Chapter 5

Linear maps on a vector space

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

5.1 Projections and direct sums

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

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

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

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

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

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

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

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

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

□

There is a converse to this result.

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

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

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

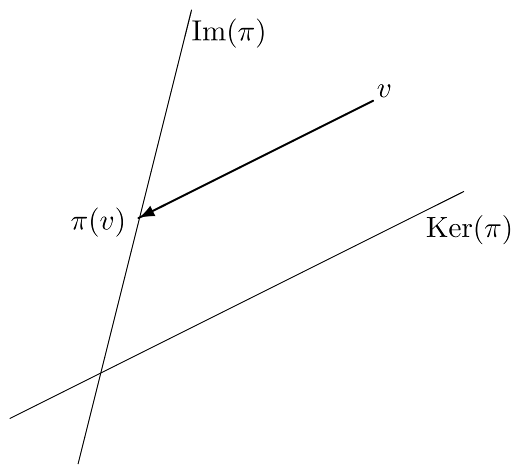

Figure 5.1: A projection

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

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

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

Proposition 5.4. *Suppose that $\pi_1, \pi_2, \ldots, \pi_r$ are projections on V satisfying*

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

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

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

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

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

where $u_i = \pi_i(v) \in \mathrm{Im}(\pi_i) = U_i$ for $i = 1, \ldots, r$. Since v was an arbitrary member of V, this shows that $V = U_1 + U_2 + \cdots + U_r$.

Now we need to show that V is actually a *direct* sum of the subspaces $\{U_i\}$, which amounts to showing that the expression (5.1) for v is *unique*. Suppose that we have any expression

$$v = u_1' + u_2' + \cdots + u_r', \quad \text{where } u_i' \in U_i \text{ for } i = 1, \ldots, r. \tag{5.2}$$

We observed earlier that if a vector u is in $\mathrm{Im}(\pi_i)$ then $\pi_i(u) = u$. Now observe that if u is in $\mathrm{Im}(\pi_j)$, for some $j \neq i$, then $\pi_j(u) = u$ and hence $\pi_i(u) = \pi_i(\pi_j(u)) = \pi_j\pi_i(u) = \mathbf{0}$.

With these observations in mind, apply π_i to (5.1) to obtain $\pi_i(v) = u_i$ and apply π_i to (5.2) to obtain $\pi_i(v) = u_i'$. We see that $u_i' = u_i$, and it follows that the expression for v is unique. □

There is a converse to the above.

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

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

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

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

Proof. (Sketch.) Since $V = U_1 \oplus U_2 \oplus \cdots \oplus U_r$, any vector $v \in V$ has a unique expression as

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

with $u_i \in U_i$ for $i = 1, \ldots, r$. Then we may define $\pi_i(v) = u_i$, for $i = 1, \ldots, r$. In a similar manner to the Proof of Proposition 5.5, we may check that $\{\pi_i\}$ are projections with the required properties. (Try this!) □

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

5.2 Linear maps and matrices

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

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

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

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

Proposition 5.6. *Let α be a linear map on V which is represented by the matrix A relative to a basis $\mathcal{B}$, and by the matrix A' relative to a basis $\mathcal{B}'$. Let $P = P_{\mathcal{B},\mathcal{B}'}$ be the transition matrix between the two bases. Then*

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

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

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

Thus similarity is an equivalence relation, and

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

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

5.3 Eigenvalues and eigenvectors

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

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

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

Example 5.9. Let

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

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

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

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

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

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

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

5.4 Diagonalisability

Some linear maps have a particularly simple representation by matrices.

Definition 5.10. The linear map α on V is *diagonalisable* if there is a basis of V relative to which the matrix representing α is a diagonal matrix.

Suppose that $v_1, \dots, v_n$ is such a basis showing that α is diagonalisable, and that $A = (a_{ij})$ is the matrix representing α in this basis. Since $a_{ij} = 0$ whenever $i \neq j$, we have $\alpha(v_i) = a_{ii}v_i$ for $i = 1, \dots, n$. Thus, the basis vectors are eigenvectors. Conversely, if we have a basis of eigenvectors, then the matrix representing α is diagonal. So:

Proposition 5.11. *The linear map α on V is diagonalisable if and only if there is a basis of V consisting of eigenvectors of α.*

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

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

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

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

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

$$c_1v_1 + \cdots + c_rv_r = \mathbf{0}, \tag{5.3}$$

with coefficients c_i not all zero. Some of these coefficients may be zero; choose a relation with the smallest number of non-zero coefficients. It is clear that there must be at least two non-zero coefficients. Suppose that $c_1 \neq 0$. (If $c_1 = 0$ just re-number the eivenvectors and their coefficients.) Now, applying α to both sides of (5.3) and using the fact that $\alpha(v_i) = \lambda_i v_i$, we get

$$\alpha(c_1v_1 + \cdots + c_rv_r) = c_1\alpha(v_1) + \cdots + c_r\alpha(v_r) = c_1\lambda_1v_1 + \cdots + c_r\lambda_rv_r = \mathbf{0}.$$

Subtracting λ_1 times equation (5.3) from the last equation we get

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

Now this equation has one fewer non-zero coefficient than the one we started with, which was assumed to have the smallest possible number. And since we started with at least two non-zero coefficients, not all the coefficients in this new identity are zero. So the linear dependency (5.3) is not minimal, contrary to our assumption. So the eigenvectors must have been linearly independent. □

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

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

(a) α is diagonalisable;

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

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

Proof. (a) $\Rightarrow$ (b). If α is diagonalisable, then there is a basis of V composed of eigenvectors of α. Each of these basis vectors lies in one of the eigenspaces; thus, $V = E(\lambda_1, \alpha) + \cdots + E(\lambda_r, \alpha)$. We need to show that this sum is actually a direct sum. If some vector $v \in V$ may be expressed in two different ways $u_1 + \cdots + u_r = u'_1 + \cdots + u'_r$, with $u_i, u'_i \in E(\lambda_i, \alpha)$, for $i = 1, \dots, r$, then $(u_1 - u'_1) + \cdots + (u_r - u'_r) = \mathbf{0}$. Each of these terms must be zero, otherwise we would have a non-trivial linear dependency between eigenvectors with distinct eigenvectors, which is disallowed by Lemma 5.13.

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

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

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

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

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

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

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

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

Example 5.15. Continuing our previous exercise, our matrix $A = \begin{bmatrix} -6 & 6 \\ -12 & 11 \end{bmatrix}$ is diagonalisable, since the eigenvectors $\begin{bmatrix} 3 \\ 4 \end{bmatrix}$ and $\begin{bmatrix} 2 \\ 3 \end{bmatrix}$ are linearly independent, and so form a basis for $\mathbb{R}$. Indeed, we see that

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

so that $AP = PD$ where

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

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

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

Furthermore, we can find two projection matrices as follows:

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

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

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

Proposition 5.16. *Let*

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

be the expression for the diagonalisable matrix A in terms of projections Π_i satisfying the conditions of Theorem 5.14, that is, $\sum_{i=1}^{r} \Pi_i = I$ and $\Pi_i\Pi_j = O$ for $i \neq j$. Then

(a) for any non-negative integer m, we have

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

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

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

As usual, we employ the convention $A^0 = I$ and $\lambda_i^0 = 1$ (even when $\lambda_i = 0$).

Proof. (a) The proof is by induction on m, the base case being

$$A^0 = I = \sum_{i=1}^{r} \Pi_i = \sum_{i=1}^{r} \lambda_i^0 \Pi_i.$$

Suppose now that the result holds for $m = k - 1$. Then

$$\begin{aligned} A^k &= A^{k-1}A \\ &= \left(\sum_{i=1}^{r} \lambda_i^{k-1}\Pi_i\right)\left(\sum_{j=1}^{r} \lambda_j \Pi_j\right). \end{aligned}$$

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

(b) Suppose $f(x) = \sum_{m=0}^{d} a_m x^m$ is a polynomial of degree d. We obtain the expression for $f(A)$ by multiplying the identity from part (a) by a_m and summing over m:

$$f(A) = \sum_{m=0}^{d} a_m A^m = \sum_{m=0}^{d} a_m \sum_{i=1}^{r} \lambda_i^m \Pi_i = \sum_{i=1}^{r} \Pi_i \sum_{m=0}^{d} a_m \lambda_i^m = \sum_{i=1}^{r} f(\lambda_i)\Pi_i.$$

□

5.5 Characteristic and minimal polynomials

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

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

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

Definition 5.17. (a) The *determinant* $\det(\alpha)$ of a linear map $\alpha : V \to V$ is the determinant of any matrix representing α.

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

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

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

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

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

Proposition 5.18. *For any linear map α on V, its minimal polynomial $m_\alpha(x)$ divides its characteristic polynomial $p_\alpha(x)$ (as polynomials).*

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

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

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

Theorem 5.19. *Let α be a linear map on V. Then the following conditions are equivalent for an element λ of $\mathbb{K}$:*

(a) λ is an eigenvalue of α;

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

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

Example 5.20. This gives us a recipe to find the eigenvalues of α: take a matrix A representing α; write down its characteristic polynomial $p_A(x) = \det(xI - A)$; and find the roots of this polynomial. In our earlier example,

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

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

Proof of Theorem 5.19. (a) $\Rightarrow$ (c). Let λ be an eigenvalue of α with eigenvector v. We have $\alpha(v) = \lambda v$. By induction, $\alpha^k(v) = \lambda^k v$ for any k, and so $f(\alpha)(v) = f(\lambda)\,v$ for any polynomial f. Choosing $f = m_\alpha$, we have $m_\alpha(\alpha)(v) = m_\alpha(\lambda)\,v$. But $m_\alpha(\alpha) = 0$ by definition, so $m_\alpha(\lambda)\,v = \mathbf{0}$. Since $v \neq \mathbf{0}$, we have $m_\alpha(\lambda) = 0$, as required.

(c) $\Rightarrow$ (b). Suppose that λ is a root of $m_\alpha(x)$. Then $(x - \lambda)$ divides $m_\alpha(x)$. But $m_\alpha(x)$ divides $p_\alpha(x)$, by Proposition 5.18, so $(x - \lambda)$ divides $p_\alpha(x)$, whence λ is a root of $p_\alpha(x)$.

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

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

Theorem 5.21. *The linear map α on V is diagonalisable if and only if its minimal polynomial is the product of distinct linear factors, that is, its roots all have multiplicity* 1.

Bonus material

The following proof is beyond the scope of the module and is non-examinable.

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

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

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

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

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

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

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

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

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

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

End of bonus material

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

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

Example 5.22. Consider first the matrix

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

The characteristic polynomial is

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

The polynomial $p_A(x)$ does not factor further over $\mathbb{R}$, as two of the roots are complex. So, viewed as a linear map on $\mathbb{R}^3$, the matrix A does not diagonalise.

This problem can be fixed by extending the field to the complex numbers $\mathbb{C}$. Then the characteristic polynomial is a product of linear factors, namely, $p_A(x) = (x-1)(x-\omega)(x-\omega^2)$, where $\omega = -\frac{1}{2} + \frac{\sqrt{3}}{2}i$. (Note that 1, ω and ω^2 are the cube roots of unity.) By Theorem 5.19, the minimal polynomial $m_A(x)$ divides $p_A(x)$ and hence $m_A(x)$ also is a product of distinct linear factors. Thus, viewed as a linear map on $\mathbb{C}^3$, the matrix A is diagonalisable, and its diagonal form is

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

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

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

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

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

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

Its characteristic polynomial is $p_B(x) = (x-2)^2(x-1)$. The minimal polynomial divides $p_B(x)$ and has the same roots, so the possibilties are either $m_B(x) = (x-2)(x-1)$ or

$m_B(x) = (x-2)^2(x-1)$. Can it be the former? Evaluating $(B-2I)(B-I)$ we find that

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

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

5.6 Jordan form

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

Definition 5.23. (a) A *Jordan block* $J(n, \lambda)$ is a matrix of the form

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

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

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

Theorem 5.24. *Over $\mathbb{C}$, any matrix is similar to a matrix in Jordan form; that is, any linear map can be represented by a matrix in Jordan form relative to a suitable basis. Moreover, the Jordan form of a matrix or linear map is unique apart from putting the Jordan blocks in a different order on the diagonal.*

Remark 5.25. A matrix over $\mathbb{C}$ is diagonalisable if and only if all the Jordan blocks in its Jordan form have size 1.

Example 5.26. Any 3×3 matrix over $\mathbb{C}$ is similar to one of

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

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

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

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

5.7 Trace

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

Definition 5.27. The *trace* $\mathrm{Tr}(A)$ of a square matrix A is the sum of its diagonal entries.

Proposition 5.28. *(a) For any two $n \times n$ matrices A and B, we have $\mathrm{Tr}(AB) = \mathrm{Tr}(BA)$.*

(b) Similar matrices have the same trace.

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

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

by the rules for matrix multiplication.

For part (b), we just observe that $\mathrm{Tr}(P^{-1}AP) = \mathrm{Tr}(APP^{-1}) = \mathrm{Tr}(AI) = \mathrm{Tr}(A)$. □

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

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

Proposition 5.29. *Let $\alpha : V \to V$ be a linear map, where $\dim(V) = n$, and let p_α be the characteristic polynomial of α, a polynomial of degree n with leading term x^n.*

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

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

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

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

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

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

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

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

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