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*

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

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*

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

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

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

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 $\lambda_1, \ldots, \lambda_r$ 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

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

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 $\lambda_1, \ldots, \lambda_r$ 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*

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

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

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

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):

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

And here is the check for (S):

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

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:

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

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

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

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

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

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 $Z = (z_{ij})$. □

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:*

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

(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

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

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

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

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

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

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

(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. □

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. □

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. □

To complete the section, we define the *least common multiple* $\mathrm{lcm}(f, g)$ 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 $\mathrm{lcm}(f, g)$ 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.