# Printed Notes for MTH6140 Linear Algebra II

2020-21 Currently edited by Shahn Majid
(based on notes due to Peter Cameron)

# Contents

# Chapter 1

# Vector spaces

In the *Linear Algebra I* module, we encountered two kinds of vector space, namely real and complex. The real numbers and the complex numbers are both examples of an algebraic structure called a *field*, which we encountered back in the *Introduction to Algebra* module. Vector spaces can be defined relative to any field, and we shall do so here. There are practical reasons to do this beyond the desire for maximum generality. For example, vector spaces over the two-element field $\mathbb{F}_2$ arise in coding theory and in computer science.

## 1.1 Axiomatic definition of a vector space

Let's assume we remember the definition of *group* from *Introduction to Algebra*. We have to start somewhere after all. If not, you should pause to consult your notes from two years back or Wikipedia.

Fields were defined in *Introduction to Algebra*, but let's just refresh our memory.

**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;

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

(D) the *distributive law*

$$a \cdot (b + c) = a \cdot b + a \cdot c$$

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

The explicit symbol "$\cdot$" for multiplication is needed only for the purpose of making sense of the definition, and we drop it right away.

In fact, the only fields we'll encounter 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,

so if you are comfortable handling these particular fields you should be just fine. We will not stop to prove that the above structures really are fields. You may have seen $\mathbb{F}_p$ referred to as $\mathbb{Z}_p$.

Specific examples of fields have properties beyond merely satisfying the axioms listed above. You may like to pause at this point to consider properties that distinguish the fields $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$ and $\mathbb{F}_2$ that we mentioned above. (Think of cardinality, order, roots of polynomials and limits of sequences.) However, many of the theorems in this module apply to vector spaces over an arbitrary field, so, particularly in the earlier stages of the module, we should keep a careful eye on our manipulations, to ensure that we don't go beyond the field axioms! However there are a lot of axioms, and a good survival technique is to have in mind a concrete field, say $\mathbb{R}$ that we are familiar with.

Having recalled the basics from *Introduction to Algebra* we are ready to introduce the our main object of study.

**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 $\mathbf{0}$ (the *zero vector*).

(VM) Rules for scalar multiplication:

- (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*. Typically we'll use use letters around $u$, $v$, $w$ in the alphabet to stand for vectors, and letters around $a$, $b$ and $c$ for scalars. (In *Linear Algebra I*, boldface was used to distinguish vectors from scalars. We retain that convention only in the context of the zero of the field 0 and the zero vector $\mathbf{0}$.)

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*. In some sections of the course, we'll be thinking specifically of real or complex vector spaces; in others, of vector spaces over general fields. As we noted, vector spaces over other fields are very useful in some applications, for example in coding theory, combinatorics and computer science.

**Example 1.3.** 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 $a < 0$.

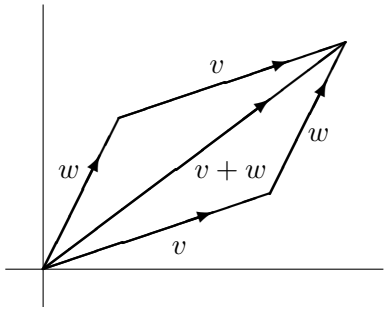


Figure 1.1: The parallelogram law

- The *algebraic* definition. We represent the points of the plane by Cartesian coordinates. Thus, a vector $v$ is just a pair $(a_1, a_2)$ of real numbers. Now we define addition and scalar multiplication by

$$(a_1, a_2) + (b_1, b_2) = (a_1 + b_1, a_2 + b_2),$$
$$c(a_1, a_2) = (ca_1, ca_2).$$

  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 may check the law $c(v + w) = cv + cw$. Let $v = (a_1, a_2)$ and $w = (b_1, b_2)$. Then we have

$$\begin{aligned} c(v + w) &= c((a_1, a_2) + (b_1, b_2)) \\ &= c(a_1 + b_1, a_2 + b_2) \\ &= (ca_1 + cb_1, ca_2 + cb_2) \\ &= (ca_1, ca_2) + (cb_1, cb_2) \\ &= cv + cw. \end{aligned}$$

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 1.4.** 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:

$$(a_1, a_2, \ldots, a_n) + (b_1, b_2, \ldots, b_n) = (a_1 + b_1, a_2 + b_2, \ldots, a_n + b_n),$$
$$c(a_1, a_2, \ldots, a_n) = (ca_1, ca_2, \ldots, ca_n).$$

**Example 1.5.** The set $\mathbb{R}^S$ of all real functions on a set $S$ is a vector space over $\mathbb{R}$. Vector addition is just addition of functions. Scalar multiplication is just scaling of a function by a real number.

**Example 1.6.** The set of all polynomials of degree $n - 1$ with coefficients in a field $\mathbb{K}$ is a vector space over $\mathbb{K}$. Vector addition is just usual addition of polynomials; scalar multiplication is just scaling of a polynomial by an element of $\mathbb{K}$. Equivalently, one can say that vector addition is coefficientwise addition, and scalar multiplication is multiplication of all coefficients by a field element. Note that from this perspective, this example is a disguised version of Example 1.4. This example was a favourite in *Linear Algebra I*!

## 1.2 Bases

Example 1.4 is much more general than it appears: *Every finite-dimensional vector space looks like Example 1.4.* (The meaning of "finite-dimensional" will become apparent shortly.) In *Linear Algebra I* we already verified that $\mathbb{K}^n$ is an example of a vector space over $\mathbb{K}$; in this section we go on to prove that that there are essentially no further examples.

**Definition 1.7.** 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 dependent* if there are scalars $c_1, c_2, \ldots, c_n$, not all zero, satisfying

$$c_1 v_1 + c_2 v_2 + \cdots + c_n v_n = \mathbf{0}.$$

The vectors $v_1, v_2, \ldots, v_n$ are *linearly independent* if they are not linearly dependent. Equivalently, they are linearly independent if, whenever we have scalars $c_1, c_2, \ldots, c_n$ satisfying

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

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

(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

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

(c) The list of vectors $v_1, v_2, \ldots, v_n$ is a *basis* for $V$ if it is linearly independent and spanning.

**Remark 1.8.** 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.

**Definition 1.9.** The *span* $\langle v_1, \ldots, v_n \rangle$ of vectors $v_1, \ldots, v_n$ is the set of all vectors that can be written as linear combinations of vectors from $v_1, \ldots, v_n$:

$$\langle v_1, \ldots, v_n \rangle = \left\{ c_1 v_1 + c_2 v_2 + \cdots + c_n v_n : (c_1, \ldots, c_n) \in \mathbb{K}^n \right\}.$$

So vectors $v_1, v_2, \ldots, v_n$ are spanning if $V = \langle v_1, v_2, \ldots, v_n \rangle$. We will see later that the span of vectors is a vector space (or you can verify it now from the definitions).

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

**Definition 1.10.** Let $V$ be a vector space over the field $\mathbb{K}$. We say that $V$ is *finite-dimensional* if we can find vectors $v_1, v_2, \ldots, v_n \in V$ that form a basis for $V$.

**Remark 1.11.** In these notes (apart from in this chapter) we are only concerned with finite-dimensional vector spaces. However, it should be noted that in various contexts, in mathematics and physics, we encounter vector spaces which are not finite dimensional.

A linearly dependent list of vectors has redundancy. It is possible to remove at least one vector from the list while keeping the span of the list the same. Here is a systematic way to do so.

**Lemma 1.12.** *Suppose $v_1, \ldots, v_m$ is a linearly dependent list of vectors in $V$. There exists an index $i \in \{1, \ldots, m\}$ such that*

*(a) $v_i \in \langle v_1, \ldots, v_{i-1} \rangle$, and*

*(b) $\langle v_1, \ldots, v_{i-1}, v_{i+1}, \ldots, v_m \rangle = \langle v_1, \ldots, v_m \rangle$.*

*Proof.* Since $v_1, \ldots, v_m$ are linearly dependent, there exist scalars $c_1, \ldots, c_m$, not all zero, such that $c_1 v_1 + \cdots + c_m v_m = \mathbf{0}$. Choose $i$ to be the largest index such that $c_i \neq 0$. Then

$$v_i = -\Big(\frac{c_1}{c_i}\Big) v_1 - \cdots - \Big(\frac{c_{i-1}}{c_i}\Big) v_{i-1} \tag{1.1}$$

is an explicit expresion for $v_i$ in terms of $v_1, \ldots, v_{i-1}$, demonstrating that $v_i \in \langle v_1, \ldots, v_{i-1} \rangle$. This deals with item (a).

For item (b), suppose $v$ is any vector in $\langle v_1, \ldots, v_m \rangle$; by definition of span, $v = a_1 v_1 + \cdots + a_m v_m$, for some $a_1, \ldots, a_m \in \mathbb{K}$. Now substitute for $v_i$, using (1.1), to obtain an expression for $v$ as a linear combination of vectors in $v_1, \ldots, v_{i-1}, v_{i+1}, \ldots, v_m$. This expression demonstrates that $v \in \langle v_1, \ldots, v_{i-1}, v_{i+1}, \ldots, v_m \rangle$. Since $v$ was arbitrary, item (b) follows. □

**Lemma 1.13.** *The length of any linearly independent list of vectors in $V$ is less than or equal to the length of any spanning list of vectors.*

*Proof.* Suppose $v_1, \ldots, v_n$ are linearly independent and $w_1, \ldots, w_m$ are spanning. Start with the list $w_1, \ldots, w_m$ and repeat the following step, which adds some vector $v_i$ to the list and removes some $w_j$. For the first step, add vector $v_1$ to the front of the list to obtain $v_1, w_1, \ldots, w_m$. Since the original list was spanning, the new one is linearly dependent as well as spanning. By Lemma 1.12, we may remove some $w_j$ so that the remaining list is still spanning. By reindexing some of the $w_j$'s we may write the resulting list as $v_1, w_2, w_3, \ldots, w_m$.

In general, suppose, after some number of steps, the procedure has reached the spanning list $v_1, \ldots, v_{k-1}, w_k, \ldots, w_m$ (where some reindexing of vectors in $w_1, \ldots, w_m$ has taken place). Add the vector $v_k$ between $v_{k-1}$ and $w_k$ in the list. As before, the new list is linearly dependent, and we may apply Lemma 1.12 to remove one of the vectors in the list while retaining the property that the list is spanning. The important observation is the following: because $v_1, \ldots, v_k$ are linearly independent, the removed vector cannot be one of the $v_i$'s and so must be one of the $w_j$'s. (See part (a) of Lemma 1.12.)

At each step we add one vector and remove one vector keeping the length of the list unchanged. We end up with a list of the form $v_1, \ldots, v_n, w_{n+1}, \ldots, w_m$. It follows that $m \geq n$. □

**Remark 1.14.** The proof establishes a little more than we needed. In fact we have essentially proved the *Steinitz Exchange Lemma*. (See, e.g., Wikipedia.)

**Theorem 1.15.** *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 spanning list of vectors can be shortened (by removing some vectors) to a basis;*

*(c) any linearly independent list of vectors can be extended (by adding some vectors) to a basis.*

*Proof.* (a) Suppose $\mathcal{B}_1$ and $\mathcal{B}_2$ are any two bases for $V$, of lengths $n_1$ and $n_2$ respectively. By Lemma 1.13, since $\mathcal{B}_1$ is linearly independent and $\mathcal{B}_2$ is spanning, $n_1 \leq n_2$. Also, since $\mathcal{B}_2$ is linearly independent and $\mathcal{B}_1$ is spanning, $n_2 \leq n_1$.

(b) Suppose $v_1, \ldots, v_m$ is any spanning list for $V$. By Lemma 1.12, if this list is linearly dependent, we can remove some vector $v_i$ from it, leaving a smaller spanning list. By repeating this step we must eventually reach a basis.

(c) Suppose $v_1, \ldots, v_m$ is a linearly independent list of vectors. If this list is not spanning then there must exist a vector $v_{m+1} \in V$ such that $v_{m+1} \notin \langle v_1, \ldots, v_m \rangle$. The extended list $v_1, \ldots, v_m, v_{m+1}$ remains linearly independent. (To see this, assume to the contrary that there exist scalars $a_1, \ldots, a_{m+1}$, not all zero, such that $a_1 v_1 + \cdots + a_{m+1} v_{m+1} = \mathbf{0}$. Since $v_1, \ldots, v_m$ are linearly independent, $a_{m+1}$ cannot be 0. Then $v_{m+1} = -(a_0/a_{m+1})v_1 - \cdots - (a_m/a_{m+1})v_n$, and $v_{m+1} \in \langle v_1, \ldots v_m \rangle$ contrary to assumption.) By repeating this step we must eventually reach a basis. (Note that the process must terminate, since the vector space $V$ is finite dimensional.) □

**Definition 1.16.** The number of elements in a basis of a vector space $V$ is called the *dimension* of $V$. Theorem 1.15 assures us that this parameter is well defined.

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 $\dim(V)$.

**Remark 1.17.** We allow the possibility that a vector space has dimension zero. Such a vector space contains just one vector, the zero vector $\mathbf{0}$; a basis for this vector space consists of the empty set.

Since the notion of basis of a vector space is so fundamental, it is useful in what follows to note some equivalent characterisations. These alternatives are not too difficult to verify, given Theorem 1.15.

**Proposition 1.18.** *The following five conditions are equivalent for a list $\mathcal{B}$ of vectors from vector space $V$ of dimension $n$ over $\mathbb{K}$*

*(a) $\mathcal{B}$ is a basis;*

*(b) $\mathcal{B}$ is a maximal linearly independent list (that is, if we add any vector to the list, then the resulting list is linearly dependent);*

*(c) $\mathcal{B}$ is a minimal spanning list (that is, if we remove any vector from the list, then the result is no longer spanning);*

*(d) $\mathcal{B}$ is linearly independent and has length $n$;*

*(e) $\mathcal{B}$ is spanning and has length $n$.*

## 1.3 Coordinate representations

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 $c_1, c_2, \ldots, c_n \in \mathbb{K}$. These scalars are 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

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

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,

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

In light of this discussion, we can make the following definition.

**Definition 1.19.** Let $V$ be a vector space with a basis $\mathcal{B} = (v_1, v_2, \ldots, v_n)$. If $v = c_1 v_1 + c_2 v_2 + \cdots + c_n v_n$, then the *coordinate representation* of $v$ relative to the basis $\mathcal{B}$ is

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

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

$$[v]_{\mathcal{B}} = \begin{bmatrix} c_1 & c_2 & \ldots & c_n \end{bmatrix}^\top,$$

where the symbol $\top$ is read "transpose".

**Remark 1.20.** In this course, the notation $v_i$, $w_i$, etc., stands for the $i$th vector in a sequence of vectors. It will not be used to denote the $i$th coordinate of the vector $v$ (which would be a scalar). We'll use different letters for the vector and for its coordinates.

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

**Theorem 1.21.** *Any $n$-dimensional vector space over a field $\mathbb{K}$ is isomorphic to the vector space $\mathbb{K}^n$.*

Note that the coordinate representation of a vector is always relative to a basis. The choice of basis is essentially arbitrary, and the development of the theory of vector spaces is cleaner if we avoid introducing a specific basis. (In this module, we only go half-way in this direction.) However, in order to compute with vectors and linear transformations, it is inevitable that we will be working relative to a basis. Different bases give rise to different coordinate representations, and it is important to know how to transform between them.

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 we could decide to represent an $n$-tuple: as a row, or as a column. Thus, the vector with components 1, 2 and $-3$ could be represented as a *row vector*

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

or as a *column vector*

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

Following *Linear Algebra I* and most (all?) textbooks on vector spaces, we stick to column vectors. However, presumably though a historical accident, vectors in coding theory are generally written as row vectors, and Cartesian coordinates for points 2- or 3-dimensional Euclidean space are also written horizontally. The choice of row or column vectors makes some technical differences in the statements of the theorems, so care is needed.

One reason to prefer column vectors in linear algebra is that we can represent a system of linear equations, say

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

in matrix form as

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

This is neater and used less space then the opposite convention. In any case, we will use column vectors in these notes.

Let $\mathcal{B} = (v_1, \ldots, v_n)$ and $\mathcal{B}' = (v'_1, \ldots, v'_n)$ be different bases of an $n$-dimensional vector space $V$ over the field $\mathbb{K}$. Recall from *Linear Algebra I* that there is an $n \times n$ *transitition matrix* $P_{\mathcal{B},\mathcal{B}'}$ that translates coordinate representations relative to $\mathcal{B}'$ to coordinate representations relative to $\mathcal{B}$. Specifically, $[v]_{\mathcal{B}} = P_{\mathcal{B},\mathcal{B}'}\,[v]_{\mathcal{B}'}$ for all vectors $v \in V$. In this course, we will see several ways in which matrices arise in linear algebra. Here is the first occurrence: matrices arise as transition matrices between bases of a vector space.

Let $I$ denote 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 $P$, that is to say, 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.

We recall from *Linear Algebra I* some facts about transition matrices, which come directly from the definition, using uniqueness of the coordinate representation. Let $\mathcal{B}, \mathcal{B}', \mathcal{B}''$ be bases of the vector space $V$. Then

(a) $P_{\mathcal{B},\mathcal{B}} = I$,

(b) $P_{\mathcal{B}',\mathcal{B}} = (P_{\mathcal{B},\mathcal{B}'})^{-1}$; in particular, the transition matrix is invertible, and

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

To see that (b) holds, let's transform the coordinate representation of $u$ relative to basis $\mathcal{B}$ by multiplication by $P_{\mathcal{B}',\mathcal{B}}$:

$$P_{\mathcal{B}',\mathcal{B}}[u]_{\mathcal{B}} = P_{\mathcal{B}',\mathcal{B}}\big(P_{\mathcal{B},\mathcal{B}'}[u]_{\mathcal{B}'}\big) = \big(P^{-1}_{\mathcal{B},\mathcal{B}'}P_{\mathcal{B},\mathcal{B}'}\big)[u]_{\mathcal{B}'} = [u]_{\mathcal{B}'}.$$

We obtain the coordinate representation of $u$ relative to basis $\mathcal{B}'$, as desired.

To see that (c) holds, transform the coordinate representation of $u$ relative to basis $\mathcal{B}''$ by multiplication by $P_{\mathcal{B},\mathcal{B}''}$:

$$P_{\mathcal{B},\mathcal{B}''}[u]_{\mathcal{B}''} = \big(P_{\mathcal{B},\mathcal{B}'}P_{\mathcal{B}',\mathcal{B}''}\big)[u]_{\mathcal{B}''} = P_{\mathcal{B},\mathcal{B}'}\big(P_{\mathcal{B}',\mathcal{B}''}[u]_{\mathcal{B}''}\big) = P_{\mathcal{B},\mathcal{B}'}[u]_{\mathcal{B}'} = [u]_{\mathcal{B}}.$$

We obtain the coordinate representation of $u$ relative to basis $\mathcal{B}$, as desired.

**Example 1.22.** Suppose that $\mathcal{B} = (v_1, v_2)$ and $\mathcal{B}' = (v_1', v_2')$ are different bases of a 2-dimensional vector space $V$ over $\mathbb{R}$. Since $\mathcal{B}$ is a basis of $V$ we can express the basis vectors of $\mathcal{B}'$ in terms of $\mathcal{B}$. Suppose, in fact, that

$$v_1' = v_1 + v_2 \qquad \text{and} \qquad v_2' = 2v_1 + 3v_2.$$

Then the transition matrix from $\mathcal{B}'$ to $\mathcal{B}$ is

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

Note that the first column of $P_{\mathcal{B},\mathcal{B}'}$ is just $[v_1']_{\mathcal{B}}$, i.e., the coordinate representation of the vector $v_1'$ relative to the basis $\mathcal{B}$, and the second column is just $[v_2']_{\mathcal{B}}$. This gives an easy way to write down $P_{\mathcal{B},\mathcal{B}'}$.

Suppose that the coordinate representation of some vector $u$ relative to the basis $\mathcal{B}'$ is $[u]_{\mathcal{B}'} = \begin{bmatrix} a & b \end{bmatrix}^\mathsf{T}$. Then, from the definition of transition matrix, we should have

$$[u]_{\mathcal{B}} = \begin{bmatrix} 1 & 2 \\ 1 & 3 \end{bmatrix}\begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} a+2b \\ a+3b \end{bmatrix}.$$

We can check the result as follows:

$$u = av_1' + bv_2' = a(v_1 + v_2) + b(2v_1 + 3v_2) = (a+2b)v_1 + (a+3b)v_2.$$

So indeed $[u]_{\mathcal{B}} = \begin{bmatrix} a+2b & a+3b \end{bmatrix}^\mathsf{T}$ as expected.

The transition matrix from $\mathcal{B}$ to $\mathcal{B}'$ is the inverse of $P_{\mathcal{B},\mathcal{B}'}$:

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

Finally, suppose $\mathcal{B}'' = (v_1'', v_2'')$ is a third basis of $V$, related to $\mathcal{B}'$ by $v_1'' = 3v_1' - 2v_2'$ and $v_2'' = -2v_1' + v_2'$. Then

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

and

$$P_{\mathcal{B},\mathcal{B}''} = P_{\mathcal{B},\mathcal{B}'}P_{\mathcal{B}',\mathcal{B}''} = \begin{bmatrix} 1 & 2 \\ 1 & 3 \end{bmatrix}\begin{bmatrix} 3 & -2 \\ -2 & 1 \end{bmatrix} = \begin{bmatrix} -1 & 0 \\ -3 & 1 \end{bmatrix}.$$

Note that this example provides additional insight into why matrix multiplication is defined the way it is: in this instance, it provides the correct rule for composing transition matrices.

## 1.4 Subspaces and direct sums

**Definition 1.23.** Suppose $V$ is a vector space over $\mathbb{K}$. We say that $U$ is a *subspace* of $V$ if $U$ is a subset of $V$, and $U$ is itself a vector space (with respect the same operations of vector addition and scalar multiplication).

**Lemma 1.24.** *Suppose $U$ is a non-empty subset of a vector space $V$. The following conditions are equivalent:*

*(a) $U$ is a subspace of $V$;*

*(b) $U$ is closed under vector addition and scalar multiplication. (That is to say, $u+u' \in U$ and $cu \in U$ for any vectors $u, u' \in U$ and scalar $c \in \mathbb{K}$.)*

*Proof.* Since any vector space is closed under vector addition and scalar multiplication, it is clear that (a) implies (b).

Suppose now that (b) holds. For any vector $u \in U$, we know that $-u = (-1)u$ is in $U$ (by closure under scalar multiplication). Also, since $U$ is non-empty, the additive identity $\mathbf{0} = u - u$ is in $U$. So (b) assures us that the operations of vector addition, taking the inverse of a vector, and scalar multiplication all make sense in $U$; moreover, $U$ contains an additive identity. The vector space axioms (VA) and (VM) for $U$ are inherited from $V$: since they hold in the larger set, they certainly hold in the smaller. (Go through all five axioms and convince yourself of this fact.). Assertion (a) follows. □

Subspaces can be constructed in various ways:

(a) Recall that the span of vectors $v_1, \ldots, v_k \in V$ is the set

$$\langle v_1, v_2, \ldots, v_k \rangle = \{c_1 v_1 + c_2 v_2 + \cdots + c_k v_k : c_1, \ldots, c_k \in \mathbb{K}\}.$$

This is a subspace of $V$. Moreover, vectors $v_1, \ldots, v_k$ are spanning in this subspace.

(b) Let $U$ and $W$ be subspaces of $V$. Then

- the *intersection* $U \cap W$ is the set of all vectors belonging to both $U$ and $W$;
- the *sum* $U + W$ is the set $\{u + w : u \in U, w \in W\}$ of all sums of vectors from the two subspaces.

Both $U \cap W$ and $U + W$ are subspaces of $V$.

We will just check (a) here, leaving (b) as an exercise. By Lemma 1.24, we just need to check closure under vector addition and scalar multiplication. So suppose $v = c_1 v_1 + \cdots + c_k v_k$ and $v' = c'_1 v_1 + \cdots + c'_k v_k$ are vectors in the span $\langle v_1, \ldots, v_k \rangle$ of $v_1, \ldots, v_k \in V$. Then $v + v' = (c_1 v_1 + \cdots + c_k v_k) + (c'_1 v_1 + \cdots + c'_k v_k) = (c_1 + c'_1)v_1 + \cdots + (c_k + c'_k)v_k$, which is clearly also in the span $\langle v_1, \ldots, v_k \rangle$. Also for any $a \in \mathbb{K}$, we have $av = a(c_1 v_1) + \cdots + a(c_k v_k) = (ac_1)v_1 + \cdots + (ac_k)v_k$, which is again clearly in $\langle v_1, \ldots, v_k \rangle$.

**Theorem 1.25.** *Let $V$ be a vector space over $\mathbb{K}$. For any two subspaces $U$ and $W$ of $V$, we have*

$$\dim(U \cap W) + \dim(U + W) = \dim(U) + \dim(W).$$

*Proof.* Let $v_1, \ldots, v_i$ be a basis for $U \cap W$. By Theorem 1.15(c) we can extend this basis to a basis $v_1, \ldots, v_i, u_1, \ldots, u_j$ of $U$ and a basis $v_1, \ldots, v_i, w_1, \ldots, w_k$ of $W$. If we can show that $v_1, \ldots, v_i, u_1, \ldots, u_j, w_1, \ldots, w_k$ is a basis of $U + W$ then we are done, since then

$$\dim(U \cap W) = i, \quad \dim(U) = i + j, \quad \dim(W) = i + k, \quad \text{and} \quad \dim(U + W) = i + j + k,$$

and both sides of the identity we are aiming to prove are equal to $2i + j + k$.

Since every vector in $U$ (respectively $W$) can be expressed as a linear combination of $v_1, \ldots, v_i, u_1, \ldots, u_j$ (respectively $v_1, \ldots, v_1, w_1, \ldots, w_k$), it is clear that the list of vectors $v_1, \ldots, v_1, u_1, \ldots, u_j, w_1, \ldots, w_k$ spans $U + W$. So we just need to show that the list $v_1, \ldots, v_1, u_1, \ldots, u_j, w_1, \ldots, w_k$ is linearly independent.

Consider any linear relationship

$$a_1 v_1 + \cdots + a_i v_i + b_1 u_1 + \cdots + b_j u_j + c_1 w_1 + \cdots + c_k w_k = \mathbf{0};$$

we need to show that $a_1, \ldots, a_i, b_1, \ldots, b_j, c_1, \ldots, c_k$ are all zero. Writing

$$c_1 w_1 + \cdots + c_k w_k = -a_1 v_1 - \cdots - a_i v_i - b_1 u_1 - \cdots - b_j u_j,$$

we see that $c_1 w_1 + \cdots + c_k w_k \in U$. But, by construction, $c_1 w_1 + \cdots + c_k w_k \in W$, so in fact $c_1 w_1 + \cdots + c_k w_k \in U \cap W$. Since $v_1, \ldots, v_i$ is a basis for $U \cap W$ we have

$$c_1 w_1 + \cdots + c_k w_k = d_1 v_1 + \cdots d_i v_i,$$

for some scalars $d_1, \ldots d_i$. But this implies that $c_1 = \cdots = c_k = 0$ (and, incidentally, $d_1 = \cdots = d_i = 0$), since $v_1, \ldots, v_i, w_1, \ldots, w_k$ is a basis for $W$ and hence linearly independent. A similar argument establishes $b_1 = \cdots = b_k = 0$. But now $a_1 = \cdots = a_i = 0$, since the list $v_1, \ldots, v_i$ is linearly independent. □

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

$$v = u + w = u' + w', \qquad \text{for some } u, u' \in U \text{ and } w, w' \in W.$$

Then

$$u - u' = w' - w.$$

But $u - u' \in U$, and $w' - w \in W$, and hence

$$u - u' = w' - w \in U \cap W = \{\mathbf{0}\}.$$

It follows that $u = u'$ and $w = w'$; that is, the two expressions for $v$ are not different after all! In this case we say that $U + W$ is the *direct sum* of the subspaces $U$ and $W$, and write it as $U \oplus W$. Note that

$$\dim(U \oplus W) = \dim(U) + \dim(W).$$

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 1.26.** 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

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

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 $i = 1, \ldots, r$.

There is an equivalent characterisation of direct sum that will be useful later.

**Lemma 1.27.** *Suppose $U_1, \ldots, U_r$ are subspaces of $V$, and $V = U_1 + \cdots + U_r$. Then the following are equivalent:*

*(a) $V$ is the* direct sum *of $U_1, \ldots, U_r$.*

*(b) For all vectors $u_1 \in U_1, \ldots, u_r \in U_r$, it is the case that $u_1 + \cdots + u_r = \mathbf{0}$ implies $u_1 = \cdots = u_r = \mathbf{0}$.*

*Proof.* (a) $\implies$ (b). Suppose $u_1 + \cdots + u_r = \mathbf{0}$, where $u_1 \in U_1, \ldots, u_r \in U_r$. Certainly $u_1 = \cdots = u_r = \mathbf{0}$ is one way this situation may occur. But the definition of direct sum tells us that such an expression is unique. So, indeed, $u_1 = \cdots = u_r = \mathbf{0}$ as required.

(b) $\implies$ (a). Suppose $v \in V$ and that $v = u_1 + \cdots + u_r$ and $v = u_1' + \cdots + u_r'$ are two ways of expressing $v$, with $u_1, u_1' \in U_1, \ldots, u_r, u_r' \in U_r$. Then

$$(u_1 - u_1') + \cdots + (u_r - u_r') = (u_1 + \cdots + u_r) - (u_1' + \cdots + u_r') = v - v = \mathbf{0}.$$

From condition (b), we deduce that $u_1 - u_1' = \cdots = u_r - u_r' = \mathbf{0}$. Thus, $u_1 = u_1', \ldots, u_r = u_r'$ as required. □

Note the similarity between the condition described in Lemma 1.27(b) and the definition of linear independence. In fact, $v_1, \ldots, v_n$ is a basis for a vector space $V$ if and only if $V = \langle v_1 \rangle \oplus \cdots \oplus \langle v_n \rangle$. In a sense, a direct sum generalises the concept of basis.

**Lemma 1.28.** *If $V = U_1 \oplus \cdots \oplus U_r$, then*

*(a) if $\mathcal{B}_i$ is a basis for $U_i$ for $i = 1, \ldots, r$, then $\mathcal{B} = (\mathcal{B}_1, \cdots, \mathcal{B}_r)$, i.e., the concatenation of the lists $\mathcal{B}_1, \ldots, \mathcal{B}_r$, is a basis for $V$;*

*(b) $\dim(V) = \dim(U_1) + \cdots + \dim(U_r)$.*

*Proof.* Since every vector $v \in V$ may be expressed as $v = u_1 + \cdots + u_r$ with $u_i \in U_i$, and every $u_i \in U_i$ may be expressed as a linear combination of basis vectors in $\mathcal{B}_i$, we see that $V$ is contained in the span of $\mathcal{B}$. So we just need to verify that the list $\mathcal{B}$ is linearly independent.

Let $d_i = \dim(U_i)$ and $\mathcal{B}_i = (u_{i,1}, \ldots, u_{i,d_i})$, for $1 \leq i \leq r$, be an explicit enumeration of the basis vectors $\mathcal{B}_i$. Suppose that some linear combination of the basis vectors $\mathcal{B}$ sums to $\mathbf{0}$. We can express this linear combination as $u_1 + \cdots + u_r = \mathbf{0}$, where $u_i = a_{i,1}u_{i,1} + \cdots + a_{i,d_i}u_{i,d_i}$ for some scalars $a_{i,1}, \ldots, a_{i,d_i} \in \mathbb{K}$.

By Lemma 1.27, $u_i = \mathbf{0}$ for all $1 \leq i \leq r$. Then, since $\mathcal{B}_i$ is a basis and hence linearly independent, $a_{i,1} = \cdots = a_{i,d_i} = 0$. Since the linear combination of basis vectors $\mathcal{B}$ was arbitrary, we deduce that $\mathcal{B}$ is linearly independent.

This deals with part (a). Part (b) follows immediately, since

$$\dim(V) = |\mathcal{B}| = |\mathcal{B}_1| + \cdots + |\mathcal{B}_r| = \dim(U_1) + \cdots + \dim(U_r).$$

□

**Remark 1.29.** The results in this chapter apply to all finite dimensional vector spaces over $\mathbb{K}$, regardless of the field $\mathbb{K}$. In our proofs, we used nothing beyond the general axioms of a field. In some later chapters we need restrict our attention to particular fields, typically $\mathbb{R}$ or $\mathbb{C}$.

## Summary

- A vector space is a mathematical structure that can be defined axiomatically.
- A basis is a list of vectors that is linearly independent (not redundant) and spanning (sufficient).
- A vector space is finite dimensional if it has a basis of finite cardinality. Every basis of a finite-dimensional vector space $V$ has the same cardinality (Exchange Lemma). The cardinality of any basis of $V$ is the dimension of $V$.
- There are many possible characterisations of a basis; it is good to know several (Proposition 1.18).
- Having chosen a basis for a vector space, every vector has a unique coordinate representation relative to that basis.
- Different bases lead to different coordinate representations, and we can translate between them using transition matrices.
- A subspace of a vector space $V$ is a subset $U$ of $V$ that is itself a vector space. To verify that a subset $U$ is a subspace, we just need to check non-emptyness together with closure under vector addition and scalar multiplication.
- We can combine subspaces through the operations of sum and intersection. The dimension of a subspace is a measure of the "size" of the subspace, which behaves in some ways like the cardinality of a finite set (even though the subspaces themselves are generally infinite). Refer to Theorem 1.25.
- Direct sums are particularly nice sums of vector spaces. A direct sum corresponds to a decomposition of a vector space with no redundancy.

# Chapter 2

# Matrices

In this chapter, we review matrix algebra from *Linear Algebra I*, consider row and column operations on matrices, and define the rank of a matrix. Along the way prove that the "row rank" and "column rank" defined in *Linear Algebra I* are in fact equal.

## 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 $\mathbb{K}$. The matrix will typically be denoted by an upper case letter, and its entries by the corresponding lower case letter. Thus, for $1 \leq i \leq m$ and $1 \leq j \leq n$, the entry in row $i$ and column $j$ of matrix $A$ is denoted by $a_{ij}$, and referred to as the $(i, j)$ entry of $A$.

**Example 2.2.** 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 2.3.** We define addition and multiplication of matrices as follows.

(a) Let $A = (a_{ij})$ and $B = (b_{ij})$ be matrices of the same size $m \times n$ over $\mathbb{K}$. Then the sum $C = A + B$ is defined by adding corresponding entries of $A$ and $B$; thus $C = (c_{ij})$ is given by

$$c_{ij} = a_{ij} + b_{ij}.$$

(b) Let $A$ be an $m \times n$ matrix and $B$ an $n \times p$ matrix over $\mathbb{K}$. Then the product $C = AB$ is the $m \times p$ matrix whose $(i, j)$ entry is obtained by multiplying each element in the $i$th row of $A$ by the corresponding element in the $j$th column of $B$ and summing:

$$c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}.$$

**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. Technically, the set $M_n(\mathbb{K})$ of $n \times n$ matrices over $\mathbb{K}$ together with matrix addition and multiplication is a ring (with identity). 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 2.4.** *Elementary row operations.* There are three types:

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

Type 2. Multiply the $i$th row by a non-zero scalar.

Type 3. Interchange the $i$th and $j$th rows, where $j \neq i$.

*Elementary column operations.* There are three types:

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

Type 2. Multiply the $i$th column by a non-zero scalar.

Type 3. Interchange the $i$th and $j$th columns, where $j \neq i$.

We can describe the elementary row and column operations in a different way. For each elementary row operation on an $m \times n$ matrix $A$, we define a corresponding *elementary matrix* by applying the same operation to the $m \times m$ identity matrix $I$. Similarly for each elementary column operation we define a corresponding elementary matrix by applying the same operation to the $n \times n$ identity matrix.

We don't have to distinguish between rows and columns for our elementary matrices: each matrix can be considered either as a row or a column operation. This observation will be important later. 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 2.5.** *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 2.6.** Let $A$ be a $2 \times 3$ real matrix. The matrices corresponding to the elementary row operation of subtracting 4 times row 1 from row 2, and the elementary column operation of subtracting twice column 1 from column 2 are

$$\begin{bmatrix} 1 & 0 \\ -4 & 1 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 1 & -2 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

respectively. If $A$ is the matrix

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

then the matrix that results from applying the above two elementary operations ought to be

$$\begin{bmatrix} 1 & 0 \\ -4 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \begin{bmatrix} 1 & -2 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 3 \\ 0 & -3 & -6 \end{bmatrix}.$$

You should check that this is indeed the case.

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 column to the second is undone by adding $-2$ times the first column to the second, so that

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

## 2.3 Rank

Recall from *Linear Algebra I* the definitions of row space, column space, row rank and column rank of a matrix.

**Definition 2.7.** Let $A$ be an $m \times n$ matrix over a field $\mathbb{K}$. The *row space* of $A$ is the vector space spanned by rows of $A$ and the *column space* the vector space spanned by columns. The *row rank* of $A$ is the dimension of the row space, and the *column rank* of $A$ the dimension of the column space of $A$. (We regard columns or rows as vectors in $\mathbb{K}^m$ and $\mathbb{K}^n$ respectively.)

**Remark 2.8.** Since a maximal linearly independent set of vectors is a basis, we could alternatively define row rank as the maximum number of linearly independent rows, and the column rank analogously.

Recall also that elementary row operations preserve row rank, and elementary column operations preserve column rank. In *Linear Algebra I*, the rank of a matrix was defined as its row rank. Why? The definition privileges rows over columns, and hence seems somewhat arbitrary. In any case, why should the dimension of the row space be a significant parameter?

The next lemma goes beyond *Linear Algebra I* by showing that elementary row operations preserve column rank, not just row rank.

**Lemma 2.9.** *(a) Elementary column operations preserve the column space of a matrix (and hence don't change the column rank).*

*(b) Elementary row operations preserve the row space of a matrix (and hence don't change the row rank).*

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

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

*Proof.* First note that if $A$ is an $m \times n$ matrix and $u = [c_1, c_2, \ldots, c_n]^\top$ an $n$-vector, then $Au$ can be thought of as expressing a linear combination of the columns of $A$. Specifically, $Au = c_1v_1 + c_2v_2 + \cdots + c_nv_n$, where $m$-vectors $v_1, v_2, \ldots, v_n$ are the $n$ columns of $A$, taken in order.

(a) Suppose $A'$ is obtained from matrix $A$ by some elementary column operation. Equivalently, $A' = AC$ for some elementary matrix $C$. Consider any vector $v$ in the column space of $A'$; as we observed, the condition for $v$ to be in the column space of $A'$ is that there exists a vector $u$ such that $v = A'u$. Then $v = A'u = (AC)u = A(Cu) = Au'$, where $u' = Cu$. Thus $v$ also is in the column space of $A$. It follows that the column space of $A'$ is contained in the column space of $A$. Finally, elementary column operations are invertible, so the inclusion holds also in the other direction. We deduce that the column spaces of $A$ and $A'$ are equal.

(b) Follows by symmetry from (a).

(c) Suppose $A'$ is obtained from matrix $A$ by some elementary row operation. Equivalently, $A' = RA$ for some elementary matrix $R$. Consider any linear dependancy among the columns of $A$, say $Au = \mathbf{0}$. Then $A'u = (RA)u = R(Au) = R\,\mathbf{0} = \mathbf{0}$ and so the same linear dependency exists among the columns of $A'$. Let $S \subseteq \{1, \ldots, n\}$ be a subset of columns. If the columns of $A$ indexed by $S$ are linearly dependent then the same columns in $A'$ are linearly dependent. Equivalently, if the $S$-indexed columns of $A'$ are linearly independent, then so are the $S$-indexed columns of $A$. Let $k$ be the column rank of $A'$. Choose a basis for the column space among the columns of $A'$. These columns are linearly independent in $A'$ and hence also in $A$. It follows that the rank of the column space of $A$ is at least $k$. Thus the column rank of $A$ is at least as large as that of $A'$. As before, column operations are invertible, so the inequality holds also in the other direction.

(d) Follows from (c) by symmetry.

□

It is important to note that elementary row operations do *not* in general preserve the column space of a matrix, only the column rank. Provide a counterexample to illustrate this fact. (An elementary row operation on a $2 \times 2$ matrix is enough for this purpose.)

By applying elementary row and column operations, we can reduce any matrix to a particularly simple form:

**Theorem 2.10.** *Let $A$ be an $m \times n$ matrix over the field $\mathbb{K}$. Then it is possible to transform $A$ by elementary row and column operations into a matrix $D = (d_{ij})$ of the same size as $A$, with the following special form: there is an $r \leq \min\{m, n\}$, such that $d_{ii} = 1$ for $1 \leq i \leq r$, and $d_{ij} = 0$ otherwise.*

*The matrix $D$ (and hence the number $r$), is uniquely defined: if $A$ can be reduced to two matrices $D$ and $D'$, both of the above form, by different sequences of elementary operations then $D = D'$.*

**Definition 2.11.** The number $r$ in the above theorem is called the *rank* of $A$; while a matrix of the form described for $D$ is said to be in the *canonical form for equivalence*. We can write the canonical form matrix in "block form" as

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

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 $O$s). Note that some or all of these $O$s may be missing: for example, if $r = m$, we just have $\begin{bmatrix} I_m & O \end{bmatrix}$.

*Proof of Theorem 2.10.* We first outline the proof that the reduction is possible. The proof is by induction on the size of the matrix $A = (a_{ij})$. Specifically, 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 $i$th rows, and the first and $j$th 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 $j$th gives a matrix with $a_{1j} = 0$. Repeat this until all non-zero elements have been removed.
- Now let $A'$ be the matrix obtained by deleting the first row and column of $A$. Then $A'$ is smaller than $A$ and so, by the inductive hypothesis, we can reduce $A'$ to canonical form by elementary row and column operations. The same sequence of operations applied to $A$ now finishes the job.

Suppose that we reduce $A$ to canonical form $D$ by elementary operations, where $D$ has $r$ 1s on the diagonal. These elementary operations don't change the row or column rank, by Lemma 2.9. Therefore, the row ranks of $A$ and $D$ are equal, and their column ranks are equal. But it is not difficult to see that, if

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

then the row and column ranks of $D$ are both equal to $r$. It doesn't matter which elementary operations we use to reduce to canonical form, we will always obtain the same matrix $D$. So the theorem is proved. □

**Corollary 2.12.** *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.*

**Example 2.13.** 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. So first we subtract 4 times the first row from the second, then subtract twice the first column from the second, and

then 3 times the first column from the third. These steps yield the following sequence of matrices

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \xrightarrow{\mathrm{R}_2 - 4\mathrm{R}_1} \begin{bmatrix} 1 & 2 & 3 \\ 0 & -3 & -6 \end{bmatrix} \xrightarrow{\mathrm{C}_2 - 2\mathrm{C}_1} \begin{bmatrix} 1 & 0 & 3 \\ 0 & -3 & -6 \end{bmatrix} \xrightarrow{\mathrm{C}_3 - 3\mathrm{C}_1} \begin{bmatrix} 1 & 0 & 0 \\ 0 & -3 & -6 \end{bmatrix}.$$

At this point we have successfully set to zero the first row and column of the matrix, except for the top left entry. 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 of the matrix by $-\frac{1}{3}$ and finally subtract twice the second column from the thord. Picking up from where we left off, this yields the sequence

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & -3 & -6 \end{bmatrix} \xrightarrow{-\frac{1}{3}\mathrm{R}_2} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 2 \end{bmatrix} \xrightarrow{\mathrm{C}_3 - 2\mathrm{C}_2} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} = D.$$

For compactness, se are using (as in *Linear Algebra I*) shorthand such as $\mathrm{R}_2 - 4\mathrm{R}_1$ for $\mathrm{R}_2 := \mathrm{R}_2 - 4\mathrm{R}_1$ and $-\frac{1}{3}\mathrm{R}_2$ for $\mathrm{R}_2 := -\frac{1}{3}\mathrm{R}_2$. We have finished the reduction to canonical form, and we conclude that the rank of the original matrix $A$ is equal to 2.

**Theorem 2.14.** *For any $m \times n$ matrix $A$ there are invertible matrices $P$ and $Q$ of sizes $m \times m$ and $n \times n$ respectively, such that $D = PAQ$ is in the canonical form for equivalence. The rank of $A$ is equal to the rank of $D$. Moreover, $P$ and $Q$ are products of elementary matrices.*

*Proof.* We know from Theorem 2.10 that there is a sequence of elementary row and column operations that reduces $A$ to $D$. These operations correspond to certain elementary matrices. Take the matrices $R_1, R_2, \ldots, R_s$ corresponding to the row operations and multiply them together (right to left). This is the matrix $P = R_s R_{s-1} \cdots R_1$. Take the matrices $C_1, C_2, \ldots, C_t$ corresponding to the column operations and multiply them together (left to right). This is the matrix $Q = C_1 C_2 \ldots C_t$. □

**Example 2.15.** We illustrate the construction of $P$ and $Q$ in the above proof, in a continuation of our previous example. In order, here is the list of elementary matrices corresponding to the operations we applied to $A$. (Here, $2 \times 2$ matrices are row operations while $3 \times 3$ matrices are column operations).

$$R_1 = \begin{bmatrix} 1 & 0 \\ -4 & 1 \end{bmatrix}, \quad C_1 = \begin{bmatrix} 1 & -2 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad C_2 = \begin{bmatrix} 1 & 0 & -3 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

$$R_2 = \begin{bmatrix} 1 & 0 \\ 0 & -1/3 \end{bmatrix}, \quad C_3 \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -2 \\ 0 & 0 & 1 \end{bmatrix}.$$

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

$$D = R_2 R_1 A C_1 C_2 C_3 = \begin{bmatrix} 1 & 0 \\ 0 & -1/3 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ -4 & 1 \end{bmatrix} A \begin{bmatrix} 1 & -2 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & -3 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -2 \\ 0 & 0 & 1 \end{bmatrix}$$

or more simply

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

where, as before,

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

There is a slightly easier (for humans) method for constructing the matrices $P$ and $Q$, which we examined in the lectures. Let's recall how it works in the context of computing the matrix $Q$. The idea is to use the same column operations we applied to $A$, but *starting instead with the* $3 \times 3$ *identity matrix* $I_3$ :

$$I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \xrightarrow{C_2 - 2C_1} \begin{bmatrix} 1 & -2 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \xrightarrow{C_3 - 3C_1} \begin{bmatrix} 1 & -2 & -3 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \xrightarrow{C_3 - 2C_2} \begin{bmatrix} 1 & -2 & 1 \\ 0 & 1 & -2 \\ 0 & 0 & 1 \end{bmatrix} = Q.$$

Think about why this method works. It is doing essentially the same calculation, but arranging it in a more human-friendly way.

**Definition 2.16.** 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.

**Remark 2.17.** 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.

**Corollary 2.18.** *An $n \times n$ matrix is invertible if and only if it has rank $n$.*

*Proof.* Suppose that $n \times n$ matrices $A$ and $B$ are equivalent. Then $A$ is invertible if and only if $B$ is invertible. (If $A$ is invertible and $B = PAQ$, then $Q^{-1}A^{-1}P^{-1}$ is the inverse of $B$, and similarly in the other direction.) We know from Theorem 2.14 that every matrix $A$ is equivalent to some matrix $D$ in the canonical form for equivalence. Moreover the rank of $A$ is equal to the rank of $D$. Thus, we have the the following chain of implications:

$$A \text{ is invertible} \iff D \text{ is invertible} \iff D = I_n \iff A \text{ has rank } n.$$

□

**Corollary 2.19.** *Every invertible square matrix is a product of elementary matrices.*

*Proof.* If $A$ is an invertible $n \times n$ matrix, then it has rank $n$ and its canonical form is the identity matrix $I_n$. Thus there are invertible matrices $P$ and $Q$, each a product of elementary matrices, such that

$$PAQ = I_n.$$

From this we deduce that

$$A = P^{-1}I_nQ^{-1} = P^{-1}Q^{-1}.$$

Since the elementary matrices are closed under taking inverses, the above is an expression for $A$ as a product of elementary matrices. □

**Corollary 2.20.** *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).*

*Proof.* We observed, when we defined elementary matrices, that they can represent either elementary column operations or elementary row operations. In the previous corollary, we saw that $A$ can be written as a product of elementary matrices, say $A = C_1C_2 \dots C_t$. We can transform $A$ to the identity by multiplying on the right by $C_t^{-1}, \dots, C_2^{-1}, C_1^{-1}$ in turn. This is equivalent to applying a sequence of column operations. Equally, we can transform $A$ to the identity by multiplying on the left by $C_1^{-1}, C_2^{-1}, \dots, C_t^{-1}$ in turn. This is equivalent to applying a sequence of row operations. □

**Theorem 2.21.** *Two $m \times n$ matrices are equivalent if and only if they have the same rank.*

*Proof.* Suppose $A$ and $B$ are (not necessarily square) equivalent matrices, i.e., $B = PAQ$ for some invertible matrices $P$ and $Q$. By Corollary 2.19 we can write $P$ and $Q$ as the product of elementary matrices. It follows that we can transform $A$ to $B$ by elementary row and column operations, and hence the ranks of $A$ and $B$ are the same. (Elementary operations preserve the rank.)

Conversely, if the ranks of $A$ and $B$ are the same then we can transform one to the other (e.g., via the common canonical form $D$) by elementary row and column operations, and hence $A$ and $B$ are equivalent. □

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. Theorem 2.21 says that in this case there are $\min\{m, n\} + 1$ equivalence classes, and the canonical form for equivalence is a canonical form in this sense.

**Remark 2.22.** As with Chapter 1, the results in this chapter apply to all fields $\mathbb{K}$.

## Summary

- Matrices (not necessarily square) can be acted upon by elementary row and column operations. The elementary row operations are of three types, as are the elementary column operations.

- To each elementary row operation on an $m \times n$ matrix $A$ there corresponds an elementary $m \times m$ matrix $R$. Multiplying $A$ on the left by $R$ is equivalent to applying the row operation. Similarly, to each elementary column operation there is an $n \times n$ matrix $C$ such that multiplication on the right by $C$ is equivalent to applying the column operation.

- The row (column) space of a matrix is the vector space spanned by the rows (columns) of the matrix. The row (column) rank is the dimension of the row (column) space.

- Row operations preserve the row space and leave the column rank unchanged. Ditto with rows and columns interchanged.

- Any matrix can be reduced to the *canonical form for equivalence* by elementary row and column operations. As a corollary, row rank and column rank are equal, so we can just talk about "rank".

- An $n \times n$ matrix is invertible iff it has rank $n$. We also say in this case that the matrix is non-singular. A non-singular matrix can be reduced to canonical form [for equivalence] using just elementary row operations (or just column operations).
- A non-singular matrix can be written as the product of elementary matrices.

# Chapter 3

# Determinants

We recall the Leibniz formula for the determinant of a square matrix, and show that the function it defines is the unique function on square matrices satisfying certain nice properties. This provides an axiomatic definition of the determinant, and demystifies, to a certain extent, why the determinant is defined the way it is. We examine methods of calculating the determinant and some of its properties. We study two polynomials associated with a matrix, the minimal and characteristic polynomials. Finally we prove the Cayley-Hamilton Theorem, that states that every matrix satisfies its own characteristic equation.

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:

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

## 3.1 Definitions: explicit and axiomatic

The formula for the determinant involves some background notation.

**Definition 3.1.** 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 $\operatorname{sign}(\pi) = \pm 1$, computed as follows: write $\pi$ as a product of disjoint cycles; if there are $k$ cycles (including cycles of length 1), then $\operatorname{sign}(\pi) = (-1)^{n-k}$. A *transposition* is a permutation which interchanges two symbols and leaves all the others fixed. Thus, if $\tau$ is a transposition, then $\operatorname{sign}(\tau) = -1$.

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 $\operatorname{sign}(\tau) = (-1)^{n-(n-1)} = -1$. We need one more fact about signs: if $\pi$ is any permutation and $\tau$ is a transposition, then $\operatorname{sign}(\pi\tau) = \operatorname{sign}(\tau\pi) = -\operatorname{sign}(\pi)$, where $\pi\tau$ denotes the composition of $\pi$ and $\tau$ (apply first $\tau$, then $\pi$).

**Definition 3.2.** Let $A = (a_{ij})$ be an $n \times n$ matrix over $\mathbb{K}$. The *determinant* of $A$ is

defined by the Leibniz formula

$$\det(A) = \sum_{\pi \in S_n} \operatorname{sign}(\pi) a_{1,\pi(1)} a_{2,\pi(2)} \cdots a_{n,\pi(n)}.$$

This gives us a nice mathematical formula for the determinant of a matrix. Unfortunately, it is a terrible formula for practical computation, 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, $10! = 3628800$.)

Let's come at this from another direction. Consider the following three properties of a function $D$ defined on $n \times n$ matrices.

(D1) For every $1 \leq i \leq n$, $D(A)$ is linear in the $i$th row of the matrix $A$. (We'll spell out below what this means.)

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

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

Some clarification of property (D1). Suppose we have any matrices $A$ and $B$ such that $B$ agrees with $A$, except that row $i$ is multiplied by some scalar $c$. Thus,

$$A = \begin{bmatrix} v_1 \\ \vdots \\ v_{i-1} \\ v_i \\ v_{i+1} \\ \vdots \\ v_n \end{bmatrix} \qquad B = \begin{bmatrix} v_1 \\ \vdots \\ v_{i-1} \\ cv_i \\ v_{i+1} \\ \vdots \\ v_n \end{bmatrix}, \tag{3.1}$$

where $v_1, \ldots, v_n$ are *row* vectors. Then (D1) legislates that $D(B) = cD(A)$. Furthermore, suppose we have three matrices $A$, $A'$ and $B$, such that $A$ and $A'$ agree except at the $i$th row, and such that the $i$th row of $B$ is the sum of the corresponding rows of $A$ and $A'$:

$$A = \begin{bmatrix} v_1 \\ \vdots \\ v_{i-1} \\ v_i \\ v_{i+1} \\ \vdots \\ v_n \end{bmatrix} \qquad A' = \begin{bmatrix} v_1 \\ \vdots \\ v_{i-1} \\ v'_i \\ v_{i+1} \\ \vdots \\ v_n \end{bmatrix} \qquad B = \begin{bmatrix} v_1 \\ \vdots \\ v_{i-1} \\ v_i + v'_i \\ v_{i+1} \\ \vdots \\ v_n \end{bmatrix}. \tag{3.2}$$

Then (D1) legislates that $D(B) = D(A) + D(A')$.

Why are these natural? Well, condition (D1) says that if we fix all the entries of $A$ apart from those in the $i$th row, then $D$ is some linear function of the remaining entries $a_{1i}, \ldots, a_{ni}$. This is a *linear* algebra course, so this property seems reasonable enough. A matrix $A$ with two equal rows has rank less than $n$. Property (D2) says that the function $D(A)$ is zero on at least some (in fact all) matrices of rank less than $n$. If we are looking for a function that is non-zero exactly for invertible matrices, this is a reasonable condition to impose. The conditions (D1) and (D2) cannot define a unique

function, since if $D$ satisfies (D1) and (D2) then so does any multiple of $D$. So if we want to pin down the function $D$ precisely, we need some condition like (D3) to fix the function at a particular point.

If we believe (D1)–(D3) are nice conditions, then the determinant is a nice function.

**Lemma 3.3.** *The function* det() *satisfies (D1)–(D3).*

*Proof.* (D1) Suppose $A = (a_{k\ell})$ is an $n \times n$ matrix, and $A'$ and $B$ are matrices agreeing with $A$ apart from in the $i$th row. Furthermore, suppose matrices $A$, $A'$ and $B$ are related as in (3.2): thus the $i$th row of $B$ is the sum of the $i$th rows of $A$ and $A'$. Then, denoting the $i$th row of $A'$ by $a'_{i1}, a'_{i2}, \ldots, a'_{in}$, we obtain, by the Leibniz formula,

$$\begin{aligned}
\det(B) &= \sum_{\pi\in S_n} \operatorname{sign}(\pi)\, a_{1,\pi(1)} \cdots a_{i-1,\pi(i-1)} \underbrace{(a_{i,\pi(i)} + a'_{i,\pi(i)})}_{b_{i,\pi(i)}} a_{i+1,\pi(i+1)} \cdots a_{n,\pi(n)} \\
&= \sum_{\pi\in S_n} \operatorname{sign}(\pi)\, a_{1,\pi(1)} \cdots a_{i-1,\pi(i-1)} a_{i,\pi(i)} a_{i+1,\pi(i+1)} \cdots a_{n,\pi(n)} \\
&\qquad + \sum_{\pi\in S_n} \operatorname{sign}(\pi)\, a_{1,\pi(1)} \cdots a_{i-1,\pi(i-1)} a'_{i,\pi(i)} a_{i+1,\pi(i+1)} \cdots a_{n,\pi(n)} \\
&= \det(A) + \det(A').
\end{aligned}$$

The case (3.1), where $B$ is obtained from $A$ by multiplying the $i$th row of $A$ by $c$, is similar, but easier, and is left as an exercise. Thus (D1) holds for the determinant.

(D2) Suppose that the $i$th and $j$th rows of $A$ are equal. Let $\tau$ be the transposition that interchanges $i$ and $j$ and leaves the other numbers fixed. Then,

$$a_{i,\pi\tau(i)} = a_{i,\pi(j)} = a_{j,\pi(j)} \quad \text{and} \quad a_{j,\pi\tau(j)} = a_{j,\pi(i)} = a_{i,\pi(i)},$$

where the second equality in each case uses the fact that the $i$th and $j$th rows of $A$ are identical. For any $k \notin \{i, j\}$ we naturally have $a_{k,\pi\tau(k)} = a_{k,\pi(k)}$. Thus, we see that the products

$$a_{1,\pi(1)} a_{2,\pi(2)} \cdots a_{n,\pi(n)} \quad \text{and} \quad a_{1,\pi\tau(1)} a_{2,\pi\tau(2)} \cdots a_{n,\pi\tau(n)}$$

are equal. But $\operatorname{sign}(\pi\tau) = -\operatorname{sign}(\pi)$. 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 $\pi, \pi\tau$ (the natural way to do this is to let $\pi$ range over the subgroup of sign $+1$). 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 $\pi$ which contributes to the sum is the identity permutation $\iota$; any other permutation $\pi$ satisfies $\pi(i) \neq i$ for some $i$, so that $a_{i\pi(i)} = 0$. The sign of $\iota$ is $+1$, and all the factors $a_{ii}$ are equal to 1, so $\det(A) = 1$, as required.

□

So there exists at least one function that satisfies (D1)–(D3). We now show, perhaps surprisingly, that there is only one.

**Theorem 3.4.** *There is a unique function $D$ on $n \times n$ matrices satisfying (D1)–(D3). That function is* det().

*Proof.* Suppose that $D$ is any function on square matrices satisfying (D1)–(D3). First, we show that applying elementary row operations to matrix $A$ has a well-defined effect on $D(A)$.

(a) If $B$ is obtained from $A$ by adding $c$ times the $j$th row to the $i$th, then $D(B) = D(A)$.

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

(c) If $B$ is obtained from $A$ by interchanging two rows $i$ and $j$, then $D(B) = -D(A)$.

For (a), let $A'$ be the matrix which agrees with $A$ in all rows except the $i$th, which is equal to the $j$th row of $A$. By rule (D2), $D(A') = 0$. By rule (D1),

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

Part (b) follows immediately from condition (D1).

To prove part (c), we observe that we can interchange the $i$th and $j$th rows by the following sequence of operations:

- add the $i$th row to the $j$th;
- multiply the $i$th row by $-1$;
- add the $j$th row to the $i$th;
- subtract the $i$th row from the $j$th.

Symbolically,

$$\begin{bmatrix} \vdots \\ v_i \\ \vdots \\ v_j \\ \vdots \end{bmatrix} \xrightarrow{\mathrm{R}_j+\mathrm{R}_i} \begin{bmatrix} \vdots \\ v_i \\ \vdots \\ v_i+v_j \\ \vdots \end{bmatrix} \xrightarrow{-1\times\mathrm{R}_i} \begin{bmatrix} \vdots \\ -v_i \\ \vdots \\ v_i+v_j \\ \vdots \end{bmatrix} \xrightarrow{\mathrm{R}_i+\mathrm{R}_j} \begin{bmatrix} \vdots \\ v_j \\ \vdots \\ v_i+v_j \\ \vdots \end{bmatrix} \xrightarrow{\mathrm{R}_j-\mathrm{R}_i} \begin{bmatrix} \vdots \\ v_j \\ \vdots \\ v_i \\ \vdots \end{bmatrix}$$

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

We now understand how elementary row operations on the matrix $A$ affect the value of $D(A)$. The proof now proceeds in two cases, depending on whether $A$ is invertible.

- If $A$ is not invertible, then its row rank is less than $n$ (Corollary 2.18). So the rows of $A$ are linearly dependent, and one row can be written as a linear combination of the others. Suppose, without loss of generality that the first row $v_1$ can be written $v_1 = c_2v_2 + c_3v_3 + \cdots + c_nv_n$. Applying property (D1), we see that

$$D\begin{bmatrix} c_2v_2+c_3v_3+\cdots+c_nv_n \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = c_2D\begin{bmatrix} v_2 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} + c_3D\begin{bmatrix} v_3 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} + \cdots + c_nD\begin{bmatrix} v_n \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = 0.$$

Note that each of the terms in the above sum is zero by (D2), as each matrix has a repeated row. So, assuming (D1)–(D3) we have shown $D(A) = 0$. Since det satisfies (D1)–(D3) (Lemma 3.3), we know in particular that $\det(A) = 0$. So $D(A)$ and $\det(A)$ agree on non-invertible matrices $A$: they are both zero.

- If $A$ is invertible, then we can reduce it to the identity by applying elementary row operations (Corollary 2.20). Suppose that these row operations correspond to elementary matrices $R_1, R_2, \ldots, R_t$, so that $R_t \ldots R_2 R_1 A = I$. Each row operation $R_i$ multiplies $D()$ by a certain factor $c_i$, determined by (a)–(c). Thus, on the one hand, $D(R_t \ldots R_2 R_1 A) = c_1 c_2 \ldots c_t D(A) = cD(A)$, where $c = c_1 c_2 \ldots c_t$. On the other hand $R_t \ldots R_2 R_1 A = I$, and so $D(R_t \ldots R_2 R_1 A) = D(I) = 1$, by (D3). It follows that $D(A) = c^{-1}$. Again, we deduce in particular that $\det(A) = c^{-1}$. Thus, $D(A)$ and $\det(A)$ agree on invertible matrices.

Putting together the two cases, we see that if $D$ is any function satisfying (D1)–(D3), then $D(A) = \det(A)$ for all $A$. □

## 3.2 Properties of determinants

**Corollary 3.5.** *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 Theorem 3.4. □

Note that Theorem 3.4 immediately yields a result from *Linear Algebra I.*

**Lemma 3.6.** *(a) If $B$ is obtained from $A$ by adding $c$ times the $j$th row to the $i$th, then* $\det(B) = \det(A)$.

*(b) If $B$ is obtained from $A$ by multiplying the $i$th row by a scalar $c$, then* $\det(B) = c\det(A)$.

*(c) If $B$ is obtained from $A$ by interchanging two rows, then* $\det(B) = -\det(A)$.

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

**Theorem 3.7.** *If $A$ and $B$ are $n \times n$ matrices over $\mathbb{K}$, then* $\det(AB) = \det(A)\det(B)$.

*Proof.* Suppose first that $A$ is not invertible. Then $\det(A) = 0$. Also, $AB$ is not invertible. (For, suppose that $(AB)^{-1} = X$, so that $(AB)X = I = A(BX)$. Then $BX$ is the inverse of $A$.) So $\det(AB) = 0$, and the theorem holds in this case.

In the other case, $A$ is invertible, so by Corollary 2.19 we can write it as a product of elementary matrices $A = E_1 E_2 \cdots E_k$. Now observe that the theorem holds for the product of an elementary matrix $E$ and a general matrix $B$. Multiplying a matrix on the left by $E$ has the effect of multiplying the determinant by a certain factor $c$, depending only on $E$; thus $\det(EB) = c\det(B)$. For example, when $E$ is Type 3, $c = -1$ and $\det(EB) = -\det(B)$. It can be checked — do this now for all three types of elementary matrices! — that $c = \det(E)$. It follows that $\det(EB) = \det(E)\det(B)$.

More generally

$$\begin{aligned}\det(AB) &= \det(E_1E_2\dots E_kB)\\ &= \det(E_1)\det(E_2\dots E_kB) = \cdots\\ &= \det(E_1)\det(E_2)\dots\det(E_k)\det(B)\\ &= \det(E_1)\dots\det(E_{k-1}E_k)\det(B) = \cdots\\ &= \det(E_1E_2\dots E_k)\det(B)\\ &= \det(A)\det(B).\end{aligned}$$

as required. □

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

**Corollary 3.8.** *The determinant is the unique function $D$ of $n \times n$ matrices which satisfies the conditions*

(D1′) *for $1 \le i \le n$, $D$ is a linear function of the $i$th column;*

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

(D3′) $D(I_n) = 1$.

*Proof.* Swapping the roles of rows and columns in the Proof of Theorem 3.4 shows that there is a unique function satisfying (D1′)–(D3′) given by the formula

$$\det(A) = \sum_{\pi\in S_n} \operatorname{sign}(\pi) a_{\pi(1),1}a_{\pi(2),2}\cdots a_{\pi(n),n},$$

which is the usual formula, but with the role of rows and columns reversed. But this formula contains the same terms as the usual one, but in a different order. (Check this. The term corresponding to $\pi$ in the usual formula is equal, after rearrangement, to the term corresponding to $\pi^{-1}$ in the above formula. Furthermore, $\operatorname{sign}(\pi^{-1}) = \operatorname{sign}(\pi)$.) □

Since det() is the unique function on matrices satisfying (D1′)–(D3′) and (D1)–(D3), and since these properties are invariant under interchange of rows and columns, the same must be true of det() itself.

**Corollary 3.9.** *If $A^\top$ denotes the transpose of $A$, then* $\det(A^\top) = \det(A)$.

**Example 3.10.** Right at the end of *Geometry I* it was shown that the area of a parallelogram can be expressed as a $2 \times 2$ determinant. This fact extends to higher dimensions. The area of a parallelogram (see Figure 3.1) defined by vectors $u$ and $v_1$ is given by the length $|u|$ of the base $u$ times the height $h_1$, and similarly for the parallelograms defined by $u$ and $v_2$, and by $u$ and $v$. The height of the third parallelogram is the sum $h_1 + h_2$ of the heights of the other two. So its area is also the sum of the areas of the other two. (This can also be seen by dissecting the figure.)

Suppose we represent the parallelogram defined by $u$ and $v$ as a $2 \times 2$ matrix whose rows are $u$ and $v$ thus:

$$\begin{bmatrix} u \\ v \end{bmatrix}.$$

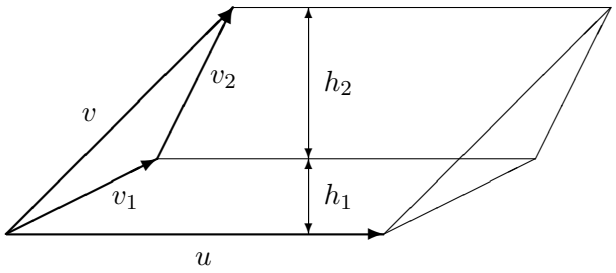


Figure 3.1: Area of a parallelogram

Note that all vectors will be *row* vectors in this example. Then we have seen that

$$\text{Area}\begin{bmatrix} u \\ v_1+v_2 \end{bmatrix} = \text{Area}\begin{bmatrix} u \\ v \end{bmatrix} = \text{Area}\begin{bmatrix} u \\ v_1 \end{bmatrix} + \text{Area}\begin{bmatrix} u \\ v_2 \end{bmatrix}.$$

Likewise,

$$= \text{Area}\begin{bmatrix} u \\ av \end{bmatrix} = a \times \text{Area}\begin{bmatrix} u \\ v \end{bmatrix}.$$

But these two observations together are telling us that "Area" satisfies property (D1)! Since the area of the degenerate parallelogram $\begin{bmatrix} u \\ u \end{bmatrix}$ is 0, we have that Area satisfies property (D2). Finally, the identity matrix represents the unit square, which has area 1, and property (D3) is satisfied also. But the determinant is the only function that satisfies (D1)–(D3), so the area of a parallelogram is given by the determinant of a $2 \times 2$ matrix.

Warning: this notion of area is signed, since the height of the parallelogram may be negative. If we negate one of the vectors, the sign of the area will be flip and if we transpose (in the sense of interchange!) the vectors then again the sign will flip, in accordance for the rules for elementary row operations.

All the above extends to $n$ dimensions. The *parallelotope* in $\mathbb{R}^n$ specified by vectors $v_1, v_2, \ldots, v_n$ is defined by

$$P = \{a_1v_1 + a_2v_2 + \cdots + a_nv_n : 0 \leq a_1, a_2, \ldots, a_n \leq 1\}.$$

Suppose we represent $P$ as an $n \times n$ matrix $A$ whose rows are the vectors $v_1, v_2, \ldots, v_n$. Then the $n$-dimensional volume of $P$ satisfies (D1)–(D3), and so $\text{vol}_n(P) = |\det(A)|$.

## 3.3 Cofactor (Laplace) expansion

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

**Definition 3.11.** Let $A$ be an $n \times n$ matrix. For $1 \leq i, j \leq n$, denote by $A_{i,j}$ the $(n-1) \times (n-1)$ matrix obtained by deleting the $i$th row and $j$th column of $A$. The $(i,j)$ *cofactor* of $A$ is defined to be $(-1)^{i+j}\det(A_{i,j})$. (These signs have a chessboard pattern, starting with sign + in the top left corner.) We denote the $(i,j)$ cofactor of $A$ by $K_{ij}(A)$. Note that the cofactor is a scalar, even though we've denoted it by an upper case latter! Finally, the *adjugate* of $A$ is the $n \times n$ matrix $\text{Adj}(A)$ whose $(i,j)$ entry is the $(j,i)$ cofactor $K_{ji}(A)$ of $A$. (Note the transposition!)

**Theorem 3.12.** (a) *For $1 \le i \le n$, we have*

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

(b) *For $1 \le j \le n$, we have*

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

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 $A$. The expressions (a) and (b) appearing in Theorem 3.12 are the *cofactor* or *Laplace expansion* along row $i$ and column $j$ respectively.

**Example 3.13.** 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 $2 \times 2$ determinant.

Theorem 3.12 looks plausible. Note that $K_{ij}(A)$ is an $(n-1) \times (n-1)$ determinant. Expanding $a_{ij}K_{ij}(A)$ by the Leibniz formula yields $(n-1)!$ terms that *ought* to correspond to those $(n-1)!$ terms in the Leibniz formula for $A$ that satisfy $\pi(i) = j$. But keeping track of the subscripts and the signs is fiddly and not very edifying, so we won't go into that here. The details can be found on the Wikipedia page on the "Laplace expansion".

Another way of going about proving Theorem 3.12 is to show that the expression $\sum_{j=1}^{n} a_{ij}K_{ij}(A)$ satisfies (D1)–(D3). The issue here is how to show (D2) when one of the two equal rows is row $i$. Again, there is no essential problem but we won't go into the details here.

At first sight, the Laplace expansion looks like a simple formula for the determinant, since it is just the sum of $n$ terms, rather than $n!$ as in the Leibniz formula. 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 3.14.** *For any $n \times n$ matrix $A$, we have*

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

**Remark 3.15.** In the above identity, the $A \cdot \mathrm{Adj}(A)$ and $\mathrm{Adj}(A) \cdot A$ are *matrix* products, while $\det(A)\, I$ is the product of a *scalar* with a matrix. We can get into big trouble by ignoring this distinction and using matrices where scalars should go. Just in this section, we'll use dots to emphasise *matrix* multiplication.

*Proof of Theorem 3.14.* 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 as that of $A$ except for the $j$th row, which has been replaced by the $i$th row of $A$. (Note that changing the $j$th 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 $\det(A)\, I$.

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

**Corollary 3.16.** *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 $a_{11}, \ldots, a_{nn}$, and all off-diagonal entries are zero, then $\det(A)$ is just the product $a_{11} \cdots a_{nn}$.

**Example 3.17.** 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 $2 \times 2$ matrix rather than do any further reduction.)

## 3.4 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 3.18.** Let $A$ be a $n \times n$ matrix. The *characteristic polynomial* of $A$ is the polynomial

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

This is a polynomial in $x$ of degree $n$.

For example, if

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

then

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

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

**Theorem 3.19** (Cayley–Hamilton Theorem)**.** *Let $A$ be an $n \times n$ matrix with characteristic polynomial $p_A(x)$. Then $p_A(A) = O$.*

**Example 3.20.** Let us just check the theorem for $2 \times 2$ matrices. If

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

then

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

and so

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

after a small amount of calculation.

*Proof.* We use the theorem

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

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

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

Now it is very tempting (for lesser beings than the MTH6140 class) just to substitute $x = A$ into this formula: on the right we have $p_A(A)\, I = p_A(A)$, while on the left there is a factor $AI - A = O$. Unfortunately this is not valid, and the reason is connected to the remark following the statement of Theorem 3.14. The expression $p_A(A)$ is a matrix, and not valid in this context, where a scalar is expected. (A similar problem exists on the left side of the incorrect identity.)

Instead, we argue as follows. $\mathrm{Adj}(xI - A)$ 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

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

So, if we let

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

then we read off that

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

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

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

So $p_A(A) = O$, as claimed. □

## Summary

- The determinant of an $n \times n$ matrix can be defined by an explicit formula (Leibniz) with $n!$ terms.
- It can also be defined axiomatically, as the unique function on square matrices satisfying certain simple properties.
- The determinant of a matrix transforms in predictable ways under the action of elementary row and column operations (equivalently, left or right multiplication by elementary matrices).
- The determinant of the product of matrices is the product of their determinants.
- The determinant of a matrix can also be defined recursively by the cofactor (Laplace) expansion.
- The inverse of a non-singular matrix can be expressed explicitly in terms of its cofactors (adjugate matrix).
- The determinant can be used to define the characteristic polynomial of a matrix.
- The Cayley-Hamilton Theorem states that every square matrix satisfies its own characteristic polynomial. (It is important to know precisely what this informal statement means.)

# Chapter 4

# 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.

## 4.1 Definition and basic properties

**Definition 4.1.** Let $V$ and $W$ be vector spaces over a field $\mathbb{K}$. A function $\alpha$ 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 $c \in \mathbb{K}$.

**Remark 4.2.** (a) We can combine the two conditions into one as follows:

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

(b) In other literature the term "linear transformation" is often used instead of "linear map".

**Definition 4.3.** Let $\alpha : V \to W$ be a linear map. The *image* of $\alpha$ is the set

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

and the *kernel* of $\alpha$ is

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

**Proposition 4.4.** *Let $\alpha : V \to W$ be a linear map. Then the image of $\alpha$ is a subspace of $W$ and the kernel is a subspace of $V$.*

*Proof.* We have to show that each subset is closed under addition and scalar multiplication, and is non-empty. Non-emptiness is immediate: the zero vector $\mathbf{0}$ is in both $\mathrm{Im}(\alpha)$ and $\mathrm{Ker}(\alpha)$ since $\alpha(\mathbf{0}) = \mathbf{0}$.

Suppose that $w_1, w_2$ are vectors in the image of $\alpha$. By definition of $\mathrm{Im}(\alpha)$, there exist $v_1, v_2 \in V$ such that $w_1 = \alpha(v_1)$ and $w_2 = \alpha(v_2)$. Then

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

by linearity of $\alpha$. It follows that $w_1 + w_2 \in \mathrm{Im}(\alpha)$. Now suppose $w \in \mathrm{Im}(\alpha)$ and $c \in \mathbb{K}$. By definition of $\mathrm{Im}(\alpha)$, there exists $v \in V$ such that $w = \alpha(v)$. Then

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

demonstrating that $cw \in \mathrm{Im}(\alpha)$.

Next suppose $v_1, v_2$ are vectors in the kernel of $\alpha$. By definition of $\mathrm{Ker}(\alpha)$, we know that $\alpha(v_1) = \alpha(v_2) = \mathbf{0}$. Thus

$$\alpha(v_1 + v_2) = \alpha(v_1) + \alpha(v_2) = \mathbf{0} + \mathbf{0} = \mathbf{0},$$

from which it follows that $v_1 + v_2 \in \mathrm{Ker}(\alpha)$. Finally, suppose $v \in \mathrm{Ker}(\alpha)$ and $c \in \mathbb{K}$. Then $\alpha(v) = \mathbf{0}$ and

$$\alpha(cv) = c\alpha(v) = c\,\mathbf{0} = \mathbf{0},$$

demonstrating that $cv \in \mathrm{Ker}(\alpha)$. □

**Definition 4.5.** We define the *rank* of $\alpha$ to be $\varrho(\alpha) = \dim(\mathrm{Im}(\alpha))$ and the *nullity* of $\alpha$ 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 4.6** (Rank–Nullity Theorem)**.** *Let $\alpha : V \to W$ be a linear map. Then $\varrho(\alpha) + \nu(\alpha) = \dim(V)$.*

*Proof.* Choose a basis $u_1, u_2, \ldots, u_q$ for $\mathrm{Ker}(\alpha)$, where $q = \dim(\mathrm{Ker}(\alpha)) = \nu(\alpha)$. 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 $c_1\alpha(v_1) + \cdots + c_s\alpha(v_s) = \mathbf{0}$. We need to show that $c_1 = \cdots = c_s = 0$. Applying the linear map $\alpha$ we have

$$\alpha(c_1v_1 + \cdots + c_sv_s) = c_1\alpha(v_1) + \cdots + c_s\alpha(v_s) = \mathbf{0},$$

so that $c_1v_1 + \cdots + c_sv_s \in \mathrm{Ker}(\alpha)$. The vector $c_1v_1 + \cdots + c_sv_s$ 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

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

But the list $(u_1, \ldots, u_q, v_1, \ldots, v_s)$ is a basis for $V$, and hence is linearly independent. It follows that $c_1 = \cdots = c_s = 0$ (and incidentally $a_1 = \cdots = a_q = 0$), as required.

Spanning: Take any vector in $\mathrm{Im}(\alpha)$, say $w$. We need to show that $w \in \langle \alpha(v_1), \ldots, \alpha(v_s) \rangle$. Since $w \in \mathrm{Im}(\alpha)$ we know that $w = \alpha(v)$ for some $v \in V$. Write $v$ in terms of the basis for $V$:

$$v = a_1 u_1 + \cdots + a_q u_q + c_1 v_1 + \cdots + c_s v_s$$

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

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

where we used the fact that $u_i \in \mathrm{Ker}(\alpha)$ and hence $\alpha(u_i) = \mathbf{0}$. So the vectors $\alpha(v_1), \ldots, \alpha(v_s)$ span $\mathrm{Im}(\alpha)$.

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

## 4.2 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 $\dim(V) = n$ and $\dim(W) = m$. Let $v_1, \ldots, v_n$ be a basis for $V$ and $w_1, \ldots, w_m$ a basis for $W$. Then for $j = 1, \ldots, n$, the vector $\alpha(v_j)$ belongs to $W$, so we can write it as a linear combination of $w_1, \ldots, w_m$.

**Definition 4.7.** The matrix representing the linear map $\alpha : V \to W$ relative to the bases $(v_1, \ldots, v_n)$ for $V$ and $(w_1, \ldots, w_m)$ for $W$ is the $m \times n$ matrix whose $(i, j)$ entry is $a_{ij}$, where

$$\alpha(v_j) = \sum_{i=1}^{m} a_{ij} w_i$$

for $j = 1, \ldots, n$. (The indices on the right hand side are reversed from what you might expect by analogy with matrix multiplication, but it will all turn out right in the end!)

In practice this means the following. Take $\alpha(v_j)$ and write it as a as a linear combination $\alpha(v_j) = a_{1j}w_1 + \cdots + a_{mj}w_m$ of basis vectors of $W$. Then the column vector $\begin{bmatrix} a_{1j} & a_{2j} & \cdots & a_{mj} \end{bmatrix}^\top$ is the $j$th column of the matrix representing $\alpha$. So, for example, if $n = 3$, $m = 2$, and

$$\alpha(v_1) = w_1 + w_2, \quad \alpha(v_2) = 2w_1 + 5w_2, \quad \alpha(v_3) = 3w_1 - w_2,$$

then the matrix representing $\alpha$ is

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

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

**Proposition 4.8.** *Let $\alpha : V \to W$ be a linear map. Choose bases $\mathcal{B}$ for $V$ and $\mathcal{B}'$ for $W$ and let $A$ be the matrix representing $\alpha$ relative to these bases. Then*

$$[\alpha(v)]_{\mathcal{B}'} = A[v]_{\mathcal{B}}.$$

*Proof.* Let $\mathcal{B} = (v_1, \ldots, v_n)$ be the basis for $V$, and $\mathcal{B}' = (w_1, \ldots, w_m)$ the basis for $W$. Suppose $v = \sum_{j=1}^{n} c_j v_j \in V$, so that in coordinates

$$[v]_{\mathcal{B}} = \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix}.$$

Then

$$\alpha(v) = \sum_{j=1}^{n} c_j \alpha(v_j) = \sum_{j=1}^{n} c_j \sum_{i=1}^{m} a_{ij} w_i = \sum_{i=1}^{m} w_i \sum_{j=1}^{n} a_{ij} c_j,$$

so the $i$th coordinate of $[\alpha(v)]_{\mathcal{B}'}$ is $\sum_{j=1}^{n} a_{ij} c_j$, which is precisely the $i$th coordinate in the matrix product $A[v]_{\mathcal{B}}$. □

In our example, if $v = 2v_1 + 3v_2 + 4v_3$, so that the coordinate representation of $v$ relative to the basis $(v_1, v_2, v_3)$ is $\begin{bmatrix} 2 & 3 & 4 \end{bmatrix}^\top$, then

$$[\alpha(v)]_{\mathcal{B}'} = A[v]_{\mathcal{B}} = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 5 & -1 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix} = \begin{bmatrix} 20 \\ 13 \end{bmatrix}.$$

The column vector on the right gives the coordinate representation of $\alpha(v)$ relative to the basis $(w_1, w_2)$, that is, $\alpha(v) = 20w_1 + 13w_2$.

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

**Definition 4.9.** Let $\alpha$ and $\beta$ 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 routine to check that $\alpha + \beta$ is a linear map.

**Proposition 4.10.** *If $\alpha$ and $\beta$ are linear maps represented relative to some basis by matrices $A$ and $B$, respectively, then $\alpha + \beta$ is represented by the matrix $A + B$, relative to the same basis.*

The proof of this is not too difficult: just apply the definitions as in the Proof of Proposition 4.12 below.

**Definition 4.11.** Let $U$, $V$ and $W$ 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 routine to check that $\beta\alpha$ is a linear map. Note that the order is important: we take a vector $u \in U$, apply $\alpha$ to it to get a vector in $V$, and then apply $\beta$ to get a vector in $W$. So $\beta\alpha$ means "apply $\alpha$, then $\beta$".

**Proposition 4.12.** *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 $BA$.*

*Proof.* Suppose linear maps $\alpha$ and $\beta$ are represented by matrices $A$ and $B$ relative to bases $\mathcal{B}$ of $U$, $\mathcal{B}'$ of $V$, and $\mathcal{B}''$ of $W$. Then

$$[(\beta\alpha)u]_{\mathcal{B}''} = [\beta(\alpha(u))]_{\mathcal{B}''} = B\,[\alpha(u)]_{\mathcal{B}'} = B(A[u]_{\mathcal{B}}) = (BA)[u]_{\mathcal{B}},$$

where we have used, in turn, the definition of product of maps, Proposition 4.8 (twice) and associatively of matrix multiplication. □

**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 4.12 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.

## 4.3 Change of basis

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

Recall the notion of *transition matrix* from Chapter 1. If $\mathcal{B} = (v_1, \ldots, v_n)$ and $\mathcal{B}' = (v'_1, \ldots, v'_n)$ are two bases for a vector space $V$ of dimension $n$, then the transition matrix $P_{\mathcal{B},\mathcal{B}'}$ is the matrix whose $j$th column is the coordinate representation of $v'_j$ relative to the basis $\mathcal{B}$. We saw that

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

where $[v]_{\mathcal{B}}$ is the coordinate representation of an arbitrary vector $v$ relative to the basis $\mathcal{B}$, and similarly for $\mathcal{B}'$. The transition matrix $P_{\mathcal{B}',\mathcal{B}}$ that transforms $[v]_{\mathcal{B}}$ back to $[v]_{\mathcal{B}'}$ is just the inverse of the matrix $P_{\mathcal{B},\mathcal{B}'}$.

**Proposition 4.13.** *Let $\alpha : V \to W$ be a linear map represented by matrix $A$ relative to the bases $\mathcal{B}$ for $V$ and $\mathcal{C}$ for $W$, and by the matrix $A'$ relative to the bases $\mathcal{B}'$ for $V$ and $\mathcal{C}'$ for $W$. If $P = P_{\mathcal{C}',\mathcal{C}}$ and $Q = P_{\mathcal{B},\mathcal{B}'}$ are transition matrices relating the unprimed to the primed bases, then*

$$A' = PAQ.$$

*Proof.* At a high level the claim seems reasonable. Suppose we apply the matrix $A'$ to a coordinate representation of some vector relative to the primed basis for $V$. Multiplication by $Q$ will transform from the primed to unprimed basis, multiplication by $A$ will apply the linear transformation relative to the unprimed bases, and finally $P$ will transform back to the primed basis.

We just need to write that scheme down in symbols, which is not too difficult:

$$(PAQ)[v]_{\mathcal{B}'} = PA(Q\,[v]_{\mathcal{B}'}) = P(A\,[v]_{\mathcal{B}}) = P[\alpha(v)]_{\mathcal{C}} = [\alpha(v)]_{\mathcal{C}'}.$$

So $A' = PAQ$ is the representation of the linear map $\alpha$ relative to the primed bases. □

In practical terms, the above result is needed for explicit calculations. For theoretical purposes its importance lies 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.

**Corollary 4.14.** *Any two matrices that represent the same linear map relative to different bases are equivalent.*

Although we shall not need the fact, the converse is also true: any two equivalent matrices can be viewed representations of a single linear map relative to two different bases. This is so because any invertible matrix can be viewed as a transition matrix.

## 4.4 Canonical form revisited

We return Theorem 2.3 about canonical forms for equivalence with a view to showing that rank of a linear map and rank of a matrix are essentially the same thing.

**Theorem 4.15.** *Let $\alpha : V \to W$ be a linear map of rank $r = \varrho(\alpha)$. Then there are bases for $V$ and $W$ such that the matrix representing $\alpha$ is, in block form,*

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

*Proof.* As in the proof of Theorem 4.6, choose a basis $v_1, \ldots, v_r, v_{r+1}, \ldots, v_n$ for $V$ such that $v_{r+1}, \ldots, v_n$ is a basis for $\mathrm{Ker}(\alpha)$. (We can do this by choosing the basis $v_{r+1}, \ldots, v_n$ of $\mathrm{Ker}(\alpha)$ first, and then extending it to a basis for the whole space $V$.)

As we saw earlier, $w_1 = \alpha(v_1), \ldots, w_r = \alpha(v_r)$ is a basis for $\mathrm{Im}(\alpha)$, and can be extended to a basis $w_1, \ldots, w_r, w_{r+1}, \ldots, w_m$ of $W$. We have

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

so the matrix of $\alpha$ 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. It is now not difficult to see that rank of a linear map and rank of a matrix are consistent.

**Corollary 4.16.** *Suppose $\alpha : V \to W$ is a linear map of rank $r$. For any choice of bases $\mathcal{B}$ for $V$ and $\mathcal{B}'$ for $W$, the rank of the matrix representing $\alpha$ relative to $\mathcal{B}$ and $\mathcal{B}'$ is also $r$.*

*Proof.* We know from Theorem 4.15 that there is some choice of bases for which the matrix $A$ representing $\alpha$ takes the canonical form. In this case the rank of the linear map $\alpha$ and the matrix $A$ certainly agree. Any other matrix $A'$ representing $\alpha$ will be equivalent to $A$ by Proposition 4.13. Equivalent matrices have the same rank (Theorem 2.21), so the rank of $A'$ is also $r$. □

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 $\min\{m, n\} + 1$. Thus the number of distinct linear maps from a vector space of dimension $n$ to one of dimension $m$ is also $\min\{m, n\} + 1$.

## Summary

- A map $\alpha : V \to W$ is linear if it interacts nicely with vector addition and scalar multiplication.

- The kernel and image are important subspaces associated with a linear map $\alpha$. The kernel is a subspace of the domain of $\alpha$ and the image is a subspace of the of the codomain.
- The dimensions of the kernel and the image are related by the rank-nullity theorem.
- Relative to a given basis, a linear map $\alpha$ is represented by a matrix $A$.
- Addition of linear maps corresponds to addition of matrices; the product of linear maps corresponds to multiplication of matrices.
- The matrix $A$ depends on the basis. The transition matrices from Chapter 2 can be used to transform the matrix from one basis to another.
- Matrices representing the same linear map relative to different bases are equivalent.
- If $A$ represents $\alpha$ relative to a certain basis, then the rank or $A$ is equal to the rank of $\alpha$.

# Chapter 5

# Linear maps on a vector space

In this chapter we consider a linear map $\alpha$ from a vector space $V$ to itself. If $\dim(V) = n$ then, as in the last chapter, we can represent $\alpha$ 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!

## 5.1 Projections and direct sums

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

**Definition 5.1.** The linear map $\pi : V \to V$ is a *projection* if $\pi^2 = \pi$ (where, as usual, $\pi^2$ is defined by $\pi^2(v) = \pi(\pi(v))$).

**Proposition 5.2.** *If $\pi : V \to V$ is a projection, then $V = \mathrm{Im}(\pi) \oplus \mathrm{Ker}(\pi)$.*

Before starting the proof, it is worth making a tiny observation that will simplify our task here and later in the chapter. Suppose $\pi$ is a projection on $V$ and $v \in V$. Then we claim that $v \in \mathrm{Im}(\pi)$ if and only if $\pi(v) = v$. The "if" direction is immediate: there *is* a vector $u \in V$, namely $u = v$, such that $v = \pi(u)$. The "only if" direction is hardly more difficult: $v \in \mathrm{Im}(\pi)$ implies that there exists $u \in V$ such that $v = \pi(u)$. Then $\pi(v) = \pi(\pi(u)) = \pi^2(u) = \pi(u) = v$.

*Proof of Proposition 5.2.* We have two things to show:

$\mathrm{Im}(\pi) + \mathrm{Ker}(\pi) = V$: 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

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

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

$\mathrm{Im}(\pi) \cap \mathrm{Ker}(\pi) = \{\mathbf{0}\}$: Take $v \in \mathrm{Im}(\pi) \cap \mathrm{Ker}(\pi)$. Since $v$ is in $\mathrm{Im}(\pi)$ we know that $\pi(v) = v$ (see above). Also, since $v$ is in $\mathrm{Ker}(\pi)$, we have $\pi(v) = \mathbf{0}$. Putting these facts together yields $v = \mathbf{0}$.

□

There is a converse to this result.

**Proposition 5.3.** *If $V = U \oplus W$, then there is a projection $\pi : V \to V$ with $\mathrm{Im}(\pi) = U$ and $\mathrm{Ker}(\pi) = W$.*

*Proof.* (Sketch.) Every vector $v \in V$ can be uniquely written as $v = u + w$, where $u \in U$ and $w \in W$; we define $\pi$ by the rule that $\pi(v) = u$. You should check that with this definition for $\pi$ it is indeed the case that $\mathrm{Im}(\pi) = U$ and $\mathrm{Ker}(\pi) = W$, and that $\pi$ is indeed a projection. □

The diagram in Figure 5.1 shows geometrically what a projection is. It moves any vector $v$ in a direction parallel to $\mathrm{Ker}(\pi)$ to a vector lying in $\mathrm{Im}(\pi)$.

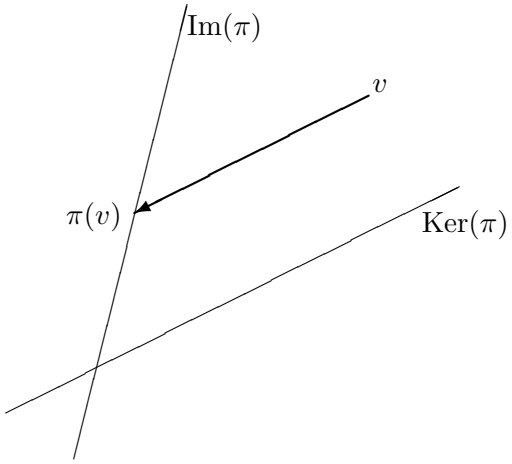


Figure 5.1: A projection

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

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

Note also that $\pi + \pi' = I$ and $\pi\pi' = \pi(I - \pi) = \pi - \pi^2 = 0$. It follows (as we shall see below) that $\mathrm{Ker}(\pi) = \mathrm{Im}(\pi')$, and hence $V = \mathrm{Im}(\pi) \oplus \mathrm{Im}(\pi')$. These observations show the way to generalise Proposition 5.2.

**Proposition 5.4.** *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 map;*

*(b) $\pi_i \pi_j = 0$ for $i \neq j$.*

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

*Proof.* (Sketch.) Let $v$ be any vector in $V$. Using the fact that $\pi_1 + \pi_2 + \cdots + \pi_r = I$ we have

$$\begin{aligned} v = I(v) &= (\pi_1 + \pi_2 + \cdots + \pi_r)(v) = \pi_1(v) + \pi_2(v) + \cdots + \pi_r(v) \\ &= u_1 + u_2 + \cdots + u_r, \end{aligned} \tag{5.1}$$

where $u_i = \pi_i(v) \in \mathrm{Im}(\pi_i) = U_i$ for $i = 1, \ldots, r$. To complete the proof, we need to check that the decomposition $v = u_1 + \cdots + u_r$ is unique. □

There is a converse to the above result.

**Proposition 5.5.** *Suppose $V$ is a vector space which is the direct sum of $r$ subspaces: $V = U_1 \oplus U_2 \oplus \cdots \oplus U_r$. Then there exists projections $\pi_1, \pi_2, \ldots, \pi_r$ on $V$ satisfying*

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

*(b) $\pi_i \pi_j = 0$ for $i \neq j$; and*

*(c) $U_i = \mathrm{Im}(\pi_i)$ for all $i$.*

*Proof.* (Sketch.) Since $V = U_1 \oplus U_2 \oplus \cdots \oplus U_r$, 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 may define $\pi_i(v) = u_i$, for $i = 1, \ldots, r$. To complete the proof, we need to verify that $\{\pi_i\}$ are projections with the required properties. □

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

## 5.2 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 $\alpha$ is represented by a matrix $A$, say, where

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

Then just as in the last section, the action of $\alpha$ 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 matrices representing them: 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 5.6.** *Let $\alpha$ be a linear map on $V$ which is represented by the matrix $A$ relative to a basis $\mathcal{B}$, and by the matrix $A'$ relative to a basis $\mathcal{B}'$. Let $P = P_{\mathcal{B},\mathcal{B}'}$ be the transition matrix between the two bases. Then*

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

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

**Definition 5.7.** Two $n \times n$ matrices $A$ and $B$ are said to be *similar* if $B = P^{-1} A P$ 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.

## 5.3 Eigenvalues and eigenvectors

**Definition 5.8.** Let $\alpha$ be a linear map on $V$. A vector $v \in V$ is said to be an *eigenvector* of $\alpha$, with *eigenvalue* $\lambda \in \mathbb{K}$, if $v \neq \mathbf{0}$ and $\alpha(v) = \lambda v$. The set $\{v : \alpha(v) = \lambda v\}$ consisting of the zero vector and the eigenvectors with eigenvalue $\lambda$ is called the $\lambda$-*eigenspace* of $\alpha$, and we'll denote it by $E(\lambda, \alpha)$.

It is not difficult to check that an eigenspace $E(\lambda, \alpha)$ as defined above is a linear subspace of $V$. (Do this!) Note that we require that $v \neq \mathbf{0}$ for any eigenvector of $\alpha$, otherwise the zero vector would be an eigenvector of $\alpha$ for any value of $\lambda$. With this requirement, each eigenvector has a unique eigenvalue: for if $\alpha(v) = \lambda v = \mu v$, then $(\lambda - \mu)v = \mathbf{0}$, and so (since $v \neq \mathbf{0}$) 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 $\alpha$, and we have to extract it by some procedure. We'll see how to do this soon.

**Example 5.9.** 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.

## 5.4 Diagonalisability

Some linear maps have a particularly simple representation by matrices.

**Definition 5.10.** The linear map $\alpha$ on $V$ is *diagonalisable* if and only if there is a basis of $V$ consisting of eigenvectors of $\alpha$.

Suppose that $v_1, \ldots, v_n$ is such a basis showing that $\alpha$ is diagonalisable, and that $A = (a_{ij})$ is the matrix representing $\alpha$ in this basis. Since $a_{ij} = 0$ whenever $i \neq j$, we have that $A$ is diagonal. (Note that the diagonal entries of $A$ are the eigenvalues of $\alpha$.) Conversely, if the matrix $A$ is diagonal then all the basis vectors are eigenvectors. So we have:

**Proposition 5.11.** *The linear map $\alpha$ on $V$ is diagonalisable if there is a basis of $V$ relative to which the matrix representing $\alpha$ is a diagonal matrix.*

**Example 5.12.** The matrix from Example 5.9 is diagonalisable, as the two eigenvectors we computed there do form a basis of $\mathbb{R}^2$.

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 $\begin{bmatrix} 1 & 0 \end{bmatrix}^\top$. So we cannot find a basis of eigenvectors.

Before looking at some equivalent characterisations of diagonalisability, we require a preparatory lemma.

**Lemma 5.13.** *Let $v_1, \ldots, v_r$ be eigenvectors of $\alpha$ with distinct eigenvalues $\lambda_1, \ldots, \lambda_r$. Then $v_1, \ldots, v_r$ are linearly independent.*

*Proof.* Suppose to the contrary that $v_1, \ldots, v_r$ are linearly dependent, so that there exists a linear relation

$$c_1 v_1 + \cdots + c_r v_r = \mathbf{0}, \tag{5.2}$$

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. It is clear that there must be at least two non-zero coefficients. Suppose that $c_1 \neq 0$. (If $c_1 = 0$ just re-number the eivenvectors and their coefficients.) Now, applying $\alpha$ to both sides of (5.2) and using the fact that $\alpha(v_i) = \lambda_i v_i$, we get

$$\alpha(c_1 v_1 + \cdots + c_r v_r) = c_1 \alpha(v_1) + \cdots + c_r \alpha(v_r) = c_1 \lambda_1 v_1 + \cdots + c_r \lambda_r v_r = \mathbf{0}.$$

Subtracting $\lambda_1$ times equation (5.2) from the last equation we get

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

Now this equation has one fewer non-zero coefficient than the one we started with, which was assumed to have the smallest possible number. And since we started with at least two non-zero coefficients, not all the coefficients in this new identity are zero. So the linear dependency (5.2) is not minimal, contrary to our assumption. So the eigenvectors must have been linearly independent. □

Note that Lemma 5.13 implies, in particular, that a linear map $\alpha : V \to V$ has at most $n$ distinct eigenvalues, where $n = \dim(V)$.

**Theorem 5.14.** *Suppose $\alpha : V \to V$ is a linear map, and let $\lambda_1, \ldots, \lambda_r$ be the distinct eigenvalues of $\alpha$. Then the following are equivalent:*

*(a)* $\alpha$ *is diagonalisable;*

*(b)* $V = E(\lambda_1, \alpha) \oplus \cdots \oplus E(\lambda_r, \alpha)$ *is the direct sum of eigenspaces of* $\alpha$*;*

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

*Proof.* We just prove the equivalence of (a) and (b) in this module.

(a) $\Rightarrow$ (b). If $\alpha$ is diagonalisable, then there is a basis of $V$ composed of eigenvectors of $\alpha$. Each of these basis vectors lies in one of the eigenspaces; thus, $V = E(\lambda_1, \alpha) + \cdots + E(\lambda_r, \alpha)$. We need to show that this sum is actually a direct sum. If some vector $v \in V$ may be expressed in two different ways $u_1 + \cdots + u_r = u_1' + \cdots + u_r'$, with $u_i, u_i' \in E(\lambda_i, \alpha)$, for $i = 1, \ldots, r$, then $(u_1 - u_1') + \cdots + (u_r - u_r') = \mathbf{0}$. Each of these terms must be zero, otherwise we would have a non-trivial linear dependency between eigenvectors with distinct eigenvectors, which is disallowed by Lemma 5.13.

(b) $\Rightarrow$ (a). Let $\mathcal{B}_i$ be a basis for $E(\lambda_i, \alpha)$ for $i = 1, \ldots, r$. Then, by Proposition 1.28, $\mathcal{B} = \mathcal{B}_1 \cup \cdots \cup \mathcal{B}_r$ is a basis for $V$; it is clearly composed of eigenvectors of $\alpha$. □

**Example 5.15.** Continuing our previous exercise, 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

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

so that $AP = PD$ where

$$P = \begin{bmatrix} 3 & 2 \\ 4 & 3 \end{bmatrix} \quad \text{and} \quad D = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}.$$

Note that the columns of $P$ are the eigenvectors of $A$, and $D$ is a diagonal matrix formed from the eigenvalues of $A$. (Of course, we must list the eigenvectors and the eigenvalues in a consistent order!) Also note that $P^{-1}AP = D$, so $A$ is similar to a diagonal matrix. Since $A = PDP^{-1}$, we may write $A$ as

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

Furthermore, we can find two projection matrices as follows:

$$\Pi_1 = \begin{bmatrix} 3 & 2 \\ 4 & 3 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 3 & -2 \\ -4 & 3 \end{bmatrix} = \begin{bmatrix} 9 & -6 \\ 12 & -8 \end{bmatrix}$$
$$\Pi_2 = \begin{bmatrix} 3 & 2 \\ 4 & 3 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 3 & -2 \\ -4 & 3 \end{bmatrix} = \begin{bmatrix} -8 & 6 \\ -12 & 9 \end{bmatrix}.$$

(Note that we have replaced $D$ in the previous expression for $A$ by a matrices with a single 1 on the diagonal.) You can check directly that $\Pi_1^2 = \Pi_1$, $\Pi_2^2 = \Pi_2$, $\Pi_1\Pi_2 = \Pi_2\Pi_1 = O$, $\Pi_1 + \Pi_2 = I$, and $2\Pi_1 + 3\Pi_2 = A$. You should stop for a moment to think about why this calculational method works.

## 5.5 Characteristic and minimal polynomials

We defined the determinant of a square matrix $A$. Now we want to define the determinant of a linear map $\alpha$. The obvious way to do this is to take the determinant of any matrix representing $\alpha$. 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 5.16.** (a) The *determinant* $\det(\alpha)$ of a linear map $\alpha : V \to V$ is the determinant of any matrix representing $\alpha$.

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

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

The second part of the definition is OK, by the same reasoning as the first, since $p_A(x)$ is just a determinant. Specifically, the characteristic polynomial of a matrix $A' = P^{-1}AP$ similar to $A$ is

$$\begin{aligned} p_{A'}(x) &= \det(xI - P^{-1}AP) \\ &= \det(P^{-1}(xI - A)P) \\ &= \det(P^{-1})\det(xI - A)\det(P) \\ &= \det(xI - A) \\ &= p_A(x). \end{aligned}$$

The third part of the definition also requires care. We know from that Cayley-Hamilton Theorem that there is some polynomial (namely the characteristic polynomial) that is satisfied by $\alpha$. But is the minimal polynomial, as defined, unique? Well, suppose that there were two different monic polynomials $m_\alpha(x)$ and $m'_\alpha(x)$ of minimum degree satisfying $m_\alpha(\alpha) = m'_\alpha(\alpha) = 0$. Then the polynomial $(m_\alpha - m'_\alpha)(x)$ satisfies $(m_\alpha - m'_\alpha)(\alpha) = m_\alpha(\alpha) - m'_\alpha(\alpha) = 0$, and is of lower degree than $m_\alpha(x)$ or $m'_\alpha(x)$. Since we can make this polynomial monic by multiplication by an appropriate scalar, this is a contradiction to minimality of $m_\alpha(x)$. The next result gives more information.

**Proposition 5.17.** *For any linear map $\alpha$ on $V$, its minimal polynomial $m_\alpha(x)$ divides its characteristic polynomial $p_\alpha(x)$ (as polynomials).*

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

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

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

**Theorem 5.18.** *Let $\alpha$ be a linear map on $V$. Then the following conditions are equivalent for an element $\lambda$ of $\mathbb{K}$:*

*(a) $\lambda$ is an eigenvalue of $\alpha$;*

*(b) $\lambda$ is a root of the characteristic polynomial of $\alpha$;*

*(c) $\lambda$ is a root of the minimal polynomial of $\alpha$.*

**Example 5.19.** This gives us a recipe to find the eigenvalues of $\alpha$: take a matrix $A$ representing $\alpha$; write down its characteristic polynomial $p_A(x) = \det(xI - A)$; and find the roots of this polynomial. In our earlier example,

$$\begin{vmatrix} x+6 & -6 \\ 12 & x-11 \end{vmatrix} = (x+6)(x-11) + 72 = x^2 - 5x + 6 = (x-2)(x-3),$$

so the eigenvalues are 2 and 3, as we found.

*Proof of Theorem 5.18.* (a) $\Rightarrow$ (c). Let $\lambda$ be an eigenvalue of $\alpha$ 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 $m_\alpha(\alpha)(v) = m_\alpha(\lambda)\, v$. But $m_\alpha(\alpha) = 0$ by definition, so $m_\alpha(\lambda)\, v = \mathbf{0}$. Since $v \neq \mathbf{0}$, we have $m_\alpha(\lambda) = 0$, as required.

(c) $\Rightarrow$ (b). Suppose that $\lambda$ is a root of $m_\alpha(x)$. Then $(x - \lambda)$ divides $m_\alpha(x)$. But $m_\alpha(x)$ divides $p_\alpha(x)$, by Proposition 5.17, so $(x - \lambda)$ divides $p_\alpha(x)$, whence $\lambda$ is a root of $p_\alpha(x)$.

(b) $\Rightarrow$ (a). Suppose that $p_\alpha(\lambda) = 0$, that is, $\det(\lambda I - \alpha) = 0$. Then $\lambda I - \alpha$ is not of full rank (i.e., the dimension of $\mathrm{Im}(\lambda I - \alpha)$ is strictly less than $\dim(V)$), so kernel of $\lambda I - \alpha$ has dimension greater than zero. Pick a non-zero vector $v$ in $\mathrm{Ker}(\lambda I - \alpha)$. Then $(\lambda I - \alpha)v = \mathbf{0}$, so that $\alpha(v) = \lambda v$; that is, $\lambda$ is an eigenvalue of $\alpha$. □

Using this result, we can give a necessary and sufficient condition for $\alpha$ to be diagonalisable.

**Theorem 5.20.** *The linear map $\alpha$ on $V$ is diagonalisable if and only if its minimal polynomial is the product of distinct linear factors.*

*Proof.* Suppose first that $\alpha$ is diagonalisable. By definition, there is a basis $v_1, \ldots, v_n$ for $V$ consisting of eigenvectors of $\alpha$. Suppose the distinct eigenvalues of $\alpha$ are $\lambda_1, \ldots, \lambda_r$. (As the eigenspaces of $\alpha$ may have dimension greater than one, $r$ may be strictly smaller than $n$.) Consider the polynomial $p(x) = (x - \lambda_1)(x - \lambda_2) \cdots (x - \lambda_r)$, which is certainly the product of distinct linear factors. We claim that $p(x)$ is in fact the minimal polynomial of $\alpha$.

Let $v_i$ be any of the basis vectors, and suppose its corresponding eigenvalue is $\lambda_j$. Then

$$(\alpha - \lambda_1 I)(\alpha - \lambda_2 I) \cdots (\alpha - \lambda_r I)v_i = (\lambda_j - \lambda_1)(\lambda_j - \lambda_2) \cdots (\lambda_j - \lambda_r)v_i = \mathbf{0},$$

since one of the factors in the product is 0. Since $p(\alpha)$ takes all of the basis vectors to to the zero vector, we must have $p(\alpha) = 0$. All the eigenvalues of $\alpha$ are roots of the minimal polynomial, so the minimal polynomial must have degree at least $r$. Since $p(x)$ has degree $r$, it is the minimal polynomial. This completes the "only if" part.

Now suppose that the minimal polynomial is a product of distinct linear factors, i.e.,

$$m_\alpha(x) = (x - \lambda_1)(x - \lambda_2)\cdots(x - \lambda_r),$$

where $\lambda_1, \ldots, \lambda_r$ are the distinct eigenvalues of $\alpha$. (Recall that the roots of the minimal polynomial are precisely the eigenvalues of $\alpha$.) Since $m_\alpha(x)$ is the minimal polynomial, we have $m_\alpha(\alpha) = 0$, in other words

$$(\alpha - \lambda_1 I)(\alpha - \lambda_2 I)\cdots(\alpha - \lambda_r I) = 0. \tag{5.3}$$

Recall that the nullity $\nu(\beta)$ of a linear map $\beta$ is the dimension of its kernal. Then

$$\nu(\alpha-\lambda_1 I)+\nu(\alpha-\lambda_2 I)+\cdots+\nu(\alpha-\lambda_r I) \geq \nu\big((\alpha-\lambda_1 I)(\alpha-\lambda_2 I)\cdots(\alpha-\lambda_r I)\big) = n, \tag{5.4}$$

where we use the inequality $\nu(\beta)+\nu(\gamma) \geq \nu(\beta\gamma)$ for the product of linear maps $\beta, \gamma$ (see question 5(a) of Assignment 6 from a previous year). The equality is from (5.3).

Observe that $\mathrm{Ker}(\alpha - \lambda_j I)$ is equal to $E(\lambda_j, \alpha)$, the eigenspace corresponding to eigenvalue $\lambda_j$. Let $W = E(\lambda_1, \alpha)+E(\lambda_2, \alpha)+\cdots+E(\lambda_r, \alpha)$. Suppose $u_1 \in E(\lambda_1, \alpha)$, $u_2 \in E(\lambda_2, \alpha)$, …, $u_r \in E(\lambda_r, \alpha)$ are vectors satisfying $u_1+u_2+\cdots+u_r = \mathbf{0}$. By Lemma 5.13, we must have $u_1 = u_2 = \cdots = u_r = \mathbf{0}$. Then we deduce from Lemma 1.27 that $W$ is actually a direct sum of the subspaces, i.e., $W = E(\lambda_1, \alpha) \oplus E(\lambda_2, \alpha) \oplus \cdots \oplus E(\lambda_r, \alpha)$. Finally,

$$n \leq \dim(E(\lambda_1, \alpha)) + \dim(E(\lambda_2, \alpha)) + \cdots + \dim(E(\lambda_r, \alpha)) = \dim(W) \leq \dim(V) = n.$$

where the first inequality is from (5.4) and the equality is from Lemma 1.28(b). Thus $W = V$ and $V$ is the direct sum of the eigenspaces of $\alpha$. It follows from Theorem 5.14 that $\alpha$ is diagonalisable. This completes the "if" part, and the proof. □

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 procedure 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.

**Example 5.21.** Consider first the matrix

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

The characteristic polynomial is

$$p_A(x) = \det(xI - A) = \begin{vmatrix} x & -1 & 0 \\ 0 & x & -1 \\ -1 & 0 & x \end{vmatrix} = x^3 - 1 = (x-1)(x^2 + x + 1).$$

The polynomial $p_A(x)$ does not factor further over $\mathbb{R}$, as two of the roots are complex. So it seems that, as a linear map on $\mathbb{R}^3$, the matrix $A$ does not diagonalise. Indeed the minimal polynomial is this case is either $(x-1)(x^2+x+1)$ or $(x-1)$. (It must contain $x-1$ as a factor, as 1 is a root of the characteristic polynomial.) Since $A - I \neq O$, the minimal polynomial is in fact $(x-1)(x^2+x+1)$, which is not the product of distinct *linear* factors.

[An aside. In fact, any irreducible factor of the characteristic polynomial must always be a factor of the minimal polynomial, so we didn't really need the case analysis. However, this fact does not follow easily from anything we have covered in the module.]

This problem can be fixed by extending the field to the complex numbers $\mathbb{C}$. Then the characteristic polynomial is a product of linear factors, namely, $p_A(x) = (x-1)(x-\omega)(x-\omega^2)$, where $\omega = -\frac{1}{2} + \frac{\sqrt{3}}{2}i$. (Note that 1, $\omega$ and $\omega^2$ are the cube roots of unity.) By Theorem 5.18, the minimal polynomial $m_A(x)$ divides $p_A(x)$ and hence $m_A(x)$ also is a product of distinct linear factors. Thus, viewed as a linear map on $\mathbb{C}^3$, the matrix $A$ is diagonalisable, and its diagonal form is

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

You can check that the eigenvectors are $\begin{bmatrix} 1 & 1 & 1 \end{bmatrix}^\top$, $\begin{bmatrix} 1 & \omega & \omega^2 \end{bmatrix}^\top$, and $\begin{bmatrix} 1 & \omega^2 & \omega \end{bmatrix}^\top$. So the matrix $P$ that diagonalises $A$, in the sense that $D = P^{-1}AP$, is

$$P = \begin{bmatrix} 1 & 1 & 1 \\ 1 & \omega & \omega^2 \\ 1 & \omega^2 & \omega \end{bmatrix};$$

its columns are just the eigenvectors of $A$ taken in order.

In the example just considered, the obstacle to diagonalisation is that the characteristic polynomial did not have a full set of roots over $\mathbb{R}$; this problem can be dealt with by extending the field to $\mathbb{C}$. The next example illustrates a deeper problem that can arise. Consider the matrix

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

Its characteristic polynomial is $p_B(x) = (x-2)^2(x-1)$. The minimal polynomial divides $p_B(x)$ and has the same roots, so the possibilities are either $m_B(x) = (x-2)(x-1)$ or $m_B(x) = (x-2)^2(x-1)$. Can it be the former? Evaluating $(B-2I)(B-I)$ we find that

$$(B-2I)(B-I) = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \neq O.$$

By a (short!) process of elimination we have found that $m_A(x) = p_A(x) = (x-2)^2(x-1)$. The minimal polynomial is not a product of *distinct* linear factors, so the matrix $B$ is not diagonalisable. This is a more fundamental problem, which cannot be solved by extending the field.

## 5.6 Jordan form

We briefly consider, without proof, a canonical form for matrices over the complex numbers that deals to some extent with the problem identified iin the previous exercise.

**Definition 5.22.** (a) A *Jordan block* $J(n, \lambda)$ is a matrix of the form

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

that is, it is an $n \times n$ matrix with $\lambda$ on the main diagonal, 1 in positions immediately above the main diagonal, and 0 elsewhere. (We take $J(1, \lambda)$ to be the $1 \times 1$ 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 5.23.** *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 5.24.** A matrix over $\mathbb{C}$ is diagonalisable if and only if all the Jordan blocks in its Jordan form have size 1.

**Example 5.25.** Any $3 \times 3$ 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).

Notice that the matrix $B$ from the previous example (the one that is not diagonalisable) is already in Jordan form.

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.

## 5.7 Trace

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

**Definition 5.26.** The *trace* $\mathrm{Tr}(A)$ of a square matrix $A$ is the sum of its diagonal entries.

**Proposition 5.27.** *(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.* Let $A = (a_{ij})$ and $B = (b_{ij})$. For part (a), note that

$$\mathrm{Tr}(AB) = \sum_{i=1}^{n}(AB)_{ii} = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}b_{ji} = \sum_{j=1}^{n}\sum_{i=1}^{n} b_{ji}a_{ij} = \sum_{j=1}^{n}(BA)_{jj} = \mathrm{Tr}(BA),$$

by the rules for matrix multiplication.

For part (b), we just observe that for any invertible matrix $P$,

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

□

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

The trace and determinant of $\alpha$ are coefficients in the characteristic polynomial of $\alpha$.

**Proposition 5.28.** *Let $\alpha : V \to V$ be a linear map, where $\dim(V) = n$, and let $p_\alpha$ be the characteristic polynomial of $\alpha$, 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 $\alpha$ is diagonalisable, then the sum of its eigenvalues (taking account of multiplicities) is $\mathrm{Tr}(\alpha)$ and their product is $\det(\alpha)$.*

*Proof.* Let $A = (a_{ij})$ be a matrix representing $\alpha$. We have

$$p_\alpha(x) = \det(xI - A) = \begin{vmatrix} x - a_{1,1} & -a_{1,2} & \dots & -a_{1,n} \\ -a_{2,1} & x - a_{2,2} & \dots & -a_{2,n} \\ & & \dots & \\ -a_{n,1} & -a_{n,2} & \dots & x - a_{n,n} \end{vmatrix}.$$

The only way to obtain a term in $x^{n-1}$ in the determinant is from the product $(x - a_{1,1})(x - a_{2,2}) \cdots (x - a_{n,n})$ of diagonal entries, taking $-a_{i,i}$ from the $i$th 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 $p_\alpha(0) = \det(-A) = (-1)^n \det(A)$.

If $\alpha$ is diagonalisable, choose a basis relative to which the matrix $A$ representing $\alpha$ is diagonal. The diagonal entries $\lambda_1, \dots, \lambda_n$ are the eigenvalues of $\alpha$ counted according to multiplicity. Then $\mathrm{Tr}(\alpha) = \mathrm{Tr}(A) = \lambda_1 + \cdots + \lambda_n$ and $\det(\alpha) = \det(A) = \lambda_1\lambda_2 \dots \lambda_n$. □

In part (b) of Proposition 5.28 we don't actually need that $\alpha$ is diagonslisable; it enough that the characteristic polynomial is a product of linear factors. Write

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

Note that 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. Now use part (a) of the proposition.

We conclude with the missing parts of the proof of Theorem 5.14. This does not form part of the module, and is included for interest only.

*Proof of the missing equivalences in Theorem 5.14.* (b) $\Rightarrow$ (c). Proposition 5.5 shows that there are projections $\pi_1, \ldots, \pi_r$ satisfying the conditions of (c), with $\mathrm{Im}(\pi_i) = E(\lambda_i, \alpha)$. We just need to check that $\alpha$ and $\lambda_1\pi_1 + \cdots + \lambda_r\pi_r$ are equal. Let $v \in V$ be arbitrary. Then,

$$\begin{aligned}\alpha(v) &= \alpha((\pi_1 + \cdots + \pi_r)(v)) \\ &= \alpha(\pi_1(v) + \cdots + \pi_r(v)) \\ &= \alpha(\pi_1(v)) + \cdots + \alpha(\pi_r(v)) \\ &= \lambda_1\pi_1(v) + \cdots + \lambda_r\pi_r(v) \\ &= (\lambda_1\pi_1 + \cdots + \lambda_r\pi_r)(v),\end{aligned}$$

where the penultimate equality comes from the fact that $\pi_i(v) \in \mathrm{Im}(\pi_i) = E(\lambda_i, \alpha)$, for $i = 1, \ldots, r$. So $\alpha = \lambda_1\pi_1 + \cdots + \lambda_r\pi_r$, as required.

(c) $\Rightarrow$ (a). Since the projections $\pi_i$ satisfy the conditions of Proposition 5.4, $V$ is the direct sum of the subspaces $\mathrm{Im}(\pi_i)$. We now observe that $\mathrm{Im}(\pi_i) \subseteq E(\lambda_i, \alpha)$. To see this, take any $u \in \mathrm{Im}(\pi_i)$ and consider $\alpha(u)$:

$$\alpha(u) = (\lambda_1\pi_1 + \cdots + \lambda_r\pi_r)(u) = \lambda_1\pi_1(u) + \cdots + \lambda_r\pi_r(u) = \lambda_i u,$$

where we have used the facts that $\pi_i(u) = u$ and $\pi_j(u) = \mathbf{0}$, for $j \neq i$. Thus $\mathrm{Im}(\pi_i) \subseteq E(\lambda_i, \alpha)$ and

$$V = \mathrm{Im}(\pi_1) + \cdots + \mathrm{Im}(\pi_r) \subseteq E(\lambda_1, \alpha) + \cdots + E(\lambda_r, \alpha) \subseteq V.$$

(The containments must of course be equality.) Thus we can choose a basis of $V$ consisting entirely of eigenvectors of $\alpha$. □

## Summary

- When we consider a linear map from a vector space $V$ to itself, there is only one basis involved, and this makes the situation more interesting than the one considered in the previous chapter.
- A projection is a linear map whose square is equal to itself (applying a projection twice gives the same result as applying it once).
- If $\pi$ is a projection on $V$ then $V$ is the direct sum of $\mathrm{Ker}(\pi)$ and $\mathrm{Im}(\pi)$.
- Suppose $V$ is the direct sum of subspaces $U_1, U_2, \ldots, U_r$. We can view the subspaces as images of projections on $V$ with nice properties. The converse is also true.
- Eigenvectors of a linear map $\alpha$ are special vectors: the effect of applying $\alpha$ to an eigenvector $v$ is to scale $v$ (by the associated eigenvalue). Eigenvectors sharing a common eigenvalue $\lambda$ live in a subspace of $V$: the eigenspace corresponding to $\lambda$.
- A linear map $\alpha$ on $V$ is diagonalisable if there is a basis for $V$ consisting entirely of eigenvectors of $\alpha$.
- Equivalently, there is a basis of $V$ relative to which the matrix representing $\alpha$ is diagonal.
- Equivalently, $V$ is the direct sum of eigenspaces of $\alpha$.
- The characteristic polynomial $p_\alpha$ of a linear map $\alpha$ is well defined (it is independent of the choice of basis). The minimal polynomial $m_\alpha$ divides $p_\alpha$ and is the smallest degree monic polynomial satisfying $m_\alpha(\alpha) = 0$.

- $\lambda$ is an eigenvalue of $\alpha$ iff it is a root of the characteristic polynomial of $\alpha$ iff it is a root of the minimal polynomial of $\alpha$.
- A linear map is diagonalisable iff its minimal polynomial factors into distinct linear factors.
- Matrices that are not diagonalisable can at least be transformed into Jordan (near-diagonal) form.
- The trace of a matrix is another scalar associated with a matrix that is invariant under similarity (and hence makes sense for linear maps).

# Chapter 6

# Linear and quadratic forms

**Cut out the dual space.**

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.

Before looking at quadratic forms, what is a linear form?

**Definition 6.1.** 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 $f = \begin{bmatrix} a_1 & a_2 & \dots & a_n \end{bmatrix}$, then for $v = \begin{bmatrix} x_1 & x_2 & \dots & x_n \end{bmatrix}^\top$ we have

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

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

**Definition 6.2.** 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 $V^*$.

**Cut out the material on the dual space, basis of dual space, adjoints. Note that this means we have to take more care when introducing adjoint operator later on.**

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.

## 6.1 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 6.3.** 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 $cx_i x_j$, 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 6.4.** *A basis change with transition matrix $P$ replaces the symmetric matrix $A$ representing a quadratic form by the matrix $P^\top AP$.*

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

**Definition 6.5.** Two symmetric matrices $A, A'$ over a field $\mathbb{K}$ are *congruent* if $A' = P^\top AP$ for some invertible matrix $P$.

**Proposition 6.6.** *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.

## 6.2 Reduction of quadratic forms

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

**Theorem 6.7.** *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*

$$q = c_1 y_1^2 + c_2 y_2^2 + \cdots + c_n y_n^2$$

*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

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

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

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

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

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

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 6.8.** 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 $P^\top A P = A'$?

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 $\alpha_i$ 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

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

we get

$$q = y_1^2 + \cdots + y_s^2 - y_{s+1}^2 - \cdots - y_{s+t}^2.$$

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*.]

## 6.3 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 6.9.** (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 $V$.

**Remark 6.10.** 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 $\alpha$ is represented by a matrix $A$ relative to the bases $B$ and $B^*$.

Summarising:

**Proposition 6.11.** *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 $\alpha$ 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. □

## 6.4 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 6.12.** *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 6.13.** *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 $s + t = \mathrm{rank}(A)$, 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 $s + t = s' + t'$. 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 7

# 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 some adjustments need to be made. To avoid overloading the module, we will concentrate on real inner product spaces, and mention the adjustments required for complex spaces only briefly.

## 7.1 Inner products and orthonormal bases

**Definition 7.1.** An *inner product* on a real vector space $V$ is a function that takes each pair of vectors $v, w \in V$ to a real number $v \cdot w$ satisfying the following conditions:

- The inner product is *symmetric*, that is, $v \cdot w = w \cdot v$ for all $v, w \in V$.
- The inner product is *bilinear*, that is, linear in the first variable when the second is kept constant and *vice versa*. (Symbolically, $(v + v') \cdot w = v \cdot w + v' \cdot w$ and $(av) \cdot w = a(v \cdot w)$ for all $v, v', w \in V$ and $a \in \mathbb{R}$.)
- The inner product is *positive definite*, that is, $v \cdot v \geq 0$ for all $v \in V$, and $v \cdot v = 0$ if and only if $v = \mathbf{0}$.

An inner product is sometimes called a *dot product* because of this notation.

Note that we don't need to insist that $v \cdot (w + w') = v \cdot w + v \cdot w'$ and $v \cdot (aw) = a(v \cdot w)$, for all $v, w, w' \in V$ and $a \in \mathbb{R}$, since these facts follow by symmetry from linearity in the first variable.

Geometrically, in a real vector space, we might define $v \cdot w = |v|\,|w| \cos\theta$, where $|v|$ and $|w|$ are the lengths of $v$ and $w$, and $\theta$ is the angle between $v$ and $w$. But we can easily reverse the order of doing things and define lengths and angles in terms of the inner product. Given an inner product on $V$, we define the *length* of any vector $v \in V$ to be

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

and, for any vectors $v, w \in V \setminus \{\mathbf{0}\}$, we define the *angle* between $v$ and $w$ to be $\theta$, 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 7.2.** *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 $(v + aw) \cdot (v + aw) \geq 0$ for any real number $a$. Expanding, we obtain

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

This is a quadratic function in $a$. Since it is non-negative for all real $a$, 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. □

Two non-zero vectors $u, v$ are said to be *orthogonal* if $u \cdot v = 0$. Certain bases in an inner product space are particularly convenient to use.

**Definition 7.3.** A basis $(v_1, \ldots, v_n)$ for an inner product space is called *orthonormal* if $v_i \cdot v_j = \delta_{i,j}$ (the Kronecker delta) for $1 \leq i, j \leq n$. Thus, the basis vectors have unit length and are pairwise orthogonal.

**Lemma 7.4.** *If vectors $v_1, \ldots, v_n$ satisfy $v_i \cdot v_j = \delta_{i,j}$, then they are necessarily linearly independent.*

*Proof.* Suppose that $c_1v_1 + \cdots + c_nv_n = \mathbf{0}$ for some scalars $c_1, \ldots, c_n \in \mathbb{K}$. Taking the inner product of this equation with $v_i$, we have on the left hand side

$$v_i \cdot (c_1v_1 + \cdots + c_nv_n) = c_1v_i \cdot v_1 + \cdots + c_iv_i \cdot v_i + \cdots c_nv_i \cdot v_n = c_i,$$

and on the left hand side $v_i \cdot \mathbf{0} = 0$. Thus $c_i = 0$, and this holds for all $i$. □

**Theorem 7.5.** *Let $\cdot$ be an inner product on a real vector space $V$. Then there is an orthonormal basis $\mathcal{B} = (v_1, \ldots, v_n)$ for $V$.*

The proof involves a constructive method for finding an orthonormal basis, known as the *Gram-Schmidt process*. The Gram-Schmidt process was covered in *Linear Algebra I*, so we won't repeat it here. But we should remind ourselves exactly what this algorithm accomplishes, as it will be important for us later. The input to the algorithm is an arbitrary basis $w_1, \ldots, w_n$ for a vector space $V$. The output is an orthonormal basis $v_1, \ldots, v_n$ for $V$ that satisfies $\langle w_1, \ldots, w_i \rangle = \langle v_1, \ldots, v_i \rangle$ for all $1 \leq i \leq n$.

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

**Proposition 7.6.** *Suppose $\mathcal{B}$ is an orthonormal basis for the inner product space $V$ of dimension $n$. If we represent vectors in coordinates with respect to $\mathcal{B}$, say $[v]_\mathcal{B} = \begin{bmatrix} a_1 & a_2 & \ldots & a_n \end{bmatrix}^\top$ and $[w]_\mathcal{B} = \begin{bmatrix} b_1 & b_2 & \ldots & b_n \end{bmatrix}^\top$, then*

$$v \cdot w = a_1b_1 + a_2b_2 + \cdots + a_nb_n.$$

*Proof.* Let $\mathcal{B} = (v_1, \ldots, v_n)$. If $v = a_1v_1 + \cdots + a_nv_n$ and $w = b_1v_1 + \cdots + b_nv_n$, then

$$v \cdot w = (a_1v_1 + \cdots + a_nv_n) \cdot (b_1v_1 + \cdots + b_nv_n) = a_1b_1 + \cdots + a_nb_n,$$

since all the cross terms are zero. □

**Definition 7.7.** The inner product on $\mathbb{R}^n$ for which the standard basis is orthonormal (that is, the one given in the above proposition) is called the *standard inner product* on $\mathbb{R}^n$.

**Remark 7.8.** If we reflect for a moment, we see that some changes must be made to deal with inner product spaces over $\mathbb{C}$. If $V$ is a vector space over $\mathbb{C}$ and $v \in V$ any non-zero vector, then bilinearity of the inner product would imply $(iv) \cdot (iv) = i^2(v \cdot v) = -(v \cdot v)$. But then it cannot be the case that both $(iv) \cdot (iv)$ and $v \cdot v$ are positive real numbers, violating the requirement that the inner product should be positive definite. The fix is demand that the inner product satisfies *conjugate symmetry*, i.e., $v \cdot w = \overline{w \cdot v}$ rather than symmetry. (Overline here denotes complex conjugation.) A knock-on effect is that if we demand $(av) \cdot w = a(v \cdot w)$ then, necessarily, we must have $v \cdot (aw) = \bar{a}(v \cdot w)$. This follows from the chain of equalities

$$v \cdot (aw) = \overline{(aw) \cdot v} = \overline{a(w \cdot v)} = \bar{a}(\overline{w \cdot v}) = \bar{a}(v \cdot w).$$

We describe this situation by saying that the inner product is "sesquilinear". Note that these changes solve the problem we identified earlier, since now $(iv) \cdot (iv) = i(-i)v \cdot v = v \cdot v$. Note that the inner product of a vector with itself is certainly real, since $v \cdot v = \overline{v \cdot v}$.

## 7.2 Adjoints and orthogonal linear maps

**Definition 7.9.** Let $V$ be an inner product space, and $\alpha : V \to V$ a linear map. Then the *adjoint* of $\alpha$ is the linear map $\alpha^* : V \to V$ defined by

$$v \cdot \alpha^*(w) = \alpha(v) \cdot w, \quad \text{for all } v \in V. \tag{7.1}$$

Given $w$, why should there exist a vector $\alpha^*(w)$ satisfying (7.1), and why should it be unique? If we fix $w$ then the right hand side of (7.1) is a *linear functional* of $v$, that is, a function $V \to \mathbb{R}$ that is linear in its argument. There is a result, the *Riesz Representation Theorem*, that states that every linear functional in $v$ has a unique representation as $v \cdot u$ for some $u \in V$. Thus $\alpha^*(w)$ exists and is unique. The Riesz Representation Theorem is beyond the scope of the course (though it is not so hard to prove).

Having seen that $\alpha^*$ is well defined, we need to show that $\alpha^*$ is linear. This is a short exercise (on the final previous year assignment), where you are also invited to show also that $\alpha^{**} = (\alpha^*)^*$ satisfies $\alpha^{**} = \alpha$.

**Proposition 7.10.** *If $\alpha$ is represented by the matrix $A$ relative to an orthonormal basis of $V$, then $\alpha^*$ is represented by the transposed matrix $A^\top$.*

*Proof.* Denote the orthonormal basis by $\mathcal{B}$, and let the coordinate representations of vectors $v$ and $w$ in the basis $\mathcal{B}$ be $[v]_\mathcal{B} = \begin{bmatrix} b_1, \ldots, b_n \end{bmatrix}^\top$ and $[w]_\mathcal{B} = \begin{bmatrix} c_1, \ldots, c_n \end{bmatrix}^\top$; also let $A = (a_{i,j})$ and $A^* = (a^*_{i,j})$ be the representations of $\alpha$ and $\alpha^*$ in the basis $\mathcal{B}$. Then (7.1) expressed in the basis $\mathcal{B}$ becomes $[v]_\mathcal{B} \cdot (A^* [w]_\mathcal{B}) = (A [v]_\mathcal{B}) \cdot [w]_\mathcal{B}$, i.e.,

$$\sum_{i=1}^{n} b_i \sum_{j=1}^{n} a^*_{i,j} c_j = \sum_{j=1}^{n} \left( \sum_{i=1}^{n} a_{j,i} b_i \right) c_j,$$

i.e.,

$$\sum_{i=1}^{n}\sum_{j=1}^{n} b_i c_j a^*_{i,j} = \sum_{i=1}^{n}\sum_{j=1}^{n} b_i c_j a_{j,i}, \tag{7.2}$$

Since (7.1) holds for all $v$ and $w$, and hence (7.2) holds for all $[v]_{\mathcal{B}}$ and $[w]_{\mathcal{B}}$, we can set $b_i = 1$ and $b_k = 0$ for all $k \neq i$, and also $c_j = 1$ and $c_k = 0$ for all $k \neq j$, to deduce from the above equation that $a^*_{i,j} = a_{j,i}$. Thus $A^*$ is the transpose of $A$. □

**Example 7.11.** The matrix

$$A = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} & 0 \\ -1/\sqrt{2} & 1/\sqrt{2} & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

represents a clockwise (looking down on the $x, y$-plane) rotation by $\pi/4$ about the $z$-axis in $\mathbb{R}^3$. Its adjoint is

$$A^* = A^\top = \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} & 0 \\ 1/\sqrt{2} & 1/\sqrt{2} & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

and represents a rotation by $\pi/4$ anticlockwise.

Now we define two important classes of linear maps on $V$.

**Definition 7.12.** Let $\alpha$ be a linear map on an inner product space $V$.

(a) $\alpha$ is *self-adjoint* if $\alpha^* = \alpha$.

(b) $\alpha$ is *orthogonal* if it is invertible and $\alpha^* = \alpha^{-1}$.

There are several ways to look at orthogonal maps.

**Theorem 7.13.** *The following are equivalent for a linear map $\alpha$ on an inner product space $V$:*

*(a) $\alpha$ is orthogonal;*

*(b) $\alpha$ preserves the inner product, that is, $\alpha(v) \cdot \alpha(w) = v \cdot w$;*

*(c) $\alpha$ maps any orthonormal basis of $V$ to an orthonormal basis.*

*Proof.* (a) ⇒ (b). Suppose $\alpha$ is orthogonal, so that $\alpha^*\alpha$ is the identity map. By the definition of adjoint,

$$\alpha(v) \cdot \alpha(w) = v \cdot \alpha^*(\alpha(w)) = v \cdot w.$$

(b) ⇒ (c). Suppose that $(v_1, \ldots, v_n)$ is an orthonormal basis, that is, $v_i \cdot v_j = \delta_{i,j}$ for all $i, j$. If (b) holds, then $\alpha(v_i) \cdot \alpha(v_j) = v_i \cdot v_j = \delta_{i,j}$, so that $(\alpha(v_1), \ldots, \alpha(v_n))$ is an orthonormal basis, and (c) holds.

(c) ⇒ (a). Suppose that $\alpha$ maps orthonormal basis $(v_1, \ldots, v_n)$ to some other orthonormal basis $(\alpha(v_1), \ldots, \alpha(v_n))$. We want to show that $\alpha^*\alpha$ is the identity map. Apply $\alpha^*\alpha$ to basis vector $v_i$ and write the result in terms of the basis vectors:

$$\alpha^*\alpha(v_i) = c_1 v_1 + c_2 v_2 + \cdots + c_n v_n.$$

Now take the inner product of both sides with a basis vector $v_j$:

$$v_j \cdot \alpha^*\alpha(v_i) = v_j \cdot (c_1v_1 + c_2v_2 + \cdots + c_nv_n) = c_j.$$

On the other hand,

$$v_j \cdot \alpha^*\alpha(v_i) = \alpha(v_j) \cdot \alpha(v_i) = \delta_{ij},$$

since $\alpha(v_1), \ldots, \alpha(v_n)$ is an orthonormal basis. Thus, $c_i = 1$, and $c_j = 0$ if $j \neq i$. In other words, $\alpha^*\alpha(v_i) = v_i$. Since $\alpha^*\alpha$ maps every basis vector to itself, it must be the identity. □

What do self-adjoint and orthogonal linear maps look like from the matrix perspective? The following is immediate from Proposition 7.10.

**Corollary 7.14.** *If $\alpha$ is represented by a matrix $A$ (relative to an orthonormal basis), then*

*(a) $\alpha$ is self-adjoint if and only if $A$ is symmetric;*

*(b) $\alpha$ is orthogonal if and only if $A^\top A = I$.*

**Example 7.15.** Returning to Example 7.11, the matrix $A^\top$ there is the inverse of the matrix $A$: the former is an anticlockwise rotation that undoes the clockwise rotation of the latter. Thus the matrix $A$ represents an orthogonal linear map. Note that $A$ preserves lengths and angles (see Theorem 7.13(b)) and maps orthonormal bases to orthonormal bases (see Theorem 7.13(c)).

A convenient characterisation of matrices representing orthogonal maps is the following.

**Corollary 7.16.** *Suppose the linear map $\alpha : V \to V$ is represented by the matrix $A$ with respect to some orthonormal basis $\mathcal{B}$. Then $\alpha$ is orthogonal if and only if the columns of $A$ (viewed as coordinate representations of vectors relative to basis $\mathcal{B}$) form an orthonormal basis for $V$.*

*Proof.* Let $A$ be the representation of linear map $\alpha$ with respect to some orthonormal basis. Let the columns of $A$ be $\bar{v}_1, \ldots, \bar{v}_n$, so that

$$A = \begin{bmatrix} \bar{v}_1 & \bar{v}_2 & \ldots & \bar{v}_n \end{bmatrix} \quad \text{and} \quad A^\top = \begin{bmatrix} \bar{v}_1^\top \\ \bar{v}_2^\top \\ \vdots \\ \bar{v}_n^\top \end{bmatrix}.$$

We view the column vector $\bar{v}_j$ as the coordinate representation of vector $v_j$ in $V$; in symbols, $\bar{v}_j = [v_j]_{\mathcal{B}}$. Then

$$A^\top A = \begin{bmatrix} \bar{v}_1^\top\bar{v}_1 & \ldots & \bar{v}_1^\top\bar{v}_n \\ \vdots & & \vdots \\ \bar{v}_n^\top\bar{v}_1 & \ldots & \bar{v}_n^\top\bar{v}_n \end{bmatrix} = \begin{bmatrix} v_1 \cdot v_1 & \ldots & v_1 \cdot v_n \\ \vdots & & \vdots \\ v_n \cdot v_1 & \ldots & v_n \cdot v_n \end{bmatrix},$$

where the second equality uses Proposition 7.6. It is clear that $A^\top A = I$ if and only if $v_i \cdot v_j = \delta_{i,j}$, for $1 \leq i, j \leq n$, i.e., if and only if the vectors $v_1, \ldots, v_n$ are orthonormal. □

**Example 7.17.** Returning again to Example 7.11, it can be checked that the columns of $A$ (and hence the rows of $A$) are an orthonormal basis (viewed as coordinate representations relative to the standard basis). Specifically, denoting the columns of $A$ by $\bar{v}_1, \bar{v}_2, \bar{v}_3$, we have $\bar{v}_1^\top \bar{v}_1 = \bar{v}_2^\top \bar{v}_2 = \bar{v}_3^\top \bar{v}_3 = 1$ and $\bar{v}_1^\top \bar{v}_2 = \bar{v}_1^\top \bar{v}_3 = \bar{v}_2^\top \bar{v}_3 = 0$.

The above definitions suggest an equivalence relation on real matrices:

**Definition 7.18.** Two real $n \times n$ matrices $A$ and $A'$ are called *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. Note that orthogonal similarity is a refinement of similarity. From Theorem 7.13 we know that orthogonal maps take orthonormal bases to orthonormal bases; in other words, orthogonal matrices are transition matrices between orthonormal bases. Thus, two real matrices $A$ and $A'$ are orthogonally similar if they represent the same linear map with respect to different orthonormal bases.

It is natural to ask when it is the case that a matrix is orthogonally similar to a diagonal matrix, and this is the question we turn to in the final chapter.

## Summary

- An inner product space is a vector space $V$ equipped with an inner product $\cdot$ assigning a scalar from $\mathbb{K}$ to every pair of vectors. The inner product for real vector spaces is symmetric, bilinear and positive definite.
- From the inner product, we can define lengths and angles.
- Two vectors are *orthogonal* if their inner product is 0. An *orthonormal basis* of a vector space $V$ is one in which the basis vectors have unit length and are pairwise orthogonal.
- An orthonormal basis for $V$ always exists and can be found using Gram-Schmidt process.
- Relative to a orthonormal basis, an inner product looks like the scalar product from Linear Algebra I.
- To each linear map $\alpha : V \to V$ there corresponds an adjoint map $\alpha^* : V \to V$.
- The linear map $\alpha$ is *self-adjoint* if $\alpha^* = \alpha$, and *orthogonal* if $\alpha^* = \alpha^{-1}$.
- If $\alpha$ is represented by matrix $A$ relative to an orthonormal basis, the adjoint $\alpha^*$ is represented by $A^\top$, the transpose of $A$. Thus, the linear map $\alpha$ is self adjoint if $A$ is symmetric and orthogonal if $AA^\top = A^\top A = I$.
- An orthogonal linear map preserves the inner product and maps any orthonormal basis to an orthonormal basis.
- Suppose $\alpha$ is represented by matrix $A$ relative to an orthonormal basis. Then $\alpha$ is orthogonal iff the columns of $A$ for an orthonormal basis.
- Two square matrices $A$ and $A'$ are defined to be orthogonally similar iff there exists an orthogonal matrix $P$ such that $A' = P^\top AP$.

# Chapter 8

# The Spectral Theorem

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!

## 8.1 Orthogonal projections and orthogonal decompositions

**Definition 8.1.** We say that two vectors $u, w$ in an inner product space $V$ are *orthogonal* if $u \cdot w = 0$. We say that two subspaces $U$ and $W$ of $V$ are orthogonal if $u \cdot w = 0$ for all $u \in U$ and $w \in W$.

**Definition 8.2.** 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 that are orthogonal to everything in $U$:

$$U^\perp = \{w \in V : w \cdot u = 0 \text{ for all } u \in U\}.$$

Thus, the orthogonal complement of $U$ is the largest subspace of $V$ that is orthogonal to $U$.

**Proposition 8.3.** *If $V$ is a real 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 $(u_1, u_2, \ldots, u_r)$ for $U$ and extend it to a basis $(u_1, u_2, \ldots, u_n)$ for $V$. Then apply the Gram-Schmidt process to this basis (processing the vectors in the order $u_1, u_2, \ldots, u_n$), to obtain an orthonormal basis $(v_1, \ldots, v_n)$ of $V$. As we noted in the previous section, the Gram-Schmidt process has the property that $\langle v_1, \ldots, v_i \rangle = \langle u_1, \ldots, u_i \rangle$ for all $1 \leq i \leq n$. In particular, the the first $r$ vectors $v_1, \ldots, v_r$ in the resulting basis form an orthonormal basis for $U$. The last $n - r$ vectors are orthogonal to $U$, and so lie in $U^\perp$. Summarising, we have $v_1, \ldots, v_r \in U$ and $v_{r+1}, \ldots, v_n \in U^\perp$. Since $v_1, \ldots, v_n$ is a basis for $V$, it follows that every vector in $V$ can be written as the sum of a vector in $U$ and a vector in $U^\perp$ or, equivalently, $V = U + U^\perp$.

To show that $V$ is actually a *direct* sum of $U$ and $U^\perp$ we just need to show that $U \cap U^\perp = \{\mathbf{0}\}$. But if $u \in U$ and $u \in U^\perp$ then $u \cdot u = 0$ which implies $u = \mathbf{0}$.

The claim about the dimension of subspaces follows from Lemma 1.28. □

Recall the connection between direct sum decompositions and projections. If we have projections $\pi_1, \ldots, \pi_r$ whose sum is the identity and which satisfy $\pi_i \pi_j = 0$ 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 8.4.** Let $V$ be an inner product space, A linear map $\pi : V \to V$ is an *orthogonal projection* if

(a) $\pi$ is a projection, that is, $\pi^2 = \pi$, and

(b) $\pi$ is self-adjoint, that is, $\pi^* = \pi$.

**Definition 8.5.** Suppose $V$ is an inner product space, and $U_1, \ldots, U_r$ are subspaces of $V$. A direct sum $V = U_1 \oplus \cdots \oplus U_r$ is an *orthogonal decomposition of* $V$ if $U_i$ is orthogonal to $U_j$ for all $i \neq j$.

**Proposition 8.6.** *Suppose $\pi_1, \pi_2, \ldots, \pi_r$ are orthogonal projections on an inner product space $V$, satisfying*

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

*(b) $\pi_i \pi_j = 0$, for $i \neq j$.*

*Let $U_i = \mathrm{Im}(\pi_i)$, for $i = 1, \ldots, r$. Then $V = U_1 \oplus U_2 \oplus \cdots \oplus U_r$ is an orthogonal decomposition of $V$.*

*Proof.* The fact that $V$ is the direct sum of the images of the $\pi_i$ follows from Proposition 5.4. We only have to prove that $U_i$ and $U_j$ are orthogonal for all $i \neq j$. Recall that if $\pi$ is a projection, then $v \in \mathrm{Im}(\pi)$ if and only if $\pi(v) = v$. So take $u_i \in U_i$ and $u_j \in U_j$ with $i \neq j$. Then $\pi_i(u_i) = u_i$ and $\pi_j(u_j) = u_j$ and hence

$$u_i \cdot u_j = \pi_i(u_i) \cdot \pi_j(u_j) = u_i \cdot \pi_i^*(\pi_j(u_j)) = u_i \cdot \pi_i(\pi_j(u_j)) = 0,$$

where the second equality is the definition of the adjoint, and the third holds because $\pi_i$ is self-adjoint. □

As with Proposition 5.4, there is a converse.

**Proposition 8.7.** *Suppose $V = U_1 \oplus \cdots \oplus U_r$ is an orthogonal decomposition of an inner product space $V$. Then there exist orthogonal projections $\pi_1, \pi_2, \ldots, \pi_r$ on $V$ satisfying*

*(a) $\pi_1 + \pi_2 + \cdots + \pi_r = I$,*

*(b) $\pi_i \pi_j = 0$, for $i \neq j$, and*

*(c) $\mathrm{Im}(\pi_i) = U_i$, for all $i$.*

*Proof (sketch).* From Proposition 5.5 we know that there are projections $\pi_i$, for $1 \leq i \leq r$, satisfying conditions (a)–(c). Only one extra thing needs to be checked, namely that these projections are orthogonal, i.e., that the $\pi_i$ are self adjoint. □

## 8.2 The Spectral Theorem

The main theorem can be stated in different ways. We list three alternatives here.

**Theorem 8.8.** *If $\alpha$ is a self-adjoint linear map on a real inner product space $V$, then there is an orthonormal basis of $V$ consisting of eigenvectors of $\alpha$. Thus, the eigenspaces of $\alpha$ form an orthogonal decomposition of $V$.*

Equivalently, we can state the result as follows.

**Corollary 8.9.** *Suppose $\alpha$ and $V$ are as in the previous theorem, and $\lambda_1, \ldots, \lambda_r$ are the distinct eigenvalues of $\alpha$. Then there exist orthogonal projections $\pi_1, \ldots, \pi_r$ satisfying*

*(a)* $\pi_1 + \cdots + \pi_r = I$,

*(b)* $\pi_i \pi_j = 0$, *whenever* $i \neq j$, *and*

*(c)* $\alpha = \lambda_1 \pi_1 + \cdots + \lambda_r \pi_r$.

*Proof of Corollary 8.9.* By Theorem 8.8 we known that $V$ has an orthogonal decomposition $V = E(\lambda_1, \alpha) \oplus \cdots \oplus E(\lambda_r, \alpha)$, where $E(\lambda_i, \alpha)$ is the eigenspace correponding to the eigenvector $\lambda_i$. Then, by Proposition 8.7, there exist orthogonal projections, satisfying (a) and (b), such that $\mathrm{Im}(\pi_i) = E(\lambda_i, \alpha)$ for $1 \leq i \leq r$. Condition (c) then follows from the following chain of equalities:

$$\alpha(v) = \alpha(\pi_1(v) + \cdots + \pi_r(v)) = \lambda_1 \pi_1(v) + \cdots + \lambda_r \pi_r(v) = (\lambda_1 \pi_1 + \cdots + \lambda_r \pi_r)(v).$$

□

Yet another statement of the spectral theorem is in terms of matrices. Since a symmetric matrix represents a self-adjoint linear map with respect to some orthonormal basis, e.g., the standard basis of $\mathbb{R}^n$:

**Corollary 8.10.** *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.*

In tackling the proof of Theorem 8.8, we need (briefly) to dip into complex inner product spaces. Suppose $V$ is a vector space of dimension $n$ over $\mathbb{R}$. We can extend $V$ to a vector space $V^{\mathbb{C}}$ over $\mathbb{C}$ as follows. We set $V^{\mathbb{C}} = \{v' + iv'' : v', v'' \in V\}$ and define vector addition and scalar multiplication in the natural way. Thus, if $v = v' + iv''$ and $w = w' + iw''$ then

$$v + w = (v' + iv'') + (w' + iw'') = (v' + w') + i(v'' + w'');$$

and if $a = a' + ia'' \in \mathbb{C}$ then

$$av = (a' + ia'')(v' + iv'') = (a'v' - a''v'') + i(a'v'' + a''v').$$

Similarly, there is a natural (indeed unique) way to extend the inner product from $V$ to $V^{\mathbb{C}}$:

$$\begin{aligned}(v' + iv'') \cdot (w' + iw'') &= v' \cdot w' + v' \cdot (iw'') + (iv'') \cdot w' + (iv'') \cdot (iw'') \\ &= v' \cdot w' - i(v' \cdot w'') + i(v'' \cdot w') + v'' \cdot w''.\end{aligned}$$

(The minus sign in front of one of the terms arises from the fact that an inner product on a complex inner product space must be sesquilinear: $(av)\cdot w = a(v\cdot w)$ but $v\cdot(aw) = \bar{a}(v\cdot w)$, where $\bar{a}$ is the complex conjugate of $a$!)

Suppose $\alpha : V \to V$ is a linear map. The map $\alpha$ extends naturally (indeed uniquely) to a linear map $\alpha : V^{\mathbb{C}} \to V^{\mathbb{C}}$ defined by $\alpha(v) = \alpha(v') + i\alpha(v'')$. (We are slightly abusing notation here, by using $\alpha$ to denote both the linear map on $V$ and its extension to $V^{\mathbb{C}}$.) If $\alpha$ is self-adjoint as a linear map on $V$ then it is self-adjoint as a linear map on $V^{\mathbb{C}}$. This follows from the identity $(v' + iv'') \cdot \alpha(w' + iw'') = \alpha(v' + iv'') \cdot (w' + iw'')$ which may be verified by expanding both sides. Take care when doing this, to ensure that the signs are all correct!

*Proof of Theorem 8.8.* 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 $\alpha$ has an eigenvector.

As mentioned earlier, we may extend $\alpha$ to a linear map $\alpha : V^{\mathbb{C}} \to V^{\mathbb{C}}$ on a complex vector space. The characteristic polynomial of $\alpha$ viewed as a linear map on $V^{\mathbb{C}}$ has a root $\lambda$ over the complex numbers. (The so-called "Fundamental Theorem of Algebra" asserts that any polynomial over $\mathbb{C}$ has a root.) Let $v \in \mathbb{C}^n$ be an eigenvector corresponding to the eigenvalue $\lambda$. Since $\alpha$ is self-adjoint,

$$\lambda(v \cdot v) = (\lambda v) \cdot v = \alpha(v) \cdot v = v \cdot \alpha(v) = v \cdot (\lambda v) = \bar{\lambda}(v \cdot v).$$

(The complex conjugation of $\lambda$ arises because $\cdot$ is sesquilinear.) Since $v \cdot v \neq 0$ we see that $\lambda = \bar{\lambda}$ and hence that $\lambda$ is real.

Now since $\alpha$ has a real eigenvalue $\lambda$, we may choose a real eigenvector $v$ corresponding to $\lambda$, for example, by taking the real part of any complex eigenvector. By multiplying by a scalar if necessary we can assume that $|v| = 1$. We are now back in the real world, and will remain there for the rest of the proof. Let $U$ be the subspace $U = \{u \in V : v\cdot u = 0\}$. This is a subspace of $V$ of dimension $n-1$, by Proposition 8.3. We claim that $\alpha : U \to U$. To see this, take any $u \in U$. Then

$$\alpha(u) \cdot v = u \cdot \alpha^*(v) = u \cdot \alpha(v) = \lambda(u \cdot v) = 0,$$

where we use the fact that $\alpha$ is self-adjoint. Hence $\alpha(u) \in U$.

So $\alpha$ restricted to $U$ is a self-adjoint linear map on an $(n-1)$-dimensional inner product space. By the inductive hypothesis, $U$ has an orthonormal basis consisting of eigenvectors of $\alpha$. They are all orthogonal to the unit vector $v$; so, adding $v$ to the basis, we get an orthonormal basis for $V$, as required.

The fact that $V$ is a direct sum of eigenspaces comes from Theorem 5.14, so for the final part of the theorem we just need to show that these eigenspaces are orthogonal. We could use the orthonormal basis just constructed to prove this but it is easier to go directly. Suppose $v \in E(\lambda, \alpha)$ and $w \in E(\mu, \alpha)$ are vectors in distinct eigenspaces. Then $\alpha(v) = \lambda v$ and $\alpha(w) = \mu w$, and

$$\lambda(v \cdot w) = \lambda v \cdot w = \alpha(v) \cdot w = v \cdot \alpha^*(w) = v \cdot \alpha(w) = v \cdot \mu w = \mu(v \cdot w),$$

so, since $\lambda \neq \mu$, we see that $v \cdot w = 0$. □

**Remark 8.11.** 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.

**Example 8.12.** Let

$$A = \begin{bmatrix} 10 & 2 & 2 \\ 2 & 13 & 4 \\ 2 & 4 & 13 \end{bmatrix}.$$

The characteristic polynomial of $A$ is

$$p_A(x) = \begin{vmatrix} x-10 & -2 & -2 \\ -2 & x-13 & -4 \\ -2 & -4 & x-13 \end{vmatrix} = (x-9)^2(x-18),$$

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 $\begin{bmatrix} 1 & 2 & 2 \end{bmatrix}^\top$. Normalising, we can choose a unit eigenvector $\begin{bmatrix} \frac{1}{3} & \frac{2}{3} & \frac{2}{3} \end{bmatrix}^\top$.

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 $\begin{bmatrix} 0 & 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix}^\top$ and $\begin{bmatrix} -4/3\sqrt{2} & 1/3\sqrt{2} & 1/3\sqrt{2} \end{bmatrix}^\top$. 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

$$P^\top A P = \begin{bmatrix} 18 & 0 & 0 \\ 0 & 9 & 0 \\ 0 & 0 & 9 \end{bmatrix}.$$

You might like to check that the orthogonal matrix in the example in the last chapter of the notes also diagonalises $A$.

## Summary

- Two subspaces $U$ and $W$ are said to be orthogonal if every vector in $U$ is orthogonal to every vector in $W$.
- The orthogonal complement $U^\perp$ of a subspace $U$ of a vector space $V$ is the set of vectors in $V$ that are orthogonal to all vectors in $U$. It is a subspace of $V$.

- If $U$ is a subspace of $V$ then $V$ is the direct sum of $U$ and $U^{\perp}$.
- A projection is said to be orthogonal if it is self-adjoint.
- A direct sum of subspaces forms an orthogonal decomposition if the subspaces are orthogonal to each other.
- There is a correspondence between orthogonal decompositions and collections of orthogonal projections satisfying certain conditions.
- If $\alpha$ is a self-adjoint linear map on a real vector space $V$ then there there is an orthonormal basis of $V$ composed of eigenvectors of $\alpha$. (The Spectral Theorem.)
- Equivalently, a real symmetric matrix $A$ is orthogonally similar to a diagonal matrix: $P^{-1}AP = D$, where $P$ is an orthogonal matrix and $D$ is diagonal.