*)$ by

$$T^*\phi = \phi \circ T$$

for all $\phi \in V^*$. (The mapping T^* is frequently written T^t.) In other words, for any $u \in U$ we have

$$(T^*\phi)(u) = (\phi \circ T)(u) = \phi(T(u)) \in \mathcal{F}.$$

To show $T^*\phi$ is indeed an element of U^*, we simply note that for u_1, $u_2 \in U$ and a, $b \in \mathcal{F}$ we have (using the linearity of T and ϕ)

$$\begin{aligned}(T^*\phi)(au_1 + bu_2) &= \phi(T(au_1 + bu_2)) \\ &= \phi(aT(u_1) + bT(u_2)) \\ &= a\phi(T(u_1)) + b\phi(T(u_2)) \\ &= a(T^*\phi)(u_1) + b(T^*\phi)(u_2)\end{aligned}$$

Furthermore, it is easy to see the mapping T^* is linear since for any ϕ, $\theta \in V^*$ and a, $b \in \mathcal{F}$ we have

$$T^*(a\phi + b\theta) = (a\phi + b\theta) \circ T = a(\phi \circ T) + b(\theta \circ T) = a(T^*\phi) + b(T^*\theta).$$

Hence we have proved the next result.

Theorem 8.11. *Suppose $T \in L(U, V)$, and define the mapping $T^* : V^* \to U^*$ by $T^*\phi = \phi \circ T$ for all $\phi \in V^*$. Then $T^* \in L(V^*, U^*)$.*

The linear mapping T^* defined in this theorem is called the **transpose** of the linear transformation T. The reason for the name transpose is shown in the next theorem.

Theorem 8.12. *Let $T \in L(U, V)$ have matrix representation $A = (a_{ij})$ with respect to the bases $\{e_1, \dots, e_m\}$ for U and $\{f_1, \dots, f_n\}$ for V. Let the dual spaces U^* and V^* have the corresponding dual bases $\{e^i\}$ and $\{f^i\}$. Then the matrix representation of $T^* \in L(V^*, U^*)$ with respect to these bases for V^* and U^* is given by A^T.*

Proof. By definition of $A = (a_{ij})$ we have

$$Te_i = \sum_{j=1}^{n} f_j a_{ji}$$

for each $i = 1, \dots, m$. Define the matrix representation $B = (b_{ij})$ of T^* by

$$T^* f^i = \sum_{j=1}^{m} e^j b_{ji}$$

for each $i = 1, \dots, n$. Applying the left side of this equation to an arbitrary basis vector e_k, we find

$$(T^* f^i)e_k = f^i(Te_k) = f^i\Big(\sum_j f_j a_{jk}\Big) = \sum_j f^i(f_j)a_{jk} = \sum_j \delta^i_j a_{jk} = a_{ik}$$

while the right side yields

$$\sum_j b_{ji} e^j(e_k) = \sum_j b_{ji} \delta^j_k = b_{ki}.$$

Therefore $b_{ki} = a_{ik} = a^T_{ki}$, and thus $B = A^T$. ∎

An alternative way to arrive at this same conclusion that also illustrates a useful technique is to take the result $(T^* f^i)e_k = a_{ik}$ shown above and write

$$(T^* f^i)e_k = a_{ik} = \sum_j a_{ij} \delta_{jk} = \sum_j a_{ij} e^j(e_k)\,.$$

Because the $\{e_k\}$ form a basis, this implies directly that $T^* f^i = \sum_j a_{ij} e^j$. Since this now sums over the column index of (a_{ij}), it shows that the matrix representation of T^* is the transpose of the representation of T.

Going back to the summation convention, Theorem 8.12 says that

$$T^* f^i = a^i{}_j e^j \tag{8.16}$$

We will use this result frequently below. In fact, referring back to Example 8.1, we know that the dual basis consists of *row* vectors. Then just as we showed that a linear transformation T takes the ith basis vector e_i to the ith *column* of the matrix representation of T, we now see that

$$a^i{}_j e^j = a^i{}_1 \begin{bmatrix} 1 & \cdots & 0 \end{bmatrix} + \cdots + a^i{}_n \begin{bmatrix} 0 & \cdots & 1 \end{bmatrix} = \begin{bmatrix} a^i{}_1 & \cdots & a^i{}_n \end{bmatrix}$$

which is the ith *row* of the matrix representation of T. In other words, while $(a^i{}_j)$ acts on the left of the basis vectors e_i, it acts to the *right* of the basis covectors e^i: if $e^i = [0 \cdots 1 \cdots 0]$ has a 1 in the ith position, then

$$\begin{bmatrix} 0 & \cdots & 1 & \cdots & 0 \end{bmatrix} \begin{bmatrix} a^1{}_1 & \cdots & a^1{}_n \\ \vdots & & \vdots \\ a^n{}_1 & \cdots & a^n{}_n \end{bmatrix} = \begin{bmatrix} a^i{}_1 & \cdots & a^i{}_n \end{bmatrix}.$$

Summarizing, we see that T^* *takes the ith basis covector to the ith row of* $(a^i{}_j)$.

Since A^T is the matrix representation of T^*, certain properties of T^* follow naturally. For example, if $T_1 \in L(V, W)$ and $T_2 \in L(U, V)$, then $(T_1 \circ T_2)^* = T_2^* \circ T_1^*$ (Theorem 2.15), and if T is nonsingular, then $(T^{-1})^* = (T^*)^{-1}$ (Corollary 4 of Theorem 2.20).

We now generalize our definition of the transpose. If $\phi \in L(U, V)$ and $T \in \mathcal{T}_r(V)$, we define the **pull-back** $\phi^* \in L(\mathcal{T}_r(V), \mathcal{T}_r(U))$ by

$$(\phi^* T)(u_1, \ldots, u_r) = T(\phi(u_1), \ldots, \phi(u_r))$$

where $u_1, \ldots, u_r \in U$. Note that in the particular case of $r = 1$, the mapping ϕ^* is just the transpose of ϕ. It should also be clear from the definition that ϕ^* is indeed a linear transformation, and hence

$$\phi^*(aT_1 + bT_2) = a\phi^* T_1 + b\phi^* T_2.$$

We also emphasize that ϕ need not be an isomorphism for us to define ϕ^*.

The main properties of the pull-back are given in the next theorem.

Theorem 8.13. *If $\phi \in L(U,V)$ and $\psi \in L(V,W)$ then*

(i) $(\psi \circ \phi)^* = \phi^* \circ \psi^*$.

(ii) If $I \in L(U)$ is the identity map, then I^* is the identity in $L(\mathcal{T}_r(U))$.

(iii) If ϕ is an isomorphism, then so is ϕ^*, and $(\phi^*)^{-1} = (\phi^{-1})^*$.

(iv) If $T_1 \in \mathcal{T}_{r_1}(V)$ and $T_2 \in \mathcal{T}_{r_2}(V)$, then

$$\phi^*(T_1 \otimes T_2) = (\phi^* T_1) \otimes (\phi^* T_2).$$

(v) Let U have basis $\{e_1, \dots, e_m\}$, V have basis $\{f_1, \dots, f_n\}$ and suppose that $\phi(e_i) = f_j a^j{}_i$. If $T \in \mathcal{T}_r(V)$ has components $T_{i_1 \cdots i_r} = T(f_{i_1}, \dots, f_{i_r})$, then the components of $\phi^* T$ relative to the basis $\{e_i\}$ are given by

$$(\phi^* T)_{j_1 \cdots j_r} = T_{i_1 \cdots i_r} a^{i_1}{}_{j_1} \cdots a^{i_r}{}_{j_r}.$$

Proof. (i) Note that $\psi \circ \phi : U \to W$, and hence $(\psi \circ \phi)^* : \mathcal{T}_r(W) \to \mathcal{T}_r(U)$. Thus for any $T \in \mathcal{T}_r(W)$ and $u_1, \dots, u_r \in U$ we have

$$\begin{aligned}((\psi \circ \phi)^* T)(u_1, \dots, u_r) &= T(\psi(\phi(u_1)), \dots, \psi(\phi(u_r))) \\ &= (\psi^* T)(\phi(u_1), \dots, \phi(u_r)) \\ &= ((\phi^* \circ \psi^*) T)(u_1, \dots, u_r).\end{aligned}$$

(ii) Obvious from the definition of I^*.

(iii) If ϕ is an isomorphism, then ϕ^{-1} exists and we have (using (i) and (ii))

$$\phi^* \circ (\phi^{-1})^* = (\phi^{-1} \circ \phi)^* = I^*.$$

Similarly $(\phi^{-1})^* \circ \phi^* = I^*$. Hence $(\phi^*)^{-1}$ exists and is equal to $(\phi^{-1})^*$.

(iv) This follows directly from the definitions (see Exercise 8.7.1).

(v) Using the definitions, we have

$$\begin{aligned}(\phi^* T)_{j_1 \cdots j_r} &= (\phi^* T)(e_{j_1}, \dots, e_{j_r}) \\ &= T(\phi(e_{j_1}), \dots, \phi(e_{j_r})) \\ &= T(f_{i_1} a^{i_1}{}_{j_1}, \dots, f_{i_r} a^{i_r}{}_{j_r}) \\ &= T(f_{i_1}, \dots, f_{i_r}) a^{i_1}{}_{j_1} \cdots a^{i_r}{}_{j_r} \\ &= T_{i_1 \cdots i_r} a^{i_1}{}_{j_1} \cdots a^{i_r}{}_{j_r}.\end{aligned}$$

Alternatively, if $\{e^i\}$and $\{f^j\}$ are the bases dual to $\{e_i\}$ and $\{f_j\}$ respectively, then $T = T_{i_1 \cdots i_r} e^{i_1} \otimes \cdots \otimes e^{i_r}$ and consequently (using the linearity of ϕ^*, part (iv) and equation (8.16)),

$$\begin{aligned}\phi^* T &= T_{i_1 \cdots i_r} \phi^* e^{i_1} \otimes \cdots \otimes \phi^* e^{i_r} \\ &= T_{i_1 \cdots i_r} a^{i_1}{}_{j_1} \cdots a^{i_r}{}_{j_r} f^{j_1} \otimes \cdots \otimes f^{j_r}\end{aligned}$$

which therefore yields the same result. ∎

For our present purposes, we will only need to consider the pull-back as defined on the space $\bigwedge^r(V)$ rather than on $\mathcal{T}_r(V)$. Therefore, if $\phi \in L(U,V)$ then $\phi^* \in L(\mathcal{T}_r(V), \mathcal{T}_r(U))$, and hence we see that for $\omega \in \bigwedge^r(V)$ we have $(\phi^*\omega)(u_1, \ldots, u_r) = \omega(\phi(u_1), \ldots, \phi(u_r))$. This shows $\phi^*(\bigwedge^r(V)) \subset \bigwedge^r(U)$. Parts (iv) and (v) of Theorem 8.13 applied to the space $\bigwedge^r(V)$ yield the following special cases. (Recall that $|i_1, \ldots, i_r|$ means the sum is over increasing indices $i_1 < \cdots < i_r$.)

Theorem 8.14. *Suppose $\phi \in L(U,V)$, $\alpha \in \bigwedge^r(V)$ and $\beta \in \bigwedge^s(V)$. Then*

(i) $\phi^*(\alpha \wedge \beta) = (\phi^*\alpha) \wedge (\phi^*\beta)$.

(ii) Let U and V have bases $\{e_i\}$ and $\{f_i\}$ respectively, and let U^* and V^* have bases $\{e^i\}$ and $\{f^i\}$. If we write $\phi(e_i) = f_j a^j{}_i$ and $\phi^*(f^i) = a^i{}_j e^j$, and if $\alpha = a_{|i_1 \cdots i_r|} f^{i_1} \wedge \cdots \wedge f^{i_r} \in \bigwedge^r(V)$, then

$$\phi^*\alpha = \hat{a}_{|k_1 \cdots k_r|} e^{k_1} \wedge \cdots \wedge e^{k_r}$$

where

$$\hat{a}_{|k_1 \cdots k_r|} = a_{|i_1 \cdots i_r|} \varepsilon^{j_1 \cdots j_r}_{k_1 \cdots k_r} a^{i_1}{}_{j_1} \cdots a^{i_r}{}_{j_r}.$$

Thus we may write

$$\hat{a}_{k_1 \cdots k_r} = a_{|i_1 \cdots i_r|} \det(a^I{}_K)$$

where

$$\det(a^I{}_K) = \begin{vmatrix} a^{i_1}{}_{k_1} & \cdots & a^{i_1}{}_{k_r} \\ \vdots & & \vdots \\ a^{i_r}{}_{k_1} & \cdots & a^{i_r}{}_{k_r} \end{vmatrix}$$

Proof. (i) For simplicity, we will write $(\phi^*\alpha)(u_J) = \alpha(\phi(u_J))$ instead of the more complete $(\phi^*\alpha)(u_1, \ldots, u_r) = \alpha(\phi(u_1), \ldots, \phi(u_r))$. Then we have

$$\begin{aligned}
[\phi^*(\alpha \wedge \beta)](u_I) &= (\alpha \wedge \beta)(\phi(u_I)) \\
&= \sum_{\overrightarrow{J}, \overrightarrow{K}} \varepsilon_I^{JK} \alpha(\phi(u_J)) \beta(\phi(u_K)) \\
&= \sum_{\overrightarrow{J}, \overrightarrow{K}} \varepsilon_I^{JK} (\phi^*\alpha)(u_J)(\phi^*\beta)(u_K) \\
&= [(\phi^*\alpha) \wedge (\phi^*\beta)](u_I).
\end{aligned}$$

By induction, this also obviously applies to the wedge product of a finite number of forms.

(ii) From $\alpha = a_{|i_1 \cdots i_r|} f^{i_1} \wedge \cdots \wedge f^{i_r}$ and $\phi^*(f^i) = a^i{}_j e^j$ we have (using part (i) and the linearity of ϕ^*)

$$\begin{aligned}
\phi^*\alpha &= a_{|i_1 \cdots i_r|} \phi^*(f^{i_1}) \wedge \cdots \wedge \phi^*(f^{i_r}) \\
&= a_{|i_1 \cdots i_r|} a^{i_1}{}_{j_1} \cdots a^{i_r}{}_{j_r} e^{j_1} \wedge \cdots \wedge e^{j_r}.
\end{aligned}$$

But
$$e^{j_1} \wedge \cdots \wedge e^{j_r} = \sum_{\underrightarrow{K}} \varepsilon^{j_1 \cdots j_r}_{k_1 \cdots k_r} e^{k_1} \wedge \cdots \wedge e^{k_r}$$
and hence we have
$$\begin{aligned} \phi^* \alpha &= a_{|i_1 \cdots i_r|} \sum_{\underrightarrow{K}} \varepsilon^{j_1 \cdots j_r}_{k_1 \cdots k_r} a^{i_1}{}_{j_1} \cdots a^{i_r}{}_{j_r} e^{k_1} \wedge \cdots \wedge e^{k_r} \\ &= \hat{a}_{|k_1 \cdots k_r|} e^{k_1} \wedge \cdots \wedge e^{k_r} \end{aligned}$$
where
$$\hat{a}_{|k_1 \cdots k_r|} = a_{|i_1 \cdots i_r|} \varepsilon^{j_1 \cdots j_r}_{k_1 \cdots k_r} a^{i_1}{}_{j_1} \cdots a^{i_r}{}_{j_r}.$$
Finally, from the definition of determinant we see that
$$\varepsilon^{j_1 \cdots j_r}_{k_1 \cdots k_r} a^{i_1}{}_{j_1} \cdots a^{i_r}{}_{j_r} = \begin{vmatrix} a^{i_1}{}_{k_1} & \cdots & a^{i_1}{}_{k_r} \\ \vdots & & \vdots \\ a^{i_r}{}_{k_1} & \cdots & a^{i_r}{}_{k_r} \end{vmatrix}$$
∎

Example 8.12. (This is a continuation of Example 8.3.) An important example of $\phi^* \alpha$ is related to the change of variables formula in multiple integrals. While we are not in any position to present this material in detail, the idea is this. Suppose we consider the spaces $U = \mathbb{R}^3(u, v, w)$ and $V = \mathbb{R}^3(x, y, z)$ where the letters in parentheses tell us the coordinate system used for that particular copy of $\mathbb{R}^3$. Note that if we write $(x, y, z) = (x^1, x^2, x^3)$ and $(u, v, w) = (u^1, u^2, u^3)$, then from elementary calculus we know that $dx^i = (\partial x^i / \partial u^j) du^j$ and $\partial / \partial u^i = (\partial x^j / \partial u^i)(\partial / \partial x^j)$.

Now recall from Example 8.3 that at each point of $\mathbb{R}^3(u, v, w)$ the tangent space has the basis $\{e_i\} = \{\partial / \partial u^i\}$ and the cotangent space has the corresponding dual basis $\{e^i\} = \{du^i\}$, with a similar result for $\mathbb{R}^3(x, y, z)$. Let us define $\phi : \mathbb{R}^3(u, v, w) \to \mathbb{R}^3(x, y, z)$ by
$$\phi\left(\frac{\partial}{\partial u^i}\right) = \left(\frac{\partial x^j}{\partial u^i}\right)\left(\frac{\partial}{\partial x^j}\right) = a^j{}_i \left(\frac{\partial}{\partial x^j}\right).$$
It is then apparent that (see equation (8.16))
$$\phi^*(dx^i) = a^i{}_j du^j = \left(\frac{\partial x^i}{\partial u^j}\right) du^j$$
as we should have expected.

We now apply this to the 3-form
$$\alpha = a_{123}\, dx^1 \wedge dx^2 \wedge dx^3 = dx \wedge dy \wedge dz \in \bigwedge^3(V).$$
Since we are dealing with a 3-form in a 3-dimensional space, we must have
$$\phi^* \alpha = \hat{a}\, du \wedge dv \wedge dw$$

where $\hat{a} = \hat{a}_{123}$ consists of the single term given by the determinant

$$\begin{vmatrix} a^1{}_1 & a^1{}_2 & a^1{}_3 \\ a^2{}_1 & a^2{}_2 & a^2{}_3 \\ a^3{}_1 & a^3{}_2 & a^3{}_3 \end{vmatrix} = \begin{vmatrix} \partial x^1/\partial u^1 & \partial x^1/\partial u^2 & \partial x^1/\partial u^3 \\ \partial x^2/\partial u^1 & \partial x^2/\partial u^2 & \partial x^2/\partial u^3 \\ \partial x^3/\partial u^1 & \partial x^3/\partial u^2 & \partial x^3/\partial u^3 \end{vmatrix}$$

which the reader may recognize as the so-called **Jacobian** of the transformation. This determinant is usually written as $\partial(x, y, z)/\partial(u, v, w)$, and hence we see that

$$\phi^*(dx \wedge dy \wedge dx) = \frac{\partial(x, y, z)}{\partial(u, v, w)} du \wedge dv \wedge dw.$$

This is precisely how volume elements transform (at least locally), and hence we have formulated the change of variables formula in quite general terms.

This formalism allows us to define the determinant of a linear transformation in an interesting abstract manner. To see this, suppose $\phi \in L(V)$ where $\dim V = n$. Since $\dim \bigwedge^n(V) = 1$, we may choose any nonzero $\omega_0 \in \bigwedge^n(V)$ as a basis. Then $\phi^* : \bigwedge^n(V) \to \bigwedge^n(V)$ is linear, and hence for any $\omega = c_0\omega_0 \in \bigwedge^n(V)$ we have

$$\phi^*\omega = \phi^*(c_0\omega_0) = c_0\phi^*\omega_0 = c_0 c\omega_0 = c(c_0\omega_0) = c\omega$$

for some scalar c (since $\phi^*\omega_0 \in \bigwedge^n(V)$ is necessarily of the form $c\omega_0$). Noting that this result did not depend on the scalar c_0 and hence is independent of $\omega = c_0\omega_0$, we see the scalar c must be unique. We therefore define the **determinant** of ϕ to be the unique scalar, denoted by $\det \phi$, such that

$$\phi^*\omega = (\det \phi)\omega.$$

It is important to realize this definition of the determinant does not depend on any choice of basis for V. However, let $\{e_i\}$ be a basis for V, and define the matrix $(a^i{}_j)$ of ϕ by $\phi(e_i) = e_j a^j{}_i$. Then for any nonzero $\omega \in \bigwedge^n(V)$ we have

$$(\phi^*\omega)(e_1, \dots, e_n) = (\det \phi)\omega(e_1, \dots, e_n).$$

On the other hand, Example 8.5 shows us that

$$\begin{aligned} (\phi^*\omega)(e_1, \dots, e_n) &= \omega(\phi(e_1), \dots, \phi(e_n)) \\ &= a^{i_1}{}_1 \cdots a^{i_n}{}_n \omega(e_{i_1}, \dots, e_{i_n}) \\ &= a^{i_1}{}_1 \cdots a^{i_n}{}_n \varepsilon_{i_1 \cdots i_n} \omega(e_1, \dots, e_n) \\ &= (\det(a^i{}_j))\omega(e_1, \dots, e_n)\,. \end{aligned}$$

Since $\omega \neq 0$, we have therefore proved the next result.

Theorem 8.15. *If V has basis $\{e_1, \dots, e_n\}$ and $\phi \in L(V)$ has the matrix representation $(a^i{}_j)$ defined by $\phi(e_i) = e_j a^j{}_i$, then* $\det \phi = \det(a^i{}_j)$.

In other words, our abstract definition of the determinant is exactly the same as our earlier classical definition. In fact, it is now easy to derive some of the properties of the determinant that were not exactly simple to prove in the more traditional manner.

Theorem 8.16. *If V is finite-dimensional and $\phi, \psi \in L(V, V)$, then*

(i) $\det(\phi \circ \psi) = (\det \phi)(\det \psi)$.

(ii) If ϕ is the identity transformation, then $\det \phi = 1$.

(iii) ϕ is an isomorphism if and only if $\det \phi \neq 0$, and if this is the case, then $\det \phi^{-1} = (\det \phi)^{-1}$.

Proof. (i) By definition we have $(\phi \circ \psi)^* \omega = \det(\phi \circ \psi)\omega$. On the other hand, by Theorem 8.13(i) we know that $(\phi \circ \psi)^* = \psi^* \circ \phi^*$, and hence

$$\begin{aligned}(\phi \circ \psi)^* \omega &= \psi^*(\phi^* \omega) = \psi^*[(\det \phi)\omega] = (\det \phi)\psi^* \omega \\ &= (\det \phi)(\det \psi)\omega.\end{aligned}$$

(ii) If $\phi = 1$ then $\phi^* = 1$ also (by Theorem 8.13(ii)), and therefore $\omega = \phi^* \omega = (\det \phi)\omega$ implies $\det \phi = 1$.

(iii) First assume ϕ is an isomorphism so that ϕ^{-1} exists. Then by parts (i) and (ii) we see that

$$1 = \det(\phi\phi^{-1}) = (\det \phi)(\det \phi^{-1})$$

which implies $\det \phi \neq 0$ and $\det \phi^{-1} = (\det \phi)^{-1}$. Conversely, suppose ϕ is not an isomorphism. Then $\operatorname{Ker} \phi \neq 0$ and there exists a nonzero $e_1 \in V$ such that $\phi(e_1) = 0$. By Theorem 1.10, we can extend this to a basis $\{e_1, \dots, e_n\}$ for V. But then for any nonzero $\omega \in \bigwedge^n(V)$ we have

$$\begin{aligned}(\det \phi)\omega(e_1, \dots, e_n) &= (\phi^* \omega)(e_1, \dots, e_n) \\ &= \omega(\phi(e_1), \dots, \phi(e_n)) \\ &= \omega(0, \phi(e_2), \dots, \phi(e_n)) \\ &= 0\end{aligned}$$

and hence we must have $\det \phi = 0$. ∎

Exercises

1. Prove Theorem 8.13(iv).

The next three exercises are related. For notational consistency, let $\alpha^i \in U^*, \beta^i \in V^*, \gamma^* \in W^*$, and let $u_i \in U, v_i \in V$ and $w_i \in W$.

2. Let $\phi \in L(U,V)$ be an isomorphism, and suppose $T \in \mathcal{T}_r^s(U)$. Define the **push-forward** $\phi_* \in L(\mathcal{T}_r^s(U), \mathcal{T}_r^s(V))$ by

$$(\phi_* T)(\beta^1, \dots, \beta^s, v_1, \dots, v_r) = T(\phi^*\beta^1, \dots, \phi^*\beta^s, \phi^{-1}v_1, \dots, \phi^{-1}v_r)$$

where $\beta^1, \dots, \beta^s \in V^*$ and $v_1, \dots, v_r \in V$. If $\psi \in L(V,W)$ is also an isomorphism, prove the following:

(a) $(\psi \circ \phi)_* = \psi_* \circ \phi_*$.
(b) If $I \in L(U)$ is the identity map, then so is $I_* \in L(\mathcal{T}_r^s(U))$.
(c) ϕ_* is an isomorphism, and $(\phi_*)^{-1} = (\phi^{-1})_*$.
(d) Since $(\phi^{-1})_*$ maps "backward," we denote it by ϕ^*. Show that in the particular case of $T \in \mathcal{T}_r(V)$, this definition of ϕ^* agrees with our previous definition of pull-back.
(e) If $T_1 \in \mathcal{T}_{r_1}^{s_1}$ and $T_2 \in \mathcal{T}_{r_2}^{s_2}$, then

$$\phi_*(T_1 \otimes T_2) = (\phi_* T_1) \otimes (\phi_* T_2).$$

3. Let $\phi \in L(U,V)$ be an isomorphism, and let U and V have bases $\{e_i\}$ and $\{f_i\}$ respectively, with corresponding dual bases $\{e^i\}$ and $\{f^i\}$. Define the matrices $(a^i{}_j)$ and $(b^i{}_j)$ by $\phi(e_i) = f_j a^j{}_i$ and $\phi^{-1}(f_i) = e_j b^j{}_i$. Suppose $T \in \mathcal{T}_r^s(U)$ has components $T^{i_1 \cdots i_s}{}_{j_1 \cdots j_r}$ relative to $\{e_i\}$, and $S \in \mathcal{T}_r^s(V)$ has components $S^{i_1 \cdots i_s}{}_{j_1 \cdots j_r}$ relative to $\{f_i\}$.

(a) Show that $(b^i{}_j) = (a^{-1})^i{}_j$.
(b) Show that the components of $\phi_* T$ are given by

$$(\phi_* T)^{i_1 \cdots i_s}{}_{j_1 \cdots j_r} = a^{i_1}{}_{k_1} \cdots a^{i_s}{}_{k_s} T^{k_1 \cdots k_s}{}_{l_1 \cdots l_r} b^{l_1}{}_{j_1} \cdots b^{l_r}{}_{j_r} \,.$$

(c) Show that the components of $\phi^* S$ are given by

$$(\phi^* S)^{i_1 \cdots i_s}{}_{j_1 \cdots j_r} = b^{i_1}{}_{k_1} \cdots b^{i_s}{}_{k_s} S^{k_1 \cdots k_s}{}_{l_1 \cdots l_r} a^{l_1}{}_{j_1} \cdots a^{l_r}{}_{j_r}$$

Hint: You will need to find the matrix representation of $(\phi^{-1})^*$.

4. (a) Let U have basis $\{e_i\}$ and dual basis $\{e^i\}$. If $\phi \in L(U)$, show that

$$\phi_* e_i = e_j a^j{}_i = \phi(e_i)$$

and

$$\phi_* e^i = b^i{}_k e^k \,.$$

(b) Let $\{e^i\}$ be the basis dual to the standard basis $\{e_i\}$ for $\mathbb{R}^2$. Let $T \in \mathcal{T}^2(\mathbb{R}^2)$ be defined by

$$T = e_1 \otimes e_1 + 2e_1 \otimes e_2 - e_2 \otimes e_1 + 3e_2 \otimes e_2$$

and let $\phi \in L(\mathbb{R}^2)$ have the matrix representation

$$A = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}.$$

Compute $\phi^* T$ using Problem 3(c).

(c) Let $T \in \mathcal{T}_1^1(\mathbb{R}^2)$ be defined by

$$T = e_1 \otimes e^2 - 2e_2 \otimes e^2$$

and let ϕ be as in part (a). Compute $\phi_* T$ using Problem 2(e).

(d) Let $\{e_i\}$ be the standard basis for $\mathbb{R}^2$ with corresponding dual basis $\{e^i\}$. Let $\{f_i\}$ be the standard basis for $\mathbb{R}^3$ with dual basis $\{f^i\}$. Let $T = -2e^1 \otimes e^2 \in \mathcal{T}_2(\mathbb{R}^2)$, and let $\phi \in L(\mathbb{R}^3, \mathbb{R}^2)$ have matrix representation

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

Compute $\phi^* T$ using Theorem 8.13(iv).

8.8 Orientations and Volumes

Suppose $\dim V = n$ and consider the space $\bigwedge^n(V)$. Since this space is 1-dimensional, we consider the n-form

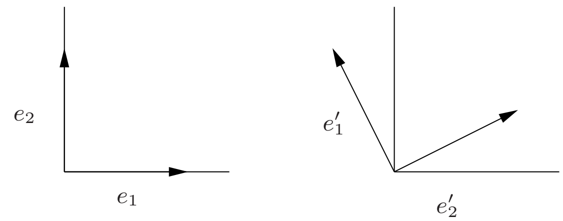
$$\omega = e^1 \wedge \cdots \wedge e^n \in \bigwedge^n(V)$$

where the basis $\{e^i\}$ for V^* is dual to the basis $\{e_i\}$ for V. If $\{v_i = e_j v^j{}_i\}$ is any set of n linearly independent vectors in V then, according to Examples 8.5 and 8.10, we have

$$\omega(v_1, \ldots, v_n) = \det(v^j{}_i)\omega(e_1, \ldots, e_n) = \det(v^j{}_i).$$

However, from the corollary to Theorem 8.9, this is just the oriented n-volume of the n-dimensional parallelepiped in $\mathbb{R}^n$ spanned by the vectors $\{v_i\}$. Therefore, we see that an n-form in some sense represents volumes in an n-dimensional space. We now proceed to make this definition precise, beginning with a careful definition of the notion of orientation on a vector space.

In order to try and make the basic idea clear, let us first consider the space $\mathbb{R}^2$ with all possible orthogonal coordinate systems. For example, we may consider the usual "right-handed" coordinate system $\{e_1, e_2\}$ shown below, or we may consider the alternative "left-handed" system $\{e'_1, e'_2\}$ also shown.



In the first case, we see that rotating e_1 into e_2 through the smallest angle between them involves a counterclockwise rotation, while in the second case, rotating e'_1 into e'_2 entails a clockwise rotation. This effect is shown in the elementary vector cross product, where the direction of $e_1 \times e_2$ is *defined* by the "right-hand rule" to point *out* of the page, while $e'_1 \times e'_2$ points *into* the page.

We now ask whether or not it is possible to continuously rotate e'_1 into e_1 and e'_2 into e_2 while maintaining a basis at all times. In other words, we ask if these two bases are in some sense equivalent. Without being rigorous, it should be clear that this can not be done because there will always be one point where the vectors e'_1 and e'_2 will be co-linear, and hence linearly dependent. This observation suggests that we consider the determinant of the matrix representing this change of basis.

To formulate this idea precisely, let us take a look at the matrix relating our two bases $\{e_i\}$ and $\{e'_i\}$ for $\mathbb{R}^2$. We thus write $e'_i = e_j a^j{}_i$ and investigate the determinant $\det(a^i{}_j)$. From the above figure, we see geometrically that

$$\begin{aligned} e'_1 &= e_1 a^1{}_1 + e_2 a^2{}_1 &\qquad& \text{where } a^1{}_1 < 0 \text{ and } a^2{}_1 > 0 \\ e'_2 &= e_1 a^1{}_2 + e_2 a^2{}_2 &\qquad& \text{where } a^1{}_2 > 0 \text{ and } a^2{}_2 > 0 \end{aligned}$$

and hence $\det(a^i{}_j) = a^1{}_1 a^2{}_2 - a^1{}_2 a^2{}_1 < 0$.

Now suppose that we view this transformation as a continuous modification of the identity transformation. This means we consider the basis vectors e'_i to be continuous functions $e'_i(t)$ of the matrix $a^j{}_i(t)$ for $0 \le t \le 1$ where $a^j{}_i(0) = \delta^j{}_i$ and $a^j{}_i(1) = a^j{}_i$, so that $e'_i(0) = e_i$ and $e'_i(1) = e'_i$. In other words, we write $e'_i(t) = e_j a^j{}_i(t)$ for $0 \le t \le 1$. Now note that $\det(a^i{}_j(0)) = \det(\delta^i{}_j) = 1 > 0$, while $\det(a^i{}_j(1)) = \det(a^i{}_j) < 0$. Therefore, since the determinant is a continuous function of its entries, there must be some value $t_0 \in (0, 1)$ where $\det(a^i{}_j(t_0)) = 0$. It then follows that the vectors $e'_i(t_0)$ will be linearly dependent.

What we have just shown is that if we start with any pair of linearly independent vectors, and then transform this pair into another pair of linearly independent vectors by moving along any continuous path of linear transformations that always maintains the linear independence of the pair, then every linear transformation along this path must have positive determinant. Another way of saying this is that if we have two bases that are related by a transformation with negative determinant, then it is impossible to continuously transform one into the other while maintaining their independence. This argument clearly applies to $\mathbb{R}^n$ and is not restricted to $\mathbb{R}^2$.

Conversely, suppose we had assumed that $e'_i = e_j a^j{}_i$, but this time with $\det(a^i{}_j) > 0$. We want to show that $\{e_i\}$ may be continuously transformed into $\{e'_i\}$ while maintaining linear independence all the way. We first assume that both $\{e_i\}$ and $\{e'_i\}$ are orthonormal bases. After treating this special case, we will show how to take care of arbitrary bases.

(Unfortunately, the argument we are about to give relies on the topological concept of path connectedness. Since this discussion is only motivation, the reader should not get too bogged down in the details of this argument. Those readers who know some topology should have no trouble filling in the necessary details if desired.)

Since $\{e_i\}$ and $\{e'_i\}$ are orthonormal, it follows from Theorem 6.6 (applied to $\mathbb{R}$ rather than $\mathbb{C}$) that the transformation matrix $A = (a^i{}_j)$ defined by $e'_i = e_j a^j{}_i$ must be orthogonal, and hence det A $= +1$ (by Theorem 5.16(i) and the fact that we are assuming $\{e_i\}$ and $\{e'_i\}$ are related by a transformation with positive determinant). Now, it is a consequence of Theorem 6.15 that there exists a nonsingular matrix S such that $S^{-1}AS = M_\theta$ where M_θ is the block diagonal form consisting of 2×2 rotation matrices $R(\theta_i)$ given by

$$R(\theta_i) = \begin{bmatrix} \cos\theta_i & -\sin\theta_i \\ \sin\theta_i & \cos\theta_i \end{bmatrix}$$

and at most a single $+1$ and/or a single -1. The fact that there can be at most a single $+1$ or -1 is due to the fact that a pair of $+1$'s can be written in the form of a rotation matrix with $\theta = 0$, and a pair of -1's can be written as a rotation matrix with $\theta = \pi$.

Since $\det R(\theta_i) = +1$ for any θ_i, we see that (using Theorem 5.10) $\det M_\theta = +1$ if there is no -1, and $\det M_\theta = -1$ if there is a single -1. From $A = SM_\theta S^{-1}$ we have $\det A = \det M_\theta$, and since we are requiring $\det A > 0$, we must have the case where there is no -1 in M_θ.

Since $\cos\theta_i$ and $\sin\theta_i$ are continuous functions of $\theta_i \in [0, 2\pi)$ (where the interval $[0, 2\pi)$ is a path connected set), we note that by parametrizing each θ_i by $\theta_i(t) = (1-t)\theta_i$, the matrix M_θ may be continuously connected to the identity matrix I (i.e., at $t = 1$). In other words, we consider the matrix $M_{\theta(t)}$ where $M_{\theta(0)} = M_\theta$ and $M_{\theta(1)} = I$. Hence every such M_θ (i.e., any matrix of the same form as our particular M_θ, but with a different set of θ_i's) may be continuously connected to the identity matrix. (For those readers who know some topology, note all we have said is that the torus $[0, 2\pi) \times \cdots \times [0, 2\pi)$ is path connected, and hence so is its continuous image which is the set of all such M_θ.)

We may write the (infinite) collection of all such M_θ as $M = \{M_\theta\}$. Clearly M is a path connected set. Since $A = SM_\theta S^{-1}$ and $I = SIS^{-1}$, we see that both A and I are contained in the collection $SMS^{-1} = \{SM_\theta S^{-1}\}$. But SMS^{-1} is also path connected since it is just the continuous image of a path connected set (matrix multiplication is obviously continuous). Thus we have shown that both

A and I lie in the path connected set SMS^{-1}, and hence A may be continuously connected to I. Note also that every transformation along this path has positive determinant since $\det SM_\theta S^{-1} = \det M_\theta = 1 > 0$ for every $M_\theta \in M$.

If we now take any path in SMS^{-1} that starts at I and goes to A, then applying this path to the basis $\{e_i\}$ we obtain a continuous transformation from $\{e_i\}$ to $\{e_i'\}$ with everywhere positive determinant. This completes the proof for the special case of orthonormal bases.

Now suppose $\{v_i\}$ and $\{v_i'\}$ are arbitrary bases related by a transformation with positive determinant. Starting with the basis $\{v_i\}$, we first apply the Gram-Schmidt process (Theorem 1.21) to $\{v_i\}$ to obtain an orthonormal basis $\{e_i\} = \{v_j b^j{}_i\}$. This orthonormalization process may be visualized as a sequence $v_i(t) = v_j b^j{}_i(t)$ (for $0 \le t \le 1$) of continuous scalings and rotations that always maintain linear independence such that $v_i(0) = v_i$ (i.e., $b^j{}_i(0) = \delta^j{}_i$) and $v_i(1) = e_i$ (i.e., $b^j{}_i(1) = b^j{}_i$). Hence we have a continuous transformation $b^j{}_i(t)$ taking $\{v_i\}$ into $\{e_i\}$ with $\det(b^j{}_i(t)) > 0$ (the transformation starts with $\det(b^j{}_i(0)) = \det I > 0$, and since the vectors are always independent, it must maintain $\det((b^j{}_i(t)) \neq 0)$.

Similarly, we may transform $\{v_i'\}$ into an orthonormal basis $\{e_i'\}$ by a continuous transformation with positive determinant. (Alternatively, it was shown in Exercise 4.4.12 that the Gram-Schmidt process is represented by an upper-triangular matrix with all positive diagonal elements, and hence its determinant is positive.) Now $\{e_i\}$ and $\{e_i'\}$ are related by an orthogonal transformation that must also have determinant equal to $+1$ because $\{v_i\}$ and $\{v_i'\}$ are related by a transformation with positive determinant, and both of the Gram-Schmidt transformations have positive determinant. This reduces the general case to the special case treated above.

With this discussion as motivation, we make the following definition. Let $\{v_1, \dots, v_n\}$ and $\{v_1', \dots, v_n'\}$ be two ordered bases for a real vector space V, and assume that $v_i' = v_j a^j{}_i$. These two bases are said to be **similarly oriented** if $\det(a^i{}_j) > 0$, and we write this as $\{v_i\} \approx \{v_i'\}$. In other words, $\{v_i\} \approx \{v_i'\}$ if $v_i' = \phi(v_i)$ with $\det \phi > 0$. We leave it to the reader to show this defines an equivalence relation on the set of all ordered bases for V (see Exercise 8.8.1). We denote the equivalence class of the basis $\{v_i\}$ by $[v_i]$.

It is worth pointing out that had we instead required $\det(a^i{}_j) < 0$, then this would not have defined an equivalence relation. This is because if $(b^i{}_j)$ is another such transformation with $\det(b^i{}_j) < 0$, then

$$\det(a^i{}_j b^j{}_k) = \det(a^i{}_j)\det(b^j{}_k) > 0.$$

Intuitively this is quite reasonable since a combination of two reflections (each of which has negative determinant) is not another reflection.

We now define an **orientation** of V to be an equivalence class of ordered bases. The space V together with an orientation $[v_i]$ is called an **oriented vector space** $(V, [v_i])$. Since the determinant of a linear transformation that relates any two bases must be either positive or negative, we see that V has exactly two orientations. In particular, if $\{v_i\}$ is any given basis, then every other basis belonging to the equivalence class $[v_i]$ of $\{v_i\}$ will be related to $\{v_i\}$ by a transformation with positive determinant, while those bases related to $\{v_i\}$ by a transformation with negative determinant will be related to each other by a transformation with positive determinant (see Exercise 8.8.1).

Now recall we have seen that n-forms seem to be related to n-volumes in an n-dimensional space V. To precisely define this relationship, we formulate orientations in terms of n-forms.

To begin with, the nonzero elements of the 1-dimensional space $\bigwedge^n(V)$ are called **volume forms** (or sometimes **volume elements**) on V. If ω_1 and ω_2 are volume forms, then ω_1 is said to be **equivalent** to ω_2 if $\omega_1 = c\,\omega_2$ for some real $c > 0$, and in this case we also write $\omega_1 \approx \omega_2$. Since every element of $\bigwedge^n(V)$ is related to every other element by a relationship of the form $\omega_1 = a\omega_2$ for some real a (i.e., $-\infty < a < \infty$), it is clear that this equivalence relation divides the set of all nonzero volume forms into two distinct groups (i.e., equivalence classes). We can relate any ordered basis $\{v_i\}$ for V to a specific volume form by defining

$$\omega = v^1 \wedge \cdots \wedge v^n$$

where $\{v^i\}$ is the basis dual to $\{v_i\}$. That this association is meaningful is shown in the next result.

Theorem 8.17. *Let $\{v_i\}$ and $\{\bar{v}_i\}$ be bases for V, and let $\{v^i\}$ and $\{\bar{v}^i\}$ be the corresponding dual bases. Define the volume forms*

$$\omega = v^1 \wedge \cdots \wedge v^n$$

and

$$\bar{\omega} = \bar{v}^1 \wedge \cdots \wedge \bar{v}^n.$$

Then $\{v_i\} \approx \{\bar{v}_i\}$ if and only if $\omega \approx \bar{\omega}$.

Proof. First suppose $\{v_i\} \approx \{\bar{v}_i\}$. Then $\bar{v}_i = \phi(v_i)$ where $\det\phi > 0$, and hence using the result

$$\omega(v_1, \ldots, v_n) = v^1 \wedge \cdots \wedge v^n(v_1, \ldots, v_n) = 1$$

shown in Example 8.10, we have

$$\begin{aligned}
\omega(\bar{v}_1, \ldots, \bar{v}_n) &= \omega(\phi(v_1), \ldots, \phi(v_n)) \\
&= (\phi^*\omega)(v_1, \ldots, v_n) \\
&= (\det\phi)\omega(v_1, \ldots, v_n) \\
&= \det\phi
\end{aligned}$$

If we assume that $\omega = c\,\bar{\omega}$ for some $-\infty < c < \infty$, then using $\bar{\omega}(\bar{v}_1, \ldots, \bar{v}_n) = 1$ we see that our result implies $c = \det\phi > 0$ and therefore $\omega \approx \bar{\omega}$.

Conversely, if $\omega = c\,\bar{\omega}$ where $c > 0$, then assuming $\bar{v}_i = \phi(v_i)$, the above calculation shows that $\det\phi = c > 0$, and hence $\{v_i\} \approx \{\bar{v}_i\}$. ∎

What this theorem shows us is that an equivalence class of bases uniquely determines an equivalence class of volume forms and conversely. Therefore it is consistent with our earlier definitions to say that an equivalence class $[\omega]$ of volume forms on V defines an **orientation** on V, and the space V together with an orientation $[\omega]$ is called an **oriented vector space** $(V, [\omega])$. A basis $\{v_i\}$ for $(V, [\omega])$ is now said to be **positively oriented** if $\omega(v_1, \dots, v_n) > 0$. Not surprisingly, the equivalence class $[-\omega]$ is called the **reverse orientation**, and the basis $\{v_i\}$ is said to be **negatively oriented** if $\omega(v_1, \dots, v_n) < 0$.

Note that if the ordered basis $\{v_1, v_2, \dots, v_n\}$ is negatively oriented, then the basis $\{v_2, v_1, \dots, v_n\}$ will be positively oriented because $\omega(v_2, v_1, \dots, v_n) = -\omega(v_1, v_2, \dots, v_n) > 0$. By way of additional terminology, the **standard orientation** on $\mathbb{R}^n$ is that orientation defined by either the standard ordered basis $\{e_1, \dots, e_n\}$, or the corresponding volume form $e^1 \wedge \cdots \wedge e^n$.

In order to proceed any further, we must introduce the notion of a metric on V. This is the subject of the next section.

Exercises

1. (a) Show the collection of all similarly oriented bases for V defines an equivalence relation on the set of all ordered bases for V.
   (b) Let $\{v_i\}$ be a basis for V. Show that all other bases related to $\{v_i\}$ by a transformation with negative determinant will be related to each other by a transformation with positive determinant.

2. Let (U, ω) and (V, μ) be oriented vector spaces with chosen volume elements. We say that $\phi \in L(U, V)$ is **volume preserving** if $\phi^*\mu = \omega$. If $\dim U = \dim V$ is finite, show ϕ is an isomorphism.

3. Let $(U, [\omega])$ and $(V, [\mu])$ be oriented vector spaces. We say that $\phi \in L(U, V)$ is **orientation preserving** if $\phi^*\mu \in [\omega]$. If $\dim U = \dim V$ is finite, show ϕ is an isomorphism. If $U = V = \mathbb{R}^3$, give an example of a linear transformation that is orientation preserving but not volume preserving.

8.9 The Metric Tensor and Volume Forms

We now generalize slightly our definition of inner products on V. In particular, recall from Section 1.5 (and Section 8.1) that property (IP3) of an inner product requires that $\langle u, u\rangle \geq 0$ for all $u \in V$ and $\langle u, u\rangle = 0$ if and only if $u = 0$. If we drop this condition entirely, then we obtain an **indefinite** inner product on V. (In fact, some authors define an inner product as obeying only (IP1) and (IP2), and then refer to what we have called an inner product as a "positive definite inner product.") If we replace (IP3) by the weaker requirement

(IP3′) $\langle u, v\rangle = 0$ for all $v \in V$ if and only if $u = 0$

then our inner product is said to be **nondegenerate**. (Note that every example of an inner product given in this book up to now has been nondegenerate.)

Thus a real nondegenerate indefinite inner product is just a real nondegenerate symmetric bilinear map. We will soon see an example of an inner product with the property that $\langle u, u\rangle = 0$ for some $u \neq 0$ (see Example 8.16 below).

Throughout the remainder of this chapter, we will assume that our inner products are indefinite and nondegenerate unless otherwise noted. We furthermore assume that we are dealing exclusively with real vector spaces.

Let $\{e_i\}$ be a basis for an inner product space V. Since in general we will not have $\langle e_i, e_j\rangle = \delta_{ij}$, we define the scalars g_{ij} by

$$g_{ij} = \langle e_i, e_j\rangle.$$

In terms of the g_{ij}, we have for any $X, Y \in V$

$$\langle X, Y\rangle = \langle x^i e_i, y^j e_j\rangle = x^i y^j \langle e_i, e_j\rangle = g_{ij} x^i y^j.$$

If $\{\bar{e}_i\}$ is another basis for V, then we will have $\bar{e}_i = e_j a^j{}_i$ for some nonsingular transition matrix $A = (a^j{}_i)$. Hence, writing $\bar{g}_{ij} = \langle \bar{e}_i, \bar{e}_j\rangle$ we see that

$$\bar{g}_{ij} = \langle \bar{e}_i, \bar{e}_j\rangle = \langle e_r a^r{}_i, e_s a^s{}_j\rangle = a^r{}_i a^s{}_j \langle e_r, e_s\rangle = a^r{}_i a^s{}_j g_{rs}$$

which shows that the g_{ij} transform like the components of a second-rank covariant tensor. Indeed, defining the tensor $g \in \mathcal{T}_2(V)$ by

$$g(X, Y) = \langle X, Y\rangle$$

results in

$$g(e_i, e_j) = \langle e_i, e_j\rangle = g_{ij}$$

as it should.

We are therefore justified in defining the (**covariant**) **metric tensor**

$$g = g_{ij}\omega^i \otimes \omega^j \in \mathcal{T}_2(V)$$

(where $\{\omega^i\}$ is the basis dual to $\{e_i\}$) by $g(X, Y) = \langle X, Y\rangle$. In fact, since the inner product is nondegenerate and symmetric (i.e., $\langle X, Y\rangle = \langle Y, X\rangle$), we see that g is a nondegenerate symmetric tensor (i.e., $g_{ij} = g_{ji}$).

Next, we notice that given any vector $A \in V$, we may define a linear functional $\langle A, \cdot\rangle$ on V by the assignment $B \mapsto \langle A, B\rangle$. In other words, for any $A \in V$, we associate the 1-form α defined by $\alpha(B) = \langle A, B\rangle$ for every $B \in V$. Note the kernel of the mapping $A \mapsto \langle A, \cdot\rangle$ (which is easily seen to be a vector space homomorphism) consists of only the zero vector (since $\langle A, B\rangle = 0$ for every $B \in V$ implies $A = 0$), and hence this association is an isomorphism.

Given any basis $\{e_i\}$ for V, the components a_i of $\alpha \in V^*$ are given in terms of those of $A = a^i e_i \in V$ by

$$a_i = \alpha(e_i) = \langle A, e_i\rangle = \langle a^j e_j, e_i\rangle = a^j \langle e_j, e_i\rangle = a^j g_{ji}.$$

Thus, to any contravariant vector $A = a^i e_i \in V$, we can associate a unique covariant vector $\alpha \in V^*$ by

$$\alpha = a_i \omega^i = (a^j g_{ji})\omega^i$$

where $\{\omega^i\}$ is the basis for V^* dual to the basis $\{e_i\}$ for V. In other words, we write

$$a_i = a^j g_{ji}$$

and we say that a_i arises by **lowering the index** j of a^j.

Example 8.13. If we consider the space $\mathbb{R}^n$ with a Cartesian coordinate system $\{e_i\}$, then we have $g_{ij} = \langle e_i, e_j\rangle = \delta_{ij}$, and hence $a_i = \delta_{ij}a^j = a^i$. Therefore, *in a Cartesian coordinate system*, there is no distinction between the components of covariant and contravariant vectors. This explains why 1-forms never arise in elementary treatments of vector analysis.

Since the metric tensor is nondegenerate, the matrix (g_{ij}) must be nonsingular (or else the mapping $a^j \mapsto a_i$ would not be an isomorphism). We can therefore define the inverse matrix (g^{ij}) by

$$g^{ij}g_{jk} = g_{kj}g^{ji} = \delta^i{}_k.$$

Using (g^{ij}), we see the inverse of the mapping $a^j \mapsto a_i$ is given by

$$g^{ij}a_j = a^i.$$

This is called, naturally enough, **raising an index**. We will show below that the g^{ij} do indeed form the components of a tensor.

It is worth remarking that the "tensor" $g^i{}_j = g^{ik}g_{kj} = \delta^i{}_j (= \delta_j{}^i)$ is unique in that it has the same components in any coordinate system. Indeed, if $\{e_i\}$ and $\{\bar{e}_i\}$ are two bases for a space V with corresponding dual bases $\{\omega^i\}$ and $\{\bar{\omega}^i\}$, then $\bar{e}_i = e_j a^j{}_i$ and $\bar{\omega}^j = b^j{}_i\omega^i = (a^{-1})^j{}_i\omega^i$ (see the discussion following Theorem 8.5). Therefore, if we define the tensor δ to have the same values in the first coordinate system as the Kronecker delta, then $\delta^i{}_j = \delta(\omega^i, e_j)$. If we now define the symbol $\bar{\delta}^i{}_j$ by $\bar{\delta}^i{}_j = \delta(\bar{\omega}^i, \bar{e}_j)$, then we see that

$$\begin{aligned}
\bar{\delta}^i{}_j &= \delta(\bar{\omega}^i, \bar{e}_j) = \delta((a^{-1})^i{}_k\omega^k, e_r a^r{}_j) = (a^{-1})^i{}_k a^r{}_j \delta(\omega^k, e_r) \\
&= (a^{-1})^i{}_k a^r{}_j \delta^k{}_r = (a^{-1})^i{}_k a^k{}_j = \delta^i{}_j.
\end{aligned}$$

This shows the $\delta^i{}_j$ are in fact the components of a tensor, and these components are the same in any coordinate system.

We would now like to show that the scalars g^{ij} are indeed the components of a tensor. There are several ways this can be done. First, let us write $g_{ij}g^{jk} = \delta^k{}_i$ where we know that both g_{ij} and $\delta^k{}_i$ are tensors. Multiplying both sides of this equation by $(a^{-1})^r{}_k a^i{}_s$ and using $(a^{-1})^r{}_k a^i{}_s \delta^k{}_i = \delta^r{}_s$ we find

$$g_{ij}g^{jk}(a^{-1})^r{}_k a^i{}_s = \delta^r{}_s.$$

Now substitute $g_{ij} = g_{it}\delta^t{}_j = g_{it}a^t{}_q(a^{-1})^q{}_j$ to obtain

$$[a^i{}_s a^t{}_q g_{it}][(a^{-1})^q{}_j (a^{-1})^r{}_k g^{jk}] = \delta^r{}_s.$$

Since g_{it} is a tensor, we know that $a^i{}_s a^t{}_q g_{it} = \bar{g}_{sq}$. If we write

$$\bar{g}^{qr} = (a^{-1})^q{}_j (a^{-1})^r{}_k g^{jk}$$

then we will have defined the g^{jk} to transform as the components of a tensor, and furthermore, they have the requisite property that $\bar{g}_{sq}\bar{g}^{qr} = \delta^r{}_s$. Therefore we have defined the (**contravariant**) **metric tensor** $G \in \mathcal{T}_0^2(V)$ by

$$G = g^{ij} e_i \otimes e_j$$

where $g^{ij} g_{jk} = \delta^i{}_k$.

There is another interesting way for us to define the tensor G. We have already seen that a vector $A = a^i e_i \in V$ defines a unique linear form $\alpha = a_j \omega^j \in V^*$ by the association $\alpha = g_{ij} a^i \omega^j$. If we denote the inverse of the matrix (g_{ij}) by (g^{ij}) so that $g^{ij} g_{jk} = \delta^i{}_k$, then to any linear form $\alpha = a_i \omega^i \in V^*$ there corresponds a unique vector $A = a^i e_i \in V$ defined by $A = g^{ij} a_i e_j$. We can now use this isomorphism to define an inner product on V^*. In other words, if $\langle \cdot\,, \cdot \rangle$ is an inner product on V, we define an inner product $\langle \cdot\,, \cdot \rangle$ on V^* by

$$\langle \alpha, \beta \rangle = \langle A, B \rangle$$

where A, $B \in V$ are the vectors corresponding to the 1-forms α, $\beta \in V^*$.

Let us write an arbitrary basis vector e_i in terms of its components relative to the basis $\{e_i\}$ as $e_i = \delta^j{}_i e_j$. Therefore, in the above isomorphism, we may define the linear form $\hat{e}_i \in V^*$ corresponding to the basis vector e_i by

$$\hat{e}_i = g_{jk} \delta^j{}_i \omega^k = g_{ik} \omega^k$$

and hence using the inverse matrix we find

$$\omega^k = g^{ki} \hat{e}_i.$$

Applying our definition of the inner product in V^* we have $\langle \hat{e}_i, \hat{e}_j \rangle = \langle e_i, e_j \rangle = g_{ij}$, and therefore we obtain

$$\langle \omega^i, \omega^j \rangle = \langle g^{ir} \hat{e}_r, g^{js} \hat{e}_s \rangle = g^{ir} g^{js} \langle \hat{e}_r, \hat{e}_s \rangle = g^{ir} g^{js} g_{rs} = g^{ir} \delta^j{}_r = g^{ij}$$

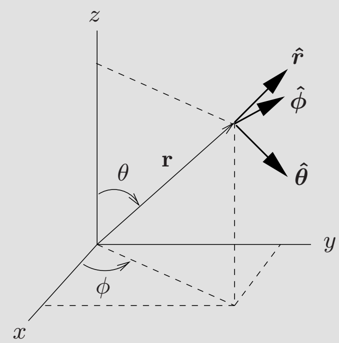
which is the analogue in V^* of the definition $g_{ij} = \langle e_i, e_j \rangle$ in V.

Lastly, since $\bar{\omega}^j = (a^{-1})^j{}_i \omega^i$, we see that

$$\begin{aligned} \bar{g}^{ij} &= \langle \bar{\omega}^i, \bar{\omega}^j \rangle = \langle (a^{-1})^i{}_r \omega^r, (a^{-1})^j{}_s \omega^s \rangle = (a^{-1})^i{}_r (a^{-1})^j{}_s \langle \omega^r, \omega^s \rangle \\ &= (a^{-1})^i{}_r (a^{-1})^j{}_s g^{rs} \end{aligned}$$

so the scalars g^{ij} may be considered to be the components of a symmetric tensor $G \in \mathcal{T}_2^0(V)$ defined as above by $G = g^{ij} e_i \otimes e_j$.

Example 8.14. Let us compute both (g_{ij}) and (g^{ij}) for the usual spherical coordinate system (r, θ, ϕ) as shown below.



Consider an infinitesimal displacement $d\mathbf{x}$. To write this in terms of the three unit vectors, first suppose that θ and ϕ are held constant and we vary r. Then the displacement is just $dr\,\hat{\boldsymbol{r}}$. Now hold r and ϕ constant and vary θ. This gives a displacement $rd\theta\,\hat{\boldsymbol{\theta}}$. Finally, holding r and θ constant and varying ϕ we have the displacement $r\sin\theta d\phi\,\hat{\boldsymbol{\phi}}$. Putting this together, we see that a general displacement $d\mathbf{x}$ is given by

$$d\mathbf{x} = dr\,\hat{\boldsymbol{r}} + rd\theta\,\hat{\boldsymbol{\theta}} + r\sin\theta\,d\phi\,\hat{\boldsymbol{\phi}}$$

where $\hat{\boldsymbol{r}}, \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\phi}}$ are orthonormal unit vectors.

The square of the magnitude of this displacement is $\|d\mathbf{x}\|^2$ and is conventionally denoted by ds^2. Writing this out we have

$$ds^2 = \langle d\mathbf{x}, d\mathbf{x}\rangle = dr^2 + r^2 d\theta^2 + r^2\sin^2\theta\,d\phi^2.$$

If we write $d\mathbf{x} = dx^i e_i$ (where $dx^1 = dr, dx^2 = d\theta$ and $dx^3 = d\phi$ so that $e_1 = \hat{\boldsymbol{r}}, e_2 = r\hat{\boldsymbol{\theta}}$ and $e_3 = r\sin\theta\hat{\boldsymbol{\phi}}$), then

$$\begin{aligned} ds^2 &= \langle dx^i e_i, dx^j e_j\rangle = dx^i dx^j \langle e_i, e_j\rangle = g_{ij}dx^i dx^j \\ &= g_{rr}dr^2 + g_{\theta\theta}d\theta^2 + g_{\phi\phi}d\phi^2. \end{aligned}$$

where $g_{11} := g_{rr}, g_{22} := g_{\theta\theta}$ and $g_{33} := g_{\phi\phi}$.

Comparing these last two equations shows that $g_{rr} = 1, g_{\theta\theta} = r^2$ and $g_{\phi\phi} = r^2\sin^2\theta$ which we can write in (diagonal) matrix form as

$$(g_{ij}) = \begin{bmatrix} 1 & & \\ & r^2 & \\ & & r^2\sin^2\theta \end{bmatrix}.$$

Since this is diagonal, the inverse matrix is clearly given by

$$(g^{ij}) = \begin{bmatrix} 1 & & \\ & 1/r^2 & \\ & & 1/r^2\sin^2\theta \end{bmatrix}$$

so that the nonzero elements are given by $g^{ij} = 1/g_{ij}$.

Example 8.15. We now show how the gradient of a function depends on the metric for its definition. In particular, we will see that the gradient is the contravariant version of the usual differential. This is exactly the same as the association between vectors and 1-forms that we made earlier when we discussed raising and lowering indices.

Referring again to Example 8.3, let $\{\partial/\partial r, \partial/\partial\theta, \partial/\partial\phi\}$ be a basis for our space $\mathbb{R}^3$ with the standard spherical coordinate system. This is an example of what is called a **coordinate basis**, i.e., a basis of the form $\{\partial_i\} = \{\partial/\partial x^i\}$ where $\{x^i\}$ is a (local) coordinate system.

Recalling that the differential of a function is defined by its action on a vector (i.e., $df(v) = v(f) = v^i \partial f/\partial x^i$), we see that the corresponding dual basis is $\{dx^i\} = \{dr, d\theta, d\phi\}$ (so that $dx^i(\partial/\partial x^j) = \partial x^i/\partial x^j = \delta^i_j$). Since $dx^i(v) = v(x^i) = v^j \partial_j x^i = v^i$ we have $df(v) = (\partial f/\partial x^i) dx^i(v)$ and therefore $df = (\partial f/\partial x^i) dx^i$. For our spherical coordinates this becomes

$$df = \frac{\partial f}{\partial r} dr + \frac{\partial f}{\partial \theta} d\theta + \frac{\partial f}{\partial \phi} d\phi. \tag{8.17}$$

We now *define* the gradient of a function f by

$$df(v) = \langle \nabla f, v \rangle. \tag{8.18}$$

If we write $\nabla f = (\nabla f)^i \partial_i$ then $\langle \nabla f, v \rangle = (\nabla f)^i v^j \langle \partial_i, \partial_j \rangle = (\nabla f)^i v^j g_{ij}$ which by definition is equal to $df(v) = (\partial f/\partial x^j) v^j$. Since v was arbitrary, equating these shows that $(\nabla f)^i g_{ij} = \partial f/\partial x^i$. Using the inverse metric, this may be written

$$(\nabla f)^i = g^{ij} \frac{\partial f}{\partial x^j} = g^{ij} \partial_j f.$$

Since (g^{ij}) is diagonal, it is now easy to use the previous example to write the gradient in spherical coordinates as

$$\nabla f = (g^{rr} \partial_r f) \partial_r + (g^{\theta\theta} \partial_\theta f) \partial_\theta + (g^{\phi\phi} \partial_\phi f) \partial_\phi$$

or

$$\nabla f = \frac{\partial f}{\partial r} \partial_r + \frac{1}{r^2} \frac{\partial f}{\partial \theta} \partial_\theta + \frac{1}{r^2 \sin^2\theta} \frac{\partial f}{\partial \phi} \partial_\phi \tag{8.19}$$

Note however, that the ∂_i are not unit vectors in general because $\langle \partial_i, \partial_j \rangle = g_{ij} \neq \delta_{ij}$. Therefore we define a **non-coordinate** *orthonormal* basis $\{\bar{e}_i\}$ by

$$\bar{e}_i = \frac{\partial_i}{\|\partial_i\|}.$$

But $\|\partial_i\| = \langle \partial_i, \partial_i \rangle^{1/2} = \sqrt{g_{ii}} := h_i$ and we have

$$\bar{e}_i = \frac{1}{h_i}\partial_i.$$

In particular, from the previous example these orthonormal basis vectors become

$$\bar{e}_r = \frac{\partial}{\partial r} \qquad \bar{e}_\theta = \frac{1}{r}\frac{\partial}{\partial \theta} \qquad \bar{e}_\phi = \frac{1}{r\sin\theta}\frac{\partial}{\partial \phi}.$$

Actually, these are just what are referred to as $\hat{\boldsymbol{r}}, \hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\phi}}$ in elementary courses. To see this, note that a point $\mathbf{x}$ has coordinates given by

$$\mathbf{x} = (r\sin\theta\cos\phi, r\sin\theta\sin\phi, r\cos\theta)$$

so that

$$\begin{aligned}
\frac{\partial \mathbf{x}}{\partial r} &= (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta) \\
\frac{\partial \mathbf{x}}{\partial \theta} &= (r\cos\theta\cos\phi, r\cos\theta\sin\phi, -r\sin\theta) \\
\frac{\partial \mathbf{x}}{\partial \phi} &= (-r\sin\theta\sin\phi, r\sin\theta\cos\phi, 0).
\end{aligned}$$

Therefore

$$\begin{aligned}
\hat{\boldsymbol{r}} &= \frac{\partial \mathbf{x}/\partial r}{\|\partial \mathbf{x}/\partial r\|} = \frac{\partial \mathbf{x}}{\partial r} \\
\hat{\boldsymbol{\theta}} &= \frac{\partial \mathbf{x}/\partial \theta}{\|\partial \mathbf{x}/\partial \theta\|} = \frac{1}{r}\frac{\partial \mathbf{x}}{\partial \theta} \\
\hat{\boldsymbol{\phi}} &= \frac{\partial \mathbf{x}/\partial \phi}{\|\partial \mathbf{x}/\partial \phi\|} = \frac{1}{r\sin\theta}\frac{\partial \mathbf{x}}{\partial \phi}
\end{aligned}$$

and what we have written as $\partial/\partial x^i$ is classically written as $\partial \mathbf{x}/\partial x^i$.

Anyway, writing an arbitrary vector v in terms of both bases we have

$$v = v^i \partial_i = \bar{v}^i \bar{e}_i = \bar{v}^i \frac{1}{h_i}\partial_i$$

and hence (with obviously no sum)

$$v^i = \frac{\bar{v}^i}{h_i} \qquad \text{or} \qquad \bar{v}^i = h_i v^i. \tag{8.20}$$

The corresponding dual basis is defined by $\omega^i(\bar{e}_j) = \delta^i_j = (1/h_j)\omega^i(\partial_j)$ and comparing this with $dx^i(\partial_j) = \delta^i_j$ we see that $\omega^i = h_i dx^i$ (no sum). This yields

$$\omega^r = dr \qquad \omega^\theta = r\,d\theta \qquad \omega^\phi = r\sin\theta\,d\phi$$

and we can now use these to write df in terms of the non-coordinate basis as

$$df = \frac{\partial f}{\partial r}\omega^r + \frac{1}{r}\frac{\partial f}{\partial\theta}\omega^\theta + \frac{1}{r\sin\theta}\frac{\partial f}{\partial\phi}\omega^\phi. \tag{8.21}$$

Since the gradient is just a vector, to find its components in terms of the orthonormal basis we use equation (8.20) to write $(\overline{\nabla f})^i = h_i(\nabla f)^i$ so that

$$\nabla f = \frac{\partial f}{\partial r}\bar{e}_r + \frac{1}{r}\frac{\partial f}{\partial\theta}\bar{e}_\theta + \frac{1}{r\sin\theta}\frac{\partial f}{\partial\phi}\bar{e}_\phi \tag{8.22}$$

or

$$\nabla f = \frac{\partial f}{\partial r}\hat{\boldsymbol{r}} + \frac{1}{r}\frac{\partial f}{\partial\theta}\hat{\boldsymbol{\theta}} + \frac{1}{r\sin\theta}\frac{\partial f}{\partial\phi}\hat{\boldsymbol{\phi}}.$$

Equations (8.19) and (8.22) express the gradient in terms of coordinate and (orthonormal) non-coordinate bases respectively, while equations (8.17) and (8.21) express the differential of a function in terms of these bases.

Now let $g = \langle\cdot\,,\cdot\rangle$ be an arbitrary (i.e., possibly degenerate) real symmetric bilinear form on the inner product space V. It follows from Sylvester's theorem (the corollary to Theorem 7.6) that there exists a basis $\{e_i\}$ for V in which the matrix (g_{ij}) of g takes the unique diagonal form

$$g_{ij} = \begin{bmatrix} I_r & & \\ & -I_s & \\ & & 0_t \end{bmatrix}$$

where $r + s + t = \dim V = n$. Thus

$$g(e_i, e_i) = \begin{cases} 1 & \text{for } 1 \le i \le r \\ -1 & \text{for } r+1 \le i \le r+s \\ 0 & \text{for } r+s+1 \le i \le n \end{cases}.$$

If $r + s < n$, the inner product is degenerate and we say the space V is **singular** (with respect to the given inner product). If $r + s = n$, then the inner product is nondegenerate, and the basis $\{e_i\}$ is orthonormal. In the orthonormal case, if either $r = 0$ or $r = n$, the space is said to be **ordinary Euclidean**, and if $0 < r < n$, then the space is called **pseudo-Euclidean**. Recall that the number $r - s = r - (n - r) = 2r - n$ is called the **signature** of g (which is therefore just the trace of (g_{ij})). Moreover, the number of -1's is called the **index** of g and is denoted by $\operatorname{Ind}(g)$. If $g = \langle\cdot\,,\cdot\rangle$ is to be a metric on V, then by definition we must have $r + s = n$ so that the inner product is nondegenerate. In this case, the basis $\{e_i\}$ is called **g-orthonormal**.

Example 8.16. If the metric g represents a positive definite inner product on V, then we must have $\operatorname{Ind}(g) = 0$, and such a metric is said to be **Riemannian**. Alternatively, another well-known metric is the Lorentz metric used in the theory of special relativity. By definition, a **Lorentz metric** η has $\operatorname{Ind}(\eta) = 1$. Therefore, if η is a Lorentz metric, an η-orthonormal basis $\{e_1, \dots, e_n\}$ ordered in such a way that $\eta(e_i, e_i) = +1$ for $i = 1, \dots, n-1$ and $\eta(e_n, e_n) = -1$ is called a **Lorentz frame**.

Thus, in terms of a g-orthonormal basis, a Riemannian metric has the form

$$(g_{ij}) = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$

while in a Lorentz frame, a Lorentz metric takes the form

$$(\eta_{ij}) = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & -1 \end{bmatrix}.$$

It is worth remarking that a Lorentz metric is also frequently defined as having $\operatorname{Ind}(\eta) = n - 1$. In this case we have $\eta(e_1, e_1) = 1$ and $\eta(e_i, e_i) = -1$ for each $i = 2, \dots, n$. We also point out that a vector $v \in V$ is called **timelike** if $\eta(v, v) < 0$, **lightlike** (or **null**) if $\eta(v, v) = 0$, and **spacelike** if $\eta(v, v) > 0$. Note that a Lorentz inner product is clearly indefinite since, for example, the nonzero vector v with components $v = (0, 0, 1, 1)$ has the property that $\langle v, v \rangle = \eta(v, v) = 0$.

We now show that introducing a metric on V leads to a unique volume form on V.

Theorem 8.18. *Let g be a metric on an n-dimensional oriented vector space $(V, [\omega])$. Then corresponding to the metric g there exists a unique volume form $\mu = \mu(g) \in [\omega]$ such that $\mu(e_1, \dots, e_n) = 1$ for every positively oriented g-orthonormal basis $\{e_i\}$ for V. Moreover, if $\{v_i\}$ is any (not necessarily g-orthonormal) positively oriented basis for V with dual basis $\{v^i\}$, then*

$$\mu = |\det(g(v_i, v_j))|^{1/2}\, v^1 \wedge \cdots \wedge v^n.$$

In particular, if $\{v_i\} = \{e_i\}$ is a g-orthonormal basis, then $\mu = e^1 \wedge \cdots \wedge e^n$.

Proof. Since $\omega \neq 0$, there exists a positively oriented g-orthonormal basis $\{e_i\}$ such that $\omega(e_1, \ldots, e_n) > 0$ (we can multiply e_1 by -1 if necessary in order that $\{e_i\}$ be positively oriented). We now define $\mu \in [\omega]$ by

$$\mu(e_1, \ldots, e_n) = 1.$$

That this defines a unique μ follows by multilinearity. We claim that if $\{f_i\}$ is any other positively oriented g-orthonormal basis, then $\mu(f_1, \ldots, f_n) = 1$ also. To show this, we first prove a simple general result.

Suppose $\{v_i\}$ is any other basis for V related to the g-orthonormal basis $\{e_i\}$ by $v_i = \phi(e_i) = e_j a^j{}_i$ where, by Theorem 8.15, we have $\det \phi = \det(a^i{}_j)$. We then have $g(v_i, v_j) = a^r{}_i a^s{}_j g(e_r, e_s)$ which in matrix notation is $[g]_v = A^T [g]_e A$, and hence

$$\det(g(v_i, v_j)) = (\det \phi)^2 \det(g(e_r, e_s)). \tag{8.23}$$

However, since $\{e_i\}$ is g-orthonormal we have $g(e_r, e_s) = \pm\delta_{rs}$, and therefore $|\det(g(e_r, e_s))| = 1$. In other words

$$|\det(g(v_i, v_j))|^{1/2} = |\det \phi| \tag{8.24}$$

Returning to our problem, we have $\det(g(f_i, f_j)) = \pm 1$ also since $\{f_i\} = \{\phi(e_i)\}$ is g-orthonormal. Thus equation (8.24) implies that $\det \phi = 1$. But $\{f_i\}$ is positively oriented so that $\mu(f_1, \ldots, f_n) > 0$ by definition. Therefore

$$\begin{aligned} 0 < \mu(f_1, \ldots, f_n) &= \mu(\phi(e_1), \ldots, \phi(e_n)) = (\phi^*\mu)(e_1, \ldots, e_n) \\ &= (\det \phi)\mu(e_1, \ldots, e_n) = \det \phi \end{aligned}$$

so that we must in fact have $\det \phi = +1$. In other words, $\mu(f_1, \ldots, f_n) = 1$ as claimed.

Now suppose $\{v_i\}$ is an arbitrary positively oriented basis for V such that $v_i = \phi(e_i)$. Then, analogously to what we have just shown, we see that $\mu(v_1, \ldots, v_n) = \det \phi > 0$. Hence equation (8.24) shows that (using Example 8.10)

$$\begin{aligned} \mu(v_1, \ldots, v_n) &= \det \phi \\ &= |\det(g(v_i, v_j))|^{1/2} \\ &= |\det(g(v_i, v_j))|^{1/2}\, v^1 \wedge \cdots \wedge v^n(v_1, \ldots, v_n) \end{aligned}$$

which implies

$$\mu = |\det(g(v_i, v_j))|^{1/2}\, v^1 \wedge \cdots \wedge v^n. \qquad \blacksquare$$

The unique volume form μ defined in Theorem 8.18 is called the g-**volume**, or sometimes the **metric volume form**. A common (although rather careless) notation is to write $\det(g(v_i, v_j))^{1/2} = \sqrt{|g|}$. In this notation, the g-volume is written as

$$\sqrt{|g|} v^1 \wedge \cdots \wedge v^n$$

where $\{v_1, \ldots, v_n\}$ must be positively oriented.

If the basis $\{v_1, v_2, \dots, v_n\}$ is negatively oriented, then clearly $\{v_2, v_1, \dots, v_n\}$ will be positively oriented. Furthermore, even though the matrix of g relative to each of these oriented bases will be different, the determinant actually remains unchanged (see the discussion following the corollary to Theorem 8.9). Therefore, for this negatively oriented basis, the g-volume is

$$\sqrt{|g|}v^2 \wedge v^1 \wedge \cdots \wedge v^n = -\sqrt{|g|}v^1 \wedge v^2 \wedge \cdots \wedge v^n.$$

We thus have the following corollary to Theorem 8.18.

Corollary. *Let $\{v_i\}$ be any basis for the n-dimensional oriented vector space $(V, [\omega])$ with metric g. Then the g-volume form on V is given by*

$$\pm\sqrt{|g|}v^1 \wedge \cdots \wedge v^n$$

where the "+" sign is for $\{v_i\}$ positively oriented, and the "−" sign is for $\{v_i\}$ negatively oriented.

Example 8.17. From Example 8.16, we see that for a Riemannian metric g and g-orthonormal basis $\{e_i\}$ we have $\det(g(e_i, e_j)) = +1$. Hence, from equation (8.23), we see that $\det(g(v_i, v_j)) > 0$ for any basis $\{v_i = \phi(e_i)\}$. Thus the g-volume form on a Riemannian space is given by $\pm\sqrt{g}v^1 \wedge \cdots \wedge v^n$.

For a Lorentz metric we have $\det(\eta(e_i, e_j)) = -1$ in a Lorentz frame, and therefore $\det(g(v_i, v_j)) < 0$ in an arbitrary frame. Thus the g-volume in a Lorentz space is given by $\pm\sqrt{-g}v^1 \wedge \cdots \wedge v^n$.

Let us point out that had we defined $\operatorname{Ind}(\eta) = n - 1$ instead of $\operatorname{Ind}(\eta) = 1$, then $\det(\eta(e_i, e_j)) < 0$ only in an even dimensional space. In this case, we would have to write the g-volume as in the above corollary.

Example 8.18. (This example is a continuation of Example 8.12.) We remark that these volume elements are of great practical use in the theory of integration on manifolds. To see an example of how this is done, let us use Examples 8.3 and 8.12 to write the volume element as (remember this applies only locally, and hence the metric depends on the coordinates)

$$d\tau = \sqrt{|g|}\, dx^1 \wedge \cdots \wedge dx^n.$$

If we go to a new coordinate system $\{\bar{x}^i\}$, then

$$\bar{g}_{ij} = \frac{\partial x^r}{\partial \bar{x}^i}\frac{\partial x^s}{\partial \bar{x}^j} g_{rs}$$

so that $|\bar{g}| = (J^{-1})^2\, |g|$ where $J^{-1} = \det(\partial x^r / \partial \bar{x}^i)$ is the determinant of the inverse Jacobian matrix of the transformation. But using $d\bar{x}^i = (\partial \bar{x}^i / \partial x^j) dx^j$

and the properties of the wedge product, it is easy to see that

$$\begin{aligned} d\bar{x}^1 \wedge \cdots \wedge d\bar{x}^n &= \frac{\partial \bar{x}^1}{\partial x^{i_1}} \cdots \frac{\partial \bar{x}^n}{\partial x^{i_n}} dx^{i_1} \wedge \cdots \wedge dx^{i_n} \\ &= \det\left(\frac{\partial \bar{x}^i}{\partial x^j}\right) dx^1 \wedge \cdots \wedge dx^n \end{aligned}$$

and hence

$$d\bar{x}^1 \wedge \cdots \wedge d\bar{x}^n = J dx^1 \wedge \cdots \wedge dx^n$$

where J is the determinant of the Jacobian matrix. (Note that the proper transformation formula for the volume element in multiple integrals arises naturally in the algebra of exterior forms.) We now have

$$\begin{aligned} d\bar{\tau} &= \sqrt{|\bar{g}|}\, d\bar{x}^1 \wedge \cdots \wedge d\bar{x}^n = J^{-1}\sqrt{|g|} J dx^1 \wedge \cdots \wedge dx^n \\ &= \sqrt{|g|}\, dx^1 \wedge \cdots \wedge dx^n = d\tau \end{aligned}$$

and hence $d\tau$ is a scalar called the **invariant volume element**. In the case of $\mathbb{R}^4$ as a Lorentz space, this result is used in the theory of relativity.

Exercises

1. Suppose V has a metric g_{ij} defined on it. Show that for any A, $B \in V$ we have $\langle A, B\rangle = a_i b^i = a^i b_i$.

2. According to the special theory of relativity, the speed of light is the same for all unaccelerated observers regardless of the motion of the source of light relative to the observer. Consider two observers moving at a constant velocity β with respect to each other, and assume that the origins of their respective coordinate systems coincide at $t = 0$. If a spherical pulse of light is emitted from the origin at $t = 0$, then (in units where the speed of light is equal to 1) this pulse satisfies the equation $x^2 + y^2 + z^2 - t^2 = 0$ for the first observer, and $\bar{x}^2 + \bar{y}^2 + \bar{z}^2 - \bar{t}^2 = 0$ for the second observer. We shall use the common notation $(t, x, y, z) = (x^0, x^1, x^2, x^3)$ for our coordinates, and hence the Lorentz metric takes the form

$$\eta_{\mu\nu} = \begin{bmatrix} -1 & & & \\ & 1 & & \\ & & 1 & \\ & & & 1 \end{bmatrix}$$

where $0 \le \mu, \nu \le 3$.

(a) Let the Lorentz transformation matrix be Λ so that $\bar{x}^\mu = \Lambda^\mu{}_\nu x^\nu$. Show the Lorentz transformation must satisfy $\Lambda^T \eta \Lambda = \eta$.

(b) If the $\{\bar{x}^\mu\}$ system moves along the x^1-axis with velocity β, then it turns out the Lorentz transformation is given by

$$\begin{aligned}\bar{x}^0 &= \gamma(x^0 - \beta x^1)\\ \bar{x}^1 &= \gamma(x^1 - \beta x^0)\\ \bar{x}^2 &= x^2\\ \bar{x}^3 &= x^3\end{aligned}$$

where $\gamma^2 = 1/(1-\beta^2)$. Using $\Lambda^\mu{}_\nu = \partial\bar{x}^\mu/\partial x^\nu$, write out the matrix $(\Lambda^\mu{}_\nu)$, and verify explicitly that $\Lambda^T\eta\Lambda = \eta$.

(c) The electromagnetic field tensor is given by

$$F_{\mu\nu} = \begin{bmatrix} 0 & -E_x & -E_y & -E_z \\ E_x & 0 & B_z & -B_y \\ E_y & -B_z & 0 & B_x \\ E_z & B_y & -B_x & 0 \end{bmatrix}.$$

Using this, find the components of the electric field $\vec{E}$ and magnetic field $\vec{B}$ in the $\{\bar{x}^\mu\}$ coordinate system. In other words, find $\overline{F}_{\mu\nu}$. (The actual definition of $F_{\mu\nu}$ is given by $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ where $\partial_\mu = \partial/\partial x^\mu$ and $A_\mu = (\phi, A_1, A_2, A_3)$ is related to $\vec{E}$ and $\vec{B}$ through the classical equations $\vec{E} = -\nabla\phi - \partial\vec{A}/\partial t$ and $\vec{B} = \nabla\times\vec{A}$. See also Exercise 8.2.5.)

3. Let V be an n-dimensional vector space with a Lorentz metric η, and let W be an $(n-1)$-dimensional subspace of V. Note that

$$W^\perp = \{v \in V : \eta(v, w) = 0 \text{ for all } w \in W\}$$

is the 1-dimensional subspace of all normal vectors for W. We say that W is **timelike** if every normal vector is spacelike, **null** if every normal vector is null, and **spacelike** if every normal vector is timelike. Prove that η restricted to W is

(a) Positive definite if W is spacelike.
(b) A Lorentz metric if W is timelike.
(c) Degenerate if W is null.

4. (a) Let D be a 3×3 determinant considered as a function of three contravariant vectors $A_{(1)}$, $A_{(2)}$ and $A_{(3)}$. Show that under a change of coordinates with positive Jacobian determinant, D does not transform as a scalar, but that $D\sqrt{\det g} := D\sqrt{|g|}$ does transform as a proper scalar. (Assume that $\operatorname{Ind}(g) = 0$ throughout this problem.)

(b) According to Theorem 3.3, for any matrix $A = (a^i{}_j) \in M_n(\mathcal{F})$ we have

$$\varepsilon_{j_1\cdots j_n} \det A = \varepsilon_{i_1\cdots i_n} a^{i_1}{}_{j_1} \cdots a^{i_n}{}_{j_n}\,.$$

If we apply this to the Jacobian matrix $(\partial x^i/\partial \bar{x}^j)$ we obtain

$$\varepsilon_{j_1\cdots j_n} \det\!\left(\frac{\partial x}{\partial \bar{x}}\right) = \varepsilon_{i_1\cdots i_n} \frac{\partial x^{i_1}}{\partial \bar{x}^{j_1}} \cdots \frac{\partial x^{i_n}}{\partial \bar{x}^{j_n}}$$

or

$$\varepsilon_{j_1\cdots j_n} = \varepsilon_{i_1\cdots i_n} \frac{\partial x^{i_1}}{\partial \bar{x}^{j_1}} \cdots \frac{\partial x^{i_n}}{\partial \bar{x}^{j_n}} \det\!\left(\frac{\partial \bar{x}}{\partial x}\right)$$

where the matrix $(\partial \bar{x}/\partial x)$ is the inverse to $(\partial x/\partial \bar{x})$. Except for the determinant factor on the right, this looks like the transformation rule for a tensor going from the coordinate system x^i to the coordinate system $\bar{x}^j$. We may denote the left-hand side by $\bar{\varepsilon}_{j_1\cdots j_n}$, but be sure to realize that by definition of the Levi-Civita *symbol* this has the same values as $\varepsilon_{i_1\cdots i_n}$.

If we take this result as defining the transformation law for the Levi-Civita *symbol*, show that the quantity

$$e_{j_1\cdots j_n} := \varepsilon_{j_1\cdots j_n} \sqrt{|g|}$$

transforms like a *tensor*. (This is the Levi-Civita tensor in general coordinates. Note that in a g-orthonormal coordinate system this reduces to the Levi-Civita symbol.)

(c) Using the metric to raise the components of the tensor $e_{j_1\cdots j_n}$, what is the contravariant version of the tensor in part (b)?

Appendix A

Elementary Real Analysis

In this appendix, we briefly go through some elementary concepts from analysis dealing with numbers and functions. While most readers will probably be familiar with this material, it is worth summarizing the basic definitions that we will be using throughout this text, and thus ensure that everyone is on an equal footing to begin with. This has the additional advantage in that it also makes this text virtually self-contained and all the more useful for self-study. The reader should feel free to skim this appendix now, and return to certain sections if and when the need arises.

A.1 Sets

For our purposes, it suffices to assume that the concept of a set is intuitively clear, as is the notion of the set of integers. In other words, a **set** is a collection of objects, each of which is called a **point** or an **element** of the set. For example, the set of integers consists of the numbers $0, \pm 1, \pm 2, \dots$ and will be denoted by $\mathbb{Z}$. Furthermore, the set $\mathbb{Z}^+$ consisting of the numbers $1, 2, \dots$ will be called the set of **positive integers**, while the collection $0, 1, 2, \dots$ is called the set of **natural numbers** (or **nonnegative integers**). If m and $n \neq 0$ are integers, then the set of all numbers of the form m/n is called the set of **rational numbers** and will be denoted by $\mathbb{Q}$. We shall shortly show that there exist real numbers not of this form. The most important sets of numbers that we shall be concerned with are the set $\mathbb{R}$ of real numbers and the set $\mathbb{C}$ of complex numbers (both of these sets will be discussed below).

If S and T are sets, then S is said to be a **subset** of T if every element of S is also an element of T, i.e., $x \in S$ implies $x \in T$. If in addition $S \neq T$, then S is said to be a **proper** subset of T. To denote the fact that S is a subset of T, we write $S \subset T$ (or sometimes $T \supset S$ in which case T is said to be a **superset** of S). Note that if $S \subset T$ and $T \subset S$, then $S = T$. This fact will be extremely useful in many proofs. The set containing no elements at all is called the **empty set** and will be denoted by $\varnothing$.

Next, consider the set of all elements which are members of T but not members of S. This defines the set denoted by $T-S$ and called the **complement** of S in T. (Many authors denote this set by $T\backslash S$, but we shall not use this notation.) In other words, $x \in T-S$ means that $x \in T$ but $x \notin S$. If (as is usually the case) the set T is understood and $S \subset T$, then we write the complement of S as S^c.

Example A.1. Let us prove the useful fact that if $A, B \in X$ with $A^c \subset B$, then it is true that $B^c \subset A$. To show this, we simply note that $x \in B^c$ implies $x \notin B$, which then implies $x \notin A^c$, and hence $x \in A$. This observation is quite useful in proving many identities involving sets.

Now let $S_1, S_2, \ldots$ be a collection of sets. (Such a collection is called a **family** of sets.) For simplicity we write this collection as $\{S_i\}, i \in I$. The set I is called an **index set**, and is most frequently taken to be the set $\mathbb{Z}^+$. The **union** $\cup_{i\in I} S_i$ of the collection $\{S_i\}$ is the set of all elements that are members of at least one of the S_i. Since the index set is usually understood, we will simply write this as $\cup S_i$. In other words, we write

$$\cup S_i = \{x : x \in S_i \text{ for at least one } i \in I\}.$$

This notation will be used throughout this text, and is to be read as "the set of all x such that x is an element of S_i for at least one $i \in I$."

Similarly, the **intersection** $\cap S_i$ of the S_i is given by

$$\cap S_i = \{x : x \in S_i \text{ for all } i \in I\}.$$

For example, if $S, T \subset X$, then $S - T = S \cap T^c$ where $T^c = X - T$. Furthermore, two sets S_1 and S_2 are said to be **disjoint** if $S_1 \cap S_2 = \varnothing$.

We now use these concepts to prove the extremely useful "**De Morgan Formulas**."

Theorem A.1. *Let $\{S_i\}$ be a family of subsets of some set T. Then*
(i) $\cup {S_i}^c = (\cap S_i)^c$.
(ii) $\cap {S_i}^c = (\cup S_i)^c$.

Proof. (i) $x \in \cup {S_i}^c$ if and only if x is an element of some ${S_i}^c$, hence if and only if x is not an element of some S_i, hence if and only if x is not an element of $\cap S_i$, and therefore if and only if $x \in (\cap S_i)^c$.

(ii) $x \in \cap {S_i}^c$ if and only if x is an element of every ${S_i}^c$, hence if and only if x is not an element of any S_i, and therefore if and only if $x \in (\cup S_i)^c$. ∎

While this may seem like a rather technical result, it is in fact directly useful not only in mathematics, but also in many engineering fields such as digital electronics where it may be used to simplify various logic circuits.

Finally, if $S_1, S_2, \ldots, S_n$ is a collection of sets, we may form the (ordered) set of all n-tuples $(x_1, \ldots, x_n)$ where each $x_i \in S_i$. This very useful set is denoted by $S_1 \times \cdots \times S_n$ and called the **Cartesian product** of the S_i.

Example A.2. Probably the most common example of the Cartesian product is the plane $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$. Each point $\vec{x} \in \mathbb{R}^2$ has **coordinates** (x, y) where $x, y \in \mathbb{R}$. In order to facilitate the generalization to $\mathbb{R}^n$, we will generally write $\vec{x} = (x_1, x_2)$ or $\vec{x} = (x^1, x^2)$. This latter notation is used extensively in more advanced topics such as tensor analysis, and there is usually no confusion between writing the components of $\vec{x}$ as superscripts and their being interpreted as exponents (see Chapter 8)

Exercises

1. Let A, B and C be sets. Prove that
   (a) $(A - B) \cap C = (A \cap C) - (B \cap C)$.
   (b) $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$.
   (c) $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$.
   (d) $(A - B) - C = A - (B \cup C)$.
   (e) $A - (B \cup C) = (A - B) \cap (A - C)$.

2. The **symmetric difference** $A \triangle B$ of two sets A and B is defined by
$$A \triangle B = (A - B) \cup (B - A).$$
   Show that
   (a) $A \triangle B = (A \cup B) - (A \cap B) = B \triangle A$.
   (b) $A \cap (B \triangle C) = (A \cap B) \triangle (A \cap C)$.
   (c) $A \cup B = (A \triangle B) \triangle (A \cap B)$.
   (d) $A - B = A \triangle (A \cap B)$.

3. Let $\mathcal{R}$ be a nonempty collection of sets with the property that $A, B \in \mathcal{R}$ implies that both $A \cap B$ and $A \triangle B$ are also in $\mathcal{R}$. Show that $\mathcal{R}$ must contain the empty set, $A \cup B$ and $A - B$. (The collection $\mathcal{R}$ is called a **ring of sets**, and is of fundamental importance in measure theory and Lebesgue integration.)

A.2 Mappings

Given two sets S and T, a **mapping** or **function** f from S to T is a rule which assigns a *unique* element $y \in T$ to each element $x \in S$. Symbolically, we write this mapping as $f : S \to T$ or $f : x \mapsto f(x)$ (this use of the colon should not be confused with its usage meaning "such that"). The set S is called the **domain** of f and T is called the **range** of f. Each point $f(x) \in T$ is called the **image** of x under f (or the **value** of f at x), and the collection $f(x) \in T : x \in S$ of all such image points is called the **image** of f.

In general, whenever a new mapping is given, we must check to see that it is in fact **well-defined**. This means that a given point $x \in S$ is mapped into a *unique* point $f(x) \in T$. In other words, we must verify that $f(x) \neq f(y)$ implies $x \neq y$. An equivalent way of saying this is the **contrapositive** statement that $x = y$ implies $f(x) = f(y)$. We will use this requirement several times throughout the text.

If $A \subset S$, the set $f(x) : x \in A$ is called the **image** of A under f and is denoted by $f(A)$. If f is a mapping from S to T and $A \subset S$, then the **restriction** of f to A, denoted by $f|A$ (or sometimes f_A), is the function from A to T defined by $f|A : x \in A \mapsto f(x) \in T$. If $x' \in T$, then any element $x \in S$ such that $f(x) = x'$ is called an **inverse image** of x' (this is sometimes also called a **preimage** of x'). Note that in general there may be more than one inverse image for any particular $x' \in T$. Similarly, if $A' \subset T$, then the inverse image of A' is the subset of S given by $x \in S : f(x) \in A'$. We will denote the inverse image of A' by $f^{-1}(A')$.

Let f be a mapping from S to T. Note that every element of T need not necessarily be the image of some element of S. However, if $f(S) = T$, then f is said to be **onto** or **surjective**. In other words, f is surjective if given any $x' \in T$ there exists $x \in S$ such that $f(x) = x'$. In addition, f is said to be **one-to-one** or **injective** if $x \neq y$ implies that $f(x) \neq f(y)$. An alternative characterization is to say that f is injective if $f(x) = f(y)$ implies that $x = y$.

If f is both injective and surjective, then f is said to be **bijective**. In this case, given any $x' \in T$ there exists a *unique* $x \in S$ such that $x' = f(x)$. If f is bijective, then we may define the **inverse mapping** $f^{-1} : T \to S$ in the following way. For any $x' \in T$, we let $f^{-1}(x')$ be that (unique) element $x \in S$ such that $f(x) = x'$.

Example A.3. Consider the function $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x^2$. This mapping is clearly not surjective since $f(x) \geq 0$ for any $x \in \mathbb{R}$. Furthermore, it is also not injective. Indeed, it is clear that $2 \neq -2$ but $f(2) = f(-2) = 4$. Note also that both the domain and range of f are the whole set $\mathbb{R}$, but that the image of f is just the subset of all nonnegative real numbers (i.e., the set of all $x \in \mathbb{R}$ with $x \geq 0$).

On the other hand, it is easy to see that the mapping $g : \mathbb{R} \to \mathbb{R}$ defined by $g(x) = ax + b$ for any $a, b \in \mathbb{R}$ (with $a \neq 0$) is a bijection. In this case the inverse mapping is simply given by $g^{-1}(x') = (x' - b)/a$.

Example A.4. If f is a mapping defined on the collections $\{A_i\}$ and $\{B_i\}$ of sets, then we claim that

$$f(\cup A_i) = \cup f(A_i)$$

and

$$f^{-1}(\cup B_i) = \cup f^{-1}(B_i).$$

To prove these relationships we proceed in our usual manner. Thus we have $x' \in f(\cup A_i)$ if and only if $x' = f(x)$ for some $x \in \cup A_i$, hence if and only if x' is in some $f(A_i)$, and therefore if and only if $x' \in \cup f(A_i)$. This proves the first statement. As to the second statement, we have $x \in f^{-1}(\cup B_i)$ if and only if $f(x) \in \cup B_i$, hence if and only if $f(x)$ is in some B_i, hence if and only if x is in some $f^{-1}(B_i)$, and therefore if and only if $x \in \cup f^{-1}(B_i)$.

Several similar relationships that will be referred to again are given in the exercises.

Now consider the sets S, T and U along with the mappings $f : S \to T$ and $g : T \to U$. We define the **composite mapping** (sometimes also called the **product**) $g \circ f : S \to U$ by

$$(g \circ f)(x) = g(f(x))$$

for all $x \in S$. In general, $f \circ g \neq g \circ f$, and we say that the composition of two functions is not **commutative**. However, if we also have a mapping $h : U \to V$, then for any $x \in S$ we have

$$\begin{aligned}(h \circ (g \circ f))(x) &= h((g \circ f)(x)) = h(g(f(x))) = (h \circ g)(f(x)) \\ &= ((h \circ g) \circ f)(x)\end{aligned}$$

This means that

$$h \circ (g \circ f) = (h \circ g) \circ f$$

and hence the composition of mappings is **associative**.

As a particular case of the composition of mappings, note that if $f : S \to T$ is a bijection and $f(x) = x' \in T$ where $x \in S$, then

$$(f \circ f^{-1})(x') = f(f^{-1}(x')) = f(x) = x'$$

and

$$(f^{-1} \circ f)(x) = f^{-1}(f(x)) = f^{-1}(x') = x.$$

If we write $f \circ f^{-1} = I_T$, then the mapping I_T has the property that $I_T(x') = x'$ for every $x' \in T$. We call I_T the **identity mapping** on T. Similarly, the composition mapping $f^{-1} \circ f = I_S$ is called the **identity mapping** on S. In the particular case that $S = T$, then $f \circ f^{-1} = f^{-1} \circ f = I$ is also called the **identity mapping**.

An extremely important result follows by noting that (even if $S \neq T$)

$$\begin{aligned}(f^{-1} \circ g^{-1})(g \circ f)(x) &= (f^{-1} \circ g^{-1})(g(f(x))) = f^{-1}(g^{-1}(g(f(x)))) \\ &= f^{-1}(f(x)) = x\end{aligned}$$

Since it is also easy to see that $(g \circ f)(f^{-1} \circ g^{-1})(x') = x'$, we have shown that

$$(g \circ f)^{-1} = f^{-1} \circ g^{-1}.$$

Exercises

1. Let f be a mapping of sets. For each of the following, state any conditions on f that may be required (e.g., surjective or injective), and then prove the statement:
   (a) $A_1 \subset A_2$ implies $f(A_1) \subset f(A_2)$.
   (b) $f(A)^c \subset f(A^c)$ is true if and only if f is surjective.
   (c) $f(\cap A_i) \subset \cap f(A_i)$.
   (d) $B_1 \subset B_2$ implies $f^{-1}(B_1) \subset f^{-1}(B_2)$.
   (e) $f^{-1}(\cap B_i) = \cap f^{-1}(B_i)$.
   (f) $f^{-1}(B^c) = f^{-1}(B)^c$.
2. Given a nonempty set A, we define the **identity mapping** $i_A : A \to A$ by $i_A(a) = a$ for every $a \in A$. Let $f : A \to A$ be any mapping.
   (a) Show that $f \circ i_A = i_A \circ f = f$.
   (b) If f is bijective (so that f^{-1} exists), show that $f \circ f^{-1} = f^{-1} \circ f = i_A$.
   (c) Let f be a bijection, and suppose that g is any other mapping with the property that $g \circ f = f \circ g = i_A$. Show that $g = f^{-1}$.

A.3 Orderings and Equivalence Relations

Given any two sets S and T, a subset R of $S \times T$ is said to be a **relation** between S and T. If $R \subset S \times T$ and $(x, y) \in R$, then it is common to write xRy to show that x and y are "R-related." In particular, consider the relation symbolized by $\preceq$ and defined as having the following properties on a set S:

(a) $x \preceq x$ (reflexivity);
(b) $x \preceq y$ and $y \preceq x$ implies $x = y$ for all $x, y \in S$ (antisymmetry);
(c) $x \preceq y$ and $y \preceq z$ implies $x \preceq z$ for all $x, y, z \in S$ (transitivity).

Any relation on a non-empty set S having these three properties is said to be a **partial ordering**, and S is said to be a **partially ordered set**. We will sometimes write $y \succeq x$ instead of the equivalent notation $x \preceq y$. The reason for including the qualifying term "partial" in this definition is shown in our next example.

Example A.5. Let S be any set, and let $\mathcal{P}(S)$ be the collection of all subsets of S (this is sometimes called the **power set** of S). If A, B and C are subsets of S, then clearly $A \subset A$ so that (a) is satisfied; $A \subset B$ and $B \subset A$ implies $A = B$ then satisfies (b); and $A \subset B$ and $B \subset C$ implies $A \subset C$ satisfies (c). Therefore $\subset$ defines a partial ordering on $\mathcal{P}(S)$, and the subsets of S are said to be **ordered by inclusion**. Note however, that if $A \subset S$ and $B \subset S$ but $A \not\subset B$ and $B \not\subset A$, then there is no relation between A and B, and we say that A and B are not **comparable**.

The terminology used in this example is easily generalized as follows. If S is any partially ordered set and $x, y \in S$, then we say that x and y are **comparable** if either $x \preceq y$ or $y \preceq x$.

If, in addition to properties (a)–(c), a relation R also has the property that any two elements are comparable, then R is said to be a **total ordering**. In other words, a total ordering also has the property that

(d) either $x \preceq y$ or $y \preceq x$ for all $x, y \in S$.

Let S be a set partially ordered by $\preceq$ and suppose $A \subset S$. It should be clear that A may be considered to be a partially ordered set by defining $a \preceq b$ for all $a, b \in A$ if $a \preceq b$ where a and b are considered to be elements of S. (This is similar to the restriction of a mapping.) We then say that A has a partial ordering $\preceq$ **induced** by the ordering on S. If A is *totally* ordered by the ordering induced by $\preceq$, then A is frequently called a **chain** in S.

Let A be a non-empty subset of a partially ordered set S. An element $x \in S$ is called an **upper bound** for A if $a \preceq x$ *for all* $a \in A$. If it so happens that x is an element of A, then x is said to be a **largest element** of A. Similarly, $y \in S$ is called a **lower bound** for A if $y \preceq a$ *for all* $a \in A$, and y is a **smallest element** of A if $y \in A$. If A has an upper (lower) bound, then we say that A is **bounded above** (**below**). Note that largest and smallest elements need not be unique.

Suppose that A is bounded above by $\alpha \in S$, and in addition, suppose that for any other upper bound x of A we have $\alpha \preceq x$. Then we say that α is a **least upper bound** (or **supremum**) of A, and we write $\alpha = \operatorname{lub} A = \sup A$. As expected, if A is bounded below by $\beta \in S$, and if $y \preceq \beta$ for all other lower bounds $y \in S$, then β is called a **greatest lower bound** (or **infimum**), and we write $\beta = \operatorname{glb} A = \inf A$. In other words, if it exists, the least upper (greatest lower) bound for A is a smallest (largest) element of the set of all upper (lower) bounds for A.

From property (b) above and the definitions of inf and sup we see that, if they exist, the least upper bound and the greatest lower bound are unique. (For example, if β and β' are both greatest lower bounds, then $\beta \preceq \beta'$ and $\beta' \preceq \beta$ implies that $\beta = \beta'$.) Hence it is meaningful to talk about *the* least upper bound and *the* greatest lower bound.

Let S be a partially ordered set, and suppose $A \subset S$. An element $\alpha \in A$ is said to be **maximal in** A if for any element $a \in A$ with $\alpha \preceq a$, we have $a = \alpha$. In other words, no element of A other than α itself is greater than or equal to α. Similarly, an element $\beta \in A$ is said to be **minimal in** A if for any $b \in A$ with $b \preceq \beta$, we have $b = \beta$. Note that a maximal element may not be a largest element (since two elements of a partially ordered set need not be comparable), and there may be many maximal elements in A.

We now state Zorn's lemma, one of the most fundamental results in set theory, and hence in all of mathematics. While the reader can hardly be expected to appreciate the significance of this lemma at the present time, it is in fact extremely powerful.

Zorn's Lemma. *Let S be a partially ordered set in which every chain has an upper bound. Then S contains a maximal element.*

It can be shown (see any book on set theory) that Zorn's lemma is logically equivalent to the **axiom of choice**, which states that given any non-empty family of non-empty disjoint sets, a set can be formed which contains precisely one element taken from each set in the family. Although this seems like a rather obvious statement, it is important to realize that either the axiom of choice or some other statement equivalent to it must be postulated in the formulation of the theory of sets, and thus Zorn's lemma is not really provable in the usual sense. In other words, Zorn's lemma is frequently taken as an axiom of set theory. However, it is an indispensable part of some of what follows although we shall have little occasion to refer to it directly.

Up to this point, we have only talked about one type of relation, the partial ordering. We now consider another equally important relation. Let S be any set. A relation $\approx$ on S is said to be an **equivalence relation** if it has the following properties for all $x, y, z \in S$:

(a) $x \approx x$ for all $x \in S$ (reflexivity);
(b) $x \approx y$ implies $y \approx x$ (symmetry);
(c) $x \approx y$ and $y \approx z$ implies $x \approx z$ for all $x, y, z \in S$ (transitivity).

Note that only (b) differs from the defining relations for a partial ordering.

A **partition** of a set S is a family $\{S_i\}$ of non-empty subsets of S such that $\cup S_i = S$ and $S_i \cap S_j \neq \varnothing$ implies $S_i = S_j$. Suppose $x \in S$ and let $\approx$ be an equivalence relation on S. The subset of S defined by $[x] = \{y : y \approx x\}$ is called the **equivalence class** of x. The most important property of equivalence relations is contained in the following theorem.

Theorem A.2. *The family of all distinct equivalence classes of a set S forms a partition of S. (This is called the partition* ***induced*** *by $\approx$.) Moreover, given any partition of S, there is an equivalence relation on S that induces this partition.*

Proof. Let $\approx$ be an equivalence relation on a set S, and let x be any element of S. Since $x \approx x$, it is obvious that $x \in [x]$. Thus each element of S lies in at least one non-empty equivalence class. We now show that any two equivalence classes are either disjoint or are identical. Let $[x_1]$ and $[x_2]$ be two equivalence classes, and let y be a member of both classes. In other words, $y \approx x_1$ and $y \approx x_2$. Now choose any $z \in [x_1]$ so that $z \approx x_1$. But this means that $z \approx x_1 \approx y \approx x_2$ so that any element of $[x_1]$ is also an element of $[x_2]$, and hence $[x_1] \subset [x_2]$. Had we chosen $z \in [x_2]$ we would have found that $[x_2] \subset [x_1]$. Therefore $[x_1] = [x_2]$, and we have shown that if two equivalence classes have any element in common, then they must in fact be identical. Let $\{S_i\}$ be any partition of S. We define an equivalence relation on S by letting $x \approx y$ if $x, y \in S_i$ for any $x, y \in S$. It should be clear that this does indeed satisfy the three conditions for an equivalence relation, and that this equivalence relation induces the partition $\{S_i\}$. ∎

Exercises

1. Let $\mathbb{Z}^+$ denote the set of positive integers. We write $m|n$ to denote the fact that m divides n, i.e., $n = km$ for some $k \in \mathbb{Z}^+$.
   (a) Show that $|$ defines a partial ordering on $\mathbb{Z}^+$.
   (b) Does $\mathbb{Z}^+$ contain either a maximal or minimal element relative to this partial ordering?
   (c) Prove that any subset of $\mathbb{Z}^+$ containing exactly two elements has a greatest lower bound and a least upper bound.
   (d) For each of the following subsets of $\mathbb{Z}^+$, determine whether or not it is a chain in $\mathbb{Z}^+$, find a maximal and minimal element, an upper and lower bound, and a least upper bound:
      (i) $\{1, 2, 4, 6, 8\}$.
      (ii) $\{1, 2, 3, 4, 5\}$.
      (iii) $\{3, 6, 9, 12, 15, 18\}$.
      (iv) $\{4, 8, 16, 32, 64, 128\}$.

2. Define a relation $\approx$ on $\mathbb{R}$ by requiring that $a \approx b$ if $|a| = |b|$. Show that this defines an equivalence relation on $\mathbb{R}$.

3. For any $a, b \in \mathbb{R}$, let $a \sim b$ mean $ab > 0$. Does $\sim$ define an equivalence relation? What happens if we use $ab \geq 0$ instead of $ab > 0$?

A.4 Cardinality and the Real Number System

We all have an intuitive sense of what it means to say that two finite sets have the same number of elements, but our intuition leads us astray when we come to consider infinite sets. For example, there are as many perfect squares $\{1, 4, 9, 16, \dots\}$ among the positive integers as there are positive integers. That this is true can easily be seen by writing each positive integer paired with its square:

$$\begin{array}{ccccc} 1, & 2, & 3, & 4, & \dots \\ 1^2, & 2^2, & 3^2, & 4^2, & \dots \end{array}$$

While it seems that the perfect squares are only sparsely placed throughout the integers, we have in fact constructed a bijection of all positive integers with all of the perfect squares of integers, and we are forced to conclude that in this sense they both have the "same number of elements."

In general, two sets S and T are said to have the same **cardinality**, or to possess the same number of elements, if there exists a bijection from S to T. A set S is **finite** if it has the same cardinality as either $\varnothing$ or the set $\{1, 2, \dots, n\}$ for some positive integer n; otherwise, S is said to be **infinite**. However, there are varying degrees of "infinity." A set S is **countable** if it has the same cardinality as a subset of the set $\mathbb{Z}^+$ of positive integers. If this is not the case, then we say that S is **uncountable**. Any infinite set which is numerically equivalent to (i.e., has the same cardinality as) $\mathbb{Z}^+$ is said to be **countably infinite**. We therefore say that a set is **countable** if it is countably infinite or if it is non-empty and finite.

It is somewhat surprising (as was first discovered by Cantor) that the set $\mathbb{Q}^+$ of all positive rational numbers is in fact countable. The elements of $\mathbb{Q}^+$ can not be listed in order of increasing size because there is no smallest such number, and between any two rational numbers there are infinitely many others (see Theorem A.4 below). To show that $\mathbb{Q}^+$ is countable, we shall construct a bijection from $\mathbb{Z}^+$ to $\mathbb{Q}^+$.

To do this, we first consider all positive rationals whose numerator and denominator add up to 2. In this case we have only $1/1 = 1$. Next we list those positive rationals whose numerator and denominator add up to 3. If we agree to always list our rationals with numerators in increasing order, then we have $1/2$ and $2/1 = 2$. Those rationals whose numerator and denominator add up to 4 are then given by $1/3$, $2/2 = 1$, $3/1 = 3$. Going on to 5 we obtain $1/4$, $2/3$, $3/2$, $4/1 = 4$. For 6 we have $1/5$, $2/4 = 1/2$, $3/3 = 1$, $4/2 = 2$, $5/1 = 5$. Continuing with this procedure, we list together all of our rationals, omitting any number already listed. This gives us the sequence

$$1,\ 1/2,\ 2,\ 1/3,\ 3,\ 1/4,\ 2/3,\ 3/2,\ 4,\ 1/5,\ 5,\ \dots$$

which contains each positive rational number exactly once, and provides our desired bijection.

We have constructed several countably infinite sets of real numbers, and it is natural to wonder whether there are in fact any uncountably infinite sets. It was another of Cantors discoveries that the set $\mathbb{R}$ of all real numbers is actually uncountable. To prove this, let us assume that we have listed (in some manner similar to that used for the set $\mathbb{Q}^+$) all the real numbers in decimal form. What we shall do is construct a decimal $.d_1d_2d_3\cdots$ that is not on our list, thus showing that the list can not be complete.

Consider only the portion of the numbers on our list to the right of the decimal point, and look at the first number on the list. If the first digit after the decimal point of the first number is a 1, we let $d_1 = 2$; otherwise we let $d_1 = 1$. No matter how we choose the remaining d's, our number will be different from the first on our list. Now look at the second digit after the decimal point of the second number on our list. Again, if this second digit is a 1, we let $d_2 = 2$; otherwise we let $d_2 = 1$. We have now constructed a number that differs from the first two numbers on our list. Continuing in this manner, we construct a decimal $.d_1 d_2 d_3 \cdots$ that differs from every other number on our list, contradicting the assumption that all real numbers can be listed, and proving that $\mathbb{R}$ is actually uncountable.

Since it follows from what we showed above that the set $\mathbb{Q}$ of all rational numbers on the real line is countable, and since we just proved that the set $\mathbb{R}$ is uncountable, it follows that a set of **irrational numbers** must exist and be uncountably infinite.

From now on we will assume that the reader understands what is meant by the real number system, and we proceed to investigate some of its most useful properties. A complete axiomatic treatment that justifies what we already know would take us too far afield, and the interested reader is referred to, e.g., [45].

Let S be any ordered set, and let $A \subset S$ be non-empty and bounded above. We say that S has the **least upper bound property** if $\sup A$ exists in S. In the special case of $S = \mathbb{R}$, we have the following extremely important axiom.

Archimedean Axiom. *Every non-empty set of real numbers which has an upper (lower) bound has a least upper bound (greatest lower bound).*

The usefulness of this axiom is demonstrated in the next rather obvious though important result, sometimes called the **Archimedean property** of the real number line.

Theorem A.3. *Let $a, b \in \mathbb{R}$ and suppose $a > 0$. Then there exists $n \in \mathbb{Z}^+$ such that $na > b$.*

Proof. Let S be the set of all real numbers of the form na where n is a positive integer. If the theorem were false, then b would be an upper bound for S. But by the Archimedean axiom, S has a least upper bound $\alpha = \sup S$. Since $a > 0$, we have $\alpha - a < \alpha$ and $\alpha - a$ can not be an upper bound of S (by definition of α). Therefore, there exists an $m \in \mathbb{Z}^+$ such that $ma \in S$ and $\alpha - a < ma$. But then $\alpha < (m+1)a \in S$ which contradicts the fact that $\alpha = \sup S$. ∎

One of the most useful facts about the real line is that the set $\mathbb{Q}$ of all rational numbers is **dense** in $\mathbb{R}$. By this we mean that given any two distinct real numbers, we can always find a rational number between them. This means that any real number may be approximated to an arbitrary degree of accuracy by a rational number. It is worth proving this using Theorem A.3.

Theorem A.4. *Suppose $x, y \in \mathbb{R}$ and assume that $x < y$. Then there exists a rational number $p \in \mathbb{Q}$ such that $x < p < y$.*

Proof. Since $x < y$ we have $y - x > 0$. In Theorem A.3, choose $a = y - x$ and $b = 1$ so there exists $n \in \mathbb{Z}^+$ such that $n(y - x) > 1$ or, alternatively,

$$1 + nx < ny.$$

Applying Theorem A.3 again, we let $a = 1$ and both $b = nx$ and $b = -nx$ to find integers $m_1, m_2 \in \mathbb{Z}^+$ such that $m_1 > nx$ and $m_2 > -nx$. Rewriting the second of these as $-m_2 < nx$, we combine the two inequalities to obtain

$$-m_2 < nx < m_1$$

so that nx lies between two integers. But if nx lies between two integers, it must lie between two consecutive integers $m - 1$ and m for some $m \in \mathbb{Z}$ where $-m_2 \leq m \leq m_1$. Thus $m - 1 \leq nx < m$ implies that $m \leq 1 + nx$ and $nx < m$. We therefore obtain

$$nx < m \leq 1 + nx < ny$$

or, equivalently (since $n \neq 0$), $x < m/n < y$. ∎

Corollary. *Suppose $x, y \in \mathbb{R}$ and assume that $x < y$. Then there exist integers $m \in \mathbb{Z}$ and $k \geq 0$ such that $x < m/2k < y$.*

Proof. Simply note that the proof of Theorem A.4 could be carried through if we choose an integer $k \geq 0$ so that $2k(y - x) > 1$, and replace n by $2k$ throughout. ∎

In addition to the real number system $\mathbb{R}$ we have been discussing, it is convenient to introduce the **extended real number system** as follows. To the real number system $\mathbb{R}$, we adjoin the symbols $+\infty$ and $-\infty$ which are *defined* to have the property that $-\infty < x < +\infty$ *for all* $x \in \mathbb{R}$. This is of great notational convenience. We stress however, that neither $+\infty$ or $-\infty$ are considered to be elements of $\mathbb{R}$.

Suppose A is a non-empty set of real numbers. We have already defined $\sup A$ in the case where A has an upper bound. If A is non-empty and has no upper bound, then we say that $\sup A = +\infty$, and if $A = \varnothing$, then $\sup A = -\infty$. Similarly, if $A \neq \varnothing$ and has no lower bound, then $\inf A = -\infty$, and if $A = \varnothing$, then $\inf A = +\infty$.

Suppose $a, b \in \mathbb{R}$ with $a \leq b$. Then the **closed interval** $[a, b]$ from a to b is the subset of $\mathbb{R}$ defined by

$$[a, b] = \{x \in \mathbb{R} : a \leq x \leq b\}.$$

Similarly, the **open interval** (a, b) is defined to be the subset

$$(a, b) = \{x : a < x < b\}.$$

We may also define the **open-closed** and **closed-open** intervals in the obvious way. The **infinity symbols** $\pm\infty$ thus allow us to talk about intervals of the form $(-\infty, b], [a, +\infty)$ and $(-\infty, +\infty)$.

Another property of the sup that will be needed later on is contained in the following theorem. By way of notation, we define $\mathbb{R}^+$ to be the set of all real numbers > 0, and $\overline{\mathbb{R}}^+ = \mathbb{R}^+ \cup \{0\}$ to be the set of all real numbers ≥ 0.

Theorem A.5. *Let A and B be non-empty bounded sets of real numbers, and define the sets*

$$A + B = \{x + y : x \in A \text{ and } y \in B\}$$

and

$$AB = \{xy : x \in A \text{ and } y \in B\}.$$

Then

(i) For all $A, B \subset \mathbb{R}$ we have $\sup(A + B) = \sup A + \sup B$.
(ii) For all $A, B \in \overline{\mathbb{R}}^+$ we have $\sup(AB) \leq (\sup A)(\sup B)$.

Proof. (i) Let $\alpha = \sup A$, $\beta = \sup B$, and suppose $x + y \in A + B$. Then

$$x + y \leq \alpha + y \leq \alpha + \beta$$

so that $\alpha+\beta$ is an upper bound for $A+B$. Now note that given $\varepsilon > 0$, there exists $x \in A$ such that $\alpha - \varepsilon/2 < x$ (or else α would not be the least upper bound). Similarly, there exists $y \in B$ such that $\beta - \varepsilon/2 < y$. Then $\alpha + \beta - \varepsilon < x + y$ so that $\alpha + \beta$ must be the least upper bound for $A + B$.

(ii) If $x \in A \subset \overline{\mathbb{R}}^+$ we must have $x \leq \sup A$, and if $y \in B \subset \overline{\mathbb{R}}^+$ we have $y \leq \sup B$. Hence $xy \leq (\sup A)(\sup B)$ for all $xy \in AB$, and therefore $A \neq \varnothing$ and $B \neq \varnothing$ implies

$$\sup(AB) \leq (\sup A)(\sup B).$$

The reader should verify that strict equality holds if $A \subset \mathbb{R}^+$ and $B \subset \mathbb{R}^+$. ∎

The last topic in our treatment of real numbers that we wish to discuss is the absolute value. Note that if $x \in \mathbb{R}$ and $x^2 = a$, then we also have $(-x)^2 = a$. We *define* $\sqrt{a}$, for $a \geq 0$, to be the unique *positive* number x such that $x^2 = a$, and we call x the **square root** of a.

Suppose $x, y \geq 0$ and let $x^2 = a$ and $y^2 = b$. Then $x = \sqrt{a}$, $y = \sqrt{b}$ and we have $(\sqrt{a}\sqrt{b})^2 = (xy)^2 = x^2y^2 = ab$ which implies that

$$\sqrt{ab} = \sqrt{a}\sqrt{b}.$$

For any $a \in \mathbb{R}$, we define its **absolute value** $|a|$ by $|a| = \sqrt{a^2}$. It then follows that $|-a| = |a|$, and hence

$$|a| = \begin{cases} a & \text{if } a \geq 0 \\ -a & \text{if } a < 0 \end{cases}$$

This clearly implies that

$$a \leq |a| \, .$$

In addition, if $a, b \geq 0$ and $a \leq b$, then we have $(\sqrt{a})^2 = a \leq b = (\sqrt{b})^2$ so that $\sqrt{a} \leq \sqrt{b}$.

The absolute value has two other useful properties. First, we note that

$$|ab| = \sqrt{(ab)^2} = \sqrt{a^2b^2} = \sqrt{a^2}\sqrt{b^2} = |a|\,|b| \, .$$

Second, we see that

$$\begin{aligned} |a+b|^2 &= (a+b)^2 \\ &= a^2 + b^2 + 2ab \\ &\leq |a|^2 + |b|^2 + 2\,|ab| \\ &= |a|^2 + |b|^2 + 2\,|a|\,|b| \\ &= (|a| + |b|)^2 \end{aligned}$$

and therefore

$$|a+b| \leq |a| + |b| \, .$$

Using these results, many other useful relationships may be obtained. For example, $|a| = |a+b-b| \leq |a+b| + |-b| = |a+b| + |b|$ so that

$$|a| - |b| \leq |a+b| \, .$$

Others are to be found in the exercises.

Example A.6. Let us show that if $\varepsilon > 0$, then $|x| < \varepsilon$ if and only if $-\varepsilon < x < \varepsilon$. Indeed, we see that if $x > 0$, then $|x| = x < \varepsilon$, and if $x < 0$, then $|x| = -x < \varepsilon$ which implies $-\varepsilon < x < 0$ (we again use the fact that $a < b$ implies $-b < -a$). Combining these results shows that $|x| < \varepsilon$ implies $-\varepsilon < x < \varepsilon$. We leave it to the reader to reverse the argument and complete the proof.

A particular case of this result that will be of use later on comes from letting $x = a-b$. We then see that $|a-b| < \varepsilon$ if and only if $-\varepsilon < a-b < \varepsilon$. Rearranging, this may be written in the form $b - \varepsilon < a < b + \varepsilon$. The reader should draw a picture of this relationship.

Exercises

1. Prove that if A and B are countable sets, then $A \times B$ is countable.

2. (a) A real number x is said to be **algebraic** (over the rationals) if it satisfies some polynomial equation of positive degree with rational coefficients:
$$x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0 = 0.$$
Given the fact (see Section 5.1) that each polynomial equation has only finitely many roots, show that the set of all algebraic numbers is countable.
(b) We say that a real number x is **transcendental** if it is not algebraic (the most common transcendental numbers are π and e). Using the fact that the reals are uncountable, show that the set of all transcendental numbers is also uncountable.

3. If $a, b \ge 0$, show that $\sqrt{ab} \le (a+b)/2$.

4. For any $a, b \in \mathbb{R}$, show that:
(a) $\big| |a| - |b| \big| \le |a + b|$.
(b) $\big| |a| - |b| \big| \le |a - b|$.

5. (a) If $A \subset \mathbb{R}$ is nonempty and bounded below, show $\sup(-A) = -\inf A$.
(b) If $A \subset \mathbb{R}$ is nonempty and bounded above, show $\inf(-A) = -\sup A$.

A.5 Induction

Another important concept in the theory of sets is called "well-ordering." In particular, we say that a totally ordered set S is **well-ordered** if *every* non-empty subset A of S has a smallest element. For example, consider the set S of all rational numbers in the interval $[0, 1]$. It is clear that 0 is the smallest element of S, but the subset of S consisting of all rational numbers > 0 has no smallest element (this is a consequence of Theorem A.4).

For our purposes, it is an (apparently obvious) axiom that every non-empty set of natural numbers has a smallest element. In other words, the natural numbers are well-ordered. The usefulness of this axiom is that it allows us to prove an important property called **induction**.

Theorem A.6. *Assume that for all $n \in \mathbb{Z}^+$ we are given an assertion $A(n)$, and assume it can be shown that:*
(i) $A(1)$ is true;
(ii) If $A(n)$ is true, then $A(n+1)$ is true.
Then $A(n)$ is true for all $n \in \mathbb{Z}^+$.

Proof. If we let S be that subset of $\mathbb{Z}^+$ for which $A(n)$ is not true, then we must show that $S = \varnothing$. According to our well-ordering axiom, if $S \neq \varnothing$ then S contains a least element which we denote by N. By assumption (1), we must have $N \neq 1$ and hence $N > 1$. Since N is a least element, $N - 1 \notin S$ so that $A(N-1)$ must be true. But then (ii) implies that $A(N)$ must be true which contradicts the definition of N. ∎

Example A.7. Let $n > 0$ be an integer. We define n-**factorial**, written $n!$, to be the number

$$n! = n(n-1)(n-2)\cdots(2)(1)$$

with $0!$ defined to be 1. The **binomial coefficient** $\binom{n}{k}$ is defined by

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

where n and k are nonnegative integers. We leave it to the reader (see Exercise A.6.1) to show that

$$\binom{n}{k} = \binom{n}{n-k}$$

and

$$\binom{n}{k-1} + \binom{n}{k} = \binom{n+1}{k}$$

What we wish to prove is the **binomial theorem**:

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}.$$

We proceed by induction as follows.

For $n = 1$, we have

$$\binom{1}{0} x^0 y^1 + \binom{1}{1} x^1 y^0 = (x+y)^1$$

so the assertion is true for $n = 1$. We now assume the theorem holds for n, and proceed to show that it also holds for $n+1$. We have

$$\begin{aligned}(x+y)^{n+1} &= (x+y)(x+y)^n = (x+y)\left[\sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}\right] \\ &= \sum_{k=0}^{n} \binom{n}{k} x^{k+1} y^{n-k} + \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k+1} \qquad (*)\end{aligned}$$

By relabelling the summation index, we see that for any function f with domain equal to $\{0, 1, \ldots, n\}$ we have

$$\sum_{k=0}^{n} f(k) = f(0) + f(1) + \cdots + f(n) = \sum_{k=1}^{n+1} f(k-1).$$

We use this fact in the first sum in (*), and separate out the $k = 0$ term in the second to obtain

$$(x+y)^{n+1} = \sum_{k=1}^{n+1} \binom{n}{k-1} x^k y^{n-k+1} + y^{n+1} + \sum_{k=1}^{n} \binom{n}{k} x^{k+1} y^{n-k+1}.$$

We now separate out the $k = n + 1$ term from the first sum in this expression and group terms to find

$$\begin{aligned}(x+y)^{n+1} &= x^{n+1} + y^{n+1} + \sum_{k=1}^{n} \left[\binom{n}{k-1} + \binom{n}{k} \right] x^k y^{n-k+1} \\ &= x^{n+1} + y^{n+1} + \sum_{k=1}^{n} \binom{n+1}{k} x^k y^{n+1-k} \\ &= \sum_{k=0}^{n+1} \binom{n+1}{k} x^k y^{n+1-k}\end{aligned}$$

as was to be shown.

A.6 Complex Numbers

At this time we wish to formally define the complex number system $\mathbb{C}$, although most readers should already be familiar with its basic properties. The motivation for the introduction of such numbers comes from the desire to solve equations such as $x^2 + 1 = 0$ which leads to the square root of a negative number. We may proceed by manipulating square roots of negative numbers as if they were square roots of positive numbers. However, a consequence of this is that on the one hand, $(\sqrt{-1})^2 = -1$, while on the other hand

$$(\sqrt{-1})^2 = \sqrt{-1}\sqrt{-1} = \sqrt{(-1)(-1)} = \sqrt{+1} = 1.$$

In order to avoid paradoxical manipulations of this type, the symbol i was introduced by Euler (in 1779) with the defining property that $i^2 = -1$. Then, if $a > 0$, we have $\sqrt{-a} = i\sqrt{a}$. Using this notation, a **complex number** $z \in \mathbb{C}$ is a number of the form $z = x + iy$ where $x \in \mathbb{R}$ is called the **real part** of z (written $\operatorname{Re} z$), and $y \in \mathbb{R}$ is called the **imaginary part** of z (written $\operatorname{Im} z$).

Two complex numbers $x + iy$ and $u + iv$ are said to be equal if $x = u$ and $y = v$. Algebraic operations in $\mathbb{C}$ are *defined* as follows:

addition:	$(x + iy) + (u + iv) = (x + u) + i(y + v)$.
subtraction:	$(x + iy) - (u + iv) = (x - u) + i(y - v)$.
multiplication:	$(x + iy)(u + iv) = (xu - yv) + i(xv + yu)$.

division:
$$\frac{(x+iy)}{(u+iv)} = \frac{(x+iy)(u-iv)}{(u+iv)(u-iv)} = \frac{(xu+yv)+i(yu-vx)}{u^2+v^2}.$$

It should be clear that the results for multiplication and division may be obtained by formally multiplying out the terms and using the fact that $i^2 = -1$.

The **complex conjugate** z^* (or sometimes $\bar{z}$) of a complex number $z = x + iy$ is defined to be the complex number $z^* = x - iy$. Note that if $z, w \in \mathbb{C}$ we have

$$\begin{aligned}
(z+w)^* &= z^* + w^* \\
(zw)^* &= z^* w^* \\
z + z^* &= 2\operatorname{Re} z \\
z - z^* &= 2i \operatorname{Im} z
\end{aligned}$$

The **absolute value** (or **modulus**) $|z|$ of a complex number $z = x + iy$ is defined to be the real number

$$|z| = \sqrt{x^2+y^2} = (zz^*)^{1/2}.$$

By analogy to the similar result for real numbers, if $z, w \in \mathbb{C}$ then (using the fact that $z = x + iy$ implies $\operatorname{Re} z = x \leq \sqrt{x^2+y^2} = |z|$)

$$\begin{aligned}
|z+w|^2 &= (z+w)(z+w)^* \\
&= zz^* + zw^* + z^*w + ww^* \\
&= |z|^2 + 2\operatorname{Re}(zw^*) + |w|^2 \\
&\leq |z|^2 + 2\,|zw^*| + |w|^2 \\
&= |z|^2 + 2\,|z|\,|w| + |w|^2 \\
&= (|z| + |w|)^2
\end{aligned}$$

and hence taking the square root of both sides yields

$$|z+w| \leq |z| + |w|\,.$$

Let the sum $z_1 + \cdots + z_n$ be denoted by $\sum_{i=1}^n z_i$. The following theorem is known as **Schwartz's inequality**.

Theorem A.7 (**Schwartz's Inequality**). *Let $a_1, \ldots, a_n$ and $b_1, \ldots, b_n$ be complex numbers. Then*

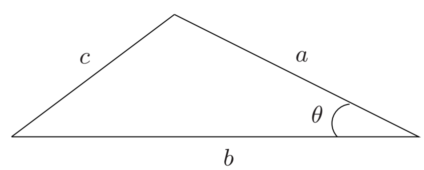
$$\left|\sum_{j=1}^n a_j b_j^*\right| \leq \left(\sum_{j=1}^n |a_j|^2\right)^{1/2} \left(\sum_{j=1}^n |b_j|^2\right)^{1/2}$$

Proof. Write (suppressing the limits on the sum) $A = \sum_j |a_j|^2$, $B = \sum_j |b_j|^2$ and $C = \sum_j a_j b_j^*$. If $B = 0$, then $b_j = 0$ for all $j = 1, \ldots, n$ and there is nothing to prove, so we assume that $B \neq 0$. We then have

$$\begin{aligned}
0 &\leq \sum_i |Ba_i - Cb_i|^2 \\
&= \sum_i (Ba_i - Cb_i)(Ba_i^* - C^* b_i^*) \\
&= B^2 \sum_i |a_i|^2 - BC^* \sum_i a_i b_i^* - BC \sum_i a_i^* b_i + |C|^2 \sum_i |b_i|^2 \\
&= B^2 A - B\,|C|^2 - B\,|C|^2 + |C|^2\, B \\
&= B(AB - |C|^2).
\end{aligned}$$

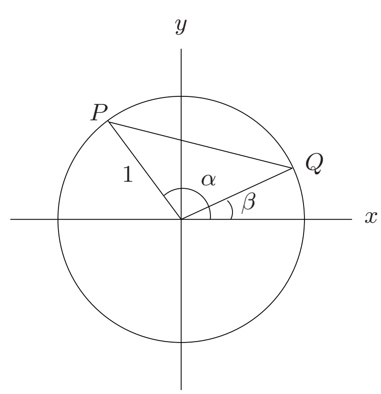
But $B \geq 0$ so that $AB - |C|^2 \geq 0$ and hence $|C|^2 \leq AB$. Taking the square root completes the proof. ∎

It is worth our going through some additional elementary properties of complex numbers that will be needed on occasion throughout this text. Purely for the sake of logical consistency, let us first prove some basic trigonometric relationships. Our starting point will be the so-called "**law of cosines**" which states that $c^2 = a^2 + b^2 - 2ab\cos\theta$ (see the figure below).



A special case of this occurs when $\theta = \pi/2$, in which case we obtain the famous **Pythagorean theorem** $a^2 + b^2 = c^2$. (While most readers should already be familiar with these results, we prove them in Section 1.5.)

Now consider a triangle inscribed in a *unit* circle as shown below:



The point P has coordinates $(x_P, y_P) = (\cos\alpha, \sin\alpha)$, and Q has coordinates $(x_Q, y_Q) = (\cos\beta, \sin\beta)$. Applying the Pythagorean theorem to the right triangle with hypotenuse defined by the points P and Q (and noting $x_Q^2 + y_Q^2 = x_P^2 + y_P^2 = 1$), we see that the square of the distance between the points P and Q is given by

$$\begin{aligned}(PQ)^2 &= (x_Q - x_P)^2 + (y_Q - y_P)^2 \\ &= (x_Q^2 + y_Q^2) + (x_P^2 + y_P^2) - 2(x_P x_Q + y_P y_Q) \\ &= 2 - 2(\cos\alpha\cos\beta + \sin\alpha\sin\beta).\end{aligned}$$

On the other hand, we can apply the law of cosines to obtain the distance PQ, in which case we find that $(PQ)^2 = 2 - 2\cos(\alpha - \beta)$. Equating these expressions yields the basic result

$$\cos(\alpha - \beta) = \cos\alpha\cos\beta + \sin\alpha\sin\beta.$$

Replacing β by $-\beta$ we obtain

$$\cos(\alpha + \beta) = \cos\alpha\cos\beta - \sin\alpha\sin\beta.$$

If we let $\alpha = \pi/2$, then we have $\cos(\pi/2 - \beta) = \sin\beta$, and if we now replace β by $\pi/2 - \beta$, we find that $\cos\beta = \sin(\pi/2 - \beta)$. Finally, we can use these last results to obtain formulas for $\sin(\alpha \pm \beta)$. In particular, we replace β by $\alpha + \beta$ to obtain

$$\begin{aligned}\sin(\alpha + \beta) &= \cos(\pi/2 - (\alpha + \beta)) \\ &= \cos(\pi/2 - \alpha - \beta) \\ &= \cos(\pi/2 - \alpha)\cos\beta + \sin(\pi/2 - \alpha)\sin\beta \\ &= \sin\alpha\cos\beta + \cos\alpha\sin\beta\end{aligned}$$

Again, replacing β by $-\beta$ yields

$$\sin(\alpha - \beta) = \sin\alpha\cos\beta - \cos\alpha\sin\beta.$$

(The reader may already know that these results are simple to derive using the Euler formula $\exp(\pm i\theta) = \cos\theta \pm i\sin\theta$ which follows from the Taylor series expansions of $\exp x$, $\sin x$ and $\cos x$, along with the definition $i^2 = -1$.)

It is often of great use to think of a complex number $z = x + iy$ as a point in the xy-plane. If we define

$$r = |z| = \sqrt{x^2 + y^2}$$

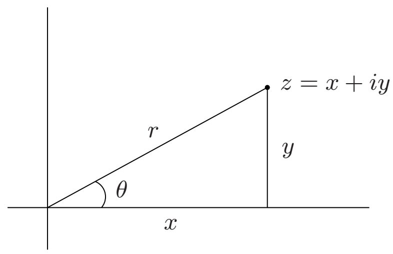
and

$$\tan\theta = y/x$$

then a complex number may also be written in the form

$$z = x + iy = r(\cos\theta + i\sin\theta) = r\exp(i\theta)$$

(see the figure below).



Given two complex numbers

$$z_1 = r_1(\cos\theta_1 + i\sin\theta_1)$$

and

$$z_2 = r_2(\cos\theta_2 + i\sin\theta_2)$$

we can use the trigonometric addition formulas derived above to show that

$$z_1 z_2 = r_1 r_2[\cos(\theta_1 + \theta_2) + i\sin(\theta_1 + \theta_2)].$$

In fact, by induction it should be clear that this can be generalized to (see Exercise A.6.5)

$$z_1 z_2 \cdots z_n = r_1 r_2 \cdots r_n[\cos(\theta_1 + \theta_2 + \cdots + \theta_n) + i\sin(\theta_1 + \theta_2 + \cdots + \theta_n)].$$

In the particular case where $z_1 = \cdots = z_n$, we find that

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

This is often called **De Moivre's theorem**.

One of the main uses of this theorem is as follows. Let w be a complex number, and let $z = w^n$ (where n is a positive integer). We say that w is an *nth* **root** of z, and we write this as $w = z^{1/n}$. Next, we observe from De Moivre's theorem that writing $z = r(\cos\theta + i\sin\theta)$ and $w = s(\cos\phi + i\sin\phi)$ yields (assuming that $z \neq 0$)

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

But $\cos\theta = \cos(\theta \pm 2k\pi)$ for $k = 0, \pm1, \pm2, \ldots$, and therefore $r = s^n$ and $n\phi = \theta \pm 2k\pi$. (This follows from the fact that if $z_1 = x_1 + iy_1 = r_1(\cos\theta_1 + i\sin\theta_1)$ and $z_2 = x_2 + iy_2 = r_2(\cos\theta_2 + i\sin\theta_2)$, then $z_1 = z_2$ implies $x_1 = x_2$ and $y_1 = y_2$ so that $r_1 = r_2$, and hence $\theta_1 = \theta_2$.) Then s is the real positive nth root of r, and $\phi = \theta/n \pm 2k\pi/n$. Since this expression for ϕ is the same if any two integers k differ by a multiple of n, we see that there are precisely n distinct solutions of $z = w^n$ (when $z \pm 0$), and these are given by

$$w = r^{1/n}[\cos(\theta + 2k\pi)/n + i \sin(\theta + 2k\pi)/n]$$

where $k = 0, 1, \ldots, n-1$.

Exercises

1. (a) Show
$$\binom{n}{k} = \binom{n}{n-k}.$$
   (b) Show
$$\binom{n+1}{k} = \binom{n}{k} + \binom{n}{k-1}.$$

2. Prove by induction the formula $1 + 2 + \cdots + n = n(n+1)/2$.

3. Prove by induction the formula
$$1 + x + x^2 + \cdots + x^{n-1} = \frac{1 - x^n}{1 - x}$$
   where x is any real number $\neq 1$.

4. Prove by induction that for any complex numbers $z_1, \ldots, z_n$ we have:

   (a)
$$|z_1 z_2 \cdots z_n| = |z_1| \, |z_2| \cdots |z_n| \, .$$
   (b)
$$\left| \sum_{i=1}^{n} z_i \right| \leq \sum_{i=1}^{n} |z_i| \, .$$
   (c)
$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} |z_i| \leq \left(\sum_{i=1}^{n} |z_i|^2 \right)^{1/2} \leq \sum_{i=1}^{n} |z_i| \, .$$
   [*Hint*: For the first half of this inequality you will need to show that $2\,|z_1|\,|z_2| \leq |z_1|^2 + |z_2|^2$.]

5. Prove by induction that for any complex numbers $z_1, \ldots, z_n$ we have
$$\begin{aligned} &z_1 z_2 \cdots z_n \\ &\qquad = r_1 r_2 \cdots r_n [\cos(\theta_1 + \theta_2 + \cdots + \theta_n) + i \sin(\theta_1 + \theta_2 + \cdots + \theta_n)] \end{aligned}$$
   where $z_j = r_j \exp(i\theta_j)$.

A.7 Additional Properties of the Integers

The material of this section is generalized to the theory of polynomials in Section 5.1. Most of what we cover here should be familiar to the reader from very elementary courses.

Our first topic is the division of an arbitrary integer $a \in \mathbb{Z}$ by a positive integer $b \in \mathbb{Z}^+$. For example, we can divide 11 by 4 to obtain $11 = 2 \cdot 4 + 3$. As another example, -7 divided by 2 yields $-7 = -4 \cdot 2 + 1$. Note that each of these examples may be written in the form $a = qb + r$ where $q \in \mathbb{Z}$ and $0 \leq r < b$. The number q is called the **quotient**, and the number r is called the **remainder** in the division of a by b. In the particular case that $r = 0$, we say that b **divides** a and we write this as $b \mid a$. If $r \neq 0$, then b does not divide a, and this is written as $b \nmid a$. If an integer $p \in \mathbb{Z}^+$ is not divisible by any positive integer other than 1 and p itself, then p is said to be **prime**.

It is probably worth pointing out the elementary fact that if $a \mid b$ and $a \mid c$, then $a \mid (mb + nc)$ for any $m, n \in \mathbb{Z}$. This is because $a \mid b$ implies $b = q_1 a$, and $a \mid c$ implies $c = q_2 a$. Thus $mb + nc = (mq_1 + nq_2)a$ so that $a \mid (mb + nc)$.

Theorem A.8 (**Division Algorithm**). *If $a \in \mathbb{Z}$ and $b \in \mathbb{Z}^+$, then there exist unique integers q and r such that $a = qb + r$ where $0 \leq r < b$.*

Proof. Define $S = \{a - nb \geq 0 : n \in \mathbb{Z}\}$. In other words, S consists of all nonnegative integers of the form $a - bn$. It is easy to see that $S \neq \varnothing$. Indeed, if $a \geq 0$ we simply choose $n = 0$ so that $a \in S$, and if $a < 0$ we choose $n = a$ so that $a - ba = a(1 - b) \in S$ (since $a < 0$ and $1 - b \leq 0$). Since S is a nonempty subset of the natural numbers, we may apply the well-ordering property of the natural numbers to conclude that S contains a least element $r \geq 0$. If we let q be the value of n corresponding to this r, then we have $a - qb = r$ or $a = qb + r$ where $0 \leq r$. We must show that $r < b$. To see this, suppose that $r \geq b$. Then

$$a - (q + 1)b = a - qb - b = r - b \geq 0$$

so that $a - (q + 1)b \in S$. But $b > 0$ so that

$$a - (q + 1)b = (a - qb) - b < a - qb = r$$

which contradicts the definition of r as the least element of S. Hence $r < b$.

To prove uniqueness, we suppose that we may write $a = q_1 b + r_1$ and $a = q_2 b + r_2$ where $0 \leq r_1 < b$ and $0 \leq r_2 < b$. Equating these two formulas yields $q_1 b + r_1 = q_2 b + r_2$ or $(q_1 - q_2)b = r_2 - r_1$, and therefore $b \mid (r_2 - r_1)$. Using the fact that $0 \leq r_1 < b$ and $0 \leq r_2 < b$, we see that $r_2 - r_1 < b - r_1 \leq b$. Similarly we have $r_1 - r_2 < b - r_2 \leq b$ or $-b < r_2 - r_1$. This means that $-b < r_2 - r_1 < b$. Therefore $r_2 - r_1$ is a multiple of b that lies strictly between $-b$ and b, and thus we must have $r_2 - r_1 = 0$. Then $(q_1 - q_2)b = 0$ with $b \neq 0$, and hence $q_1 - q_2 = 0$ also. This shows that $r_1 = r_2$ and $q_1 = q_2$ which completes the proof of uniqueness. ∎

Suppose we are given two integers $a, b \in \mathbb{Z}$ where we assume that a and b are not both zero. We say that an integer $d \in \mathbb{Z}^+$ is the **greatest common divisor** of a and b if $d \mid a$ and $d \mid b$, and if c is any other integer with the property that $c \mid a$ and $c \mid b$, then $c \mid d$. We denote the greatest common divisor of a and b by $\gcd\{a, b\}$. Our next theorem shows that the gcd always exists and is unique. Furthermore, the method of proof shows us how to actually compute the gcd.

Theorem A.9 (Euclidean Algorithm). *If $a, b \in \mathbb{Z}$ are not both zero, then there exists a unique positive integer $d \in \mathbb{Z}^+$ such that*

(i) $d \mid a$ and $d \mid b$.

(ii) If $c \in \mathbb{Z}$ is such that $c \mid a$ and $c \mid b$, then $c \mid d$.

Proof. First assume $b > 0$. Applying the division algorithm, there exist unique integers q_1 and r_1 such that

$$a = q_1 b + r_1 \qquad \text{with } 0 \leq r_1 < b.$$

If $r_1 = 0$, then $b \mid a$ and we may take $d = b$ to satisfy both parts of the theorem. If $r_1 \neq 0$, then we apply the division algorithm again to b and r_1, obtaining

$$b = q_2 r_1 + r_2 \qquad \text{with } 0 \leq r_2 < r_1.$$

Continuing this procedure, we obtain a sequence of nonzero remainders r_1, r_2, $\dots$, r_k where

$$\begin{aligned}
a &= q_1 b + r_1 && \text{with } 0 \leq r_1 < b \\
b &= q_2 r_1 + r_2 && \text{with } 0 \leq r_2 < r_1 \\
r_1 &= q_3 r_2 + r_3 && \text{with } 0 \leq r_3 < r_2 \\
&\vdots \\
r_{k-2} &= q_k r_{k-1} + r_k && \text{with } 0 \leq r_k < r_{k-1} \\
r_{k-1} &= q_{k+1} r_k
\end{aligned} \tag{*}$$

That this process must terminate with a zero remainder as shown is due to the fact that each remainder is a nonnegative integer with $r_1 > r_2 > \cdots$. We have denoted the last nonzero remainder by r_k.

We now claim that $d = r_k$. Since $r_{k-1} = q_{k+1} r_k$, we have $r_k \mid r_{k-1}$. Then, using $r_{k-2} = q_k r_{k-1} + r_k$ along with $r_k \mid r_k$ and $r_k \mid r_{k-1}$, we have $r_k \mid r_{k-2}$. Continuing in this manner, we see that $r_k \mid r_{k-1}, r_k \mid r_{k-2}, \dots, r_k \mid r_1, r_k \mid b$ and $r_k \mid a$. This shows that r_k is a common divisor of a and b. To show that r_k is in fact the greatest common divisor, we first note that if $c \mid a$ and $c \mid b$, then $c \mid r_1$ because $r_1 = a - q_1 b$. But now we see in the same way that $c \mid r_2$, and working our way through the above set of equations we eventually arrive at $c \mid r_k$. Thus r_k is a gcd as claimed.

If $b < 0$, we repeat the above process with a and $-b$ rather than a and b. Since b and $-b$ have the same divisors, it follows that a gcd of $\{a, -b\}$ will be a gcd of $\{a, b\}$. (Note we have not yet shown the uniqueness of the gcd.) If $b = 0$, then we can simply let $d = |a|$ to satisfy both statements in the theorem.

As to uniqueness of the gcd, suppose we have integers d_1 and d_2 that satisfy both statements of the theorem. Then applying statement (ii) to both d_1 and d_2, we must have $d_1 \mid d_2$ and $d_2 \mid d_1$. But both d_1 and d_2 are positive, and hence $d_1 = d_2$. ∎

Corollary. *If $d = \gcd\{a, b\}$ where a and b are not both zero, then $d = am + bn$ for some $m, n \in \mathbb{Z}$.*

Proof. Referring to equations (*) in the proof of Theorem A.9, we note that the equation for $r_k - 2$ may be solved for r_k to obtain $r_k = r_{k-2} - r_{k-1}q_k$. Next, the equation $r_{k-3} = q_{k-1}r_{k-2} + r_{k-1}$ may be solved for r_{k-1}, and this is then substituted into the previous equation to obtain $r_k = r_{k-2}(1+q_{k-1}q_k) - r_{k-3}q_k$. Working our way up equations (*), we next eliminate r_{k-2} to obtain r_k in terms of r_{k-3} and r_{k-4}. Continuing in this manner, we eventually obtain r_k in terms of b and a. ∎

If $a, b \in \mathbb{Z}$ and $\gcd\{a, b\} = 1$, then we say that a and b are **relatively prime** (or sometimes **coprime**). The last result on integers that we wish to prove is the result that if p is prime and $p \mid ab$ (where $a, b \in \mathbb{Z}$), then either $p \mid a$ or $p \mid b$.

Theorem A.10. *(i) Suppose $a, b, c \in \mathbb{Z}$ where $a \mid bc$ and a and b are relatively prime. Then $a \mid c$.*

(ii) If p is prime and $a_1, \dots, a_n \in \mathbb{Z}$ with $p \mid a_1 \cdots a_n$, then $p \mid a_i$ for some $i = 1, \dots, n$.

Proof. (i) By the corollary to Theorem A.9 we have $\gcd\{a, b\} = 1 = am + bn$ for some $m, n \in \mathbb{Z}$. Multiplying this equation by c we obtain $c = amc + bnc$. But $a \mid bc$ by hypothesis so clearly $a \mid bnc$. Since it is also obvious that $a \mid amc$, we see that $a \mid c$.

(ii) We proceed by induction on n, the case $n = 1$ being trivial. We therefore assume that $n > 1$ and $p \mid a_1 \cdots a_n$. If $p \mid a_1 \cdots a_{n-1}$, then $p \mid a_i$ for some $i = 1, \dots, n-1$ by our induction hypothesis. On the other hand, if $p \nmid a_1 \cdots a_{n-1}$ then $\gcd\{p, a_1, \dots, a_{n-1}\} = 1$ since p is prime. We then apply part (i) with $a = p$, $b = a_1 \cdots a_{n-1}$ and $c = a_n$ to conclude that $p \mid a_n$. ∎

Exercises

1. Find the gcd of the following sets of integers:

(a) $\{6, 14\}$.
(b) $\{-75, 105\}$.
(c) $\{14, 21, 35\}$.

2. Find the gcd of each set and write it in terms of the given integers:

   (a) $\{1001, 33\}$.
   (b) $\{-90, 1386\}$.
   (c) $\{-2860, -2310\}$.

3. Suppose p is prime and $p \nmid a$ where $a \in \mathbb{Z}$. Prove that a and p are relatively prime.

4. Prove that if $\gcd\{a, b\} = 1$ and $c \mod a$, then $\gcd\{b, c\} = 1$.

5. If $a, b \in \mathbb{Z}^+$, then $m \in \mathbb{Z}^+$ is called the **least common multiple** (abbreviated lcm) if $a \mid m$ and $b \mid m$, and if $c \in \mathbb{Z}$ is such that $a \mid c$ and $b \mid c$, then $m \mid c$. Suppose $a = {p_1}^{s_1} \cdots {p_k}^{s_k}$ and $b = {p_1}^{t_1} \cdots {p_k}^{t_k}$ where $p_1, \ldots, p_k$ are distinct primes and each s_i and t_i are ≥ 0.

   (a) Prove that $a \mid b$ if and only if $s_i \leq t_i$ for all $i = 1, \ldots, k$.
   (b) For each $i = 1, \ldots, k$ let $u_i = \min\{s_i, t_i\}$ and $v_i = \max\{s_i, t_i\}$. Prove that $\gcd\{a, b\} = {p_1}^{u_1} \cdots {p_k}^{u_k}$ and $\operatorname{lcm}\{a, b\} = {p_1}^{v_1} \cdots {p_k}^{v_k}$.

6. Prove the **Fundamental Theorem of Arithmetic**: Every integer > 1 can be written as a unique (except for order) product of primes. Here is an outline of the proof:

   (a) Let

   $$S = \{a \in \mathbb{Z} : a > 1 \text{ and } a \text{ can not be written as a product of primes.}\}$$

   (In particular, note that S contains no primes.) Show that $S = \varnothing$ by assuming the contrary and using the well-ordered property of the natural numbers.

   (b) To prove uniqueness, assume that $n > 1$ is an integer that has two different expansions as $n = p_1 \cdots p_s = q_1 \cdots q_t$ where all the p_i and q_j are prime. Show that $p_1 \mid q_j$ for some $j = 1, \ldots, t$ and hence that $p_1 = q_j$. Thus p_1 and q_j can be canceled from each side of the equation. Continue this argument to cancel one p_i with one q_j, and then finally concluding that $s = t$.

Bibliography

[1] Abraham, R., Marsden, J. E. and Ratiu, T., *Manifolds, Tensor Analysis, and Applications,* Addison-Wesley, Reading, MA, 1983.

[2] Adler, R., Bazin, M. and Schiffer, M., *Introduction to General Relativity,* McGraw-Hill, New York, 1965.

[3] Arnold, V. I., *Mathematical Methods of Classical Mechanics,* Springer-Verlag, New York, 1978.

[4] Biggs, N. L., *Discrete Mathematics,* Clarendon Press, Oxford, 1985.

[5] Bishop, R. L. and Goldberg, S. I., *Tensor Analysis on Manifolds,* Macmillan, New York, 1968.

[6] Boothby, W. M., *An Introduction to Differentiable Manifolds and Riemannian Geometry,* Academic Press, New York, 1975.

[7] Byron, F. W. Jr., and Fuller, R. W., *Mathematics of Classical and Quantum Physics,* Addison-Wesley, Reading, MA, 1969.

[8] Curtis, C. W., *Linear Algebra,* Springer-Verlag, New York, 1984.

[9] Curtis, W. D. and Miller, F. R., *Differential Manifolds and Theoretical Physics,* Academic Press, Orlando, FL, 1985.

[10] Dennery, P. and Krzywicki, A., *Mathematics for Physicists,* Harper and Row, New York, 1967.

[11] Durbin, J. R., *Modern Algebra,* John Wiley and Sons, New York, 1985.

[12] Fetter, Alexander L. and Walecka, John Dirk, *Theoretical Mechanics of Particles and Continua* Dover Publications, Mineola, NY, 2003

[13] Flanders, H., *Differential Forms,* Academic Press, New York, 1963.

[14] Frankel, T., *The Geometry of Physics,* 2nd edition, Cambridge University Press, New York, NY, 2004.

[15] Frankel, T., *Linear Algebra,* unpublished lecture notes, University of California, San Diego, 1980.

[16] Friedberg, S. H. and Insel, A. J., *Introduction to Linear Algebra with Applications,* Prentice-Hall, Englewood Cliffs, NJ, 1986.

[17] Friedberg, S. H., Insel, A. J. and Spence, L. E., *Linear Algebra,* Prentice-Hall, Englewood Cliffs, NJ, 1979.

[18] Gemignani, M. C., *Elementary Topology,* 2nd edition, Addison-Wesley, Reading, MA, 1972.

[19] Geroch, R., *Mathematical Physics,* University of Chicago Press, Chicago, IL, 1985.

[20] Goldstein, H., *Classical Mechanics,* 2nd edition, Addison-Wesley, Reading, MA, 1980.

[21] Halmos, P. R., *Finite-Dimensional Vector Spaces,* Springer-Verlag, New York, 1974.

[22] Herstein, I. N., *Topics in Algebra,* Xerox College Publishing, Lexington, MA, 1964.

[23] Hoffman, K. and Kunze, R., *Linear Algebra,* 2nd edition, Prentice-Hall, Englewood Cliffs, NJ, 1971.

[24] Jackson, J. D., *Classical Electrodynamics,* 2nd edition, John Wiley and Sons, New York, 1975.

[25] Johnson, R. E., *Linear Algebra,* Prindle, Weber and Schmidt, Boston, MA, 1967.

[26] Knopp, K., *Infinite Sequences and Series,* Dover Publications, New York, 1956.

[27] Kolmogorov, A. N. and Fomin, S. V., *Functional Analysis,* Graylock Press, Rochester, NY, 1957.

[28] Lang, S., *Analysis I,* Addison-Wesley, Reading, MA, 1968.

[29] Lang, S., *Linear Algebra,* 2nd edition, Addison-Wesley, Reading, MA, 1971.

[30] Lang, S., *Real Analysis,* 2nd edition, Addison-Wesley, Reading, MA, 1983.

[31] Lipschutz, S., *Linear Algebra,* Schaums Outline Series, McGraw-Hill, New York, 1968.

[32] Marcus, M. and Minc, H., *A Survey of Matrix Theory and Matrix Inequalities,* Allyn and Bacon, Boston, MA, 1964.

[33] Marcus, M., *Finite-Dimensional Multilinear Algebra, Parts I and II,* Marcel Dekker, New York, 1973.

[34] Marcus, M., *Introduction to Linear Algebra,* Dover Publications, New York, 1988.

[35] Marion, J. B., *Classical Dynamics of Particles and Systems,* 2nd edition, Academic Press, Orlando, FL, 1970.

[36] Marsden, J. E., *Elementary Classical Analysis,* W. H. Freeman, San Francisco, CA, 1974.

[37] Misner, C. W., Thorne, K. S. and Wheeler, J. A., *Gravitation,* W. H. Freeman, San Francisco, CA, 1973.

[38] Munkres, J. R., *Topology,* Prentice-Hall, Englewood Cliffs, NJ, 1975.

[39] Murdoch, D. C., *Linear Algebra,* John Wiley and Sons, New York, 1970.

[40] Prugovecki, E., *Quantum Mechanics in Hilbert Space,* 2nd edition, Academic Press, New York, 1981.

[41] Reed, M. and Simon, B., *Functional Analysis,* Revised and Enlarged edition, Academic Press, Orlando, FL, 1980.

[42] Roach, G. F., *Greens Functions,* 2nd edition, Cambridge University Press, Cambridge, 1982.

[43] Royden, H. L., *Real Analysis,* 2nd edition, Macmillan, New York, 1968.

[44] Rudin, W., *Functional Analysis,* 2nd edition, McGraw-Hill, New York, 1974.

[45] Rudin, W., *Principles of Mathematical Analysis,* 3rd edition, McGraw-Hill, New York, 1976.

[46] Ryder, L. H., *Quantum Field Theory,* Cambridge University Press, Cambridge, 1985.

[47] Schutz, B., *Geometrical Methods of Mathematical Physics,* Cambridge University Press, Cambridge, 1980.

[48] Shankar, R., *Principles of Quantum Mechanics,* Plenum Press, New York, 1980.

[49] Simmons, G. F., *Introduction to Topology and Modern Analysis,* McGraw-Hill, New York, 1963.

[50] Talman, J. D., *Special Functions,* W. A. Benjamin, New York, 1968.

[51] Taylor, J. R., *Scattering Theory,* John Wiley and Sons, New York, 1972.

[52] Tung, W., *Group Theory in Physics,* World Scientific, Philadelphia, PA, 1985.

Index