LINEAR ALGEBRA, MATH 50003: Lecture Notes

Lecturer: Martin Liebeck

1 Course Overview

Most of the course will consist of basic results on matrices, vector spaces and linear maps. The last part of the course will have a more geometrical flavour.

1.1 Matrix results

Let's begin with a survey of some of the highlights among the matrix results in the course. We start with a definition.

Definition Let A, B be $n \times n$ matrices over a field F. We say A is *similar* to B if there exists an invertible $n \times n$ matrix P such that $B = P^{-1}AP$.

Note that if we define a relation $\sim$ on $n \times n$ matrices by

$$A \sim B \Leftrightarrow A \text{ is similar to } B,$$

then $\sim$ is an equivalence relation (question on Problem Sheet 1).

Two similar matrices A, B share many basic properties: for example, they have

- the same determinant
- the same characteristic polynomial
- the same eigenvalues
- the same rank
- the same trace

(question on Problem Sheet 1). One of the major aims of the subject is:

Major Aim For an arbitrary $n \times n$ matrix A, find a "nice" matrix B such that $A \sim B$.

In the course we'll prove three famous therorems, in each of which the meaning of the word "nice" will be apparent.

Example Probably the nicest matrices are the diagonal ones. Recall that an $n \times n$ matrix A is *diagonalisable* if it is similar to a diagonal matrix $D = \text{diag}(\lambda_1, \ldots, \lambda_n)$ (the diagonal matrix with diagonal entries $\lambda_1, \ldots, \lambda_n$, the eigenvalues of A). This property can be used to do many computations with A, such as calculating any power A^k: a matrix P such that $D = P^{-1}AP$ can be computed (its columns are a basis of eigenvectors of A). Then $A = PDP^{-1}$, so

$$A^k = (PDP^{-1})\,(PDP^{-1}) \cdots (PDP^{-1}) = PD^kP^{-1},$$

and D^k is the diagonal matrix $D = \text{diag}(\lambda_1^k, \ldots, \lambda_n^k)$.

However, many matrices are not diagonalisable, for example

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

To see this, suppose that A is diagonalisable. Then since the only eigenvalue is 1, there exist P such that $P^{-1}AP = \mathrm{diag}(1,1) = I$, so $A = PIP^{-1} = I$, a contradiction.

So not every matrix can be diagonalised. However, every complex matrix can be *triangularised*. This is one of the first main results of the course:

Triangularisation Theorem *If A is an $n \times n$ matrix over $\mathbb{C}$, then A is similar to an upper triangular matrix, i.e. there exists P such that*

$$P^{-1}AP = \begin{pmatrix} \lambda_1 & & & & \\ 0 & \lambda_2 & & * & \\ & & \cdot & & \\ & & & \cdot & \\ 0 & 0 & & & \lambda_n \end{pmatrix}.$$

Note that this result does not hold for matrices aver arbitrary fields: for example over the real numbers $\mathbb{R}$, the matrix $\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ has complex eigenvalues $\pm i$, so is not similar to a real upper triangular matrix.

The theorem has a more serious drawback though: there is nothing unique about an upper triangular matrix similar to A. For example, for any $a, b, a', b' \neq 0$,

$$\begin{pmatrix} 1 & a & b \\ & 1 & 0 \\ & & 1 \end{pmatrix} \sim \begin{pmatrix} 1 & a' & b' \\ & 1 & 0 \\ & & 1 \end{pmatrix},$$

(question on Sheet 1), so if A is similar to one such matrix, it is similar to all of them.

It is very desirable to have a *unique* matrix of a nice form that is similar to A, and that is provided by the next main result.

Jordan Canonical Form Theorem *If A is an $n \times n$ matrix over $\mathbb{C}$, then A is similar to a matrix of the form*

$$J = \begin{pmatrix} J_1 & & & & \\ & J_2 & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & J_k \end{pmatrix},$$

a block-diagonal matrix with blocks

$$J_i = \begin{pmatrix} \lambda_i & 1 & 0 & \dots & 0 & 0 \\ 0 & \lambda_i & 1 & \dots & 0 & 0 \\ 0 & 0 & \lambda_i & \dots & 0 & 0 \\ & & & \dots & & \\ 0 & 0 & 0 & \dots & \lambda_i & 1 \\ 0 & 0 & 0 & \dots & 0 & \lambda_i \end{pmatrix}$$

(these are called Jordan blocks). The collection of Jordan blocks $J_1, \dots, J_k$ is uniquely determined by A.

We call the matrix J the Jordan Canonical Form (JCF) of A. Its uniqueness is a vital part of the theorem, since it gives a powerful test for the similarity of two arbitrary complex matrices A and B: find the JCFs of A and B, call them J and J'. If J and J' are the same (apart from changing the order in which the Jordan blocks appear), then $A \sim B$; if not, then $A \not\sim B$. This test can be programmed very efficiently, and can be used for huge matrices.

The Jordan Canonical Form Theorem is an ideal result for complex matrices. But what about matrices over other fields, such as $\mathbb{R}$ or $\mathbb{Q}$ or the finite field $\mathbb{F}_p$ (the field of prime order p consisting of the integers $0, 1, \ldots, p-1$ with addition and multiplication modulo p)? The JCF theorem does not hold for arbitrary matrices over these fields, for the same reason that the Triangularisation theorem does not hold.

However we will prove another canonical form theorem – the Rational Canonical Form – that holds over arbitrary fields. To state this, we need a bit of notation. Let F be a field, and denote by $F[x]$ the set of polynomials in x over F. We can add and multiply polynomials (indeed, under addition and multiplication they form what is called a *ring*).

We call a polynomial $p(x) \in F[x]$ *monic* if it has degree $r \geq 1$ and its leading coefficient is 1, i.e.

$$p(x) = x^r + a_{r-1}x^{r-1} + \cdots + a_0. \tag{1}$$

Definition Let $p(x)$ be a monic polynomial of degree r as in (1). The *companion matrix* of $p(x)$ is the $r \times r$ matrix $C(p(x))$ defined as follows:

$$C(p(x)) = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 & -a_0 \\ 1 & 0 & 0 & \cdots & 0 & -a_1 \\ 0 & 1 & 0 & \cdots & 0 & -a_2 \\ & & & \cdots & & \\ 0 & 0 & 0 & \cdots & 1 & -a_{r-1} \end{pmatrix}.$$

For example,

$$C(x^3 - x + 1) = \begin{pmatrix} 0 & 0 & -1 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}.$$

Note that $C(p(x))$ has characteristic polynomial $p(x)$ (question on Sheet 1).

Rational Canonical Form Theorem *Let A be an $n \times n$ matrix over F, with characteristic polynomial $p(x)$.*

(i) There exists a factorization $p(x) = p_1(x) \cdots p_k(x)$ such that A is similar to a block-diagonal matrix with blocks $C(p_i(x))$ for $i = 1, \ldots k$.

(ii) Under some conditions, the polynomials $p_1(x), \ldots, p_k(x)$ are uniquely determined by A.

The "conditions" in part (ii) will be spelled out when we state and prove the theorem in the lectures.

1.2 Geometry

The last part of the course will be concerned with some geometrical aspects of linear algebra.

Recall the *dot product* on $\mathbb{R}^n$: if $u = (u_1, \ldots, u_n)$ and $v = (v_1, \ldots, v_n) \in \mathbb{R}^n$, then

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

Much of the geometry of $\mathbb{R}^n$ is based on the dot product. For example, the length $||u|| = \sqrt{u.u}$, and the distance between u and v is $||u - v||$. Various types of $n \times n$ matrices fit naturally into this geometrical picture, for example

- P is *orthogonal* if $P^T P = I$ (which implies that $Pu.Pv = u.v$ for all u, v)
- A is *symmetric* if $A^T = A$ (which implies that $Au.v = u.Av$ for all u, v).

It is useful to axiomatise the basic properties of the dot product, to obtain the theory of *inner product spaces*: an inner product space is a real vector space with a map sending any pair of vectors u, v to a scalar (u, v) satisfying the following axioms:

(1) the map is linear in each variable u, v

(2) the map is symmetric, i.e. $(v, u) = (u, v)$ for all u, v

(3) $(u, u) > 0$ for all nonzero vectors u.

We shall develop the theory of inner product spaces. In order to extend the geometrical notions to vector spaces over arbitrary fields, we shall also develop the theory of bilinear forms.

2 Some revision from 1st Year Linear Algebra

This chapter is a summary of some of the theory of matrices and linear maps from the 1st year course that we'll need.

Let V be a finite dimensional vector space over a field F and $T : V \to V$ a linear map. If $B = \{v_1, \ldots, v_n\}$ is a basis of V, let

$$\begin{aligned} T(v_1) &= a_{11}v_1 + \ldots + a_{n1}v_n, \\ &\vdots \\ T(v_n) &= a_{1n}v_1 + \ldots + a_{nn}v_n \end{aligned}$$

where all the coefficients $a_{ij} \in F$. The *matrix of T with respect to B* is

$$[T]_B = (a_{ij}) = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix}.$$

Proposition 2.1 *Let $S : V \to V$ and $T : V \to V$ be linear transformations and let B be a basis of V. Then*

$$[ST]_B = [S]_B[T]_B,$$

where ST is the composition of S and T.

As a consequence of the proposition, the map $T \to [T]_B$ from linear maps to $n \times n$ matrices has many nice properties. For example, if $[T]_B = A$ then $[T^2]_B = A^2$ and similarly $\left[T^k\right]_B = A^k$ for any positive integer k. More generally, for a polynomial $q(x) = a_r x^r + \cdots + a_1 x + a_0$ ($a_i \in F$), define

$$q(A) = a_r A^r + \cdots + a_1 A + a_0 I$$

and

$$q(T) = a_r T^r + \cdots + a_1 T + a_0 I_V$$

where $I_V : V \to V$ is the identity map. Then Proposition 2.1 implies that

$$[q(T)]_B = q(A).$$

Change of basis

Let V be n-dimensional, and let bases $E = \{e_1, \ldots, e_n\}$ and $F = \{f_1, \ldots, f_n\}$ be two bases of V. Write

$$\begin{aligned} f_1 &= p_{11}e_1 + \cdots + p_{n1}e_n, \\ &\vdots \\ f_n &= p_{1n}e_1 + \cdots + p_{nn}e_n. \end{aligned}$$

and define P to be the $n \times n$ matrix (p_{ij}). We call P the *change of basis matrix* from E to F.

Proposition 2.2 (i) *The change of basis matrix P is invertible.*

(ii) *If $T : V \to V$ is a linear map, then $[T]_F = P^{-1}[T]_E P$ (so $[T]_E$ and $[T]_F$ are similar matrices).*

Determinants

As we already noted in Chapter 1, if A, B are similar $n \times n$ matrices, then they have the same determinant. Hence if $T : V \to V$ is a linear map, and E, F are two bases of V, then the matrices $[T]_E$ and $[T]_F$ have the same determinant (by Proposition 2.2(ii)). Therefore we can define the determinant $\det(T)$ of a linear map T to be the determinant of the matrix $[T]_E$ for any basis E of V. The *characteristic polynomial* of T is defined to be $\det(xI_V - T)$. This is a polynomial in x of degree $n = \dim V$.

Proposition 2.3 (i) *The eigenvalues of T are the roots of the characteristic polynomial of T.*

(ii) *If λ is an eigenvalue of T, the eigenvectors corresponding to λ are the nonzero vectors in*

$$E_\lambda = \{v \in V : (\lambda I_V - T)(v) = 0\} = \ker(\lambda I_V - T).$$

(iii) *The matrix $[T]_B$ is a diagonal matrix iff B consists of eigenvectors of T.*

Definition We call E_λ the λ-*eigenspace* of T. Note that E_λ is a subspace of V (since it is the kernel of the linear map $\lambda I_V - T$).

Proposition 2.4 *Let V a finite-dimensional vector space over $\mathbb{C}$, and let $T : V \to V$ be a linear map. Then T has an eigenvalue $\lambda \in \mathbb{C}$.*

Proof The characteristic polynomial of T has a root $\lambda \in \mathbb{C}$ by the Fundamental theorem of Algebra. $\square$

Note that Proposition 2.4 is not necessarily true for vector spaces over other fields. For example $T : \mathbb{R}^2 \to \mathbb{R}^2$ defined by $T(x_1, x_2) = (x_2, -x_1)$ has characteristic polynomial $x^2 + 1$, which has no real roots.

Diagonalisation

Recall that a linear map $T : V \to V$ is diagonalisable iff there exists a basis of V consisting of eigenvectors of T. Here is a very useful result on eigenvectors.

Proposition 2.5 *Let $T : V \to V$ be a linear map. Suppose $v_1, \ldots, v_k$ are eigenvectors of T corresponding to* distinct *eigenvalues $\lambda_1, \ldots, \lambda_k$. Then $v_1, \ldots, v_k$ are linearly independent.*

Corollary 2.6 *Let V be n-dimensional over F, and let $T : V \to V$ a linear map. Suppose the characteristic polynomial of T has n distinct roots in F. Then T is diagonalisable.*

Example Let

$$A = \begin{pmatrix} \lambda_1 & & & \\ 0 & \lambda_2 & * & \\ \vdots & & \ddots & \\ 0 & \cdots & 0 & \lambda_n \end{pmatrix}$$

be upper triangular, with diagonal entries $\lambda_1, \ldots, \lambda_n$, all distinct. The characteristic polynomial of A is $\prod_{i=1}^{n}(x - \lambda_i)$, which has roots $\lambda_1, \ldots, \lambda_n$. Hence by Corollary 2.6, A is diagonalisable, so there exists P such that $P^{-1}AP = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$.

Note that this is not necessarily true if the diagonal entries are not distinct, e.g. $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ is not diagonalisable.

As a final point about diagonalisation, it is sometimes important to specify which field we are working over. If A is an $n \times n$ matrix over a field F, we say A is diagonalisable *over F* if it is similar to a diagonal matrix with entries in F. For example, the matrix $\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ is not diagonalisable over $\mathbb{R}$, but it *is* diagonalisable over $\mathbb{C}$.

3 Algebraic and geometric multiplicities of eigenvalues

In this chapter we introduce and study two types of eigenvalue multiplicity.

Definition Let $T : V \to V$ be a linear map with characteristic polynomial $p(x)$. Let λ be a root of $p(x)$ (i.e. an eigenvalue of T). Then there is a positive integer $a(\lambda)$ such that

$$p(x) = (x - \lambda)^{a(\lambda)} q(x),$$

where λ is not a root of $q(x)$. We call $a(\lambda)$ the *algebraic multiplicity* of λ as an eigenvalue of T.

The *geometric multiplicity* of λ is defined to be

$$g(\lambda) = \dim E_\lambda,$$

where E_λ is the λ-eigenspace of T.

We adopt similar definitions for $n \times n$ matrices.

Example For $A = \begin{pmatrix} 1 & 1 \\ 0 & 2 \end{pmatrix}$, we have

$$a(1) = g(1) = 1, \;\; a(2) = g(2) = 1.$$

And for $B = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$, we have

$$a(1) = 2, g(1) = 1.$$

Proposition 3.1 *If λ is an eigenvalue of $T : V \to V$, then $g(\lambda) \leq a(\lambda)$.*

Proof Let $r = g(\lambda) = \dim E_\lambda$ and let $v_1, \ldots, v_r$ be a basis of E_λ. Extend to a basis of V:

$$B = \{v_1, \ldots, v_r, w_1, \ldots, w_s\}.$$

We work out the matrix $[T]_B$:

$$\begin{aligned} T(v_1) &= \lambda v_1, \\ &\vdots \\ T(v_r) &= \lambda v_r, \\ T(w_1) &= a_{11}v_1 + \cdots + a_{r1}v_r + b_{11}w_1 + \cdots + b_{s1}w_s, \\ &\vdots \\ T(w_s) &= a_{1s}v_1 + \cdots + a_{rs}v_r + b_{1s}w_1 + \cdots + b_{ss}w_s. \end{aligned}$$

So

$$[T]_B = \left(\begin{array}{cccc|ccc} \lambda & 0 & \cdots & 0 & a_{11} & \cdots & a_{1s} \\ 0 & \lambda & \cdots & 0 & \vdots & & \vdots \\ \vdots & \vdots & \ddots & & \vdots & & \vdots \\ 0 & 0 & \cdots & \lambda & a_{r1} & \cdots & a_{rs} \\ \hline 0 & \cdots & \cdots & 0 & b_{11} & \cdots & b_{1s} \\ \vdots & & & \vdots & \vdots & & \vdots \\ \vdots & & & \vdots & \vdots & & \vdots \\ 0 & \cdots & \cdots & 0 & b_{s1} & \cdots & b_{ss} \end{array}\right) = \begin{pmatrix} \lambda I_r & A \\ 0 & B \end{pmatrix}.$$

The characteristic polynomial of this is

$$p(x) = \det \left(\begin{array}{c|c} (x-\lambda)I_r & -A \\ \hline 0 & xI_s - B \end{array}\right).$$

Using Q4 on Sheet 1, this is

$$p(x) = \det((x-\lambda)I_r)\det(xI_s - B) = (x-\lambda)^r s(x),$$

where $s(x)$ is the characteristic polynomial of B. Hence the algebraic multiplicity $a(\lambda) \geq r = g(\lambda)$. $\square$

Using this we can prove the following basic criterion for diagonalisation.

Theorem 3.2 *Let* $\dim V = n$*, let* $T : V \to V$ *be a linear map, let* $\lambda_1, \dots, \lambda_r$ *be the distinct eigenvalues of* T*, and let the characteristic polynomial of* T *be*

$$p(x) = \prod_{i=1}^{r} (x - \lambda_i)^{a(\lambda_i)}$$

(so $\sum_{i=1}^{r} a(\lambda_i) = n$*). The following statements are equivalent:*

(1) T *is diagonalisable.*

(2) $\sum_{i=1}^{r} g(\lambda_i) = n$*.*

(3) $g(\lambda_i) = a(\lambda_i)$ *for all* i*.*

Proof We first prove (1) $\Rightarrow$ (2). Suppose (1) holds, so V has a basis B consisting of eigenvectors of T. Each vector in B is in some eigenspace E_{λ_i}, so

$$\sum_{i=1}^{r} g(\lambda_i) = \sum_{i=1}^{r} \dim E_{\lambda_i} \geq |B| = n.$$

By 3.1, $\sum_{i=1}^{r} g(\lambda_i) \leq \sum_{i=1}^{r} a(\lambda_i) = n$. Hence $\sum g(\lambda_i) = n$.

Next we show that (2) $\Leftrightarrow$ (3). This is easy, as

$$\sum g(\lambda_i) = n \Leftrightarrow \sum g(\lambda_i) = \sum a(\lambda_i) \Leftrightarrow g(\lambda_i) = a(\lambda_i) \ \forall i$$

(using 3.1 for the last implication).

To complete the proof, we show that (2) $\Rightarrow$ (1). Suppose (2) holds, so $\sum_{i=1}^{r} \dim E_{\lambda_i} = n$. Let B_i be a basis of E_{λ_i} and let $B = \bigcup_{i=1}^{r} B_i$, so $|B| = n$ (the B_i's are disjoint as they consist of eigenvectors for different eigenvalues).

We claim that B is a basis of V (hence (1) holds). Since $|B| = n = \dim V$, it is enough to show that B is linearly independent. Suppose there is a linear relation on the vectors in B, and write it as

$$\sum_{a \in B_1} \alpha_a a + \cdots + \sum_{z \in B_r} \alpha_z z = 0. \qquad (2)$$

Write

$$\begin{aligned} v_1 &= \textstyle\sum_{a \in B_1} \alpha_a a, \\ &\vdots \\ v_r &= \textstyle\sum_{z \in B_r} \alpha_z z, \end{aligned}$$

so $v_i \in E_{\lambda_i}$ and $v_1 + \cdots + v_r = 0$. As $\lambda_1, \dots, \lambda_r$ are distinct, the set of nonzero v_i's is linearly independent by 2.5. Therefore there can't be any nonzero v_i's, and so $v_i = 0$ for all i. Then $v_1 = \sum_{a \in B_1} \alpha_a a = 0$, so as B_1 is linearly independent (it is a basis of E_{λ_1}) all the coefficients $\alpha_a = 0$. Similarly all the other α's in (2) are 0. This completes the proof that B is linearly independent, hence a basis of V. $\square$

Using 3.2 we obtain a test to check whether a given $n \times n$ matrix or linear map is diagonalisable:

1. Find the characteristic polynomial, and factorise it as

$$\prod_{i=1}^{r} (x - \lambda_i)^{a(\lambda_i)}.$$

2. Calculate each $g(\lambda_i) = \dim E_{\lambda_i}$.
3. If $g(\lambda_i) = a(\lambda_i)$ for all i, YES.
   If $g(\lambda_i) < a(\lambda_i)$ for some i, NO.

Example Let $A = \begin{pmatrix} -3 & 1 & -1 \\ -7 & 5 & -1 \\ -6 & 6 & -2 \end{pmatrix}$. Check that

(1) Characteristic polynomial is $(x+2)^2(x-4)$.

(2) For eigenvalue 4: $a(4) = 1, g(4) = 1$ (as it is $\leq a(4)$).
For eigenvalue -2: $a(-2) = 2$, $g(-2) = \dim E_{-2} = 1$.

So $g(-2) < a(-2)$ and A is not diagonalisable.

4 Direct sums

Recall that if $U_1, \ldots, U_k$ are subspaces of a vector space V, we can form their *sum*

$$U_1 + \cdots + U_k = \{u_1 + \cdots + u_k : u_i \in U_i \text{ for all } i\},$$

which is another subspace of V. A *direct sum* of subspaces is a particular case of this, defined as follows.

Definition Let V be a vector space, and let $V_1, \ldots, V_k$ be subspaces of V. We write

$$V = V_1 \oplus V_2 \oplus \cdots \oplus V_k \tag{3}$$

if every vector $v \in V$ can be expressed as $v = v_1 + \cdots + v_k$ for *unique* vectors $v_i \in V_i$. The uniqueness statement means that if $v_1 + \cdots + v_k = v_1' + \cdots + v_k'$ with $v_i, v_i' \in V_i$, then $v_i = v_i'$ for all i. If (3) holds, we say that V is the *direct sum* of the subspaces $V_1, \ldots, V_k$.

As an obvious first example, $\mathbb{R}^2 = \mathrm{Sp}(1,0) \oplus \mathrm{Sp}(0,1)$. (Here, and throughout these notes, "Sp" is an abbreviation for "Span".)

It will be important for us to be able to check quickly whether the direct sum condition (3) holds. For a direct sum of two subspaces (the case $k = 2$), this is easy:

Proposition 4.1 *The following statements are equivalent:*

(1) $V = V_1 \oplus V_2$.

(2) $V_1 \cap V_2 = \{0\}$ *and* $\dim V_1 + \dim V_2 = \dim V$.

Proof First we show (1) $\Rightarrow$ (2). Assume (1), so that $V = V_1 \oplus V_2$. If there exists $0 \neq v \in V_1 \cap V_2$, then

$$v = v + 0 = 0 + v$$

gives two different expressions for v as a sum of vectors in V_1 and V_2, contradicting the uniqueness statement in the definition of a direct sum. Therefore $V_1 \cap V_2 = \{0\}$. It follows that

$$\dim V = \dim(V_1 + V_2) = \dim V_1 + \dim V_2 - \dim V_1 \cap V_2 = \dim V_1 + \dim V_2.$$

Hence (2) holds.

Now we show (2) $\Rightarrow$ (1). Assume that (2) holds. Then

$$\dim(V_1 + V_2) = \dim V_1 + \dim V_2 - \dim V_1 \cap V_2 = \dim V_1 + \dim V_2 = \dim V.$$

Hence $V = V_1 + V_2$. To show uniqueness, suppose $v_1 + v_2 = v_1' + v_2'$ with $v_i, v_i' \in V_i$. Then

$$v_1 - v_1' = v_2' - v_2 \in V_1 \cap V_2.$$

Since $V_1 \cap V_2 = \{0\}$, this implies that $v_1 = v_1', v_2 = v_2'$. Hence $V = V_1 \oplus V_2$. $\square$

The next result shows how to check the direct sum condition (3) for arbitrary values of k.

Proposition 4.2 *The following statements are equivalent:*

(1) $V = V_1 \oplus \cdots \oplus V_k$.

(2) $\dim V = \sum_{i=1}^k \dim V_i$, *and if* B_i *is a basis for* V_i *for* $1 \leq i \leq k$, *then* $B = B_1 \cup \cdots \cup B_k$ *is a basis of* V.

Proof First we prove (1) $\Rightarrow$ (2). Assume that $V = V_1 \oplus \cdots \oplus V_k$. Let B_i be a basis of V_i for $1 \leq i \leq k$, and let $B = B_1 \cup \cdots \cup B_k$.

Claim B is a basis of V.

Proof of Claim: Clearly B spans V, since $V = V_1 + \cdots + V_k$. Now we show linear independence. Suppse there is a linear relation on the vectors in B, and write this as

$$\sum_{a \in B_1} \alpha_a a + \cdots + \sum_{z \in B_r} \alpha_z z = 0. \tag{4}$$

Now $V = V_1 \oplus \cdots \oplus V_k$, hence $0 = 0 + \cdots + 0$ is the *unique* expression for the zero vector as a sum of vectors in $V_1, \ldots, V_k$. Hence each sum in the left hand side of (4) is equal to 0, and so all the α's in (4) are 0. This proves that B is linearly independent, hence is a basis, proving the Claim.

As in the proof of 4.1 we see that $V_i \cap V_j = \{0\}$ for $i \neq j$, and hence $B_i \cap B_j = \emptyset$ and B is the disjoint union of the B_i. By the Claim, therefore, we have

$$\dim V = |B| = \sum_{i=1}^k |B_i| = \sum_{i=1}^k \dim V_i,$$

so that (2) holds.

Now we prove that (2) $\Rightarrow$ (1). Assume that (2) holds. For each i let B_i be a basis of V_i, and let $B = \bigcup_1^k B_i$, a basis of V. As $\dim V = \sum_1^k \dim V_i$, we have $|B| = \sum |B_i|$, so the B_i's are disjoint sets. Every vector in V is in the span of B, hence is a sum of vectors in $V_1, \ldots, V_k$, so $V = V_1 + \cdots + V_k$. To prove uniqueness, suppose that

$$v_1 + \cdots + v_k = v_1' + \cdots + v_k'$$

where each $v_i, v_i' \in V_i$. Then

$$0 = (v_1 - v_1') + \cdots + (v_k - v_k').$$

If any term $v_i - v_i'$ is nonzero, this equation will give a nontrivial linear relation on the vectors in the basis B, a contradiction. Hence $v_i = v_i'$ for all i, proving uniqueness, and so $V = V_1 \oplus \cdots \oplus V_k$. $\square$

Example In $\mathbb{R}^4$ let $V_1 = \mathrm{sp}((1,1,0,0),(0,-1,1,0))$, $V_2 = \mathrm{sp}(2,1,2,1)$, $V_3 = \mathrm{sp}(0,0,1,1)$. Is $\mathbb{R}^4 = V_1 \oplus V_2 \oplus V_3$?

Answer: no, as $\{(1,1,0,0),(0,-1,1,0),(2,1,2,1),(0,0,1,1)\}$ is not a basis of $\mathbb{R}^4$. (The simplest way to check this is to write the vectors as the rows of a 4×4 matrix and show that this can be reduced by row operations to a matrix with a zero row.)

To complete this chapter, we demonstrate an important link between direct sums and linear maps. First we need a definition.

Definition Let $T : V \to V$ be a linear map, and W a subspace of V. We say that W is *T-invariant* if $T(W) \subseteq W$, where $T(W) = \{T(w) : w \in W\}$ (in other words, T maps $W \to W$). If W is T-invariant, write $T_W : W \to W$ for the *restriction* of T to W. Thus T_W is the linear map $W \to W$ defined by $T_W(w) = T(w)$ for all $w \in W$.

Proposition 4.3 *Let $T : V \to V$ be a linear map, and suppose that $V = V_1 \oplus \cdots \oplus V_k$, where each subspace V_i is T-invariant. For each i let B_i be a basis of V_i, and let A_i be the matrix of the restriction $[T_{V_i}]_{B_i}$. Then if B is the basis $\bigcup_1^k B_i$ of V, the matrix $[T]_B$ is the block-diagonal matrix*

$$[T]_B = \begin{pmatrix} A_1 & & & \\ & \cdot & & \\ & & \cdot & \\ & & & A_k \end{pmatrix}. \tag{5}$$

Proof Let $B_1 = \{v_1, \ldots, v_r\}$. Then $T(v_1) = T_{V_1}(v_1)$ is a vector in V_1, say $T(v_1) = a_{11}v_1 + \cdots + a_{r1}v_r$. Similarly for $T(v_2), \ldots,$ up to $T(v_r) = T_{V_1}(v_r) = a_{1r}v_1 + \cdots + a_{rr}v_r$. So we see that the top left hand block of $[T]_B$ is the $r \times r$ matrix (a_{ij}), which is $[T_{V_1}]_{B_1}$. Carrying on like this, we see that the next diagonal block is $[T_{V_2}]_{B_2}$, and so on. $\square$

Notation In view of the proposition, and for convenience of notation, we shall denote the block-diagonal matrix in (5) by $A_1 \oplus \cdots \oplus A_k$. Thus for $n_i \times n_i$ matrices A_i $(1 \le i \le k)$, we write

$$A_1 \oplus \cdots \oplus A_k = \begin{pmatrix} A_1 & & & \\ & \cdot & & \\ & & \cdot & \\ & & & A_k \end{pmatrix},$$

an $n \times n$ block-diagonal matrix, where $n = \sum_{i=1}^k n_i$.

5 Quotient spaces

Let V be a vector space over a field F, and W a subspace of V. In this section we define the *quotient space* V/W. Its vectors are the *cosets* $W + v$ for $v \in V$, where

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

These are just cosets of the additive subgroup W of the group $(V, +)$, as seen in 1st Year Group Theory. (They are *right* cosets, but the right coset $W + v$ is the same as the left coset $v + W$ because addition is commutative, so we just call them cosets.) It is of course possible to have $W + v = W + v'$ for different vectors v, v'; it is easy to tell when this happens:

$$W + v = W + v' \Leftrightarrow v - v' \in W.$$

You will have seen this fact in the 1st Year, but I have also set it as a question on Sheet 2 to make sure.

To make V/W into a vector space, we need to define addition and scalar multiplication of cosets. The natural definitions are:

(A) $(W + v_1) + (W + v_2) = W + v_1 + v_2$

(S) $\lambda(W + v) = W + \lambda v$

for all $v_i, v \in V$, $\lambda \in F$. We must check that these operations are well-defined. Here is the check for (A):

$$\begin{aligned} W + v_1 = W + v_1',\ W + v_2 = W + v_2' \quad &\Rightarrow v_1 - v_1',\ v_2 - v_2' \in W \\ &\Rightarrow v_1 + v_2 - (v_1' + v_2') \in W \\ &\Rightarrow W + v_1 + v_2 = W + v_1' + v_2'. \end{aligned}$$

And here is the check for (S):

$$\begin{aligned} W + v = W + v' \quad &\Rightarrow v - v' \in W \\ &\Rightarrow \lambda(v - v') \in W \\ &\Rightarrow \lambda v - \lambda v' \in W \\ &\Rightarrow W + \lambda v = W + \lambda v'. \end{aligned}$$

Proposition 5.1 *Let V/W be the set of cosets $W + v$ for $v \in V$. Then with addition and scalar multiplication defined by (A) and (S) as above, V/W is a vector space over F.*

Proof. We need to check the vector space axioms for V/W. These are:
Addition axioms: these amount to saying that $(V/W, +)$ is an abelian group, with identity element the zero vector $W + 0 = W$.
Scalar multiplication axioms – these are

(S1) $\lambda\left((W + v_1) + (W + v_2)\right) = \lambda(W + v_1) + \lambda(W + v_2)$

(S2) $(\lambda + \mu)(W + v) = \lambda(W + v) + \mu(W + v)$

(S3) $(\lambda(\mu)(W + v)) = (\lambda\mu)(W + v)$

(S4) $1(W + v) = W + v$.

Checking all the axioms is a routine exercise. I will just do (S1) and leave the rest to you to check:

$$\begin{aligned} \lambda\left((W + v_1) + (W + v_2)\right) \quad &= \lambda(W + v_1 + v_2) \\ &= W + \lambda(v_1 + v_2) \\ &= W + \lambda v_1 + \lambda v_2 \\ &= (W + \lambda v_1) + (W + \lambda v_2) \\ &= \lambda(W + v_1) + \lambda(W + v_2). \ \square \end{aligned}$$

We call the vector space V/W the *quotient space* of V by W. Its dimension is given by the next result.

Proposition 5.2 *Let V be finite-dimensional, and let W be a subspace of V. Then* $\dim V/W = \dim V - \dim W$.

Proof. Let $w_1, \ldots, w_r$ be a basis of W. Extend this to a basis of V:

$$w_1, \ldots, w_r,\ v_1, \ldots, v_s.$$

So $\dim W = r$ and $\dim V = r + s$.

Claim $W + v_1, \ldots, W + v_s$ is a basis of V/W.

Proof of Claim We first show the given set of vectors is linearly independent. Suppose

$$\sum_{i=1}^{s} \lambda_i (W + v_i) = W \quad \text{(the zero vector of } V/W).$$

Then LHS $= W + \sum \lambda_i v_i = W$, so $\sum \lambda_i v_i \in W$. Hence there exist scalars μ_j such that

$$\sum_{i=1}^{s} \lambda_i v_i = \sum_{j=1}^{r} \mu_j w_j.$$

As $w_1, \ldots, w_r,\ v_1, \ldots, v_s$ is a basis, this implies that $\lambda_i = 0$ for all i, proving that the set of vectors in the Claim is linearly independent.

Now we prove the set spans V/W. Let $W + v \in V/W$. There are scalars λ_i, μ_j such that

$$v = \sum_{j=1}^{r} \mu_j w_j + \sum_{i=1}^{s} \lambda_i v_i = w + \sum_{i=1}^{s} \lambda_i v_i,$$

where $w \in W$ is the first sum. Hence

$$W + v = W + \sum_{i=1}^{s} \lambda_i v_i = \sum_{i=1}^{s} \lambda_i (W + v_i).$$

This proves the spanning assertion, and so the Claim is proved.

By the Claim, we have

$$\dim V/W = s = \dim V - \dim W. \ \square$$

Example Let $V = \mathbb{R}^3$ and $W = \mathrm{Sp}(e_1 + e_2 + e_3)$. To find a basis of V/W, extend the basis $w = e_1 + e_2 + e_3$ of W to a basis of V – say w, e_1, e_2. Then by the Claim in the above proof, $W + e_1, W + e_2$ is a basis of V/W.

Quotient spaces and linear maps

Let $T : V \to V$ be a linear map. Suppose that W is a T-invariant subspace of V (recall this means that $T(W) \subseteq W$). Then we can define the restriction $T_W : W \to W$. We can also define a *quotient map* $\bar{T} : V/W \to V/W$ as follows:

$$\bar{T}(W + v) = W + T(v) \quad \forall v \in V.$$

We need to check that $\bar{T}$ is well-defined; here is the check:

$$\begin{aligned} W + v = W + v' \ &\Rightarrow v - v' \in W \\ &\Rightarrow T(v - v') \in W \text{ (since } T(W) \subseteq W) \\ &\Rightarrow T(v) - T(v') \in W \\ &\Rightarrow W + T(v) = W + T(v') \\ &\Rightarrow \bar{T}(W + v) = \bar{T}(W + v'). \end{aligned}$$

We now show that there is close relationship between the matrices of T, T_W and $\bar{T}$ with respect to certain bases. Choose of basis B_W of W:

$$B_W = \{w_1, \ldots, w_r\}.$$

Extend this to a basis B of V:

$$B = \{w_1, \ldots, w_r, v_1, \ldots v_s\}.$$

As in 5.2, we have a basis $\bar{B}$ of V/W:

$$\bar{B} = \{W + v_1, \ldots, W + v_s\}.$$

Proposition 5.3 *Let $X = [T_W]_{B_W}$ (an $r \times r$ matrix) and $Y = [\bar{T}]_{\bar{B}}$ (an $s \times s$ matrix). Then*

$$[T]_B = \begin{pmatrix} X & Z \\ 0 & Y \end{pmatrix},$$

where Z is $r \times s$.

Proof. Let

$$\begin{array}{l} T(w_i) = \sum_{j=1}^{r} x_{ji} w_j \quad (1 \le i \le r), \\ T(v_i) = \sum_{j=1}^{r} z_{ji} w_j + \sum_{j=1}^{s} y_{ji} v_j \quad (1 \le i \le s) \end{array}$$

Then

$$\begin{aligned} \bar{T}(W + v_i) &= W + \sum_{j=1}^{r} z_{ji} w_j + \sum_{j=1}^{s} y_{ji} v_j \\ &= W + \sum_{j=1}^{s} y_{ji} v_j \\ &= \sum_{j=1}^{s} y_{ji}(W + v_j). \end{aligned}$$

Hence $[T_W]_{B_W} = (x_{ij}) = X$, $[\bar{T}]_{\bar{B}} = (y_{ij}) = Y$ and

$$[T]_B = \begin{pmatrix} X & Z \\ 0 & Y \end{pmatrix},$$

where $Z = (z_{ij})$. $\square$

Example Let $V = \mathbb{R}^3$ and $T : V \to V$ be given by $T(v) = Av$ for all $v \in V$, where

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

Let $w = (1,1,1)^T$. Then $T(w) = 0$, so $W = \mathrm{Sp}(w)$ is a T-invariant subspace. We extend the basis $\{w\}$ of W to a basis $B = \{w, e_1, e_2\}$ of V, so we have a basis $\bar{B} = \{W + e_1, W + e_2\}$ of V/W. Check that

$$T(e_1) = (1, -2, 1)^T = w - 3e_2, \;\; T(e_2) = (-2, 0, 1)^T = w - 3e_1 - e_2.$$

Hence $\bar{T}(W + e_1) = W - 3e_2$, $\bar{T}(W + e_2) = W - 3e_1 - e_2$, and so

$$[\bar{T}]_{\bar{B}} = \begin{pmatrix} 0 & -3 \\ -3 & -1 \end{pmatrix}.$$

Finally,

$$[T]_B = \begin{pmatrix} 0 & 1 & 1 \\ 0 & 0 & -3 \\ 0 & -3 & -1 \end{pmatrix} = \begin{pmatrix} [T_W]_{B_W} & Z \\ 0 & [\bar{T}]_{\bar{B}} \end{pmatrix},$$

where $Z = (1, 1)$.

Corollary 5.4 *Let $T : V \to V$ be a linear map, and let W be a T-invariant subspace of V. Let $c(x)$, $c_1(x)$ and $c_2(x)$ be the characteristic polynomials of T, T_W and $\bar{T}$, respectively, Then $c(x) = c_1(x)c_2(x)$.*

Proof. In the notation of Prop. 5.3,

$$\begin{aligned} c(x) &= det \begin{pmatrix} xI_r - X & -Z \\ 0 & xI_s - Y \end{pmatrix} \\ &= det(xI_r - X)\, det(xI_s - Y) \\ &= c_1(x)c_2(x). \quad \square \end{aligned}$$

6 Triangularisation

Triangular matrices are not as easy to compute with as diagonal matrices, but they do have many nice properties. Here are a couple that will be familiar to you from 1st Year.

Proposition 6.1 *Let A and B be upper triangular $n \times n$ matrices:*

$$A = \begin{pmatrix} \lambda_1 & & & & \\ 0 & \lambda_2 & & * & \\ & & \cdot & & \\ & & & \cdot & \\ 0 & 0 & & & \lambda_n \end{pmatrix}, \quad B = \begin{pmatrix} \mu_1 & & & & \\ 0 & \mu_2 & & * & \\ & & \cdot & & \\ & & & \cdot & \\ 0 & 0 & & & \mu_n \end{pmatrix}.$$

(i) *The characteristic polynomial of A is $\prod_{i=1}^n (x - \lambda_i)$, the eigenvalues are $\lambda_1, \ldots, \lambda_n$ and the determinant is $\prod_{i=1}^n \lambda_i$.*

(ii) *The product AB is also upper triangular, with diagonal entries $\lambda_1\mu_1, \ldots, \lambda_n\mu_n$.*

So the characteristic polynomial of a triangular matrix is $\prod_1^n (x - \lambda_i)$, a product of linear factors. The triangularisation theorem shows that the converse is true:

Theorem 6.2 (Triangularisation Theorem) *Let V be an n-dimensional vector space over a field F and let $T : V \to V$ be a linear map. Suppose that characteristic polynomial $c(x)$ of T factorizes as a product of linear factors, so that $c(x) = \prod_1^n (x - \lambda_i)$ with all $\lambda_i \in F$. Then there is a basis B of V such that the matrix $[T]_B$ is upper triangular.*

We will prove this after making a few remarks on it. First we state the corresponding matrix version:

Corollary 6.3 *Let A be an $n \times n$ matrix over a field F, and suppose the characteristic polynomial of A factorizes as a product of linear factors. Then A is similar to an upper triangular matrix over F.*

Proof. Let $V = F^n$ and apply 6.2 to the linear map $T : V \to V$ given by $T(v) = Av$ for all $v \in V$. $\square$

Remarks (1) If $F = \mathbb{C}$ then by the Fundamental Theorem of Algebra, every polynomial over F factorizes as a product of linear factors. So Corollary 6.3 shows that every $n \times n$ matrix over $\mathbb{C}$ can be triangularised.

(2) For other fields this may not be the case; for example for $F = \mathbb{R}$, the matrix $\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ has characteristic polynomial $x^2 + 1$ which has no roots in $\mathbb{R}$, hence is not similar to a real triangular matrix.

Proof of Theorem 6.2. The proof goes by induction on $n = \dim V$. The result is obvious for $\dim V = 1$.

Now assume the result for vector spaces of dimension $n - 1$. Let $n = \dim V$, and $T : V \to V$ a linear map whose characteristic polynomial $c(x)$ factorizes as a product of linear factors. Then $c(x)$ has a root $\lambda \in F$. Let $w_1 \in V$ be a corresponding eigenvector with $T(w_1) = \lambda w_1$, and let $W = \mathrm{Sp}(w_1)$, a T-invariant subspace.

The quotient space V/W has dimension $n - 1$ by Prop. 5.2. Consider the quotient map $\bar{T} : V/W \to W/W$ (defined by $\bar{T}(W + v) = W + T(v)$ for $v \in V$). By Cor. 5.4, the characteristic polynomial of $\bar{T}$ divides $c(x)$, hence is also a product of linear factors. Hence by the induction assumption, V/W has a basis

$$\bar{B} = \{W + v_2, \ldots, W + v_n\}$$

such that the matrix $[\bar{T}]_{\bar{B}}$ is upper triangular. Let $Y = [\bar{T}]_{\bar{B}}$. Then $B = \{w_1, v_2, \ldots, v_n\}$ is a basis of V, and by Prop. 5.3,

$$[T]_B = \begin{pmatrix} \lambda & Z \\ 0 & Y \end{pmatrix}$$

(where Z is $1 \times n - 1$ and 0 is $n - 1 \times 1$). This matrix $[T]_B$ is upper triangular, so the induction proof is complete. □

The above proof gives an algorithm for triangularising a linear map $T : V \to V$ (assuming its characteristic polynomial factorizes):

(1) Find an eigenvector w_1 for T; let $W = \mathrm{Sp}(w_1)$.

(2) Find an eigenvector $W + w_2$ for $\bar{T} : V/W \to V/W$. Let $W' = \mathrm{Sp}(w_1, w_2)$.

(3) Find an eigenvector $W + w_3$ for $\bar{\bar{T}} : V/W' \to V/W'$.

(4) Continue, until we have a basis $B = \{w_1, w_2, w_3, \ldots, w_n\}$ of V. Then $[T]_B$ is upper triangular.

Here is a an example.

Example Let $V = \mathbb{R}^3$ and let $T : V \to V$ be defined by $T(v) = Av$ for all $v \in V$, where

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

Check that the characteristic polynomial of T is $(x - 1)^3$.

(1) We find an eigenvector $w_1 = (1, -1, 0)^T$. Let $W = \mathrm{Sp}(w_1)$.

(2) Extend w_1 to a basis $C = \{w_1, e_2, e_3\}$ of V. Then $\bar{C} = \{W + e_2, W + e_3\}$ is a basis of V/W. Compute that

$$[\bar{T}]_{\bar{C}} = \begin{pmatrix} 2 & 1 \\ -1 & 0 \end{pmatrix}.$$

This matrix has an eigenvector $(1,-1)^T$, which corresponds to an eigenvector $W+e_2-e_3$ of $\bar{T}$. So in the algorithm we can take $w_2 = e_2 - e_3$.

(3) Thus our final triangularising basis is $B = \{w_1, w_2, e_3\}$ (the third vector can be any vector that makes a basis with w_1, w_2): the matrix $[T]_B$ is upper triangular (with 1's on the diagonal, as 1 is the only eigenvalue of T). Also, if P is the matrix with columns w_1, w_2, e_3, then $P^{-1}AP$ is upper triangular.

7 The Cayley-Hamilton theorem

Recall that if $T : V \to V$ is a linear transformation and $p(x) = a_k x^k + \cdots + a_1 x + a_0$ is a polynomial, then $p(T) : V \to V$ is defined by

$$p(T) = a_k T^k + a_{k-1} T^k + \cdots + a_1 T + a_0 I_V.$$

Likewise if A is $n \times n$ matrix,

$$p(A) = a_k A^k + \cdots a_1 A + a_0 I.$$

In this chapter we prove one of the most fundamental results in the whole of linear algebra:

Theorem 7.1 (Cayley-Hamilton Theorem) *Let V be a finite-dimensional vector space over a field F, and let $T : V \to V$ be a linear map with characteristic polynomial $p(x)$. Then $p(T) = 0$.*

An immeadiate consequence is the corresponding statement for matrices:

Corollary 7.2 *If A is an $n \times n$ matrix over a field F with characteristic polynomial $p(x)$, then $p(A) = 0$.*

Remarks (1) Here is a "proof" of the corollary: by definition

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

Substitute $x = A$: this gives $p(A) = \det(AI - A) = 0!$

Is this a valid proof? No, of course not: the substitution $x = A$ makes no sense, as x is a scalar variable and A is a matrix.

(2) Note that Corollary 7.2 is obvious for diagonal matrices $A = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$: the characteristic polynomial of A is $\prod_{i=1}^n (x - \lambda_i)$, and $p(A) = \mathrm{diag}(p(\lambda_1), \ldots, p(\lambda_n)) = 0$.

(3) Proving Corollary 7.2 for upper triangular matrices is also not too difficult (set as a question on Problem Sheet 3). Combined with the Triangularisation Theorem 6.2, this gives a proof of the Cayley-Hamilton theorem for matrices over $\mathbb{C}$, but not for arbitrary fields.

(4) What about a direct proof of the Cayley-Hamilton theorem? Consider the 2×2 case: let $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. This has characteristic polynomial $p(x) = x^2 - (a+d)x + ad - bc$, so

$$p(A) = A^2 - \mathrm{tr}(A)\, A + \det(A)\, I.$$

We can verify by direct calculation that this is 0. But for $3 \times 3, \ldots, n \times n$ matrices, this is not a pleasant approach, and we need a better idea.

There are several different proofs of the Cayley-Hamilton theorem. I have chosen to present my favourite proof, which also has the merit of introducing some material that will be needed in later chapters.

Proof of Theorem 7.1

Let $T : V \to V$ be a linear map with characteristic polynomial $p(x)$. The proof proceeds by induction on $n = \dim V$. The result is trivial for $n = 1$. Now assume it is true for vector spaces of dimension at most $n - 1$.

(A) Assume first that there exists a T-invariant subspace W such that $W \neq 0$ or V. As in Proposition 5.3, choose a basis B_W of W, and extend it to a basis B of V such that

$$[T]_B = \begin{pmatrix} X & Z \\ 0 & Y \end{pmatrix},$$

where $X = [T_W]_{B_W}$, $Y = [\bar{T}]_{\bar{B}}$. By Corollary 5.4,

$$p(x) = p_X(x)p_Y(x),$$

where p_X, p_Y are the characteristic polynomials of X and Y. Now X is $r \times r$ and Y is $s \times s$, where $r = \dim W < n$, $s = \dim V/W < n$. Hence by the induction hypothesis,

$$p_X(X) = 0, \; p_Y(Y) = 0.$$

It follows that if we let $A = [T]_B = \begin{pmatrix} X & Z \\ 0 & Y \end{pmatrix}$, then

$$\begin{aligned} p(A) &= p_X(A)p_Y(A) \\ &= \begin{pmatrix} p_X(X) & Z_1 \\ 0 & p_X(Y) \end{pmatrix} \begin{pmatrix} p_Y(X) & Z_2 \\ 0 & p_Y(Y) \end{pmatrix} \\ &= \begin{pmatrix} 0 & Z_1 \\ 0 & p_X(Y) \end{pmatrix} \begin{pmatrix} p_Y(X) & Z_2 \\ 0 & 0 \end{pmatrix} \\ &= 0. \end{aligned}$$

(B) By (A), we can now assume that

$$V \text{ has no } T\text{-invariant subspaces apart from } 0 \text{ and } V. \tag{6}$$

Claim Let $0 \neq v \in V$, and let $B = \{v, T(v), \ldots, T^{n-1}(v)\}$. Then B is a basis of V.

Proof Since $\dim V = n$, it is enough to show that B is linearly independent. Let j be the largest integer such that the set

$$S = \{v, T(v), \ldots, T^{j-1}(v)\}$$

is linearly independent. Since $v \neq 0$ we have $j \geq 1$, and obviously $j \leq n$. Let $X = \mathrm{Sp}(S)$, so that $\dim X = j$.

By the choice of j, the set $\{v, T(v), \ldots, T^j(v)\}$ is linearly dependent. Hence $T^j(v) \in \mathrm{Sp}(S) = X$, and so X is T-invariant. Therefore by (6), we have $X = V$. Hence $j = n$, proving the Claim.

Now we work out the matrix $[T]_B$, where B is as in the Claim. Since $T^n(v) \in \mathrm{Sp}(B)$, we can write

$$T^n(v) = -a_0 v - a_1 T(v) - \cdots - a_{n-1} T^{n-1}(v) \tag{7}$$

for some scalars $a_i \in F$. Then

$$[T]_B = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 & -a_0 \\ 1 & 0 & 0 & \cdots & 0 & -a_1 \\ 0 & 1 & 0 & \cdots & 0 & -a_2 \\ & & & \cdots & & \\ 0 & 0 & 0 & \cdots & 1 & -a_{n-1} \end{pmatrix}. \tag{8}$$

By Q7 of Problem Sheet 1, the characteristic polynomial of this matrix is

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

Hence by (7),

$$p(T)(v) = T^n(v) + a_{n-1}T^{n-1}(v) + \cdots + a_0 v = 0.$$

This is true for any $v \in V$ (since the choice of v in the Claim was arbitrary). Hence $p(T) = 0$, and the proof is complete. □

Definition For $p(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_0 \in F[x]$, we call the $n \times n$ matrix in (8) the *companion matrix* of $p(x)$, denoted $C(p(x))$ (or just $C(p)$).

8 Polynomials

Let F be a field. A *polynomial* in x over F is an expression

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

where each $a_i \in F$. We denote the set of all polynomials over F by $F[x]$. Addition and multiplication are defined on $F[x]$ as follows: if $p(x) = \sum a_i x^i$, $q(x) = \sum b_j x^j$, then

$$p(x) + q(x) = \textstyle\sum (a_i + b_i)x^i,$$

$$p(x)\, q(x) = \textstyle\sum c_k x^k, \text{ where } c_k = \sum_{i+j=k} a_i b_j.$$

The *zero* polynomial is the one with all coefficients equal to 0, and is also denoted as 0. For $p(x) \neq 0$, the *degree* $\deg(p(x))$ is the highest power of x occuring in $p(x)$ with a nonzero coefficient. (The degree of the zero polynomial is undefined.) I leave it as an exercise for you to show that

$$\deg(p(x)q(x)) = \deg(p(x)) + \deg(q(x)).$$

We say that $p(x)$ *divides* $q(x)$ if there exists $r(x) \in F[x]$ such that $q(x) = p(x)r(x)$. Note that if $p(x)$ divides $q(x)$, then also $\lambda p(x)$ divides $q(x)$ for any scalar $\lambda \neq 0$, since $q(x) = (\lambda p(x))\,(\lambda^{-1}r(x))$. We write $p(x)|q(x)$ to denote that $p(x)$ divides $q(x)$. Finally, $p(x)$ is *monic* if its leading coefficient (that is, the coefficient of the highest power of x) is 1.

In what follows, we shall often write just f, g instead of $f(x), g(x)$, etc. for notational convenience. We aim to develop a theory of factorization of polynomials analogous to the theory of prime factorization of the integers. The main result is the Unique Factorization Theorem for polynomials, Theorem 8.7 below.

The theory starts with the following basic result.

Proposition 8.1 *(***Euclidean Algorithm***) Let $f, g \in F[x]$ with $\deg(g) \geq 1$. Then there exist polynomials $q, r \in F[x]$ such that*

$$f = qg + r,$$

where either $r = 0$ or $\deg(r) < \deg(g)$.

Proof The proof goes by induction on $n = \deg(f)$. The result is clear if $\deg(f) = 0$ (just take $q = 0$, $r = f$).

Now let $n = \deg(f)$, $m = \deg(g)$, and write

$$f = a_n x^n + \cdots + a_0, \quad g = b_m x^m + \cdots + b_0$$

(so that $a_n, b_m \neq 0$). If $n < m$, take $q = 0, r = f$ and the conclusion holds. So assume that $n \geq m$. Let

$$f_1 = f - a_n b_m^{-1} x^{n-m} g.$$

Then $\deg(f_1) < \deg(f) = n$, so by induction hypothesis, there are polynomials q_1, r_1 such that

$$f_1 = q_1 g + r_1$$

and either $r_1 = 0$ or $\deg(r_1) < \deg(g)$. Then

$$\begin{aligned} f &= f_1 + a_n b_m^{-1} x^{n-m} g \\ &= \left(q_1 + a_n b_m^{-1} x^{n-m}\right) g + r_1. \end{aligned}$$

Hence the result holds by induction. □

Definition Let $f, g \in F[x] \setminus \{0\}$. We say that $d \in F[x]$ is a *greatest common divisor* (gcd) of f, g if the following two conditions hold:

(1) $d | f$ and $d | g$,

(2) if $e(x) \in F[x]$ and $e | f$ and $e | g$, then $e | d$.

Note that if d is a gcd of f, g, then so is λd for any nonzero $\lambda \in F$. But apart from this, $\gcd(f, g)$ is unique, if it exists (Q on Sheet 3). In fact it *does* exist:

Proposition 8.2 *If $f, g \in F[x] \setminus \{0\}$, then $\gcd(f, g)$ exists, and is unique up to scalar multiplication.*

Proof We can assume that $\deg(f) \geq \deg(g)$, and repeatedly apply the Euclidean Algorithm 8.1:

$$\begin{aligned} &f = qg + r_1, \ \deg(r_1) < \deg(g), \\ &g = q_1 r_1 + r_2, \ \deg(r_2) < \deg(r_1), \\ &r_1 = q_2 r_2 + r_3, \ \deg(r_3) < \deg(r_2), \\ &\ldots \\ &r_{n-1} = q_n r_n + r_{n+1}, \ \deg(r_{n+1}) < \deg(r_n), \\ &r_n = q_{n+1} r_{n+1}. \end{aligned}$$

Then $r_{n+1} = \gcd(f, g)$. □

Definition We say that the polynomials $f, g \in F[x]$ are *coprime* if $\gcd(f, g) = 1$.

Proposition 8.3 *If $d = \gcd(f, g)$, then there exist $r, s \in F[x]$ such that $d = rf + sg$.*

Proof Referring to the previous proof, start with the equation $d = r_{n+1} = r_{n-1} - q_n r_n$. Substitute for r_n using the previous equation; then substitute for r_{n-1}, and so on. □

Factorization

First we define what are the "primes" in $F[x]$.

Definition A polynomial $p(x) \in F[x]$ is *irreducible* over F if $\deg(p) \geq 1$, and $p(x)$ cannot be factorized as a product of polynomials in $F[x]$ of smaller degree.

Note that there are always factorizations of the form $p(x) = (\lambda p(x))(\lambda^{-1})$ with $\lambda \in F \setminus \{0\}$. A polynomial that is not irreducible is called *reducible*.

Examples (1) The irreducibility of a polynomial depends on the field: for example $x^2 + 1$ is irreducible over $\mathbb{R}$, but not over $\mathbb{C}$ (since $x^2 + 1 = (x + i)(x - i)$).

(2) Every polynomial in $\mathbb{C}[x]$ of degree at least 1 has a root in $\mathbb{C}$, by the Fundamental Theorem of Algebra. So the only irreducible polynomials in $\mathbb{C}[x]$ are linear polynomials $ax + b$. The irreducibles in $\mathbb{R}[x]$ are linear polynomials, and also quadratic polynomials with no real roots (Q on Sheet 3).

(3) Here are the irreducibles of small degree in $\mathbb{F}_2[x]$ (where $\mathbb{F}_2 = \{0, 1\}$, the field of 2 elements):

degree 1: x, $x + 1$

degree 2: $x^2 + x + 1$ (this is irreducible as it has no roots in $\mathbb{F}_2$)

degree 3: $x^3 + x + 1$, $x^3 + x^2 + 1$ (these are irreducible as they have no roots in $\mathbb{F}_2$)

In Q on Sheet 3 you are asked to find all the irreducibles of degree 4.

Let me now briefly discuss irreducible polynomials in $\mathbb{Q}[x]$, an interesting and tricky topic. Given $p(x) \in \mathbb{Q}[x]$, it is usually hard to decide whether it is irreducible. The next result is a useful tool for monic polynomials that happen to have integer coefficients.

Proposition 8.4 *Let $p(x) \in \mathbb{Q}[x]$ be a monic polynomial with integer coefficients.*

(1) *If $\alpha \in \mathbb{Q}$ is a root of $p(x)$, then $\alpha \in \mathbb{Z}$.*

(2) *If $p(x)$ is reducible over $\mathbb{Q}$, then it has a factorization $p = ab$, where $a(x), b(x)$ are also monic with integer coefficients.*

Proof Part (1) is Q on Sheet 3. Part (2) is a famous result called *Gauss's Lemma.* We won't prove it here – if you are interested, you can find a proof in the recommended textbook by I N Herstein. □

Example We show that $x^3 + x + 1$ is irreducible over $\mathbb{Q}$. Suppose it is reducible: then it has a linear factor, hence has a root $\alpha \in \mathbb{Q}$. Then $\alpha \in \mathbb{Z}$ by Prop. 8.4(1), and α divides the constant term 1, hence $\alpha = \pm 1$. But 1 and -1 are not roots of $x^3 + x + 1$, contradiction.

Irreducible polynomials have several properties which are analogous to those of prime numbers. Here is one such basic property.

Proposition 8.5 *Let $p(x) \in F[x]$ be irreducible, and let $a(x), b(x) \in F[x]$. If $p|ab$, then either $p|a$ or $p|b$.*

Proof Suppose that $p|ab$ and also $p \not| a$. As p is irreducible, $\gcd(p, a) = 1$, and so by Proposition 8.3, there exist $r, s \in F[x]$ such that

$$1 = rp + sa.$$

Multiplying through by b, this gives $b = rpb + sab$. As p divides ab, it divides the RHS of this equation, hence it divides b. $\square$

Corollary 8.6 *If $p(x) \in F[x]$ is irreducible and $p|g_1 \cdots g_r$ (where each $g_i \in F[x]$), then $p|g_i$ for some i.*

Proof This is by induction on r, using Proposition 8.5. $\square$

Theorem 8.7 *(***Unique Factorization Theorem***) Let $f(x) \in F[x]$ with $\deg(f) \geq 1$.*

(1) *Then f factorizes as a product*

$$f = p_1 \cdots p_r,$$

where each $p_i \in F[x]$ is irreducible.

(2) *The factorization is unique (apart from multiplying factors by scalars).*

Proof (1) The proof is by induction on $\deg(f)$. The result is obvious if $\deg(f) = 1$.

Let $n = \deg(f)$, and assume the result holds for polynomials of degree less than n. If f is irreducible, the result holds, taking $p_1 = f$. And if f is reducible, then $f = ab$ where $a, b \in F[x]$ both have degree less than n. By induction hypothesis, a and b are products of irreducibles, hence so is f.

(2) Again we proceed by induction on $\deg(f)$. Suppose

$$f = p_1 \cdots p_r = q_1 \cdots q_s, \tag{9}$$

where all the polynomials p_i, q_i are irreducible. Then $p_1 | q_1 \cdots q_s$, so by Corollary 8.6, $p_1 | q_i$ for some i. Re-label the q's to take $i = 1$. Hence $q_1 = bp_1$ for some $b \in F[x]$, and as q_1 is irreducible, b is a scalar. Reaplce q_1 by $b^{-1}q_1$ (and q_2 by bq_2), so that $p_1 = q_1$. Now we can cancel these factors in (9), giving

$$p_2 \cdots p_r = q_2 \cdots q_s.$$

By the induction hypothesis, $r = s$ and (re-ordering the factors), $p_i = q_i$ for all $i \geq 2$, up to scalar multiplication of factors. Hence $p_i = q_i$ for all $i \geq 1$ (up to scalar mult.), completing the proof by induction. $\square$

To complete the section, we define the *least common multiple* $\operatorname{lcm}(f, g)$ of two polynomials $f, g \in F[x]$: this is a polynomial $h \in F[x]$ such that

(1) f and g both divide h, and

(2) if f and g both divide a polynomial $k \in F[x]$, then $h|k$.

Q of Sheet 3 shows that $\operatorname{lcm}(f, g)$ exists and is equal to $\frac{fg}{\gcd(f,g)}$. It can also be computed using the factorizations of f and g as products of irreducibles.

9 The minimal polynomial of a linear map

Let V be a vector space of dimension n over a field F, and $T@V \to V$ a linear map. We know that there are nonzero polynomials $f(x) \in F[x]$ such that $f(T) = 0$ – for example, $f(x) = c_T(x)$, the characteristic polynomial of T (by the Cayley-Hamilton theorem).

Definition We say that a polynomial $m(x) \in F[x]$ is a *minimal polynomial* for $T : V \to V$ if the following three conditions hold:

(1) $m(T) = 0$,

(2) $m(x)$ is monic,

(3) $\deg(m)$ is as small as possible such that (1) and (2) hold.

Our first result shows that the minimal polynomial of T is unique.

Proposition 9.1 *Let $T : V \to V$ be a linear map.*

(1) *T has a unique minimal polynomial: denote it as $m_T(x)$.*

(2) *For $p(x) \in F[x]$,*
$$p(T) = 0 \Leftrightarrow m_T(x)|p(x).$$

Proof (1) Suppose $m(x)$ and $m_1(x)$ satisfy conditions (1)-(3) of the definition. Then m and m_1 are monic of the same degree, so $\deg(m - m_1) < \deg(m)$ and $(m - m_1)(T) - m(T) - m_1(T) = 0$. Hence by the minimality of the degree, $m - m_1 = 0$ and so $m = m_1$.

(2) ($\Leftarrow$) For $p(x) \in F[x]$,
$$m_T(x)|p(x) \Rightarrow p(x) = m_T(x)q(x) \Rightarrow p(T) = m_T(T)q(T) = 0.$$

($\Rightarrow$) Suppose $p(x) \in F[x]$ and $p(T) = 0$. By the Euclidean Algorithm, there exist $q, r \in F[x]$ such that
$$p(x) = q(x)m_T(x) + r(x)$$
and either $r = 0$ or $\deg(r) < \deg(m_T)$. Then
$$0 = p(T) = q(T)m_T(T) + r(T) = r(T).$$
As $\deg(r) < \deg(m_T)$ this implies $r = 0$, hence $m_T|p$. $\square$

We adopt the same definition as above for the minimal polynomial $m_A(x)$ of an $n \times n$ matrix A. Note that if A and B are similar, they have the same minimal polynomial (Q on Sheet 4).

It will be important for us to be able to compute the minimal polynomial of a linear map or a matrix. The next result is useful for this.

Proposition 9.2 *Let $T : V \to V$ be a linear map.*

(1) *$m_T(x)$ divides $c_T(x)$, the characteristic polynomial of T.*

(2) *If $\lambda \in F$ is a root of $c_T(x)$ (i.e. an eigenvalue of T), then λ is also a root of $m_T(x)$.*

Proof (1) This follows from Proposition 9.1(2), since $c_T(T) = 0$ by Cayley-Hamilton.

(2) Let v be an eigenvector of T with $T(v) = \lambda v$. Then $0 = m_T(T)(v) = m_T(\lambda)(v)$. Hence $m_T(\lambda) = 0$. □

Examples (1) Let A be a diagonal matrix, with characteristic polynomial $\prod_{i=1}^r (x - \lambda_i)^{a_i}$, where $\lambda_1, \ldots, \lambda_r$ are the distinct diagonal entries with multiplicities $a_1, \ldots, a_r$. Then

$$m_A(x) = \prod_{i=1}^{r} (x - \lambda_i),$$

a product of distinct linear factors (Q on Sheet 4).

(2) Let us find the minimal polynomial of the matrix

$$A = \begin{pmatrix} 2 & 2 & -5 \\ 3 & 7 & -15 \\ 1 & 2 & -4 \end{pmatrix}.$$

We first compute the characteristic polynomial $c_A(x) = (x-1)^2(x-3)$. By Prop. 9.2, $m_A(x)$ divides this and has the same roots. Hence $m_A(x) = (x-1)(x-3)$ or $(x-1)^2(x-3)$. We compute the matrix $(A-I)(A-3I)$ and find that it is 0. Hence $m_A(x) = (x-1)(x-3)$.

(3) Recall that for $p(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_0 \in F[x]$, the companion matrix $C(p(x))$ is defined by

$$C(p(x)) = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 & -a_0 \\ 1 & 0 & 0 & \cdots & 0 & -a_1 \\ 0 & 1 & 0 & \cdots & 0 & -a_2 \\ & & & \cdots & & \\ 0 & 0 & 0 & \cdots & 1 & -a_{n-1} \end{pmatrix}.$$

Then

(a) this has characteristic polynomial $p(x)$ (Q7 on Sheet 1)

(b) it also has minimal polynomial $p(x)$ (Q on Sheet 4).

By Proposition 9.2, $m_T(x)$ and $c_T(x)$ have the same *linear* factors. What about other irreducible factors? The answer is the same:

Theorem 9.3 *Let $T : V \to V$ be a linear map. If $p(x) \in F[x]$ is an irreducible factor of $c_T(x)$, then $p(x)$ divides $m_T(x)$.*

For the proof we need to recall some facts from Sections 4 and 5 about *T-invariant subspaces* W (ie. subspaces W such that $T(W) \subseteq W$). There are two associated linear maps:

$T_W : W \to W$, the restriction of T to W

$\bar{T} : V/W \to V/W$, the quotient map $\bar{T}(W + v) = W + T(v)$ for $v \in V$.

Proposition 9.4 (1) *We have $c_T(x) = c_{T_W}(x)\, c_{\bar{T}}(x)$.*

(2) *The minimal polynomials $m_{T_W}(x)$ and $m_{\bar{T}}(x)$ both divide $m_T(x)$.*

Proof (1) is Corollary 5.4.

(2) For $w \in W$,

$$m_T(T_W)(w) = m_T(T)(w) = 0.$$

And for $v \in V$,

$$m_T(\bar{T})(W + v) = W + m_T(T)(v) = W + 0 = W.$$

Hence $m_T(T_W) = 0$ and $m_T(\bar{T}) = 0$, so m_{T_W} and $m_{\bar{T}}$ divide m_T by Prop. 9.1(2). $\square$

Proof of Theorem 9.3

Let $T : V \to V$ be a linear map, and let $p(x) \in F[x]$ be an irreducible factor of $c_T(x)$. We need to show that $p(x)$ divides $m_T(x)$. The proof proceeds by induction on $\dim V$; it is trivial for $\dim V = 1$. We follows a similar approach to the proof of the Cayley-Hamilton theorem 7.1.

(A) Assume first that there exists a T-invariant subspace W that is not equal to V or 0. Then by Prop. 9.4(1), $c_T(x) = c_{T_W}(x)\, c_{\bar{T}}(x)$. By Prop. 8.5, $p(x)$ divides either $c_{T_W}(x)$ or $c_{\bar{T}}(x)$. Since both W and V/W have dimension less than $\dim V$, the induction hypothesis therefore implies that $p(x)$ divides either $m_{T_W}(x)$ or $m_{\bar{T}}(x)$. Both of these divide $m_T(x)$ by Prop. 9.4(2), so $p(x)|m_T(x)$, as required.

(B) By (A), we may now assume that V has no T-invariant subspaces apart from 0 and V. Let $0 \neq v \in V$, and define

$$B = \{v, T(v), \ldots, T^{n-1}(v)\}.$$

By the proof of the Cayley-Hamilton theorem 7.1, B is a basis of V, and

$$[T]_B = C(c_T(x)),$$

the companion matrix of $c_T(x)$. The minimal polynomial of this matrix is also $c_T(x)$, by Q of Sheet 4, so $m_T(x) = c_T(x)$. Hence $p(x)$ divides $m_T(x)$, and the proof is complete. $\square$

Example Let A be the following 5×5 matrix over the field $\mathbb{F}_2 = \{0, 1\}$:

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

Find the minimal polynomial $m_A(x)$.

Answer First compute the characteristic polynomial $c_A(x)$ and factorize it as a product of irreducibles in $\mathbb{F}_2[x]$:

$$c_A(x) = (x^2 + x + 1)^2(x + 1).$$

Hence by Theorem 9.3, $m_A(x) = (x^2 + x + 1)^i(x + 1)$ with $i = 1$ or 2. Now compute that $(A^2 + A + I)(A + I) \neq 0$. Hence

$$m_A(x) = c_A(x) = (x^2 + x + 1)^2(x + 1).$$

10 Primary Decomposition

Recall from Section 1: we are aiming to prove "Canonical Form" theorems. These say that any $n \times n$ matrix A over a field F is similar to a block-diagonal matrix

$$M_1 \oplus \cdots \oplus M_r = \begin{pmatrix} M_1 & & & \\ & \cdot & & \\ & & \cdot & \\ & & & M_k \end{pmatrix}$$

where the M_i are "nice" matrices (Jordan blocks or companion matrices). To prove these theorems, we need methods for decomposing a vector space V as $V = V_1 \oplus \cdots \oplus V_r$, a direct sum of A-invariant subspaces. In this section we prove a fundamental such decomposition theorem.

Theorem 10.1 (**Primary Decomposition Theorem**) *Let V be a finite-dimensional vector space over a field F, and let $T : V \to V$ be a linear map with minimal polynomial $m_T(x)$. Let the factorization of $m_T(x)$ into irreducible polynomials be*

$$m_T(x) = \prod_{i=1}^{k} f_i(x)^{n_i},$$

where $f_1(x), \ldots, f_k(x)$ are distinct irreducible polynomials in $F[x]$. For $1 \leq i \leq k$, define

$$V_i = \ker\left(f_i(T)^{n_i}\right).$$

Then

(1) $V = V_1 \oplus \cdots \oplus V_k$,

(2) *each V_i is T-invariant,*

(3) *each restriction T_{V_i} has minimal polynomial $f_i(x)^{n_i}$.*

Definition We call the decomposition $V = V_1 \oplus \cdots \oplus V_k$ in Theorem 10.1 the *primary decomposition* of V with respect to T.

Before starting the proof of the theorem, we make some remarks on the important special case where every irreducible $f_i(x)$ is *linear*, say $f_i(x) = x - \lambda_i$ (eg. this will be the case if $F = \mathbb{C}$). In this case, the factorization is

$$m_T(x) = \prod_{i=1}^{k} (x - \lambda_i)^{n_i}$$

where $\lambda_1, \ldots, \lambda_k$ are the distinct eigenvalues of T, and

$$V_i = \ker(T - \lambda_i I)^{n_i}.$$

We call V_i the *generalized λ_i-eigenspace* of T.

Example Let $A = \begin{pmatrix} 2 & 0 & 0 \\ -1 & -3 & -1 \\ -1 & 4 & 1 \end{pmatrix}$ and let $T : V \to V$ be the linear map $T(v) = Av$, where $V = \mathbb{R}^3$. Let us compute the primary decomposition of V.

First find that $m_A(x) = c_A(x) = (x-2)(x+1)^2$. So in this case V_1 and V_2 are the generalized eigenspaces $\ker(A-2I)$ and $\ker(A+I)^2$.

Compute that $V_1 = \ker(A-2I) = \mathrm{Sp}(v_1)$, where $v_1 = (-1,0,1)^T$, and $V_2 = \ker(A+I)^2 = \mathrm{Sp}(e_2, e_3)$. So $V = V_1 \oplus V_2$ is the primary decomposition, and with respect to the basis $B = \{v_1, e_2, e_3\}$, we have

$$[T]_B = \begin{pmatrix} 2 & 0 & 0 \\ 0 & -3 & -1 \\ 0 & 4 & 1 \end{pmatrix}.$$

The diagonal blocks (2) and $\begin{pmatrix} -3 & -1 \\ 4 & 1 \end{pmatrix}$ are the matrices of the restrictions T_{V_1} and T_{V_2}.

Corollary 10.2 *A linear map $T : V \to V$ is diagonalisable if and only if $m_T(x) = \prod_{i=1}^k (x-\lambda_i)$, a product of distinct linear factors.*

Proof $(\Rightarrow)$ Suppose T is diagonalisable, and let B be a basis of V consisting of eigenvectors of T. Let $\lambda_1 \ldots, \lambda_k$ be the distinct eigenvalues of T, and let $f(x) = \prod_{i=1}^k (x-\lambda_i)$. Then $f(T) = \prod_1^k (T - \lambda_i I)$ maps each basis vector of B to 0, and hence $f(T) = 0$. Hence $m_T(x)$ divides $f(x)$, and so $m_T(x)$ is a product of distinct linear factors.

$(\Leftarrow)$ Suppose $m_T(x) = \prod_{i=1}^k (x-\lambda_i)$, a product of distinct linear factors. By Theorem 10.1, we have $V = V_1 \oplus \cdots \oplus V_k$, where each $V_i = \ker(T - \lambda_i I) = E_{\lambda_i}$, the λ_i-eigenspace of T. By Prop. 4.2, the union of bases of $V_1, \ldots, V_k$ is a basis of V, and it consists of eigenvectors of T. Hence T is diagonalisable. $\square$

We now begin working towards the proof of Theorem 10.1. This is based on the following result.

Proposition 10.3 *Let $T : V \to V$ be a linear map, and suppose $g_1(x), g_2(x) \in F[x]$ are coprime polynomials such that $g_1(T)g_2(T) = 0$.*

(1) *Then $V = V_1 \oplus V_2$, where $V_i = \ker g_i(T)$ for $i = 1, 2$; also each V_i is T-invariant.*

(2) *Suppose also that $m_T(x) = g_1(x)g_2(x)$. Then $m_{T_{V_i}}(x) = g_i(x)$ for $i = 1, 2$.*

Proof (1) As $g_1(x), g_2(x)$ are coprime, there exist $s_1(x), s_2(x) \in F[x]$ such that

$$s_1(x)g_1(x) + s_2(x)g_2(x) = 1.$$

Then

$$s_1(T)g_1(T) + s_2(T)g_2(T) = I_V.$$

Let $v \in V$. Then

$$v = I_V(v) = s_1(T)g_1(T)(v) + s_2(T)g_2(T)(v).$$

So $v = v_1 + v_2$, where $v_i = s_i(T)g_i(T)(v)$ for $i = 1, 2$. Since $g_1(T)g_2(T) = 0$, we see that $v_1 \in \ker g_2(T) = V_2$ and $v_2 \in \ker g_1(T) = V_1$. Hence

$$V = V_1 + V_2.$$

Also

$$v \in V_1 \cap V_2 \Rightarrow v = s_1(T)g_1(T)(v) + s_2(T)g_2(T)(v) = 0,$$

and so $V_1 \cap V_2 = \{0\}$. Therefore $V = V_1 \oplus V_2$ by Prop. 4.1. Finally, each V_i is T-invariant since

$$v \in V_i \Rightarrow g_i(T)(v) = 0 \Rightarrow g_i(T)\,T(v) = T\,g_i(T)(v) = 0 \Rightarrow T(v) \in \ker g_i(T) = V_i.$$

(2) Let $m_i(x) = m_{T_{V_i}}(x)$ for $i = 1, 2$. As $V_i = \ker g_i(T)$, we have $g_i(T_{V_i}) = 0$, so $m_i(x)$ divides $g_i(x)$ by Prop. 9.1(2). As g_1, g_2 are coprime, so are m_1, m_2. Therefore by Q on Sheet 4,

$$m_T(x) = \mathrm{lcm}(m_1(x), m_2(x)) = m_1(x)m_2(x).$$

Since by the hypothesis of (2) we have $m_T(x) = g_1(x)g_2(x)$, it follows that $m_i(x) = g_i(x)$ for $i = 1, 2$. $\square$

Proof of Theorem 10.1

Let $T : V \to V$ be a linear map with $m_T(x) = \prod_{i=1}^{k} f_i(x)^{n_i}$, where $f_1(x), \ldots, f_k(x)$ are distinct irreducible polynomials in $F[x]$. The proof proceeds by induction on k. It is trivial for $k = 1$, so assume $k \geq 2$.

In Proposition 10.3, take

$$g_1(x) = f_1(x)^{n_1}, \quad g_2(x) = \prod_{i=2}^{k} f_i(x)^{n_i}.$$

These are coprime, so by 10.3, we have

$$V = V_1 \oplus W$$

where $V_1 = \ker g_1(T)$, $W = \ker g_2(T)$, and also

minimal poly. of T_{V_1} is $g_1(x) = f_1(x)^{n_1}$,

minimal poly. of T_W is $g_2(x) = \prod_{i=2}^{k} f_i(x)^{n_i}$.

Applying the induction hypothesis to the restriction $T_W : W \to W$, we obtain

$$W = V_2 \oplus \cdots \oplus V_k,$$

where for $i = 2, \ldots, k$ we have $V_i = \ker f_i(T_W)^{n_i}$ and also the minimal poly. of $(T_W)_{V_i}$ is $f_i(x)^{n_i}$. Note that $\ker f_i(T_W)^{n_i} = \ker f_i(T)^{n_i}$, since the RHS of this equation is contained in $\ker g_2(T) = W$. Also $(T_W)_{V_i} = T_{V_i}$. Hence we have shown that the following conditions hold:

- $V = V_1 \oplus W = V_1 \oplus V_2 \oplus \cdots \oplus V_k$,
- each $V_i = \ker f_i(T)^{n_i}$,
- each T_{V_i} has minimal poly. $f_i(x)^{n_i}$.

These are the conditions specified in the conclusion of the theorem , so this completes the proof by induction. $\square$

11 Jordan Canonical Form

In this chapter we prove the first of the canonical form theorems mentioned in the introductory chapter 1. This is the Jordan Canonical Form theorem, one of the main results in the whole of linear algebra.

Definition Let F be a field and let $\lambda \in F$. Define the $n \times n$ matrix

$$J_n(\lambda) = \begin{pmatrix} \lambda & 1 & 0 & \dots & 0 & 0 \\ 0 & \lambda & 1 & \dots & 0 & 0 \\ 0 & 0 & \lambda & \dots & 0 & 0 \\ & & & \dots & & \\ 0 & 0 & 0 & \dots & \lambda & 1 \\ 0 & 0 & 0 & \dots & 0 & \lambda \end{pmatrix}$$

Such a matrix is called a *Jordan block*.

For example

$$J_2(5) = \begin{pmatrix} 5 & 1 \\ 0 & 5 \end{pmatrix}, \; J_3(0) = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \; J_1(\lambda) = (\lambda).$$

Here are some basic properties of Jordan blocks.

Proposition 11.1 *Let $J = J_n(\lambda)$.*

(1) *Both the characteristic and the minimal polynomials of J are equal to $(x - \lambda)^n$.*

(2) *λ is the only eigenvalue of J: its algebraic multiplicity is n and its geometric multiplicity is 1.*

(3) *$J - \lambda I = J_n(0)$, and multiplication by $J - \lambda I$ sends the standard basis vectors*

$$e_n \to e_{n-1} \to \cdots \to e_2 \to e_1 \to 0.$$

(4) *$(J - \lambda I)^n = 0$, and for $i < n$, $(J - \lambda I)^i$ has rank $n - i$ and sends $e_n \to e_{n-i}$, $e_{n-1} \to e_{n-i-1}$ and so on.*

Proof (1) As J is upper triangular, the characteristic polynomial $c_J(x) = (x - \lambda)^n$. Hence $m_J(x) = (x - \lambda)^i$ for some $i \leq n$. As $(J - \lambda I)^{n-1} \neq 0$ by part (4), $m_J(x)$ must be $(x - \lambda)^n$.

(2) The eigenspace $E_\lambda(J)$ is the solution space of $(J - \lambda I)v = 0$, which is $\mathrm{Sp}(e_1)$, of dimension 1. Hence the geometric multiplicity $g(\lambda) = 1$.

Finally, (3) is clear, and it follows that $(J - \lambda I)^i = J_n(0)^i$ sends $e_n \to e_{n-i}$, $e_{n-1} \to e_{n-i-1}$ and so on, giving (4), $\square$

Recall the definition of a *block diagonal* matrix: if $A_1, \dots, A_k$ are square matrices, where A_i is $n_i \times n_i$, define

$$A_1 \oplus A_2 \oplus \cdots \oplus A_k = \begin{pmatrix} A_1 & 0 & \dots & 0 \\ 0 & A_2 & \dots & 0 \\ & & \dots & \\ 0 & 0 & \dots & A_k \end{pmatrix}$$

This is $n \times n$, where $n = \sum n_i$.

Proposition 11.2 *Let $A = A_1 \oplus \cdots \oplus A_k$ and for each i let A_i have characteristic polynomial $c_i(x)$ and minimal polynomial $m_i(x)$.*

(1) *The characteristic polynomial $c_A(x) = \prod_1^k c_i(x)$.*

(2) *The minimal polynomial $m_A(x) = \operatorname{lcm}(m_1(x), \dots, m_k(x))$.*

(3) *For any eigenvalue λ of A, $\dim E_\lambda(A) = \sum_1^k \dim E_\lambda(A_i)$.*

(4) *For any polynomial $q(x)$, we have $q(A) = q(A_1) \oplus \cdots \oplus q(A_k)$.*

Proof Parts (1) and (4) are clear; part (3) is in Q5 of Sheet 2; and part (2) is Q on Sheet 4. □

Here is the great theorem.

Theorem 11.3 (Jordan Canonical Form) *Let A be an $n \times n$ matrix over a field F, and suppose the characteristic polynomial of A is a product of linear factors over F. Then*

(1) *A is similar to a matrix of the form of the form*

$$J = J_{n_1}(\lambda_1) \oplus J_{n_2}(\lambda_2) \oplus \cdots \oplus J_{n_k}(\lambda_k) \tag{10}$$

where $\sum n_i = n$. (Note that the eiegnvalues λ_i are not necessarily distinct.)

(2) *The matrix J in (10) is uniquely determined by A, apart from changing the order in which the Jordan blocks appear.*

Definition We call the block-diagonal matrix J in (10) the *Jordan Canonical Form* (JCF) of A.

There is of course an equivalent statement of Theorem 11.3 for linear maps $T : V \to V$, where V is an n-dimensional vector space over F. This states that if $c_T(x)$ is a product of linear factors, then there is a basis B of V such that $[T]_B = J$, a unique JCF matrix.

Note that the condition on $c_A(x)$ in the hypothesis of the theorem says that all the eigenvalues of A lie in F. This condition is obviously necessary for the conclusion to hold (as it was for the Triangularisation Theorem). It always holds when the field $F = \mathbb{C}$, by the Fundamental Theorem of Algebra.

Example Here are a few examples of JCFs:

$$A = J_2(1) \oplus J_2(1) = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

$$B = J_3(1) \oplus J_1(1) = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

$$C = J_1(1) \oplus J_3(1) = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

The uniqueness part (2) of Theorem 11.3 implies A is not similar to B or C. But note that B is similar to C (see Q5 on Sheet 2).

Notice that the only diagonal JCF matrices are of the form $J_1(\lambda_1) \oplus \cdots \oplus J_1(\lambda_k)$ – so in some sense "most" matrices are not diagonalisable.

How to compute the JCF of a matrix

We shall prove the JCF Theorem 11.3 later. First we make some remarks on how to compute the JCF of any given matrix. Let A be an $n \times n$ matrix such that $c_A(x)$ is a product of linear factors. The JCF Theorem tells us that $A \sim J$, a JCF matrix as in (10) (where as usual we use $\sim$ to denote similarity of matrices). How can we compute J?

First note that A and J have the same characteristic polynomial, minimal polynomial, eigenvalues and geometric multiplicities, and that $q(A) \sim q(J)$ for any polynomial $q(x)$. For each eigenvalue λ, collect up all the Jordan blocks with evalue λ, and change the order of the blocks to re-write

$$J = (J_{n_1}(\lambda) \oplus \cdots \oplus J_{n_a}(\lambda)) \oplus (J_{m_1}(\mu) \oplus \cdots \oplus J_{m_b}(\mu)) \oplus \cdots$$

We call the first bracket the λ-blocks of J, then the μ-blocks, and so on.

Proposition 11.4 *Let J be as above, and λ an eigenvalue.*

(1) $n_1 + \cdots + n_a = a(\lambda)$*, the algebraic multiplicity of* λ.

(2) $a =$*number of* λ*-blocks* $= g(\lambda)$*, the geometric multiplicity of* λ.

(3) $\max(n_1, \ldots, n_a) = r$*, where* $(x-\lambda)^r$ *is the highest power of* $x - \lambda$ *dividing* $m_A(x)$*, the minimal polynomial of* A.

Proof (1) The power of $x - \lambda$ dividing the characteristic polynomial of J is $\prod_{i=1}^{a}(x-\lambda)^{n_i}$, so $a(\lambda) = \sum_1^a n_i$.

(2) Each λ-block has geometric multiplicity 1 by Prop. 11.1(2), so by Prop. 11.2(3), we have $a = g(\lambda)$.

(3) By Prop. 11.1(1), the minimal polynomial of $J_{n_i}(\lambda)$ is $(x-\lambda)^{n_i}$. Hence by Prop. 11.2(2), the power of $x - \lambda$ dividing $m_J(x)$ is $\operatorname{lcm}((x-\lambda)^{n_1}, \ldots, (x-\lambda)^{n_a})$, which is equal to $(x-\lambda)^{\max(n_1,\ldots,n_a)}$. $\square$

So computing the multiplicities $a(\lambda), g(\lambda)$ and the minimal polynomial of A gives a lot of information about the JCF of A. Often – but not always – this is enough to determine the JCF. Here are some examples.

Examples (1) Find the JCF of

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

Answer The characteristic poly $c_A(x) = (x+1)^2(x-1)^3$, so the eigenvalues are $-1, 1$ with $a(-1) = 2$, $a(1) = 3$. Calculate that $\operatorname{rank}(A + I) = 4$ and $\operatorname{rank}(A - I) = 3$, so

$g(-1) = 1$ and $g(1) = 2$. This means that the JCF of A has one -1-block and two 1-blocks, which is already enough to determine it uniquely as

$$J_2(-1) \oplus J_2(1) \oplus J_1(1).$$

(2) Find the JCF of

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

Answer Here $c_A(x) = (x-1)^4$, and $\text{rank}(A - I) = 2$, so $g(1) = 2$ and so the number of 1-blocks is 2. Hence the JCF is either $J_2(1) \oplus J_2(1)$ or $J_3(1) \oplus J_1(1)$. Which ?

To determine which, we need to compute $m_A(x)$: check that $(A - I)^2 = 0$, so $m_A(x) = (x-1)^2$. Hence by Prop. 11.4(3), the largest block has size 2, so the JCF of A is $J_2(1) \oplus J_2(1)$.

(3) Suppose we are given the following information about a matrix A:

$$c_A(x) = x^7,\ m_A(x) = x^3,\ g(0) = 3. \tag{11}$$

Can we compute the JCF of A?

Well, by Prop. 11.4 we know that there are 3 blocks of sizes adding up to 7, and the maximum size is 3. There are two JCFs satisying these conditions:

$$J = J_3(0) \oplus J_3(0) \oplus J_1(0), \text{ and } J' = J_3(0) \oplus J_2(0) \oplus J_2(0).$$

So the information in (11) is not sufficent to determine the JCF.

What further information about A is needed to determine the JCF? Well, (11) determines the ranks of A and A^3: we have $\text{rank}(A) = 7 - g(0) = 4$, and $\text{rank}(A^3) = 0$ since $m_A(x) = x^3$. If we are given also $\text{rank}(A^2)$, we can determine the JCF, since $\text{rank}(J^2) = 2$, whereas $\text{rank}(J'^2) = 1$.

A completely general method for computing the JCF of a given matrix will be provided by the proof of uniqueness part (2) of the JCF Theorem, coming up right now....

Uniqueness of JCF

Here we prove the uniqueness part (2) of the JCF Theorem .11.3:

Theorem 11.5 *Suppose A is an $n \times n$ matrix over a field F, and A is similar to a JCF matrix J, where*

$$J = J_{n_1}(\lambda_1) \oplus J_{n_2}(\lambda_2) \oplus \cdots \oplus J_{n_k}(\lambda_k).$$

Then J is uniquely determined by A, apart from changing the order in which the Jordan blocks appear.

Proof (A) First we handle the case where A has only *one* eigenvalue λ – so all $\lambda_i = \lambda$ and $c_A(x) = (x-\lambda)^n$. Re-order the blocks to take

$$J = J_1(\lambda)^{a_1} \oplus J_2(\lambda)^{a_2} \oplus J_r(\lambda)^{a_r},$$

where all $a_i \geq 0$ (some can be 0) – meaning that J has a_1 blocks of size 1, a_2 blocks of size 2, and so on. For $i \geq 1$, define

$$m_i = \text{ rank}\,(A - \lambda I)^i = \text{ rank}\,(J - \lambda I)^i.$$

We shall show that the a_i's can be expressed in terms of the m_i's. Observe that

$$J - \lambda I = J_1(0)^{a_1} \oplus J_2(0)^{a_2} \oplus J_r(0)^{a_r}.$$

Hence using Prop. 11.1(4), we see that

$$\begin{aligned}
&m_r = \text{ rank}\,(J - \lambda I)^r = 0, \\
&m_{r-1} = \text{ rank}\,(J - \lambda I)^{r-1} = a_r, \\
&m_{r-2} = \text{ rank}\,(J - \lambda I)^{r-2} = 2a_r + a_{r-1}, \\
&m_{r-3} = \text{ rank}\,(J - \lambda I)^{r-3} = 3a_r + 2a_{r-1} + a_{r-2}, \\
&\vdots \\
&m_2 = (r-2)a_r + \cdots + 2a_4 + a_3, \\
&m_1 = (r-1)a_r + \cdots + 2a_3 + a_2.
\end{aligned}$$

Hence, given the m_i's, we can determine the a_i's uniquely.

This proves the uniqueness statement of the theorem for the case of one eigenvalue.

(B) Now we handle the general case. We are given that $A \sim J$, a JCF matrix. Let λ be an eigenvalue of A, let J_λ be the block-diagonal sum of all the λ-blocks in J, and let L be the sum of the other blocks of J. So re-ordering the blocks, we have

$$J = J_\lambda \oplus L,$$

where λ is not an eigenvalue of L. So $L - \lambda I$ is invertible and $\text{rank}\,(L - \lambda I)^i = l$ for all $i \geq 1$, where L is $l \times l$.

For $i \geq 1$, define

$$r_i = \text{ rank}\,(A - \lambda I)^i = \text{ rank}\,(J - \lambda I)^i.$$

Then

$$r_i = \text{ rank}\,(J_\lambda - \lambda I)^i + l,$$

and so for each i we can compute

$$m_i = \text{ rank}\,(J_\lambda - \lambda I)^i = r_i - l.$$

Hence, as in (A), we can determine uniquely the sizes of all the λ-blocks in J_λ. Now repeat this for all the other eigenvalues of A, and the proof is complete. □

Existence of JCF

Now we prove part (1) of the JCF Theorem 11.3. It is convenient to prove it for linear maps rather than matrices. Here is the statement.

Theorem 11.6 *Let $T : V \to V$ be a linear map, and suppose that $c_T(x)$ is a product of linear factors. Then there exists a basis B of V such that $[T]_B$ is a JCF matrix.*

First we shall reduce the proof of this theorem to the case where T has only one eigenvalue.

Let $T : V \to V$ be as in the theorem, and let

$$c_T(x) = \prod_{i=1}^{k}(x-\lambda_i)^{a_i}, \;\; m_T(x) = \prod_{i=1}^{k}(x-\lambda_i)^{n_i},$$

where $\lambda_1 \ldots, \lambda_k$ are the distinct eigenvalues of T, and $a_i \geq n_i \geq 1$. We apply the Primary Decomposition Theorem 10.1. If we define $V_i = \ker(T-\lambda_i)^{n_i}$ for $1 \leq i \leq k$, this tells us that

$$V = V_1 \oplus \cdots \oplus V_k.$$

Let B_i be a basis of V_i. Then $B = B_1 \cup \cdots \cup B_k$ is a basis of V by Prop. 4.2. Let $A_i = [T_{V_i}]_{B_i}$. Then by Prop. 4.3,

$$[T]_B = A_1 \oplus \cdots \oplus A_k,$$

and by Theorem 10.1(3), each A_i has minimal polynomial $(x-\lambda_i)^{n_i}$. Hence if we prove Theorem 11.6 for each restriction T_{V_i}, the theorem will follow in general.

We have now shown that it is enough to establish Theorem 11.6 for the *case where T has only one eigenvalue.*

The case of one eigenvalue

Let $\dim V = n$ and let $T : V \to V$ be a linear map with $c_T(x) = (x-\lambda)^n$, so that T has only one eigenvalue λ. Define $S = T - \lambda I_V$. Then

$$S^n = (T - \lambda I_V)^n = 0,$$

so S has only one eigenvalue 0. Such a linear map is said to be *nilpotent*.

Here is the JCF Theorem 11.6 for S:

Theorem 11.7 *Let $S : V \to V$ be a nilpotent linear map. Then there exists a basis B of V such that*

$$[S]_B = J_{n_1}(0) \oplus \cdots \oplus J_{n_k}(0).$$

Corollary 11.8 *Then $T = S + \lambda I_V$ has $[T]_B = J_{n_1}(\lambda) \oplus \cdots \oplus J_{n_k}(\lambda)$. In other words, Theorem* 11.6 *holds for any linear map T having only one eigenvalue.*

So to complete the proof of Theorem 11.6 it remains to prove Theorem 11.7.

Proof of Theorem 11.7

Let $n = \dim V$ and $S : V \to V$ with S nilpotent. We are aiming to find a basis B such that $[S]_B = J_{n_1}(0) \oplus \cdots$. So if $v_{n_1}, \ldots, v_1$ are the first n_1 vectors of B in that order, we require

$$S(v_1) = v_2,\; S(v_2) = v_3, \ldots, S(v_{n_1}) = 0.$$

In other words, the first n_1 vectors of B should be (in reverse order):

$$v_1,\; S(v_1), \ldots, S^{n_1-1}(v_1),$$

where $S^{n_1}(v_1) = 0$. Thus we are looking for a basis B of V of the form

$$v_1, S(v_1), \ldots, S^{n_1-1}(v_1), \ldots, v_k, S(v_k), \ldots, S^{n_k-1}(v_k), \tag{12}$$

where $S^{n_i}(v_i) = 0$ for $i = 1, \ldots, k$. Then (after reversing each subsequence $v_i, \ldots, S^{n_i-1}(v_i)$ in B), the matrix $[S]_B$ will be the JCF matrix in the conclusion of the theorem. We call such a basis a *Jordan basis* of V.

We prove by induction on $n = \dim V$ that a Jordan basis of V exists. It is obvious for $n = 1$. Now assume it is true for vector spaces of dimension less than n.

Consider $\mathrm{Im}(S) = S(V) \subseteq V$. As 0 is an eigenvalue of S, we have $\ker(S) \neq 0$, and so by the Rank-Nullity theorem, $S(V) \neq V$. Hence

$$\dim S(V) < n.$$

Let $W = S(V)$. Then W is S-invariant of dimension less than n, and the restriction $S_W : W \to W$ is clearly nilpotent, so we can apply the induction hypothesis to S_W. This implies that there is a Jordan basis of W; write it as

$$u_1, S(u_1), \ldots, S^{m_1-1}(u_1), \ldots, u_r, S(u_r), \ldots, S^{m_r-1}(u_r), \tag{13}$$

where $S^{m_i}(u_i) = 0$ for $i = 1, \ldots, r$ and $\sum_1^r m_i = \dim W$.

Now add vectors to the list in (13) as follows:

(1) for each i, add a vector $v_i \in V$ such that $u_i = S(v_i)$;

(2) note that $\ker(S)$ contains the linearly independent vectors

$$S_{m_1-1}(u_1), \ldots S^{m_r-1}(u_r);$$

extend these to a basis of $\ker(S)$ by adding further vectors $w_1, \ldots, w_s$ (so $\dim\ker(S) = r + s$).

With these additions to (13), we now have a list of vectors

$$v_1, S(v_1), \ldots, S^{m_1}(v_1), \ldots, v_r, S(v_r), \ldots, S^{m_r}(v_r),\ w_1, \ldots, w_s. \tag{14}$$

Claim The list (14) is a basis of V.

Proof We first prove the list is linearly independent. Suppose we have a linear relation

$$\alpha_1 v_1 + \cdots + \alpha_{m_1+1} S^{m_1}(v_1) + \cdots + \gamma_1 v_r + \cdots + \gamma_{m_r+1} S^{m_r}(v_r) + \sum_{i=1}^{s} \delta_i w_i = 0. \tag{15}$$

Apply S to this, noting that $S^{m_i+1}(v_i) = S^{m_i}(u_i) = 0$ and also $S(w_i) = 0$. This gives

$$\alpha_1 S(v_1) + \cdots + \alpha_{m_1} S^{m_1}(v_1) + \cdots + \gamma_1 S(v_r) + \cdots + \gamma_{m_r} S^{m_r}(v_r) = 0.$$

This is a linear relation on the basis (13) of W. Hence

$$\alpha_1 = \cdots = \alpha_{m_1} = \cdots = \gamma_1 = \cdots = \gamma_{m_r} = 0.$$

Hence (15) is now

$$\alpha_{m_1+1} S^{m_1}(v_1) + \cdots + \gamma_{m_r+1} S^{m_r}(v_r) + \sum_1^s \delta_i w_i = 0.$$

This is a linear relation on our basis of $\ker(S)$, so all the coefficients are 0:

$$\alpha_{m_1+1} = \cdots = \gamma_{m_r+1} = \delta_i = 0 \ \ \forall i.$$

Hence all the coefficients in the linear relation (15) are 0, proving the linear independence of (14).

To complete the proof of the Claim, we count the number of vectors in the list (14): the number is

$$\begin{aligned} (m_1+1)+\cdots+(m_r+1)+s &= (\textstyle\sum_1^r m_i) + r + s \\ &= \dim W + r + s \\ &= \dim \mathrm{Im}(S) + \dim \ker(S) \\ &= \dim V. \end{aligned}$$

Hence the list (14) consists of $n = \dim V$ linearly independent vectors, so it is a basis of V, proving the Claim.

Now let B be the basis (14) of V, and reverse each of the subsequences $v_i, \ldots, S^{m_i}(v_i)$. Then

$$[S]_B = J_{m_1+1}(0) \oplus \cdots \oplus J_{m_r+1}(0) \oplus J_1(0)^s.$$

Hence B is a Jordan basis of V, and the proof of the theorem by induction is complete. $\square$

Computing a Jordan basis

Given any linear map $T : V \to V$ with $c_T(x)$ a product of linear factors, a basis B for which $[T]_B$ is a JCF matrix is called a *Jordan basis* of V. We can use the method of the inductive proof of Theorem 11.7 to give an algorithm for computing a Jordan basis for any such linear map T. As we argued before, the Primary Decomposition Theorem gives

$$V = V_1 \oplus \cdots \oplus V_k,$$

where each restriction T_{V_i} has only one eigenvalue λ_i. And if we let $S_i = T_{V_i} - \lambda_i I$, then S_i is nilpotent. Hence we just need to find a Jordan basis for *nilpotent* linear maps $S : V \to V$. Here is the algorithm for this.

Let $S : V \to V$ be nilpotent.

Step 1 Compute the subspaces

$$V \supset S(V) \supset S^2(V) \supset \cdots \supset S^r(V) \supset 0,$$

where $S^{r+1}(V) = 0$.

Step 2 Find a basis of $S^r(V)$. Then use rules (1) and (2) in the proof of Theorem 11.7 to add vectors to get a Jordan basis of $S^{r-1}(V)$.

Step 3 Repeat: successively find Jordan bases of $S^{r-2}(V), \ldots, S(V), V$.

Example Here is an example carrying out this algorithm. Let F be a field, let $V = F^5$ and let $S : V \to V$ be defined by $S(v) = Av$ for all $v \in V$, where

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

Find a Jordan basis for this map.

Answer Observe that $c_A(x) = x^5$, $m_A(x) = x^4$ and the geometric multiplicity $g(0) = 2$. Hence the JCF of A is $J_4(0) \oplus J_1(0)$.

To find a Jordan basis, we use the algorithm.

First check that

$$\begin{aligned} S(V) &= \mathrm{Sp}(e_1, e_2, e_3 + e_4), \\ S^2(V) &= \mathrm{Sp}(e_1, e_2), \\ S^3(V) &= \mathrm{Sp}(e_1), \\ S^4(V) &= 0. \end{aligned}$$

We now perform Steps 2 and 3. A basis of $S^3(V)$ is e_1.

Add vectors to get a Jordan basis of $S^2(V)$: e_2, e_1 (since $S(e_2) = e_1$).

To get a Jordan basis of $S(V)$: add $v_1 \in S(V)$ such that $S(v_1) = e_2$: take $v_1 = \frac{1}{2}(-e_2 + e_3 + e_4)$. So Jordan basis of $S(V)$ is

$$v_1, e_2, e_1.$$

Finally get a Jordan basis of V:

1) add $x_1 \in V$ such that $S(x_1) = v_1$: take $x_1 = \frac{1}{2}(-e_3 + e_5)$;

2) add w_1 such that e_1, w_1 is a basis of $\ker(S)$: take $w_1 = e_2 + e_3 - e_4$.

So our final Jordan basis of V is

$$x_1, v_1, e_2, e_1, w_1.$$

To get the ordered basis B such that $[S]_B$ is the JCF matrix $J_4(0) \oplus J_1(0)$, we must reverse the sequence of the first four vectors in this basis.

12 Cyclic Decomposition and Rational Canonical Form

Let V be a finite-dimensional vector space over a field F, and $T : V \to V$ a linear map. If $F = \mathbb{C}$, then the characteristic polynomial $c_T(x)$ factorizes as a product of linear factors, so the JCF Theorem applies to T. But for other fields, such as $\mathbb{R}$, $\mathbb{Q}$ or $\mathbb{F}_p$ (p prime), many polynomials do not factorize into linear factors so the JCF Theorem does not apply. We need a more general canonical form theory. In this section we will prove the Rational Canonical Form Theorem. This works over any field, and states that there is a basis B of V such that

$$[T]_B = C(f_1) \oplus \cdots \oplus C(f_k),$$

where the matrices $C(f_i)$ are the companion matrices of uniquely determined polynomials $f_i \in F[x]$. The theory behind this result is based on the notion of *cyclic* subspaces, which we now introduce.

Cyclic subspaces

Let V be a finite-dimensional vector space over F, and $T : V \to V$ a linear map.

Definition Let $v \in V$ with $v \neq 0$, and define

$$\begin{aligned} Z(v, T) &= \{f(T)(v) : f(x) \in F[x]\} \\ &= \mathrm{Sp}\left(v, T(v), T^2(v), \ldots\right). \end{aligned}$$

Call $Z(v,T)$ the T*-cyclic subspace* of V generated by v (or slightly more briefly, the *cyclic* subspace generated by v). Clearly $Z(v,T)$ is T-invariant; we write T_v to denote the restriction of T to $Z(v,T)$.

Similarly, if A is an $n \times n$ matrix over F, and $0 \neq v \in F^n$, we define $Z(v,A) = \mathrm{Sp}(A^i v : i \geq 0)$.

Example Let

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

The 1-eigenspace of A is $E_1 = \mathrm{Sp}(e_1, e_3)$, so for any $v \in E_1$ we have $Z(v,A) = \mathrm{Sp}(v)$. All other cyclic subspaces $Z(w,A)$ are 2-dimensional: for $w \notin E_1$, we have $Z(w,A) = \mathrm{Sp}(w, e_1)$.

We next prove some basic facts about cyclic subspaces. Let v, T be as above. In the sequence

$$v, T(v), T^2(v), \ldots$$

let $T^k(v)$ be the first vector that is in the span of the previous ones. So we can express

$$T^k(v) = -a_0 v - a_1 T(v) - \cdots - a_{k-1} T^{k-1}(v) \tag{16}$$

for some $a_i \in F$. Define

$$m_v(x) = x^k + a_{k-1}x^{k-1} + \cdots + a_0 \in F[x].$$

By the choice of k, this is the monic polynomial of smallest degree with the property that $m_v(T)(v) = 0$. Note that also $m_v(T)(w) = 0$ for all $w \in Z(v,T)$.

Definition We call the polynomial $m_v(x)$ the T*-annihilator* of v and $Z(v,T)$.

Proposition 12.1 *With the above notation, the following hold:*

(1) $B = \{v, T(v), \ldots, T^{k-1}(v)\}$ *is a basis of* $Z(v,T)$ *(so* $\dim Z(v,T) = k$*).*

(2) *The matrix* $[T_v]_B$ *is the companion matrix*

$$C(m_v) = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 & -a_0 \\ 1 & 0 & 0 & \cdots & 0 & -a_1 \\ 0 & 1 & 0 & \cdots & 0 & -a_2 \\ & & & \cdots & & \\ 0 & 0 & 0 & \cdots & 1 & -a_{k-1} \end{pmatrix}.$$

(3) *The minimal polynomial of* T_v *is* $m_v(x)$*.*

Proof (1) By the choice of k, no vector in B is in the span of the previous ones, hence B is linearly independent. Now we show that B spans $Z(v,T)$. By (16), $T^k(v) \in \mathrm{Sp}(B)$. Hence, applying T to both sides of (16), we see that $T^{k+1}(v) \in \mathrm{Sp}(B)$. Continuing like this (or using induction), we see that $T^r(v) \in \mathrm{Sp}(B)$ for all $r \geq 0$, and hence $\mathrm{Sp}(B) = Z(v,T)$.

(2) This is clear.

(3) By Q on Sheet 4, the minimal polynomial of the companion matrix $C(m_v)$ is $m_v(x)$. $\square$

Example Let A be the following matrix over the field $\mathbb{F}_2$:

$$A = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{pmatrix}. \tag{17}$$

We compute the cyclic subspace $Z(e_1, A)$. The list of vectors $e_1, Ae_1, \ldots$ is

$$e_1,\ e_3 + e_4,\ e_1 + e_3 + e_4, \ldots$$

Hence $Z(e_1, A)$ has dimension 2 and basis $B = \{e_1, e_3 + e_4\}$, and $m_{e_1}(x) = x^2 + x + 1$. Finally, denoting also by A the linear map sending $v \to Av$, we have

$$[A_{e_1}]_B = C(m_{e_1}) = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}.$$

Recall the Primary Decomposition Theorem 10.1: if $m_T(x) = \prod_{i=1}^k f_i(x)^{n_i}$ where $f_1(x), \ldots, f_k(x) \in F[x]$ are distinct irreducible polynomials, then $V = V_1 \oplus \cdots \oplus V_k$, where each restriction T_{V_i} has minimal polynomial $f_i(x)^{n_i}$. Hence, as for the JCF Theorem, to decompose V further we need to focus on the case where $m_T(x) = f(x)^k$ with $f(x)$ irreducible. This is the content of the next result, which is the main theorem of this chapter.

Theorem 12.2 *(***Cyclic Decomposition Theorem***) Let V be a finite-dimensioanal vector space over a field F, let $T : V \to V$ be a linear map, and suppose the minimal polynomial $m_T(x) = f(x)^k$, where $f(x) \in F[x]$ is irreducible. Then there exist vectors $v_1, \ldots, v_r \in V$ such that*

$$V = Z(v_1, T) \oplus \cdots \oplus Z(v_r, T),$$

where

(1) *each $Z(v_i, T)$ has T-annihilator $f(x)^{k_i}$ for $1 \leq i \leq r$, and $k = k_1 \geq k_2 \geq \cdots \geq k_r$,*

(2) *the numbers r and $k_1, \ldots, k_r$ are uniquely determined by T.*

Before proving this, we deduce two corollaries. The first is just the matrix version of the theorem, which follows using Prop. 12.1.

Corollary 12.3 *Let T be as in Theorem* 12.2*. Then there is a basis B of V such that*

$$[T]_B = C\left(f(x)^{k_1}\right) \oplus \cdots \oplus C\left(f(x)^{k_r}\right),$$

where $k = k_1 \geq k_2 \geq \cdots \geq k_r$, uniquely determined by T.

Example Let A be the matrix over $\mathbb{F}_2$ as in (17) in the previous example. The characteristic polynomial $c_A(x) = (x^2 + x + 1)^2$. Hence (as $x^2 + x + 1 \in \mathbb{F}_2[x]$ is irreducible), $m_A(x) = (x^2 + x + 1)^i$ with $i = 1$ or 2. Check that $A^2 + A + I = 0$, hence $m_A(x) = x^2 + x + 1$. So it follows from Cor. 12.3 that

$$A \sim C(x^2 + x + 1) \oplus C(x^2 + x + 1) = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix}.$$

Corollary 12.3 implies one of the main results in our proof of the JCF Theorem, namely the nilpotent case (which was covered in Theorem 11.7):

Corollary 12.4 *Let A be an $n \times n$ matrix over F, and suppose $m_A(x) = x^k$. Then*

$$A \sim C(x^{k_1}) \oplus \cdots \oplus C(x^{k_r}),$$

where $k = k_1 \geq k_2 \geq \cdots \geq k_r$, uniquely determined by A.

Note that

$$C(x^k) = \begin{pmatrix} 0 & 0 & \cdots & 0 & 0 \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ & & \cdots & & \\ & & \cdots & 0 & 0 \\ & & \cdots & 1 & 0 \end{pmatrix} = J_k(0)^T$$

and $J_k(0)^T \sim J_k(0)$, so this does indeed imply Theorem 11.7. We chose to give a different proof of that theorem, since the method provided us with an algorithm for computing a Jordan basis.

Proof of Theorem 12.2

The proof proceeds by induction on $\dim V$. The result is obvious for $\dim V = 1$.

Now let $n = \dim V$, and assume the result is true for vector spaces of dimension less than n. The minimal polynomial $m_T(x) = f(x)^k$ with $f(x) \in F[x]$ irreducible. Hence there exists $v_1 \in V$ such that $f(T)^{k-1}(v_1) \neq 0$. The T-annihilator of v_1 is therefore $f(x)^k$. Define

$$Z_1 = Z(v_1, T),$$

a cyclic subspace with T-annihilator $f(x)^k$.

Let $\bar{V} = V/Z_1$, and let $\bar{T} : \bar{V} \to \bar{V}$ be the quotient map (defined by $\bar{T}(Z_1 + v) = Z_1 + T(v)$ for $v \in V$). By Prop. 9.4, the minimal polynomial $m_{\bar{T}}(x)$ divides $f(x)^k$, so is $f(x)^{k_2}$ for some $k_2 \leq k$. So we can apply the induction hypothesis to the map $\bar{T} : \to \bar{V}$: this impies that there are cosets $\bar{w}_2 = Z_1 + w_2, \ldots \bar{w}_r = Z_1 + w_r \in \bar{V} = V/Z_1$ such that the following hold:

(a) $\bar{V} = Z(\bar{w}_2, \bar{T}) \oplus \cdots \oplus Z(\bar{w}_r, \bar{T})$, and

(b) for $2 \leq i \leq r$, $\bar{w}_r$ has $\bar{T}$-annihilator $f(x)^{k_i}$, where $k_2 \geq \cdots \geq k_r$.

Claim 1 There exists a vector $v_2 \in Z_1 + w_2$ with T-annihilator $f(x)^{k_2}$.

Proof Let $v \in Z_1 + w_2 = \bar{w}_2$. Since $f(\bar{T})^{k_2}(\bar{w}_2) = Z_1$ (the zero vector of $\bar{V} = V/Z_1$), and $f(\bar{T})^{k_2}(Z_1 + v) = Z_1 + f(T)^{k_2}(v)$ by definition of $\bar{T}$, we have

$$f(T)^{k_2}(v) \in Z_1.$$

Hence by definition of $Z_1 = Z(v_1, T)$, there exists $g(x) \in F[x]$ such that

$$f(T)^{k_2}(v) = g(T)(v_1). \tag{18}$$

Then

$$0 = f(T)^k(v) = f(T)^{k-k_2} g(T)(v_1).$$

The T-annihilator of v_1 is $f(x)^k$, so $f(x)^k$ divides $f(x)^{k-k_2}g(x)$. Hence there exists $h(x) \in F[x]$ such that $g(x) = f(x)^{k_2}h(x)$. Define

$$v_2 = v - h(T)(v_1).$$

Then $v_2 \in Z_1 + v = Z_1 + w_2$, and

$$f(T)^{k_2}(v_2) = f(T)^{k_2}(v) - g(T)(v_1) = 0 \ \text{ (by (18))}.$$

Hence v_2 has T-annhilator $f(x)^{k_2}$, proving Claim 1.

Similarly, for $i = 2, \ldots, r$, there exists $v_i \in Z_1 + w_i$ with T-annhilator $f(x)^{k_i}$. Define

$$Z_i = Z(v_i, T) \quad (2 \le i \le r).$$

Claim 2 We have $V = Z_1 \oplus Z_2 \oplus \cdots \oplus Z_r$ (and so part (1) of the Theorem 12.2 is proved).

Proof We shall prove

(i) $\dim V = \sum_{i=1}^{r} \dim Z_i$, and

(ii) $V = Z_1 + Z_2 + \cdots + Z_r$.

By Prop. 4.2, Claim 2 follows from (i) and (ii).

Let us define a little more notation. Write $d = \deg(f)$. For $v \in V$, let $\bar{v} = Z_1 + v \in V/Z_1 = \bar{V}$. And for $i \ge 2$, define

$$\bar{Z}_i = \{\bar{z} : z \in Z_i\} = Z(\bar{w}_i, \bar{T}).$$

First note that for $i \ge 2$, both $\bar{Z}_i$ and Z_i have annihilator $f(x)^{k_i}$. Hence by Prop. 12.1(1),

$$\dim \bar{Z}_i = \dim Z_i = dk_i.$$

Also Z_1 has annihilator $f(x)^{k_1}$ (where $k_1 = k$), so $\dim Z_1 = dk_1$. As $\bar{V} = \bar{Z}_2 \oplus \cdots \oplus \bar{Z}_r$ (by (a) above), we have $\dim \bar{V} = \sum_{i=2}^{r} \dim \bar{Z}_i$, and it follows that

$$\dim V = \dim \bar{V} + \dim Z_1 = \sum_{i=1}^{r} \dim Z_i.$$

Finally, $\bar{V} = \bar{Z}_2 \oplus \cdots \oplus \bar{Z}_r$ implies that $V = Z_1 + Z_2 + \cdots + Z_r$. Thus (i) and (ii) are established, proving Claim 2.

We have now proved part (1) of Theorem 12.2, so it remains to prove the uniqueness statement (2). From Claim 2, we have

$$V = Z_1 \oplus \cdots \oplus Z_r, \tag{19}$$

where each Z_i has T-annihilator $f(x)^{k_i}$, and $k = k_1 \geq \cdots \geq k_r$. For $1 \leq i \leq r$, let n_i be the number of subspaces Z_i having annihilator $f(x)^{k_i}$. If we apply $f(T)^{k-1}$ to both sides of (19), we get

$$f(T)^{k-1}(V) = f(T)^{k-1}(Z_1) \oplus \cdots \oplus f(T)^{k-1}(Z_{n_1}).$$

By Q on Sheet 6, each subspace $f(T)^{k-1}(Z_i)$ (for $1 \leq i \leq n_1$) is cyclic with T-annihilator $f(x)$, and hence by Prop. 12.1(1) has dimension d. Hence

$$\dim f(T)^{k-1}(V) = dn_1.$$

Thus the value of n_1 is uniquely determined by T.

Next, apply $f(T)^{k-2}$ to both sides of (19):

$$\begin{aligned} f(T)^{k-2}(V) = \; & \left(f(T)^{k-2}(Z_1) \oplus \cdots \oplus f(T)^{k-2}(Z_{n_1})\right) \oplus \\ & \left(f(T)^{k-2}(Z_{n_1+1}) \oplus \cdots \oplus f(T)^{k-2}(Z_{n_1+n_2})\right). \end{aligned}$$

By Q on Sheet 6, on the right hand side, the n_1 subspaces in the first bracket have annihilator $f(x)^2$, and the n_2 subspaces in the second bracket have annihilator $f(x)$. Hence

$$\dim f(T)^{k-2}(V) = 2dn_1 + dn_2,$$

showing that n_2 is uniquely determined. Continuing in this fashion, we see that all of the n_i are determined uniquely, completing the proof of part (2) of Theorem 12.2. □

Rational Canonical Form

We are now ready to state and prove the Rational Canonical Form Theorem. The great thing about it is that it applies completely generally – to any linear map of any finite-dimensional vector space over any field.

Theorem 12.5 *(***Rational Canonical Form Theorem***) Let V be finite-dimensional over a field F, and let $T : V \to V$ be a linear map. Let the minimal polynomial $m_T(x)$ factorize as*

$$m_T(x) = \prod_{i=1}^{t} f_i(x)^{k_i}, \tag{20}$$

where $f_1(x), \ldots, f_t(x) \in F[x]$ are distinct irreducible polynomials. Then there exists a basis B of V such that

$$\begin{aligned} [T]_B = \; & C\left(f_1(x)^{k_{11}}\right) \oplus \cdots \oplus C\left(f_1(x)^{k_{1r_1}}\right) \oplus \cdots \\ & \oplus C\left(f_t(x)^{k_{t1}}\right) \oplus \cdots \oplus C\left(f_t(x)^{k_{tr_t}}\right), \end{aligned} \tag{21}$$

where for each i,

$$k_i = k_{i1} \geq \cdots \geq k_{ir_i}.$$

The numbers r_i and $k_{i1}, \ldots, k_{ir_i}$ are uniquely determined by T.

Corollary 12.6 *If A is an $n \times n$ matrix over a field F, with minimal polynomial as in* (20)*, then A is similar over F to a unique matrix of the form* (21)*.*

Definition In the situation of Corollary 12.6, we call the matrix (21) the *Rational Canonical Form* (RCF) of A.

Proof of Theorem 12.5

Let $T : V \to V$ be as in the hypothesis of the theorem. By the Primary Decomposition Theorem 10.1, if we let $V_i = \ker f_i(T)^{k_i}$ for $1 \le i \le t$, then

$$V = V_1 \oplus \cdots \oplus V_t,$$

where each restriction T_{V_i} has minimal polynomial $f_i(x)^{k_i}$. By Corollary 12.3, each V_i has a basis B_i such that

$$[T_{V_i}]_{B_i} = C\left(f_i(x)^{k_{i1}}\right) \oplus \cdots \oplus C\left(f_i(x)^{k_{ir_i}}\right),$$

where $k_i = k_{i1} \ge \cdots \ge k_{ir_i}$, and the numbers r_i and $k_{i1}, \ldots, k_{ir_i}$ are unique. Hence if B is the basis $B_1 \cup \cdots \cup B_t$ of V, then $[T]_B$ is as in (21) in the statement of the theorem, with uniqueness. $\square$

Remarks (1) The polynomials $f_i(x)^{k_{ij}}$ are called the *elementary divisors* of T.

(2) There is another version of the RCF Theorem: it states that every $n \times n$ matrix over F is similar to a unique matrix of the form

$$C(g_1) \oplus \cdots \oplus C(g_k),$$

where $g_i(x) \in F[x]$ are monic polynomials such that $g_i | g_{i+1}$ for all i. This can be deduced from Theorem 12.5 using the fact that if $f(x)$ and $g(x)$ are coprime polynomials in $F[x]$, then

$$C(f) \oplus C(g) \sim C(fg)$$

(see Q of Sheet 7).

We shall close this chapter by first giving a nice application of the RCF Theorem to a topic in group theory, and then discussing how to compute the RCF of any given matrix.

An application to group theory

Recall the *general linear group* $GL(n, F)$ is the group of all invertible $n \times n$ matrices over a field F (where the binary operation is of course matrix multiplication). Let $G = GL(n, F)$ and let $g \in G$. Using the symbol $\sim$ as usual for the relation of similarity of matrices, the similarity class of g is

$$\begin{aligned} [g] &= \{y \in G : y \sim g\} \\ &= \{y \in G : y = x^{-1}gx \text{ for some } x \in G\}. \end{aligned}$$

In the language of group theory, this is also called the *conjugacy class* of g in G.

When studying a group, one of the first things one needs to understand is its conjugacy classes. For the group $GL(n, F)$, this problem is solved by the RCF Theorem and its corollary 12.6, which impies that each conjugacy class has a unique representative that is an RCF matrix. In particular, the total number of conjugacy classes of $GL(n, F)$ is equal to the number of distinct RCFs of invertible $n \times n$ matrices over F.

Example Let

$$G = GL(3, \mathbb{F}_2),$$

the group of all invertible 3×3 matrices over the field $\mathbb{F}_2 = \{0, 1\}$. Let us compute the number of conjugacy classes of G.

The irreducible polynomials in $\mathbb{F}_2[x]$ of degree at most 3 are

$$x,\ x+1,\ x^2+x+1,\ x^3+x+1,\ x^3+x^2+1.$$

The possible characteristic polynomials of elements of G are products of these irreducibles that have total degree 3, but with no factor x (as matrices in G are invertible). There are four such polynomials, listed in column 1 of Table 1 below. The possible minimal polynomials divide these, and have the same irreducible factors; there are six possible minimal polynomials, listed in column 2 of the table. For each possible minimal polynomial, Corollary 12.6 shows that there is only one RCF matrix, as listed in column 3 of the table. We conclude that $GL(3, \mathbb{F}_2)$ has 6 conjugacy classes, and representatives of each of these classes are give by the matrices in column 3.

Table 1: Conjugacy classes of $GL(3, \mathbb{F}_2)$

char. poly.	possible min. polys.	RCF
$(x+1)^3$	$(x+1)$, $(x+1)^2$, $(x+1)^3$	I, $\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$, $\begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix}$
$(x+1)(x^2+x+1)$	$(x+1)(x^2+x+1)$	$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix}$
x^3+x+1	x^3+x+1	$\begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$
x^3+x^2+1	x^3+x^2+1	$\begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 1 \end{pmatrix}$

How to compute the RCF

Let $T : V \to V$ have characteristic and minimal polynomials

$$c_T(x) = \prod_{i=1}^{t} f_i(x)^{n_i}, \quad m_T(x) = \prod_{i=1}^{t} f_i(x)^{k_i},$$

where $f_1(x), \ldots, f_t(x) \in F[x]$ are distinct irreducible polynomials. To compute the RCF, it is enough to know, for each $i = 1, \ldots, t$, the values of

$$\text{rank}\,(f_i(T)^r) \quad (1 \le r \le k_i),$$

since then the RCF can be calculated by the method given in the last part of the proof of Theorem 12.2 (the proof of the uniqueness part (2) of the theorem).

But often much less information than this is needed. Here is an example.

Example Let A be a 15×15 matrix over $\mathbb{F}_2$, and suppose we are given the following information:

(1) $c_A(x) = (x+1)^5(x^2+x+1)^5$

(2) $m_A(x) = (x+1)^3(x^2+x+1)^2$

(3) $\mathrm{rank}(A+I) = 13$

(4) $\mathrm{rank}(A^2+A+I) = 9$.

Compute the RCF of A.

Answer Let $V = \mathbb{F}_2^{15}$ and denote the map $v \to Av$ also by A.

By the Primary Decomposition Theorem, we know that $V = V_1 \oplus V_2$, where $V_1 = \ker(A+I)^3$, $V_2 = \ker(A^2+A+I)^2$, of dimensions 5, 10 respectively (using the information in (1) and (2)).

By (2), the RCF of the restriction A_{V_1} is either

(a) $C(x+1)^3 \oplus C(x+1)^2$, or (b) $C(x+1)^3 \oplus C(x+1) \oplus C(x+1)$.

By (3), $\mathrm{rank}(A_{V_1} + I_{V_1}) = 3$, so there are 2 blocks. Hence case (a) holds.

By (2), writing $f(x) = x^2+x+1$, the RCF of A_{V_2} is either

(c) $C(f^2) \oplus C(f^2) \oplus C(f)$, or (b) $C(f^2) \oplus C(f) \oplus C(f) \oplus C(f)$.

By (4), $\mathrm{rank}(A_{V_2}^2 + A_{V_2} + I_{V_2}) = 4$, so (c) holds.

Hence the RCF of A is

$$C(x+1)^3 \oplus C(x+1)^2 \oplus C(f^2) \oplus C(f^2) \oplus C(f),$$

where $f(x) = x^2+x+1$.

13 The dual space

In this chapter we begin the geometric part of the course – inner product spaces, bilinear forms etc, as sketched in the Introduction. An important tool in this theory is the notion of a *dual space*, which we introduce here.

Definition Let V be a vector space over a field F. A *linear functional* on V is a linear map $\phi : V \to F$, ie. a map such that

$$\phi(\alpha v_1 + \beta v_2) = \alpha\phi(v_1) + \beta\phi(v_2) \quad \forall v_i \in V,\ \alpha, \beta \in F.$$

Examples (1) Let $V = F^n$ and define $\pi_i : V \to F$ by

$$\pi_i(x_1, \ldots, x_n) = x_i.$$

Then π_i is a linear functional, called the i^{th} *projection* map.

(2) Let $V = M_n(F)$, the vector space of $n \times n$ matrices over F. The trace map sending a matrix $A \to \mathrm{tr}(A)$ for $A \in V$ is a linear functional.

(3) The zero map $0 : V \to F$ is a linear functional.

We can add and scalar multiply linear functionals ϕ_1, ϕ_2 in the usual way: for any $v \in V$ and $\lambda \in F$,

$$(\phi_1 + \phi_2)(v) = \phi_1(v) + \phi_2(v),$$
$$(\lambda\phi)(v) = \lambda\phi(v).$$

Definition Let

$$V^* = \{\phi \mid \phi : V \to F \text{ a linear functional}\}.$$

With the above addition and scalar multiplication, V^* is a vector space over F (a routine exercise for the reader – you need to check all the vector space axioms, what fun). It is called the *dual space* of V.

Dimension

Observe that if $v_1, \ldots, v_n$ is a basis of V, and $\lambda_1, \ldots, \lambda_n \in F$, then there is a unique $\phi \in V^*$ that sends $v_i \to \lambda_i$ for all i (namely, $\phi\left(\sum \alpha_i v_i\right) = \sum \alpha_i \lambda_i$). In the following proposition, we use the "Kronecker delta" notation δ_{ij} – you have probably seen this: δ_{ij} is defined to be 1 if $i = j$ and 0 if $i \neq j$.

Proposition 13.1 *Let $n = \dim V$, and let $B = \{v_1, \ldots, v_n\}$ be a basis of V. For each $i = 1, \ldots, n$, define $\phi_i \in V^*$ by*

$$\phi_i(v_j) = \delta_{ij} \quad \textit{for } 1 \leq j \leq n$$

(so $\phi_i\left(\sum \alpha_j v_j\right) = \alpha_i$). Then $\{\phi_1, \ldots, \phi_n\}$ is a basis of V^*, called the dual basis of B. Hence $\dim V^* = n = \dim V$.

Proof If $\sum \lambda_i \phi_i = 0$, then for any j we have $0 = \sum \lambda_i \phi_i(v_j) = \lambda_j$. Hence $\phi_1, \ldots, \phi_n$ are linearly independent. To see then they span V^*, let $\sigma \in V^*$ and observe that

$$\sigma = \sum_{i=1}^{n} \sigma(v_i)\phi_i,$$

since both sides give the same value when applied to any basis vector v_j. $\square$

Examples (1) Let $V = F^n$ with standard basis $e_1, \ldots, e_n$. The dual basis is $\pi_1, \ldots, \pi_n$, where π_i is the projection map defined in Example (1) above.

(2) Let $V = \mathbb{R}^2$, with basis $v_1 = (2, 1)$, $v_2 = (3, 1)$. The dual basis is ϕ_1, ϕ_2 where

$$\phi_1(x_1, x_2) = -x_1 + 3x_2, \quad \phi_2(x_1, x_2) = x_1 - 2x_2.$$

Annihilators

Let V be a finite-dimensional vector space over a field F, and V^* the dual space.

Definition For a subset $X \subseteq V$, define the *annihilator* X^0 of X:

$$X^0 = \{\phi \in V^* : \phi(x) = 0 \ \forall x \in X\}.$$

I leave it as an easy exercise for you check that X^0 is a subspace of V^*.

Proposition 13.2 *If W is a subspace of V, then* $\dim W^0 = \dim V - \dim W$.

Proof Let $r = \dim W$ and let $w_1, \ldots, w_r$ be a basis of W. Extend this to a basis of V:

$$w_1, \ldots, w_r,\ v_1, \ldots, v_s.$$

Let the dual basis of V^* be $\phi_1, \ldots, \phi_r, \sigma_1, \ldots, \sigma_s$. Then each $\sigma_i \in W^0$.

Claim $\sigma_1, \ldots, \sigma_s$ is a basis of W^0.

Proof of Claim Obviously $\sigma_1, \ldots, \sigma_s$ are linearly independent as they are part of a basis. To show they span W^0, let $\sigma \in W^0$. We can express σ in terms of the dual basis:

$$\sigma = \sum_{i=1}^{r} \lambda_i \phi_i + \sum_{i=1}^{s} \mu_i \sigma_i.$$

As $\sigma \in W^0$, we have $\sigma(w_j) = 0$ for $1 \leq j \leq r$, so

$$0 = \sum_{1}^{r} \lambda_i \phi_i(w_j) = \lambda_j.$$

Hence $\sigma = \sum_{i=1}^{s} \mu_i \sigma_i$, showing that $\sigma_1, \ldots, \sigma_s$ span W^0 and proving the Claim.

The Claim shows that $\dim W^0 = s = \dim V = \dim W$, completing the proof. □

14 Inner Product Spaces

We now expore the geometry of vector spaces. The geometry of the Euclidean space $\mathbb{R}^n$ or the complex space $\mathbb{C}^n$ begins with the definition of the *dot product*: for vectors $x = (x_1, \ldots, x_n)^T$, $y = (y_1, \ldots, y_n)^T$,

$$x.y = \sum_{1}^{n} x_i \bar{y}_i \quad (= x^T \bar{y}).$$

Our first aim is to extend this notion to arbitrary vector spaces over $\mathbb{R}$ or $\mathbb{C}$. To do this we encapsulate the basic properties of the dot product in some axioms as follows.

Definition Let $F = \mathbb{R}$ or $\mathbb{C}$, and let V be a vector space over F. An *inner product* on V is a map $V \times V \to F$, denoted simply by $(u, v) \in F$ for any $u, v, \in V$, satisfying the following properties:

(1) $(\lambda_1 v_1 + \lambda_2 v_2, w) = \lambda_1 (v_1, w) + \lambda_2 (v_2, w)$,

(2) $(w, v) = \overline{(v, w)}$,

(3) $(v, v) > 0$ if $v \neq 0$,

for all $v_i, v, w \in V$ and $\lambda_i \in F$. We call such a vector space V with an inner product $(\ ,\ )$ an *inner product space* (real or complex).

Notes Here are some remarks about this definition.

(a) Condition (1) says that the inner product $(\ ,\ )$ is *left-linear*. Note that by (1) and (2),

$$(v, \lambda_1 w_1 + \lambda_2 w_2) = \bar{\lambda}_1 (v, w_1) + \bar{\lambda}_2 (v, w_2),$$

so the inner product is right-linear if $F = \mathbb{R}$, but not if $F = \mathbb{C}$.

(b) By (2) we have $(v, v) \in \mathbb{R}$, so condition (3) makes sense.

(c) We have $(0, v) = 0$ for all $v \in V$ (where of course the first 0 is the zero vector and the second is the zero scalar: this is because $(0, v) = (0v, v) = 0(v, v)$ (using (1)).

(d) If $F = \mathbb{R}$ then (2) says $(w, v) = (v, w)$, meaing that the inner product $(\,,\,)$ is *symmetric.*

(e) An elementary but important observation is that if $(v, w) = (v, x)$ for all $v \in V$, then $w = x$ (Q on Sheet 8).

Examples (1) The dot product on $\mathbb{R}^n$ or $\mathbb{C}^n$ is an inner product.

(2) Let V be the vector space over $\mathbb{R}$ of continuous functions $f : [0, 1] \to \mathbb{R}$, and for $f, g \in V$ define

$$(f, g) = \int_0^1 f(x)g(x)\,dx.$$

This is an inner product on V (exercise).

(3) Let V be the vector space consisting of all $m \times n$ matrices over $\mathbb{C}$, and for $A, B \in V$ define

$$(A, B) = \operatorname{tr}(B^T \bar{A}).$$

This is an inner product (Q on Sheet 8).

(4) Let $V = \mathbb{R}^2$, and for $x, y \in V$ define

$$\begin{aligned}(x, y) &= x_1y_1 - x_1y_2 - x_2y_1 + 3x_2y_2 \\ &= x^T \begin{pmatrix} 1 & -1 \\ -1 & 3 \end{pmatrix} y.\end{aligned}$$

We check that this is an inner product: axioms (1) and (2) are clear, and for (3), if $x \neq 0$,

$$(x, x) = x_1^2 - 2x_1x_2 + 3x_2^2 = (x_1 - x_2)^2 + 2x_2^2 > 0.$$

Matrix of an inner product

Let V be a finite-dimensional inner product space, let $B = \{v_1, \dots, v_n\}$ be a basis, and fro $1 \leq i, j \leq n$ define

$$a_{ij} = (v_i, v_j).$$

By axiom (2) we have $a_{ji} = \bar{a}_{ij}$, so the $n \times n$ matrix $A = (a_{ij})$ satisfies

$$A^T = \bar{A}.$$

If $F = \mathbb{R}$ such a matrix A is symmetric; and if $F = \mathbb{C}$ we call such a matrix A a *Hermitian* matrix. (We shall use the term Hermitian to cover both cases.) For $v, w \in V$ we have

$$(v, w) = [v]_B^T A [\bar{w}]_B,$$

where as usual $[v]_B$ is the coordinate vector of v with respect to B (see Q on Sheet 8). Hence by axiom (3), we have $x^T A\bar{x} > 0$ for all nonzero vectors $x \in F^n$.

Definition A Hermitian matrix A is said to be *positive-definite* if $x^T A\bar{x} > 0$ for all nonzero vectors $x \in F^n$ (where $F = \mathbb{R}$ or $\mathbb{C}$).

For example, as in Example (4) above, the symmetric matrix $\begin{pmatrix} 1 & -1 \\ -1 & 3 \end{pmatrix}$ is positive-definite.

In general, the eigenvalues of a Hermitian matrix A are all real, and A is positive-definite if and only if all its eigenvalues are positive (Q on Sheet 8).

Geometry

Let V be an inner product space over $F = \mathbb{R}$ or $\mathbb{C}$. For $u, v \in V$ define

the *length* $||u|| = \sqrt{(u,u)}$,

the *distance* $d(u,v) = ||u - v||$.

We say that u is a *unit vector* if $||u|| = 1$.

Here is our first basic geometric result.

Proposition 14.1 *For $u, v, w \in V$ the following hold.*

(1) $|(u,v)| \leq ||u||\,||v||$ *(Cauchy-Schwarz Inequality)*

(2) $||u + v|| \leq ||u|| + ||v||$

(3) $||u - v|| \leq ||u - w|| + ||w - v||$ *(Triangle Inequality).*

Proof (1) The result is trivial if $v = 0$, so assume $v \neq 0$. Let $v' = \frac{v}{||v||}$, a unit vector, and let $\lambda = (u, v')$. Then

$$\begin{aligned} 0 \leq ||u - \lambda v'||^2 &= (u - \lambda v', u - \lambda v') \\ &= ||u||^2 + \lambda\bar{\lambda}||v'||^2 - \lambda(v', u) - \bar{\lambda}(u, v') \\ &= ||u||^2 + \lambda\bar{\lambda} - \lambda\bar{\lambda} - \bar{\lambda}\lambda \\ &= ||u||^2 - |\lambda|^2. \end{aligned}$$

Hence

$$||u||^2 \geq |\lambda|^2 = \left| \left(u, \frac{v}{||v||} \right) \right|^2.$$

Now multiply through by $||v||^2$ to obtain part (1).

Parts (2) and (3) are simple deductions from (1), set as Q on Sheet 8. □

Dual space

Let V be an inner product space over $F = \mathbb{R}$ or $\mathbb{C}$. For $v \in V$ define $f_v : V \to F$ by

$$f_v(w) = (w, v) \quad \forall w \in V.$$

Then f_v is linear, so $f_v \in V^*$, the dual space of V.

We are interested in the map $V \to V^*$ sending $v \to f_v$. This map is not linear when $F = \mathbb{C}$, since $f_{\lambda v} = \bar{\lambda} f_v$ for $\lambda \in \mathbb{C}$. To remedy this, we define another vector space which we denote by $\bar{V}$. The vectors in $\bar{V}$ are the vectors of V, and addition and scalar multiplication are defined as follows:

addition in $\bar{V}$ is the same addition $u + v$ as in V

scalar multiplication $\lambda * v$ in $\bar{V}$ is different from V: for $\lambda \in F$, $v \in V$,

$$\lambda * v = \bar{\lambda} v$$

(where of course $\bar{\lambda} v$ is the scalar multiplication in V).

It is a routine exercise to check that $\bar{V}$ is a vector space over F of the same dimension as V (Q on Sheet 8).

Proposition 14.2 *Assume V is finite-dimensional, and define $\pi : \bar{V} \to V^*$ by*

$$\pi(v) = f_v \quad \forall v \in V.$$

Then π is a vector space isomorphism.

Proof We first check that π is linear. Clearly $f_{v_1+v_2} = f_{v_1} + f_{v_2}$, so we just need to check that $f_{\lambda * v} = \lambda f_v$ for $\lambda \in F, v \in V$: well, for $w \in V$,

$$\begin{aligned} f_{\lambda * v}(w) &= (w, \lambda * v) \\ &= (w, \bar{\lambda} v) \\ &= \lambda (w, v) \\ &= \lambda f_v(w). \end{aligned}$$

Hence $f_{\lambda * v} = \lambda f_v$ and so π is linear.

Next we show that $\ker \pi = 0$: well,

$$\begin{aligned} v \in \ker \pi &\Rightarrow f_v = 0 \\ &\Rightarrow (w, v) = 0 \;\; \forall w \in V \\ &\Rightarrow (v, v) = 0 \\ &\Rightarrow v = 0. \end{aligned}$$

Hence $\ker \pi = 0$. As $\dim \bar{V} = \dim V = \dim V^*$, it follows that π is an isomorphism. □

Corollary 14.3 *For any $f \in V^*$, there exists a unique $v \in V$ such that $f = f_v$.*

Orthogonality

Continue to assume that V is an inner product space over $F = \mathbb{R}$ or $\mathbb{C}$. We say the vectors $u, v \in V$ are *orthogonal* if $(u, v) = 0$. Note that by axiom (2) of inner product spaces, $(u, v) = 0 \Leftrightarrow (v, u) = 0$.

Definition A set of vectors $\{v_1, \ldots, v_k\}$ is *orthogonal* if $(v_i, v_j) = 0$ for all i, j with $i \neq j$. It is *orthonormal* if it is orthogonal and also $||v_i|| = 1$ for all i.

Examples (1) $e_1, \ldots, e_n$ is an orthonormal basis of F^n.

Here is another orthonormal basis of $\mathbb{C}^2$: $\frac{1}{\sqrt{2}}(1, i)$, $\frac{1}{\sqrt{2}}(i, 1)$.

(2) Let V be the vector space over $\mathbb{R}$ of continuous functions $f : [0, \pi] \to \mathbb{R}$, with inner product

$$(f, g) = \int_0^\pi f(x) g(x) \, dx.$$

The set $\{1, \cos x, \cos 2x, \ldots, \cos nx\}$ is orthogonal (Q on Sheet 8).

Definition For $W \subseteq V$, define

$$W^\perp = \{u \in V : (u, w) = 0 \; \forall w \in W\}.$$

It is a routine exercise to check that $W^\perp$ is a subspace of V.

Example Let $V = \mathbb{R}^3$ with the standard inner product (ie. the dot product). If $0 \neq w \in V$, then $w^\perp$ is the plane through 0 perpendicular to w.

Proposition 14.4 *Let V be a finite-dimensional inner product space, and let W be a subspace of V. Then*

$$V = W \oplus W^\perp.$$

Proof Consider the annihilator space $W^0 \subseteq V^*$:

$$\begin{aligned} W^0 &= \{f \in V^* : f(w) = 0 \; \forall w \in W\} \\ &= \{f_v \in V^* : (v, w) = 0 \; \forall w \in W\} \text{ (by Cor. 14.3)} \\ &= \{f_v : v \in W^\perp\}. \end{aligned}$$

The last subspace has the same dimension as $W^\perp$, by Prop. 14.2, and hence $\dim W^0 = \dim W^\perp$. Therefore by Prop. 13.2,

$$\dim W^\perp = \dim V - \dim W.$$

Finally, $W \cap W^\perp = 0$, since

$$v \in W \cap W^\perp \Rightarrow (v, v) = 0 \Rightarrow v = 0.$$

Hence $V = W \oplus W^\perp$ by Prop. 4.1. □

Here is one of the most fundamental results about inner product spaces.

Theorem 14.5 *Let V be a finite-dimensional inner product space.*

(1) *V has an orthonormal basis.*

(2) *Any orthonormal set of vectors $\{w_1, \ldots, w_r\}$ can be extended to an orthonormal basis of V.*

Proof (1) We proceed by induction on $n = \dim V$. The result is true for $n = 1$, since in this case $V = \mathrm{Sp}(v)$ for a nonzero vector v, hence V has an orthonormal basis $\frac{v}{||v||}$.

Now assume the result is true for inner product spaces of dimension $\leq n-1$, and let $n = \dim V$. Let $v_1 \in V$ be a unit vector, and let $W = \mathrm{Sp}(v_1)$. By Prop. 14.4 we have $V = W \oplus W^\perp$. Now $W^\perp$ is an inner product space of dimension $n-1$, so by the induction hypothesis, $W^\perp$ has an orthonormal basis $v_2, \ldots, v_n$. Then $v_1, v_2, \ldots, v_n$ is an orthonormal basis of V.

(2) Let $W = \mathrm{Sp}(w_1, \ldots, w_r)$. By Prop. 14.4 we have $V = W \oplus W^\perp$, and by (1), $W^\perp$ has an orthonormal basis $v_1, \ldots, v_s$. Then $w_1, \ldots, w_r, v_1, \ldots, v_s$ is an orthonormal basis of V. □

Gram-Schmidt Process

Another way of proving the existence of orthonormal bases is to use the *Gram-Schmidt Process* (you saw this in Year 1). This is a process to construct an orthonormal basis of an inner product space V. The steps are as follows:

Step 1 Start with any basis $v_1, \dots, v_n$ of V.

Step 2 Let $u_1 = \frac{v_1}{||v_1||}$, a unit vector, and define

$$w_2 = v_2 - (v_2, u_1)\, u_1.$$

Then $(w_2, u_1) = 0$. Let

$$u_2 = \frac{w_2}{||w_2||}.$$

Then $\{u_1, u_2\}$ is an orthonormal set of vectors.

Step 3 Let

$$w_3 = v_3 - (v_3, u_1)\, u_1 - (v_3, u_2)\, u_2$$

and $u_3 = \frac{w_3}{||w_3||}$. Then $\{u_1, u_2, u_3\}$ is an orthonormal set.

Step 4 Continue this process: at the i^{th} step let

$$w_i = v_i - (v_i, u_1)\, u_1 - \cdots - (v_i, u_{i-1})\, u_{i-1}$$

and $u_i = \frac{w_i}{||w_i||}$. After n steps, end up with an orthonormal basis $\{u_1, \dots, u_n\}$ with the property that

$$\mathrm{Sp}(u_1, \dots, u_i) = \mathrm{Sp}(v_1, \dots, v_i)$$

for all $i = 1, \dots, n$.

Note I chose the different method of proof given for Theorem 14.5, for two reasons:

(a) the method shows the basic connection between the inner product and the dual space

(b) the method (using the correspondence between $W^\perp$ and W^0) can be applied more generally, when we have a vector space with a *bilinear form* – we will do later this in Section 16.

Some applications of orthonormal bases

Orthonormal bases of inner product spaces have many applications. We will give two major ones.

(1) Fourier coefficients

Given an orthonormal basis, the Fourier coefficients of an arbitrary vector are the coefficients in its expression as a linear combination of the basis vectors. These can be computed using the following basic result.

Proposition 14.6 *Let V be an inner product space with an orthonormal basis $u_1, \dots, u_n$, and let $v \in V$.*

(1) *Then $v = \sum_{i=1}^n \lambda_i u_i$, where $\lambda_i = (v, u_i)$ (the Fourier coefficients of v).*

(2) $||v||^2 = \sum_{i=1}^n |\lambda_i|^2$.

Proof (1) We know that $v = \sum_{j=1}^n \lambda_j u_j$ for some scalars λ_j. Taking the inner product of both sides with u_i gives

$$(v, u_i) = \left(\sum \lambda_j u_j, u_i\right) = \lambda_i.$$

(2) We have

$$||v||^2 = \left(\sum \lambda_i u_i, \sum \lambda_j u_j\right) = \sum \lambda_i \bar{\lambda}_i = \sum |\lambda_i|^2. \quad \square$$

The reason these are called Fourier coefficients is because of the connection of all this with *Fourier series*. To decribe this, let V be the vector space over $\mathbb{R}$ of continuous functions $f : [0, \pi] \to \mathbb{R}$ (this is of course infinite-dimensional). As we have seen, V has an inner product

$$(f, g) = \frac{2}{\pi} \int_0^\pi f(x)g(x)\, dx.$$

Then the set of functions

$$\frac{1}{2}, \cos x, \cos 2x, \ldots, \cos nx, \ldots$$

is an orthonormal set in V. For $f \in V$, the Fourier coefficients are

$$\lambda_n = (f, \cos nx) = \frac{2}{\pi} \int_0^\pi f(x) \cos nx\, dx.$$

Fourier's famous theorem says that for $x \in [0, \pi]$, the series $\sum_{n=0}^\infty \lambda_n \cos nx$ is equal to $f(x)$. We call $\sum_{n=0}^\infty \lambda_n \cos nx$ the *Fourier cosine series* for $f(x)$.

(2) Projections

Let V be an inner product space, and let $v, w \in V \setminus 0$. The *projection of v along w* is defined to be the vector λw, where $\lambda = \frac{(v,w)}{(w,w)}$: this is the vector we hit when we drop a perpendicular from v to the line $\mathrm{Sp}(w)$. (This is easily seen by drawing a simple diagram as in lectures, but I am not capable of doing that in Latex.)

More generally, for a subspace W of V, and $v \in V$, we define the projection of v along W as follows: by Prop. 14.4 we have $V = W \oplus W^\perp$, so we can write

$$v = w + w'$$

for unique $w \in W$, $w' \in W^\perp$. Define $\pi_W : V \to W$ by

$$\pi_W(v) = w.$$

Definition We call π_W the *orthogonal projection* map along W.

Again, the geometry of this map is rather clear via a simple diagram, as shown in the lecture.

The projection π_W has nice geometrical properties:

Proposition 14.7 *Let V, W, π_W be as above.*

(1) *Let $v \in V$. Then $\pi_W(v)$ is the vector in W closest to v – in other words, for $w \in W$, the distance $||w - v||$ is minimal for $w = \pi_W(v)$.*

(2) *If $\text{dist}(v, W)$ denotes the shortest distance from v to any vector in W, then*

$$\text{dist}(v, W) = ||v - \pi_W(v)||.$$

(3) *If $v_1, \dots, v_r$ is an orthonormal basis of W, then*

$$\pi_W(v) = \sum_{j=1}^{r} (v, v_j)\, v_j.$$

Proof This is set as Q on Sheet 9. □

Change of orthonormal basis

The change of basis matrix from one orthonormal basis to another has a very special form, as shown in the next result.

Proposition 14.8 *Let V be an inner product space, and let $E = \{e_1, \dots, e_n\}$ and $F = \{f_1, \dots, f_n\}$ be orthonormal bases of V. Let $P = (p_{ij})$ be the change of basis matrix, so that for $1 \le i \le n$,*

$$f_i = \sum_{j=1}^{n} p_{ji} e_j.$$

Then $P^T \bar{P} = I$ (where $\bar{P}$ is the matrix $(\bar{p}_{ij})$).

Proof For any r, s we have

$$\begin{aligned} (f_r, f_s) &= \left(\textstyle\sum_{j=1}^{n} p_{jr} e_j,\ \sum_{k=1}^{n} p_{ks} e_k \right) \\ &= \textstyle\sum_{j=1}^{n} p_{jr} \bar{p}_{js} \\ &= (P^T \bar{P})_{rs}. \end{aligned}$$

Hence $(P^T \bar{P})_{rs} = \delta_{rs}$, and so $P^T \bar{P} = I$. □

Definition A real $n \times n$ matrix P such thayt $P^T P = I$ is called an *orthogonal* matrix. A complex $n \times n$ matrix P such that $P^T \bar{P} = I$ is called a *unitary* matrix.

These are very important classes of matrices. Here are two reasons why:

(1) They are the length-preserving maps of $\mathbb{R}^n$ and $\mathbb{C}^n$ (also called *isometries*), by which I mean that

$$||Pv|| = ||v||\ \forall v \in \mathbb{C}^n \quad \Leftrightarrow P \text{ is unitary},$$

with a similar statement for $\mathbb{R}^n$ and orthogonal matrices. (See Q on Sheet 9.)

(2) The set of all such isometries forms a group, known as a *classical* group:
orthogonal group $O(n, \mathbb{R}) = \{P \text{ real } n \times n \ :\ P^T P = I\}$,
unitary group $U(n, \mathbb{C}) = \{P \text{ complex } n \times n \ :\ P^T \bar{P} = I\}$.

These classical groups play a role in many parts of mathematics. There are some questions involving them on Sheet 9.

15 Linear maps on inner product spaces

Recall one of the basic theorems from 1st Year Linear Algebra: if A is a real symmetric matrix, then there is an orthogonal matrix P such that $P^{-1}AP$ is diagonal. This is often referred to as the "Spectral Theorem". Our aim in this chapter is to prove a generalization of the Spectral Theorem which applies to linear maps on inner product spaces. First, we need to define the analogue of a symmetric matrix for linear maps. To do this we will use the following result. As in the previous chapter, our inner product spaces are always over the field F, where $F = \mathbb{R}$ or $\mathbb{C}$.

Proposition 15.1 *Let V be a (f.d.) inner product space, and $T : V \to V$ a linear map. Then there is a unique linear map $T^* : V \to V$ such that for all $u, v, \in V$,*

$$(T(u), v) = (u, T^*(v)) .$$

Proof Let $v \in V$. The map $h : V \to F$ defined by

$$h(u) = (T(u), v) \quad \forall u \in V$$

is linear, so $h \in V^*$. Hence by Corollary 14.3, there is a unique $v' \in V$ such that $h = f_{v'}$, so that $h(u) = (u, v')$ for all $u \in V$. Define $T^* : V \to V$ by letting

$$T^*(v) = v'.$$

Then

$$(T(u), v) = (u, T^*(v)) \quad \forall u, v, \in V.$$

Finally, we must show that T^* is linear: for $\alpha, \beta \in F$,

$$\begin{aligned}(u, T^*(\alpha v_1 + \beta v_2)) &= (T(u), \alpha v_1 + \beta v_2)\\ &= \bar{\alpha}\,(T(u), v_1) + \bar{\beta}\,(T(u), v_2)\\ &= \bar{\alpha}\,(u, T^*(v_1)) + \bar{\beta}\,(u, T^*(v_2))\\ &= (u, \alpha T^*(v_1) + \beta T^*(v_2)) .\end{aligned}$$

This holds for all $u \in V$. Hence (using Q on Sheet 8), $T^*(\alpha v_1 + \beta v_2) = \alpha T^*(v_1) + \beta T^*(v_2)$. □

Definition The linear map T^* is called the *adjoint* of T. We say that T is *self-adjoint* if $T = T^*$.

Example Let $V = \mathbb{R}^n$ with the usual inner product (ie. the dot product), and let $T : V \to V$ be the linear map $T(v) = Av$, where A is a real $n \times n$ matrix. Then for $u, v \in V$,

$$\begin{aligned}(T(u), v) &= (Au)^T v\\ &= u^T A^T v\\ &= \left(u, A^T v\right) .\end{aligned}$$

Hence $T^*(v) = A^T v$, and T is self-adjoint iff $A = A^T$. ie. A is a symmetric matrix.

The last example generalizes to arbitrary inner product spaces:

Proposition 15.2 *Let V be an inner product space with orthonormal basis $E = \{v_1, \ldots, v_n\}$. Let $T : V \to V$ be a linear map, and let $A = [T]_E$. Then*

$$[T^*]_E = \bar{A}^T.$$

Proof By Prop. 14.6,

$$T(v_i) = \sum_{j=1}^{n} (T(v_i), v_j)\ v_j.$$

Hence the ij-entry of the matrix $A = [T]_E$ is

$$a_{ij} = (T(v_j), v_i)\,.$$

Therefore, if we let $B = [T^*]_E$, we have

$$\begin{aligned} b_{ij} &= (T^*(v_j), v_i) \\ &= \overline{(v_i, T^*(v_j))} \\ &= \overline{(T(v_i), v_j)} \\ &= \overline{a_{ji}}. \end{aligned}$$

Hence $[T^*]_E = \bar{A}^T$. $\square$

By the proposition, if $T = T^*$ and $A = [T]_E$, then $A = \bar{A}^T$. Hence if the field $F = \mathbb{R}$, then A is real symmetric; and if $F = \mathbb{C}$, then A is a complex *Hermitian* matrix.

Here is the main result of this chapter.

Theorem 15.3 (Spectral Theorem) *Let V be an inner product space, and let $T : V \to V$ be a self-adjoint linear map. Then V has an orthonormal basis of T-eigenvectors.*

Corollary 15.4 (1) *If A is an $n \times n$ real symmetric matrix, there exists an orthogonal matrix P such that $P^{-1}AP$ is diagonal.*
(2) *If A is an $n \times n$ complex Hermitian matrix, there exists a unitary matrix P such that $P^{-1}AP$ is diagonal.*

Proof Apply Theorem 15.3 to the linear map defined by $T(v) = Av$ for $v \in V = F^n$. $\square$

For the proof of the Spectral Theorem, we need the following lemma.

Lemma 15.5 *Let $T : V \to V$ be self-adjoint.*

(1) *The eigenvalues of T are all real.*

(2) *Eigenvectors for distinct eigenvalues are orthogonal to each other.*

(3) *If $W \subseteq V$ is T-invariant, so is $W^\perp$.*

Proof (1) Let v be an eigenvector with $T(v) = \lambda v$. Then as $T = T^*$, we have

$$\begin{aligned} & (T(v), v) = (v, T^*(v)) = (v, T(v)) \\ \Rightarrow\ & (\lambda v, v) = (v, \lambda v) \\ \Rightarrow\ & \lambda(v, v) = \bar{\lambda}(v, v) \\ \Rightarrow\ & \lambda = \bar{\lambda}\ \text{ (as } (v, v) > 0). \end{aligned}$$

(2) Let $T(u) = \lambda u$, $T(v) = \mu v$ with $\lambda \neq \mu$ (and both real, by (1)). Then

$$\begin{aligned} (T(u), v) = (u, T(v)) &\Rightarrow (\lambda u, v) = (u, \mu v) \\ &\Rightarrow \lambda(u, v) = \mu(u, v)\ \text{ (as } \mu \in \mathbb{R}) \\ &\Rightarrow (u, v) = 0\ \text{ (as } \lambda \neq \mu). \end{aligned}$$

(3) Let $x \in W^{\perp}$. Then for $w \in W$,

$$\begin{aligned}(w, T(x)) &= (T(w), x) \quad (\text{as } T = T^*) \\ &= 0 \quad (\text{as } T(w) \in W).\end{aligned}$$

Hence $T(x) \in W^{\perp}$. $\square$

Proof of Theorem 15.3

The proof proceeds by induction on $n = \dim V$, the case $n = 1$ being trivial.

Let $T : V \to V$ be self-adjoint. By Lemma 15.5(1), T has a real eigenvalue λ. Let u_1 be a unit eigenvector with $T(u_1) = \lambda u_1$, and define $W = \mathrm{Sp}(u_1)$. Then $\dim W^{\perp} = n - 1$, and $W^{\perp}$ is T-invariant by Lemma 15.5(3). The restriction $T_{W^{\perp}}$ is self-adjoint (as $(T(u), v) = (u, T(v))$ for all $u, v \in W^{\perp}$). Hence by the induction hypothesis, $W^{\perp}$ has an orthonormal basis of T-eigenvectors $u_2, \ldots, u_n$. Then $u_1, u_2, \ldots, u_n$ is an orthonormal basis of V consisting of T-eigenvectors. This completes the proof by induction. $\square$

There is a simple algorithm for computing an orthonormal basis of eigenvectors for a self-adjoint linear map T:

Step 1 Compute the eigenspaces E_{λ_i} of T.

Step 2 Use Gram-Schmidt to find an orthonormal basis B_i of each E_{λ_i}.

Step 3 By Lemma 15.5(2), for $i \neq j$, the eigenspaces E_{λ_i} and E_{λ_j} are orthogonal to each other. Hence the union of the bases B_i is an orthonormal basis of V.

You will find questions on Sheet 9 where you can use this algorithm.

16 Bilinear and Quadratic Forms

In this chapter we shall define and study some analogues of inner products over arbitrary fields. Since the axiom $(v, v) > 0$ does not make sense over an arbitrary field, we drop this condition.

Definition Let V be a vector space over a field F. A *bilinear form* on V is a map $(\,,\,) : V \times V \to F$ (ie. $(u, v) \in F$ for all $u, v \in V$) which is both left-linear and right-linear; in other words, for any $\alpha, \beta \in F$

$$\begin{aligned}(\alpha v_1 + \beta v_2, w) &= \alpha(v_1, w) + \beta(v_2, w), \text{ and} \\ (v, \alpha w_1 + \beta w_2) &= \alpha(v, w_1) + \beta(v, w_2).\end{aligned}$$

Examples (1) $V = \mathbb{R}^n$ with $(u, v) = u^T v$, the usual dot product.

However, for $\mathbb{C}^n$ the dot product $(u, v) = u^T \bar{v}$ is not bilinear as it is not right-linear.

(2) Let $V = \mathbb{R}^2$ and define

$$(u, v) = u_1 v_2 - u_2 v_1 = u^T \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} v.$$

This is a bilinear form. Notice that for $u = (1, 1)^T$ we have $(u, u) = 0$, so it is not an inner product.

(3) Here is a general example. Let F be any field, let $V = F^n$, and let A be an $n \times n$ matrix over F. Then

$$(u, v) = u^T A v \quad \forall u, v \in V$$

defines a bilinear form on V. (To get the usual dot product, we take $A = I$.)

In fact, all bilinear forms on $V = F^n$ arise as in Example (3), as we'll see next.

Matrices Let $(\ ,\ )$ be a bilinear form on a finite-dimensional vector space V, and let $B = \{v_1, \ldots, v_n\}$ be a basis of V. Define the matrix of $(\ ,\ )$ with respect to B to be the $n \times n$ matrix $A = (a_{ij})$, where

$$a_{ij} = (v_i, v_j).$$

Then for $u, v \in V$ we have

$$(u, v) = [u]_B^T A [v]_B$$

(exercise).

We shall focus on two particular types of bilinear forms that appear in many different parts of mathematics:

Definition A bilinear form $(\ ,\ )$ on V is

symmetric if $(v, u) = (u, v)$ for all $u, v \in V$

skew-symmetric if $(v, u) = -(u, v)$ for all $u, v \in V$.

If $(\ ,\ )$ is symmetric, then defining a_{ij} as above, we have $a_{ij} = (v_i, v_j) = (v_j, v_i) = a_{ji}$, so the matrix $A = A^T$ is symmetric. And if $(\ ,\ )$ is skew-symmetric, then $a_{ij} = -a_{ji}$, so $A^T = -A$ and A is a *skew-symmetric* matrix. For example, the form in Example (2) above is skew-symmetric.

Observe that if a bilinear form $(\ ,\ )$ is skew-symmetric, then taking $u = v$ we have $(v, v) = -(v, v)$ for all $v \in V$, so $2(v, v) = 0$. Provided $2 \neq 0$ in the field F, this implies that $(v, v) = 0$ for all $v \in V$.

Fields in which $2 = 0$ (for example the field $\mathbb{F}_2$) are called fields of *characteristic* 2. In general, the characteristic of a field F is the smallest positive integer n such that $n = 0$ in F if such an integer exists; if no such integer exists, we say F has characteristic 0. For example, $\mathbb{C}$ and $\mathbb{R}$ have characteristic 0, and $\mathbb{F}_p$ has characteristic p. Denote by $\mathrm{char}(F)$ the characteristic of F.

The above observation about skew-symmetric bilinear forms, although elementary, is important, so we record it in a lemma.

Lemma 16.1 *Let V be a vector space over a field F with $\mathrm{char}(F) \neq 2$, and let $(\ ,\ )$ be a skew-symmetric bilinear form on V, Then $(v, v) = 0$ for all $v \in V$.*

Orthogonality

In order to bring some geometrical ideas into the picture, we want to define perpendicular spaces $W^\perp$ and so on for a bilinear form. This only makes sense if we have the condition

$$(v, w) = 0 \Leftrightarrow (w, v) = 0. \tag{22}$$

Obviously this condition holds if $(\ ,\ )$ is symmetric or skew-symmetric. Less obviously, the converse holds:

Theorem 16.2 *A bilinear form $(\ ,\ )$ satisfies the condition (22) if and only if it is symmetric or skew-symmetric.*

The proof of this is straightforward, but rather long. I have included the theorem to motivate our focus on symmetric and skew-symmetric forms. A proof can be found in Theorem 1.17 of some nice online notes of Keith Conrad:

https://kconrad.math.uconn.edu/blurbs/linmultialg/bilinearform.pdf

From now on, we will consider only symmetric and skew-symmetric bilinear forms $(\ ,\ )$ on a finite-dimensional vector space V. As before in our study of inner products, for $W \subseteq V$ we define

$$W^{\perp} = \{v \in V : (v, w) = 0\ \forall w \in W\}.$$

This is a subspace of V (exercise).

Instead of the inner product axiom $(v, v) > 0$, we shall frequently impose the following condition.

Definition A bilinear form $(\ ,\ )$ on V is *non-degenerate* if $V^{\perp} = \{0\}$ – in other words, if for any $u \in V$,

$$(u, v) = 0\ \forall v \in V \Rightarrow u = 0.$$

Note that $V^{\perp} = \{0\}$ iff the matrix of $(\ ,\ )$ with respect to some basis is invertible (Q on Sheet 10).

Examples (1) For $V = F^2$, the bilinear form $(u, v) = u^T \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} v$ is non-degenerate.

(2) For $V = F^n$ with $n \geq 2$, the bilinear form $(u, v) = \sum_{i,j} u_i v_j$ is degenerate, since $(u, v) = u^T A v$ where A is the matrix with $a_{ij} = 1$ for all i, j.

Dual space

Here is a result for bilinear forms similar to Props. 14.2 and 14.4 for inner products.

Proposition 16.3 *Suppose $(\ ,\ )$ is a non-degenerate bilinear form (symmetric or skew-symmetric) on a finite-dimensional vector space V.*

(1) *For $v \in V$, define $f_v \in V^*$ by $f_v(u) = (v, u)$ for all $u \in V$. Then the map*

$$\phi : v \to f_v \quad (v \in V)$$

is an isomorphism $V \to V^*$.

(2) *For a subspace W of V, we have $\dim W^{\perp} = \dim V - \dim W$.*

Proof (1) The map ϕ is linear, and

$$v \in \ker(\phi) \Leftrightarrow f_v = 0 \Leftrightarrow (v, u) = 0\ \forall u \in V \Leftrightarrow u = 0$$

(using the fact that $(\ ,\ )$ is non-degenerate for the last deduction). Hence $\ker(\phi) = 0$, and so ϕ is injective. Since $\dim V = \dim V^*$ by 13.1, it follows that ϕ is an isomorphism.

(2) As in the proof of Prop 14.4, we have

$$W^0 = \{f_v : v \in W^{\perp}\}.$$

Hence $\dim W^\perp = \dim W^0 = \dim V - \dim W$ by 13.2. $\square$

Note that, unlike the case of inner product spaces, it is not always that case that $V = W \oplus W^\perp$ for a subspace W. For example, if $W = \mathrm{Sp}(v)$, where v is a nonzero vector v such that $(v,v) = 0$, then $v \in W \cap W^\perp$ so $V \neq W \oplus W^\perp$.

Bases

For any inner product on $\mathbb{R}^n$ there is an orthonormal basis (by Theorem 14.5), and the matrix of the inner product with respect to this basis is of course I, the identity. Obviously there cannot be such a basis in general for a bilinear form – for example, for a skew-symmetric form the matrix is always skew-symmetric. But can we find a "nice" basis? We'll see below that the answer is yes.

First we need to discuss what happens to the matrix of a bilinear form when we change the basis. Let $(\ ,\ )$ be a bilinear form on V and let B_1, B_2 be bases of V. Let A be the matrix of the form with respect to B_1. To aid notation, for $v \in V$ and $i = 1,2$ write $[v]_{B_i} = [v]_i$. If P is the change of basis matrix, so that $[v]_1 = P[v]_2$ for all $v \in V$, then

$$\begin{aligned} (u,v) &= [u]_1^T A [v]_1 \\ &= (P[u]_2)^T A (P[v]_2) \\ &= [u]_2^T P^T A P [v]_2. \end{aligned}$$

Hence the matrix of the form with respect to B_2 is $P^T AP$.

Definition Two $n \times n$ matrices A, B over F are said to be *congruent* if there exists an invertible matrix P over F such that $B = P^T AP$.

If A, B are congruent, the corresponding bilinear forms $(u,v)_1 = u^T Av$ and $(u,v)_2 = u^T Bv$ on F^n are said to be *equivalent*.

Check that congruence is an equivalence relation on $n \times n$ matrices. By the above discussion, our question becomes this: given a matrix A (symmetric or skew-symmetric), can we find an invertible P such that $P^T AP$ is a "nice" matrix? Theorems 16.4 and 16.6 below provide some answers. Perhaps surprisingly, the answer is much more precise for the skew-symmetric case.

Skew-symmetric bilinear forms

Theorem 16.4 *Let V be a finite-dimensional vector space over a field F, where* $\mathrm{char}(F) \neq 2$, *and let $(\ ,\ )$ be a non-degenerate skew-symmetric bilinear form on V. Then*

(1) $\dim V$ *is even, and*

(2) *There is a basis $B = \{e_1, f_1, \ldots, e_m, f_m\}$ of V such that the matrix of $(\ ,\ )$ with respect to B is the block-diagonal matrix*

$$J_m = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \oplus \cdots \oplus \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \quad (m \text{ blocks}), \tag{23}$$

(so that $(e_i, f_i) = -(f_i, e_i) = 1$ and $(e_i, e_j) = (f_i, f_j) = (e_i, f_j) = (f_j, e_i) = 0$ for all $i \neq j$).

Corollary 16.5 *If A is an invertible skew-symmetric $n \times n$ matrix over F, where* $\mathrm{char}(F) \neq 2$, *then $n = 2m$ for some m, and A is congruent to the matrix J_m in (23).*

Another way of stating this is to say that any non-degenerate skew-symmetric bilinear form on F^n is equivalent to the form

$$(x,y) = x^T J_m y = (x_1y_2 - x_2y_1) + \cdots + (x_{m-1}y_m - x_my_{m-1}).$$

Proof of Theorem 16.4

By Lemma 16.1, we have $(v,v) = 0$ for all $v \in V$. Hence $\dim V > 1$, as $(\, , \,)$ is non-degenerate. The proof proceeds by induction on $n = \dim V$.

Let $e_1 \in V \setminus 0$. Then $(e_1,e_1) = 0$. By 16.3, $\dim e_1^\perp = n-1$, so there exists $f \in V \setminus e_1^\perp$. Let $\lambda = (e_1,f)$ and $f_1 = \lambda^{-1}f$. Then $(e_1,f_1) = 1$; also $(f_1,e_1) = -1$ and $(e_1,e_1) = (f_1,f_1) = 0$. If $n = \dim V = 2$, then e_1, f_1 is the required basis, so the induction starts at $n = 2$. Now suppose $n > 2$.

Let $W = \mathrm{Sp}(e_1,f_1)$, a 2-dimensional subspace. We claim that $W \cap W^\perp = \{0\}$. To see this, let $w \in W \cap W^\perp$, and write $w = \alpha e_1 + \beta f_1$. Then

$$0 = (e_1,w) = \beta, \;\; 0 = (f_1,w) = -\alpha,$$

and hence $w = 0$, proving the claim.

Now $\dim W + \dim W^\perp = n$ by 16.3, so by the claim, we have

$$V = W \oplus W^\perp.$$

Therefore, if we restrict the form $(\, , \,)$ to $W^\perp$ it is non-degenerate, and so by the induction hypothesis, $W^\perp$ has even dimension and has a basis $e_2, f_2, \ldots, e_m, f_m$ such that the matrix of the restriction of $(\, , \,)$ with respect to this basis is

$$J_{m-1} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \oplus \cdots \oplus \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \quad (m-1 \text{ blocks}).$$

Then $e_1, f_1, e_2, f_2, \ldots, e_m, f_m$ is the required basis of V, completing the proof by induction. $\square$

Remark In the literature, a non-degenerate skew-symmetric form on V is often called a *symplectic form*. By Theorem 16.4, for any even-dimensional vector space V over any field of characteristic $\neq 2$, there is, up to congruence, a *unique* symplectic form on V.

Symmetric bilinear forms

Theorem 16.6 *Let V be a finite-dimensional vector space over a field F, where $\mathrm{char}(F) \neq 2$, and let $(\, , \,)$ be a non-degenerate symmetric bilinear form on V. Then V has an orthogonal basis $B = \{v_1, \ldots, v_n\}$, ie. a basis such that $(v_i,v_j) = 0$ for $i \neq j$ and $(v_i,v_i) = \alpha_i \neq 0$ for all i. The matrix of $(\, , \,)$ with respect to B is the diagonal matrix $\mathrm{diag}(\alpha_1, \ldots, \alpha_n)$.*

Corollary 16.7 *If A is an invertible symmetric matrix over a field F, where $\mathrm{char}(F) \neq 2$, then A is congruent to a diagonal matrix.*

Proof of Theorem 16.6

As usual we proceed by induction on $n = \dim V$. The result is clear for $n = 1$.

Claim 1 There exists $v \in V$ such that $(v, v) \neq 0$.

Proof Suppose $(v, v) = 0$ for all $v \in V$. Let $u, w \in V$. Then

$$\begin{aligned}(u+w, u+w) = 0 \quad &\Rightarrow (u,u) + (w,w) + (u,w) + (w,u) = 0 \\ &\Rightarrow 2(u,w) = 0 \\ &\Rightarrow (u,w) = 0 \text{ (since } \operatorname{char}(F) \neq 2).\end{aligned}$$

Hence $(u, w) = 0$ for all $u, w \in V$, which is a contradiction. Hence Claim 1 is proved.

By Claim 1, we can choose $v_1 \in V$ such that $(v_1, v_1) \neq 0$. Let $W = \operatorname{Sp}(v_1)$.

Claim 2 We have $V = W \oplus W^{\perp}$.

Proof Observe that

$$w \in W \cap W^{\perp} \Rightarrow w = \lambda v_1 \text{ and } (\lambda v_1, v_1) = \lambda(v_1, v_1) = 0 \Rightarrow \lambda = 0.$$

Hence $W \cap W^{\perp} = 0$, and Claim 2 follows.

We can now conclude as in the previous proof. By Claim 2, the restriction of the form $(\, , \,)$ to $W^{\perp}$ is non-degenerate. Hence by the induction hypothesis, $W^{\perp}$ has an orthogonal basis $v_2, \ldots, v_n$. Then $v_1, v_2, \ldots, v_n$ is an orthogonal basis of V, completing the proof by induction. □

Remarks (1) The conclusion of Therorem 16.6 may not hold if $\operatorname{char}(F) = 2$. For example, let $F = \mathbb{F}_2$ and $V = F^2$, and for $x, y \in V$ define

$$(x, y) = x^T \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} y = x_1 y_2 + x_2 y_1.$$

This is a non-degenerate symmetric bilinear form, but there is no orthogonal basis since $(x, x) = 0$ for all vectors $x \in V$.

(2) How can we compute an orthogonal basis? Gram-Schmidt does not necessarily work, since we might start with a basis $w_1, \ldots, w_n$ such that $(w_1, w_1) = 0$. The most obvious algorithm is simply to follow along the lines of the proof of the theorem:

1) find v_1 such that $(v_1, v_1) \neq 0$
2) compute $v_1^{\perp}$ and find $v_2 \in v_1^{\perp}$ such that $(v_2, v_2) \neq 0$
3) compute $\operatorname{Sp}(v_1, v_2)^{\perp}$ and find $v_3 \in \operatorname{Sp}(v_1, v_2)^{\perp}$ such that $(v_3, v_3) \neq 0$
4) carry on choosing vectors like this until we get an orthogonal basis.

Example Let $V = F^2$ with $\operatorname{char}(F) \neq 2$, and bilinear form $(x, y) = x^T \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} y$.

1) Take $v_1 = (1, 1)^T$. Then $(v_1, v_1) = 2$.
2) Compute $v_1^{\perp} = \operatorname{Sp}(1, -1)^T$. Take $v_2 = (1, -1)^T$, so $(v_2, v_2) = -2$.

Now have orthogonal basis v_1, v_2. With respect to this basis, the form has matrix $\begin{pmatrix} 2 & 0 \\ 0 & -2 \end{pmatrix}$.

By 16.7, to classify symmetric forms (up to congruence), we need to be able to answer the following question: given diagonal matrices

$$D_1 = \operatorname{diag}(\alpha_1, \ldots, \alpha_n), \; D_2 = \operatorname{diag}(\beta_1, \ldots, \beta_n), \quad (\alpha_i, \beta_i \in F),$$

are D_1 and D_2 congruent over F? (ie. does there exist an invertible P over F such that $D_2 = P^T D_1 P$?) This can be a complicated question, and depends on properties of the field F. Here is a simple example. We use the notation $A \equiv B$ to mean that A is congruent to B.

Example Let $D_1 = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$, $D_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$. Then

$$\begin{aligned} D_1 &\equiv D_2 \text{ over } \mathbb{C} \;\; \text{(for example, take } P = \operatorname{diag}(1, \tfrac{1}{\sqrt{2}})) \\ &\equiv D_2 \text{ over } \mathbb{R} \\ &\not\equiv D_2 \text{ over } \mathbb{Q} \;\; \text{(Ex)} \\ &\equiv D_2 \text{ over } \mathbb{F}_7 \;\; \text{(Ex)} \\ &\not\equiv D_2 \text{ over } \mathbb{F}_3 \;\; \text{(Ex)} \end{aligned}$$

To discuss this question further, we introduce the next topic.

Quadratic forms

Assume from now on that F is a field with $\operatorname{char}(F) \neq 2$, and V is a finite-dimensional vector space over F.

Definition A *quadratic form* on V is a map $Q : V \to F$ of the form

$$Q(v) = (v, v) \quad \forall v \in V,$$

where $(\ ,\ )$ is a symmetric bilinear form on V. We also say that Q is non-degenerate if $(\ ,\ )$ is non-degenerate.

Remarks (1) We can determine the form $(\ ,\ )$, given the map Q, since for any $u, v \in V$,

$$(u, v) = \frac{1}{2}\left(Q(u+v) - Q(u) - Q(v)\right).$$

(2) On $V = F^n$, every symmetric bilinear form takes the form

$$(x, y) = x^T A y \quad (x, y \in V),$$

where $A = A^T$. So for $x = (x_1, \ldots, x_n)^T$, we have

$$\begin{aligned} Q(x) &= x^T A x \\ &= \textstyle\sum_{i,j} a_{ij} x_i x_j \\ &= \textstyle\sum_{i=1}^n a_{ii} x_i^2 + 2 \sum_{i<j} a_{ij} x_i x_j. \end{aligned}$$

This is a general *homogeneous* quadratic polynomial in $x_1, \ldots, x_n$ (the term "homogeneous" means that every term has the same degree, namely 2).

Example For $n = 2$, $V = F^2$ and

$$\begin{aligned} Q(x_1, x_2) &= a x_1^2 + b x_1 x_2 + c x_2^2 \\ &= x^T \begin{pmatrix} a & b/2 \\ b/2 & c \end{pmatrix} x. \end{aligned}$$

If $F = \mathbb{R}$, the equation $Q(x_1, x_2)$ defines a *conic* in $\mathbb{R}^2$.

Change of variables

Let $V = F^n$ and $Q : V \to F$ a quadratic form, so

$$Q(x) = x^T Ax \quad \forall x \in V,$$

where A is symmetric. If we change variables to $y = (y_1, \dots, y_n)^T$, where $x = Py$ with P invertible, then

$$Q(x) = (Py)^T A(Py) = y^T P^T APy = Q'(y). \tag{24}$$

If there exists P such that (24) holds, we say the quadratic forms Q and Q' are *equivalent.* Check that this is an equivalence relation on quadratic forms. Note that the congruent matrices A and $P^T AP$ are *not* in general similar to each other (but they are of course similar if P happens to be an orthogonal matrix, since then $P^T = P^{-1}$).

Example Consider the quadratic forms Q, Q' on F^2 defined by

$$Q(x_1, x_2) = 4x_1x_2 = x^T \begin{pmatrix} 0 & 2 \\ 2 & 0 \end{pmatrix} x, \;\; Q'(x_1, x_2) = x_1^2 - x_2^2 = x^T \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

These are equivalent forms, since

$$Q(x_1, x_2) = 4x_1x_2 = (x_1 + x_2)^2 - (x_1 - x_2)^2 = y_1^2 - y_2^2 = Q'(y_1, y_2),$$

where $y_1 = x_1 + x_2$, $y_2 = x_1 - x_2$.

By Theorem 16.6, every non-degenerate quadratic form on F^n is equivalent to a form

$$Q(x) = x^T \operatorname{diag}(\alpha_1, \dots, \alpha_n)x = \alpha_1 x_1^2 + \cdots + \alpha_n x_n^2. \tag{25}$$

Using this we can classify (up to equivalence) all quadratic forms over $\mathbb{C}$ and $\mathbb{R}$, and also say something about forms over $\mathbb{Q}$:

Theorem 16.8 *Let $V = F^n$, and let $Q : V \to F$ be a non-degenerate quadratic form.*

(1) *If $F = \mathbb{C}$, then Q is equivalent to the form*

$$Q_0(x) = x_1^2 + \cdots + x_n^2 \quad (x \in \mathbb{C}^n).$$

This form has matrix I_n.

(2) *If $F = \mathbb{R}$, then Q is equivalent to a unique form $Q_{p,q}$, where $p + q = n$ and*

$$Q_{p,q}(x) = x_1^2 + \cdots + x_p^2 - x_{p+1}^2 - \cdots - x_{p+q}^2 \quad (x \in \mathbb{C}^n).$$

This form has matrix $I_{p,q} = \begin{pmatrix} I_p & 0 \\ 0 & -I_q \end{pmatrix}$.

(3) *If $F = \mathbb{Q}$, there are infinitely many inequivalent non-degenerate quadratic forms on $\mathbb{Q}^n$.*

Proof (1) Start with Q as in (25). Over $\mathbb{C}$, we have

$$\operatorname{diag}(\alpha_1, \dots, \alpha_n) = P^T IP,$$

where $P = \operatorname{diag}(\sqrt{\alpha_1}, \dots, \sqrt{\alpha_n})$ (all these square roots exist in $\mathbb{C}$). Hence Q is equivalent to $Q_0(x) = x^T Ix = \sum_1^n x_i^2$.

(2) Again start with Q as in (25). Note that all the diagonal entries α_i are nonzero, as Q is non-degenerate. Re-order the α_i's so that

$$\alpha_1, \ldots, \alpha_p > 0, \;\; \alpha_{p+1}, \ldots, \alpha_{p+q} < 0.$$

Then

$$\mathrm{diag}(\alpha_1, \ldots, \alpha_n) = P^T I_{p,q} P,$$

where

$$P = \;\mathrm{diag}(\sqrt{\alpha_1}, \ldots, \sqrt{\alpha_p}, \sqrt{-\alpha_{p+1}}, \ldots, \sqrt{-\alpha_{p+q}}).$$

Hence Q is equivalent to the form $Q_{p,q}$ defined in part (2).

It remains to prove the uniqueness assertion in part (2). (This is a famous property of real quadratic forms known as "Sylvester's Law of Inertia".) Suppose that

$$Q \sim Q_{p,q} \sim Q_{p',q'},$$

where of course $\sim$ denotes equivalence of quadratic forms. Let $(\;,\;)$ be the bilinear form on V corresponding to Q (ie. $(x,y) = \frac{1}{2}(Q(x+y) - Q(x) - Q(y))$). As $Q \sim Q_{p,q}$, there is an orthogonal basis $v_1 \ldots, v_n$ of V such that

$$(v_i, v_i) = \begin{cases} 1, \text{ for } 1 \le i \le p \\ -1, \text{ for } p+1 \le i \le n. \end{cases}$$

Likewise, as $Q \sim Q_{p',q'}$, there is another orthogonal basis $w_1, \ldots, w_n$ such that

$$(w_i, w_i) = \begin{cases} 1, \text{ for } 1 \le i \le p' \\ -1, \text{ for } p'+1 \le i \le n. \end{cases}$$

Let

$$U = \mathrm{Sp}(v_1, \ldots, v_p), \;\; W = \mathrm{Sp}(w_{p'+1}, \ldots, w_n).$$

Then

$$Q(u) = (u,u) > 0 \text{ for } u \in U \setminus 0,$$

$$Q(w) < 0 \text{ for } w \in W \setminus 0.$$

Hence $U \cap W = 0$. Consequently

$$n \ge \dim(U + W) = \dim U + \dim W = p + n - p'.$$

It follows that $p' \ge p$. Similarly (by a symmetrical argument) we have $p \ge p'$. Hence $p' = p$, proving uniqueness.

(3) Let $F = \mathbb{Q}$. By a *square-free* integer d, we mean an integer that is not divisible by any integer square (apart from 1) – this amounts to saying that d is a product of distinct primes. For d a positive square-free integer, define a quadratic form Q_d on $\mathbb{Q}^n$ by

$$Q_d(x) = x_1^2 + \cdots + x_{n-1}^2 + dx_n^2 \quad (x \in \mathbb{Q}^n).$$

This form has matrix $A_d = \mathrm{diag}(1, \ldots, 1, d)$. If $Q_d \sim Q_{d'}$, then there exists a rational invertible matrix P such that

$$P^T A_d P = A_{d'}.$$

Take determinants, to get

$$(\det P)^2 d = d'.$$

Hence $\frac{d'}{d}$ is the square of a rational number. Since d and d' are square-free, this forces $d = d'$. We conclude that if d and d' are square-free with $d \neq d'$, then $Q_d \not\sim Q_{d'}$, and part (3) follows. $\square$

Some applications of bilinear and quadratic forms

Having read through this chapter, you may ask what is the point of all this theory of bilinear and quadratic forms. The generalised answer is that these occur naturally in many branches of mathematics. Let me mention just a few here, and leave it at that. A quick internet search will lead you to many more such instances.

(1) *Special relativity* The general setting for this theory is *Minkowski spacetime*, which is $\mathbb{R}^4$ together with the bilinear form $(x, y) = x_1y_1 + x_2y_2 + x_3y_3 - x_4y_4$ and associated quadratic form $Q(x) = x_1^2 + x_2^2 + x_3^2 - x_4^2$.

(2) *Number theory* A classical question in number theory asks the following. Given a rational quadratic form $Q : \mathbb{Q}^n \to \mathbb{Q}$, and a rational k, does the equation $Q(x) = k$ have a rational solution $x \in \mathbb{Q}^n$? Even more classically, one asks for the integer solutions of such equations – for example the Pythagorean equation $x^2 + y^2 = k$, or Pell's equation $x^2 - dy^2 = 1$. There is a huge amount of theory arising from such questions. See for example the book "Rational Quadratic Forms " by J W S Cassels.

(3) *Classical groups* Just as we did for inner product spaces, one can define isometries of bilinear and quadratic forms and get interesting groups. Here is a quick sketch.

Definition Let $f = (\ ,\ )$ be a non-degenerate symmetric or skew-symmetric bilinear form on a finite-dimensional vector space V. An *isometry* of f is a linear map $T : V \to V$ such that

$$(T(u), T(v)) = (u, v) \quad \forall u, v \in V.$$

Note that T is invertible, since f is non-degenerate. Define further

$$I(V, f) = \{T : T \text{ an isometry}\}.$$

This is a subgroup of the general linear group $GL(V)$. (Q on Sheet 10)

We can also define these groups in terms of matrices. Fix a basis B of V, and let A be the matrix of f with respect to B. If $[T]_B = X$, then $T \in I(V, f)$ iff $X^TAX = A$ (Sheet 10). Hence $I(V, f)$ is isomorphic to a group of matrices:

$$I(V, f) \cong \{X \in GL(n, F) : X^TAX = A\}.$$

If f is skew-symmetric, there is only one form (up to equivalence) by Theorem 16.4, and so we get one isometry group – the classical *symplectic group* $\mathrm{Sp}(V, f)$.

If f is symmetric, there are in general many possible forms (see Theorem 16.8), and the corresponding isometry groups are the classical *orthogonal groups* $O(V, f)$.

These families of classical groups (together with the ones we saw earlier in Chapter 14) play a huge role in various parts of mathematics such as geometry, algebra and number theory. I hope you will see some of them again in your future studies.

That is the end of the course. Thank you for your attention, hope you enjoyed it!