Notes on Linear Algebra

Peter J. Cameron

Preface

Linear algebra has two aspects. Abstractly, it is the study of vector spaces over fields, and their linear maps and bilinear forms. Concretely, it is matrix theory: matrices occur in all parts of mathematics and its applications, and everyone working in the mathematical sciences and related areas needs to be able to diagonalise a real symmetric matrix. So in a course of this kind, it is necessary to touch on both the abstract and the concrete aspects, though applications are not treated in detail.

On the theoretical side, we deal with vector spaces, linear maps, and bilinear forms. Vector spaces over a field $\mathbb{K}$ are particularly attractive algebraic objects, since each vector space is completely determined by a single number, its dimension (unlike groups, for example, whose structure is much more complicated). Linear maps are the structure-preserving maps or homomorphisms of vector spaces.

On the practical side, the subject is really about one thing: matrices. If we need to do some calculation with a linear map or a bilinear form, we must represent it by a matrix. As this suggests, matrices represent several different kinds of things. In each case, the representation is not unique, since we have the freedom to change bases in our vector spaces; so many different matrices represent the same object. This gives rise to several equivalence relations on the set of matrices, summarised in the following table:

Equivalence	Similarity	Congruence	Orthogonal similarity
Same linear map $\alpha : V \to W$	Same linear map $\alpha : V \to V$	Same bilinear form b on V	Same self-adjoint $\alpha : V \to V$ w.r.t. orthonormal basis
$A' = Q^{-1}AP$ P, Q invertible	$A' = P^{-1}AP$ P invertible	$A' = P^{\top}AP$ P invertible	$A' = P^{-1}AP$ P orthogonal

The power of linear algebra in practice stems from the fact that we can choose bases so as to simplify the form of the matrix representing the object in question. We will see several such "canonical form theorems" in the notes.

These lecture notes correspond to the course Linear Algebra II, as given at Queen Mary, University of London, in the first sememster 2005–6.

The course description reads as follows:

> This module is a mixture of abstract theory, with rigorous proofs, and concrete calculations with matrices. The abstract component builds on the notions of subspaces and linear maps to construct the theory of bilinear forms i.e. functions of two variables which are linear in each variable, dual spaces (which consist of linear mappings from the original space to the underlying field) and determinants. The concrete applications involve ways to reduce a matrix of some specific type (such as symmetric or skew-symmetric) to as near diagonal form as possible.

In other words, students on this course have met the basic concepts of linear algebra before. Of course, some revision is necessary, and I have tried to make the notes reasonably self-contained. If you are reading them without the benefit of a previous course on linear algebra, you will almost certainly have to do some work filling in the details of arguments which are outlined or skipped over here.

The notes for the prerequisite course, Linear Algebra I, by Dr Francis Wright, are currently available from

`http://centaur.maths.qmul.ac.uk/Lin_Alg_I/`

I have by-and-large kept to the notation of these notes. For example, a general field is called $\mathbb{K}$, vectors are represented as column vectors, linear maps (apart from zero and the identity) are represented by Greek letters.

I have included in the appendices some extra-curricular applications of linear algebra, including some special determinants, the method for solving a cubic equation, the proof of the "Friendship Theorem" and the problem of deciding the winner of a football league, as well as some worked examples.

Peter J. Cameron
September 5, 2008

Contents

Chapter 1

Vector spaces

These notes are about linear maps and bilinear forms on vector spaces, how we represent them by matrices, how we manipulate them, and what we use this for.

1.1 Definitions

Definition 1.1 A *field* is an algebraic system consisting of a non-empty set $\mathbb{K}$ equipped with two binary operations $+$ (addition) and $\cdot$ (multiplication) satisfying the conditions:

(A) $(\mathbb{K}, +)$ is an abelian group with identity element 0 (called *zero*);

(M) $(\mathbb{K} \setminus \{0\}, \cdot)$ is an abelian group with identity element 1;

(D) the *distributive law*

$$a(b+c) = ab + ac$$

holds for all $a, b, c \in \mathbb{K}$.

If you don't know what an abelian group is, then you can find it spelled out in detail in Appendix A. In fact, the only fields that I will use in these notes are

- $\mathbb{Q}$, the field of rational numbers;
- $\mathbb{R}$, the field of real numbers;
- $\mathbb{C}$, the field of complex numbers;
- $\mathbb{F}_p$, the field of integers mod p, where p is a prime number.

I will not stop to prove that these structures really are fields. You may have seen $\mathbb{F}_p$ referred to as $\mathbb{Z}_p$.

Definition 1.2 A *vector space* V over a field $\mathbb{K}$ is an algebraic system consisting of a non-empty set V equipped with a binary operation $+$ (vector addition), and an operation of scalar multiplication

$$(a,v) \in \mathbb{K} \times V \mapsto av \in V$$

such that the following rules hold:

(VA) $(V,+)$ is an abelian group, with identity element 0 (the *zero vector*).

(VM) Rules for scalar multiplication:

- (VM0) For any $a \in \mathbb{K}$, $v \in V$, there is a unique element $av \in V$.
- (VM1) For any $a \in \mathbb{K}$, $u,v \in V$, we have $a(u+v) = au+av$.
- (VM2) For any $a,b \in \mathbb{K}$, $v \in V$, we have $(a+b)v = av+bv$.
- (VM3) For any $a,b \in \mathbb{K}$, $v \in V$, we have $(ab)v = a(bv)$.
- (VM4) For any $v \in V$, we have $1v = v$ (where 1 is the identity element of $\mathbb{K}$).

Since we have two kinds of elements, namely elements of $\mathbb{K}$ and elements of V, we distinguish them by calling the elements of $\mathbb{K}$ *scalars* and the elements of V *vectors*.

A vector space over the field $\mathbb{R}$ is often called a *real vector space*, and one over $\mathbb{C}$ is a *complex vector space*.

Example 1.1 The first example of a vector space that we meet is the *Euclidean plane* $\mathbb{R}^2$. This is a real vector space. This means that we can add two vectors, and multiply a vector by a scalar (a real number). There are two ways we can make these definitions.

- The *geometric* definition. Think of a vector as an arrow starting at the origin and ending at a point of the plane. Then addition of two vectors is done by the *parallelogram law* (see Figure 1.1). The scalar multiple av is the vector whose length is $|a|$ times the length of v, in the same direction if $a > 0$ and in the opposite direction if $a < 0$.

- The *algebraic* definition. We represent the points of the plane by Cartesian coordinates (x,y). Thus, a vector v is just a pair (x,y) of real numbers. Now we define addition and scalar multiplication by

$$\begin{aligned} (x_1,y_1)+(x_2,y_2) &= (x_1+x_2,y_1+y_2), \\ a(x,y) &= (ax,ay). \end{aligned}$$

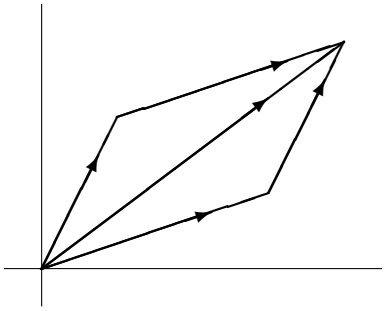


Figure 1.1: The parallelogram law

Not only is this definition much simpler, but it is much easier to check that the rules for a vector space are really satisfied! For example, we check the law $a(v+w) = av+aw$. Let $v = (x_1, y_1)$ and $w = (x_2, y_2)$. Then we have

$$\begin{aligned} a(v+w) &= a((x_1,y_1)+(x_2,y_2)) \\ &= a(x_1+x_2, y_1+y_2) \\ &= (ax_1+ax_2, ay_1+ay_2) \\ &= (ax_1, ay_1)+(ax_2, ay_2) \\ &= av+aw. \end{aligned}$$

In the algebraic definition, we say that the operations of addition and scalar multiplication are *coordinatewise*: this means that we add two vectors coordinate by coordinate, and similarly for scalar multiplication.

Using coordinates, this example can be generalised.

Example 1.2 Let n be any positive integer and $\mathbb{K}$ any field. Let $V = \mathbb{K}^n$, the set of all n-tuples of elements of $\mathbb{K}$. Then V is a vector space over $\mathbb{K}$, where the operations are defined coordinatewise:

$$\begin{aligned} (a_1, a_2, \ldots, a_n)+(b_1, b_2, \ldots, b_n) &= (a_1+b_1, a_2+b_2, \ldots, a_n+b_n), \\ c(a_1, a_2, \ldots, a_n) &= (ca_1, ca_2, \ldots, ca_n). \end{aligned}$$

1.2 Bases

This example is much more general than it appears: *Every finite-dimensional vector space looks like Example 1.2.* Here's why.

Definition 1.3 Let V be a vector space over the field $\mathbb{K}$, and let $v_1, \ldots, v_n$ be vectors in V.

(a) The vectors $v_1, v_2, \ldots, v_n$ are *linearly independent* if, whenever we have scalars $c_1, c_2, \ldots, c_n$ satisfying

$$c_1 v_1 + c_2 v_2 + \cdots + c_n v_n = 0,$$

then necessarily $c_1 = c_2 = \cdots = 0$.

(b) The vectors $v_1, v_2, \ldots, v_n$ are *spanning* if, for every vector $v \in V$, we can find scalars $c_1, c_2, \ldots, c_n \in \mathbb{K}$ such that

$$v = c_1 v_1 + c_2 v_2 + \cdots + c_n v_n.$$

In this case, we write $V = \langle v_1, v_2, \ldots, v_n \rangle$.

(c) The vectors $v_1, v_2, \ldots, v_n$ form a *basis* for V if they are linearly independent and spanning.

Remark Linear independence is a property of a *list* of vectors. A list containing the zero vector is never linearly independent. Also, a list in which the same vector occurs more than once is never linearly independent.

I will say "Let $B = (v_1, \ldots, v_n)$ be a basis for V" to mean that the list of vectors $v_1, \ldots, v_n$ is a basis, and to refer to this list as B.

Definition 1.4 Let V be a vector space over the field $\mathbb{K}$. We say that V is *finite-dimensional* if we can find vectors $v_1, v_2, \ldots, v_n \in V$ which form a basis for V.

Remark In these notes we are only concerned with finite-dimensional vector spaces. If you study Functional Analysis, Quantum Mechanics, or various other subjects, you will meet vector spaces which are not finite dimensional.

Proposition 1.1 *The following three conditions are equivalent for the vectors $v_1, \ldots, v_n$ of the vector space V over $\mathbb{K}$:*

(a) $v_1, \ldots, v_n$ is a basis;

(b) $v_1, \ldots, v_n$ is a maximal linearly independent set (that is, if we add any vector to the list, then the result is no longer linearly independent);

(c) $v_1, \ldots, v_n$ is a minimal spanning set (that is, if we remove any vector from the list, then the result is no longer spanning).

The next theorem helps us to understand the properties of linear independence.

Theorem 1.2 (The Exchange Lemma) *Let V be a vector space over $\mathbb{K}$. Suppose that the vectors $v_1,\ldots,v_n$ are linearly independent, and that the vectors $w_1,\ldots,w_m$ are linearly independent, where $m>n$. Then we can find a number i with $1\leq i\leq m$ such that the vectors $v_1,\ldots,v_n,w_i$ are linearly independent.*

In order to prove this, we need a lemma about systems of equations.

Lemma 1.3 *Given a system $(*)$*

$$\begin{aligned} a_{11}x_1+a_{12}x_2+\cdots+a_{1m}x_m &= 0,\\ a_{21}x_1+a_{22}x_2+\cdots+a_{2m}x_m &= 0,\\ &\cdots\\ a_{n1}x_1+a_{n2}x_2+\cdots+a_{nm}x_m &= 0 \end{aligned}$$

of homogeneous linear equations, where the number n of equations is strictly less than the number m of variables, there exists a non-zero solution $(x_1,\ldots,x_m)$ (that is, $x_1,\ldots,x_m$ are not all zero).

Proof This is proved by induction on the number of variables. If the coefficients $a_{11},a_{21},\ldots,a_{n1}$ of x_1 are all zero, then putting $x_1=1$ and the other variables zero gives a solution. If one of these coefficients is non-zero, then we can use the corresponding equation to express x_1 in terms of the other variables, obtaining $n-1$ equations in $m-1$ variables. By hypothesis, $n-1<m-1$. So by the induction hypothesis, these new equations have a non-zero solution. Computing the value of x_1 gives a solution to the original equations.

Now we turn to the proof of the Exchange Lemma. Let us argue for a contradiction, by assuming that the result is false: that is, assume that none of the vectors w_i can be added to the list $(v_1,\ldots,v_n)$ to produce a larger linearly independent list. This means that, for all j, the list $(v_1,\ldots,v_n,w_i)$ is linearly dependent. So there are coefficients $c_1,\ldots,c_n,d$, not all zero, such that

$$c_1v_1+\cdots+c_nv_n+dw_i=0.$$

We cannot have $d=0$; for this would mean that we had a linear combination of $v_1,\ldots,v_n$ equal to zero, contrary to the hypothesis that these vectors are linearly independent. So we can divide the equation through by d, and take w_i to the other side, to obtain (changing notation slightly)

$$w_i=a_{1i}v_1+a_{2i}v_2+\cdots+a_{ni}v_n=\sum_{j=1}^{n}a_{ji}v_j.$$

We do this for each value of $i = 1, \ldots, m$.

Now take a non-zero solution to the set of equations $(*)$ above: that is,

$$\sum_{i=1}^{m} a_{ji} x_i = 0$$

for $j = 1, \ldots, n$.

Multiplying the formula for w_i by x_i and adding, we obtain

$$x_1 w_1 + \cdots + x_m w_m = \sum_{j=1}^{n} \left(\sum_{i=1}^{m} a_{ji} x_i \right) v_j = 0.$$

But the coefficients are not all zero, so this means that the vectors $(w_1, \ldots, w_m)$ are not linearly dependent, contrary to hypothesis.

So the assumption that no w_i can be added to $(v_1, \ldots, v_n)$ to get a linearly independent set must be wrong, and the proof is complete.

The Exchange Lemma has some important consequences:

Corollary 1.4 *Let V be a finite-dimensional vector space over a field $\mathbb{K}$. Then*

(a) any two bases of V have the same number of elements;

(b) any linearly independent set can be extended to a basis.

The number of elements in a basis is called the *dimension* of the vector space V. We will say "an n-dimensional vector space" instead of "a finite-dimensional vector space whose dimension is n". We denote the dimension of V by $\dim(V)$.

Proof Let us see how the corollary follows from the Exchange Lemma.

(a) Let $(v_1, \ldots, v_n)$ and $(w_1, \ldots, w_m)$ be two bases for V. Suppose, for a contradiction, that they have different numbers of elements; say that $n < m$, without loss of generality. Both lists of vectors are linearly independent; so, according to the Exchange Lemma, we can add some vector w_i to the first list to get a larger linearly independent list. This means that $v_1, \ldots, v_n$ was not a maximal linearly independent set, and so (by Proposition 1.1) not a basis, contradicting our assumption. We conclude that $m = n$, as required.

(b) Let $(v_1, \ldots, v_n)$ be linearly independent and let $(w_1, \ldots, w_m)$ be a basis. Necessarily $n \leq m$, since otherwise we could add one of the vs to $(1, \ldots, w_m)$ to get a larger linearly independent set, contradicting maximality. But now we can add some ws to $(v_1, \ldots, v_n)$ until we obtain a basis.

Remark We allow the possibility that a vector space has dimension zero. Such a vector space contains just one vector, the zero vector 0; a basis for this vector space consists of the empty set.

Now let V be an n-dimensional vector space over $\mathbb{K}$. This means that there is a basis $v_1, v_2, \ldots, v_n$ for V. Since this list of vectors is spanning, every vector $v \in V$ can be expressed as

$$v = c_1 v_1 + c_2 v_2 + \cdots + c_n v_n$$

for some scalars $c_1, c_2, \ldots, c_n \in \mathbb{K}$. The scalars $c_1, \ldots, c_n$ are the *coordinates* of v (with respect to the given basis), and the *coordinate representation* of v is the n-tuple

$$(c_1, c_2, \ldots, c_n) \in \mathbb{K}^n.$$

Now *the coordinate representation is unique*. For suppose that we also had

$$v = c'_1 v_1 + c'_2 v_2 + \cdots + c'_n v_n$$

for scalars $c'_1, c'_2 \ldots, c'_n$. Subtracting these two expressions, we obtain

$$0 = (c_1 - c'_1)v_1 + (c_2 - c'_2)v_2 + \cdots + (c_n - c'_n)v_n.$$

Now the vectors $v_1, v_2 \ldots, v_n$ are linearly independent; so this equation implies that $c_1 - c'_1 = 0$, $c_2 - c'_2 = 0$, $\ldots$, $c_n - c'_n = 0$; that is,

$$c_1 = c'_1, \quad c_2 = c'_2, \quad \ldots \quad c_n = c'_n.$$

Now it is easy to check that, when we add two vectors in V, we add their coordinate representations in $\mathbb{K}^n$ (using coordinatewise addition); and when we multiply a vector $v \in V$ by a scalar c, we multiply its coordinate representation by c. In other words, addition and scalar multiplication in V translate to the same operations on their coordinate representations. This is why we only need to consider vector spaces of the form $\mathbb{K}^n$, as in Example 1.2.

Here is how the result would be stated in the language of abstract algebra:

Theorem 1.5 *Any n-dimensional vector space over a field $\mathbb{K}$ is isomorphic to the vector space $\mathbb{K}^n$.*

1.3 Row and column vectors

The elements of the vector space $\mathbb{K}^n$ are all the n-tuples of scalars from the field $\mathbb{K}$. There are two different ways that we can represent an n-tuple: as a row, or as

a column. Thus, the vector with components 1, 2 and -3 can be represented as a *row vector*

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

or as a *column vector*

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

(Note that we use square brackets, rather than round brackets or parentheses. But you will see the notation $(1,2,-3)$ and the equivalent for columns in other books!)

Both systems are in common use, and you should be familiar with both. The choice of row or column vectors makes some technical differences in the statements of the theorems, so care is needed.

There are arguments for and against both systems. Those who prefer row vectors would argue that we already use (x,y) or (x,y,z) for the coordinates of a point in 2- or 3-dimensional Euclidean space, so we should use the same for vectors. The most powerful argument will appear when we consider representing linear maps by matrices.

Those who prefer column vectors point to the convenience of representing, say, the linear equations

$$\begin{aligned} 2x+3y &= 5, \\ 4x+5y &= 9 \end{aligned}$$

in matrix form

$$\begin{bmatrix} 2 & 3 \\ 4 & 5 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 5 \\ 9 \end{bmatrix}.$$

Statisticians also prefer column vectors: to a statistician, a vector often represents data from an experiment, and data are usually recorded in columns on a datasheet.

I will use column vectors in these notes. So we make a formal definition:

Definition 1.5 Let V be a vector space with a basis $B = (v_1, v_2, \ldots, v_n)$. If $v = c_1v_1 + c_2v_2 + \cdots + c_nv_n$, then the *coordinate representation* of v relative to the basis B is

$$[v]_B = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix}.$$

In order to save space on the paper, we often write this as

$$[v]_B = \begin{bmatrix} c_1 & c_2 & \ldots & v_n \end{bmatrix}^\top.$$

The symbol $\top$ is read "transpose".

1.4 Change of basis

The coordinate representation of a vector is always relative to a basis. We now have to look at how the representation changes when we use a different basis.

Definition 1.6 Let $B = (v_1, \ldots, v_n)$ and $B' = (v'_1, \ldots, v'_n)$ be bases for the n-dimensional vector space V over the field $\mathbb{K}$. The *transitition matrix* P from B to B' is the $n \times n$ matrix whose jth column is the coordinate representation $[v'_j]_B$ of the jth vector of B' relative to B. If we need to specify the bases, we write $P_{B,B'}$.

Proposition 1.6 *Let B and B' be bases for the n-dimensional vector space V over the field $\mathbb{K}$. Then, for any vector $v \in V$, the coordinate representations of v with respect to B and B' are related by*

$$[v]_B = P\,[v]_{B'}.$$

Proof Let p_{ij} be the i, j entry of the matrix P. By definition, we have

$$v'_j = \sum_{i=1}^{n} p_{ij} v_i.$$

Take an arbitrary vector $v \in V$, and let

$$[v]_B = [c_1, \ldots, c_n]^\top, \qquad [v]_{B'} = [d_1, \ldots, d_n]^\top.$$

This means, by definition, that

$$v = \sum_{i=1}^{n} c_i v_i = \sum_{j=1}^{n} d_j v'_j.$$

Substituting the formula for v'_j into the second equation, we have

$$v = \sum_{j=1}^{n} d_j \left(\sum_{i=1}^{n} p_{ij} v_i \right).$$

Reversing the order of summation, we get

$$v = \sum_{i=1}^{n} \left(\sum_{j=1}^{n} p_{ij} d_j \right) v_i.$$

Now we have two expressions for v as a linear combination of the vectors v_i. By the uniqueness of the coordinate representation, they are the same: that is,

$$c_i = \sum_{j=1}^{n} p_{ij} d_j.$$

In matrix form, this says

$$\begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix} = P \begin{bmatrix} d_1 \\ \vdots \\ d_n \end{bmatrix},$$

or in other words

$$[v]_B = P\,[v]_{B'},$$

as required.

In this course, we will see four ways in which matrices arise in linear algebra. Here is the first occurrence: **matrices arise as transition matrices between bases of a vector space.**

The next corollary summarises how transition matrices behave. Here I denotes the *identity matrix*, the matrix having 1s on the main diagonal and 0s everywhere else. Given a matrix P, we denote by P^{-1} the *inverse* of P, the matrix Q satisfying $PQ = QP = I$. Not every matrix has an inverse: we say that P is *invertible* or *non-singular* if it has an inverse.

Corollary 1.7 *Let B, B', B'' be bases of the vector space V.*

(a) $P_{B,B} = I$.

(b) $P_{B',B} = (P_{B,B'})^{-1}$.

(c) $P_{B,B''} = P_{B,B'}P_{B',B''}$.

This follows from the preceding Proposition. For example, for (b) we have

$$[v]_B = P_{B,B'}\,[v]_{B'}, \qquad [v]_{B'} = P_{B',B}\,[v]_B,$$

so

$$[v]_B = P_{B,B'}P_{B',B}\,[v]_B.$$

By the uniqueness of the coordinate representation, we have $P_{B,B'}P_{B',B} = I$.

Corollary 1.8 *The transition matrix between any two bases of a vector space is invertible.*

This follows immediately from (b) of the preceding Corollary.

Remark We see that, to express the coordinate representation w.r.t. the new basis in terms of that w.r.t. the old one, we need the inverse of the transition matrix:

$$[v]_{B'} = P_{B,B'}^{-1}[v]_B.$$

Example Consider the vector space $\mathbb{R}^2$, with the two bases

$$B = \left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right), \qquad B' = \left(\begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 3 \end{bmatrix} \right).$$

The transition matrix is

$$P_{B,B'} = \begin{bmatrix} 1 & 2 \\ 1 & 3 \end{bmatrix},$$

whose inverse is calculated to be

$$P_{B',B} = \begin{bmatrix} 3 & -2 \\ -1 & 1 \end{bmatrix}.$$

So the theorem tells us that, for any $x, y \in \mathbb{R}$, we have

$$\begin{bmatrix} x \\ y \end{bmatrix} = x \begin{bmatrix} 1 \\ 0 \end{bmatrix} + y \begin{bmatrix} 0 \\ 1 \end{bmatrix} = (3x - 2y) \begin{bmatrix} 1 \\ 1 \end{bmatrix} + (-x + y) \begin{bmatrix} 2 \\ 3 \end{bmatrix},$$

as is easily checked.

1.5 Subspaces and direct sums

Definition 1.7 A non-empty subset of a vector space is called a *subspace* if it contains the sum of any two of its elements and any scalar multiple of any of its elements. We write $U \leq V$ to mean "U is a subspace of V".

A subspace of a vector space is a vector space in its own right.
Subspaces can be constructed in various ways:

(a) Let $v_1, \ldots, v_n \in V$. The *span* of $(v_1, \ldots, v_n)$ is the set

$$\{c_1 v_1 + c_2 v_2 + \cdots + c_n v_n : c_1, \ldots, c_n \in \mathbb{K}\}.$$

This is a subspace of V. Moreover, $(v_1, \ldots, v_n)$ is a spanning set in this subspace. We denote the span of $v_1, \ldots, v_n$ by $\langle v_1, \ldots, v_n \rangle$.

(b) Let U_1 and U_2 be subspaces of V. Then

- the *intersection* $U_1 \cap U_2$ is the set of all vectors belonging to both U_1 and U_2;
- the *sum* $U_1 + U_2$ is the set $\{u_1 + u_2 : u_1 \in U_1, u_2 \in U_2\}$ of all sums of vectors from the two subspaces.

Both $U_1 \cap U_2$ and $U_1 + U_2$ are subspaces of V.

The next result summarises some properties of these subspaces. Proofs are left to the reader.

Proposition 1.9 *Let V be a vector space over $\mathbb{K}$.*

(a) For any $v_1, \ldots, v_n \in V$, the dimension of $\langle v_1, \ldots, v_n \rangle$ is at most n, with equality if and only if $v_1, \ldots, v_n$ are linearly independent.

(b) For any two subspaces U_1 and U_2 of V, we have

$$\dim(U_1 \cap U_2) + \dim(U_1 + U_2) = \dim(U_1) + \dim(U_2).$$

An important special case occurs when $U_1 \cap U_2$ is the zero subspace $\{0\}$. In this case, the sum $U_1 + U_2$ has the property that each of its elements has a *unique* expression in the form $u_1 + u_2$, for $u_1 \in U_1$ and $u_2 \in U_2$. For suppose that we had two different expressions for a vector v, say

$$v = u_1 + u_2 = u_1' + u_2', \qquad u_1, u_1' \in U_1, u_2, u_2' \in U_2.$$

Then

$$u_1 - u_1' = u_2' - u_2.$$

But $u_1 - u_1' \in U_1$, and $u_2' - u_2 \in U_2$; so this vector is in $U_1 \cap U_2$, and by hypothesis it is equal to 0, so that $u_1 = u_1'$ and $u_2 = u_2'$; that is, the two expressions are not different after all! In this case we say that $U_1 + U_2$ is the *direct sum* of the subspaces U_1 and U_2, and write it as $U_1 \oplus U_2$. Note that

$$\dim(U_1 \oplus U_2) = \dim(U_1) + \dim(U_2).$$

The notion of direct sum extends to more than two summands, but is a little complicated to describe. We state a form which is sufficient for our purposes.

Definition 1.8 Let $U_1, \ldots, U_r$ be subspaces of the vector space V. We say that V is the *direct sum* of $U_1, \ldots, U_r$, and write

$$V = U_1 \oplus \ldots \oplus U_r,$$

if every vector $v \in V$ can be written uniquely in the form $v = u_1 + \cdots + u_r$ with $u_i \in U_i$ for $i = 1, \ldots, r$.

Proposition 1.10 *If $V = U_1 \oplus \cdots \oplus U_r$, then*

(a) $\dim(V) = \dim(U_1) + \cdots + \dim(U_r)$;

(b) if B_i is a basis for U_i for $i = 1, \ldots, r$, then $B_1 \cup \cdots \cup B_r$ is a basis for V.

Chapter 2

Matrices and determinants

You have certainly seen matrices before; indeed, we met some in the first chapter of the notes. Here we revise matrix algebra, consider row and column operations on matrices, and define the rank of a matrix. Then we define the determinant of a square matrix axiomatically and prove that it exists (that is, there is a unique "determinant" function satisfying the rules we lay down), and give some methods of calculating it and some of its properties. Finally we prove the Cayley–Hamilton Theorem: every matrix satisfies its own characteristic equation.

2.1 Matrix algebra

Definition 2.1 A *matrix* of size $m \times n$ over a field $\mathbb{K}$, where m and n are positive integers, is an array with m rows and n columns, where each entry is an element of $\mathbb{K}$. For $1 \leq i \leq m$ and $1 \leq j \leq n$, the entry in row i and column j of A is denoted by A_{ij}, and referred to as the (i, j) entry of A.

Example 2.1 A column vector in $\mathbb{K}^n$ can be thought of as a $n \times 1$ matrix, while a row vector is a $1 \times n$ matrix.

Definition 2.2 We define addition and multiplication of matrices as follows.

(a) Let A and B be matrices of the same size $m \times n$ over $\mathbb{K}$. Then the sum $A+B$ is defined by adding corresponding entries:

$$(A+B)_{ij} = A_{ij} + B_{ij}.$$

(b) Let A be an $m \times n$ matrix and B an $n \times p$ matrix over $\mathbb{K}$. Then the product AB is the $m \times p$ matrix whose (i, j) entry is obtained by multiplying each

element in the ith row of A by the corresponding element in the jth column of B and summing:

$$(AB)_{ij} = \sum_{k=1}^{n} A_{ik}B_{kj}.$$

Remark Note that we can only add or multiply matrices if their sizes satisfy appropriate conditions. In particular, for a fixed value of n, we can add and multiply $n \times n$ matrices. It turns out that the set $M_n(\mathbb{K})$ of $n \times n$ matrices over $\mathbb{K}$ is a *ring with identity*: this means that it satisfies conditions (A0)–(A4), (M0)–(M2) and (D) of Appendix 1. The zero matrix, which we denote by O, is the matrix with every entry zero, while the identity matrix, which we denote by I, is the matrix with entries 1 on the main diagonal and 0 everywhere else. Note that matrix multiplication is not commutative: BA is usually not equal to AB.

We already met matrix multiplication in Section 1 of the notes: recall that if $P_{B,B'}$ denotes the transition matrix between two bases of a vector space, then

$$P_{B,B'}P_{B',B''} = P_{B,B''}.$$

2.2 Row and column operations

Given an $m \times n$ matrix A over a field $\mathbb{K}$, we define certain operations on A called row and column operations.

Definition 2.3 *Elementary row operations* There are three types:

Type 1 Add a multiple of the jth row to the ith, where $j \neq i$.

Type 2 Multiply the ith row by a non-zero scalar.

Tyle 3 Interchange the ith and jth rows, where $j \neq i$.

Elementary column operations There are three types:

Type 1 Add a multiple of the jth column to the ith, where $j \neq i$.

Type 2 Multiply the ith column by a non-zero scalar.

Tyle 3 Interchange the ith and jth column, where $j \neq i$.

By applying these operations, we can reduce any matrix to a particularly simple form:

Theorem 2.1 *Let A be an $m \times n$ matrix over the field $\mathbb{K}$. Then it is possible to change A into B by elementary row and column operations, where B is a matrix of the same size satisfying $B_{ii} = 1$ for $0 \leq i \leq r$, for $r \leq \min\{m,n\}$, and all other entries of B are zero.*

If A can be reduced to two matrices B and B' both of the above form, where the numbers of non-zero elements are r and r' respectively, by different sequences of elementary operations, then $r = r'$, and so $B = B'$.

Definition 2.4 The number r in the above theorem is called the *rank* of A; while a matrix of the form described for B is said to be in the *canonical form for equivalence*. We can write the canonical form matrix in "block form" as

$$B = \begin{bmatrix} I_r & O \\ O & O \end{bmatrix},$$

where I_r is an $r \times r$ identity matrix and O denotes a zero matrix of the appropriate size (that is, $r \times (n-r)$, $(m-r) \times r$, and $(m-r) \times (n-r)$ respectively for the three Os). Note that some or all of these Os may be missing: for example, if $r = m$, we just have $\begin{bmatrix} I_m & O \end{bmatrix}$.

Proof We outline the proof that the reduction is possible. To prove that we always get the same value of r, we need a different argument.

The proof is by induction on the size of the matrix A: in other words, we assume as inductive hypothesis that any smaller matrix can be reduced as in the theorem. Let the matrix A be given. We proceed in steps as follows:

- If $A = O$ (the all-zero matrix), then the conclusion of the theorem holds, with $r = 0$; no reduction is required. So assume that $A \neq O$.

- If $A_{11} \neq 0$, then skip this step. If $A_{11} = 0$, then there is a non-zero element A_{ij} somewhere in A; by swapping the first and ith rows, and the first and jth columns, if necessary (Type 3 operations), we can bring this entry into the $(1,1)$ position.

- Now we can assume that $A_{11} \neq 0$. Multiplying the first row by A_{11}^{-1}, (row operation Type 2), we obtain a matrix with $A_{11} = 1$.

- Now by row and column operations of Type 1, we can assume that all the other elements in the first row and column are zero. For if $A_{1j} \neq 0$, then subtracting A_{1j} times the first column from the jth gives a matrix with $A_{1j} = 0$. Repeat this until all non-zero elements have been removed.

- Now let B be the matrix obtained by deleting the first row and column of A. Then B is smaller than A and so, by the inductive hypothesis, we can reduce B to canonical form by elementary row and column operations. The same sequence of operations applied to A now finish the job.

Example 2.2 Here is a small example. Let

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

We have $A_{11} = 1$, so we can skip the first three steps. Subtracting twice the first column from the second, and three times the first column from the third, gives the matrix

$$\begin{bmatrix} 1 & 0 & 0 \\ 4 & -3 & -6 \end{bmatrix}.$$

Now subtracting four times the first row from the second gives

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & -3 & -6 \end{bmatrix}.$$

From now on, we have to operate on the smaller matrix $\begin{bmatrix} -3 & -6 \end{bmatrix}$, but we continue to apply the operations to the large matrix.

Multiply the second row by $-1/3$ to get

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

Now subtract twice the second column from the third to obtain

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

We have finished the reduction, and we conclude that the rank of the original matrix A is equal to 2.

We finish this section by describing the elementary row and column operations in a different way.

For each elementary row operation on an n-rowed matrix A, we define the corresponding *elementary matrix* by applying the same operation to the $n \times n$ identity matrix I. Similarly we represent elementary column operations by elementary matrices obtained by applying the same operations to the $m \times m$ identity matrix.

We don't have to distinguish between rows and columns for our elementary matrices. For example, the matrix

$$\begin{bmatrix} 1 & 2 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

corresponds to the elementary column operation of adding twice the first column to the second, or to the elementary row operation of adding twice the second row to the first. For the other types, the matrices for row operations and column operations are identical.

Lemma 2.2 *The effect of an elementary row operation on a matrix is the same as that of multiplying on the left by the corresponding elementary matrix. Similarly, the effect of an elementary column operation is the same as that of multiplying on the right by the corresponding elementary matrix.*

The proof of this lemma is somewhat tedious calculation.

Example 2.3 We continue our previous example. In order, here is the list of elementary matrices corresponding to the operations we applied to A. (Here 2×2 matrices are row operations while 3×3 matrices are column operations).

$$\begin{bmatrix}1&-2&0\\0&1&0\\0&0&1\end{bmatrix}, \begin{bmatrix}1&0&-3\\0&1&0\\0&0&1\end{bmatrix}, \begin{bmatrix}1&0\\-4&1\end{bmatrix}, \begin{bmatrix}1&0\\0&-1/3\end{bmatrix}, \begin{bmatrix}1&0&0\\0&1&-2\\0&0&1\end{bmatrix}.$$

So the whole process can be written as a matrix equation:

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

or more simply

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

where, as before,

$$A=\begin{bmatrix}1&2&3\\4&5&6\end{bmatrix}, \qquad B=\begin{bmatrix}1&0&0\\0&1&0\end{bmatrix}.$$

An important observation about the elementary operations is that each of them can have its effect undone by another elementary operation of the same kind, and hence every elementary matrix is invertible, with its inverse being another elementary matrix of the same kind. For example, the effect of adding twice the first row to the second is undone by adding -2 times the first row to the second, so that

$$\begin{bmatrix}1&2\\0&1\end{bmatrix}^{-1}=\begin{bmatrix}1&-2\\0&1\end{bmatrix}.$$

Since the product of invertible matrices is invertible, we can state the above theorem in a more concise form. First, one more definition:

Definition 2.5 The $m \times n$ matrices A and B are said to be *equivalent* if $B = PAQ$, where P and Q are invertible matrices of sizes $m \times m$ and $n \times n$ respectively.

Theorem 2.3 *Given any $m \times n$ matrix A, there exist invertible matrices P and Q of sizes $m \times m$ and $n \times n$ respectively, such that PAQ is in the canonical form for equivalence.*

Remark The relation "equivalence" defined above is an equivalence relation on the set of all $m \times n$ matrices; that is, it is reflexive, symmetric and transitive.

When mathematicians talk about a "canonical form" for an equivalence relation, they mean a set of objects which are representatives of the equivalence classes: that is, every object is equivalent to a unique object in the canonical form. We have shown this for the relation of equivalence defined earlier, except for the uniqueness of the canonical form. This is our job for the next section.

2.3 Rank

We have the unfinished business of showing that the rank of a matrix is well defined; that is, no matter how we do the row and column reduction, we end up with the same canonical form. We do this by defining two further kinds of rank, and proving that all three are the same.

Definition 2.6 Let A be an $m \times n$ matrix over a field $\mathbb{K}$. We say that the *column rank* of A is the maximum number of linearly independent columns of A, while the *row rank* of A is the maximum number of linearly independent rows of A. (We regard columns or rows as vectors in $\mathbb{K}^m$ and $\mathbb{K}^n$ respectively.)

Now we need a sequence of four lemmas.

Lemma 2.4 *(a) Elementary column operations don't change the column rank of a matrix.*

(b) Elementary row operations don't change the column rank of a matrix.

(c) Elementary column operations don't change the row rank of a matrix.

(d) Elementary row operations don't change the row rank of a matrix.

Proof (a) This is clear for Type 3 operations, which just rearrange the vectors. For Types 1 and 2, we have to show that such an operation cannot take a linearly independent set to a linearly dependent set; the *vice versa* statement holds because the inverse of an elementary operation is another operation of the same kind.

So suppose that $v_1,\ldots,v_n$ are linearly independent. Consider a Type 1 operation involving adding c times the jth column to the ith; the new columns are $v'_1,\ldots,v'_n$, where $v'_k = v_k$ for $k \neq i$, while $v'_i = v_i + cv_j$. Suppose that the new vectors are linearly dependent. Then there are scalars $a_1,\ldots,a_n$, not all zero, such that

$$\begin{aligned} 0 &= a_1v'_1 + \cdots + a_nv'_n \\ &= a_1v_1 + \cdots + a_i(v_i + cv_j) + \cdots + a_jv_j + \cdots + a_nv_n \\ &= a_1v_1 + \cdots + a_iv_i + \cdots + (a_j + ca_i)v_j + \cdots + a_nv_n. \end{aligned}$$

Since $v_1,\ldots,v_n$ are linearly independent, we conclude that

$$a_1 = 0,\ldots,a_i = 0,\ldots,a_j + ca_i = 0,\ldots,a_n = 0,$$

from which we see that all the a_k are zero, contrary to assumption. So the new columns are linearly independent.

The argument for Type 2 operations is similar but easier.

(b) It is easily checked that, if an elementary row operation is applied, then the new vectors satisfy exactly the same linear relations as the old ones (that is, the same linear combinations are zero). So the linearly independent sets of vectors don't change at all.

(c) Same as (b), but applied to rows.

(d) Same as (a), but applied to rows.

Theorem 2.5 *For any matrix A, the row rank, the column rank, and the rank are all equal. In particular, the rank is independent of the row and column operations used to compute it.*

Proof Suppose that we reduce A to canonical form B by elementary operations, where B has rank r. These elementary operations don't change the row or column rank, by our lemma; so the row ranks of A and B are equal, and their column ranks are equal. But it is trivial to see that, if

$$B = \begin{bmatrix} I_r & O \\ O & O \end{bmatrix},$$

then the row and column ranks of B are both equal to r. So the theorem is proved.

We can get an extra piece of information from our deliberations. Let A be an invertible $n \times n$ matrix. Then the canonical form of A is just I: its rank is equal to n. This means that there are matrices P and Q, each a product of elementary matrices, such that

$$PAQ = I_n.$$

From this we deduce that

$$A = P^{-1} I_n Q^{-1} = P^{-1} Q^{-1};$$

in other words,

Corollary 2.6 *Every invertible square matrix is a product of elementary matrices.*

In fact, we learn a little bit more. We observed, when we defined elementary matrices, that they can represent either elementary column operations or elementary row operations. So, when we have written A as a product of elementary matrices, we can choose to regard them as representing column operations, and we see that A can be obtained from the identity by applying elementary column operations. If we now apply the inverse operations in the other order, they will turn A into the identity (which is its canonical form). In other words, the following is true:

Corollary 2.7 *If A is an invertible $n \times n$ matrix, then A can be transformed into the identity matrix by elementary column operations alone (or by elementary row operations alone).*

2.4 Determinants

The determinant is a function defined on square matrices; its value is a scalar. It has some very important properties: perhaps most important is the fact that a matrix is invertible if and only if its determinant is not equal to zero.

We denote the determinant function by det, so that $\det(A)$ is the determinant of A. For a matrix written out as an array, the determinant is denoted by replacing the square brackets by vertical bars:

$$\det \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} = \begin{vmatrix} 1 & 2 \\ 3 & 4 \end{vmatrix}.$$

You have met determinants in earlier courses, and you know the formula for the determinant of a 2×2 or 3×3 matrix:

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc, \qquad \begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} = aei + bfg + cdh - afh - bdi - ceg.$$

Our first job is to define the determinant for square matrices of any size. We do this in an "axiomatic" manner:

Definition 2.7 A function D defined on $n \times n$ matrices is a *determinant* if it satisfies the following three conditions:

(D1) For $1 \leq i \leq n$, D is a linear function of the ith column: this means that, if A and A' are two matrices which agree everywhere except the ith column, and if A'' is the matrix whose ith column is c times the ith column of A plus c' times the ith column of A', but agreeing with A and A' everywhere else, then

$$D(A'') = cD(A) + c'D(A').$$

(D2) If A has two equal columns, then $D(A) = 0$.

(D3) $D(I_n) = 1$, where I_n is the $n \times n$ identity matrix.

We show the following result:

Theorem 2.8 *There is a unique determinant function on $n \times n$ matrices, for any n.*

Proof First, we show that applying elementary row operations to A has a well-defined effect on $D(A)$.

(a) If B is obtained from A by adding c times the jth column to the ith, then $D(B) = D(A)$.

(b) If B is obtained from A by multiplying the ith column by a non-zero scalar c, then $D(B) = cD(A)$.

(c) If B is obtained from A by interchanging two columns, then $D(B) = -D(A)$.

For (a), let A' be the matrix which agrees with A in all columns except the ith, which is equal to the jth column of A. By rule (D2), $D(A') = 0$. By rule (D1),

$$D(B) = D(A) + cD(A') = D(A).$$

Part (b) follows immediately from rule (D3).

To prove part (c), we observe that we can interchange the ith and jth columns by the following sequence of operations:

- add the ith column to the jth;
- multiply the ith column by -1;
- add the jth column to the ith;
- subtract the ith column from the jth.

In symbols,

$$(c_i, c_j) \mapsto (c_i, c_j + c_i) \mapsto (-c_i, c_j + c_i) \mapsto (c_j, c_j + c_i) \mapsto (c_j, c_i).$$

The first, third and fourth steps don't change the value of D, while the second multiplies it by -1.

Now we take the matrix A and apply elementary column operations to it, keeping track of the factors by which D gets multiplied according to rules (a)–(c). The overall effect is to multiply $D(A)$ by a certain non-zero scalar c, depending on the operations.

- If A is invertible, then we can reduce A to the identity, so that $cD(A) = D(I) = 1$, whence $D(A) = c^{-1}$.
- If A is not invertible, then its column rank is less than n. So the columns of A are linearly dependent, and one column can be written as a linear combination of the others. Applying axiom (D1), we see that $D(A)$ is a linear combination of values $D(A')$, where A' are matrices with two equal columns; so $D(A') = 0$ for all such A', whence $D(A) = 0$.

This proves that the determinant function, if it exists, is unique. We show its existence in the next section, by giving a couple of formulae for it.

Given the uniqueness of the determinant function, we now denote it by $\det(A)$ instead of $D(A)$. The proof of the theorem shows an important corollary:

Corollary 2.9 *A square matrix is invertible if and only if* $\det(A) \neq 0$.

Proof See the case division at the end of the proof of the theorem.

One of the most important properties of the determinant is the following.

Theorem 2.10 *If A and B are $n \times n$ matrices over $\mathbb{K}$, then* $\det(AB) = \det(A)\det(B)$.

Proof Suppose first that B is not invertible. Then $\det(B) = 0$. Also, AB is not invertible. (For, suppose that $(AB)^{-1} = X$, so that $XAB = I$. Then XA is the inverse of B.) So $\det(AB) = 0$, and the theorem is true.

In the other case, B is invertible, so we can apply a sequence of elementary column operations to B to get to the identity. The effect of these operations is to multiply the determinant by a non-zero factor c (depending on the operations), so that $c\det(B) = I$, or $c = (\det(B))^{-1}$. Now these operations are represented by elementary matrices; so we see that $BQ = I$, where Q is a product of elementary matrices.

If we apply the same sequence of elementary operations to AB, we end up with the matrix $(AB)Q = A(BQ) = AI = A$. The determinant is multiplied by the same factor, so we find that $c\det(AB) = \det(A)$. Since $c = \det(B))^{-1}$, this implies that $\det(AB) = \det(A)\det(B)$, as required.

Finally, we have defined determinants using columns, but we could have used rows instead:

Proposition 2.11 *The determinant is the unique function D of $n \times n$ matrices which satisfies the conditions*

(D1′) for $1 \leq i \leq n$, D is a linear function of the ith row;

(D2′) if two rows of A are equal , then $D(A) = 0$;

(D3′) $D(I_n) = 1$.

The proof of uniqueness is almost identical to that for columns. To see that $D(A) = \det(A)$: if A is not invertible, then $D(A) = \det(A) = 0$; but if A is invertible, then it is a product of elementary matrices (which can represent either row or column operations), and the determinant is the product of the factors associated with these operations.

Corollary 2.12 *If $A^\top$ denotes the transpose of A, then* $\det(A^\top) = \det(A)$.

For, if D denotes the "determinant" computed by row operations, then $\det(A) = D(A) = \det(A^\top)$, since row operations on A correspond to column operations on $A^\top$.

2.5 Calculating determinants

We now give a couple of formulae for the determinant. This finishes the job we left open in the proof of the last theorem, namely, showing that a determinant function actually exists!

The first formula involves some background notation.

Definition 2.8 A *permutation* of $\{1,\ldots,n\}$ is a bijection from the set $\{1,\ldots,n\}$ to itself. The *symmetric group* S_n consists of all permutations of the set $\{1,\ldots,n\}$. (There are $n!$ such permutations.) For any permutation $\pi \in S_n$, there is a number $\mathrm{sign}(\pi) = \pm 1$, computed as follows: write π as a product of disjoint cycles; if there are k cycles (including cycles of length 1), then $\mathrm{sign}(\pi) = (-1)^{n-k}$. A *transposition* is a permutation which interchanges two symbols and leaves all the others fixed. Thus, if τ is a transposition, then $\mathrm{sign}(\tau) = -1$.

The last fact holds because a transposition has one cycle of size 2 and $n-2$ cycles of size 1, so $n-1$ altogether; so $\text{sign}(\tau) = (-1)^{n-(n-1)} = -1$.

We need one more fact about signs: if π is any permutation and τ is a transposition, then $\text{sign}(\pi\tau) = -\text{sign}(\pi)$, where $\pi\tau$ denotes the composition of π and τ (apply first τ, then π).

Definition 2.9 Let A be an $n \times n$ matrix over $\mathbb{K}$. The *determinant* of A is defined by the formula

$$\det(A) = \sum_{\pi \in S_n} \text{sign}(\pi) A_{1\pi(1)} A_{2\pi(2)} \cdots A_{n\pi(n)}.$$

Proof In order to show that this is a good definition, we need to verify that it satisfies our three rules (D1)–(D3).

(D1) According to the definition, $\det(A)$ is a sum of $n!$ terms. Each term, apart from a sign, is the product of n elements, one from each row and column. If we look at a particular column, say the ith, it is clear that each product is a linear function of that column; so the same is true for the determinant.

(D2) Suppose that the ith and jth columns of A are equal. Let τ be the transposition which interchanges i and j and leaves the other symbols fixed. Then $\pi(\tau(i)) = \pi(j)$ and $\pi(\tau(j)) = \pi(i)$, whereas $\pi(\tau(k)) = \pi(k)$ for $k \neq i, j$. Because the elements in the ith and jth columns of A are the same, we see that the products $A_{1\pi(1)} A_{2\pi(2)} \cdots A_{n\pi(n)}$ and $A_{1\pi\tau(1)} A_{2\pi\tau(2)} \cdots A_{n\pi\tau(n)}$ are equal. But $\text{sign}(\pi\tau) = -\text{sign}(\pi)$. So the corresponding terms in the formula for the determinant cancel one another. The elements of S_n can be divided up into $n!/2$ pairs of the form $\{\pi, \pi\tau\}$. As we have seen, each pair of terms in the formula cancel out. We conclude that $\det(A) = 0$. Thus (D2) holds.

(D3) If $A = I_n$, then the only permutation π which contributes to the sum is the identity permutation ι: for any other permutation π satisfies $\pi(i) \neq i$ for some i, so that $A_{i\pi(i)} = 0$. The sign of ι is $+1$, and all the terms $A_{i\iota(i)} = A_{ii}$ are equal to 1; so $\det(A) = 1$, as required.

This gives us a nice mathematical formula for the determinant of a matrix. Unfortunately, it is a terrible formula in practice, since it involves working out $n!$ terms, each a product of matrix entries, and adding them up with $+$ and $-$ signs. For n of moderate size, this will take a very long time! (For example, $10! = 3628800$.)

Here is a second formula, which is also theoretically important but very inefficient in practice.

Definition 2.10 Let A be an $n \times n$ matrix. For $1 \leq i, j \leq n$, we define the (i, j) *minor* of A to be the $(n-1) \times (n-1)$ matrix obtained by deleting the ith row and jth column of A. Now we define the (i, j) *cofactor* of A to be $(-1)^{i+j}$ times the determinant of the (i, j) minor. (These signs have a chessboard pattern, starting with sign $+$ in the top left corner.) We denote the (i, j) cofactor of A by $K_{ij}(A)$. Finally, the *adjugate* of A is the $n \times n$ matrix $\mathrm{Adj}(A)$ whose (i, j) entry is the (j, i) cofactor $K_{ji}(A)$ of A. (Note the transposition!)

Theorem 2.13 *(a) For $j \leq i \leq n$, we have*

$$\det(A) = \sum_{i=1}^{n} A_{ij} K_{ij}(A).$$

(b) For $1 \leq i \leq n$, we have

$$\det(A) = \sum_{j=1}^{n} A_{ij} K_{ij}(A).$$

This theorem says that, if we take any column or row of A, multiply each element by the corresponding cofactor, and add the results, we get the determinant of A.

Example 2.4 Using a cofactor expansion along the first column, we see that

$$\begin{aligned}
\begin{vmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 10 \end{vmatrix} &= \begin{vmatrix} 5 & 6 \\ 8 & 10 \end{vmatrix} - 4 \begin{vmatrix} 2 & 3 \\ 8 & 10 \end{vmatrix} + 7 \begin{vmatrix} 2 & 3 \\ 5 & 6 \end{vmatrix} \\
&= (5 \cdot 10 - 6 \cdot 8) - 4(2 \cdot 10 - 3 \cdot 8) + 7(2 \cdot 6 - 3 \cdot 5) \\
&= 2 + 16 - 21 \\
&= -3
\end{aligned}$$

using the standard formula for a 2×2 determinant.

Proof We prove (a); the proof for (b) is a simple modification, using rows instead of columns. Let $D(A)$ be the function defined by the right-hand side of (a) in the theorem, using the jth column of A. We verify rules (D1)–(D3).

(D1) It is clear that $D(A)$ is a linear function of the jth column. For $k \neq j$, the cofactors are linear functions of the kth column (since they are determinants), and so $D(A)$ is linear.

(D2) If the kth and lth columns of A are equal, then each cofactor is the determinant of a matrix with two equal columns, and so is zero. The harder case is when the jth column is equal to another, say the kth. Using induction, each cofactor can be expressed as a sum of elements of the kth column times $(n-2)\times(n-2)$ determinants. In the resulting sum, it is easy to see that each such determinant occurs twice with opposite signs and multiplied by the same factor. So the terms all cancel.

(D3) Suppose that $A = I$. The only non-zero cofactor in the jth column is $K_{jj}(I)$, which is equal to $(-1)^{j+j}\det(I_{n-1}) = 1$. So $D(I) = 1$.

By the main theorem, the expression $D(A)$ is equal to $\det(A)$.

At first sight, this looks like a simple formula for the determinant, since it is just the sum of n terms, rather than $n!$ as in the first case. But each term is an $(n-1)\times(n-1)$ determinant. Working down the chain we find that this method is just as labour-intensive as the other one.

But the cofactor expansion has further nice properties:

Theorem 2.14 *For any $n\times n$ matrix A, we have*

$$A\cdot\mathrm{Adj}(A) = \mathrm{Adj}(A)\cdot A = \det(A)\cdot I.$$

Proof We calculate the matrix product. Recall that the (i,j) entry of $\mathrm{Adj}(A)$ is $K_{ji}(A)$.

Now the (i,i) entry of the product $A\cdot\mathrm{Adj}(A)$ is

$$\sum_{k=1}^{n} A_{ik}(\mathrm{Adj}(A))_{ki} = \sum_{k=1}^{n} A_{ik}K_{ik}(A) = \det(A),$$

by the cofactor expansion. On the other hand, if $i \neq j$, then the (i,j) entry of the product is

$$\sum_{k=1}^{n} A_{ik}(\mathrm{Adj}(A))_{kj} = \sum_{k=1}^{n} A_{ik}K_{jk}(A).$$

This last expression is the cofactor expansion of the matrix A' which is the same of A except for the jth row, which has been replaced by the ith row of A. (Note that changing the jth row of a matrix has no effect on the cofactors of elements in this row.) So the sum is $\det(A')$. But A' has two equal rows, so its determinant is zero.

Thus $A\cdot\mathrm{Adj}(A)$ has entries $\det(A)$ on the diagonal and 0 everywhere else; so it is equal to $\det(A)\cdot I$.

The proof for the product the other way around is the same, using columns instead of rows.

Corollary 2.15 *If the $n \times n$ matrix A is invertible, then its inverse is equal to*

$$(\det(A))^{-1}\,\mathrm{Adj}(A).$$

So how can you work out a determinant efficiently? The best method in practice is to use elementary operations.

Apply elementary operations to the matrix, keeping track of the factor by which the determinant is multiplied by each operation. If you want, you can reduce all the way to the identity, and then use the fact that $\det(I) = 1$. Often it is simpler to stop at an earlier stage when you can recognise what the determinant is. For example, if the matrix A has diagonal entries $a_1, \ldots, a_n$, and all off-diagonal entries are zero, then $\det(A)$ is just the product $a_1 \cdots a_n$.

Example 2.5 Let

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 10 \end{bmatrix}.$$

Subtracting twice the first column from the second, and three times the second column from the third (these operations don't change the determinant) gives

$$\begin{bmatrix} 1 & 0 & 0 \\ 4 & -3 & -6 \\ 7 & -6 & -11 \end{bmatrix}.$$

Now the cofactor expansion along the first row gives

$$\det(A) = \begin{vmatrix} -3 & -6 \\ -6 & -11 \end{vmatrix} = 33 - 36 = -3.$$

(At the last step, it is easiest to use the formula for the determinant of a 2×2 matrix rather than do any further reduction.)

2.6 The Cayley–Hamilton Theorem

Since we can add and multiply matrices, we can substitute them into a polynomial. For example, if

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

then the result of substituting A into the polynomial $x^2 - 3x + 2$ is

$$A^2 - 3A + 2I = \begin{bmatrix} -2 & 3 \\ -6 & 7 \end{bmatrix} + \begin{bmatrix} 0 & -3 \\ 6 & -9 \end{bmatrix} + \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

We say that the matrix A *satisfies the equation* $x^2 - 3x + 2 = 0$. (Notice that for the constant term 2 we substituted $2I$.)

It turns out that, for every $n \times n$ matrix A, we can calculate a polynomial equation of degree n satisfied by A.

Definition 2.11 Let A be a $n \times n$ matrix. The *characteristic polynomial* of A is the polynomial

$$c_A(x) = \det(xI - A).$$

This is a polynomial in x of degree n.

For example, if

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

then

$$c_A(x) = \begin{vmatrix} x & -1 \\ 2 & x-3 \end{vmatrix} = x(x-3) + 2 = x^2 - 3x + 2.$$

Indeed, it turns out that this is the polynomial we want in general:

Theorem 2.16 (Cayley–Hamilton Theorem) *Let A be an $n \times n$ matrix with characteristic polynomial $c_A(x)$. Then $c_A(A) = O$.*

Example 2.6 Let us just check the theorem for 2×2 matrices. If

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix},$$

then

$$c_A(x) = \begin{vmatrix} x-a & -b \\ -c & x-d \end{vmatrix} = x^2 - (a+d)x + (ad - bc),$$

and so

$$c_A(A) = \begin{bmatrix} a^2+bc & ab+bd \\ ac+cd & bc+d^2 \end{bmatrix} - (a+d)\begin{bmatrix} a & b \\ c & d \end{bmatrix} + (ad-bc)\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = O,$$

after a small amount of calculation.

Proof We use the theorem

$$A \cdot \mathrm{Adj}(A) = \det(A) \cdot I.$$

In place of A, we put the matrix $xI - A$ into this formula:

$$(xI - A)\,\mathrm{Adj}(xI - A) = \det(xI - A)I = c_A(x)I.$$

Now it is very tempting just to substitute $x = A$ into this formula: on the right we have $c_A(A)I = c_A(A)$, while on the left there is a factor $AI - A = O$. Unfortunately this is not valid; it is important to see why. The matrix $\mathrm{Adj}(xI - A)$ is an $n \times n$ matrix whose entries are determinants of $(n-1) \times (n-1)$ matrices with entries involving x. So the entries of $\mathrm{Adj}(xI - A)$ are polynomials in x, and if we try to substitute A for x the size of the matrix will be changed!

Instead, we argue as follows. As we have said, $\mathrm{Adj}(xI - A)$ is a matrix whose entries are polynomials, so we can write it as a sum of powers of x times matrices, that is, as a polynomial whose coefficients are matrices. For example,

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

The entries in $\mathrm{Adj}(xI - A)$ are $(n-1) \times (n-1)$ determinants, so the highest power of x that can arise is x^{n-1}. So we can write

$$\mathrm{Adj}(xI - A) = x^{n-1}B_{n-1} + x^{n-2}B_{n-2} + \cdots + xB_1 + B_0,$$

for suitable $n \times n$ matrices $B_0, \ldots, B_{n-1}$. Hence

$$\begin{aligned} c_A(x)I &= (xI - A)\,\mathrm{Adj}(xI - A) \\ &= (xI - A)(x^{n-1}B_{n-1} + x^{n-2}B_{n-2} + \cdots + xB_1 + B_0) \\ &= x^n B_{n-1} + x^{n-1}(-AB_{n-1} + B_{n-2}) + \cdots + x(-AB_1 + B_0) - AB_0. \end{aligned}$$

So, if we let

$$c_A(x) = x^n + c_{n-1}x^{n-1} + \cdots + c_1 x + c_0,$$

then we read off that

$$\begin{aligned} & & B_{n-1} &= I, \\ -AB_{n-1} &+ B_{n-2} &= c_{n-1}I, \\ & & \cdots & \\ -AB_1 &+ B_0 &= c_1 I, \\ -AB_0 & & = c_0 I. & \end{aligned}$$

We take this system of equations, and multiply the first by A^n, the second by A^{n-1}, …, and the last by $A^0 = I$. What happens? On the left, all the terms cancel in pairs: we have

$$A^n B_{n-1} + A^{n-1}(-AB_{n-1} + B_{n-2}) + \cdots + A(-AB_1 + B_0) + I(-AB_0) = O.$$

On the right, we have

$$A^n + c_{n-1}A^{n-1} + \cdots + c_1 A + c_0 I = c_A(A).$$

So $c_A(A) = O$, as claimed.

Chapter 3

Linear maps between vector spaces

We return to the setting of vector spaces in order to define linear maps between them. We will see that these maps can be represented by matrices, decide when two matrices represent the same linear map, and give another proof of the canonical form for equivalence.

3.1 Definition and basic properties

Definition 3.1 Let V and W be vector spaces over a field $\mathbb{K}$. A function α from V to W is a *linear map* if it preserves addition and scalar multiplication, that is, if

- $\alpha(v_1+v_2) = \alpha(v_1)+\alpha(v_2)$ for all $v_1, v_2 \in V$;
- $\alpha(cv) = c\alpha(v)$ for all $v \in V$ and $c \in \mathbb{K}$.

Remarks 1. We can combine the two conditions into one as follows:

$$\alpha(c_1v_1+c_2v_2) = c_1\alpha(v_1)+c_2\alpha(v_2).$$

2. In other literature the term "linear transformation" is often used instead of "linear map".

Definition 3.2 Let $\alpha : V \to W$ be a linear map. The *image* of α is the set

$$\mathrm{Im}(\alpha) = \{w \in W : w = \alpha(v) \text{ for some } v \in V\},$$

and the *kernel* of α is

$$\mathrm{Ker}(\alpha) = \{v \in V : \alpha(v) = 0\}.$$

Proposition 3.1 *Let $\alpha : V \to W$ be a linear map. Then the image of α is a subspace of W and the kernel is a subspace of V.*

Proof We have to show that each is closed under addition and scalar multiplication. For the image, if $w_1 = \alpha(v_1)$ and $w_2 = \alpha(v_2)$, then

$$w_1 + w_2 = \alpha(v_1) + \alpha(v_2) = \alpha(v_1 + v_2),$$

and if $w = \alpha(v)$ then

$$cw = c\alpha(v) = \alpha(cv).$$

For the kernel, if $\alpha(v_1) = \alpha(v_2) = 0$ then

$$\alpha(v_1 + v_2) = \alpha(v_1) + \alpha(v_2) = 0 + 0 = 0,$$

and if $\alpha(v) = 0$ then

$$\alpha(cv) = c\alpha(v) = c0 = 0.$$

Definition 3.3 We define the *rank* of α to be $\rho(\alpha) = \dim(\mathrm{Im}(\alpha))$ and the *nullity* of α to be $\nu(\alpha) = \dim(\mathrm{Ker}(\alpha))$. (We use the Greek letters 'rho' and 'nu' here to avoid confusing the rank of a linear map with the rank of a matrix, though they will turn out to be closely related!)

Theorem 3.2 (Rank–Nullity Theorem) *Let $\alpha : V \to W$ be a linear map. Then $\rho(\alpha) + \nu(\alpha) = \dim(V)$.*

Proof Choose a basis $u_1, u_2, \ldots, u_q$ for $\mathrm{Ker}(\alpha)$, where $r = \dim(\mathrm{Ker}(\alpha)) = \nu(\alpha)$. The vectors $u_1, \ldots, u_q$ are linearly independent vectors of V, so we can add further vectors to get a basis for V, say $u_1, \ldots, u_q, v_1, \ldots, v_s$, where $q + s = \dim(V)$.

We claim that the vectors $\alpha(v_1), \ldots, \alpha(v_s)$ form a basis for $\mathrm{Im}(\alpha)$. We have to show that they are linearly independent and spanning.

Linearly independent: Suppose that $c_1\alpha(v_1) + \cdots + c_s\alpha(v_s) = 0$. Then $\alpha(c_1v_1 + \cdots + c_sv_s) = 0$, so that $c_1v_1 + \cdots + c_sv_s \in \mathrm{Ker}(\alpha)$. But then this vector can be expressed in terms of the basis for $\mathrm{Ker}(\alpha)$:

$$c_1v_1 + \cdots + c_sv_s = a_1u_1 + \cdots + a_qu_q,$$

whence

$$-a_1u_1 - \cdots - a_qu_q + c_1v_1 + \cdots + c_sv_s = 0.$$

But the us and vs form a basis for V, so they are linearly independent. So this equation implies that all the as and cs are zero. The fact that $c_1 = \cdots = c_s = 0$ shows that the vectors $\alpha(v_1, \ldots, \alpha(v_s)$ are linearly independent.

Spanning: Take any vector in $\mathrm{Im}(\alpha)$, say w. Then $w = \alpha(v)$ for some $v \in V$. Write v in terms of the basis for V:

$$v = a_1u_1 + \cdots + a_qu_q + c_1v_1 + \cdots + c_sv_s$$

for some $a_1, \ldots, a_q, c_1, \ldots, c_s$. Applying α, we get

$$\begin{aligned} w &= \alpha(v) \\ &= a_1\alpha(u_1) + \cdots + a_q\alpha(u_q) + c_1\alpha(v_1) + \cdots + c_s\alpha(v_s) \\ &= c_1w_1 + \cdots + c_sw_s, \end{aligned}$$

since $\alpha(u_i) = 0$ (as $u_i \in \mathrm{Ker}(\alpha)$) and $\alpha(v_i) = w_i$. So the vectors $w_1, \ldots, w_s$ span $\mathrm{Im}(\alpha)$.

Thus, $\rho(\alpha) = \dim(\mathrm{Im}(\alpha)) = s$. Since $\nu(\alpha) = q$ and $q + s = \dim(V)$, the theorem is proved.

3.2 Representation by matrices

We come now to the second role of matrices in linear algebra: **they represent linear maps between vector spaces**.

Let $\alpha : V \to W$ be a linear map, where $\dim(V) = m$ and $\dim(W) = n$. As we saw in the first section, we can take V and W in their coordinate representation: $V = \mathbb{K}^m$ and $W = \mathbb{K}^n$ (the elements of these vector spaces being represented as column vectors). Let $e_1, \ldots, e_m$ be the standard basis for V (so that e_i is the vector with ith coordinate 1 and all other coordinates zero), and $f_1, \ldots, f_n$ the standard basis for V. Then for $i = 1, \ldots, m$, the vector $\alpha(e_i)$ belongs to W, so we can write it as a linear combination of $f_1, \ldots, f_n$.

Definition 3.4 The matrix representing the linear map $\alpha : V \to W$ relative to the bases $B = (e_1, \ldots, e_m)$ for V and $C = (f_1, \ldots, f_n)$ for W is the $n \times m$ matrix whose (i, j) entry is a_{ij}, where

$$\alpha(e_i) = \sum_{j=1}^{n} a_{ji} f_j$$

for $j = 1, \ldots, n$.

In practice this means the following. Take $\alpha(e_i)$ and write it as a column vector $[a_{1i} \quad a_{2i} \quad \cdots \quad a_{ni}]^\top$. This vector is the ith column of the matrix representing α. So, for example, if $m = 3$, $n = 2$, and

$$\alpha(e_1) = f_1 + f_2, \quad \alpha(e_2) = 2f_1 + 5f_2, \quad \alpha(e_3) = 3f_1 - f_2,$$

then the vectors $\alpha(e_i)$ as column vectors are

$$\alpha(e_1) = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \alpha(e_2) = \begin{bmatrix} 2 \\ 5 \end{bmatrix}, \quad \alpha(e_3) = \begin{bmatrix} 3 \\ -1 \end{bmatrix},$$

and so the matrix representing T is

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

Now the most important thing about this representation is that the action of α is now easily described:

Proposition 3.3 *Let $\alpha : V \to W$ be a linear map. Choose bases for V and W and let A be the matrix representing α. Then, if we represent vectors of V and W as column vectors relative to these bases, we have*

$$\alpha(v) = Av.$$

Proof Let $e_1, \ldots, e_m$ be the basis for V, and $f_1, \ldots, f_n$ for W. Take $v = \sum_{i=1}^{m} c_i e_i \in V$, so that in coordinates

$$v = \begin{bmatrix} c_1 \\ \vdots \\ c_m \end{bmatrix}.$$

Then

$$\alpha(v) = \sum_{i=1}^{m} c_i \alpha(e_i) = \sum_{i=1}^{m} \sum_{j=1}^{n} c_i a_{ji} f_j,$$

so the jth coordinate of $\alpha(v)$ is $\sum_{i=1}^{n} a_{ji} c_i$, which is precisely the jth coordinate in the matrix product Av.

In our example, if $v = 2e_1 + 3e_2 + 4e_3 = [2 \quad 3 \quad 4]^\top$, then

$$\alpha(v) = Av = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 5 & -1 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix} = \begin{bmatrix} 20 \\ 13 \end{bmatrix}.$$

Addition and multiplication of linear maps correspond to addition and multiplication of the matrices representing them.

Definition 3.5 Let α and β be linear maps from V to W. Define their sum $\alpha + \beta$ by the rule

$$(\alpha + \beta)(v) = \alpha(v) + \beta(v)$$

for all $v \in V$. It is easy to check that $\alpha + \beta$ is a linear map.

Proposition 3.4 *If α and β are linear maps represented by matrices A and B respectively, then $\alpha+\beta$ is represented by the matrix $A+B$.*

The proof of this is not difficult: just use the definitions.

Definition 3.6 Let U,V,W be vector spaces over $\mathbb{K}$, and let $\alpha : U \to V$ and $\beta : V \to W$ be linear maps. The product $\beta\alpha$ is the function $U \to W$ defined by the rule

$$(\beta\alpha)(u) = \beta(\alpha(u))$$

for all $u \in U$. Again it is easily checked that $\beta\alpha$ is a linear map. Note that the order is important: we take a vector $u \in U$, apply α to it to get a vector in V, and then apply β to get a vector in W. So $\beta\alpha$ means "apply α, then β".

Proposition 3.5 *If $\alpha : U \to V$ and $\beta : V \to W$ are linear maps represented by matrices A and B respectively, then $\beta\alpha$ is represented by the matrix BA.*

Again the proof is tedious but not difficult. Of course it follows that a linear map is invertible (as a map; that is, there is an inverse map) if and only if it is represented by an invertible matrix.

Remark Let $l = \dim(U)$, $m = \dim(V)$ and $n = \dim(W)$, then A is $m \times l$, and B is $n \times m$; so the product BA is defined, and is $n \times l$, which is the right size for a matrix representing a map from an l-dimensional to an n-dimensional space.

The significance of all this is that the strange rule for multiplying matrices is chosen so as to make Proposition 3.5 hold. The definition of multiplication of linear maps is the natural one (composition), and we could then say: what definition of matrix multiplication should we choose to make the Proposition valid? We would find that the usual definition was forced upon us.

3.3 Change of basis

The matrix representing a linear map depends on the choice of bases we used to represent it. Now we have to discuss what happens if we change the basis.

Remember the notion of *transition matrix* from Chapter 1. If $B = (v_1,\ldots,v_m)$ and $B' = (v'_1,\ldots,v'_m)$ are two bases for a vector space V, the transition matrix $P_{B,B'}$ is the matrix whose jth column is the coordinate representation of v'_j in the basis B. Then we have

$$[v]_B = P[v]_{B'},$$

where $[v]_B$ is the coordinate representation of an arbitrary vector in the basis B, and similarly for B'. The inverse of $P_{B,B'}$ is $P_{B',B}$. Let p_{ij} be the (i,j) entry of $P = P_{B,B'}$.

Now let $C = (w_1, \ldots, w_n)$ and $C' = (w'_1, \ldots, w'_n)$ be two bases for a space W, with transition matrix $Q_{C,C'}$ and inverse $Q_{C',C}$. Let $Q = Q_{C,C'}$ and let $R = Q_{C',C}$ be its inverse, with (i,j) entry r_{ij}.

Let α be a linear map from V to W. Then α is represented by a matrix A using the bases B and C, and by a matrix A' using the bases B' and C'. What is the relation between A and A'?

We just do it and see. To get A', we have to represent the vectors $\alpha(v'_i)$ in the basis C'. We have

$$v'_j = \sum_{i=1}^{m} p_{ij} v_i,$$

so

$$\begin{aligned} \alpha(v'_j) &= \sum_{i=1}^{m} p_{ij} \alpha(v_i) \\ &= \sum_{i=1}^{m} \sum_{k=1}^{m} p_{ij} A_{ki} w_k \\ &= \sum_{i=1}^{m} \sum_{k=1}^{n} \sum_{l=1}^{n} p_{ij} A_{ki} r_{lk} w'_l. \end{aligned}$$

This means, on turning things around, that

$$(A')_{lj} = \sum_{k=1}^{n} \sum_{i=1}^{m} r_{lk} A_{ki} p_{ij},$$

so, according to the rules of matrix multiplication,

$$A' = RAP = Q^{-1}AP.$$

Proposition 3.6 *Let $\alpha : V \to W$ be a linear map represented by matrix A relative to the bases B for V and C for W, and by the matrix A' relative to the bases B' for V and C' for W. If $P = P_{B,B'}$ and $Q = P_{C,C'}$ are the transition matrices from the unprimed to the primed bases, then*

$$A' = Q^{-1}AP.$$

This is rather technical; you need it for explicit calculations, but for theoretical purposes the importance is the following corollary. Recall that two matrices A and B are equivalent if B is obtained from A by multiplying on the left and right by invertible matrices. (It makes no difference that we said $B = PAQ$ before and $B = Q^{-1}AP$ here, of course.)

Proposition 3.7 *Two matrices represent the same linear map with respect to different bases if and only if they are equivalent.*

This holds because

- transition matrices are always invertible (the inverse of $P_{B,B'}$ is the matrix $P_{B',B}$ for the transition in the other direction); and
- any invertible matrix can be regarded as a transition matrix: for, if the $n \times n$ matrix P is invertible, then its rank is n, so its columns are linearly independent, and form a basis B' for $\mathbb{K}^n$; and then $P = P_{B,B'}$, where B is the "standard basis".

3.4 Canonical form revisited

Now we can give a simpler proof of Theorem 2.3 about canonical form for equivalence. First, we make the following observation.

Theorem 3.8 *Let $\alpha : V \to W$ be a linear map of rank $r = \rho(\alpha)$. Then there are bases for V and W such that the matrix representing α is, in block form,*

$$\begin{bmatrix} I_r & O \\ O & O \end{bmatrix}.$$

Proof As in the proof of Theorem 3.2, choose a basis $u_1, \ldots, u_s$ for $\mathrm{Ker}(\alpha)$, and extend to a basis $u_1, \ldots, u_s, v_1, \ldots, v_r$ for V. Then $\alpha(v_1), \ldots, \alpha(v_r)$ is a basis for $\mathrm{Im}(\alpha)$, and so can be extended to a basis $\alpha(v_1), \ldots, \alpha(v_r), x_1, \ldots, x_t$ for W. Now we will use the bases

$$\begin{aligned} v_1, \ldots, v_r, v_{r+1} = u_1, \ldots, v_{r+s} = w_s \quad &\text{for} \quad V, \\ w_1 = \alpha(v_1), \ldots, w_r = \alpha(v_r), w_{r+1} = x_1, \ldots, w_{r+s} = x_s \quad &\text{for} \quad W. \end{aligned}$$

We have

$$\alpha(v_i) = \begin{cases} w_i & \text{if } 1 \le i \le r, \\ 0 & \text{otherwise;} \end{cases}$$

so the matrix of α relative to these bases is

$$\begin{bmatrix} I_r & O \\ O & O \end{bmatrix}$$

as claimed.

We recognise the matrix in the theorem as the canonical form for equivalence. Combining Theorem 3.8 with Proposition 3.7, we see:

Theorem 3.9 *A matrix of rank r is equivalent to the matrix*

$$\begin{bmatrix} I_r & O \\ O & O \end{bmatrix}.$$

We also see, by the way, that the rank of a linear map (that is, the dimension of its image) is equal to the rank of any matrix which represents it. So all our definitions of rank agree!

The conclusion is that

two matrices are equivalent if and only if they have the same rank.

So how many equivalence classes of $m \times n$ matrices are there, for given m and n? The rank of such a matrix can take any value from 0 up to the minimum of m and n; so the number of equivalence classes is $\min\{m,n\}+1$.

Chapter 4

Linear maps on a vector space

In this chapter we consider a linear map α from a vector space V to itself. If $\dim(V) = n$ then, as in the last chapter, we can represent α by an $n \times n$ matrix relative to any basis for V. However, this time we have less freedom: instead of having two bases to choose, there is only one. This makes the theory much more interesting!

4.1 Projections and direct sums

We begin by looking at a particular type of linear map whose importance will be clear later on.

Definition 4.1 The linear map $\pi : V \to V$ is a *projection* if $\pi^2 = \pi$ (where, as usual, π^2 is defined by $\pi^2(v) = \pi(\pi(v))$).

Proposition 4.1 *If $\pi : V \to V$ is a projection, then $V = \mathrm{Im}(\pi) \oplus \mathrm{Ker}(\pi)$.*

Proof We have two things to do:

$\mathrm{Im}(\pi) + \mathrm{Ker}(\pi) = V$: Take any vector $v \in V$, and let $w = \pi(v) \in \mathrm{Im}(\pi)$. We claim that $v - w \in \mathrm{Ker}(\pi)$. This holds because

$$\pi(v-w) = \pi(v) - \pi(w) = \pi(v) - \pi(\pi(v)) = \pi(v) - \pi^2(v) = 0,$$

since $\pi^2 = \pi$. Now $v = w + (v-w)$ is the sum of a vector in $\mathrm{Im}(\pi)$ and one in $\mathrm{Ker}(\pi)$.

$\mathrm{Im}(\pi) \cap \mathrm{Ker}(\pi) = \{0\}$: Take $v \in \mathrm{Im}(\pi) \cap \mathrm{Ker}(\pi)$. Then $v = \pi(w)$ for some vector w; and

$$0 = \pi(v) = \pi(\pi(w)) = \pi^2(w) = \pi(w) = v,$$

as required (the first equality holding because $v \in \mathrm{Ker}(\pi)$).

It goes the other way too: if $V = U \oplus W$, then there is a projection $\pi : V \to V$ with $\mathrm{Im}(\pi) = U$ and $\mathrm{Ker}(\pi) = W$. For every vector $v \in V$ can be uniquely written as $v = u + w$, where $u \in U$ and $w \in W$; we define π by the rule that $\pi(v) = u$. Now the assertions are clear.

The diagram in Figure 4.1 shows geometrically what a projection is. It moves any vector v in a direction parallel to $\mathrm{Ker}(\pi)$ to a vector lying in $\mathrm{Im}(\pi)$.

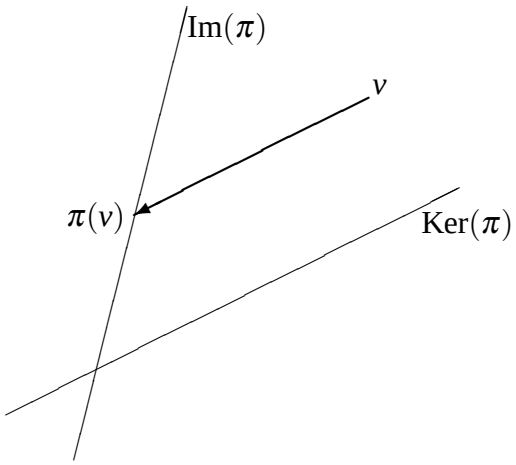


Figure 4.1: A projection

We can extend this to direct sums with more than two terms. First, notice that if π is a projection and $\pi' = I - \pi$ (where I is the identity map, satisfying $I(v) = v$ for all vectors v), then π' is also a projection, since

$$(\pi')^2 = (I-\pi)^2 = I - 2\pi + \pi^2 = I - 2\pi + \pi = I - \pi = \pi';$$

and $\pi + \pi' = I$; also $\pi\pi' = \pi(I-\pi) = \pi - \pi^2 = O$. Finally, we see that $\mathrm{Ker}(\pi) = \mathrm{Im}(\pi')$; so $V = \mathrm{Im}(\pi) \oplus \mathrm{Im}(\pi')$. In this form the result extends:

Proposition 4.2 *Suppose that $\pi_1, \pi_2, \ldots, \pi_r$ are projections on V satisfying*

(a) *$\pi_1 + \pi_2 + \cdots + \pi_r = I$, where I is the identity transformation;*

(b) *$\pi_i \pi_j = O$ for $i \neq j$.*

Then $V = U_1 \oplus U_2 \oplus \cdots \oplus U_r$, where $U_i = \mathrm{Im}(\pi_i)$.

Proof We have to show that any vector v can be *uniquely* written in the form $v = u_1 + u_2 + \cdots + u_r$, where $u_i \in U_i$ for $i = 1, \ldots, r$. We have

$$v = I(v) = \pi_1(v) + \pi_2(v) + \cdots + \pi_r(v) = u_1 + u_2 + \cdots + u_r,$$

where $u_i = \pi_i(v) \in \mathrm{Im}(\pi_i)$ for $i = 1, \ldots, r$. So any vector can be written in this form. Now suppose that we have any expression

$$v = u'_1 + u'_2 + \cdots + u'_r,$$

with $u'_i \in U_i$ for $i = 1, \ldots, r$. Since $u'_i \in U_i = \mathrm{Im}(\pi_i)$, we have $u'_i = \pi(v_i)$ for some v_i; then

$$\pi_i(u'_i) = \pi_i^2(v_i) = \pi_i(v_i) = u'_i.$$

On the other hand, for $j \neq i$, we have

$$\pi_i(u'_j) = \pi_i\pi_j(v_j) = 0,$$

since $\pi_i\pi_j = O$. So applying π_i to the expression for v, we obtain

$$\pi_i(v) = \pi_i(u'_1) + \pi_i(u'_2) + \cdots + \pi_i(u'_r) = \pi_i(u'_i) = u'_i,$$

since all terms in the sum except the ith are zero. So the only possible expression is given by $u_i = \pi_i(v)$, and the proof is complete.

Conversely, if $V = U_1 \oplus U_2 \oplus \cdots \oplus U_r$, then we can find projections $\pi_i, \pi_2, \ldots, \pi_r$ satisfying the conditions of the above Proposition. For any vector $v \in V$ has a unique expression as

$$v = u_1 + u_2 + \cdots + u_r$$

with $u_i \in U_i$ for $i = 1, \ldots, r$; then we define $\pi_i(v) = u_i$.

The point of this is that projections give us another way to recognise and describe direct sums.

4.2 Linear maps and matrices

Let $\alpha : V \to V$ be a linear map. If we choose a basis $v_1, \ldots, v_n$ for V, then V can be written in coordinates as $\mathbb{K}^n$, and α is represented by a matrix A, say, where

$$\alpha(v_i) = \sum_{j=1}^{n} a_{ji} v_j.$$

Then just as in the last section, the action of α on V is represented by the action of A on $\mathbb{K}^n$: $\alpha(v)$ is represented by the product Av. Also, as in the last chapter, sums and products (and hence arbitrary polynomials) of linear maps are represented by sums and products of the representing matrices: that is, for any polynomial $f(x)$, the map $f(\alpha)$ is represented by the matrix $f(A)$.

What happens if we change the basis? This also follows from the formula we worked out in the last chapter. However, there is only one basis to change.

Proposition 4.3 *Let α be a linear map on V which is represented by the matrix A relative to a basis B, and by the matrix A' relative to a basis B'. Let $P = P_{B,B'}$ be the transition matrix between the two bases. Then*

$$A' = P^{-1}AP.$$

Proof This is just Proposition 4.6, since P and Q are the same here.

Definition 4.2 Two $n \times n$ matrices A and B are said to be *similar* if $B = P^{-1}AP$ for some invertible matrix P.

Thus similarity is an equivalence relation, and

> two matrices are similar if and only if they represent the same linear map with respect to different bases.

There is no simple canonical form for similarity like the one for equivalence that we met earlier. For the rest of this section we look at a special class of matrices or linear maps, the "diagonalisable" ones, where we do have a nice simple representative of the similarity class. In the final section we give without proof a general result for the complex numbers.

4.3 Eigenvalues and eigenvectors

Definition 4.3 Let α be a linear map on V. A vector $v \in V$ is said to be an *eigenvector* of α, with *eigenvalue* $\lambda \in \mathbb{K}$, if $v \neq 0$ and $\alpha(v) = \lambda v$. The set $\{v : \alpha(v) = \lambda v\}$ consisting of the zero vector and the eigenvectors with eigenvalue λ is called the λ-*eigenspace* of α.

Note that we require that $v \neq 0$; otherwise the zero vector would be an eigenvector for any value of λ. With this requirement, each eigenvector has a unique eigenvalue: for if $\alpha(v) = \lambda v = \mu v$, then $(\lambda - \mu)v = 0$, and so (since $v \neq 0$) we have $\lambda = \mu$.

The name *eigenvalue* is a mixture of German and English; it means "characteristic value" or "proper value" (here "proper" is used in the sense of "property"). Another term used in older books is "latent root". Here "latent" means "hidden": the idea is that the eigenvalue is somehow hidden in a matrix representing α, and we have to extract it by some procedure. We'll see how to do this soon.

Example Let

$$A = \begin{bmatrix} -6 & 6 \\ -12 & 11 \end{bmatrix}.$$

The vector $v = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$ satisfies

$$\begin{bmatrix} -6 & 6 \\ -12 & 11 \end{bmatrix} \begin{bmatrix} 3 \\ 4 \end{bmatrix} = 2 \begin{bmatrix} 3 \\ 4 \end{bmatrix},$$

so is an eigenvector with eigenvalue 2. Similarly, the vector $w = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$ is an eigenvector with eigenvalue 3.

If we knew that, for example, 2 is an eigenvalue of A, then we could find a corresponding eigenvector $\begin{bmatrix} x \\ y \end{bmatrix}$ by solving the linear equations

$$\begin{bmatrix} -6 & 6 \\ -12 & 11 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = 2 \begin{bmatrix} x \\ y \end{bmatrix}.$$

In the next-but-one section, we will see how to find the eigenvalues, and the fact that there cannot be more than n of them for an $n \times n$ matrix.

4.4 Diagonalisability

Some linear maps have a particularly simple representation by matrices.

Definition 4.4 The linear map α on V is *diagonalisable* if there is a basis of V relative to which the matrix representing α is a diagonal matrix.

Suppose that $v_1, \ldots, v_n$ is such a basis showing that α is diagonalisable. Then $\alpha(v_i) = a_{ii}v_i$ for $i = 1, \ldots, n$, where a_{ii} is the ith diagonal entry of the diagonal matrix A. Thus, the basis vectors are eigenvectors. Conversely, if we have a basis of eigenvectors, then the matrix representing α is diagonal. So:

Proposition 4.4 *The linear map α on V is diagonalisable if and only if there is a basis of V consisting of eigenvectors of α.*

Example The matrix $\begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}$ is not diagonalisable. It is easy to see that its only eigenvalue is 1, and the only eigenvectors are scalar multiples of $[1 \quad 0]^\top$. So we cannot find a basis of eigenvectors.

Theorem 4.5 *Let $\alpha : V \to V$ be a linear map. Then the following are equivalent:*

(a) α is diagonalisable;

(b) V is the direct sum of eigenspaces of α;

(c) $\alpha = \lambda_1\pi_1 + \cdots + \lambda_r\pi_r$, where $\lambda_1, \ldots, \lambda_r$ are the distinct eigenvalues of α, and $\pi_1, \ldots, \pi_r$ are projections satisfying $\pi_1 + \cdots + \pi_r = I$ and $\pi_i\pi_j = 0$ for $i \neq j$.

Proof Let $\lambda_1, \ldots, \lambda_r$ be the distinct eigenvalues of α, and let $v_{i1}, \ldots, v_{im_i}$ be a basis for the λ_i-eigenspace of α. Then α is diagonalisable if and only if the union of these bases is a basis for V. So (a) and (b) are equivalent.

Now suppose that (b) holds. Proposition 4.2 and its converse show that there are projections $\pi_1, \ldots, \pi_r$ satisfying the conditions of (c) where $\mathrm{Im}(\pi_i)$ is the λ_i-eigenspace. Now in this case it is easily checked that T and $\sum \lambda_i\pi_i$ agree on every vector in V, so they are equal. So (b) implies (c).

Finally, if $\alpha = \sum \lambda_i\pi_i$, where the π_i satisfy the conditions of (c), then V is the direct sum of the spaces $\mathrm{Im}(\pi_i)$, and $\mathrm{Im}(\pi_i)$ is the λ_i-eigenspace. So (c) implies (b), and we are done.

Example Our matrix $A = \begin{bmatrix} -6 & 6 \\ -12 & 11 \end{bmatrix}$ is diagonalisable, since the eigenvectors $\begin{bmatrix} 3 \\ 4 \end{bmatrix}$ and $\begin{bmatrix} 2 \\ 3 \end{bmatrix}$ are linearly independent, and so form a basis for $\mathbb{R}$. Indeed, we see that

$$\begin{bmatrix} -6 & 6 \\ -12 & 11 \end{bmatrix} \begin{bmatrix} 3 & 4 \\ 2 & 3 \end{bmatrix} = \begin{bmatrix} 3 & 4 \\ 2 & 3 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix},$$

so that $P^{-1}AP$ is diagonal, where P is the matrix whose columns are the eigenvectors of A.

Furthermore, one can find two projection matrices whose column spaces are the eigenspaces, namely

$$P_1 = \begin{bmatrix} 9 & -6 \\ 12 & -8 \end{bmatrix}, \qquad P_2 = \begin{bmatrix} -8 & 6 \\ -12 & 9 \end{bmatrix}.$$

Check directly that $P_1^2 = P_1$, $P_2^2 = P_2$, $P_1P_2 = P_2P_1 = 0$, $P_1 + P_2 = I$, and $2P_1 + 3P_2 = A$.

This expression for a diagonalisable matrix A in terms of projections is useful in calculating powers of A, or polynomials in A.

Proposition 4.6 *Let*

$$A = \sum_{i=1}^{r} \lambda_i P_i$$

be the expression for the diagonalisable matrix A in terms of projections P_i satisfying the conditions of Theorem 4.5, that is, $\sum_{i=1}^{r} P_i = I$ and $P_iP_j = O$ for $i \neq j$. Then

(a) for any positive integer m, we have

$$A^m = \sum_{i=1}^{r} \lambda_i^m P_i;$$

(b) for any polynomial $f(x)$, we have

$$f(A) = \sum_{i=1}^{r} f(\lambda_i) P_i.$$

Proof (a) The proof is by induction on m, the case $m = 1$ being the given expression. Suppose that the result holds for $m = k-1$. Then

$$\begin{aligned} A^k &= A^{k-1}A \\ &= \left(\sum_{i=1}^{r} \lambda_i^{k-1} P_i\right)\left(\sum_{i=1}^{r} \lambda_i P_i\right) \end{aligned}$$

When we multiply out this product, all the terms P_iP_j are zero for $i \neq j$, and we obtain simply $\sum_{i=1}^{r} \lambda_i^{k-1}\lambda_i P_i$, as required. So the induction goes through.

(b) If $f(x) = \sum a_m x^m$, we obtain the result by multiplying the equation of part (a) by a_m and summing over m. (Note that, for $m = 0$, we use the fact that

$$A^0 = I = \sum_{i=1}^{r} P_i = \sum_{i=1}^{r} \lambda_i^0 P_i,$$

that is, part (a) holds also for $m = 0$.)

4.5 Characteristic and minimal polynomials

We defined the determinant of a square matrix A. Now we want to define the determinant of a linear map α. The obvious way to do this is to take the determinant of any matrix representing α. For this to be a good definition, we need to show that it doesn't matter which matrix we take; in other words, that $\det(A') = \det(A)$ if A and A' are similar. But, if $A' = P^{-1}AP$, then

$$\det(P^{-1}AP) = \det(P^{-1})\det(A)\det(P) = \det(A),$$

since $\det(P^{-1})\det(P) = 1$. So our plan will succeed:

Definition 4.5 (a) The *determinant* $\det(\alpha)$ of a linear map $\alpha : V \to V$ is the determinant of any matrix representing T.

(b) The *characteristic polynomial* $c_\alpha(x)$ of a linear map $\alpha : V \to V$ is the characteristic polynomial of any matrix representing α.

(c) The *minimal polynomial* $m_\alpha(x)$ of a linear map $\alpha : V \to V$ is the monic polynomial of smallest degree which is satisfied by α.

The second part of the definition is OK, by the same reasoning as the first (since $c_A(x)$ is just a determinant). But the third part also creates a bit of a problem: how do we know that α satisfies any polynomial? The Cayley–Hamilton Theorem tells us that $c_A(A) = O$ for any matrix A representing α. Now $c_A(A)$ represents $c_A(\alpha)$, and $c_A = c_\alpha$ by definition; so $c_\alpha(\alpha) = O$. Indeed, the Cayley–Hamilton Theorem can be stated in the following form:

Proposition 4.7 *For any linear map α on V, its minimal polynomial $m_\alpha(x)$ divides its characteristic polynomial $c_\alpha(x)$ (as polynomials).*

Proof Suppose not; then we can divide $c_\alpha(x)$ by $m_\alpha(x)$, getting a quotient $q(x)$ and non-zero remainder $r(x)$; that is,

$$c_\alpha(x) = m_\alpha(x)q(x) + r(x).$$

Substituting α for x, using the fact that $c_\alpha(\alpha) = m_\alpha(\alpha) = O$, we find that $r(\alpha) = 0$. But the degree of r is less than the degree of m_α, so this contradicts the definition of m_α as the polynomial of least degree satisfied by α.

Theorem 4.8 *Let α be a linear map on V. Then the following conditions are equivalent for an element λ of $\mathbb{K}$:*

(a) λ is an eigenvalue of α;

(b) λ is a root of the characteristic polynomial of α;

(c) λ is a root of the minimal polynomial of α.

Remark: This gives us a recipe to find the eigenvalues of α: take a matrix A representing α; write down its characteristic polynomial $c_A(x) = \det(xI - A)$; and find the roots of this polynomial. In our earlier example,

$$\begin{vmatrix} x-0.9 & -0.3 \\ -0.1 & x-0.7 \end{vmatrix} = (x-0.9)(x-0.7) - 0.03 = x^2 - 1.6x + 0.6 = (x-1)(x-0.6),$$

so the eigenvalues are 1 and 0.6, as we found.

Proof (b) implies (a): Suppose that $c_\alpha(\lambda) = 0$, that is, $\det(\lambda I - \alpha) = 0$. Then $\lambda I - \alpha$ is not invertible, so its kernel is non-zero. Pick a non-zero vector v in $\mathrm{Ker}(\lambda I - \alpha)$. Then $(\lambda I - \alpha)v = 0$, so that $\alpha(v) = \lambda v$; that is, λ is an eigenvalue of α.

(c) implies (b): Suppose that λ is a root of $m_\alpha(x)$. Then $(x - \lambda)$ divides $m_\alpha(x)$. But $m_\alpha(x)$ divides $c_\alpha(x)$, by the Cayley–Hamilton Theorem: so $(x - \lambda$ divides $c_\alpha(x)$, whence λ is a root of $c_\alpha(x)$.

(a) implies (c): Let λ be an eigenvalue of A with eigenvector v. We have $\alpha(v) = \lambda v$. By induction, $\alpha^k(v) = \lambda^k v$ for any k, and so $f(\alpha)(v) = f(\lambda)(v)$ for any polynomial f. Choosing $f = m_\alpha$, we have $m_\alpha(\alpha) = 0$ by definition, so $m_\alpha(\lambda)v = 0$; since $v \neq 0$, we have $m_\alpha(\lambda) = 0$, as required.

Using this result, we can give a necessary and sufficient condition for α to be diagonalisable. First, a lemma.

Lemma 4.9 *Let $v_1, \ldots, v_r$ be eigenvectors of α with distinct eigenvalues $\lambda_1, \ldots, \lambda_r$. Then $v_1, \ldots, v_r$ are linearly independent.*

Proof Suppose that $v_1, \ldots, v_r$ are linearly dependent, so that there exists a linear relation

$$c_1 v_1 + \cdots + c_r v_r = 0,$$

with coefficients c_i not all zero. Some of these coefficients may be zero; choose a relation with the smallest number of non-zero coefficients. Suppose that $c_1 \neq 0$. (If $c_1 = 0$ just re-number.) Now acting on the given relation with α, using the fact that $\alpha(v_i) = \lambda_i v_i$, we get

$$c_1\lambda_1 v_1 + \cdots + c_r\lambda_r v_r = 0.$$

Subtracting λ_1 times the first equation from the second, we get

$$c_2(\lambda_2 - \lambda_1)v_2 + \cdots + c_r(\lambda_r - \lambda_1)v_r = 0.$$

Now this equation has fewer non-zero coefficients than the one we started with, which was assumed to have the smallest possible number. So the coefficients in

this equation must all be zero. That is, $c_i(\lambda_i - \lambda_1) = 0$, so $c_i = 0$ (since $\lambda_i \neq \lambda_1$), for $i = 2, \ldots, n$. This doesn't leave much of the original equation, only $c_1 v_1 = 0$, from which we conclude that $c_1 = 0$, contrary to our assumption. So the vectors must have been linearly independent.

Theorem 4.10 *The linear map α on V is diagonalisable if and only if its minimal polynomial is the product of distinct linear factors, that is, its roots all have multiplicity* 1.

Proof Suppose first that α is diagonalisable, with eigenvalues $\lambda_1, \ldots, \lambda_r$. Then there is a basis such that α is represented by a diagonal matrix D whose diagonal entries are the eigenvalues. Now for any polynomial f, $f(\alpha)$ is represented by $f(D)$, a diagonal matrix whose diagonal entries are $f(\lambda_i)$ for $i = 1, \ldots, r$. Choose

$$f(x) = (x - \lambda_1) \cdots (x - \lambda_r).$$

Then all the diagonal entries of $f(D)$ are zero; so $f(D) = 0$. We claim that f is the minimal polynomial of α; clearly it has no repeated roots, so we will be done. We know that each λ_i is a root of $m_\alpha(x)$, so that $f(x)$ divides $m_\alpha(x)$; and we also know that $f(\alpha) = 0$, so that the degree of f cannot be smaller than that of m_α. So the claim follows.

Conversely, we have to show that if m_α is a product of distinct linear factors then α is diagonalisable. This is a little argument with polynomials. Let $f(x) = \prod(x - \lambda_i)$ be the minimal polynomial of α, with the roots λ_i all distinct. Let $h_i(x) = f(x)/(x - \lambda_i)$. Then the polynomials $h_1, \ldots, h_r$ have no common factor except 1; for the only possible factors are $(x - \lambda_i)$, but this fails to divide h_i. Now the Euclidean algorithm shows that we can write the h.c.f. as a linear combination:

$$1 = \sum_{i=1}^{r} h_i(x) k_i(x).$$

Let $U_i = \mathrm{Im}(h_i(\alpha))$. The vectors in U_i are eigenvectors of α with eigenvalue λ_i; for if $u \in U_i$, say $u = h_i(\alpha)v$, then

$$(\alpha - \lambda_i I)u_i = (\alpha - \lambda_i I)h_i(\alpha)(v) = f(\alpha)v = 0,$$

so that $\alpha(v) = \lambda_i(v)$. Moreover every vector can be written as a sum of vectors from the subspaces U_i. For, given $v \in V$, we have

$$v = Iv = \sum_{i=1}^{r} h_i(\alpha)(k_i(\alpha)v),$$

with $h_i(\alpha)(k_i(\alpha)v) \in \mathrm{Im}(h_i(\alpha))$. The fact that the expression is unique follows from the lemma, since the eigenvectors are linearly independent.

So how, in practice, do we "diagonalise" a matrix A, that is, find an invertible matrix P such that $P^{-1}AP = D$ is diagonal? We saw an example of this earlier. The matrix equation can be rewritten as $AP = PD$, from which we see that the columns of P are the eigenvectors of A. So the proceedure is: Find the eigenvalues of A, and find a basis of eigenvectors; then let P be the matrix which has the eigenvectors as columns, and D the diagonal matrix whose diagonal entries are the eigenvalues. Then $P^{-1}AP = D$.

How do we find the minimal polynomial of a matrix? We know that it divides the characteristic polynomial, and that every root of the characteristic polynomial is a root of the minimal polynomial; then it's trial and error. For example, if the characteristic polynomial is $(x-1)^2(x-2)^3$, then the minimal polynomial must be one of $(x-1)(x-2)$ (this would correspond to the matrix being diagonalisable), $(x-1)^2(x-2)$, $(x-1)(x-2)^2$, $(x-1)^2(x-2)^2$, $(x-1)(x-2)^3$ or $(x-1)^2(x-2)^3$. If we try them in this order, the first one to be satisfied by the matrix is the minimal polynomial.

For example, the characteristic polynomial of $A = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}$ is $(x-1)^2$; its minimal polynomial is not $(x-1)$ (since $A \neq I$); so it is $(x-1)^2$.

4.6 Jordan form

We finish this chapter by stating without proof a canonical form for matrices over the complex numbers under similarity.

Definition 4.6 (a) A *Jordan block* $J(n,\lambda)$ is a matrix of the form

$$\begin{bmatrix} \lambda & 1 & 0 & \cdots & 0 \\ 0 & \lambda & 1 & \cdots & 0 \\ & & \cdots & & \\ 0 & 0 & 0 & \cdots & \lambda \end{bmatrix},$$

that is, it is an $n \times n$ matrix with λ on the main diagonal, 1 in positions immediately above the main diagonal, and 0 elsewhere. (We take $J(1,\lambda)$ to be the 1×1 matrix $[\lambda]$.)

(b) A matrix is in *Jordan form* if it can be written in block form with Jordan blocks on the diagonal and zeros elsewhere.

Theorem 4.11 *Over $\mathbb{C}$, any matrix is similar to a matrix in Jordan form; that is, any linear map can be represented by a matrix in Jordan form relative to a suitable basis. Moreover, the Jordan form of a matrix or linear map is unique apart from putting the Jordan blocks in a different order on the diagonal.*

Remark A matrix over $\mathbb{C}$ is diagonalisable if and only if all the Jordan blocks in its Jordan form have size 1.

Example Any 3×3 matrix over $\mathbb{C}$ is similar to one of

$$\begin{bmatrix} \lambda & 0 & 0 \\ 0 & \mu & 0 \\ 0 & 0 & \nu \end{bmatrix}, \quad \begin{bmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \mu \end{bmatrix}, \quad \begin{bmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{bmatrix},$$

for some $\lambda, \mu, \nu \in \mathbb{C}$ (not necessarily distinct).

Example Consider the matrix $\begin{bmatrix} a & b \\ -b & a \end{bmatrix}$, with $b \neq 0$. Its characteristic polynomial is $x^2 - 2ax + (a^2 + b^2)$, so that the eigenvalues over $\mathbb{C}$ are $a + b\mathrm{i}$ and $a - b\mathrm{i}$. Thus A is diagonalisable, if we regard it as a matrix over the complex numbers. But over the real numbers, A has no eigenvalues and no eigenvectors; it is not diagonalisable, and cannot be put into Jordan form either.

We see that there are two different "obstructions" to a matrix being diagonalisable:

(a) The roots of the characteristic polynomial don't lie in the field $\mathbb{K}$. We can always get around this by working in a larger field (as above, enlarge the field from $\mathbb{R}$ to $\mathbb{C}$).

(b) Even though the characteristic polynomial factorises, there may be Jordan blocks of size bigger than 1, so that the minimal polynomial has repeated roots. This problem cannot be transformed away by enlarging the field; we are stuck with what we have.

Though it is beyond the scope of this course, it can be shown that if all the roots of the characteristic polynomial lie in the field $\mathbb{K}$, then the matrix is similar to one in Jordan form.

4.7 Trace

Here we meet another function of a linear map, and consider its relation to the eigenvalues and the characteristic polynomial.

Definition 4.7 The *trace* $\mathrm{Tr}(A)$ of a square matrix A is the sum of its diagonal entries.

Proposition 4.12 *(a) For any two $n \times n$ matrices A and B, we have* $\mathrm{Tr}(AB) = \mathrm{Tr}(BA)$.

(b) Similar matrices have the same trace.

Proof (a)

$$\mathrm{Tr}(AB) = \sum_{i=1}^{n} (AB)_{ii} = \sum_{i=1}^{n} \sum_{j=1}^{n} A_{ij} B_{ji},$$

by the rules for matrix multiplication. Now obviously $\mathrm{Tr}(BA)$ is the same thing.

(b) $\mathrm{Tr}(P^{-1}AP) = \mathrm{Tr}(APP^{-1}) = \mathrm{Tr}(AI) = \mathrm{Tr}(A)$.

The second part of this proposition shows that, if $\alpha : V \to V$ is a linear map, then any two matrices representing α have the same trace; so, as we did for the determinant, we can define the *trace* $\mathrm{Tr}(\alpha)$ of α to be the trace of any matrix representing α.

The trace and determinant of α are coefficients in the characteristic polynomial of α.

Proposition 4.13 *Let $\alpha : V \to V$ be a linear map, where $\dim(V) = n$, and let c_α be the characteristic polynomial of α, a polynomial of degree n with leading term x^n.*

(a) The coefficient of x^{n-1} is $-\mathrm{Tr}(\alpha)$, and the constant term is $(-1)^n \det(\alpha)$.

(b) If α is diagonalisable, then the sum of its eigenvalues is $\mathrm{Tr}(\alpha)$ and their product is $\det(\alpha)$.

Proof Let A be a matrix representing α. We have

$$c_\alpha(x) = \det(xI - A) = \begin{vmatrix} x - a_{11} & -a_{12} & \dots & -a_{1n} \\ -a_{21} & x - a_{22} & \dots & -a_{2n} \\ & & \dots & \\ -a_{n1} & -a_{n2} & \dots & x - a_{nn} \end{vmatrix}.$$

The only way to obtain a term in x^{n-1} in the determinant is from the product $(x-a_{11})(x-a_{22})\cdots(x-a_{nn})$ of diagonal entries, taking $-a_{ii}$ from the ith factor and x from each of the others. (If we take one off-diagonal term, we would have to have at least two, so that the highest possible power of x would be x^{n-2}.) So the coefficient of x^{n-1} is minus the sum of the diagonal terms.

Putting $x = 0$, we find that the constant term is $c_\alpha(0) = \det(-A) = (-1)^n \det(A)$.

If α is diagonalisable then the eigenvalues are the roots of $c_\alpha(x)$:

$$c_\alpha(x) = (x - \lambda_1)(x - \lambda_2)\cdots(x - \lambda_n).$$

Now the coefficient of x^{n-1} is minus the sum of the roots, and the constant term is $(-1)^n$ times the product of the roots.

Chapter 5

Linear and quadratic forms

In this chapter we examine "forms", that is, functions from a vector space V to its field, which are either linear or quadratic. The linear forms comprise the dual space of V; we look at this and define dual bases and the adjoint of a linear map (corresponding to the transpose of a matrix).

Quadratic forms make up the bulk of the chapter. We show that we can change the basis to put any quadratic form into "diagonal form" (with squared terms only), by a process generalising "completing the square" in elementary algebra, and that further reductions are possible over the real and complex numbers.

5.1 Linear forms and dual space

The definition is simple:

Definition 5.1 Let V be a vector space over $\mathbb{K}$. A *linear form* on V is a linear map from V to $\mathbb{K}$, where $\mathbb{K}$ is regarded as a 1-dimensional vector space over $\mathbb{K}$: that is, it is a function from V to $\mathbb{K}$ satisfying

$$f(v_1+v_2) = f(v_1)+f(v_2), \qquad f(cv) = cf(v)$$

for all $v_1, v_2, v \in V$ and $c \in \mathbb{K}$.

If $\dim(V) = n$, then a linear form is represented by a $1 \times n$ matrix over $\mathbb{K}$, that is, a *row vector* of length n over $\mathbb{K}$. If $f = [a_1 \quad a_2 \quad \ldots \quad a_n]$, then for $v = [x_1 \quad x_2 \quad \ldots \quad x_n]^\top$ we have

$$f(v) = [a_1 \quad a_2 \quad \ldots \quad a_n] \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = a_1x_1 + a_2x_2 + \cdots + a_nx_n.$$

Conversely, any row vector of length n represents a linear form on $\mathbb{K}^n$.

Definition 5.2 Linear forms can be added and multiplied by scalars in the obvious way:

$$(f_1 + f_2)(v) = f_1(v) + f_2(v), \qquad (cf)(v) = cf(v).$$

So they form a vector space, which is called the *dual space* of V and is denoted by V^*.

Not surprisingly, we have:

Proposition 5.1 *If V is finite-dimensional, then so is V^*, and* $\dim(V^*) = \dim(V)$.

Proof We begin by observing that, if $(v_1, \ldots, v_n)$ is a basis for V, and $a_1, \ldots, a_n$ are any scalars whatsoever, then there is a unique linear map f with the property that $f(v_i) = a_i$ for $i = 1, \ldots, n$. It is given by

$$f(c_1v_1 + \cdots + c_nv_n) = a_1c_1 + \cdots + a_nc_n,$$

in other words, it is represented by the row vector $[a_1 \quad a_2 \quad \ldots \quad a_n]$, and its action on $\mathbb{K}^n$ is by matrix multiplication as we saw earlier.

Now let f_i be the linear map defined by the rule that

$$f_i(v_j) = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j. \end{cases}$$

Then $(f_1, \ldots, f_n)$ form a basis for V^*; indeed, the linear form f defined in the preceding paragraph is $a_1f_1 + \cdots + a_nf_n$. This basis is called the *dual basis* of V^* corresponding to the given basis for V. Since it has n elements, we see that $\dim(V^*) = n = \dim(V)$.

We can describe the basis in the preceding proof as follows.

Definition 5.3 The *Kronecker delta* δ_{ij} for $i, j \in \{1, \ldots, n\}$ is defined by the rule that

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

Note that δ_{ij} is the (i, j) entry of the identity matrix. Now, if $(v_1, \ldots, v_n)$ is a basis for V, then the *dual basis* for the dual space V^* is the basis $(f_1, \ldots, f_n)$ satisfying

$$f_i(v_j) = \delta_{ij}.$$

There are some simple properties of the Kronecker delta with respect to summation. For example,

$$\sum_{i=1}^{n} \delta_{ij}a_i = a_j$$

for fixed $j \in \{1, \ldots, n\}$. This is because all terms of the sum except the term $i = j$ are zero.

5.1.1 Adjoints

Definition 5.4 Let $\alpha : V \to W$ be a linear map. There is a linear map $\alpha^* : W^* \to V^*$ (note the reversal!) defined by

$$(\alpha^*(f))(v) = f(\alpha(v)).$$

The map α^* is called the *adjoint* of α.

This definition takes a bit of unpicking. We are given $\alpha : V \to W$ and asked to define $\alpha^* : W^* \to V^*$. This means that, to any element $f \in W^*$ (any linear form on W) we must associate a linear form $g = \alpha^*(f) \in V^*$. This linear form must act on vectors $v \in V$ to produce scalars. Our definition says that $\alpha^*(f)$ maps the vector v to the scalar $f(\alpha(v))$: this makes sense because $\alpha(v)$ is a vector in W, and hence the linear form $f \in W^*$ can act on it to produce a scalar.

Now α^*, being a linear map, is represented by a matrix when we choose bases for W^* and V^*. The obvious bases to choose are the dual bases corresponding to some given bases of W and V respectively. What is the matrix? Some calculation shows the following, which will not be proved in detail here.

Proposition 5.2 *Let $\alpha : V \to W$ be a linear map. Choose bases B for V, and C for W, and let A be the matrix representing α relative to these bases. Let B^* and C^* denote the dual bases of V^* and W^* corresponding to B and C. Then the matrix representing α^* relative to the bases C^* and B^* is the transpose of A, that is, $A^\top$.*

5.1.2 Change of basis

Suppose that we change bases in V from $B = (v_1, \ldots, v_n)$ to $B' = (v'_1, \ldots, v'_n)$, with change of basis matrix $P = P_{B,B'}$. How do the dual bases change? In other words, if $B^* = (f_1, \ldots, f_n)$ is the dual basis of B, and $(B')^* = (f'_1, \ldots, f'_n)$ the dual basis of B, then what is the transition matrix $P_{B^*,(B')^*}$? The next result answers the question.

Proposition 5.3 *Let B and B' be bases for V, and B^* and $(B')^*$ the dual bases of the dual space. Then*

$$P_{B^*,(B')^*} = \left(P_{B,B'}^\top\right)^{-1}.$$

Proof Use the notation from just before the Proposition. If $P = P_{B,B'}$ has (i,j) entry p_{ij}, and $Q = P_{B^*,(B')^*}$ has (i,j) entry q_{ij}, we have

$$\begin{aligned} v'_i &= \sum_{k=1}^{n} p_{ki} v_k, \\ f'_j &= \sum_{l=1}^{n} q_{lj} f_l, \end{aligned}$$

and so

$$\begin{aligned}\delta_{ij} &= f_j'(v_i') \\ &= \left(\sum_{l=1}^{n} q_{lj} f_l\right)\left(\sum_{k=1}^{n} p_{ki} v_i\right) \\ &= \sum_{l=1}^{n}\sum_{k=1}^{n} q_{lj}\delta_{ij} p_{ki} \\ &= \sum_{k=1}^{n} q_{kj} p_{ki}.\end{aligned}$$

Now q_{kj} is the (j,k) entry of $Q^\top$, and so we have

$$I = Q^\top P,$$

whence $Q^\top = P^{-1}$, so that $Q = \left(P^{-1}\right)^\top = \left(P^\top\right)^{-1}$, as required.

5.2 Quadratic forms

A lot of applications of mathematics involve dealing with quadratic forms: you meet them in statistics (analysis of variance) and mechanics (energy of rotating bodies), among other places. In this section we begin the study of quadratic forms.

5.2.1 Quadratic forms

For almost everything in the remainder of this chapter, we assume that

the characteristic of the field $\mathbb{K}$ *is not equal to* 2.

This means that $2 \neq 0$ in $\mathbb{K}$, so that the element $1/2$ exists in $\mathbb{K}$. Of our list of "standard" fields, this only excludes $\mathbb{F}_2$, the integers mod 2. (For example, in $\mathbb{F}_5$, we have $1/2 = 3$.)

A quadratic form as a function which, when written out in coordinates, is a polynomial in which every term has total degree 2 in the variables. For example,

$$q(x,y,z) = x^2 + 4xy + 2xz - 3y^2 - 2yz - z^2$$

is a quadratic form in three variables.

We will meet a formal definition of a quadratic form later in the chapter, but for the moment we take the following.

Definition 5.5 A *quadratic form* in n variables $x_1, \ldots, x_n$ over a field K is a polynomial

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

in the variables in which every term has degree two (that is, is a multiple of $x_i x_j$ for some i, j).

In the above representation of a quadratic form, we see that if $i \neq j$, then the term in $x_i x_j$ comes twice, so that the coefficient of $x_i x_j$ is $a_{ij} + a_{ji}$. We are free to choose any two values for a_{ij} and a_{ji} as long as they have the right sum; but we will always make the choice so that the two values are equal. That is, to obtain a term $c x_i x_j$, we take $a_{ij} = a_{ji} = c/2$. (This is why we require that the characteristic of the field is not 2.)

Any quadratic form is thus represented by a *symmetric* matrix A with (i, j) entry a_{ij} (that is, a matrix satisfying $A = A^\top$). *This is the third job of matrices in linear algebra:* **Symmetric matrices represent quadratic forms.**

We think of a quadratic form as defined above as being a function from the vector space $\mathbb{K}^n$ to the field $\mathbb{K}$. It is clear from the definition that

$$q(x_1, \ldots, x_n) = v^\top A v, \text{ where } v = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}.$$

Now if we change the basis for V, we obtain a different representation for the same function q. The effect of a change of basis is a linear substitution $v = Pv'$ on the variables, where P is the transition matrix between the bases. Thus we have

$$v^\top A v = (Pv')\top A(Pv') = (v')^\top (P^\top A P) v',$$

so we have the following:

Proposition 5.4 *A basis change with transition matrix P replaces the symmetric matrix A representing a quadratic form by the matrix $P^\top A P$.*

As for other situations where matrices represented objects on vector spaces, we make a definition:

Definition 5.6 Two symmetric matrices A, A' over a field $\mathbb{K}$ are *congruent* if $A' = P^\top A P$ for some invertible matrix P.

Proposition 5.5 *Two symmetric matrices are congruent if and only if they represent the same quadratic form with respect to different bases.*

Our next job, as you may expect, is to find a canonical form for symmetric matrices under congruence; that is, a choice of basis so that a quadratic form has a particularly simple shape. We will see that the answer to this question depends on the field over which we work. We will solve this problem for the fields of real and complex numbers.

5.2.2 Reduction of quadratic forms

Even if we cannot find a canonical form for quadratic forms, we can simplify them very greatly.

Theorem 5.6 *Let q be a quadratic form in n variables $x_1,\ldots,x_n$, over a field $\mathbb{K}$ whose characteristic is not 2. Then by a suitable linear substitution to new variables $y_1,\ldots,y_n$, we can obtain*

$$q = c_1y_1^2 + c_2y_2^2 + \cdots + c_ny_n^2$$

for some $c_1,\ldots,c_n \in \mathbb{K}$.

Proof Our proof is by induction on n. We call a quadratic form which is written as in the conclusion of the theorem *diagonal*. A form in one variable is certainly diagonal, so the induction starts. Now assume that the theorem is true for forms in $n-1$ variables. Take

$$q(x_1,\ldots,x_n) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}x_ix_j,$$

where $a_{ij} = a_{ji}$ for $i \neq j$.

Case 1: Assume that $a_{ii} \neq 0$ for some i. By a permutation of the variables (which is certainly a linear substitution), we can assume that $a_{11} \neq 0$. Let

$$y_1 = x_1 + \sum_{i=2}^{n}(a_{1i}/a_{11})x_i.$$

Then we have

$$a_{11}y_1^2 = a_{11}x_1^2 + 2\sum_{i=2}^{n} a_{1i}x_1x_i + q'(x_2,\ldots,x_n),$$

where q' is a quadratic form in $x_2,\ldots,x_n$. That is, all the terms involving x_1 in q have been incorporated into $a_{11}y_1^2$. So we have

$$q(x_1,\ldots,x_n) = a_{11}y_1^2 + q''(x_2,\ldots,x_n),$$

where q'' is the part of q not containing x_1 minus q'.

By induction, there is a change of variable so that

$$q''(x_2,\ldots,x_n) = \sum_{i=2}^{n} c_i y_i^2,$$

and so we are done (taking $c_1 = a_{11}$).

Case 2: All a_{ii} are zero, but $a_{ij} \neq 0$ for some $i \neq j$. Now

$$x_{ij} = \tfrac{1}{4}\left((x_i + x_j)^2 - (x_i - x_j)^2\right),$$

so taking $x_i' = \frac{1}{2}(x_i + x_j)$ and $x_j' = \frac{1}{2}(x_i - x_j)$, we obtain a new form for q which does contain a non-zero diagonal term. Now we apply the method of Case 1.

Case 3: All a_{ij} are zero. Now q is the zero form, and there is nothing to prove: take $c_1 = \cdots = c_n = 0$.

Example 5.1 Consider the quadratic form $q(x,y,z) = x^2 + 2xy + 4xz + y^2 + 4z^2$. We have

$$(x+y+2z)^2 = x^2 + 2xy + 4xz + y^2 + 4z^2 + 4yz,$$

and so

$$\begin{aligned} q &= (x+y+2z)^2 - 4yz \\ &= (x+y+2z)^2 - (y+z)^2 + (y-z)^2 \\ &= u^2 + v^2 - w^2, \end{aligned}$$

where $u = x+y+2z$, $v = y-z$, $w = y+z$. Otherwise said, the matrix representing the quadratic form, namely

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

is congruent to the matrix

$$A' = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix}.$$

Can you find an invertible matrix P such that $P^\top AP = A'$?

Thus any quadratic form can be reduced to the diagonal shape

$$\alpha_1 x_1^2 + \cdots + \alpha_n x_n^2$$

by a linear substitution. But this is still not a "canonical form for congruence". For example, if $y_1 = x_1/c$, then $\alpha_1 x_1^2 = (\alpha_1 c^2) y_1^2$. In other words, we can multiply any α_i by any factor which is a perfect square in $\mathbb{K}$.

Over the complex numbers $\mathbb{C}$, every element has a square root. Suppose that $\alpha_1, \ldots, \alpha_r \neq 0$, and $\alpha_{r+1} = \cdots = \alpha_n = 0$. Putting

$$y_i = \begin{cases} (\sqrt{\alpha_i})x_i & \text{for } 1 \leq i \leq r, \\ x_i & \text{for } r+1 \leq i \leq n, \end{cases}$$

we have

$$q = y_1^2 + \cdots + y_r^2.$$

We will see later that r is an "invariant" of q: however we do the reduction, we arrive at the same value of r.

Over the real numbers $\mathbb{R}$, things are not much worse. Since any positive real number has a square root, we may suppose that $\alpha_1, \ldots, \alpha_s > 0$, $\alpha_{s+1}, \ldots, \alpha_{s+t} < 0$, and $\alpha_{s+t+1}, \ldots, \alpha_n = 0$. Now putting

$$y_i = \begin{cases} (\sqrt{\alpha_i})x_i & \text{for } 1 \leq i \leq s, \\ (\sqrt{-\alpha_i})x_i & \text{for } s+1 \leq i \leq s+t, \\ x_i & \text{for } s+t+1 \leq i \leq n, \end{cases}$$

we get

$$q = x_1^2 + \cdots + x + s^2 - x_{s+1}^2 - \cdots - x_{s+t}^2.$$

Again, we will see later that s and t don't depend on how we do the reduction. [This is the theorem known as *Sylvester's Law of Inertia*.]

5.2.3 Quadratic and bilinear forms

The formal definition of a quadratic form looks a bit different from the version we gave earlier, though it amounts to the same thing. First we define a bilinear form.

Definition 5.7 (a) Let $b : V \times V \to \mathbb{K}$ be a function of two variables from V with values in $\mathbb{K}$. We say that b is a *bilinear form* if it is a linear function of each variable when the other is kept constant: that is,

$$b(v, w_1 + w_2) = b(v, w_1) + b(v, w_2), \qquad b(v, cw) = cb(v, w),$$

with two similar equations involving the first variable. A bilinear form b is *symmetric* if $b(v, w) = b(w, v)$ for all $v, w \in V$.

(b) Let $q : V \to \mathbb{K}$ be a function. We say that q is a *quadratic form* if

- $q(cv) = c^2 q(v)$ for all $c \in \mathbb{K}$, $v \in V$;
- the function b defined by

$$b(v,w) = \tfrac{1}{2}(q(v+w) - q(v) - q(w))$$

is a bilinear form on V.

Remarks The bilinear form in the second part is symmetric; and the division by 2 in the definition is permissible because of our assumption that the characteristic of $\mathbb{K}$ is not 2.

If we think of the prototype of a quadratic form as being the function x^2, then the first equation says $(cx)^2 = c^2x^2$, while the second has the form

$$\tfrac{1}{2}((x+y)^2 - x^2 - y^2) = xy,$$

and xy is the prototype of a bilinear form: it is a linear function of x when y is constant, and *vice versa*.

Note that the formula $b(x,y) = \frac{1}{2}(q(x+y) - q(x) - q(y))$ (which is known as the *polarisation formula*) says that the bilinear form is determined by the quadratic term. Conversely, if we know the symmetric bilinear form b, then we have

$$2q(v) = 4q(v) - 2q(v) = q(v+v) - q(v) - q(v) = 2b(v,v),$$

so that $q(v) = b(v,v)$, and we see that the quadratic form is determined by the symmetric bilinear form. So these are equivalent objects.

If b is a symmetric bilinear form on V and $B = (v_1, \ldots, v_n)$ is a basis for V, then we can represent b by the $n \times n$ matrix A whose (i,j) entry is $a_{ij} = b(v_i, v_j)$. Note that A is a symmetric matrix. It is easy to see that this is the same as the matrix representing the quadratic form.

Here is a third way of thinking about a quadratic form. Let V^* be the dual space of V, and let $\alpha : V \to V^*$ be a linear map. Then for $v \in V$, we have $\alpha(v) \in V^*$, and so $\alpha(v)(w)$ is an element of $\mathbb{K}$. The function

$$b(v,w) = \alpha(v)(w)$$

is a bilinear form on V. If $\alpha(v)(w) = \alpha(w)(v)$ for all $v, w \in V$, then this bilinear form is symmetric. Conversely, a symmetric bilinear form b gives rise to a linear map $\alpha : V \to V^*$ satisfying $\alpha(v)(w) = \alpha(w)(v)$, by the rule that $\alpha(v)$ is the linear map $w \mapsto b(v,w)$.

Now given $\alpha : V \to V^*$, choose a basis B for V, and let B^* be the dual basis for V^*. Then α is represented by a matrix A relative to the bases B and B^*.

Summarising:

Proposition 5.7 *The following objects are equivalent on a vector space over a field whose characteristic is not* 2*:*

(a) a quadratic form on V;

(b) a symmetric bilinear form on V;

(c) a linear map $\alpha : V \to V^*$ satisfying $\alpha(v)(w) = \alpha(w)(v)$ for all $v, w \in V$.

Moreover, if corresponding objects of these three types are represented by matrices as described above, then we get the same matrix A in each case. Also, a change of basis in V with transition matrix P replaces A by $P^\top AP$.

Proof Only the last part needs proof. We have seen it for a quadratic form, and the argument for a bilinear form is the same. So suppose that $\alpha : V \to V^*$, and we change from B to B' in V with transition matrix P. We saw that the transition matrix between the dual bases in V^* is $(P^\top)^{-1}$. Now go back to the discussion of linear maps between different vector spaces in Chapter 4. If $\alpha : V \to W$ and we change bases in V and W with transition matrices P and Q, then the matrix A representing α is changed to $Q^{-1}AP$. Apply this with $Q = P^\top)^{-1}$, so that $Q^{-1} = P^\top$, and we see that the new matrix is $P^\top AP$, as required.

5.2.4 Canonical forms for complex and real forms

Finally, in this section, we return to quadratic forms (or symmetric matrices) over the real and complex numbers, and find canonical forms under congruence. Recall that two symmetric matrices A and A' are congruent if $A' = P^\top AP$ for some invertible matrix P; as we have seen, this is the same as saying that the represent the same quadratic form relative to different bases.

Theorem 5.8 *Any $n \times n$ complex symmetric matrix A is congruent to a matrix of the form*

$$\begin{bmatrix} I_r & O \\ O & O \end{bmatrix}$$

for some r. Moreover, $r = \mathrm{rank}(A)$, and so A is congruent to two matrices of this form then they both have the same value of r.

Proof We already saw that A is congruent to a matrix of this form. Moreover, if P is invertible, then so is $P^\top$, and so

$$r = \mathrm{rank}(P^\top AP) = \mathrm{rank}(A)$$

as claimed.

The next result is *Sylvester's Law of Inertia*.

Theorem 5.9 *Any $n \times n$ real symmetric matrix A is congruent to a matrix of the form*

$$\begin{bmatrix} I_s & O & O \\ O & -I_t & O \\ O & O & O \end{bmatrix}$$

for some s,t. Moreover, if A is congruent to two matrices of this form, then they have the same values of s and of t.

Proof Again we have seen that A is congruent to a matrix of this form. Arguing as in the complex case, we see that $s+t = \text{rank}(A)$, and so any two matrices of this form congruent to A have the same values of $s+t$.

Suppose that two different reductions give the values s,t and s',t' respectively, with $s+t = s'+t' = n$. Suppose for a contradiction that $s < s'$. Now let q be the quadratic form represented by A. Then we are told that there are linear functions $y_1,\ldots,y_n$ and $z_1,\ldots,z_n$ of the original variables $x_1,\ldots,x_n$ of q such that

$$q = y_1^2 + \cdots + y_s^2 - y_{s+1}^2 - \cdots - y_{s+t}^2 = z_1^2 + \cdots + z_{s'}^2 - z_{s'+1}^2 - \cdots - z_{s+t}^2.$$

Now consider the equations

$$y_1 = 0, \ldots, y_s = 0, z_{s'+1} = 0, \ldots z_n = 0$$

regarded as linear equations in the original variables $x_1,\ldots,x_n$. The number of equations is $s + (n-s') = n - (s'-s) < n$. According to a lemma from much earlier in the course (we used it in the proof of the Exchange Lemma!), the equations have a non-zero solution. That is, there are values of $x_1,\ldots,x_n$, not all zero, such that the variables $y_1,\ldots,y_s$ and $z_{s'+1},\ldots,z_n$ are all zero.

Since $y_1 = \cdots = y_s = 0$, we have for these values

$$q = -y_{s+1}^2 - \cdots - y_n^2 \leq 0.$$

But since $z_{s'+1} = \cdots = z_n = 0$, we also have

$$q = z_1^2 + \cdots + z_{s'}^2 > 0.$$

But this is a contradiction. So we cannot have $s < s'$. Similarly we cannot have $s' < s$ either. So we must have $s = s'$, as required to be proved.

We saw that $s+t$ is the rank of A. The number $s-t$ is known as the *signature* of A. Of course, both the rank and the signature are independent of how we reduce the matrix (or quadratic form); and if we know the rank and signature, we can easily recover s and t.

You will meet some further terminology in association with Sylvester's Law of Inertia. Let q be a quadratic form in n variables represented by the real symmetric matrix A. Let q (or A) have rank $s+t$ and signature $s-t$, that is, have s positive and t negative terms in its diagonal form. We say that q (or A) is

- *positive definite* if $s=n$ (and $t=0$), that is, if $q(v) \geq 0$ for all v, with equality only if $v=0$;
- *positive semidefinite* if $t=0$, that is, if $q(v) \geq 0$ for all v;
- *negative definite* if $t=n$ (and $s=0$), that is, if $q(v) \leq 0$ for all v, with equality only if $v=0$;
- *negative semi-definite* if $s=0$, that is, if $q(v) \leq 0$ for all v;
- *indefinite* if $s>0$ and $t>0$, that is, if $q(v)$ takes both positive and negative values.

Chapter 6

Inner product spaces

Ordinary Euclidean space is a 3-dimensional vector space over $\mathbb{R}$, but it is more than that: the extra geometric structure (lengths, angles, etc.) can all be derived from a special kind of bilinear form on the space known as an inner product. We examine inner product spaces and their linear maps in this chapter.

One can also define inner products for complex vector spaces, but things are a bit different: we have to use a form which is not quite bilinear. We defer this to Chapter 8.

6.1 Inner products and orthonormal bases

Definition 6.1 An *inner product* on a real vector space V is a function $b : V \times V \to \mathbb{R}$ satisfying

- b is bilinear (that is, b is linear in the first variable when the second is kept constant and *vice versa*);
- b is *positive definite*, that is, $b(v,v) \geq 0$ for all $v \in V$, and $b(v,v) = 0$ if and only if $v = 0$.

We usually write $b(v,w)$ as $v \cdot w$. An inner product is sometimes called a *dot product* (because of this notation).

Geometrically, in a real vector space, we define $v \cdot w = |v|.|w| \cos\theta$, where $|v|$ and $|w|$ are the lengths of v and w, and θ is the angle between v and w. Of course this definition doesn't work if either v or w is zero, but in this case $v \cdot w = 0$. But it is much easier to reverse the process. Given an inner product on V, we define

$$|v| = \sqrt{v \cdot v}$$

for any vector $v \in V$; and, if $v, w \neq 0$, then we define the angle between them to be θ, where

$$cos\theta = \frac{v \cdot w}{|v|.|w|}.$$

For this definition to make sense, we need to know that

$$-|v|.|w| \leq v \cdot w \leq |V|.|w|$$

for any vectors v, w (since $\cos\theta$ lies between -1 and 1). This is the content of the *Cauchy–Schwarz inequality*:

Theorem 6.1 *If v, w are vectors in an inner product space then*

$$(v \cdot w)^2 \leq (v \cdot v)(w \cdot w).$$

Proof By definition, we have $(v + xw) \cdot (v + xw) \geq 0$ for any real number x. Expanding, we obtain

$$x^2(w \cdot w) + 2x(v \cdot w) + (v \cdot v) \geq 0.$$

This is a quadratic function in x. Since it is non-negative for all real x, either it has no real roots, or it has two equal real roots; thus its discriminant is non-positive, that is,

$$(v \cdot w)^2 - (v \cdot v)(w \cdot w) \leq 0,$$

as required.

There is essentially only one kind of inner product on a real vector space.

Definition 6.2 A basis $(v_1, \ldots, v_n)$ for an inner product space is called *orthonormal* if $v_i \cdot v_j = \delta_{ij}$ (the Kronecker delta) for $1 \leq i, j \leq n$.

Remark: If vectors $v_1, \ldots, v_n$ satisfy $v_i \cdot v_j = \delta_{ij}$, then they are necessarily linearly independent. For suppose that $c_1 v_1 + \cdots + c_n v_n = 0$. Taking the inner product of this equation with v_i, we find that $c_i = 0$, for all i.

Theorem 6.2 *Let $\cdot$ be an inner product on a real vector space V. Then there is an orthonormal basis $(v_1, \ldots, v_n)$ for V. If we represent vectors in coordinates with respect to this basis, say $v = [x_1 \quad x_2 \quad \ldots \quad x_n]^\top$ and $w = [y_1 \quad y_2 \quad \ldots \quad y_n]^\top$, then*

$$v \cdot w = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n.$$

Proof This follows from our reduction of quadratic forms in the last chapter. Since the inner product is bilinear, the function $q(v) = v \cdot v = |v|^2$ is a quadratic form, and so it can be reduced to the form

$$q = x_1^2 + \cdots + x_s^2 - x_{s+1}^2 - \cdots - x_{s+t}^2.$$

Now we must have $s = n$ and $t = 0$. For, if $t > 0$, then the $s+1$st basis vector v_{s+1} satisfies $v_{s+1} \cdot v_{s+1} = -1$; while if $s + t < n$, then the nth basis vector v_n satisfies $v_n \cdot v_n = 0$. Either of these would contradict the positive definiteness of V. Now we have

$$q(x_1, \ldots, x_n) = x_1^2 + \cdots + x_n^2,$$

and by polarisation we find that

$$b((x_1, \ldots, x_n), (y_1, \ldots, y_n)) = x_1 y_1 + \cdots + x_n y_n,$$

as required.

However, it is possible to give a more direct proof of the theorem; this is important because it involves a constructive method for finding an orthonormal basis, known as the *Gram–Schmidt process*.

Let $w_1, \ldots, w_n$ be any basis for V. The Gram–Schmidt process works as follows.

- Since $w_1 \neq 0$, we have $w_1 \cdot w_1 > 0$, that is, $|w_1| > 0$. Put $v_1 = w_1/|w_1|$; then $|v_1| = 1$, that is, $v_1 \cdot v_1 = 1$.
- For $i = 2, \ldots, n$, let $w_i' = w_i - (v_1 \cdot w_i)v_1$. Then

  $$v_1 \cdot w_i' = v_1 \cdot w_i - (v_1 \cdot w_i)(v_1 \cdot v_1) = 0$$

  for $i = 2, \ldots, n$.
- Now apply the Gram–Schmidt process recursively to $(w_2', \ldots, w_n')$.

Since we replace these vectors by linear combinations of themselves, their inner products with v_1 remain zero throughout the process. So if we end up with vectors $v_2, \ldots, v_n$, then $v_1 \cdot v_i = 0$ for $i = 2, \ldots, n$. By induction, we can assume that $v_i \cdot v_j = \delta_{ij}$ for $i, j = 2, \ldots, n$; by what we have said, this holds if i or j is 1 as well.

Definition 6.3 The inner product on $\mathbb{R}^n$ for which the standard basis is orthonormal (that is, the one given in the theorem) is called the *standard inner product* on $\mathbb{R}^n$.

Example 6.1 In $\mathbb{R}^3$ (with the standard inner product), apply the Gram–Schmidt process to the vectors $w_1 = [1 \quad 2 \quad 2]^\top$, $w_2 = [1 \quad 1 \quad 0]^\top$, $w_3 = [1 \quad 0 \quad 0]^\top$.

To simplify things, I will write (a_1, a_2, a_3) instead of $[a_1 \quad a_2 \quad a_3]^\top$.

We have $w_1 \cdot w_1 = 9$, so in the first step we put

$$v_1 = \tfrac{1}{3} w_1 = (\tfrac{1}{3}, \tfrac{2}{3}, \tfrac{2}{3}).$$

Now $v_1 \cdot w_2 = 1$ and $v_1 \cdot w_3 = \frac{1}{3}$, so in the second step we find

$$\begin{aligned} w_2' &= w_2 - v_1 = (\tfrac{2}{3}, \tfrac{1}{3}, -\tfrac{2}{3}), \\ w_3' &= w_3 - \tfrac{1}{3} v_1 = (\tfrac{8}{9}, -\tfrac{2}{9}, \tfrac{2}{9}). \end{aligned}$$

Now we apply Gram–Schmidt recursively to w_2' and w_3'. We have $w_2' \cdot w_2' = 1$, so $v_2 = w_2' = (\frac{2}{3}, \frac{1}{3}, -\frac{2}{3})$. Then $v_2 \cdot w_3' = \frac{2}{3}$, so

$$w_3'' = w_3' - \tfrac{2}{3} v_2 = (\tfrac{4}{9}, -\tfrac{4}{9}, \tfrac{2}{9}).$$

Finally, $w_3'' \cdot w_3'' = \frac{4}{9}$, so $v_3 = \frac{3}{2} w_3'' = (\frac{2}{3}, -\frac{2}{3}, \frac{1}{3})$.

Check that the three vectors we have found really do form an orthonormal basis.

6.2 Adjoints and orthogonal linear maps

We saw in the last chapter that a bilinear form on V is the same thing as a linear map from V to its dual space. The importance of an inner product is that the corresponding linear map is a bijection which maps an orthonormal basis of V to its dual basis in V^*.

Recall that the linear map $\alpha : V \to V^*$ corresponding to a bilinear form b on V satisfies $\alpha(v)(w) = b(v,w)$; in our case, $\alpha(v)(w) = v \cdot w$. Now suppose that $(v_1, \ldots, v_n)$ is an orthonormal basis for V, so that $v_i \cdot v_j = \delta_{ij}$. Then, if $\alpha(v_i) = f_i$, we have $f_i(v_j) = \delta_{ij}$; but this is exactly the statement that $(f_1, \ldots, f_n)$ is the dual basis to $(v_1, \ldots, v_n)$.

So, on an inner product space V, we have a natural way of matching up V with V^*.

Recall too that we defined the *adjoint* of $\alpha : V \to V$ to be the map $\alpha^* : V^* \to V^*$ defined by $\alpha^*(f)(v) = f(\alpha(v))$, and we showed that the matrix representing α^* relative to the dual basis is the transpose of the matrix representing α relative to the original basis.

Translating all this to an inner product space, we have the following definition and result:

Definition 6.4 Let V be an inner product space, and $\alpha : V \to V$ a linear map. Then the *adjoint* of α is the linear map $\alpha^* : V \to V$ defined by

$$v \cdot \alpha^*(w) = \alpha(v) \cdot w.$$

Proposition 6.3 *If α is represented by the matrix A relative to an orthonormal basis of V, then α^* is represented by the transposed matrix $A^\top$.*

Now we define two important classes of linear maps on V.

Definition 6.5 Let α be a linear map on an inner product space V.

(a) α is *self-adjoint* if $\alpha^* = \alpha$.

(b) α is *orthogonal* if it is invertible and $\alpha^* = \alpha^{-1}$.

Proposition 6.4 *If α is represented by a matrix A (relative to an orthonormal basis), then*

(a) α is self-adjoint if and only if A is symmetric;

(b) α is orthogonal if and only if $A^\top A = I$.

Part (a) of this result shows that we have yet another equivalence relation on real symmetric matrices:

Definition 6.6 Two real symmetric matrices are called *orthogonally similar* if they represent the same self-adjoint map with respect to different orthonormal bases.

Then, from part (b), we see:

Proposition 6.5 *Two real symmetric matrices A and A' are orthogonally similar if and only if there is an orthogonal matrix P such that $A' = P^{-1}AP = P^\top AP$.*

Here $P^{-1} = P^\top$ because P is orthogonal. We see that orthogonal similarity is a refinement of both similarity and congruence. We will examine self-adjoint maps (or symmetric matrices) further in the next section.

Next we look at orthogonal maps.

Theorem 6.6 *The following are equivalent for a linear map α on an inner product space V:*

(a) α is orthogonal;

(b) α preserves the inner product, that is, $\alpha(v) \cdot \alpha(w) = v \cdot w$;

(c) α maps an orthonormal basis of V to an orthonormal basis.

Proof We have

$$\alpha(v) \cdot \alpha(w) = v \cdot \alpha^*(\alpha(w)),$$

by the definition of adjoint; so (a) and (b) are equivalent.

Suppose that $(v_1, \ldots, v_n)$ is an orthonormal basis, that is, $v_i \cdot v_j = \delta_{ij}$. If (b) holds, then $\alpha(v_i) \cdot \alpha(v_j) = \delta_{ij}$, so that $(\alpha(v_1), \ldots, \alpha(v_n)$ is an orthonormal basis, and (c) holds. Converesely, suppose that (c) holds, and let $v = \sum x_i v_i$ and $w = \sum y_i v_i$ for some orthonormal basis $(v_1, \ldots, v_n)$, so that $v \cdot w = \sum x_i y_i$. We have

$$\alpha(v) \cdot \alpha(w) = \left(\sum x_i \alpha(v_i)\right) \cdot \left(\sum y_i \alpha(v_i)\right) = \sum x_i y_i,$$

since $\alpha(v_i) \cdot \alpha(v_j) = \delta_{ij}$ by assumption; so (b) holds.

Corollary 6.7 *α is orthogonal if and only if the columns of the matrix representing α relative to an orthonormal basis themselves form an orthonormal basis.*

Proof The columns of the matrix representing α are just the vectors $\alpha(v_1), \ldots, \alpha(v_n)$, written in coordinates relative to $v_1, \ldots, v_n$. So this follows from the equivalence of (a) and (c) in the theorem. Alternatively, the condition on columns shows that $A^\top A = I$, where A is the matrix representing α; so $\alpha^* \alpha = I$, and α is orthogonal.

Example Our earlier example of the Gram–Schmidt process produces the orthogonal matrix

$$\begin{bmatrix} \frac{1}{3} & \frac{2}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{1}{3} & -\frac{2}{3} \\ \frac{2}{3} & -\frac{2}{3} & \frac{1}{3} \end{bmatrix}$$

whose columns are precisely the orthonormal basis we constructed in the example.

Chapter 7

Symmetric and Hermitian matrices

We come to one of the most important topics of the course. In simple terms, any real symmetric matrix is diagonalisable. But there is more to be said!

7.1 Orthogonal projections and orthogonal decompositions

We say that two vectors v, w in an inner product space are *orthogonal* if $v \cdot w = 0$.

Definition 7.1 Let V be a real inner product space, and U a subspace of V. The *orthogonal complement* of U is the set of all vectors which are orthogonal to everything in U:

$$U^\perp = \{w \in V : w \cdot u = 0 \text{ for all } u \in U\}.$$

Proposition 7.1 *If V is an inner product space and U a subspace of V, with $\dim(V) = n$ and $\dim(U) = r$, then $U^\perp$ is a subspace of V, and $\dim(U^\perp) = n - r$. Moreover, $V = U \oplus U^\perp$.*

Proof Proving that $U^\perp$ is a subspace is straightforward from the properties of the inner product. If $w_1, w_2 \in U^\perp$, then $w_1 \cdot u = w_2 \cdot u = 0$ for all $u \in U$, so $(w_1 + w_2) \cdot u = 0$ for all $u \in U$, whence $w_1 + w_2 \in U^\perp$. The argument for scalar multiples is similar.

Now choose a basis for U and extend it to a basis for V. Then apply the Gram–Schmidt process to this basis (starting with the elements of the basis for U), to obtain an orthonormal basis $(v_1, \ldots, v_n)$. Since the process only modifies vectors by adding multiples of earlier vectors, the first r vectors in the resulting basis will form an orthonormal basis for U. The last $n - r$ vectors will be orthogonal to

U, and so lie in $U^\perp$; and they are clearly linearly independent. Now suppose that $w \in U^\perp$ and $w = \sum c_i v_i$, where $(v_1, \ldots, v_n)$ is the orthonormal basis we constructed. Then $c_i = w \cdot v_i = 0$ for $i = 1, \ldots, r$; so w is a linear combination of the last $n-r$ basis vectors, which thus form a basis of $U^\perp$. Hence $\dim(U^\perp) = n-r$, as required.

Now the last statement of the proposition follows from the proof, since we have a basis for V which is a disjoint union of bases for U and $U^\perp$.

Recall the connection between direct sum decompositions and projections. If we have projections $P_1, \ldots, P_r$ whose sum is the identity and which satisfy $P_i P_j = O$ for $i \neq j$, then the space V is the direct sum of their images. This can be refined in an inner product space as follows.

Definition 7.2 Let V be an inner product space. A linear map $\pi : V \to V$ is an *orthogonal projection* if

(a) π is a projection, that is, $\pi^2 = \pi$;

(b) π is self-adjoint, that is, $\pi^* = \pi$ (where $\pi^*(v) \cdot w = v \cdot \pi(w)$ for all $v, w \in V$).

Proposition 7.2 *If π is an orthogonal projection, then* $\mathrm{Ker}(\pi) = \mathrm{Im}(\pi)^\perp$.

Proof We know that $V = \mathrm{Ker}(\pi) \oplus \mathrm{Im}(\pi)$; we only have to show that these two subspaces are orthogonal. So take $v \in \mathrm{Ker}(\pi)$, so that $\pi(v) = 0$, and $w \in \mathrm{Im}(\pi)$, so that $w = \pi(u)$ for some $u \in V$. Then

$$v \cdot w = v \cdot \pi(u) = \pi^*(v) \cdot u = \pi(v) \cdot u = 0,$$

as required.

Proposition 7.3 *Let $\pi_1, \ldots, \pi_r$ be orthogonal projections on an inner product space V satisfying $\pi_1 + \cdots + \pi_r = I$ and $\pi_i \pi_j = O$ for $i \neq j$. Let $U_i = \mathrm{Im}(\pi_i)$ for $i = 1, \ldots, r$. Then*

$$V = U_1 \oplus U_2 \oplus \cdots \oplus U_r,$$

and if $u_i \in U_i$ and $u_j \in U_j$, then u_i and u_j are orthogonal.

Proof The fact that V is the direct sum of the images of the π_i follows from Proposition 5.2. We only have to prove the last part. So take u_i and u_j as in the Proposition, say $u_i = \pi_i(v)$ and $u_j = \pi_j(w)$. Then

$$u_i \cdot u_j = \pi_i(v) \cdot \pi_j(w) = \pi_i^*(v) \cdot \pi_j(w) = v \cdot \pi_i(\pi_j(w)) = 0,$$

where the second equality holds since π_i is self-adjoint and the third is the definition of the adjoint.

A direct sum decomposition satisfying the conditions of the theorem is called an *orthogonal decomposition* of V.

Conversely, if we are given an orthogonal decomposition of V, then we can find orthogonal projections satisfying the hypotheses of the theorem.

7.2 The Spectral Theorem

The main theorem can be stated in two different ways. I emphasise that these two theorems are the same! Either of them can be referred to as the *Spectral Theorem*.

Theorem 7.4 *If α is a self-adjoint linear map on a real inner product space V, then the eigenspaces of α form an orthogonal decomposition of V. Hence there is an orthonormal basis of V consisting of eigenvectors of α. Moreover, there exist orthogonal projections $\pi_1, \ldots, \pi_r$ satisfying $\pi_1 + \cdots + \pi_r = I$ and $\pi_i \pi_j = O$ for $i \neq j$, such that*

$$\alpha = \lambda_1 \pi_1 + \cdots + \lambda_r \pi_r,$$

where $\lambda_1, \ldots, \lambda_r$ are the distinct eigenvalues of α.

Theorem 7.5 *Let A be a real symmetric matrix. Then there exists an orthogonal matrix P such that $P^{-1}AP$ is diagonal. In other words, any real symmetric matrix is orthogonally similar to a diagonal matrix.*

Proof The second theorem follows from the first, since the transition matrix from one orthonormal basis to another is an orthogonal matrix. So we concentrate on the first theorem. It suffices to find an orthonormal basis of eigenvectors, since all the rest follows from our remarks about projections, together with what we already know about diagonalisable maps.

The proof will be by induction on $n = \dim(V)$. There is nothing to do if $n = 1$. So we assume that the theorem holds for $(n-1)$-dimensional spaces.

The first job is to show that α has an eigenvector.

Choose an orthonormal basis; then α is represented by a real symmetric matrix A. Its characteristic polynomial has a root λ over the complex numbers. (The so-called "Fundamental Theorem of Algebra" asserts that any polynomial over $\mathbb{C}$ has a root.) We temporarily enlarge the field from $\mathbb{R}$ to $\mathbb{C}$. Now we can find a column vector $v \in \mathbb{C}^n$ such that $Av = \lambda v$. Taking the complex conjugate, remembering that A is real, we have $A\overline{v} = \overline{\lambda}\overline{v}$.

If $v = [z_1 \quad z_2 \quad \cdots \quad z_n]^\top$, then we have

$$\begin{aligned}
\overline{\lambda}(|z_1|^2 + |z_2|^2 + \cdots + |z_n|^2) &= \overline{\lambda}\overline{v}^\top v \\
&= (A\overline{v})^\top v \\
&= \overline{v}^\top A v \\
&= \overline{v}^\top (\lambda v) \\
&= \lambda(|z_1|^2 + |z_2|^2 + \cdots + |z_n|^2),
\end{aligned}$$

so $(\overline{\lambda} - \lambda)(|z_1|^2 + |z_2|^2 + \cdots + |z_n|^2) = 0$. Since v is not the zero vector, the second factor is positive, so we must have $\overline{\lambda} = \lambda$, that is, λ is real.

Now since α has a real eigenvalue, we can choose a real eigenvector v, and (multiplying by a scalar if necessary) we can assume that $|v| = 1$.

Let U be the subspace $v^\perp = \{u \in V : v \cdot u = 0\}$. This is a subspace of V of dimension $n-1$. We claim that $\alpha : U \to U$. For take $u \in U$. Then

$$v \cdot \alpha(u) = \alpha^*(v) \cdot u = \alpha(v) \cdot u = \lambda v \cdot u = 0,$$

where we use the fact that α is self-adjoint. So $\alpha(u) \in U$.

So α is a self-adjoint linear map on the $(n-1)$-dimensional inner product space U. By the inductive hypothesis, U has an orthonormal basis consisting of eigenvectors of α. They are all orthogonal to the unit vector v; so, adding v to the basis, we get an orthonormal basis for V, and we are done.

Remark The theorem is almost a canonical form for real symmetric relations under the relation of orthogonal congruence. If we require that the eigenvalues occur in decreasing order down the diagonal, then the result is a true canonical form: each matrix is orthogonally similar to a unique diagonal matrix with this property.

Corollary 7.6 *If α is self-adjoint, then eigenvectors of α corresponding to distinct eigenvalues are orthogonal.*

Proof This follows from the theorem, but is easily proved directly. If $\alpha(v) = \lambda v$ and $\alpha(w) = \mu w$, then

$$\lambda v \cdot w = \alpha(v) \cdot w = \alpha^*(v) \cdot w = v \cdot \alpha(w) = \mu v \cdot w,$$

so, if $\lambda \neq \mu$, then $v \cdot w = 0$.

Example 7.1 Let

$$A = \begin{bmatrix} 10 & 2 & 2 \\ 2 & 13 & 4 \\ 2 & 4 & 13 \end{bmatrix}.$$

The characteristic polynomial of A is

$$\begin{vmatrix} x-10 & -2 & -2 \\ -2 & x-13 & -4 \\ -2 & -4 & x-13 \end{vmatrix} = (x-9)^2(x-18),$$

so the eigenvalues are 9 and 18.

For eigenvalue 18 the eigenvectors satisfy

$$\begin{bmatrix} 10 & 2 & 2 \\ 2 & 13 & 4 \\ 2 & 4 & 13 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 18x \\ 18y \\ 18z \end{bmatrix},$$

so the eigenvectors are multiples of $[1 \quad 2 \quad 2]^\top$. Normalising, we can choose a unit eigenvector $[\frac{1}{3} \quad \frac{2}{3} \quad \frac{2}{3}]^\top$.

For the eigenvalue 9, the eigenvectors satisfy

$$\begin{bmatrix} 10 & 2 & 2 \\ 2 & 13 & 4 \\ 2 & 4 & 13 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 9x \\ 9y \\ 9z \end{bmatrix},$$

that is, $x+2y+2z=0$. (This condition says precisely that the eigenvectors are orthogonal to the eigenvector for $\lambda = 18$, as we know.) Thus the eigenspace is 2-dimensional. We need to choose an orthonormal basis for it. This can be done in many different ways: for example, we could choose $[0 \quad 1/\sqrt{2} \quad -1/\sqrt{2}]^\top$ and $[-4/3\sqrt{2} \quad 1/3\sqrt{2} \quad 1/3\sqrt{2}]^\top$. Then we have an orthonormal basis of eigenvectors. We conclude that, if

$$P = \begin{bmatrix} 1/3 & 0 & -4/3\sqrt{2} \\ 2/3 & 1/\sqrt{2} & 1/3\sqrt{2} \\ 2/3 & -1/\sqrt{2} & 1/3\sqrt{2} \end{bmatrix},$$

then P is orthogonal, and

$$P^\top AP = \begin{bmatrix} 18 & 0 & 0 \\ 0 & 9 & 0 \\ 0 & 0 & 9 \end{bmatrix}.$$

You might like to check that the orthogonal matrix in the example in the last chapter of the notes also diagonalises A.

7.3 Quadratic forms revisited

Any real quadratic form is represented by a real symmetric matrix; and, as we have seen, orthogonal similarity is a refinement of congruence. This gives us a new look at the reduction of real quadratic forms. Recall that any real symmetric matrix is congruent to one of the form

$$\begin{bmatrix} I_s & O & O \\ O & -I_t & O \\ O & O & O \end{bmatrix},$$

where the numbers s and t are uniquely determined: $s+t$ is the rank, and $s-t$ the signature, of the matrix (Sylvester's Law of Inertia).

Proposition 7.7 *The rank of a real symmetric matrix is equal to the number of non-zero eigenvalues, and the signature is the number of positive eigenvalues minus the number of negative eigenvalues (counted according to multiplicity).*

Proof Given a real symmetric matrix A, there is an orthogonal matrix P such that $P^\top AP$ is diagonal, with diagonal entries $\lambda_1, \ldots, \lambda_n$, say. Suppose that $\lambda_1, \ldots, \lambda_s$ are positive, $\lambda_{s+1}, \ldots, \lambda_{s+t}$ are negative, and the remainder are zero. Let D be a diagonal matrix with diagonal entries

$$1/\sqrt{\lambda_1}, \ldots, 1/\sqrt{\lambda_s}, 1/\sqrt{-\lambda_{s+1}}, \ldots, 1/\sqrt{-\lambda_{s+t}}, 1, \ldots, 1.$$

Then

$$(PD)^\top APD = D^\top P^\top APD = \begin{bmatrix} I_s & O & O \\ O & -I_t & O \\ O & O & O \end{bmatrix}.$$

7.4 Simultaneous diagonalisation

There are two important theorems which allow us to diagonalise more than one matrix at the same time. The first theorem we will consider just in the matrix form.

Theorem 7.8 *Let A and B be real symmetric matrices, and suppose that A is positive definite. Then there exists an invertible matrix P such that $P^\top AP = I$ and $P^\top BP$ is diagonal. Moreover, the diagonal entries of $P^\top BP$ are the roots of the polynomial* $\det(xA - B) = 0$.

Proof A is a real symmetric matrix, so there exists an invertible matrix P_1 such that $P_1^\top AP_1$ is in the canonical form for congruence (as in Sylvester's Law of Inertia). Since A is positive definite, this canonical form must be I; that is, $P_1^\top AP_1 = I$.

Now consider $P_1^\top BP = C$. This is a real symmetric matrix; so, according to the spectral theorem (in matrix form), we can find an orthogonal matrix P_2 such that $P_2^\top CP_2 = D$ is diagonal. Moreover, P_2 is orthogonal, so $P_2^\top P_2 = I$.

Let $P = P_1P_2$. Then

$$P^\top AP = P_2^\top (P_1^\top AP_1)P_2 = P_2^\top IP_2 = I,$$

and

$$P^\top BP = P_2^\top (P_1^\top BP_1)P_2 = P_2^\top CP_2 = D,$$

as required.

The diagonal entries of D are the eigenvalues of C, that is, the roots of the equation $\det(xI-C)=0$. Now we have

$$\det(P_1^\top)\det(xA-B)\det(P_1)=\det(P_1^\top(xA-B)P_1)=\det(xP_1^\top AP_1-P_1^\top BP_1)=\det(xI-C),$$

and $\det(P_1^\top)=\det(P_1)$ is non-zero; so the polynomials $\det(xA-B)$ and $\det(xI-C)$ are non-zero multiples of each other and so have the same roots.

You might meet this formula in mechanics. If a mechanical system has n coordinates $x_1,\ldots,x_n$, then the kinetic energy is a quadratic form in the velocities $\dot{x}_1,\ldots,\dot{x}_n$, and (from general physical principles) is positive definite (zero velocities correspond to minimum energy); near equilibrium, the potential energy is approximated by a quadratic function of the coordinates $x_1,\ldots,x_n$. If we simultaneously diagonalise the matrices of the two quadratic forms, then we can solve n separate differential equations rather than a complicated system with n variables!

The second theorem can be stated either for linear maps or for matrices.

Theorem 7.9 *(a) Let α and β be self-adjoint maps on an inner product space V, and suppose that $\alpha\beta=\beta\alpha$. Then there is an orthonormal basis for V which consists of vectors which are simultaneous eigenvalues for α and β.*

(b) Let A and B be real symmetric matrices satisfying $AB=BA$. Then there is an orthogonal matrix P such that both $P^\top AP$ and $P^\top BP$ are diagonal.

Proof Statement (b) is just a translation of (a) into matrix terms; so we prove (a).

Let $\lambda_1,\ldots,\lambda_r$ be the distinct eigenvalues of α. By the Spectral Theorem, have an orthogonal decomposition

$$V=U_1\oplus\cdots\oplus U_r,$$

where U_i is the λ_i-eigenspace of α.

We claim that β maps U_i to U_i. For take $u\in U_i$, so that $\alpha(u)=\lambda_i u$. Then

$$\alpha(\beta(u))=\beta(\alpha(u))=\beta(\lambda_i u)=\lambda_i\beta(u),$$

so $\beta(u)$ is also an eigenvector of α with eigenvalue λ_i. Hence $\beta(u)\in U_i$, as required.

Now β is a self-adjoint linear map on the inner product space U_i, and so by the spectral theorem again, U_i has an orthonormal basis consisting of eigenvectors of β. But these vectors are also eigenvectors of α, since they belong to U_i.

Finally, since we have an orthogonal decomposition, putting together all these bases gives us an orthonormal basis of V consisting of simultaneous eigenvectors of α and β.

Remark This theorem easily extends to an arbitrary set of real symmetric matrices such that any two commute. For a finite set, the proof is by induction on the number of matrices in the set, based on the proof just given. For an infinite set, we use the fact that they span a finite-dimensional subspace of the space of all real symmetric matrices; to diagonalise all the matrices in our set, it suffices to diagonalise the matrices in a basis.

Chapter 8

The complex case

The theory of real inner product spaces and self-adjoint linear maps has a close parallel in the complex case. However, some changes are required. In this chapter we outline the complex case. Usually, the proofs are similar to those in the real case.

8.1 Complex inner products

There are no positive definite bilinear forms over the complex numbers; for we always have $(\mathrm{i}v)\cdot(\mathrm{i}v) = -v\cdot v$.

But it is possible to modify the definitions so that everything works in the same way over $\mathbb{C}$.

Definition 8.1 A *inner product* on a complex vector space V is a map $b : V \times V \to \mathbb{C}$ satisfying

(a) b is a linear function of its second variable, keeping the first variable constant;

(b) $b(w,v) = \overline{b(v,w)}$, where $\bar{}$ denotes complex conjugation. [It follows that $b(v,v) \in \mathbb{R}$ for all $v \in V$.]

(c) $b(v,v) \geq 0$ for all $v \in V$, and $b(v,v) = 0$ if and only if $v = 0$.

As before, we write $b(v,w)$ as $v\cdot w$. This time, b is not linear as a function of its first variable; in fact we have

$$b(v_1+v_2,w) = b(v_1,w)+b(v_2,w), \qquad b(cv,w) = \bar{c}b(v,w)$$

for $v_1, v_2, v, w \in V$ and $c \in \mathbb{C}$. (Sometimes we say that b is *semilinear* (that is, "$\frac{1}{2}$"-linear) as a function of its first variable, and describe it as a *sesquilinear form*

(that is, "$1\frac{1}{2}$-linear". A form satisfying (b) is called *Hermitian*, and one satisfying (c) is *positive definite*. Thus an inner product is a positive definite Hermitian sesquilinear form.)

The definition of an orthonormal basis is exactly as in the real case, and the Gram–Schmidt process allows us to find one with only trivial modifications. The standard inner product (with respect to an orthonormal basis) is given by

$$v \cdot w = \overline{x_1} y_1 + \cdots + \overline{x_n} y_n,$$

where $v = [x_1 \quad \ldots \quad x_n]^\top$, $w = [y_1 \quad \cdots \quad y_n]^\top$.

The adjoint of $\alpha : V \to V$ is defined as before by the formula

$$\alpha^*(v) \cdot w = v \cdot \alpha(w),$$

but this time there is a small difference in the matrix representation: if α is represented by A (relative to an orthonormal basis), then its adjoint α^* is represented by $(\overline{A})^\top$. (Take the complex conjugates of all the entries in A, and then transpose.) So

- a self-adjoint linear map is represented by a matrix A satisfying $A = (\overline{A})^\top$: such a matrix is called *Hermitian*.
- a map which preserves the inner product (that is, which satisfies $\alpha(v) \cdot \alpha(w) = v \cdot w$, or $\alpha^* = \alpha^{-1}$) is represented by a matrix A satisfying $(\overline{A})^\top = A^{-1}$: such a matrix is called *unitary*.

8.2 The complex Spectral Theorem

The spectral theorem for self-adjoint linear maps on complex inner product spaces is almost identical to the real version. The proof goes through virtually unchanged.

The definition of an orthogonal projection is the same: a projection which is self-adjoint.

Theorem 8.1 *If α is a self-adjoint linear map on a complex inner product space V, then the eigenspaces of α form an orthogonal decomposition of V. Hence there is an orthonormal basis of V consisting of eigenvectors of α. Moreover, there exist orthogonal projections $\pi_1, \ldots, \pi_r$ satisfying $\pi_1 + \cdots + \pi_r = I$ and $\pi_i \pi_j = O$ for $i \neq j$, such that*

$$\alpha = \lambda_1 \pi_1 + \cdots + \lambda_r \pi_r,$$

where $\lambda_1, \ldots, \lambda_r$ are the distinct eigenvalues of α.

Theorem 8.2 *Let A be a complex Hermitian matrix. Then there exists a unitary matrix P such that $P^{-1}AP$ is diagonal.*

There is one special feature of the complex case:

Proposition 8.3 *Any eigenvalue of a self-adjoint linear map on a complex inner product space (or of a complex Hermitian matrix) is real.*

Proof Suppose that α is self-adjoint and $\alpha(v) = \lambda v$. Then

$$\lambda v \cdot v = v \cdot \alpha(v) = \alpha^*(v) \cdot v = \alpha(v) \cdot v = \overline{\lambda} v \cdot v,$$

where in the last step we use the fact that $(cv) \cdot w = \bar{c} v \cdot w$ for a complex inner product. So $(\overline{\lambda} - \lambda) v \cdot v = 0$. Since $v \neq 0$, we have $v \cdot v \neq 0$, and so $\overline{\lambda} = \lambda$; that is, λ is real.

We also have a theorem on simultaneous diagonalisation:

Proposition 8.4 *Let α and β be self-adjoint linear maps of a complex inner product space V, and suppose that $\alpha\beta = \beta\alpha$. Then there is an orthonormal basis for V consisting of eigenvectors of both α and β.*

The proof is as in the real case. You are invited to formulate the theorem in terms of commuting Hermitian matrices.

8.3 Normal matrices

The fact that the eigenvalues of a complex Hermitian matrix are real leaves open the possibility of proving a more general version of the spectral theorem. We saw that a real symmetric matrix is orthogonally similar to a diagonal matrix. In fact, the converse is also true. For if A is a real $n \times n$ matrix and P is an orthogonal matrix such that $P^\top AP = D$ is diagonal, then $A = PDP^\top$, and so

$$A^\top = PD^\top P^\top = PDP^\top = A.$$

In other words, *a real matrix is orthogonally similar to a diagonal matrix if and only if it is symmetric.*

This is not true for complex Hermitian matrices, since such matrices have real eigenvalues and so cannot be similar to non-real diagonal matrices.

What really happens is the following.

Definition 8.2 (a) Let α be a linear map on a complex inner-product space V. We say that α is *normal* if it commutes with its adjoint: $\alpha\alpha^* = \alpha^*\alpha$.

(b) Let A be an $n \times n$ matrix over $\mathbb{C}$. We say that A is *normal* if it commutes with its conjugate transpose: $A\overline{A}^\top = \overline{A}^\top A$.

Theorem 8.5 *(a) Let α be a linear map on a complex inner product space V. Then V has an orthonormal basis consisting of eigenvectors of α if and only if α is normal.*

(b) Let A be an $n \times n$ matrix over $\mathbb{C}$. Then there is a unitary matrix P such that $P^{-1}AP$ is diagonal if and only if A is normal.

Proof As usual, the two forms of the theorem are equivalent. We prove it in the first form.

If α has an orthonormal basis $(v_1, \ldots, v_n)$ consisting of eigenvectors, then $\alpha(v_i) = \lambda_i v_i$ for $i = 1, \ldots, n$, where λ_i are eigenvalues. We see that $\alpha^*(v_i) = \overline{\lambda_i} v_i$, and so

$$\alpha\alpha^*(v_i) = \alpha^*\alpha(v_i) = \overline{\lambda_i}\lambda_i v_i.$$

Since $\alpha\alpha^*$ and $\alpha^*\alpha$ agree on the vectors of a basis, they are equal; so α is normal.

Conversely, suppose that α is normal. Let

$$\beta = \tfrac{1}{2}(\alpha + \alpha^*), \qquad \gamma = \tfrac{1}{2\mathrm{i}}(\alpha - \alpha^*).$$

(You should compare these with the formulae $x = \frac{1}{2}(z+\overline{z})$, $y = \frac{1}{2\mathrm{i}}(z-\overline{z})$ for the real and imaginary parts of a quadratic form. The analogy is even closer, since clearly we have $\alpha = \beta + \mathrm{i}\gamma$.) Now we claim:

- *β and γ are Hermitian.* For

$$\begin{aligned} \beta^* &= \tfrac{1}{2}(\alpha^* + \alpha) = \beta, \\ \gamma^* &= \tfrac{1}{-2\mathrm{i}}(\alpha^* - \alpha) = \gamma, \end{aligned}$$

  where we use the fact that $(c\alpha)^* = \overline{c}\alpha^*$.

- $\beta\gamma = \gamma\beta$. For

$$\begin{aligned} \beta\gamma &= \frac{1}{4\mathrm{i}}(\alpha^2 - \alpha\alpha^* + \alpha^*\alpha - (\alpha^*)^2) = \frac{1}{4\mathrm{i}}(\alpha^2 - (\alpha^*)^2), \\ \gamma\beta &= \frac{1}{4\mathrm{i}}(\alpha^2 + \alpha\alpha^* - \alpha^*\alpha - (\alpha^*)^2) = \frac{1}{4\mathrm{i}}(\alpha^2 - (\alpha^*)^2). \end{aligned}$$

  (Here we use the fact that $\alpha\alpha^* = \alpha^*\alpha$.)

Hence, by the Proposition at the end of the last section, there is an orthonormal basis B whose vectors are eigenvectors of β and γ, and hence are eigenvectors of $\alpha = \beta + \mathrm{i}\gamma$.

Note that the eigenvalues of β and γ in this proof are the real and imaginary parts of the eigenvalues of α.

Chapter 9

Skew-symmetric matrices

We spent the last three chapters looking at symmetric matrices; even then we could only find canonical forms for the real and complex numbers. It turns out that life is much simpler for skew-symmetric matrices. We find a canonical form for these matrices under congruence which works for any field whatever. (More precisely, as we will see, this statement applies to "alternating matrices", but these are precisely the same as skew-symmetric matrices unless the characteristic of the field is 2.)

9.1 Alternating bilinear forms

Alternating forms are as far from positive definite as they can be:

Definition 9.1 Let V be a vector space over $\mathbb{K}$. A bilinear form b on V is *alternating* if $b(v,v) = 0$ for all $v \in V$.

Proposition 9.1 *An alternating bilinear form b satisfies $b(w,v) = -b(v,w)$ for all $v, w \in V$.*

Proof

$$0 = b(v+w, v+w) = b(v,v) + b(v,w) + b(w,v) + b(w,w) = b(v,w) + b(w,v)$$

for any $v, w \in V$, using the definition of an alternating bilinear form.

Now here is the analogue of the Gram–Schmidt process for alternating bilinear forms.

Theorem 9.2 *Let b be an alternating bilinear form on a vector space V. Then there is a basis $(u_1,\dots,u_s,w_1,\dots,w_s,z_1,\dots,z_t)$ for V such that $b(u_i,w_i)=1$ and $b(w_i,u_i)=-1$ for $i=1,\dots,s$ and $b(x,y)=0$ for any other choices of basis vectors x and y.*

Proof If b is identically zero, then simply choose a basis $(z_1,\dots,z_n)$ and take $s=0$, $t=n$. So suppose not.

Choose a pair of vectors u and w such that $c=b(u,w)\neq 0$. Replacing w by w/c, we have $b(u,w)=1$.

We claim that u and w are linearly independent. For suppose that $cu+dw=0$. Then

$$\begin{aligned} 0 &= b(u,cu+dw)=cb(u,u)+db(u,w)=d,\\ 0 &= b(w,cu+dw)=cb(w,u)+db(w,w)=-c, \end{aligned}$$

so $c=d=0$. We take $u_1=u$ and $w_1=v$ as our first two basis vectors.

Now let $U=\langle u,w\rangle$ and $W=\{x\in V : b(u,x)=b(w,x)=0\}$. We claim that $V=U\oplus W$. The argument just above already shows that $U\cap W=0$, so we have to show that $V=U+W$. So take a vector $v\in V$, and let $x=-b(w,v)u+b(u,v)w$. Then

$$\begin{aligned} b(u,x)&=-b(w,v)b(u,u)+b(u,v)b(u,w)=b(u,v),\\ b(w,x)&=-b(w,v)b(w,u)+b(u,v)b(w,w)=b(w,v) \end{aligned}$$

so $b(u,v-x)=b(w,v-x)=0$. Thus $v-x\in W$. But clearly $x\in U$, and so our assertion is proved.

Now b is an alternating bilinear form on W, and so by induction there is a basis of the required form for W, say $(u_2,\dots,u_s,w_2,\dots,w_s,z_1,\dots,z_t)$. Putting in u_1 and w_1 gives the required basis for V.

9.2 Skew-symmetric and alternating matrices

A matrix A is *skew-symmetric* if $A^\top=-A$.

A matrix A is *alternating* if A is skew-symmetric and has zero diagonal. If the characteristic of the field $\mathbb{K}$ is not equal to 2, then any skew-symmetric matrix is alternating; but if the characteristic is 2, then the extra condition is needed.

Recall the matrix representing a bilinear form b relative to a basis $(v_1,\dots,v_n)$: its (i,j) entry is $b(v_i,v_j)$.

Proposition 9.3 *An alternating bilinear form b on a vector space over $\mathbb{K}$ is represented by an alternating matrix; and any alternating matrix represents an alternating bilinear form. If the characteristic of $\mathbb{K}$ is not 2, we can replace "alternating matrix" by "skew-symmetric matrix".*

Proof This is obvious since if b is alternating then $a_{ji} = b(v_j, v_i) = -b(v_i, v_j) = -a_{ij}$ and $a_{ii} = b(v_i, v_i) = 0$.

So we can write our theorem in matrix form as follows:

Theorem 9.4 *Let A be an alternating matrix (or a skew-symmetric matrix over a field whose characteristic is not equal to 2). Then there is an invertible matrix P such that $P^\top AP$ is the matrix with s blocks $\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$ on the diagonal and all other entries zero. Moreover the number s is half the rank of A, and so is independent of the choice of P.*

Proof We know that the effect of a change of basis with transition matrix P is to replace the matrix A representing a bilinear form by $P^\top AP$. Also, the matrix in the statement of the theorem is just the matrix representing b relative to the special basis that we found in the preceding theorem.

This has a corollary which is a bit surprising at first sight:

Corollary 9.5 *(a) The rank of a skew-symmetric matrix (over a field of characteristic not equal to 2) is even.*

(b) The determinant of a skew-symmetric matrix (over a field of characteristic not equal to 2) is a square, and is zero if the size of the matrix is odd.

Proof (a) The canonical form in the theorem clearly has rank $2s$.

(b) If the skew-symmetric matrix A is singular then its determinant is zero, which is a square. So suppose that it is invertible. Then its canonical form has $s = n/2$ blocks $\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$ on the diagonal. Each of these blocks has determinant 1, and hence so does the whole matrix. So $\det(P^\top AP) = \det(P)^2 \det(A) = 1$, whence $\det(A) = 1/(\det(P)^2)$, which is a square.

If the size n of A is odd, then the rank cannot be n (by (a)), and so $\det(A) = 0$.

Remark There is a function defined on skew-symmetric matrices called the *Pfaffian*, which like the determinant is a polynomial in the matrix entries, and has the property that $\det(A)$ is the square of the Pfaffian of A: that is, $\det(A) = (\mathrm{Pf}(A))^2$.

For example,

$$\mathrm{Pf}\begin{bmatrix} 0 & a \\ -a & 0 \end{bmatrix} = a, \qquad \mathrm{Pf}\begin{bmatrix} 0 & a & b & c \\ -a & 0 & d & e \\ -b & -d & 0 & f \\ -c & -e & -f & 0 \end{bmatrix} = af - be + cd.$$

(Check that the determinant of the second matrix is $(af - be + cd)^2$.)

9.3 Complex skew-Hermitian matrices

What if we play the same variation that led us from real symmetric to complex Hermitian matrices? That is, we are working in a complex inner product space, and if α is represented by the matrix A, then its adjoint is represented by $\overline{A}^\top$, the conjugate transpose of A.

The matrix A is Hermitian if it is equal to its adjoint, that is, if $\overline{A}^\top = A$. So we make the following definition:

Definition 9.2 The complex $n \times n$ matrix A is *skew-Hermitian* if $\overline{A}^\top = -A$.

Actually, things are very much simpler here, because of the following observation:

Proposition 9.6 *The matrix A is skew-Hermitian if and only if $\mathrm{i}A$ is Hermitian.*

Proof Try it and see!

Corollary 9.7 *Any skew-Hermitian matrix can be diagonalised by a unitary matrix.*

Proof This follows immediately from the Proposition preceding.

Alternatively, a skew-Hermitian matrix is obviously normal, and the Corollary follows from our result about normal matrices (Theorem 8.5).

Since the eigenvalues of a Hermitian matrix are real, we see that the eigenvalues of a skew-Hermitian matrix are imaginary.

Appendix A

Fields and vector spaces

Fields

A *field* is an algebraic structure $\mathbb{K}$ in which we can add and multiply elements, such that the following laws hold:

Addition laws

(FA0) For any $a, b \in \mathbb{K}$, there is a unique element $a+b \in \mathbb{K}$.

(FA1) For all $a, b, c \in \mathbb{K}$, we have $a+(b+c) = (a+b)+c$.

(FA2) There is an element $0 \in \mathbb{K}$ such that $a+0 = 0+a = a$ for all $a \in \mathbb{K}$.

(FA3) For any $a \in K$, there exists $-a \in \mathbb{K}$ such that $a+(-a) = (-a)+a = 0$.

(FA4) For any $a, b \in \mathbb{K}$, we have $a+b = b+a$.

Multiplication laws

(FM0) For any $a, b \in \mathbb{K}$, there is a unique element $ab \in \mathbb{K}$.

(FM1) For all $a, b, c \in \mathbb{K}$, we have $a(bc) = (ab)c$.

(FM2) There is an element $1 \in \mathbb{K}$, not equal to the element 0 from (FA2), such that $a1 = 1a = a$ for all $a \in \mathbb{K}$.

(FM3) For any $a \in \mathbb{K}$ with $a \neq 0$, there exists $a^{-1} \in \mathbb{K}$ such that $aa^{-1} = a^{-1}a = 1$.

(FM4) For any $a, b \in \mathbb{K}$, we have $ab = ba$.

Distributive law

(D) For all $a, b, c \in \mathbb{K}$, we have $a(b+c) = ab+ac$.

Note the similarity of the addition and multiplication laws. We say that $(\mathbb{K},+)$ is an *abelian group* if (FA0)–(FA4) hold. Then (FM0)–(FM4) say that $(\mathbb{K}\setminus\{0\},\cdot)$ is also an abelian group. (We have to leave out 0 because, as (FM3) says, 0 does not have a multiplicative inverse.)

Examples of fields include $\mathbb{Q}$ (the rational numbers), $\mathbb{R}$ (the real numbers), $\mathbb{C}$ (the complex numbers), and $\mathbb{F}_p$ (the integers mod p, for p a prime number).

Associated with any field $\mathbb{K}$ there is a non-negative integer called its *characteristic*, defined as follows. If there is a positive integer n such that $1+1+\cdots+1=0$, where there are n ones in the sum, then the smallest such n is prime. (For if $n=rs$, with $r,s>1$, and we denote the sum of n ones by $n\cdot 1$, then

$$0 = n\cdot 1 = (r\cdot 1)(s\cdot 1);$$

by minimality of n, neither of the factors $r\cdot 1$ and $s\cdot 1$ is zero. But in a field, the product of two non-zero elements is non-zero.) If so, then this prime number is the characteristic of $\mathbb{K}$. If no such n exists, we say that the characteristic of $\mathbb{K}$ is zero.

For our important examples, $\mathbb{Q}$, $\mathbb{R}$ and $\mathbb{C}$ all have characteristic zero, while $\mathbb{F}_p$ has characteristic p.

Vector spaces

Let $\mathbb{K}$ be a field. A *vector space* V over $\mathbb{K}$ is an algebraic structure in which we can add two elements of V, and multiply an element of V by an element of $\mathbb{K}$ (this is called *scalar multiplication*), such that the following rules hold:

Addition laws

(VA0) For any $u,v\in V$, there is a unique element $u+v\in V$.

(VA1) For all $u,v,w\in V$, we have $u+(v+w)=(u+v)+w$.

(VA2) There is an element $0\in V$ such that $v+0=0+v=av$ for all $v\in V$.

(VA3) For any $v\in V$, there exists $-v\in V$ such that $v+(-v)=(-v)+v=0$.

(VA4) For any $u,v\in V$, we have $u+v=v+u$.

Scalar multiplication laws

(VM0) For any $a\in\mathbb{K}$, $v\in V$, there is a unique element $av\in V$.

(VM1) For any $a\in\mathbb{K}$, $u,v\in V$, we have $a(u+v)=au+av$.

(VM2) For any $a,b\in\mathbb{K}$, $v\in V$, we have $(a+b)v=av+bv$.

(VM3) For any $a,b\in\mathbb{K}$, $v\in V$, we have $(ab)v=a(bv)$.

(VM4) For any $v \in V$, we have $1v = v$ (where 1 is the element given by (FM2)).

Again, we can summarise (VA0)–(VA4) by saying that $(V, +)$ is an abelian group.

The most important example of a vector space over a field $\mathbb{K}$ is the set $\mathbb{K}^n$ of all n-tuples of elements of $\mathbb{K}$: the addition and scalar multiplication are defined by the rules

$$\begin{aligned}(u_1, u_2, \ldots, u_n) + (v_1, v_2, \ldots, v_n) &= (u_1 + v_1, u_2 + v_2, \ldots, u_n + v_n),\\ a(v_1, v_2, \ldots, v_n) &= (av_1, av_2, \ldots, av_n).\end{aligned}$$

The fact that $\mathbb{K}^n$ is a vector space will be assumed here. Proofs are straightforward but somewhat tedious. Here is a particularly easy one, the proof of (VM4), as an example.

If $v = (v_1, \ldots, v_n)$, then

$$1v = 1(v_1, \ldots, v_n) = (1v_1, \ldots, 1v_n) = (v_1, \ldots, v_n) = v.$$

The second step uses the definition of scalar multiplication in K^n, and the third step uses the field axiom (FM2).

Appendix B

Vandermonde and circulant matrices

The *Vandermonde matrix* $V(a_1, a_2, \ldots, a_n)$ is the $n \times n$ matrix

$$\begin{bmatrix} 1 & 1 & \ldots & 1 \\ a_1 & a_2 & \ldots & a_n \\ a_1^2 & a_2^2 & \ldots & a_n^2 \\ & & \ldots & \\ a_1^{n-1} & a_2^{n-1} & \ldots & a_n^{n-1} \end{bmatrix}.$$

This is a particularly important type of matrix. We can write down its determinant explicitly:

Theorem B.1

$$\det(V(a_1, a_2, \ldots, a_n)) = \prod_{i<j}(a_j - a_i).$$

That is, the determinant is the product of the differences between all pairs of parameters a_i. From this theorem, we draw the following conclusion:

Corollary B.2 *The matrix $V(a_1, a_2, \ldots, a_n)$ is invertible if and only if the parameters $a_1, a_2, \ldots, a_n$ are all distinct.*

For the determinant can be zero only if one of the factors vanishes.

Proof To prove the theorem, we first regard a_n as a variable x, so that the determinant Δ is a polynomial $f(x)$ of degree $n-1$ in x. We see that $f(a_i) = 0$ for $1 \leq i \leq n-1$, since the result is the determinant of a matrix with two equal columns. By the Factor Theorem,

$$\Delta = K(x - a_1)(x - a_2)\cdots(x - a_{n-1}),$$

where K is independent of x. In other words, the original determinant is $K(a_n - a_1)\cdots(a_n - a_{n-1})$. In the same way, all differences $(a_j - a_i)$ for $i < j$ are factors, so that the determinant is K_0 times the product of all these differences, where K_0 does not contain any of $a_1, \ldots, a_n$, that is, K_0 is a constant.

To find K_0, we observe that the leading diagonal of the matrix gives us a term $a_2 a_3^2 \cdots a_n^{n-1}$ in the determinant with sign $+1$; but this product is obtained by taking the term with larger index from each factor in the product, also giving sign $+1$. So $K_0 = 1$ and the theorem is proved.

Another general type of matrix whose determinant can be calculated explicitly is the *circulant matrix*, whose general form is as follows:

$$C(a_0, \ldots, a_{n-1}) = \begin{bmatrix} a_0 & a_1 & a_2 & \ldots & a_{n-1} \\ a_{n-1} & a_0 & a_1 & \ldots & a_{n-2} \\ a_{n-2} & a_{n-1} & a_0 & \ldots & a_{n-3} \\ & \ldots & & \ldots & \\ a_1 & a_2 & a_3 & \ldots & a_0 \end{bmatrix}.$$

Theorem B.3 *Let $C = C(a_0, \ldots, a_{n-1})$ be a circulant matrix over the field $\mathbb{C}$. Let $\omega = \mathrm{e}^{2\pi \mathrm{i}/n}$ be a primitive nth root of unity. Then*

(a) C is diagonalisable;

(b) the eigenvalues of C are $\sum_{j=0}^{n-1} a_j \omega^{jk}$, for $k = 0, 1, \ldots, n-1$;

(c) $\det(C)$ is the product of the eigenvalues listed in (b).

Proof We can write down the eigenvectors. For $k = 0, 1, \ldots, n-1$, let $v_k = [1 \quad \omega^k \quad \ldots \quad \omega^{(n-1)k}]^\top$. The jth entry in Cv_k is

$$\begin{aligned} & a_{n-j} + a_{n-j+1}\omega^k + \cdots + a_{n-j-1}\omega^{(n-1)k} \\ = \; & a_0\omega^{jk} + \cdots + a_{n-j-1}\omega^{(n-1)k} + a_{n-j}\omega^{nk} + \cdots + a_{n-1}\omega^{(n+j-1)k} \\ = \; & \omega^{jk}(a_0 + a_1\omega^k + \cdots + a_{n-1}\omega^{(n-1)k}), \end{aligned}$$

using the fact that $\omega^n = 1$. This is $a_0 + a_1\omega^k + \cdots + a_{n-1}\omega^{(n-1)k}$ times the jth entry in v_k. So

$$Cv_k = (a_0 + a_1\omega^k + \cdots + a_{n-1}\omega^{(n-1)k})v_k,$$

as required.

Now the vectors $v_0, \ldots, v_{n-1}$ are linearly independent. (Why? They are the columns of a Vandermonde matrix $V(1, \omega, \ldots, \omega^{n-1})$, and the powers of ω are all distinct; so the first part of this appendix shows that the determinant of this

matrix is non-zero, so that the columns are linearly independent.) Hence we have diagonalised C, and its eigenvalues are as claimed.

Finally, for part (c), the determinant of a diagonalisable matrix is the product of its eigenvalues.

Example B.1 We have the identity

$$a^3+b^3+c^3-3abc=\begin{vmatrix} a & b & c \\ c & a & b \\ b & c & a \end{vmatrix} = (a+b+c)(a+\omega b+\omega^2 c)(a+\omega^2 b+\omega c),$$

where $\omega = \mathrm{e}^{2\pi \mathrm{i}/3}$.

This formula has an application to solving cubic equations. Consider the equation

$$x^3+ax^2+bx+c=0.$$

By "completing the cube", putting $y = x+\frac{1}{3}a$, we get rid of the square term:

$$y^3+dy+e=0$$

for some d,e. Now, as above, we have

$$y^3-3uvy+u^3+v^3=(y+u+v)(y+\omega u+\omega^2 v)(y+\omega^2 u+\omega v),$$

so if we can find u and v satisfying $-3uv=d$ and $u^3+v^3=e$, then the solutions of the equation are $y=-u-v$, $y=-\omega u-\omega^2 v$, and $y=-\omega^2 u-\omega v$.

Let $U=u^3$ and $V=v^3$. Then $U+V=e$ and $UV=-d^3/27$. Thus we can find U and V by solving the quadratic equation $z^2-ez-d^3/27=0$. Now u is a cube root of U, and then $v=-d/(3u)$, and we are done.

Remark The formula for the determinant of a circulant matrix works over any field $\mathbb{K}$ which contains a primitive nth root of unity.

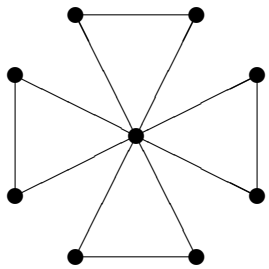
Appendix C

The Friendship Theorem

The *Friendship Theorem* states:

> *Given a finite set of people with the property that any two have a unique common friend, there must be someone who is everyone else's friend.*

The theorem asserts that the configuration must look like this, where we represent people by dots and friendship by edges:

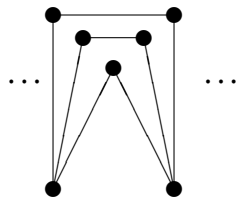


The proof of the theorem is in two parts. The first part is "graph theory", the second uses linear algebra. We argue by contradiction, and so we assume that we have a counterexample to the theorem.

Step 1: Graph theory We show that there is a number m such that everyone has exactly m friends. [In the terminology of graph theory, this says that we have a regular graph of valency m.]

To prove this, we notice first that if P_1 and P_2 are not friends, then they have the same number of friends. For they have one common friend P_3; any further

friend Q of P_1 has a common friend R with P_2, and conversely, so we can match up the common friends as in the next picture.



Now let us suppose that there are two people P and Q who have different numbers of friends. By the preceding argument, P and Q must be friends. They have a common friend R. Any other person S must have a different number of friends from either P or Q, and so must be the friend of either P or Q (but not both). Now if S is the friend of P but not Q, and T is the friend of Q but not P, then any possible choice of the common friend of S and T leads to a contradiction. So this is not possible; that is, either everyone else is P's friend, or everyone else is Q's friend. But this means that we don't have a counterexample after all.

So we conclude this step knowing that the number of friends of each person is the same, say m, as claimed.

Step 2: Linear algebra We prove that $m = 2$.

Suppose that there are n people $P_1, \ldots, P_n$. Let A be the $n \times n$ matrix whose (i, j) entry is 1 if P_i and P_j are friends, and is 0 otherwise. Then by assumption, A is an $n \times n$ symmetric matrix. Let J be the $n \times n$ matrix with every entry equal to 1; then J is also symmetric.

Consider the product AJ. Since every entry of J is equal to 1, the (i, j) entry of AJ is just the number of ones in the ith row of A, which is the number of friends of P_i; this is m, by Step 1. So every entry of AJ is m, whence $AJ = mJ$. Similarly, $JA = mJ$. Thus, A and J are commuting symmetric matrices, and so by Theorem 7.9, they can be simultaneously diagonalised. We will calculate their eigenvalues.

First let us consider J. If j is the column vector with all entries 1, then clearly $Jj = nj$, so j is an eigenvector of J with eigenvalue n. The other eigenvalues of J are orthogonal to j. Now $v \cdot j = 0$ means that the sum of the components of v is zero; this implies that $Jv = 0$. So any vector orthogonal to j is an eigenvector of J with eigenvalue 0.

Now we turn to A, and observe that

$$A^2 = (m-1)I + J.$$

For the (i, j) entry of A^2 is equal to the number of people P_k who are friends of both P_i and P_j. If $i = j$, this number is m, while if $i \neq j$ then (by assumption) it is 1. So A^2 has diagonal entries m and off-diagonal entries 1, so it is equal to $(m-1)I + J$, as claimed.

The all-one vector j satisfies $Aj = mj$, so is an eigenvector of A with eigenvalue m. This shows, in particular, that

$$m^2 j = A^2 j = ((m-1)I + J)j = (m-1+n)j,$$

so that $n = m^2 - m + 1$. (**Exercise:** Prove this by a counting argument in the graph.)

As before, the remaining eigenvectors of A are orthogonal to j, and so are eigenvectors of J with eigenvalue 0. Thus, if v is an eigenvector of A with eigenvalue λ, not a multiple of j, then

$$\lambda^2 v = A^2 v = ((m-1)I + J)v = (m-1)v,$$

so $\lambda^2 = m-1$, and $\lambda = \pm\sqrt{m-1}$.

The diagonal entries of A are all zero, so its trace is zero. So if we let f and g be the multiplicities of $\sqrt{m-1}$ and $-\sqrt{m-1}$ as eigenvalues of A, we have

$$0 = \mathrm{Tr}(A) = m + f\sqrt{m-1} + g(-\sqrt{m-1}) = m + (f-g)\sqrt{m-1}.$$

This shows that $m-1$ must be a perfect square, say $m-1 = u^2$, from which we see that m is congruent to 1 mod u. But the trace equation is $0 = m + (f-g)u$; this says that $0 \equiv 1 \bmod u$. This is only possible if $u = 1$. But then $m = 2$, $n = 3$, and we have the Three Musketeers (three individuals, any two being friends). This configuration does indeed satisfy the hypotheses of the Friendship Theorem; but it is after all not a counterexample, since each person is everyone else's friend. So the theorem is proved.

Appendix D

Who is top of the league?

In most league competitions, teams are awarded a fixed number of points for a win or a draw. It may happen that two teams win the same number of matches and so are equal on points, but the opponents beaten by one team are clearly "better" than those beaten by the other. How can we take this into account?

You might think of giving each team a "score" to indicate how strong it is, and then adding the scores of all the teams beaten by team T to see how well T has performed. Of course this is self-referential, since the score of T depends on the scores of the teams that T beats. So suppose we ask simply that the score of T should be proportional to the sum of the scores of all the teams beaten by T.

Now we can translate the problem into linear algebra. Let $T_1, \ldots, T_n$ be the teams in the league. Let A be the $n \times n$ matrix whose (i, j) entry is equal to 1 if T_I beats T_j, and 0 otherwise. Now for any vector $[x_1 \quad x_2 \quad \ldots \quad x_n]^\top$ of scores, the ith entry of Ax is equal to the sum of the scores x_j for all teams T_j beaten by T_i. So our requirement is simply that

x should be an eigenvector of A with all entries positive.

Here is an example. There are six teams A, B, C, D, E, and F. Suppose that

A beats B, C, D, E;

B beats C, D, E, F;

C beats D, E, F;

D beats E, F;

E beats F;

F beats A.

The matrix A is

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

We see that A and B each have four wins, but that A has generally beaten the stronger teams; there was one upset when F beat A. Also, E and F have the fewest wins, but F took A's scalp and should clearly be better.

Calculation with Maple shows that the vector

$$[0.7744 \quad 0.6452 \quad 0.4307 \quad 0.2875 \quad 0.1920 \quad 0.3856]^\top$$

is an eigenvector of A with eigenvalue 2.0085. This confirms our view that A is top of the league and that F is ahead of E; it even puts F ahead of D.

But perhaps there is a different eigenvalue and/or eigenvector which would give us a different result?

In fact, there is a general theorem called the *Perron–Frobenius theorem* which gives us conditions for this method to give a unique answer. Before we state it, we need a definition.

Definition D.1 Let A be an $n \times n$ real matrix with all its entries non-negative. We say that A is *indecomposable* if, for any i, j with $1 \leq i, j \leq n$, there is a number m such that the (i, j) entry of A^m is strictly positive.

This odd-looking condition means, in our football league situation, that for any two teams T_i and T_j, there is a chain $T_{k_0}, \ldots, T_{k_m}$ with $T_{k_0} = T_i$ and $T_{k_m} = T_j$, sich that each team in the chain beats the next one. Now it can be shown that the only way that this can fail is if there is a collection C of teams such that each team in C beats each team not in C. In this case, obviously the teams in C occupy the top places in the league, and we have reduced the problem to ordering these teams. So we can assume that the matrix of results is indecomposable.

In our example, we see that B beats F beats A, so the $(2,1)$ entry in A^2 is non-zero. Similarly for all other pairs. So A is indecomposable in this case.

Theorem D.1 (Perron–Frobenius Theorem) *Let A be a $n \times n$ real matrix with all its entries non-negative, and suppose that A is indecomposable. Then, up to scalar multiplication, there is a unique eigenvector $v = [x_1 \quad \ldots \quad x_n]^\top$ for A with the property that $x_i > 0$ for all i. The corresponding eigenvalue is the largest eigenvalue of A.*

So the Perron–Frobenius eigenvector solves the problem of ordering the teams in the league.

Remark Sometimes even this extra level of sophistication doesn't guarantee a result. Suppose, for example, that there are five teams A, B, C, D, E; and suppose that A beats B and C, B beats C and D, C beats D and E, D beats E and A, and E beats A and B. Each team wins two games, so the simple rule gives them all the same score. The matrix A is

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

which is easily seen to be indecomposable; and if v is the all-1 vector, then $Av = 2v$, so that v is the Perron–Frobenius eigenvector. So even with this method, all teams get the same score. In this case, it is clear that there is so much symmetry between the teams that none can be put above the others by any possible rule.

Remark Further refinements are clearly possible. For example, instead of just putting the (i, j) entry equal to 1 if T_i beats T_j, we could take it to be the number of goals by which T_i won the game.

Remark This procedure has wider application. How does an Internet search engine like Google find the most important web pages that match a given query? An important web page is one to which a lot of other web pages link; this can be described by a matrix, and we can use the Perron–Frobenius eigenvector to do the ranking.

Appendix E

Other canonical forms

One of the unfortunate things about linear algebra is that there are many types of equivalence relation on matrices! In this appendix I say a few brief words about some that we have not seen elsewhere in the course. Some of these will be familiar to you from earlier linear algebra courses, while others arise in courses on different parts of mathematics (coding theory, group theory, etc.)

Row-equivalence

Two matrices A and B of the same size over $\mathbb{K}$ are said to be *row-equivalent* if there is an invertible matrix P such that $B = PA$. Equivalently, A and B are row-equivalent if we can transform A into B by the use of elementary row operations only. (This is true because any invertible matrix can be written as a product of elementary matrices; see Corollary 2.6.)

A matrix A is said to be in *echelon form* if the following conditions hold:

- The first non-zero entry in any row (if it exists) is equal to 1 (these entries are called the *leading ones*);
- The leading ones in rows lower in the matrix occur further to the right.

We say that A is in *reduced echelon form* if, in addition to these two conditions, also

- All the other entries in the column containing a leading one are zero.

For example, the matrix

$$\begin{bmatrix} 0 & 1 & a & b & 0 & c \\ 0 & 0 & 0 & 0 & 1 & d \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

is in reduced echelon form, whatever the values of $a, \ldots, e$.

Theorem E.1 *Any matrix is row-equivalent to a unique matrix in reduced echelon form.*

Coding equivalence

In the theory of error-correcting codes, we meet a notion of equivalence which lies somewhere between row-equivalence and equivalence. As far as I know it does not have a standard name.

Two matrices A and B of the same size are said to be *coding-equivalent* if B can be obtained from A by a combination of arbitrary row operations and column operations of Types 2 and 3 only. (See page 16).

Using these operations, any matrix can be put into block form $\begin{bmatrix} I_r & A \\ O & O \end{bmatrix}$, for some matrix A. To see this, use row operations to put the matrix into reduced echelon form, then column permutations to move the columns containing the leading ones to the front of the matrix.

Unfortunately this is not a canonical form; a matrix can be coding-equivalent to several different matrices of this special form.

It would take us too far afield to explain why this equivalence relation is important in coding theory.

Congruence over other fields

Recall that two symmetric matrices A and B, over a field $\mathbb{K}$ whose characteristic is not 2, are *congruent* if $B = P^\top AP$ for some invertible matrix P. This is the natural relation arising from representing a quadratic form relative to different bases.

We saw in Chapter 5 the canonical form for this relation in the cases when $\mathbb{K}$ is the real or complex numbers.

In other cases, it is usually much harder to come up with a canonical form. Here is one of the few cases where this is possible. I state the result for quadratic forms.

Theorem E.2 *Let $\mathbb{F}_p$ be the field of integers mod p, where p is an odd prime. Let c be a fixed element of $\mathbb{F}_p$ which is not a square. A quadratic form q in n variables over $\mathbb{F}_p$ can be put into one of the forms*

$$x_1^2 + \cdots + x_r^2, \qquad x_1^2 + \cdots + x_{r-1}^2 + cx_r^2$$

by an invertible linear change of variables. Any quadratic form is congruent to just one form of one of these types.

Appendix F

Worked examples

1. Let
$$A = \begin{bmatrix} 1 & 2 & 4 & -1 & 5 \\ 1 & 2 & 3 & -1 & 3 \\ -1 & -2 & 0 & 1 & 3 \end{bmatrix}.$$

   (a) Find a basis for the row space of A.

   (b) What is the rank of A?

   (c) Find a basis for the column space of A.

   (d) Find invertible matrices P and Q such that PAQ is in the canonical form for equivalence.

(a) Subtract the first row from the second, add the first row to the third, then multiply the new second row by -1 and subtract four times this row from the third, to get the matrix

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

The first two rows clearly form a basis for the row space.

(b) The rank is 2, since there is a basis with two elements.

(c) The column rank is equal to the row rank and so is also equal to 2. By inspection, the first and third columns of A are linearly independent, so they form a basis. The first and second columns are not linearly independent, so we cannot use these! (Note that we have to go back to the original A here; row operations change the column space, so selecting two independent columns of B would not be correct.)

(d) By step (a), we have $PA = B$, where P is obtained by performing the same elementary row operations on the 3×3 identity matrix I_3:

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

Now B can be brought to the canonical form

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

by subtracting 2, 4, -1 and 5 times the first column from the second, third, fourth and fifth columns, and twice the third column from the fifth, and then swapping the second and third columns; so $C = BQ$ (whence $C = PAQ$), where Q is obtained by performing the same column operations on I_5:

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

Remark: P and Q can also be found by multiplying elementary matrices, if desired; but the above method is simpler. You may find it easier to write an identity matrix after A and perform the row operations on the extended matrix to find P, and to put an identity matrix underneath B and perform the column operations on the extended matrix to find Q.

2. A certain country has n political parties $\mathscr{P}_1, \ldots, \mathscr{P}_n$. At the beginning of the year, the percentage of voters who supported the party $\mathscr{P}_i$ was x_i. During the year, some voters change their minds; a proportion a_{ij} of former supporters of $\mathscr{P}_j$ will support $\mathscr{P}_i$ at the end of the year.

Let v be the vector $\begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}^\top$ recording support for the parties at the beginning of the year, and A the matrix whose (i, j) entry is a_{ij}.

(a) Prove that the vector giving the support for the parties at the end of the year is Av.

(b) In subsequent years, exactly the same thing happens, with the same proportions. Show that the vector giving the support for the parties at the end of m years is $A^m v$.

(c) Suppose that $n = 2$ and that

$$A = \begin{bmatrix} 0.9 & 0.3 \\ 0.1 & 0.7 \end{bmatrix}.$$

Show that, after a long time, the support for the parties will be approximately 0.75 for $\mathscr{P}_1$ to 0.25 for $\mathscr{P}_2$.

(a) Let y_i be the proportion of the population who support $\mathscr{P}_i$ at the end of the year. From what we are given, the proportion supporting $\mathscr{P}_j$ at the beginning of the year was x_j, and a fraction a_{ij} of these changed their support to $\mathscr{P}_i$. So the proportion of the whole population who supported $\mathscr{P}_j$ at the beginning of the year and $\mathscr{P}_i$ at the end is $a_{ij}x_j$. The total support for $\mathscr{P}_i$ is found by adding these up for all j: that is,

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

or $v' = Av$, where v' is the vector $[y_1 \quad \dots \quad y_n]^\top$ giving support for the parties at the end of the year.

(b) Let v_k be the column vector whose ith component is the proportion of the population supporting party $\mathscr{P}_i$ after the end of k years. In part (a), we showed that $v_1 = Av_0$, where $v_0 = v$. An exactly similar argument shows that $v_k = Av_{k-1}$ for any k. So by induction, $v_m = P^m v_0 = P^m v$, as required. (The result of (a) starts the induction with $m = 1$. If we assume that $v_{k-1} = A^{k-1}v$, then

$$v_k = Av_{k-1} = A(A^{k-1}v) = A^k v,$$

and the induction step is proved.)

(c) The matrix P has characteristic polynomial

$$\begin{vmatrix} x-0.9 & -0.3 \\ -0.7 & x-0.7 \end{vmatrix} = x^2 - 1.6x + 0.6 = (x-1)(x-0.6).$$

So the eigenvalues of P are 1 and 0.6. We find by solving linear equations that eigenvectors for the two eigenvalues are $\begin{bmatrix} 3 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} 1 \\ -1 \end{bmatrix}$ respectively. As in the text, we compute that the corresponding projections are

$$P_1 = \begin{bmatrix} 0.75 & 0.75 \\ 0.25 & 0.25 \end{bmatrix}, \qquad P_2 = \begin{bmatrix} 0.25 & -0.75 \\ -0.25 & 0.75 \end{bmatrix}.$$

(Once we have found P_1, we can find P_2 as $I - P_1$.) Then P is diagonalisable:

$$A = P_1 + 0.6P_2.$$

From this and Proposition 4.6 we see that

$$A^m = P_1 + (0.6)^m P_2.$$

As $m \to \infty$, we have $(0.6)^m \to 0$, and so $A \to P_1$. So in the limit, if $v_0 = \begin{bmatrix} x \\ y \end{bmatrix}$ is the matrix giving the initial support for the parties, with $x + y = 1$, then the matrix giving the final support is approximately

$$\begin{bmatrix} 0.75 & 0.75 \\ 0.25 & 0.25 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0.75(x+y) \\ 0.25(x+y) \end{bmatrix} = \begin{bmatrix} 0.75 \\ 0.25 \end{bmatrix}.$$

As a check, use the computer with Maple to work out P^m for some large value of m. For example, I find that

$$P^{10} = \begin{bmatrix} 0.7515116544 & 0.7454650368 \\ 0.2484883456 & 0.2545349632 \end{bmatrix}.$$

3. The vectors v_1, v_2, v_3 form a basis for $V = \mathbb{R}^3$; the dual basis of V^* is f_1, f_2, f_3. A second basis for V is given by $w_1 = v_1 + v_2 + v_3$, $w_2 = 2v_1 + v_2 + v_3$, $w_3 = 2v_2 + v_3$. Find the basis of V^* dual to w_1, w_2, w_3.

The first dual basis vector g_1 satisfies $g_1(w_1) = 1$, $g_1(w_2) = g_1(w_3) = 0$. If $g_1 = xf_1 + yf_2 + zf_3$, we find

$$\begin{aligned} x + y + z &= 1, \\ 2x + y + z &= 0, \\ 2y + z &= 0, \end{aligned}$$

giving $x = -1$, $y = -2$, $z = 4$. So $g_1 = -f_1 - 2f_2 + 4f_3$. Solving two similar sets of equations gives $g_2 = f_1 + f_2 - 2f_3$ and $g_3 = f_2 - f_3$.

Alternatively, the transition matrix P from the vs to the ws is

$$P = \begin{bmatrix} 1 & 2 & 0 \\ 1 & 1 & 2 \\ 1 & 1 & 1 \end{bmatrix},$$

and we showed in Section 5.1.2 that the transition matrix between the dual bases is

$$(P^{-1})^\top = \begin{bmatrix} -1 & 1 & 0 \\ -2 & 1 & 1 \\ 4 & -2 & -1 \end{bmatrix}.$$

The coordinates of the gs in the basis of fs are the columns of this matrix.

4. The *Fibonacci numbers* F_n are defined by the recurrence relation

$$F_0 = 0, \qquad F_1 = 1, \qquad F_{n+2} = F_n + F_{n+1} \text{ for } n \geq 0.$$

Let A be the matrix $\begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}$. Prove that

$$A^n = \begin{bmatrix} F_{n-1} & F_n \\ F_n & F_{n+1} \end{bmatrix},$$

and hence find a formula for F_n.

The equation for F_n is proved by induction on n. It is clearly true for $n = 1$. Suppose that it holds for n; then

$$A^{n+1} = A^n \cdot A = \begin{bmatrix} F_{n-1} & F_n \\ F_n & F_{n+1} \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} F_n & F_{n-1}+F_n \\ F_{n+1} & F_n + F_{n+1} \end{bmatrix} = \begin{bmatrix} F_n & F_{n+1} \\ F_{n+1} & F_{n+2} \end{bmatrix}.$$

So the induction step is proved.

To find a formula for F_n, we show that A is diagonalisable, and then write $A = \lambda_1 P_1 + \lambda_2 P_2$, where P_1 and P_1 are projection matrices with sum I satisfying $P_1 P_2 = P_2 P_1 = 0$. Then we get $A^n = \lambda_1^n P_1 + \lambda_2^n P_2$, and taking the $(1,2)$ entry we find that

$$F_n = c_1 \lambda_1^n + c_2 \lambda_2^n,$$

where c_1 and c_2 are the $(1,2)$ entries of P_1 and P_2 respectively.

From here it is just calculation. The eigenvalues of A are the roots of $0 = \det(xI - A) = x^2 - x - 1$; that is, $\lambda_1, \lambda_2 = \frac{1}{2}(1 \pm \sqrt{5})$. (Since the eigenvalues are distinct, we know that A is diagonalisable, so the method will work.) Now because $P_1 + P_2 = I$, the $(1,2)$ entries of these matrices are the negatives of each other; so we have $F_n = c(\lambda_1^n - \lambda_2^n)$. Rather than find P_1 explicitly, we can now argue as follows: $1 = F_1 = c(\lambda_1 - \lambda_2) = c\sqrt{5}$, so that $c = 1/\sqrt{5}$ and

$$F_n = \frac{1}{\sqrt{5}} \left(\left(\frac{1+\sqrt{5}}{2} \right)^n - \left(\frac{1-\sqrt{5}}{2} \right)^n \right).$$

5. Let V_n be the vector space of real polynomials of degree at most n.

(a) Show that the function

$$f \cdot g = \int_0^1 f(x) g(x) \, \mathrm{d}x$$

is an inner product on V_n.

(b) In the case $n = 3$, write down the matrix representing the bilinear form relative to the basis $1, x, x^2, x^3$ for V_3.

(c) Apply the Gram–Schmidt process to the basis $(1, x, x^2)$ to find an orthonormal basis for V_2.

(d) Let W_n be the subspace of V_n consisting of all polynomials $f(x)$ of degree at most n which satisfy $f(0) = f(1) = 0$. Let $D : W_n \to W_n$ be the linear map given by differentiation: $(Df)(x) = f'(x)$. Prove that the adjoint of D is $-D$.

(a) Put $b(f,g) = \int_0^1 f(x)g(x)\,\mathrm{d}x$. The function b is obviously symmetric. So we have to show that it is linear in the first variable, that is, that

$$\begin{aligned}\int_0^1 (f_1(x) + f_2(x))g(x)\,\mathrm{d}x &= \int_0^1 f_1(x)g(x)\,\mathrm{d}x + \int_0^1 f_2(x)g(x)\,\mathrm{d}x,\\ \int_0^1 (cf(x))g(x)\,\mathrm{d}x &= c\int_0^1 f(x)g(x)\,\mathrm{d}x,\end{aligned}$$

which are clear from elementary calculus.

We also have to show that the inner product is positive definite, that is, that $b(f,f) \geq 0$, with equality if and only if $f = 0$. This is clear from properties of integration.

(b) If the basis is $f_1 = 1$, $f_2 = x$, $f_3 = x^2$, $f_4 = x^3$, then the (i,j) entry of the matrix representing b is

$$\int_0^1 x^{i-1}x^{j-1}\,\mathrm{d}x = \frac{1}{i+j-1},$$

so the matrix is

$$\begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{3} & \frac{1}{4} \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{4} & \frac{1}{5} \\ \frac{1}{3} & \frac{1}{4} & \frac{1}{5} & \frac{1}{6} \\ \frac{1}{4} & \frac{1}{5} & \frac{1}{6} & \frac{1}{7} \end{bmatrix}.$$

(c) The first basis vector is clearly 1. To make x orthogonal to 1 we must replace it by $x + a$ for some a; doing the integral we find that $a = -\frac{1}{2}$. To make x^2 orthogonal to the two preceding is the same as making it orthogonal to 1 and x, so we replace it by $x^2 + bx + c$; we find that

$$\begin{aligned}\tfrac{1}{3} + \tfrac{1}{2}b + c &= 0,\\ \tfrac{1}{4} + \tfrac{1}{3}b + \tfrac{1}{2}c &= 0,\end{aligned}$$

so that $b = -1$ and $c = \frac{1}{6}$.

Now $1 \cdot 1 = 1$, $(x-\frac{1}{2}) \cdot (x-\frac{1}{2}) = \frac{1}{12}$, and $(x^2-x+\frac{1}{6}) \cdot (x^2-x+\frac{1}{6}) = \frac{1}{180}$; so the required basis is

$$1, \quad \frac{1}{2\sqrt{3}}(x-\frac{1}{2}), \quad \frac{1}{6\sqrt{5}}(x^2-x+\frac{1}{6}).$$

(d) Integration by parts shows that

$$\begin{aligned} f \cdot D(g) &= \int_0^1 f(x)g'(x)\,\mathrm{d}x \\ &= [f(x)g(x)]_0^1 - \int_0^1 f'(x)g(x)\,\mathrm{d}x \\ &= -(Df) \cdot g, \end{aligned}$$

where the first term vanishes because of the condition on polynomials in W_n. Thus, by definition, the adjoint of D is $-D$.

6. Let A and B be real symmetric matrices. Is each of the following statements true or false? Give brief reasons.

(a) If A and B are orthogonally similar then they are congruent.

(b) If A and B are orthogonally similar then they are similar.

(c) If A and B are congruent then they are orthogonally similar.

(d) If A and B are similar then they are orthogonally similar.

Recall that A and B are similar if $B = P^{-1}AP$ for some invertible matrix P; they are congruent if $B = P^\top AP$ for some invertible matrix P; and they are orthogonally similar if $B = P^{-1}AP$ for some orthogonal matrix P (invertible matrix satisfying $P^\top = P^{-1}$). Thus it is clear that both (a) and (b) are true.

The Spectral Theorem says that A is orthogonally congruent to a diagonal matrix whose diagonal entries are the eigenvalues. If A and B are similar, then they have the same eigenvalues, and so are orthogonally congruent to the same diagonal matrix, and so to each other. So (d) is true.

By Sylvester's Law of Inertia, any real symmetric matrix is congruent to a diagonal matrix with diagonal entries 1, -1 and 0. If we choose a symmetric matrix none of whose eigenvalues is 1, -1 or 0, then it is not orthogonally similar to the Sylvester form. For example, the matrices I and $2I$ are congruent but not orthogonally similar. So (c) is false.

7. Find an orthogonal matrix P such that $P^{-1}AP$ and $P^{-1}BP$ are diagonal, where

$$A = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix}.$$

Remark A and B are commuting symmetric matrices, so we know that the matrix P exists.

First solution We have to find an orthonormal basis which consists of eigenvectors for both matrices.

Some eigenvectors can be found by inspection. If $v_1 = (1,1,1,1)$ then $Av_1 = 4v_1$ and $Bv_1 = 2v_1$. If $v_2 = (1,-1,1,-1)$ then $Av_2 = 0$ and $Bv_2 = -2v_2$. Any further eigenvector $v = (x,y,z,w)$ should be orthogonal to both of these, that is, $x+y+z+w = 0 = x-y+z-w$. So $x+z = 0$ and $y+w = 0$. Conversely, any such vector satisfies $Av = 0$ and $Bv = 0$. So choose two orthogonal vectors satisfying these conditions, say $(1,0,-1,0)$ and $(0,1,0,-1)$. Normalising, we obtain the required basis: $(1,1,1,1)/2$, $(1,-1,1,-1)/2$, $(1,0,-1,0)/\sqrt{2}$, $(0,1,0,-1)/\sqrt{2}$. So

$$P = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & \frac{1}{\sqrt{2}} & 0 \\ \frac{1}{2} & -\frac{1}{2} & 0 & \frac{1}{\sqrt{2}} \\ \frac{1}{2} & \frac{1}{2} & -\frac{1}{\sqrt{2}} & 0 \\ \frac{1}{2} & -\frac{1}{2} & 0 & -\frac{1}{\sqrt{2}} \end{bmatrix}.$$

Second solution Observe that both A and B are circulant matrices. So we know from Appendix B that the columns of the Vandermonde matrix

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & \mathrm{i} & -1 & -\mathrm{i} \\ 1 & -1 & 1 & -1 \\ 1 & -\mathrm{i} & -1 & \mathrm{i} \end{bmatrix}$$

are eigenvectors of both matrices. The second and fourth columns have corresponding eigenvalues 0 for both matrices, and hence so do any linear combinations of them; in particular, we can replace these two columns by their real and imaginary parts, giving (after a slight rearrangement) the matrix

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

After normalising the columns, this gives the same solution as the first.

The results of Appendix B also allow us to write down the eigenvalues of A and B without any calculation. For example, the eigenvalues of B are

$$1+1 = 2, \quad \mathrm{i}-\mathrm{i} = 0, \quad -1-1 = -2, \qquad -\mathrm{i}+\mathrm{i} = 0.$$

Remark A more elegant solution is the matrix

$$\frac{1}{2}\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix}.$$

This matrix (without the factor $\frac{1}{2}$) is known as a *Hadamard matrix*. It is an $n \times n$ matrix H with all entries ± 1 satisfying $H^\top H = nI$. It is known that an $n \times n$ Hadamard matrix cannot exist unless n is 1, 2, or a multiple of 4; however, nobody has succeeded in proving that a Hadamard matrix of any size n divisible by 4 exists.

The smallest order for which the existence of a Hadamard matrix is still in doubt is (at the time of writing) $n = 668$. The previous smallest, $n = 428$, was resolved only in 2004 by Hadi Kharaghani and Behruz Tayfeh-Reziae in Tehran, by constructing an example.

As a further exercise, show that, if H is a Hadamard matrix of size n, then $\begin{bmatrix} H & H \\ H & -H \end{bmatrix}$ is a Hadamard matrix of size $2n$. (The Hadamard matrix of size 4 constructed above is of this form.)

8. Let $A = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$.
Find an invertible matrix P and a diagonal matrix D such that $P^\top AP = I$ and $P^\top BP = D$, where I is the identity matrix.

First we take the quadratic form corresponding to A, and reduce it to a sum of squares. The form is $x^2 + 2xy + 2y^2$, which is $(x+y)^2 + y^2$. (*Note:* This is the sum of two squares, in agreement with the fact that A is positive definite.)

Now the matrix that transforms (x,y) to $(x+y,y)$ is $Q = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$, since

$$\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x+y \\ y \end{bmatrix}.$$

Hence

$$\begin{bmatrix} x & y \end{bmatrix} Q^\top Q \begin{bmatrix} x \\ y \end{bmatrix} = x^2 + 2xy + 2y^2 = \begin{bmatrix} x & y \end{bmatrix} A \begin{bmatrix} x \\ y \end{bmatrix},$$

so that $Q^\top Q = A$.

Now, if we put $P = Q^{-1} = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}$, we see that $P^\top AP = P^\top (Q^\top Q)P = I$.

What about $P^\top BP$? We find that

$$P^\top BP = \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} = D,$$

the required diagonal matrix. So we are done.

Remark 1: In general it is not so easy. The reduction of the quadratic form will give a matrix P_1 such that $P_1^\top AP_1 = I$, but in general $P_1^\top BP_1$ won't be diagonal; all we can say is that it is symmetric. So by the Spectral Theorem, we can find an orthogonal matrix P_2 such that $P_1^\top (P_1^\top BP_1)P_2$ is diagonal. (P_2 is the matrix whose columns are orthonormal eigenvectors of $P_1^\top BP_1$.) Then because P_2 is orthogonal, we have

$$P_2^\top (P_1^\top AP_1)P_2 = P_2^\top IP_2 = I,$$

so that $P = P_1P_2$ is the required matrix.

Remark 2: If you are only asked for the diagonal matrix D, and not the matrix P, you can do an easier calculation. We saw in the lectures that the diagonal entries of D are the roots of the polynomial $\det(xA - B) = 0$. In our case, we have

$$\begin{vmatrix} x-1 & x-1 \\ x-1 & 2x \end{vmatrix} = x^2 - 1 = (x-1)(x+1),$$

so the diagonal entries of D are $+1$ and -1 (as we found).

Index