16
Appendix A: Linear Algebra

This appendix gives a summary of the results we need from linear algebra. Recommended for further reading are Blyth and Robertson's books *Basic Linear Algebra* [4] and *Further Linear Algebra* [5] and Halmos *Finite-Dimensional Vector Spaces* [11].

We expect that the reader will already know the definition of vector spaces and will have seen some examples. For most of this book, we deal with finite-dimensional vector spaces over the complex numbers, so the main example to bear in mind is $\mathbf{C}^n$, which we think of as a set of column vectors.

We assume that the reader knows about bases, subspaces, and direct sums. We therefore begin our account by describing quotient spaces. Next we discuss the connection between linear maps and matrices, diagonalisation of matrices, and Jordan canonical form. We conclude by reviewing the bilinear algebra needed in the main text.

16.1 Quotient Spaces

Suppose that W is a subspace of the vector space V. A *coset of* W is a set of the form

$$v + W := \{v + w : w \in W\}.$$

It is important to realise that unless $W = 0$, each coset will have many different labels; in fact, $v + W = v' + W$ if and only if $v - v' \in W$.

K. Erdmann et al., *Introduction to Lie Algebras*

The *quotient space* V/W is the set of all cosets of W. This becomes a vector space, with zero element $0 + W = W$, if addition is defined by

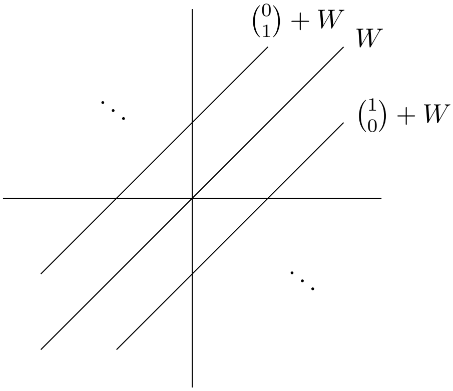
$$(v + W) + (v' + W) := (v + v') + W \quad \text{for } v, v' \in V$$

and scalar multiplication by

$$\lambda(v + W) := \lambda v + W \quad \text{for } v, v' \in V,\ \lambda \in F.$$

One must check that these operations are *well-defined*; that is, they do not depend on the choice of labelling elements. Suppose for instance that $v + W = v' + W$. Then, since $v - v' \in W$, we have $\lambda v - \lambda v' \in W$ for any scalar λ, so $\lambda v + W = \lambda v' + W$.

The following diagram shows the elements of $\mathbf{R}^2/W$, where W is the subspace of $\mathbf{R}^2$ spanned by $\binom{1}{1}$.



The cosets $\mathbf{R}^2/W$ are all the translations of the line W. One can choose a standard set of coset representatives by picking any line through 0 (other than W) and looking at its intersection points with the cosets of W; this gives a geometric interpretation of the isomorphism $\mathbf{R}^2/W \cong \mathbf{R}$.

It is often useful to consider quotient spaces when attempting a proof by induction on the dimension of a vector space. In this context, it can be useful to know that if $v_1, \ldots, v_k$ are vectors in V such that the cosets $v_1 + W, \ldots, v_k + W$ form a basis for the quotient space V/W, then $v_1, \ldots, v_k$, together with any basis for W, forms a basis for V.

16.2 Linear Maps

Let V and W be vector spaces over a field F. A *linear map* (or *linear transformation*) $x : V \to W$ is a map satisfying

$$x(\lambda u + \mu v) = \lambda x(u) + \mu x(v) \quad \text{for all } u, v \in V \text{ and } \lambda, \mu \in F.$$

A bijective linear map between two vector spaces is an *isomorphism*. We assume the reader knows about the definitions of the image and kernel of a linear map, and can prove the rank-nullity theorem,

$$\dim V = \dim \operatorname{im} x + \dim \ker x.$$

A corollary of the rank-nullity theorem is that if $\dim V = \dim W$ and $x : V \to W$ is injective, then, since $\dim \operatorname{im} x = \dim V$, x is an isomorphism. One can draw the same conclusion if instead we know that x is surjective. We refer to this type of reasoning as an argument by *dimension counting*.

We can now state the isomorphism theorems for vector spaces.

Theorem 16.1 (Isomorphism theorems for vector spaces)

(a) If $x : V \to W$ is a linear map, then $\ker x$ is a subspace of V, $\operatorname{im} x$ is a subspace of W, and

$$V/\ker x \cong \operatorname{im} x.$$

(b) If U and W are subspaces of a vector space, $(U + W)/W \cong U/(U \cap W)$.

(c) Suppose that U and W are subspaces of a vector space V such that $U \subseteq W$. Then W/U is a subspace of V/U and $(V/U)/(W/U) \cong V/W$.

Proof

For part (a), define a map $\varphi : V/\ker x \to \operatorname{im} x$ by

$$\varphi(v + \ker x) = x(v).$$

This map is well-defined since if $v + \ker x = v' + \ker x$ then $v - v' \in \ker x$, so $\varphi(v + \ker x) = x(v) = x(v') = \varphi(v' + \ker x)$. It is routine to check that φ is linear, injective, and surjective, so it gives the required isomorphism.

To prove (b), consider the composite of the inclusion map $U \to U + W$ with the quotient map $U + W \to (U + W)/W$. This gives us a linear map $U \to (U + W)/W$. Under this map, $x \in U$ is sent to $0 \in (U + W)/W$ if and only if $x \in W$, so its kernel is $U \cap W$. Now apply part (a).

Part (c) can be proved similarly; we leave this to the reader. $\square$

Parts (a), (b) and (c) of this theorem are known respectively as the *first*, *second*, and *third isomorphism theorems*. See Exercise 16.5 for one application.

16.3 Matrices and Diagonalisation

Suppose that $x : V \to V$ is a linear transformation of a finite-dimensional vector space V. Let $\{v_1, \ldots, v_n\}$ be a basis of V. Using this basis, we may define scalars a_{ij} by

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

We say that the $n \times n$ matrix A with entries (a_{ij}) is the *matrix of* x with respect to our chosen basis. Conversely, given a basis of V and a matrix A, we can use the previous equation to define a linear map x, whose matrix with respect to this basis is A.

Exercise 16.1

(i) Let $x : V \to V$ and $y : V \to V$ be linear maps with matrices A and B with respect to a basis of V. Show that, with respect to this basis, the matrix of the composite map yx is the matrix product BA.

(ii) Suppose that x has matrix A with respect to the basis $v_1, \ldots, v_n$ of V. Let $w_1, \ldots, w_n$ be another basis of V. Show that the matrix of A in this new basis is $P^{-1}AP$ where the matrix $P = (p_{ij})$ is defined by

$$w_j = \sum_{i=1}^{n} p_{ij} v_i.$$

Matrices related in this way are said to be *similar*.

It had been said that “a true gentleman never takes bases unless he really has to.” We generally agree with this sentiment, preferring to use matrices only when they are necessary for explicit computations (for example in Chapter 12 when we look at the classical Lie algebras). When we are obliged to consider matrices, then we can at least try to choose bases so that they are of a convenient form.

Recall that a non-zero vector $v \in V$ such that $x(v) = \lambda v$ is said to be an *eigenvector* of x with corresponding *eigenvalue* λ. The *eigenspace* for eigenvalue λ is the vector subspace

$$\{v \in V : x(v) = \lambda v\}.$$

It is an elementary fact that non-zero vectors in different eigenspaces are linearly independent. (This will often be useful for us; for example, see step 1 in the proof of Theorem 8.5.)

The linear map x can be represented by a diagonal matrix if and only if V has a basis consisting of eigenvectors for x. This is the same as saying that the space V is a direct sum of x-eigenspaces,

$$V = V_{\lambda_1} \oplus V_{\lambda_2} \oplus \ldots \oplus V_{\lambda_r},$$

where the λ_i are the distinct eigenvalues of x. If this is the case, we say that x is *diagonalisable*.

Note that $\lambda \in \mathbf{C}$ is an eigenvalue of x if and only if $\ker(x - \lambda 1_V)$ is non-zero, which is the case if and only if $\det(x - \lambda 1_V) = 0$. The eigenvalues of x are therefore the roots of the *characteristic polynomial* of x, defined by

$$c_x(X) = \det(x - X1_V),$$

where X is an indeterminant. Since over $\mathbf{C}$ any non-constant polynomial has a root, this shows that any linear transformation of a complex vector space has an eigenvalue.

The characteristic polynomial of x does not in itself give enough information to determine whether x is diagonalisable — consider for example the matrices

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

To get further, one needs the minimal polynomial. The *minimal polynomial* of x is the monic polynomial of least degree which kills x, so $m(X) = X^d + a_{d-1}X^{d-1} + \ldots + a_1X + a_0$ is the minimal polynomial of x if

$$x^d + a_{d-1}x^{d-1} + \ldots + a_1x + a_0 1_V = 0$$

and the degree d is as small as possible.

An important property of the minimal polynomial is that if $f(X)$ is any polynomial such that $f(x) = 0$ then $m(X)$ divides $f(X)$.

Exercise 16.2

Prove this assertion by using polynomial division to write $f(X) = a(X)m(X) + r(X)$, where the remainder polynomial $r(X)$ is either 0 or has degree less than that of $m(X)$, and then showing that $r(x) = 0$.

By the famous theorem of Cayley–Hamilton (see Exercise 16.4), the minimal polynomial of x divides the characteristic polynomial of x. We now explore some of the arguments in which the minimal polynomial is used.

16.3.1 The Primary Decomposition Theorem

Theorem 16.2 (Primary decomposition theorem)

Suppose the minimal polynomial of x factorises as

$$(X-\lambda_1)^{a_1}\dots(X-\lambda_r)^{a_r},$$

where the λ_i are distinct and each $a_i \geq 1$. Then V decomposes as a direct sum of x-invariant subspaces V_i,

$$V = V_1 \oplus V_2 \oplus \ldots \oplus V_r,$$

where $V_i = \ker(x-\lambda_i 1_V)^{a_i}$. The subspaces V_i are said to be the *generalised eigenspaces* of x.

This theorem may be proved by repeatedly applying the following lemma.

Lemma 16.3

If $f(X) \in \mathbb{C}[X]$ and $g(X) \in \mathbb{C}[X]$ are coprime polynomials such that $f(x)g(x) = 0$, then $\operatorname{im} f(x)$ and $\operatorname{im} g(x)$ are x-invariant subspaces of V. Moreover,

(i) $V = \operatorname{im} f(x) \oplus \operatorname{im} g(x)$, and

(ii) $\operatorname{im} f(x) = \ker g(x)$ and $\operatorname{im} g(x) = \ker f(x)$.

Proof

If $v = f(x)w$, then $xv = f(x)xw$, so the subspaces $\operatorname{im} f(x)$ and $\operatorname{im} g(x)$ are x-invariant. By Euclid's algorithm, there exist polynomials $a(X), b(X) \in \mathbf{C}[X]$ such that $a(X)f(X) + b(X)g(X) = 1$, so for any $v \in V$,

$$f(x)(a(x)v) + g(x)(b(x)v) = v. \tag{$\star$}$$

This shows that $V = \operatorname{im} f(x) + \operatorname{im} g(x)$. If $v \in \operatorname{im} g(x)$ with, say, $v = g(x)w$, then $f(x)v = f(x)g(x)w = 0$, so $\operatorname{im} g(x) \subseteq \ker f(x)$. On the other hand, if $f(x)v = 0$, then by ($\star$), $v = g(x)(b(x)v)$ so $v \in \operatorname{im} g(x)$. Finally, if

$$v \in \operatorname{im} f(x) \cap \operatorname{im} g(x) = \ker f(x) \cap \ker g(x),$$

then as $f(x)a(x)v = a(x)f(x)v = 0$ and similarly $b(x)g(x) = 0$, it follows from ($\star$) that $v = 0$. □

The following criterion for a linear map to be diagonalisable follows directly from the primary decomposition theorem.

Theorem 16.4

Let $x : V \to V$ be a linear map of a vector space V. Then x is diagonalisable if and only if the minimal polynomial of x splits as a product of distinct linear factors. □

Corollary 16.5

Let $x : V \to V$ be a diagonalisable linear transformation. Suppose that U is a subspace of V which is invariant under x, that is, $x(u) \in U$ for all $u \in U$.

(a) The restriction of x to U is diagonalisable.

(b) Given any basis of U consisting of eigenvectors for x, we may extend this basis to a basis of V consisting of eigenvectors for x.

Proof

Let $m(X)$ be the minimal polynomial of $x : V \to V$. Let $m_U(X)$ be the minimal polynomial of x, regarded just as a linear transformation of U. Then $m(x)(U) = 0$, so $m_U(X)$ must divide $m(X)$. Hence $m_U(X)$ is a product of distinct linear factors.

Now let $V = V_{\lambda_1} \oplus \ldots \oplus V_{\lambda_r}$ be the decomposition of V into distinct eigenspaces of x. Since x acts diagonalisably on U we have

$$U = U \cap V_{\lambda_1} \oplus \ldots \oplus U \cap V_{\lambda_r}.$$

Extend the basis of each $U \cap V_{\lambda_i}$ to a basis of V_{λ_i}. This gives us a basis of V of the required form. □

We now give another application of the primary decomposition theorem.

Lemma 16.6

Suppose that x has minimal polynomial

$$f(X) = (X - \lambda_1)^{a_1} \ldots (X - \lambda_r)^{a_r},$$

where the λ_i are pairwise distinct. Let the corresponding primary decomposition of V as a direct sum of generalised eigenspaces be

$$V = V_1 \oplus \ldots \oplus V_r,$$

where $V_i = \ker(x - \lambda_i 1_V)^{a_i}$. Then, given any $\mu_1, \ldots, \mu_r \in \mathbb{C}$, there is a polynomial $p(X)$ such that

$$p(x) = \mu_1 1_{V_1} + \mu_2 1_{V_2} \ldots + \mu_r 1_{V_r}.$$

Proof

Suppose we could find a polynomial $f(X) \in \mathbf{C}[X]$ such that

$$f(X) \equiv \mu_i \bmod (X - \lambda_i)^{a_i}.$$

Take $v \in V_i = \ker(x - \lambda_i 1_V)^{a_i}$. By our supposition, $f(X) = \mu_i + a(X)(X - \lambda_i)^{a_i}$ for some polynomial $a(X)$. Hence

$$f(x)v = \mu_i 1_{V_i} v + a(x)(x - \lambda_i)^{a_i} v = \mu_i v,$$

as required.

The polynomials $(X - \lambda_1)^{a_1} \ldots, (X - \lambda_r)^{a_r}$ are coprime. We may therefore apply the Chinese Remainder Theorem, which states that in these circumstances the map

$$\mathbf{C}[X] \to \bigoplus_{i=1}^{r} \frac{\mathbf{C}[X]}{(X - \lambda_i)^{a_i}}$$

$$f(X) \mapsto (f(X) \bmod (X - \lambda_1)^{a_1}, \ldots, f(X) \bmod (X - \lambda_r)^{a_r})$$

is surjective, to obtain a suitable $p(X)$. □

In terms of matrices, this lemma says that

$$p(x) = \begin{pmatrix} \mu_1 I_{n_1} & 0 & \ldots & 0 \\ 0 & \mu_2 I_{n_2} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \mu_r I_{n_r} \end{pmatrix},$$

where $n_i = \dim V_i$ and I_s denotes the $s \times s$ identity matrix.

16.3.2 Simultaneous Diagonalisation

In the main text, we shall several times have a finite family of linear transformations of a vector space V, each of which is individually diagonalisable. When can one find a basis of V in which they are all simultaneously diagonal?

Lemma 16.7

Let $x_1, \ldots, x_k : V \to V$ be diagonalisable linear transformations. There is a basis of V consisting of simultaneous eigenvectors for all the x_i if and only if they commute. (That is, $x_i x_j = x_j x_i$ for all pairs i, j.)

Proof

For the “only if” direction we note that diagonal matrices commute with one another, so if we can represent all the x_i by diagonal matrices, they must commute.

The main step in the “if” direction is the case $k = 2$. Write V as a direct sum of eigenspaces for x_1, say $V = V_{\lambda_1} \oplus \ldots \oplus V_{\lambda_r}$, where the λ_i are the distinct eigenvalues of x_1. If $v \in V_{\lambda_i}$ then so is $x_2(v)$, for

$$x_1 x_2(v) = x_2 x_1(v) = x_2(\lambda_i v) = \lambda_i(x_2(v)).$$

We now apply Corollary 16.5(a) to deduce that x_2 restricted to V_{λ_i} is diagonalisable. A basis of V_{λ_i} consisting of eigenvectors for x_2 is automatically a basis of eigenvectors for x_1, so if we take the union of a basis of eigenvectors for x_2 on each V_{λ_i}, we get a basis of V consisting of simultaneous eigenvectors for both x_1 and x_2.

The inductive step is left to the reader. □

In Exercise 16.6, we give a small generalisation which will be needed in the main text.

16.4 Interlude: The Diagonal Fallacy

Consider the following (fallacious) argument. Let V be a 2-dimensional vector space, say with basis v_1, v_2. Let $x : V \to V$ be the linear map whose matrix with respect to this basis is

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

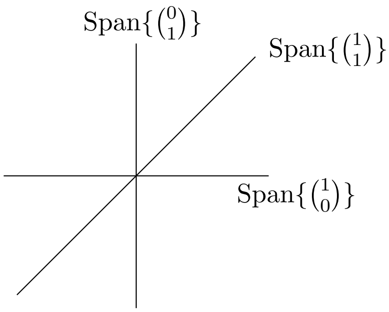
We claim that if U is a subspace of V such that $x(U) \subseteq U$, then either $U = 0$, $U = \text{Span}\{v_1\}$, or $U = V$. Clearly each of these subspaces is invariant under x, so we only need to prove that there are no others. But since $x(v_2) = v_1$, $\text{Span}\{v_2\}$ is not x-invariant. (QED?)

Here we committed the *diagonal fallacy*: We assumed that an arbitrary subspace of V would contain one of our chosen basis vectors. This assumption is very tempting — which perhaps explains why it is so often made — but it is nonetheless totally unjustified.

Exercise 16.3

Give a correct proof of the previous result.

The following diagram (which is frequently useful as a counterexample in linear algebra) illustrates how the fallacy we have been discussing gets its name.


$\mathrm{Span}\{\binom{0}{1}\}$
$\mathrm{Span}\{\binom{1}{1}\}$
$\mathrm{Span}\{\binom{1}{0}\}$

16.5 Jordan Canonical Form

Let V be a finite-dimensional complex vector space and let $x : V \to V$ be a linear map. Exercise 6.2 outlines the proof that one can always find a basis of V in which x is represented by an upper triangular matrix. For many purposes, this result is sufficient. For example, it implies that a nilpotent map may be represented by a strictly upper triangular matrix, and so nilpotent maps have trace 0.

Sometimes, however, one needs the full strength of Jordan canonical form. A general matrix in Jordan canonical form looks like

$$\begin{pmatrix} A_1 & 0 & \dots & 0 \\ 0 & A_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & A_r \end{pmatrix},$$

where each A_i is a *Jordan block matrix* $J_t(\lambda)$ for some $t \in \mathbb{N}$ and $\lambda \in \mathbb{C}$:

$$J_t(\lambda) = \begin{pmatrix} \lambda & 1 & 0 & \dots & 0 & 0 \\ 0 & \lambda & 1 & \dots & 0 & 0 \\ 0 & 0 & \lambda & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & \lambda & 1 \\ 0 & 0 & 0 & \dots & 0 & \lambda \end{pmatrix}_{t \times t} .$$

We now outline a proof that any linear transformation of a complex vector space can be represented by a matrix in Jordan canonical form.

The first step is to reduce to the case where $x^q = 0$ for some $q \geq 1$; that is, x is a nilpotent linear map.

By the primary decomposition theorem, it suffices to consider the case where x has only one eigenvalue, say λ. Then by considering $x - \lambda 1_V$, we may reduce to the case where x acts nilpotently. So it suffices to show that a nilpotent transformation can be put into Jordan canonical form.

16.5.1 Jordan Canonical Form for Nilpotent Maps

We shall work by induction on $\dim V$.

Suppose that $x^q = 0$ and $x^{q-1} \neq 0$. Let $v \in V$ be any vector such that $x^{q-1}v \neq 0$. One can check that the vectors $v, xv, \ldots, x^{q-1}v$ are linearly independent. Their span, U say, is an x-invariant subspace of V. With respect to the given basis of U, the matrix of $x : U \to U$ is the $q \times q$ matrix

$$J_q(0) = \begin{pmatrix} 0 & 1 & 0 & \ldots & 0 \\ 0 & 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & 1 \\ 0 & 0 & 0 & \ldots & 0 \end{pmatrix}.$$

Suppose we can find an x-invariant complementary subspace to U; that is, a subspace C such that x maps C into C and $V = U \oplus C$. Then, by induction, there is a basis of C in which the matrix of x restricted to C is in Jordan canonical form. Putting the bases of C and U together gives us a suitable basis for V.

To show that a suitable complement exists, we use a further induction on q. If $q = 1$, then $x = 0$ and any vector space complement to $\operatorname{Span}\{v\}$ will do. Now suppose we can find complements when $x^{q-1} = 0$.

Consider $\operatorname{im} x \subseteq V$. On $\operatorname{im} x$, x acts as a nilpotent linear map whose $q-1$ power is 0, so by induction on q we get

$$\operatorname{im} x = \operatorname{Span}\left\{xv, \ldots, x^{q-1}v\right\} \oplus W$$

for some x-invariant subspace W. Note that $U \cap W = 0$. Our task is to extend W to a suitable x-invariant complement for U in V.

Suppose first that $W = 0$. In this case, $\operatorname{im} x = \operatorname{Span}\left\{xv, \ldots, x^{q-1}v\right\}$ and $\ker x \cap \operatorname{im} x = \left\langle x^{q-1}v \right\rangle$. Extend $x^{q-1}v$ to a basis of $\ker x$, say by $v_1, \ldots, v_s$. By the rank-nullity formula

$$v, xv \ldots, x^{q-1}v, v_1, \ldots, v_s$$

is a basis of V. The subspace spanned by $v_1, \dots, v_s$ is an x-invariant complement to U.

Now suppose that $W \neq 0$. Then x induces a linear transformation, say $\bar{x}$, on V/W. Let $\bar{v} = v + W$. Since $\operatorname{im} \bar{x} = \operatorname{Span}\left\{\bar{x}\bar{v}, \dots, \bar{x}^{q-1}\bar{v}\right\}$, the first case implies that there is an $\bar{x}$-invariant complement in V/W to $\operatorname{Span}\left\{\bar{v}, \bar{x}\bar{v}, \dots \bar{x}^{q-1}\bar{v}\right\}$. The preimage of this complement in V is a suitable complement to U.

16.6 Jordan Decomposition

Any linear transformation x of a complex vector space V has a *Jordan decomposition*, $x = d + n$, where d is diagonalisable, n is nilpotent, and d and n commute.

One can see this by putting x into Jordan canonical form: Fix a basis of V in which x is represented by a matrix in Jordan canonical form. Let d be the map whose matrix in this basis has the diagonal entries of x down its diagonal, and let $n = x - d$. For example we might have

$$x = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad d = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad n = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

As n is represented by a strictly upper triangular matrix, it is nilpotent. We leave it to the reader to check that d and n commute.

In applications it is useful to know that d and n can be expressed as polynomials in x. In the following lemma, we also prove a related result that is needed in Chapter 9.

Lemma 16.8

Let x have Jordan decomposition $x = d + n$ as above, where d is diagonalisable, n is nilpotent, and d, n commute.

(a) There is a polynomial $p(X) \in \mathbf{C}[X]$ such that $p(x) = d$.

(b) Fix a basis of V in which d is diagonal. Let $\bar{d}$ be the linear map whose matrix with respect to this basis is the complex conjugate of the matrix of d. There is a polynomial $q(X) \in \mathbf{C}[X]$ such that $q(x) = \bar{d}$.

Proof

Let $\lambda_1, \dots, \lambda_r$ be the distinct eigenvalues of x. The minimal polynomial of x is then

$$m(X) = (X - \lambda_1)^{a_1} \dots (X - \lambda_r)^{a_r},$$

where a_i is the size of the largest Jordan block with eigenvalue λ_i.

We can now apply Lemma 16.6 to get the polynomials we seek. For part (a) take $\mu_i = \lambda_i$, and for part (b) take $\mu_i = \bar{\lambda}_i$. □

Part (a) of this lemma can be used to prove that the Jordan decomposition of a linear map is unique — see Exercise 16.7 below.

16.7 Bilinear Algebra

As well as the books already mentioned, we recommend Artin's *Geometric Algebra* [1] for further reading on bilinear algebra. From now on, we let V be an n-dimensional vector space over a field F.

16.7.1 Dual spaces

The *dual space* of V, denoted $V^\star$, is by definition the set of all linear maps from V to F. Thus, if $f, g \in V^\star$, then $f + g$ is defined by $(f+g)(v) = f(v) + g(v)$ for $v \in V$, and if $\lambda \in F$, then λf is defined by $(\lambda f)(v) = \lambda f(v)$.

Given a vector space basis $\{v_1, \dots, v_n\}$ of V, one defines the associated *dual basis* of V as follows. Let $f_i : V \to F$ be the linear map defined on basis elements by

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

It is not hard to check that $f_1, \dots, f_n$ is a basis for $V^\star$. In particular $\dim V = \dim V^\star$.

The dual space of V^* can be identified with V in a natural way. Given $v \in V$, we may define an *evaluation map* $\varepsilon_v : V^* \to F$ by

$$\varepsilon_v(f) := f(v) \quad \text{for all } f \in V^*.$$

It is straightforward to check that ε_v is linear and so belongs to the dual space of $V^\star$; that is, to $V^{\star\star}$. Moreover, the map $v \mapsto \varepsilon_v$ (which we might call ε) from V to $V^{\star\star}$ is itself linear. We claim that $\varepsilon : V \to V^{\star\star}$ is an isomorphism.

Since we have already shown that $\dim V = \dim V^\star = \dim V^{\star\star}$, it is sufficient to show that $\varepsilon_v = 0$ implies $v = 0$. One way to do this is as follows. If $v \neq 0$, then we may extend v to a basis of V and take the associated dual basis. Then $f_1(v) = 1$ and hence $\varepsilon_v(f_1) \neq 0$, so $\varepsilon_v \neq 0$.

If U is a subspace of V we let

$$U^\circ = \{f \in V^\star : f(u) = 0 \text{ for all } u \in U\}$$

be the *annihilator* of U in $V^\star$. One can show that U° is a subspace of $V^\star$ and that

$$\dim U + \dim U^\circ = \dim V.$$

A proof of the last statement is outlined in Exercise 16.8.

Given a subspace W of V^*, we can similarly define the annihilator of W in $V^{\star\star}$. Under the identification of V^{**} with V, the annihilator of W becomes

$$W^0 = \{v \in V : f(v) = 0 \text{ for all } f \in W\}.$$

In particular, we have $\dim W + \dim W^0 = \dim V$.

16.7.2 Bilinear Forms

Definition 16.9

A *bilinear form* on V is a map

$$(-,-) : V \times V \to F$$

such that

$$\begin{aligned}(\lambda_1 v_1 + \lambda_2 v_2, w) &= \lambda_1(v_1, w) + \lambda_2(v_2, w),\\ (v, \mu_1 w_1 + \mu_2 w_2) &= \mu_1(v, w_1) + \mu_2(v, w_2),\end{aligned}$$

for all $v, w, v_i, w_i \in V$ and $\lambda_i, \mu_i \in F$.

For example, if $F = \mathbf{R}$ and $V = \mathbf{R}^n$, then the usual dot product is a bilinear form on V.

As for linear transformations, we can represent bilinear forms by matrices. Suppose that $(-,-)$ is a bilinear form on the vector space V and that V has basis $\{v_1, \ldots, v_n\}$. The *matrix of* $(-,-)$ with respect to this basis is $A = (a_{ij})$, where $a_{ij} = (v_i, v_j)$. If we change the basis, say to $\{w_1, \ldots, w_n\}$, then the new matrix representing $(-,-)$ is $P^t A P$ where $P = (p_{ij})$ is the $n \times n$ matrix defined by

$$w_j = \sum_{i=1}^n p_{ij} v_i.$$

Matrices related in this way are said to be *congruent*.

Conversely, given an $n \times n$ matrix $S = (s_{ij})$, we may define a bilinear form on V by setting

$$(v_i, v_j) = s_{ij}$$

and extending "bilinearly" to arbitrary elements in $V \times V$. That is, if $v = \sum_i \lambda_i v_i$ and $w = \sum_j \mu_j v_j$ with λ_i and μ_j scalars, then

$$(v, w) = \sum_{i=1}^{n} \sum_{j=1}^{n} s_{ij} \lambda_i \mu_j .$$

The last equation may be written in matrix form as

$$(v, w) = (\lambda_1 \dots \lambda_n) \begin{pmatrix} s_{11} & \dots & s_{1n} \\ \vdots & \ddots & \vdots \\ s_{n1} & \dots & s_{nn} \end{pmatrix} \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_n \end{pmatrix} .$$

Given a subset U of V, we set

$$U^{\perp} := \{v \in V : (u, v) = 0 \text{ for all } u \in U\}.$$

This is always a subspace of V. We say that the form $(-,-)$ is *non-degenerate* if $V^{\perp} = \{0\}$.

Example 16.10

Let U be a $2m$-dimensional vector space with basis $u_1, \dots, u_{2m}$, and let

$$S = \begin{pmatrix} 0 & I_m \\ I_m & 0 \end{pmatrix} ,$$

where I_m is the identity matrix of size $m \times m$. The bilinear form associated to S may be shown to be non-degenerate. (For example, this follows from Exercise 16.9.) However, the restriction of the form to the subspace spanned by $u_1, \dots, u_m$ is identically zero.

For a more substantial example, see Exercise 16.10 below.

We now explain the connection between bilinear forms and dual spaces. Let $\varphi : V \to V^{\star}$ be the linear map defined by $\varphi(v) = (-, v)$. That is, $\varphi(v)$ is the linear map sending $u \in V$ to (u, v). If $(-,-)$ is non-degenerate, then $\ker \varphi = 0$, so by dimension counting, φ is an isomorphism. Thus every element of $V^{\star}$ is of the form $(-, v)$ for a unique $v \in V$; this is a special case of the *Riesz representation theorem*. A small generalisation of this argument can be used to prove the following lemma — see Exercise 16.8.

Lemma 16.11

Suppose that $(-,-)$ is a non-degenerate bilinear form on the vector space V. Then, for all subspaces U of V, we have

$$\dim U + \dim U^{\perp} = \dim V.$$

If $U \cap U^{\perp} = 0$, then $V = U \oplus U^{\perp}$ and, furthermore, the restrictions of $(-,-)$ to U and to $U^{\perp}$ are non-degenerate. □

16.7.3 Canonical Forms for Bilinear Forms

Definition 16.12

Suppose that $(-,-) : V \times V \to F$ is a bilinear form. We say that $(-,-)$ is *symmetric* if $(v,w) = (w,v)$ for all $v, w \in V$ and that $(-,-)$ is *skew-symmetric* or *symplectic* if $(v,w) = -(w,v)$ for all $v, w \in V$.

In the main text, we shall only need to deal with bilinear forms that are either symmetric or skew-symmetric. For such a form, $(v,w) = 0$ if and only if $(w,v) = 0$. When $F = \mathbf{R}$, a symmetric bilinear form with $(v,v) \geq 0$ for all $v \in V$ and such that $(v,v) = 0$ if and only if $v = 0$ is said to be an *inner product.*

A vector $v \in V$ is said to be *isotropic* with respect to a form $(-,-)$ if $(v,v) = 0$. For example, if $(-,-)$ is symplectic and the characteristic of the field is not 2, then all elements in V are isotropic. But symmetric bilinear forms can also have isotropic vectors (as long as they do not come from inner products). For example, in Example 16.10 above, the basis of U consists of isotropic vectors.

If $(-,-)$ is non-degenerate and $v \in V$ is isotropic, then there exists some $w \in V$ such that $(v,w) \neq 0$. Clearly v and w must be linearly independent. This observation motivates the following lemma (which we use in Appendix C).

Lemma 16.13

Suppose V has a non-degenerate bilinear form $(-,-)$. Suppose U_1 and U_2 are trivially-intersecting subspaces of V such that $(u,v) = 0$ for all $u, v \in U_1$ and for all $u, v \in U_2$ and that $(-,-)$ restricted to $U_1 \oplus U_2$ is non-degenerate. Then, if $\{u_1, \ldots, u_m\}$ is a basis of U_1 there is a basis $\{u_1', \ldots, u_n'\}$, of U_2 such that

$$(u_i, u_j') = \begin{cases} 1 & i = j \\ 0 & i \neq j. \end{cases}$$

Proof

Consider the map $\gamma : U_2 \to U_1^*$ defined by $\gamma(v) = (-, v)$. That is, $\gamma(v)(u) = (u, v)$ for all $u \in U_1$. This map is linear, and it is injective because the restriction of $(-,-)$ to $U_1 \oplus U_2$ is non-degenerate, so we have

$$\dim U_2 \leq \dim U_1^* = \dim U_1.$$

By symmetry, we also have $\dim U_1 \leq \dim U_2$, so γ must be an isomorphism.

Given the basis $\{u_1, \ldots, u_n\}$ of U_1, let $\{f_1, \ldots, f_n\}$ be the corresponding dual basis of U_1^*. For $1 \leq j \leq n$, let $u_j' \in U_2$ be the unique vector such that $\gamma(u_j') = f_j$. Then we have

$$(u_i, u_j') = f_j(u_i) = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

as required. □

Note that if $(-,-)$ is symmetric, then the matrix of $(-,-)$ with respect to this basis of $U_1 \oplus U_2$ is

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

An analogous result holds if $(-,-)$ is skew-symmetric.

In the following we assume that the characteristic of F is not 2.

Lemma 16.14

Let $(-,-)$ be a non-degenerate symmetric bilinear form on V. Then there is a basis $\{v_1, \ldots, v_n\}$ of V such that $(v_i, v_j) = 0$ if $i \neq j$ and $(v_i, v_i) \neq 0$.

Proof

We use induction on $n = \dim V$. If $n = 1$, then the result is obvious, so we may assume that $\dim V \geq 2$.

Suppose that $(v, v) = 0$ for all $v \in V$. Then, thanks to the identity

$$(v + w, v + w) = (v, v) + (w, w) + 2(v, w),$$

we have $(v, w) = 0$ for all $v, w \in V$, which contradicts our assumption that $(-,-)$ is non-degenerate. (This is where we need our assumption on the characteristic of F.)

We may therefore choose $v \in V$ so that $(v, v) \neq 0$. Let $U = \operatorname{Span}\{v\}$. By hypothesis $U \cap U^\perp = \{0\}$, so by Lemma 16.11 we have $V = U \oplus U^\perp$. Moreover, the restriction of $(-,-)$ to $U^{\perp}$ is non-degenerate. By the inductive hypothesis, there is a basis of $U^{\perp}$, say $\{v_2, \dots, v_n\}$, such that $(v_i, v_j) = 0$ for $i \neq j$ and $(v_i, v_i) \neq 0$ for $2 \leq i \leq n$. Since also $(v, v_j) = 0$ for $j \neq 1$, if we put $v_1 = v$ then the basis $\{v_1, \dots, v_n\}$ has the required properties. □

Depending on the field, we may be able to be more precise about the diagonal entries $d_i = (v_i, v_i)$. Suppose that $F = \mathbf{R}$. Then we may find $\lambda_i \in \mathbf{R}$ such that $\lambda_i^2 = |d_i|$. By replacing v_i with v_i/λ_i, we may assume that $(v_i, v_i) = \pm 1$. The bilinear form $(-,-)$ is an inner product if and only if $(v_i, v_i) > 0$ for all i.

If $F = \mathbf{C}$, then we can find λ_i so that $\lambda_i^2 = d_i$, and hence we may assume that $(v_i, v_i) = 1$ for all i, so the matrix representing $(-,-)$ is the $n \times n$ identity matrix.

Lemma 16.15

Suppose that $(-,-)$ is a non-degenerate symplectic bilinear form on V. Then we have $\dim V = 2m$ for some m. Moreover, there is a basis of V such that $(v_i, v_{i+n}) \neq 0$ for $1 \leq i \leq n$ and $(v_i, v_j) = 0$ if $|i - j| \neq n$.

Proof

Again we work by induction $\dim V$. Let $0 \neq v \in V$. Since $(-,-)$ is non-degenerate, we may find $w \in V$ such that $(v, w) \neq 0$. Since v, w are isotropic, it is clear that $\{v, w\}$ is linearly independent. Set $v_1 = v$ and $v_2 = w$. If $\dim V = 2$, then we are done. Otherwise, let U be the orthogonal complement of the space W spanned by v_1, v_2. One shows easily that $U \cap W = \{0\}$ and that by dimension counting $V = U \oplus W$. Now, the restriction of $(-,-)$ to U is non-degenerate and also symplectic. The result now follows by induction. □

When $F = \mathbf{R}$ or $F = \mathbf{C}$, it is again useful to scale the basis elements. In particular, when $F = \mathbf{C}$ we may arrange that the matrix representing $(-,-)$ has the form

$$\begin{pmatrix} 0 & I_m \\ -I_m & 0 \end{pmatrix},$$

where I_m is the $m \times m$ identity matrix.

EXERCISES

16.4. Let $x : V \to V$ be a linear transformation of a complex vector space. By the result mentioned at the start of §16.5, we may find a basis

$v_1, \ldots, v_n$ of V in which x is represented by an upper triangular matrix. Let $\lambda_1, \ldots, \lambda_n$ be the diagonal entries of this matrix. Show that if $1 \leq k \leq n$ then

$$(x - \lambda_k 1_V) \ldots (x - \lambda_n 1_V)V \subseteq \operatorname{Span}\{v_1, \ldots, v_{k-1}\}.$$

Hence prove the Cayley–Hamilton theorem for linear maps on complex vector spaces.

16.5. Let A be an $m \times n$ matrix with entries in a field F. Show that there is a bijective correspondence between solution sets of the equation $Ax = y$ for $y \in \operatorname{im} A$ and elements of the quotient vector space $F^n / \ker A$.

16.6.† Let V be a finite-dimensional vector space.

(i) Show that $\operatorname{Hom}(V, V)$, the set of linear transformations of V, is a vector space, and determine its dimension.

(ii) Let $A \subseteq \operatorname{Hom}(V, V)$ be a collection of commuting linear maps, each individually diagonalisable. Show that there is a basis of V in which all the elements of A are simultaneously diagonal.

(iii)* Can the assumption that V is finite-dimensional be dropped?

16.7.† Suppose that $x : V \to V$ is a linear map on a vector space V and that $x = d + n = d' + n'$ where d, d' are diagonalisable and n, n' are nilpotent, d and n commute and d' and n' commute. Show that d and d' commute. Hence show that $d - d' = n' - n = 0$. Deduce that the Jordan decomposition of a linear map is unique.

16.8. Let U be a subspace of the F-vector space V.

(i) Consider the restriction map $r : V^\star \to U^\star$, which takes a linear map $f : V \to F$ and regards it just as a map on U. Show that $\ker r = U^\circ$ and $\operatorname{im} r = U^\star$. Hence prove that

$$\dim U + \dim U^0 = \dim V.$$

(ii) Now suppose that $(-, -)$ is a non-degenerate bilinear form on V. By considering the linear map $\varphi : V \to U^\star$ defined by

$$\varphi(v)(u) = (u, v),$$

show that $\dim U + \dim U^\perp = \dim V$.

16.9. Let V be a finite-dimensional vector space with basis $\{v_1, \ldots, v_n\}$. Suppose that $(-,-)$ is a bilinear form on V, and let $a_{ij} = (v_i, v_j)$. Show that $V^\perp = \{0\}$ if and only if the matrix $A = (a_{ij})$ is non-singular.

16.10. Let V be a finite-dimensional vector space and let $\mathrm{Hom}(V, V)$ be the vector space of all linear transformations of V. Show that

$$(x, y) \mapsto \mathrm{tr}(xy)$$

defines a non-degenerate symmetric bilinear form on $\mathrm{Hom}(V, V)$. By Exercise 16.8(ii) this form induces an isomorphism

$$\mathrm{Hom}(V, V) \to \mathrm{Hom}(V, V)^\star.$$

What is the image of the identity map $1_V : V \to V$ under this isomorphism?