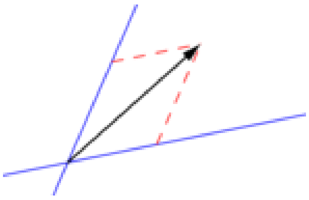

MA22020: Advanced linear algebra

Notes by Fran Burstall

Corrections by:

Roman Filippov
Alfie Gidney
Gregory Sankaran
Elise Tullett
Macsen Wyn

Contents

Chapter 1

Linear algebra: key concepts

Let us warm up by revising some of the key ideas from Algebra 1B.

1.1 Vector spaces

Recall from Algebra 1B, §1.1:

Definition. A *vector space V over a field* $\mathbb{F}$ is a set V with two operations:

addition $V \times V \to V : (v, w) \mapsto v + w$ such that:

- $v + w = w + v$, for all $v, w \in V$;
- $u + (v + w) = (u + v) + w$, for all $u, v, w \in V$;
- there is a *zero element* $0 \in V$ for which $v + 0 = v = 0 + v$, for all $v \in V$;
- each element $v \in V$ has an *additive inverse* $-v \in V$ for which $v + (-v) = 0 = (-v) + v$.

In fancy language, V with addition is an *abelian group*.

scalar multiplication $\mathbb{F} \times V \to V : (\lambda, v) \mapsto \lambda v$ such that

- $(\lambda + \mu)v = \lambda v + \mu v$, for all $v \in V$, $\lambda, \mu \in \mathbb{F}$.
- $\lambda(v + w) = \lambda v + \lambda w$, for all $v, w \in V$, $\lambda \in \mathbb{F}$.
- $(\lambda\mu)v = \lambda(\mu v)$, for all $v \in V$, $\lambda, \mu \in \mathbb{F}$.
- $1v = v$, for all $v \in V$.

We call the elements of $\mathbb{F}$ *scalars* and those of V *vectors*.

Examples.

(1) Take $V = \mathbb{F}$, the field itself, with addition and scalar multiplication the field addition and multiplication.

(2) $\mathbb{F}^n$, the n-fold Cartesian product of $\mathbb{F}$ with itself, with component-wise addition and scalar multiplication:

$$\begin{aligned}(\lambda_1, \dots, \lambda_n) + (\mu_1, \dots, \mu_n) &:= (\lambda_1 + \mu_1, \dots, \lambda_n + \mu_n)\\ \lambda(\lambda_1, \dots, \lambda_n) &:= (\lambda\lambda_1, \dots, \lambda\lambda_n).\end{aligned}$$

(3) Let $M_{m\times n}(\mathbb{F})$ denote the set of m by n matrices (thus m rows and n columns) with entries in $\mathbb{F}$. This is a vector space under entry-wise addition and scalar multiplication.
Special cases are the vector spaces of *column vectors* $M_{n\times 1}(\mathbb{F})$ and *row vectors* $M_{1\times n}(\mathbb{F})$. In computations, we often identify $\mathbb{F}^n$ with $M_{n\times 1}(\mathbb{F})$ by associating $x = (x_1, \dots, x_n) \in \mathbb{F}^n$ with the column vector

$$\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}.$$

1.2 Subspaces

Definition. A *vector* (or *linear*) *subspace* of a vector space V over $\mathbb{F}$ is a non-empty subset $U \subseteq V$ which is closed under addition and scalar multiplication: whenever $u, u_1, u_2 \in U$ and $\lambda \in \mathbb{F}$, then $u_1 + u_2 \in U$ and $\lambda u \in U$.

In this case, we write $U \leq V$.

Say that U is *trivial* if $U = \{0\}$ and *proper* if $U \neq V$.

Of course, U is now a vector space in its own right using the addition and scalar multiplication of V.

Exercise.[1] $U \subseteq V$ is a subspace if and only if U satisfies the following conditions:

(1) $0 \in U$;

(2) For all $u_1, u_2 \in U$ and $\lambda \in \mathbb{F}$, $u_1 + \lambda u_2 \in U$.

This gives a efficient recipe for checking when a subset is a subspace.

1.3 Bases

Definitions. Let $v_1, \dots, v_n$ be a list of vectors in a vector space V.

(1) The *span* of $v_1, \dots, v_n$ is

$$\operatorname{span}\{v_1, \dots, v_n\} := \{\lambda_1 v_1 + \dots + \lambda_n v_n \mid \lambda_i \in \mathbb{F}, 1 \leq i \leq n\} \leq V.$$

(2) $v_1, \dots, v_n$ *span* V (or *are a spanning list for* V) if $\operatorname{span}\{v_1, \dots, v_n\} = V$.

(3) $v_1, \dots, v_n$ are *linearly independent* if, whenever $\lambda_1 v_1 + \dots + \lambda_n v_n = 0$, then each $\lambda_i = 0$, $1 \leq i \leq n$, and *linearly dependent* otherwise.

(4) $v_1, \dots, v_n$ is a *basis* for V if they are linearly independent and span V.

Remark. Notice that any re-ordering of a basis is also a (different) basis. Example: if $v_1, v_2 v_3$ is a basis, so is v_2, v_1, v_3 and so on.

Definition. A vector space is *finite-dimensional* if it admits a finite list of vectors as basis and *infinite-dimensional* otherwise.

If V is finite-dimensional, the *dimension* of V, $\dim V$, is the number of vectors in a (any) basis of V.

Terminology. Let $v_1, \dots, v_n$ be a list of vectors.

(1) A vector of the form $\lambda_1 v_1 + \dots + \lambda_n v_n$ is called a *linear combination of the* v_i.

(2) An equation of the form $\lambda_1 v_1 + \dots + \lambda_n v_n = 0$ is called a *linear relation on the* v_i.

Example. Some lucky vector spaces come with a natural choice of basis. For instance, define $e_i := (0, \dots, 1, \dots, 0) \in \mathbb{F}^n$, $1 \leq i \leq n$ with a single 1 in the i-th place and zeros elsewhere. Then $e_1, \dots, e_n$ is a basis of $\mathbb{F}^n$ called the *standard basis*

[1] Question 1 on sheet 1.

1.3.1 Useful facts

A very useful fact about bases that we shall use many times was proved in Algebra 1B:

Proposition 1.1 (Algebra 1B, Corollary 1.5.7)**.** *Any linearly independent list of vectors in a finite-dimensional vector space can be extended to a basis.*

Here is another helpful result :

Proposition 1.2 (Algebra 1B, Corollary 1.5.6)**.** *Let V be a finite-dimensional vector space and $U \leq V$. Then*

$$\dim U \leq \dim V$$

with equality if and only if $U = V$.

1.4 Linear maps

Definitions. A map $\phi\colon V \to W$ of vector spaces over $\mathbb{F}$ is a *linear map* (or, in older books, *linear transformation*) if

$$\begin{aligned} \phi(v+w) &= \phi(v) + \phi(w) \\ \phi(\lambda v) &= \lambda\phi(v), \end{aligned}$$

for all $v, w \in V$, $\lambda \in \mathbb{F}$.

The *kernel* of ϕ is $\ker\phi := \{v \in V \mid \phi(v) = 0\} \leq V$.

The *image* of ϕ is $\operatorname{im}\phi := \{\phi(v) \mid v \in V\} \leq W$.

Remark. ϕ is linear if and only if

$$\phi(v + \lambda w) = \phi(v) + \lambda\phi(w),$$

for all $v, w \in V$, $\lambda \in \mathbb{F}$, which has the virtue of being only one thing to prove.

Examples.

(1) $A \in M_{m\times n}(\mathbb{F})$ determines a linear map $\phi_A : \mathbb{F}^n \to \mathbb{F}^m$ by $\phi_A(x) = y$ where, for $1 \leq i \leq m$,

$$y_i = \sum_{j=1}^{n} A_{ij}x_j.$$

Otherwise said, y is given by matrix multiplication: $\mathbf{y} = A\mathbf{x}$.

(2) For any vector space V, the identity map $\mathrm{id}_V : V \to V$ is linear.

(3) If $\phi : V \to W$ and $\psi : W \to U$ are linear then so is $\psi \circ \phi : V \to U$.

Definition. A linear map $\phi : V \to W$ is a *(linear) isomorphism* if there is a linear map $\psi : W \to V$ such that

$$\psi \circ \phi = \mathrm{id}_V, \qquad \phi \circ \psi = \mathrm{id}_W .$$

If there is an isomorphism $V \to W$, say that V and W are isomorphic and write $V \cong W$.

In Algebra 1B, we saw:

Lemma 1.3 (Algebra 1B, lemma 1.3.3 (4))**.** *$\phi : V \to W$ is an isomorphism if and only if ϕ is a linear bijection (and then $\psi = \phi^{-1}$).*

Notation. For vector spaces V, W over $\mathbb{F}$, denote by $L_{\mathbb{F}}(V, W)$ (or simply $L(V, W)$) the set $\{\phi : V \to W \mid \phi \text{ is linear}\}$ of linear maps from V to W.

Theorem 1.4 (Linearity is a linear condition)**.** *$L(V, W)$ is a vector space under pointwise addition and scalar multiplication. Thus*

$$\begin{aligned}(\phi + \psi)(v) &:= \phi(v) + \psi(v)\\ (\lambda\phi)(v) &:= \lambda\phi(v),\end{aligned}$$

for all $\phi, \psi \in L(V, W)$, $v \in V$ and $\lambda \in \mathbb{F}$.

Proof. There is a lot to do here but it is all easy. First we must show that $\phi + \psi$, as defined above, really is a linear map when $\phi, \psi \in L(V, W)$:

$$\begin{aligned}(\phi + \psi)(v + \lambda w) &= \phi(v + \lambda w) + \psi(v + \lambda w)\\ &= \phi(v) + \lambda\phi(w) + \psi(v) + \lambda\psi(w)\\ &= (\phi(v) + \psi(v)) + \lambda(\phi(w) + \psi(w))\\ &= (\phi + \psi)(v) + \lambda(\phi + \psi)(w),\end{aligned}$$

for all $v, w \in V$, $\lambda \in \mathbb{F}$. Here the first and last equalities are just the definition of pointwise addition while the middle equalities come from the linearity of ϕ, ψ and the vector space axioms of W.

Similarly, it is a simple exercise to see that if $\mu \in \mathbb{F}$ and $\phi \in L(V, W)$ then $\mu\phi$ is also linear.

Now we need a zero element for our proposed vector space: observe that the zero map $0 : v \mapsto 0 \in W$ is linear:

$$0(v + \lambda w) = 0 = 0 + \lambda 0 = 0(v) + \lambda 0(w).$$

We also define $-\phi$ by

$$(-\phi)(v) = -\phi(v),$$

for $v \in V$ and check that it is also linear.

Finally, we must check all the vector space axioms which all follow from those of W. For example, for any $v \in V$,

$$(\phi + \psi)(v) = \phi(v) + \psi(v) = \psi(v) + \phi(v) = (\psi + \phi)(v),$$

so that $\phi + \psi = \psi + \phi$. The remaining axioms are left as a (rather boring) exercise. □

A linear map of a finite-dimensional vector space is completely determined by its action on a basis. More precisely:

Proposition 1.5 (Extension by linearity)**.** *Let V, W be vector spaces over $\mathbb{F}$. Let $v_1, \dots, v_n$ be a basis of V and $w_1, \dots, w_n$* any *vectors in W.*

Then there is a unique *$\phi \in L(V, W)$ such that*

$$\phi(v_i) = w_i, \qquad 1 \leq i \leq n. \tag{1.1}$$

Proof. We need to prove that such a ϕ exists and that there is only one. We prove existence first.

Let $v \in V$. From Algebra 1B[2],we know there are unique $\lambda_1, \dots, \lambda_n \in \mathbb{F}$ for which

$$v = \lambda_1 v_1 + \cdots + \lambda_n v_n$$

and so we define $\phi(v)$ to be the only thing it could be:

$$\phi(v) := \lambda_1 w_1 + \cdots + \lambda_n w_n.$$

Let us show that this ϕ does the job. First, with $\lambda_i = 1$ and $\lambda_j = 0$, for $i \neq j$, we see that

$$\phi(v_i) = \sum_{j \neq i} 0 w_j + 1 w_i = w_i$$

[2] Proposition 1.4.4

so that (1.1) holds. Now let us see that ϕ is linear: let $v, w \in V$ with

$$\begin{aligned} v &= \lambda_1 v_1 + \cdots + \lambda_n v_n \\ w &= \mu_1 v_1 + \cdots + \mu_n v_n. \end{aligned}$$

Then, for $\lambda \in \mathbb{F}$,

$$v + \lambda w = (\lambda_1 + \lambda\mu_1)v_1 + \cdots + (\lambda_n + \lambda\mu_n)v_n$$

whence

$$\begin{aligned} \phi(v + \lambda w) &= (\lambda_1 + \lambda\mu_1)w_1 + \cdots + (\lambda_n + \lambda\mu_n)w_n \\ &= (\lambda_1 w_1 + \cdots + \lambda_n w_n) + \lambda(\mu_1 w_1 + \cdots + \mu_n w_n) \\ &= \phi(v) + \lambda\phi(w). \end{aligned}$$

For uniqueness, suppose that $\phi, \phi' \in L(V, W)$ both satisfy (1.1). Let $v \in V$ and write $v = \lambda_1 v_1 + \cdots + \lambda_n v_n$. Then

$$\begin{aligned} \phi(v) &= \lambda_1\phi(v_1) + \cdots + \lambda_n\phi(v_n) \\ &= \lambda_1 w_1 + \cdots + \lambda_n w_n \\ &= \lambda_1\phi'(v_1) + \cdots + \lambda_n\phi'(v_n) \\ &= \phi'(v), \end{aligned}$$

where the first and last equalities come from the linearity of ϕ, ϕ' and the middle two from (1.1) for first ϕ and then ϕ'. We conclude that $\phi = \phi'$ and we are done. □

Remark. In the context of Theorem 1.5, ϕ is an isomorphism if and only if $w_1, \dots, w_n$ is a basis for W (exercise[3]!).

Among the most important results in Algebra 1B is the famous rank-nullity theorem:

Theorem 1.6 (Rank-nullity)**.** *Let $\phi : V \to W$ be linear with V finite-dimensional. Then*

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

Using this, together with the observation that ϕ is injective if and only if $\ker \phi = \{0\}$ and surjective if and only if $\operatorname{im} \phi = W$ we have:

Proposition 1.7. *Let $\phi : V \to W$ be linear with V, W finite-dimensional vector spaces of the same dimension: $\dim V = \dim W$.*

Then the following are equivalent:

(1) *ϕ is injective.*

(2) *ϕ is surjective.*

(3) *ϕ is an isomorphism.*

[3] This is question 6 on exercise sheet 1.

Chapter 2

Sums and quotients

We will discuss various ways of building new vector spaces out of old ones.

Convention. In this chapter, all vector spaces are over the same field $\mathbb{F}$ unless we say otherwise.

2.1 Sums of subspaces

Definition. Let $V_1, \dots, V_k \leq V$. The *sum* $V_1 + \dots + V_k$ is the set

$$V_1 + \dots + V_k := \{v_1 + \dots + v_k \mid v_i \in V_i, 1 \leq i \leq k\}.$$

$V_1 + \dots + V_k$ is the smallest subspace of V that contains each V_i. More precisely:

Proposition 2.1. *Let $V_1, \dots, V_k \leq V$. Then*

(1) $V_1 + \dots + V_k \leq V$.

(2) If $W \leq V$ and $V_1, \dots, V_k \leq W$ then $V_1, \dots, V_k \leq V_1 + \dots + V_k \leq W$.

Proof. It suffices to prove (2) since (1) then follows by taking $W = V$.

For (2), first note that $V_1 + \dots + V_k$ is a subset of W: if $v_i \in V_i$ then $v_i \in W$ so that $v_1 + \dots + v_k \in W$ since W is closed under addition.

Now observe that each $V_i \leq V_1 + \dots + V_k$ since we can write any $v_i \in V_i$ as $0 + \dots + v_i + \dots + 0 \in V_1 + \dots + V_k$. In particular, $0 \in V_1 + \dots + V_k$.

Finally, we show that $V_1 + \dots + V_k$ is a subspace. If $v_1 + \dots + v_k, w_1 + \dots + w_k \in V_1 + \dots + V_k$, with $v_i, w_i \in V_i$, for all i, and $\lambda \in \mathbb{F}$ then

$$(v_1 + \dots + v_k) + \lambda(w_1 + \dots + w_k) = (v_1 + \lambda w_1) + \dots + (v_k + \lambda w_k) \in V_1 + \dots + V_k$$

since each $v_i + \lambda w_i \in V_i$. □

Remark. The union $\bigcup_{i=1}^{k} V_i$ is almost never a subspace of V so we use sums as a substitute for unions in Linear Algebra.

2.2 Direct sums

Let $V_1, \dots, V_k \leq V$. Any $v \in V_1 + \dots + V_k$ can be written

$$v = v_1 + \dots + v_k,$$

with each $v_i \in V_i$. We distinguish the case where the v_i are *unique*.

Definition. Let $V_1, \dots, V_k \leq V$. The sum $V_1 + \cdots + V_k$ is *direct* if each $v \in V_1 + \cdots + V_k$ can be written

$$v = v_1 + \cdots + v_k$$

in only one way, that is, for unique $v_i \in V_i$, $1 \leq i \leq k$.

In this case, we write $V_1 \oplus \cdots \oplus V_k$ instead of $V_1 + \cdots + V_k$.

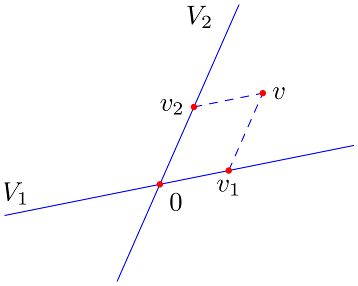

Figure 2.1: $\mathbb{R}^2 = V_1 \oplus V_2$

Example. Define $V_1, V_2 \leq \mathbb{F}^3$ by

$$\begin{aligned} V_1 &= \{(x_1, x_2, 0) \mid x_1, x_2 \in \mathbb{F}\} \\ V_2 &= \{(0, 0, x_3) \mid x_3 \in \mathbb{F}\}. \end{aligned}$$

Then $\mathbb{F}^3 = V_1 \oplus V_2$.

When is a sum direct? We begin with a useful reformulation of the property.

Proposition 2.2. *Let $V_1, \dots, V_k \leq V$. Then $V_1 + \cdots + V_k$ is direct if and only if whenever $v_1 + \cdots + v_k = 0$, with $v_i \in V_i$, $1 \leq i \leq k$, then $v_i = 0$, for all $1 \leq i \leq k$.*

Proof. Suppose that $V_1 + \cdots + V_k$ is direct and let $v_1 + \cdots + v_k = 0$, with each $v_i \in V_i$. We can also write $0 = 0 + \cdots + 0$ so that the uniqueness in the direct sum property forces each $v_i = 0$.

Conversely, if the "zero sum" property holds, suppose that, for some $v \in V_1 + \cdots + V_k$, we have

$$v = v_1 + \cdots + v_k = w_1 + \cdots + w_k,$$

with each $v_i, w_i \in V_i$. Then

$$0 = v - v = (v_1 - w_1) + \cdots + (v_k - w_k)$$

and each $v_i - w_i \in V_i$ so the zero sum property gives $v_i = w_i$. We conclude that the sum is direct. □

For the case of two summands this gives a very simple way to decide if a sum is direct:

Proposition 2.3. *Let $V_1, V_2 \leq V$. Then $V_1 + V_2$ is direct if and only if $V_1 \cap V_2 = \{0\}$.*

Proof. Suppose first that $V_1 + V_2$ is direct and let $v \in V_1 \cap V_2$. Then

$$0 = v + (-v)$$

and $v \in V_1$, $-v \in V_2$ so that $v = -v = 0$ by Theorem 2.2.

Conversely, suppose that $V_1 \cap V_2 = \{0\}$ and that $v_1 + v_2 = 0$, with $v_i \in V_i$, $i = 1, 2$. Then $v_1 = -v_2 \in V_1 \cap V_2 = \{0\}$ so that $v_1 = v_2 = 0$. Thus $V_1 + V_2$ is direct by Theorem 2.2. □

The special case $V = V_1 + V_2$ is important and deserves some terminology:

Definition. Let $V_1, V_2 \leq V$. V is the *(internal) direct sum of* V_1 *and* V_2 if $V = V_1 \oplus V_2$.

In this case, say that V_2 is a *complement* of V_1 (and V_1 is a complement of V_2).

Warning. This notion of the complement of the subspace V_1 has *nothing at all* to do with the set-theoretic complement $V \setminus V_1$ which is never a subspace.

Remarks.

(1) From Theorem 2.3, we see that $V = V_1 \oplus V_2$ if and only if $V = V_1 + V_2$ and $V_1 \cap V_2 = \{0\}$. Many people take these latter properties as the *definition* of internal direct sum.

(2) There is a related notion of *external* direct sum that we will not discuss.

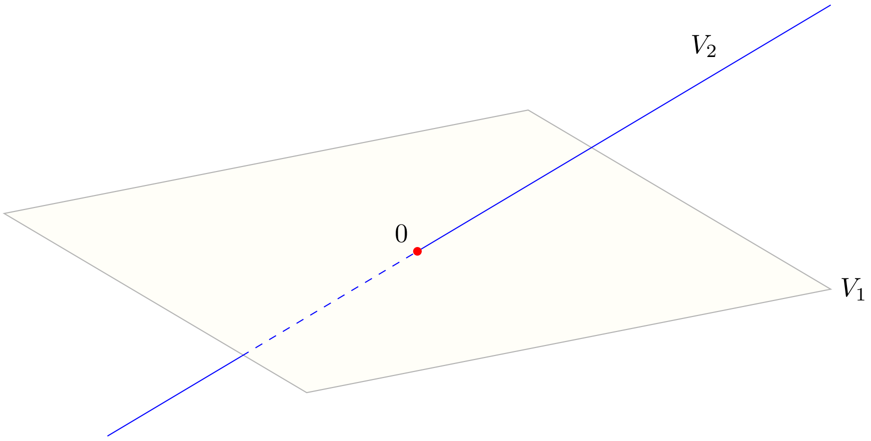

Figure 2.2: $\mathbb{R}^3$ as a direct sum of a line and a plane

When there are many summands, the condition that a sum be direct is a little more involved:

Proposition 2.4. *Let $V_1, \ldots, V_k \leq V$, $k \geq 2$. Then the sum $V_1 + \cdots + V_k$ is direct if and only if, for each $1 \leq i \leq k$, $V_i \cap (\sum_{j \neq i} V_j) = \{0\}$.*

Proof. This is an exercise in imitating the proof of Theorem 2.3. □

Remark. This is a much stronger condition than simply asking that each $V_i \cap V_j = \{0\}$, for $i \neq j$.

2.2.1 Induction from two summands

A convenient way to analyse direct sums with many summands is to induct from the two summand case. For this, we need:

Lemma 2.5. *Let $V_1, \ldots, V_k \leq V$. Then $V_1 + \cdots + V_k$ is direct if and only if $V_1 + \cdots + V_{k-1}$ is direct and $(V_1 + \cdots + V_{k-1}) + V_k$ (two summands) is direct.*

Proof. Suppose first that $V_1 + \cdots + V_k$ is direct. We use Theorem 2.2 to see that $V_1 + \cdots + V_{k-1}$ is direct: let $v_1 + \cdots + v_{k-1} = 0$ with each $v_i \in V_i$, $1 \leq i \leq k-1$. Write this as $v_1 + \cdots + v_k = 0$ where $v_k = 0 \in V_k$ and deduce that each $v_i = 0$, $1 \leq i \leq k-1$.

Again, if $v = v_1 + \cdots + v_{k-1} \in V_1 + \cdots + V_{k-1}$ and $v_k \in V_k$ with $v + v_k = 0$, then we have $v_1 + \cdots + v_k = 0$ so that each $v_i = 0$ whence $v = 0$ also. Now Theorem 2.2 tells us that $(V_1 + \cdots + V_{k-1}) + V_k$ is direct.

Conversely, suppose that both $V_1 + \cdots + V_{k-1}$ and $(V_1 + \cdots + V_{k-1}) + V_k$ are direct and that $v_1 + \cdots + v_k = 0$, with each $v_i \in V_i$. Let $v = v_1 + \cdots + v_{k-1} \in V_1 + \cdots + V_{k-1}$ so that $v + v_k = 0$. Now Theorem 2.2 and the directness of $(V_1 + \cdots + V_{k-1}) + V_k$ tell us that $v = v_k = 0$. Thus $v_1 + \cdots + v_{k-1} = 0$ and a final application of Theorem 2.2 yields $v_i = 0$, $1 \leq i \leq k-1$ since $V_1 + \cdots + V_{k-1}$ is direct. □

2.2.2 Direct sums, bases and dimension

When a sum is direct, bases of the summands fit together to give a basis of the sum:

Proposition 2.6. *Let $V_1, V_2 \leq V$ be subspaces with bases $\mathcal{B}_1\colon v_1, \dots, v_k$ and $\mathcal{B}_2\colon w_1, \dots, w_l$. Then $V_1 + V_2$ is direct if and only if the* concatenation[1] *$\mathcal{B}_1\mathcal{B}_2\colon v_1, \dots, v_k, w_1, \dots, w_l$ is a basis of $V_1 + V_2$.*

Proof. Clearly $\mathcal{B}_1\mathcal{B}_2$ spans $V_1 + V_2$ and so will be a basis exactly when it is linearly independent.

Suppose that $V_1 + V_2$ is direct and that we have a linear relation $\sum_{i=1}^{k} \lambda_i v_i + \sum_{j=1}^{l} \mu_j w_j = 0$. Then Theorem 2.2 yields

$$\sum_{i=1}^{k} \lambda_i v_i = \sum_{j=1}^{l} \mu_j w_j = 0$$

so that all the λ_i and μ_j vanish since $\mathcal{B}_1$ and $\mathcal{B}_2$ are linearly independent. We conclude that $\mathcal{B}_1\mathcal{B}_2$ is linearly independent and so a basis.

Conversely, if $\mathcal{B}_1\mathcal{B}_2$ is a basis and $v + w = 0$ with $v \in V_1$ and $w \in V_2$, write $v = \sum_{i=1}^{k} \lambda_i v_i$ and $w = \sum_{j=1}^{l} \mu_j w_j$ to get a linear relation $\sum_{i=1}^{k} \lambda_i v_i + \sum_{j=1}^{l} \mu_j w_j = 0$. By linear independence of $\mathcal{B}_1\mathcal{B}_2$, all λ_i, μ_j vanish so that $v = w = 0$. Thus $V_1 + V_2$ is direct by Theorem 2.2. □

Again, this along with Theorem 2.5 and induction on k yields the many-summand version:

Corollary 2.7. *Let $V_1, \dots, V_k \leq V$ be finite-dimensional subspaces with $\mathcal{B}_i$ a basis of V_i, $1 \leq i \leq k$. Then $V_1 + \dots + V_k$ is direct if and only if the concatenation $\mathcal{B}_1 \dots \mathcal{B}_k$ is a basis for $V_1 + \dots + V_k$.*

Proof. Our induction hypothesis at k is that $V_1 + \dots + V_k$ is direct if and only if $\mathcal{B}_1 \dots \mathcal{B}_k$ is a basis for $V_1 + \dots + V_k$. This is vacuous at $k = 1$ so let us suppose it is true for k and examine the case $k + 1$.

First suppose that $V_1 + \dots + V_{k+1}$ is direct so that $V_1 + \dots + V_k$ and $(V_1 + \dots + V_k) + V_{k+1}$ are direct by Theorem 2.5. The induction hypothesis applies to both of these so that, first, $\mathcal{B}_1 \dots \mathcal{B}_k$ is a basis of $V_1 + \dots + V_k$ and then $(\mathcal{B}_1 \dots \mathcal{B}_k)\mathcal{B}_{k+1} = \mathcal{B}_1 \dots \mathcal{B}_{k+1}$ is a basis of $(V_1 + \dots + V_k) + V_{k+1} = V_1 + \dots + V_{k+1}$.

Conversely, if $\mathcal{B}_1 \dots \mathcal{B}_{k+1}$ is a basis of $V_1 + \dots + V_{k+1}$, $\mathcal{B}_1 \dots \mathcal{B}_k$ is linearly independent and so a basis of $V_1 + \dots + V_k$. By the induction hypothesis, we learn that $V_1 + \dots + V_k$ is direct. Similarly, we see that $(V_1 + \dots + V_k) + V_{k+1}$ is direct whence, by Theorem 2.5, $V_1 + \dots + V_{k+1}$ is direct.

This establishes the induction hypothesis at $k + 1$ and so the result is proved. □

From this we see that dimensions add over direct sums:

Corollary 2.8. *Let $V_1, \dots, V_k \leq V$ be subspaces of a finite-dimensional vector space V with $V_1 + \dots + V_k$ direct. Then*

$$\dim V_1 \oplus \dots \oplus V_k = \dim V_1 + \dots + \dim V_k.$$

Proof. Let $\mathcal{B}_i$ be basis for V_i so that $\mathcal{B}_1 \dots \mathcal{B}_k$ is a basis of $V_1 + \dots + V_k$ by Theorem 2.7. Then

$$\dim V_1 + \dots + V_l = |\mathcal{B}_1 \dots \mathcal{B}_k| = |\mathcal{B}_1| + \dots + |\mathcal{B}_k| = \dim V_1 + \dots + \dim V_k.$$

□

Exercise.[2] Prove the converse of Theorem 2.8: if $\dim V_1 + \dots + V_k = \dim V_1 + \dots + \dim V_k$, then the sum is direct.

[1] The concatenation of two lists is simply the list obtained by adjoining all entries in the second list to the first.

[2] Question 2 on sheet 2.

2.2.3 Complements

For finite-dimensional vector spaces, any subspace has a complement:

Proposition 2.9 (Complements exist)**.** *Let $U \leq V$, a finite-dimensional vector space. Then there is a complement to U.*

Proof. Let $\mathcal{B}_1 : v_1, \dots, v_k$ be a basis for U and so a linearly independent list of vectors in V. By Theorem 1.1, we can extend the list to get a basis $\mathcal{B} : v_1, \dots, v_n$ of V. Set $W = \operatorname{span}\{v_{k+1}, \dots, v_n\} \leq V$: this is a complement to U.

Indeed, $\mathcal{B}_2 : v_{k+1}, \dots, v_n$ is a basis for W and $\mathcal{B} = \mathcal{B}_1\mathcal{B}_2$ so that $V = U \oplus W$ by Theorem 2.6. □

In fact, as Figure 2.3 illustrates, there are many complements to a given subspace.

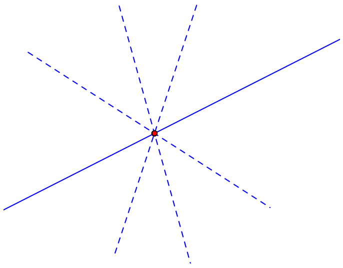

Figure 2.3: Each dashed line is a complement to the undashed subspace.

2.3 Quotients

Let $U \leq V$. We construct a new vector space from U and V which is an "abstract complement" to U. The elements of this vector space are equivalence classes for the following equivalence relation:

Definition. Let $U \leq V$. Say that $v, w \in V$ are *congruent modulo* U if $v - w \in U$. In this case, we write $v \equiv w \mod U$.

Warning. This is emphatically not the relation of congruence modulo an integer n that you studied in Algebra 1A: here the relation is between vectors in a vector space. However, both notions of congruence are examples of a general construction in group theory.

Lemma 2.10. *Congruence modulo U is an equivalence relation.*

Proof. Exercise[3]! □

Thus each $v \in V$ lies in exactly one equivalence class $[v] \subseteq V$.

What do these equivalence classes look like? Note that $w \equiv v \mod U$ if and only if $w - v \in U$ or, equivalently, $w = v + u$, for some $u \in U$.

Definition. For $v \in V$, $U \leq V$, the set $v + U := \{v + u \mid u \in U\} \subseteq V$ is called a *coset of* U and v is called a *coset representative* of $v + U$.

We conclude that the equivalence class of v modulo U is the coset $v + U$.

Remark. In geometry, cosets of vector subspaces are called *affine subspaces.* Examples include lines in $\mathbb{R}^2$ and lines and planes in $\mathbb{R}^3$ irrespective of whether they contain zero (as vector subspaces must).

[3] This is question 3 on exercise sheet 2.

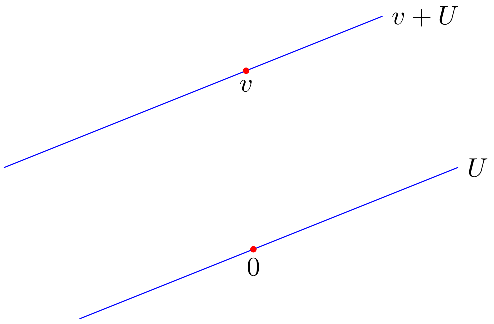

Figure 2.4: A subspace $U \leq \mathbb{R}^2$ and a coset $v+U$.

Example. Fibres of a linear map: let $\phi : V \to W$ be a linear map and let $w \in \operatorname{im}\phi$. Then the *fibre of ϕ over w* is defined by:

$$\phi^{-1}\{w\} := \{v \in V \mid \phi(v) = w\}.$$

Unless $w = 0$, this is not a linear subspace but notice that v, v' are in the same fibre if and only if $\phi(v) = \phi(v')$, or, equivalently, $\phi(v-v') = 0$ or $v - v' \in \ker\phi$. We conclude that the fibres of ϕ are exactly the cosets of $\ker\phi$:

$$\phi^{-1}\{w\} = v + \ker\phi,$$

for any $v \in \phi^{-1}\{w\}$.

We shall see below that any coset arises this way for a suitable ϕ.

Definition. Let $U \leq V$. The *quotient space V/U of V by U* is the set V/U, pronounced "V mod U", of cosets of U:

$$V/U := \{v + U \mid v \in V\}.$$

This is a subset of the *power set*[4] $\mathcal{P}(V)$ of V.

The *quotient map $q : V \to V/U$* is defined by

$$q(v) = v + U.$$

The quotient map q will be important to us. It has two key properties:

(1) q is surjective.

(2) $q(v) = q(v')$ if and only if $v \equiv v' \mod U$, that is, $v - v' \in U$.

We can add and scalar multiply cosets to make V/U into a vector space and q into a linear map:

Theorem 2.11. *Let $U \leq V$. Then, for $v, w \in V$, $\lambda \in \mathbb{F}$,*

$$\begin{aligned}(v+U) + (w+U) &:= (v+w) + U \\ \lambda(v+U) &:= (\lambda v) + U\end{aligned}$$

give well-defined operations of addition and scalar multiplication on V/U with respect to which V/U is a vector space and $q : V \to V/U$ is a linear map.

Moreover, $\ker q = U$ *and* $\operatorname{im} q = V/U$.

Proof. We phrase everything in terms of q to keep the notation under control. Since q surjects, we lose nothing by doing this: any element of V/U is of the form $q(v)$ for some $v \in V$.

[4] Recall from Algebra 1A that the power set of a set A is the set of all subsets of A.

With this understood, the proposed addition and scalar multiplication in V/U read

$$\begin{aligned} q(v)+q(w) &:= q(v+w) \\ \lambda q(v) &:= q(\lambda v) \end{aligned}$$

so that q is certainly linear so long as these operations make sense. Here the issue is that if $q(v)=q(v')$ and $q(w)=q(w')$, we must show that

$$q(v+w)=q(v'+w'), \qquad q(\lambda v)=q(\lambda v'). \tag{2.1}$$

However, in this case, we have $v-v' \in U$ and $w-w' \in U$ so that

$$\begin{aligned} (v+w)-(v'+w') &= (v-v')+(w-w') \in U \\ \lambda v - \lambda v' &= \lambda(v-v') \in U, \end{aligned}$$

since U is a subspace, and this establishes (2.1).

As for the vector space axioms, these follow from those of V. For example:

$$q(v)+q(w)=q(v+w)=q(w+v)=q(w)+q(v).$$

Here the first and third equalities are the definition of addition in V/U and the middle one comes from commutativity of addition in V. The zero element is $q(0)=0+U=U$ while the additive inverse of $q(v)$ is $q(-v)$.

The linearity of q comes straight from how we defined our addition and scalar multiplication while $v \in \ker q$ if and only if $q(v)=q(0)$ if and only if $v=v-0 \in U$ so that $\ker q = U$. □

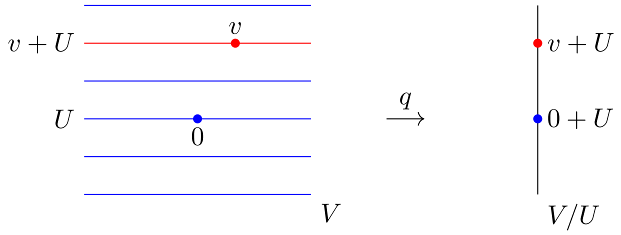

Figure 2.5: The quotient map q.

Corollary 2.12. *Let $U \leq V$. If V is finite-dimensional then so is V/U and*

$$\dim V/U = \dim V - \dim U.$$

Proof. Apply rank-nullity to q using $\ker q = U$ and $\operatorname{im} q = V/U$. □

Remark. Theorem 2.11 shows that:

(1) Any $U \leq V$ is the kernel of a linear map.

(2) Any coset $v+U$ is the fibre of a linear map: indeed

$$v+U=q^{-1}\{q(v)\}.$$

Commentary. Many people find the quotient space V/U difficult to think about: its elements are (special) subsets of V and this can be confusing.

An alternative, perhaps better way, to proceed is to concentrate instead on the *properties* of V/U in much that same way that, in Analysis, we deal with real numbers via the axioms of a complete ordered field without worrying too much what a real number actually is!

From this point of view, the quotient V/U of V by U is a vector space along with a linear map $q : V \to V/U$ such that

- q surjects;
- $\ker q = U$

and this is really all you need to know!

The content of Theorem 2.11, from this perspective, is simply that quotients exist!

Theorem 2.13 (First Isomorphism Theorem)**.** *Let $\phi : V \to W$ be a linear map of vector spaces.*

Then $V/\ker\phi \cong \operatorname{im}\phi$.

In fact, define $\bar{\phi} : V/\ker\phi \to \operatorname{im}\phi$ by

$$\bar{\phi}(q(v)) = \phi(v),$$

where $q : V \to V/\ker\phi$ is the quotient map.

Then $\bar{\phi}$ is a well-defined linear isomorphism.

Proof. First we show that $\bar{\phi}$ is well-defined: $q(v) = q(v')$ if and only if $v - v' \in \ker\phi$ if and only if $\phi(v - v') = 0$, or, equivalently, $\phi(v) = \phi(v')$. We also get a bit more: $\bar{\phi}$ injects since if $\bar{\phi}(q(v)) = \bar{\phi}(q(v'))$ then $\phi(v) = \phi(v')$ which implies that $q(v) = q(v')$.

To see that $\bar{\phi}$ is linear, we compute using the linearity of q and ϕ:

$$\bar{\phi}(q(v_1) + \lambda q(v_2)) = \bar{\phi}(q(v_1 + \lambda v_2)) = \phi(v_1 + \lambda v_2) = \phi(v_1) + \lambda\phi(v_2) = \bar{\phi}(q(v_1)) + \lambda\bar{\phi}(q(v_2)),$$

for $v_1, v_2 \in V$, $\lambda \in \mathbb{F}$.

It remains to show that $\bar{\phi}$ is surjective: but if $w \in \operatorname{im}\phi$, then $w = \phi(v) = \bar{\phi}(q(v))$, for some $v \in V$, and we are done. □

Remarks.

(1) Let $q : V \to V/\ker\phi$ be the quotient map and $i : \operatorname{im}\phi \to W$ the inclusion. Then the First Isomorphism Theorem shows that we may write ϕ as the composition $i \circ \bar{\phi} \circ q$ of a quotient map, an isomorphism and an inclusion.

(2) This whole story of cosets, quotients and the First Isomorphism Theorem has versions in many other contexts such as group theory and ring theory (see MA22017).

Chapter 3

Polynomials, operators and matrices

3.1 Polynomials

Recall from Algebra 1A (§3.2):

Definitions. A *polynomial in a variable x with coefficients in a field* $\mathbb{F}$ is a formal expression

$$p = \sum_{k=0}^{\infty} a_k x^k$$

with *coefficients* $a_k \in \mathbb{F}$ such that only finitely many a_k are non-zero.

Two polynomials are equal if all their coefficients are equal.

The zero polynomial has all coefficients zero.

The *degree* of a polynomial p is $\deg p = \max\{k \in \mathbb{N} \mid a_k \neq 0\}$. By convention, $\deg 0 = -\infty$.

The set of all polynomials in x with coefficients in $\mathbb{F}$ is denoted $\mathbb{F}[x]$.

When $\deg p = n$, we usually write

$$p = a_0 + a_1 x + \cdots + a_n x^n.$$

Thus we adopt the convention $x^0 = 1, x^1 = x$. Here $a_n x^n$ is the *leading term* of p and a_n the *leading coefficient.*

Definition. A polynomial is *monic* if its leading coefficient is 1:

$$p = a_0 + \cdots + x^n.$$

We can add and multiply polynomials: if

$$p = \sum_{k=0}^{\infty} a_k x^k, \qquad q = \sum_{k=0}^{\infty} b_k x^k$$

then

$$p + q := \sum_{k=0}^{\infty} (a_k + b_k) x^k$$

$$pq := \sum_{k=0}^{\infty} \Big(\sum_{i+j=k} a_i b_j \Big) x^k.$$

In particular, we multiply polynomials using $x^i x^j = x^{i+j}$ and collecting terms.

The usual rules of multiplication and addition apply (in the language of MA22017, $\mathbb{F}[x]$ is a *ring*) and, in particular, $\mathbb{F}[x]$ is a vector space. Moreover we have:

$$\begin{aligned}\deg(pq) &= \deg p + \deg q,\\ \deg(p+q) &\leq \max\{\deg p, \deg q\}.\end{aligned}$$

We can *evaluate* polynomials at elements of $\mathbb{F}$. For $p = a_0 + \cdots + a_n x^n$ and $t \in \mathbb{F}$, define $p(t) \in \mathbb{F}$ by

$$p(t) := a_0 + a_1 t + \cdots + a_n t^n,$$

where all the additions and multiplications take place in $\mathbb{F}$. We say that $t \in \mathbb{F}$ is a *root* of p if $p(t) = 0 \in \mathbb{F}$.

Here are the main facts about evaluation:

- Evaluation preserves addition and multiplication: for fixed $t \in \mathbb{F}$, we have

  $$\begin{aligned}(p+q)(t) &= p(t) + q(t)\\ (pq)(t) &= p(t)q(t).\end{aligned}$$

  In particular, $p \mapsto p(t)$ is a linear map $\mathbb{F}[x] \to \mathbb{F}$.
- Evaluation defines functions on $\mathbb{F}$: each $p \in \mathbb{F}[x]$ defines a function $t \mapsto p(t) : \mathbb{F} \to \mathbb{F}$.

Remark. What is a polynomial? We are used to thinking of them as the functions they define but this is not quite correct. Polynomials are simply lists of coefficients or, equivalently, sequences in $\mathbb{F}$ that are eventually zero:

$$\mathbb{F}[x] \cong \{(a_0, \dots, a_n, 0, 0, \dots)\}.$$

The role of the variable x is that of a placeholder to help keep track of things when we multiply polynomials.

For some fields, different polynomials can define the same function. For example, with $\mathbb{F} = \mathbb{Z}_2$, $p = x^2 + x$ and the zero polynomial both define the zero function[1]: $p(t) = 0$ for all $t \in \mathbb{Z}_2$.

We will need three crucial results from Algebra 1A:

Theorem 3.1 (Algebra 1A, Proposition 3.19)**.** *Let $p, q \in \mathbb{F}[x]$. Then there are unique $r, s \in \mathbb{F}[x]$ such that*

$$p = sq + r$$

with $\deg r < \deg q$.

Theorem 3.1 holds for any field $\mathbb{F}$ but the next two results show that the field $\mathbb{C}$ of complex numbers is special:

Theorem 3.2 (Fundamental Theorem of Algebra)**.** *Let $p \in \mathbb{C}[x]$ be a polynomial with $\deg p \geq 1$. Then p has a root. Thus there is $t \in \mathbb{C}$ with $p(t) = 0$.*

Together with Theorem 3.1, this yields:

Theorem 3.3. *Let $p \in \mathbb{C}[x]$ and $\lambda_1, \dots, \lambda_k$ the distinct roots of p. Then*

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

for some $a \in \mathbb{C}$ and $n_i \in \mathbb{Z}_+$, $1 \leq i \leq k$.

n_i is called the multiplicity *of the root λ_i.*

[1] This is question 3 on exercise sheet 3.

3.2 Linear operators, matrices and polynomials

3.2.1 Linear operators and matrices

Definition. Let V be a vector space over $\mathbb{F}$. A *linear operator on* V is a linear map $\phi : V \to V$.

The vector space of linear operators on V is denoted $L(V)$ (instead of $L(V, V)$).

Notation. Write $M_n(\mathbb{F})$ for $M_{n\times n}(\mathbb{F})$.

Recall from Algebra 1B §1.5 that, in the presence of a basis, there is a close relationship between linear operators and square matrices:

Definition. Let V be a finite-dimensional vector space over $\mathbb{F}$ with basis $\mathcal{B} : v_1, \dots, v_n$. Let $\phi \in L(V)$. The *matrix of* ϕ *with respect to* $\mathcal{B}$ is the matrix $A = (A_{ij}) \in M_n(\mathbb{F})$ defined by:

$$\phi(v_j) = \sum_{i=1}^{n} A_{ij} v_i, \tag{3.1}$$

for all $1 \leq j \leq n$.

Thus the recipe for computing A is: *expand* $\phi(v_j)$ *in terms of* $v_1, \dots, v_n$ *to get the* j*-th column of* A.

Equivalently, $\phi(x_1 v_1 + \cdots + x_n v_n) = y_1 v_1 + \cdots + y_n v_n$ where

$$\mathbf{y} = A\mathbf{x}.$$

The map $\phi \mapsto A$ is a linear isomorphism $L(V) \cong M_n(\mathbb{F})$ which also plays well with composition and matrix multiplication: if $\psi \in L(V)$ has matrix B with respect to $\mathcal{B}$ then $\psi \circ \phi$ has matrix BA with respect to $\mathcal{B}$. This gives us a compelling dictionary between linear maps and matrices.

Remark. There is a fancy way to say all this: recall that a basis $\mathcal{B} : v_1, \dots, v_n$ of V gives rise to a linear isomorphism $\phi_{\mathcal{B}} : \mathbb{F}^n \to V$ via

$$\phi_{\mathcal{B}}(\lambda_1, \dots, \lambda_n) = \sum_{i=1}^{n} \lambda_i v_i. \tag{3.2}$$

Now the relation between ϕ and A is that

$$\phi = \phi_{\mathcal{B}} \circ \phi_A \circ \phi_{\mathcal{B}}^{-1}$$

or, equivalently, $\phi_{\mathcal{B}} \circ \phi_A = \phi \circ \phi_{\mathcal{B}}$ so that the following diagram commutes:

$$\begin{tikzcd}
V \arrow[r, "\phi"] & V \\
\mathbb{F}^n \arrow[u, "\phi_{\mathcal{B}}"] \arrow[r, "\phi_A"] & \mathbb{F}^n \arrow[u, "\phi_{\mathcal{B}}"']
\end{tikzcd}$$

(The assertion that such a diagram commutes is simply that the two maps one builds by following the arrows in two different ways coincide. However, the diagram also helps us keep track of where the various maps go!)

3.2.2 Polynomials in linear operators and matrices

A special feature of $L(V)$ is that composition is a binary operation $(\phi, \psi) \mapsto \phi \circ \psi : L(V) \times L(V) \to L(V)$. Thus we can think of composition as a multiplication of operators which suggests the following notations:

Notation. For $\phi, \psi \in L(V)$ write $\phi\psi$ for $\phi \circ \psi \in L(V)$.

Similarly, write ϕ^n for the n-fold composition of ϕ with itself:

$$\phi^n = \underbrace{\phi \circ \cdots \circ \phi}_{n \text{ times}}$$

and define $\phi^0 := \mathrm{id}_V$, $\phi^1 := \phi$.

Finally, for $A \in M_n(\mathbb{F})$, set $A^0 = I_n$, $A^1 = A$.

With these notations and conventions, we have

$$\phi^{n+m} = \phi^n \phi^m, \qquad A^{n+m} = A^n A^m, \tag{3.3}$$

for any $\phi \in L(V)$, $A \in M_n(\mathbb{F})$ and $n, m \in \mathbb{N}$.

Note that if ϕ has matrix A with respect to a basis $\mathcal{B}$ then ϕ^n has matrix A^n with respect to $\mathcal{B}$, for all $n \in \mathbb{N}$.

We can now evaluate polynomials on operators and matrices:

Definition. Let $p \in \mathbb{F}[x]$, $p = a_0 + \cdots + a_n x^n$, $\phi \in L(V)$ and $A \in M_n(\mathbb{F})$. Then $p(\phi) \in L(V)$ and $p(A) \in M_n(\mathbb{F})$ are given by:

$$\begin{aligned} p(\phi) &:= a_0 \,\mathrm{id}_V + a_1\phi + \cdots + a_n\phi^n = \sum_{k\in\mathbb{N}} a_k\phi^k, \\ p(A) &:= a_0 I_n + a_1 A + \cdots + a_n A^n = \sum_{k\in\mathbb{N}} a_k A^k. \end{aligned}$$

Remark. If ϕ has matrix A with respect to a basis $\mathcal{B}$ then $p(\phi)$ has matrix $p(A)$ with respect to $\mathcal{B}$.

This construction plays nicely with the algebra of polynomials:

Proposition 3.4. *For $p, q \in \mathbb{F}[x]$, $\phi \in L(V)$ and $A \in M_n(\mathbb{F})$,*

$$(p+q)(\phi) = p(\phi) + q(\phi) \qquad (p+q)(A) = p(A) + q(A) \tag{3.4}$$

$$(pq)(\phi) = p(\phi)q(\phi) = q(\phi)p(\phi) \qquad (pq)(A) = p(A)q(A) = q(A)p(A). \tag{3.5}$$

Proof. We prove the formulae for ϕ. The arguments for A are very similar.

Write $p = \sum_{k\in\mathbb{N}} a_k x^k$ and $q = \sum_{k\in\mathbb{N}} b_k x^k$. Then

$$(p+q)(\phi) = \sum_{k\in\mathbb{N}} (a_k + b_k)\phi^k = \sum_{k\in\mathbb{N}} a_k\phi^k + \sum_{k\in\mathbb{N}} b_k\phi^k = p(\phi) + q(\phi)$$

which establishes (3.4) for ϕ.

Now for (3.5). We have

$$\begin{aligned} (pq)(\phi) &= \sum_{k\in\mathbb{N}} \big(\sum_{i+j=k} a_i b_j\big)\phi^k = \sum_{k\in\mathbb{N}} \big(\sum_{i+j=k} a_i b_j \phi^i\phi^j\big) \\ &= \sum_{k\in\mathbb{N}} \sum_{i+j=k} (a_i\phi^i)(b_j\phi^j) = \big(\sum_{i\in\mathbb{N}} a_i\phi^i\big)\big(\sum_{j\in\mathbb{N}} b_j\phi^j\big) = p(\phi)q(\phi). \end{aligned}$$

Here we used (3.3) for the last equality on the first line and linearity of ϕ^i to get $b_j\phi^i\phi^j = \phi^i(b_j\phi^j)$.

Finally $pq = qp$ so that

$$pq(\phi) = qp(\phi) = q(\phi)p(\phi)$$

by what we have already proved. □

Remark. The fancy way to say Theorem 3.4 is that the maps $p \mapsto p(\phi) : \mathbb{F}[x] \to L(V)$ and $p \mapsto p(A) : \mathbb{F}[x] \to M_n(\mathbb{F})$ are *homomorphisms of rings* (see MA22017).

3.3 The minimum polynomial

Proposition 3.5. *Let $A \in M_n(\mathbb{F})$. Then there is a monic polynomial $p \in \mathbb{F}[x]$ such that $p(A) = 0$.*

Similarly, if $\phi \in L(V)$ is a linear operator on a finite-dimensional vector space over $\mathbb{F}$ then there is a monic polynomial $p \in \mathbb{F}[x]$ with $p(\phi) = 0$.

Proof. We prove the result for A and then deduce that for ϕ.

We know that $\dim M_n(\mathbb{F}) = n^2$ so that the $n^2 + 1$ elements $I_n, A, \dots, A^{n^2}$ of $M_n(\mathbb{F})$ must be linearly dependent. We therefore have a linear relation

$$a_0 I_n + \cdots + a_{n^2} A^{n^2} = 0$$

with not all a_k zero. Otherwise said, $q(A) = 0$, where

$$q = a_0 + \cdots + a_{n^2} x^{n^2} \in \mathbb{F}[x].$$

Let a_m be the leading term of q (m could be less than n^2). Then $p := q/a_m$ is a monic polynomial with $p(A) = 0$.

Now let $\phi \in L(V)$ and let A be its matrix with respect to some basis. Let $p \in \mathbb{F}[x]$ be a monic polynomial with $p(A) = 0$. Then $p(\phi) = 0$ also. □

This prompts:

Definition. A *minimum polynomial* for $\phi \in L(V)$, V a vector space over $\mathbb{F}$ is a monic polynomial $p \in \mathbb{F}[x]$ of minimum degree with $p(\phi) = 0$: thus, if $r \in \mathbb{F}[x]$ has $r(\phi) = 0$ and $\deg r < \deg p$, then $r = 0$.

Similarly, a minimum polynomial for $A \in M_n(\mathbb{F})$ is a monic polynomial p of least degree with $p(A) = 0$.

Remark. If ϕ has matrix A with respect to some basis, then $p(\phi) = 0$ if and only if $p(A) = 0$ so that p is a minimum polynomial for ϕ if and only if it is one for A.

Minimum polynomials exist and are unique:

Theorem 3.6. *Let $\phi \in L(V)$ be a linear operator on a finite-dimensional vector space over a field $\mathbb{F}$. Then ϕ has a unique minimum polynomial.*

Similarly, any $A \in M_n(\mathbb{F})$ has a unique minimum polynomial.

We denote these by m_ϕ and m_A respectively.

Proof. We prove this for ϕ. The argument for A is the same.

By Theorem 3.5, the set of non-zero polynomials which vanish on ϕ is non-empty. Choose one of smallest degree and divide by the leading term if necessary to get a monic one. This settles existence.

For uniqueness, suppose that we have p_1, p_2 in the set, both monic and of smallest degree. Set $r = p_1 - p_2$. Then $\deg r < \deg p_i$, since the leading terms of the p_i cancel, while $r(\phi) = p_1(\phi) - p_2(\phi) = 0$. Thus $r = 0$ and $p_1 = p_2$. □

Remark. Unless $V = \{0\}$, $\deg m_\phi \geq 1$: the only monic polynomial of degree zero is 1 and $1(\phi) = \mathrm{id}_V \neq 0$!

Examples.

(1) $m_0 = x$.

(2) $m_{\mathrm{id}_V} = x - 1$.

(3) More generally, for $\lambda \in \mathbb{F}$, $m_{\lambda\,\mathrm{id}_V} = x - \lambda$. Thus $\deg m_\phi = 1$ if and only if $\phi = \lambda\,\mathrm{id}_V$, for some $\lambda \in \mathbb{F}$.

(4) Let $\pi \in L(V)$ be a projection[2] with $0 < \dim\ker\pi < \dim V$. Then $m_\pi = x^2 - x$ (exercise!).

How can we compute m_A? One method is to find it by brute force: for each $k \geq 1$ in turn, seek $a_0, \dots, a_{k-1}$ such that

$$a_0 I + \cdots + a_{k-1}A^{k-1} + A^k = 0.$$

This is n^2 inhomogeneous linear equations in k unknowns. They are either inconsistent, in which case you move on to $k+1$ or, the first time you find a solution, $m_A = a_0 + \cdots + x^k$.

Examples.

(1) Find m_A where

$$A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}.$$

Solution. $A \neq \lambda I$ so $\deg m_A \geq 2$. First try to find a_0, a_1 with $a_0 I + a_1 A + A^2 = 0$. This expands out to

$$\begin{pmatrix} a_0 + a_1 + 7 & 0 + 2a_1 + 10 \\ 0 + 3a_1 + 15 & a_0 + 4a_1 + 22 \end{pmatrix} = 0$$

The equation in the $(1,2)$-slot gives $a_1 = -5$ and then that in the $(1,1)$-slot gives $a_0 = -2$. These also satisfy the other two equations and so $m_A = -2 - 5x + x^2$.

(2) Find m_A where

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

Solution. We have

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

so that the $(1,3)$-slot of $a_0 I_3 + a_1 A + A^2 = 0$ gives the inconsistent equation $a_0 0 + a_1 0 + 1 = 0$ and we conclude that $\deg m_A$ is at least three. Carrying on, we compute A^3 and find that $A^3 = I_3$ which short-circuits the whole story: $A^3 - I_3 = 0$ so that $m_A = x^3 - 1$.

We will see other ways to compute the minimum polynomial later.

One reason the minimum polynomial is important:

Proposition 3.7. *Let $\phi \in L(V)$ be a linear operator on a finite-dimensional vector space over $\mathbb{F}$ and $p \in \mathbb{F}[x]$.*

Then $p(\phi) = 0$ if and only if m_ϕ divides p, that is, there is $s \in \mathbb{F}[x]$ such that $p = sm_\phi$.

Proof. If $p(\phi) = 0$ then, by Theorem 3.1, there are $s, r \in \mathbb{F}[x]$ with $\deg r < \deg m_\phi$ such that $p = sm_\phi + r$. But then

$$0 = p(\phi) = s(\phi)m_\phi(\phi) + r(\phi) = r(\phi)$$

so that $r = 0$ and $p = sm_\phi$ by the smallest degree property of m_ϕ.

Conversely, if $p = sm_\phi$ then $p(\phi) = s(\phi)m_\phi(\phi) = 0$. □

Of course, the same statement (and proof!) holds for the minimum polynomial of a matrix $A \in M_n(\mathbb{F})$.

[2] Thus $\pi \circ \pi = \pi$.

3.4 Eigenvalues and the characteristic polynomial

Recall from Chapter 3 of Algebra 1B:

Definitions. Let V be a vector space over $\mathbb{F}$ and $\phi \in L(V)$.

An *eigenvalue* of ϕ is a scalar $\lambda \in \mathbb{F}$ such that there is a *non-zero* $v \in V$ with

$$\phi(v) = \lambda v.$$

Such a vector v is called an *eigenvector of ϕ with eigenvalue λ*.

The *λ-eigenspace $E_\phi(\lambda)$ of ϕ* is given by

$$E_\phi(\lambda) := \ker(\phi - \lambda \operatorname{id}_V) \leq V.$$

Remark. Thus $E_\phi(\lambda)$ consists of all eigenvectors of ϕ with eigenvalue λ along with 0.

Definition. Let V be a finite-dimensional vector space over $\mathbb{F}$ and $\phi \in L(V)$.

The *characteristic polynomial Δ_ϕ of ϕ* is given by

$$\Delta_\phi(\lambda) := \det(\phi - \lambda \operatorname{id}_V) = \det(A - \lambda \mathrm{I}),$$

where A is the matrix of ϕ with respect to some (any!) basis of V.

Thus $\deg \Delta_\phi = \dim V$.

The characteristic polynomial is important to us because:

Lemma 3.8. *A scalar $\lambda \in \mathbb{F}$ is an eigenvalue of ϕ if and only if λ is a root of Δ_ϕ.*

This prompts:

Definitions. Let $\phi \in L(V)$ be in a linear operator on a finite-dimensional vector space V over $\mathbb{F}$ and λ an eigenvalue of ϕ. Then

(1) The *algebraic multiplicity* of λ, $\operatorname{am}(\lambda) \in \mathbb{Z}_+$, is the multiplicity of λ as a root of Δ_ϕ.

(2) The *geometric multiplicity* of λ, $\operatorname{gm}(\lambda) \in \mathbb{Z}_+$, is $\dim E_\phi(\lambda)$.

From Algebra 1B[3],we know that $\operatorname{am}(\lambda) \geq \operatorname{gm}(\lambda)$ and we will get a geometric understanding of $\operatorname{am}(\lambda)$ in the next chapter (see §4.3.2).

When $\mathbb{F} = \mathbb{C}$, Theorem 3.2, the Fundamental Theorem of Algebra, ensures that the characteristic polynomial has at least one root so we conclude from Theorem 3.8:

Theorem 3.9. *Let ϕ be a linear operator on a finite-dimensional vector space V over $\mathbb{C}$. Then ϕ has an eigenvalue.*

Remark. This was crucial in Algebra 1B for the proof of the Spectral Theorem and will be equally crucial for us in the next chapter.

Eigenvalues and eigenvectors play nicely with polynomials:

Proposition 3.10. *Let $\phi \in L(V)$ be a linear operator on a vector space over a field $\mathbb{F}$ and let $v \in V$ be an eigenvector of ϕ with eigenvalue λ:*

$$\phi(v) = \lambda v. \tag{3.6}$$

Let $p \in \mathbb{F}[x]$. Then

$$p(\phi)(v) = p(\lambda)v,$$

so that v is an eigenvector of $p(\phi)$ also with eigenvalue $p(\lambda)$.

[3] Proposition 3.4.6.

Proof. The idea is to iterate (3.6):

$$\phi^2(v) = \phi(\phi(v)) = \phi(\lambda v) = \lambda\phi(v) = \lambda^2 v$$

and so, by induction, $\phi^k(v) = \lambda^k v$, for all $k \in \mathbb{N}$.

Now, for $p = \sum_{k=0}^n a_k x^k$,

$$p(\phi)(v) = \sum_{k=0}^n a_k \phi^k(v) = \sum_{k=0}^n a_k \lambda^k v = \big(\sum_{k=0}^n a_k \lambda^k\big) v = p(\lambda)v.$$

□

This gives us something interesting: if $p(\phi) = 0$ then

$$0 = p(\phi)(v) = p(\lambda)v$$

so that, since $v \neq 0$, $p(\lambda) = 0$. Thus any eigenvalue of ϕ is a root of p. In particular:

Corollary 3.11. *Let ϕ be a linear operator on a finite-dimensional vector space V over $\mathbb{F}$. Then any eigenvalue of ϕ is a root of m_ϕ.*

3.5 The Cayley–Hamilton theorem

Theorem 3.12 (Cayley–Hamilton[4] Theorem)**.** *Let $\phi \in L(V)$ be a linear operator on a finite-dimensional vector space over a field $\mathbb{F}$.*

Then $\Delta_\phi(\phi) = 0$.

Equivalently, for any $A \in M_n(\mathbb{F})$, $\Delta_A(A) = 0$.

Before proving this, let us see what it tells us. Let

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in M_2(\mathbb{F}).$$

Then

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

So the Cayley–Hamilton theorem is telling us that

$$A^2 - (a+d)A + (ad - bc)I_2 = 0,$$

that is,

$$\begin{pmatrix} a^2 + bc & ab + bd \\ ca + dc & cb + d^2 \end{pmatrix} - (a+d)\begin{pmatrix} a & b \\ c & d \end{pmatrix} + \begin{pmatrix} ad - bc & 0 \\ 0 & ad - bc \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

This is certainly true (check it!) but is far from obvious! If you are not yet convinced, work out what the theorem says for $A \in M_3(\mathbb{F})$.

Proof of Theorem 3.12. We will prove the matrix version. So let $A \in M_n(\mathbb{F})$ and write

$$\Delta_A = a_0 + \cdots + a_n x^n.$$

Thus, our mission is to show that

$$a_0 I_n + a_1 A + \cdots + a_n A^n = 0.$$

[4] Arthur Cayley, 1821–1895; William Rowan Hamilton, 1805–1865.

The key is the adjugate formula from Algebra 1B[5]:

$$\operatorname{adj}(A - xI_n)(A - xI_n) = \det(A - xI_n)I_n. \tag{3.7}$$

Each entry of $\operatorname{adj}(A - xI_n)$ is a polynomial in x of degree at most $n-1$ so we write

$$\operatorname{adj}(A - xI_n) = B_0 + B_1 x + \dots + B_{n-1}x^{n-1},$$

with each $B_k \in M_n(\mathbb{F})$. Substitute this into (3.7) to get

$$(B_0 + B_1 x + \dots + B_{n-1}x^{n-1})(A - xI_n) = (a_0 + \dots + a_n x^n)I_n$$

and compare coefficients of x^k to get

$$B_k A - B_{k-1} = a_k I_n, \tag{3.8}$$

for $0 \leq k \leq n$, where we have set $B_{-1} = B_n = 0 \in M_n(\mathbb{F})$.

Multiply (3.8) by A^k on the right to get

$$B_k A^{k+1} - B_{k-1}A^k = a_k A^k$$

and sum:

$$\Delta_A(A) = \sum_{k=0}^n a_k A^k = \sum_{k=0}^n (B_k A^{k+1} - B_{k-1}A^k) = B_n A^{n+1} - B_{-1} = 0$$

because nearly all terms in the penultimate sum cancel. □

Corollary 3.13. *Let $\phi \in L(V)$ be a linear operator on a finite-dimensional vector space over a field $\mathbb{F}$.*

(1) *m_ϕ divides Δ_ϕ. Equivalently, m_A divides Δ_A, for any $A \in M_n(\mathbb{F})$.*

(2) *The roots of m_ϕ are exactly the eigenvalues of ϕ.*

Proof. By Theorem 3.12, $\Delta_\phi(\phi) = 0$ so m_ϕ divides Δ_ϕ by Theorem 3.7. As a result, any root of m_ϕ is a root of Δ_ϕ and so an eigenvalue. Conversely, any eigenvalue is a root of m_ϕ by Theorem 3.11. □

Let us summarise the situation when $\mathbb{F} = \mathbb{C}$ so that any polynomial is a product of linear factors. So let $\phi \in L(V)$ be a linear operator on a finite-dimensional complex vector space with distinct eigenvalues $\lambda_1, \dots, \lambda_k$. Then

$$\begin{aligned} \Delta_\phi &= \pm \prod_{i=1}^k (x - \lambda_i)^{r_i} \\ m_\phi &= \prod_{i=1}^k (x - \lambda_i)^{s_i}, \end{aligned}$$

where $r_i = \operatorname{am}(\lambda_i)$ and $1 \leq s_i \leq r_i$, for $1 \leq i \leq k$.

This gives us another way to find m_ϕ if we can factorise Δ_ϕ: m_ϕ will be of the form $p = \prod_{i=1}^k (x - \lambda_i)^{s_i}$, with each $1 \leq s_i \leq r_i$, so evaluate $p(\phi)$ to find the one of lowest degree with $p(\phi) = 0$.

Examples. Let us find m_A in the following cases:

(1) Take

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

Since A is upper triangular, we immediately see that $\Delta_A = -(x-1)^2(x-2)$ so that m_A is either $(x-1)(x-2)$ or $(x-1)^2(x-2)$.

[5] Theorem 2.4.6

We try the first of these:

$$(A - I_3)(A - 2I_3) = \begin{pmatrix} 0 & 1 & 2 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} -1 & -1 & 2 \\ 0 & -1 & 1 \\ 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & -1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \neq 0.$$

We conclude that $m_A = (x-1)^2(x-2)$.

(2) Let us try again with

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

which also has $\Delta_A = -(x-1)^2(x-2)$ so that m_A is either $(x-1)(x-2)$ or $(x-1)^2(x-2)$. However, this time

$$(A - I_3)(A - 2I_3) = \begin{pmatrix} 0 & 0 & 3 \\ 0 & 0 & 2 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} -1 & 0 & 3 \\ 0 & -1 & 2 \\ 0 & 0 & 0 \end{pmatrix} = 0$$

so that $m_A = (x-1)(x-2)$.

Chapter 4

The structure of linear operators

4.1 On normal forms

Question. Given $\phi \in L(V)$, is there a basis with respect to which ϕ has a "nice" matrix?

Of course, this does not make much sense without some idea of what "nice" should mean for matrices but a reasonable idea might be that there should be a low number of non-zero entries.

There is a matrix version of the same question. For this, recall:

Definition. Matrices $A, B \in M_n(\mathbb{F})$ are *similar* if there is an invertible matrix $P \in M_n(\mathbb{F})$ such that

$$B = P^{-1}AP.$$

We can then ask:

Question. Is A similar to a "nice" matrix?

and a very practical question:

Question (Similarity problem)**.** When are $A, B \in M_n(\mathbb{F})$ similar?

A possible answer to this last question would be to compare "nice" matrices similar to A and B (recall that similarity is an equivalence relation!).

We already know one situation where this sort of thing works out. Recall from Algebra 1B[1] that $A \in M_n(\mathbb{F})$ is *diagonalisable* if and only if it has an eigenbasis if and only if it is similar to a diagonal matrix

$$\begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}. \tag{4.1}$$

Here $\lambda_1, \dots, \lambda_n$ are the eigenvalues of A listed with their multiplicities, that is, each λ_i appears $\operatorname{am}(\lambda_i)$ times. We say that (4.1) is a *normal form* of A.

We can conclude, after reordering eigenbases if necessary:

Theorem. *Diagonalisable matrices $A, B \in M_n(\mathbb{F})$ are similar if and only if they have the same eigenvalues and multiplicities up to order.*

Our plan in this chapter is to try and generalise these ideas to arbitrary $A \in M_n(\mathbb{F})$. We encounter two difficulties almost immediately.

[1] Definition 3.3.1

(1) **Not enough eigenvalues**: Let

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

Then $\Delta_A = x^2 + 1$ which has no eigenvalues at all in $\mathbb{F} = \mathbb{R}$. We solve this problem by working over $\mathbb{C}$.

(2) **Not enough eigenvectors**: Let

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

Then $\Delta_A = x^2$ but $\ker A = \operatorname{span}\{(1, 0)\}$. We therefore do not have enough eigenvectors to span $\mathbb{C}^2$. To solve this problem will need a new idea (see §4.3).

In this chapter, we will, among other things, completely solve the similarity problem for any $A \in M_n(\mathbb{C})$. This will take quite a bit of work but here is a sneak preview: any $A \in M_n(\mathbb{C})$ is similar to a matrix of the form

$$\begin{pmatrix} \lambda_1 & * & & 0 \\ & \ddots & \ddots & \\ & & \ddots & * \\ 0 & & & \lambda_n \end{pmatrix}$$

with eigenvalues with multiplicity on the diagonal, each $*$ on the first super-diagonal either 0 or 1 and zeros elsewhere.

4.2 Invariant subspaces

Definition. Let ϕ be a linear operator on a vector space V. A subspace $U \subseteq V$ is ϕ-invariant if and only if $\phi(u) \in U$, for all $u \in U$.

The next lemma gives us lots of examples:

Lemma 4.1. *Let $\phi, \psi \in L(V)$ be linear operators and suppose that $\phi\psi = \psi\phi$ (say that ϕ and ψ* commute*). Then $\ker \psi$ and $\operatorname{im} \psi$ are ϕ-invariant.*

Proof. Let $v \in \ker \psi$ so that $\psi(v) = 0$. Then

$$\psi(\phi(v)) = \phi(\psi(v)) = \phi(0) = 0$$

so that $\phi(v) \in \ker \psi$ also.

Again, if $v \in \operatorname{im} \psi$, there is $w \in V$ with $\psi(w) = v$ and now

$$\phi(v) = \phi(\psi(w)) = \psi(\phi(w)) \in \operatorname{im} \psi,$$

as required. □

As a consequence, the following are ϕ-invariant:

- $\ker \phi$ and $\operatorname{im} \phi$ (since ϕ commutes with itself!).
- $\ker p(\phi)$, $\operatorname{im} p(\phi)$, for any $p \in \mathbb{F}[x]$ (since $xp = px$ so that $\phi p(\phi) = p(\phi)\phi$).

Also, we have

- $\operatorname{span}\{v\}$, for any eigenvector v of ϕ, since $\phi(v) = \lambda v \in \operatorname{span}\{v\}$. Thus:
- Any $U \leq E_\phi(\lambda)$ is ϕ-invariant.

Remark. If $U \leq V$ is ϕ-invariant then $\phi_{|U} : U \to U$ is in $L(U)$.

Definition. Let $V_1, \dots, V_k \leq V$ with $V = V_1 \oplus \dots \oplus V_k$ and let $\phi_i \in L(V_i)$, for $1 \leq i \leq k$.

Define $\phi : V \to V$ by

$$\phi(v) = \phi_1(v_1) + \dots + \phi_k(v_k),$$

where $v = v_1 + \dots + v_k$ with $v_i \in V_i$, for $1 \leq i \leq k$.

Call ϕ the *direct sum of the* ϕ_i and write $\phi = \phi_1 \oplus \dots \oplus \phi_k$.

There is a related notion for matrices:

Definition. Let $A_1, \dots, A_k$ be square matrices with $A_i \in M_{n_i}(\mathbb{F})$. The *direct sum of the* A_i is

$$A_1 \oplus \dots \oplus A_k := \begin{pmatrix} A_1 & & 0 \\ & \ddots & \\ 0 & & A_k \end{pmatrix} \in M_n(\mathbb{F}),$$

where $n = n_1 + \dots + n_k$.

A matrix of this type is said to be *block diagonal.*

Example.

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \oplus \begin{pmatrix} 5 \end{pmatrix} \oplus \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} = \left(\begin{array}{cc|c|cc} 1 & 2 & 0 & 0 & 0 \\ 3 & 4 & 0 & 0 & 0 \\ \hline 0 & 0 & 5 & 0 & 0 \\ \hline 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{array}\right) \in M_5(\mathbb{R}).$$

Proposition 4.2. *Let $V_1, \dots, V_k \leq V$ with $V = V_1 \oplus \dots \oplus V_k$ and let $\phi_i \in L(V_i)$, for $1 \leq i \leq k$. Let $\phi = \phi_1 \oplus \dots \oplus \phi_k$. Then*

(1) *ϕ is linear so that $\phi \in L(V)$.*

(2) *Each V_i is ϕ-invariant and $\phi_{|V_i} = \phi_i$, $1 \leq i \leq k$.*

(3) *Let $\mathcal{B}_i$ be a basis of V_i and ϕ_i have matrix A_i with respect to $\mathcal{B}_i$, $1 \leq i \leq k$. Then ϕ has matrix $A_1 \oplus \dots \oplus A_k$ with respect to the concatenated basis $\mathcal{B} = \mathcal{B}_1 \dots \mathcal{B}_k$.*

Proof. For (1), let $v, w \in V$ and write

$$v = v_1 + \dots + v_k \qquad w = w_1 + \dots + w_k,$$

with each $v_i, w_i \in V_i$. Then

$$v + \lambda w = (v_1 + \lambda w_1) + \dots + (v_k + \lambda w_k)$$

with each $v_i + \lambda w_i \in V_i$.

Then

$$\phi(v + \lambda w) = \sum_{i=1}^{k} \phi_i(v_i + \lambda w_i) = \sum_{i=1}^{k} \big(\phi_i(v_i) + \lambda\phi_i(w_i)\big) = \sum_{i=1}^{k} \phi_i(v_i) + \lambda \sum_{i=1}^{k} \phi_i(w_i) = \phi(v) + \lambda\phi(w),$$

where we used the linearity of ϕ_i in the second equality.

For (2), let $v \in V_i$ so that we can write $v = v_1 + \dots + v_k$ with $v_i = v$ and $v_j = 0$, for $i \neq j$. Then

$$\phi(v) = \phi_1(0) + \dots + \phi_i(v) + \dots + \phi_k(0) = \phi_i(v) \in V_i$$

so that V_i is ϕ-invariant and $\phi_{|V_i} = \phi_i$.

Finally, for (3), let $\mathcal{B} = \mathcal{B}_1 \dots \mathcal{B}_k = v_1, \dots, v_n$ with $\mathcal{B}_i = v_{a+1}, \dots, v_{a+r}$. Let ϕ have matrix A with respect to $\mathcal{B}$. Then, for $1 \leq j \leq r$,

$$\phi(v_{a+j}) = \sum_{b=1}^{n} A_{b,a+j} v_b.$$

On the other hand,

$$\phi(v_{a+j}) = \phi_i(v_{a+j}) = \sum_{c=1}^{r} (A_i)_{cj} v_{a+c}.$$

Now compare coefficients to see that

$$\begin{aligned} A_{a+c,a+j} &= (A_i)_{cj}, \quad 1 \leq j \leq r \\ A_{b,a+j} &= 0 \quad \text{otherwise.} \end{aligned}$$

Otherwise said, the $a+j$-th column of A has the j-th column of the $r \times r$ matrix A_i in rows $a+1, \dots, a+r$ and zeros elsewhere. This settles (3). □

Conversely, any direct sum decomposition into ϕ-invariant subspaces arises this way:

Proposition 4.3. *Let $V_1, \dots, V_k \leq V$ with $V = V_1 \oplus \dots \oplus V_k$ and let $\phi \in L(V)$. Suppose that each V_i is ϕ-invariant.*

Then $\phi = \phi_1 \oplus \dots \oplus \phi_k$ where $\phi_i := \phi_{|V_i} \in L(V_i)$.

Proof. This is almost obvious: write $v \in V$ as $v = v_1 + \dots + v_k$ with each $v_i \in V_i$. Then

$$\phi(v) = \phi(v_1) + \dots + \phi(v_k) = \phi_1(v_1) + \dots + \phi_k(v_k) = \phi_1 \oplus \dots \oplus \phi_k(v),$$

where the first equality comes from linearity of ϕ and the last from the definition of $\phi_1 \oplus \dots \oplus \phi_k$. □

The usefulness of such a decomposition comes from the fact that nearly all properties of ϕ reduce to properties of the simpler ϕ_i:

Proposition 4.4. *Let $V_1, \dots, V_k \leq V$ with $V = V_1 \oplus \dots \oplus V_k$, $\phi_i \in L(V_i)$, $1 \leq i \leq k$ and $\phi = \phi_1 \oplus \dots \oplus \phi_k$.*

Then:

1. $\ker \phi = \ker \phi_1 \oplus \dots \oplus \ker \phi_k$.
2. $\operatorname{im} \phi = \operatorname{im} \phi_1 \oplus \dots \oplus \operatorname{im} \phi_k$.
3. $p(\phi) = p(\phi_1) \oplus \dots \oplus p(\phi_k)$, *for any* $p \in \mathbb{F}[x]$.
4. $\Delta_\phi = \prod_{i=1}^{k} \Delta_{\phi_i}$.

Note that the sums in (1) and (2) are direct thanks to:

Exercise.[2] Let $V = V_1 \oplus \dots \oplus V_k$ and let $U_i \leq V_i$, $1 \leq i \leq k$. Then the sum $U_1 + \dots + U_k$ is direct.

Proof of Theorem 4.4. For (1), write $v \in \ker \phi$ as $v = v_1 + \dots + v_k$ with each $v_i \in V_i$. Then

$$\phi(v) = \phi_1(v_1) + \dots + \phi_k(v_k) = 0 = 0 + \dots + 0,$$

with $\phi_i(v_i), 0 \in V_i$. The direct sum property tells us that each $\phi_i(v_i) = 0$ so that $v \in \ker \phi_1 \oplus \dots \oplus \ker \phi_k$. Thus $\ker \phi \leq \ker \phi_1 \oplus \dots \oplus \ker \phi_k$.

Conversely, if $v = v_1 + \dots + v_k \in \ker \phi_1 \oplus \dots \oplus \ker \phi_k$ then each $\phi_i(v_i) = 0$ and

$$\phi(v) = \phi_1(v_1) + \dots + \phi_k(v_k) = 0.$$

The argument for item (2) is very similar and so left as an exercise[3].

For item (3), note that, for $v_i \in V_i$, $\phi(v_i) = \phi_i(v_i) \in V_i$ so that

$$\phi^2(v_i) = \phi(\phi_i(v_i)) = \phi_i(\phi_i(v_i)) = \phi_i^2(v_i)$$

[2] Exercise sheet 4, question 2(a)

[3] Question 2(b) on exercise sheet 4.

and so on.

Finally, for item (4), let A_i be the matrix of ϕ_i with respect to some basis $\mathcal{B}_i$ of V_i. Then ϕ has matrix $A_1 \oplus \cdots \oplus A_k$ with respect to $\mathcal{B}_1 \ldots \mathcal{B}_k$ by Theorem 4.2(3). Now Theorem 2.1.4 of Algebra 1B tells us

$$\Delta_\phi = \det(A - xI) = \begin{vmatrix} A_1 - xI & & 0 \\ & \ddots & \\ 0 & & A_k - xI \end{vmatrix} = \prod_{i=1}^{k} \det(A_i - xI) = \prod_{i=1}^{k} \Delta_{\phi_i}.$$

□

Exercise.[4] In this situation, what can you say about m_ϕ?

Here is a first example of these ideas in action:

Proposition 4.5. *Let $\phi \in L(V)$ be a linear operator on a finite-dimensional vector space over a field $\mathbb{F}$ and let $\lambda_1, \ldots, \lambda_k$ be the distinct eigenvalues of ϕ.*

Then ϕ is diagonalisable if and only if

$$V = \bigoplus_{i=1}^{k} E_\phi(\lambda_i). \tag{4.2}$$

Proof. Suppose that (4.2) holds and let $\mathcal{B}_i$ be a basis of $E_\phi(\lambda_i)$. Then, by Theorem 2.7, $\mathcal{B}_1 \ldots \mathcal{B}_k$ is a basis of V which consists of eigenvectors and so is an eigenbasis. Thus ϕ is diagonalisable.

Conversely, suppose that $\mathcal{B} = v_1, \ldots, v_n$ is an eigenbasis for ϕ so that each $\phi(v_j) = \mu_j v_j$, for some $\mu_j \in \{\lambda_1, \ldots, \lambda_k\}$.

We claim: for λ an eigenvalue,

$$U_\lambda := \operatorname{span}\{v_j \mid \mu_j = \lambda\} = E_\phi(\lambda).$$

Given this, $\mathcal{B}_i := \{v_j \mid \mu_j = \lambda_i\}$ is a basis for $E_\phi(\lambda_i)$ and then $\mathcal{B} = \mathcal{B}_1 \ldots \mathcal{B}_k$ so that (4.2) holds, again by Theorem 2.7.

It remains to prove the claim. Clearly $U_\lambda \leq E_\phi(\lambda)$. Conversely, if $v \in E_\phi(\lambda)$, write $v = \sum_{j=1}^{n} a_j v_j$. Then

$$0 = (\phi - \lambda \operatorname{id})(v) = \sum_{j \mid \mu_j = \lambda} (\mu_j - \lambda) a_j v_j + \sum_{j \mid \mu_j \neq \lambda} (\mu_j - \lambda) a_j v_j = \sum_{j \mid \mu_j \neq \lambda} (\mu_j - \lambda) a_j v_j.$$

Since the v_j are linearly independent, we see that $(\mu_j - \lambda) a_j = 0$, for all j with $\mu_j \neq \lambda$, and so all such a_j vanish. Thus

$$v = \sum_{j \mid \mu_j = \lambda} a_j v_j \in U_\lambda.$$

□

To summarise the situation: when ϕ is diagonalisable, then with $V_i := E_\phi(\lambda_i)$ and $\phi_i := \phi_{|V_i}$, we have $V = V_1 \oplus \cdots \oplus V_k$, $\phi = \phi_1 \oplus \cdots \oplus \phi_k$ and

$$\phi_i = \lambda_i \operatorname{id}_{V_i}.$$

Thus the ϕ_i are as simple as they possibly can be!

We now turn to what we can say about general ϕ.

[4] Exercise sheet 4, question 3.

4.3 Jordan decomposition

4.3.1 Powers of operators and Fitting's Lemma

Proposition 4.6 (Increasing kernels, decreasing images)**.** *Let V be a vector space over a field $\mathbb{F}$ and $\phi \in L(V)$. Then*

(1) $\ker \phi^k \leq \ker \phi^{k+1}$, *for all $k \in \mathbb{N}$. That is,*

$$\{0\} = \ker \phi^0 \leq \ker \phi \leq \ker \phi^2 \leq \dots.$$

If $\ker \phi^k = \ker \phi^{k+1}$ then $\ker \phi^k = \ker \phi^{k+n}$, for all $n \in \mathbb{N}$.

(2) $\operatorname{im} \phi^k \geq \operatorname{im} \phi^{k+1}$, *for all $k \in \mathbb{N}$. That is,*

$$V = \operatorname{im} \phi^0 \geq \operatorname{im} \phi \geq \operatorname{im} \phi^2 \geq \dots.$$

If $\operatorname{im} \phi^k = \operatorname{im} \phi^{k+1}$ then $\operatorname{im} \phi^k = \operatorname{im} \phi^{k+n}$, for all $n \in \mathbb{N}$.

Proof. We prove (1) and leave (2) as an exercise[5].

If $v \in \ker \phi^k$ then $\phi^k(v) = 0$ so that $\phi^{k+1}(v) = \phi(\phi^k(v)) = \phi(0) = 0$. Thus $v \in \ker \phi^{k+1}$ as required.

Now suppose that $\ker \phi^k = \ker \phi^{k+1}$ and induct to prove that $\ker \phi^k = \ker \phi^{k+n}$, for $n \in \mathbb{N}$. We already have the $n = 1$ case by assumption so suppose $\ker \phi^k = \ker \phi^{k+n}$, for some n and let $v \in \ker \phi^{k+n+1}$. Then

$$0 = \phi^{k+n+1}(v) = \phi^{k+1}(\phi^n(v))$$

so that $\phi^n(v) \in \ker \phi^{k+1} = \ker \phi^k$. Thus $\phi^{n+k}(v) = 0$ and $v \in \ker \phi^{n+k} = \ker \phi^k$ by the induction hypothesis. Induction now tells us that $\ker \phi^k = \ker \phi^{k+n}$, for all $n \in \mathbb{N}$. □

Corollary 4.7. *Let V be finite-dimensional with $\dim V = n$ and $\phi \in L(V)$. Then, for all $k \in \mathbb{N}$,*

$$\begin{aligned} \ker \phi^n &= \ker \phi^{n+k} \\ \operatorname{im} \phi^n &= \operatorname{im} \phi^{n+k}. \end{aligned}$$

Proof. By Theorem 4.6, we need to prove $\ker \phi^n = \ker \phi^{n+1}$ and $\operatorname{im} \phi^n = \operatorname{im} \phi^{n+1}$.

If $\ker \phi^n \neq \ker \phi^{n+1}$ then, by Theorem 4.6, we have subspaces

$$\{0\} \lneqq \ker \phi \lneqq \cdots \lneqq \ker \phi^{n+1}$$

of strictly increasing dimension so that $\dim \ker \phi^{n+1} \geq n + 1 > \dim V$: a contradiction. Thus $\ker \phi^n = \ker \phi^{n+1}$.

Rank-nullity now tells us that $\dim \operatorname{im} \phi^n = \dim \operatorname{im} \phi^{n+1}$ whence $\operatorname{im} \phi^n = \operatorname{im} \phi^{n+1}$ also. □

Theorem 4.8 (Fitting[6]'s Lemma)**.** *Let $\phi \in L(V)$ be a linear operator on a finite-dimensional vector space over a field $\mathbb{F}$. Then, with $n = \dim V$, we have*

$$V = \ker \phi^n \oplus \operatorname{im} \phi^n.$$

Proof. From Theorem 4.7, we know that $\ker \phi^n = \ker \phi^{n+k}$, $\operatorname{im} \phi^n = \operatorname{im} \phi^{n+k}$, for all $k \in \mathbb{N}$.

We start by proving that $\ker \phi^n \cap \operatorname{im} \phi^n = \{0\}$. For this, let $v \in \ker \phi^n \cap \operatorname{im} \phi^n$ so that $\phi^n(v) = 0$ and there is $w \in V$ such that $v = \phi^n(w)$. Then $0 = \phi^n(v) = \phi^{2n}(w)$ so that $w \in \ker \phi^{2n} = \ker \phi^n$. Thus $v = \phi^n(w) = 0$ as required.

It follows that $V \geq \ker \phi^n \oplus \operatorname{im} \phi^n$ but, by rank-nullity, the dimensions of these spaces coincide whence $V = \ker \phi^n \oplus \operatorname{im} \phi^n$. □

[5] Question 5 on exercise sheet 4.

[6] Hans Fitting, 1906–1938.

4.3.2 Generalised eigenspaces

Let us revisit the example of Section 4.1 of an operator with not enough eigenvectors: contemplate $\phi := \phi_A \in L(\mathbb{C}^2)$ where

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

We know that ϕ has only zero as eigenvalue and the corresponding eigenspace $E_\phi(0) = \operatorname{span}\{(1,0)\} \neq \mathbb{C}^2$. However, $A^2 = 0$ so that $\ker(\phi - 0\,\mathrm{id})^2 = \mathbb{C}^2$.

This gives us a new idea: for $\phi \in L(V)$ and $\lambda \in \mathbb{F}$ look for non-zero $v \in V$ such that

$$(\phi - \lambda\,\mathrm{id})^k(v) = 0,$$

for some $k \in \mathbb{N}$. Thanks to Theorem 4.6 and Theorem 4.7, this amounts to demanding that

$$(\phi - \lambda\,\mathrm{id})^n(v) = 0,$$

where $n = \dim V$.

Observe that this means that $(\phi - \lambda\,\mathrm{id})^k$ is not injective (it has non-trivial kernel) so that $\phi - \lambda\,\mathrm{id}$ is not injective either (and so has non-trivial kernel) and therefore λ is an eigenvalue of ϕ.

This prompts:

Definition. Let $\phi \in L(V)$ be a linear operator on an n-dimensional vector space over a field $\mathbb{F}$. A *generalised eigenvector of ϕ with eigenvalue* λ is a non-zero $v \in V$ such that

$$(\phi - \lambda\,\mathrm{id})^n(v) = 0. \tag{4.3}$$

The set of all such along with 0 is called the *generalised eigenspace of ϕ with eigenvalue* λ and denoted $G_\phi(\lambda)$. Thus

$$G_\phi(\lambda) = \ker(\phi - \lambda\,\mathrm{id}_V)^n \leq V.$$

Lemma 4.9. *$E_\phi(\lambda) \leq G_\phi(\lambda) \leq V$ and $G_\phi(\lambda)$ is ϕ-invariant.*

Proof. There are two things to prove:

(1) $E_\phi(\lambda) \leq G_\phi(\lambda)$. This is straight from Theorem 4.6: $E_\phi(\lambda) = \ker(\phi - \lambda\,\mathrm{id}_V) \leq \ker(\phi - \lambda\,\mathrm{id}_V)^n = G_\phi(\lambda)$.

(2) $G_\phi(\lambda)$ is ϕ-invariant. $G_\phi(\lambda) = \ker p(\phi)$ where $p = (x - \lambda)^n$ which is ϕ-invariant (see the examples after Theorem 4.1).

□

Lemma 4.10. *Let $\phi \in L(V)$ be a linear operator on an n-dimensional vector space over $\mathbb{F}$ and $\lambda_1, \lambda_2 \in \mathbb{F}$ distinct eigenvalues of ϕ. Then $G_\phi(\lambda_1) \cap G_\phi(\lambda_2) = \{0\}$.*

Proof. The assertion amounts to the fact that $(\phi - \lambda_1\,\mathrm{id})^n_{|G_\phi(\lambda_2)}$ is injective (having trivial kernel). It is enough then to prove that $(\phi - \lambda_1\,\mathrm{id})_{|G_\phi(\lambda_2)}$ is injective, or, equivalently, that $E_\phi(\lambda_1) \cap G_\phi(\lambda_2) = \{0\}$. For this, let $v \in E_\phi(\lambda_1) \cap G_\phi(\lambda_2)$. Then $\phi(v) = \lambda_1 v$ and, from Theorem 3.10, $(\phi - \lambda_2\,\mathrm{id})^n(v) = (\lambda_1 - \lambda_2)^n v = 0$. We conclude that $v = 0$ and we are done. □

We now arrive at the promised generalisation of Theorem 4.5.

Theorem 4.11 (Jordan[7] decomposition)**.** *Let $\phi \in L(V)$ be a linear operator on a finite-dimensional vector space over $\mathbb{C}$ with distinct eigenvalues $\lambda_1, \dots, \lambda_k$. Then*

$$V = \bigoplus_{i=1}^{k} G_\phi(\lambda_i).$$

[7] Camille Jordan, 1838–1922.

Proof. We induct on $n := \dim V$.

When $n = 1$, $\phi = \lambda \operatorname{id}$, for some $\lambda \in \mathbb{C}$, so that $V = E_\phi(\lambda) = G_\phi(\lambda)$. This settles the base case.

For the induction step, suppose that the theorem holds for spaces of dimension $< n$ and that $\dim V = n$. Now, by Theorem 3.9, ϕ has an eigenvalue λ_1, say (this is where we use $\mathbb{F} = \mathbb{C}$). Then $G_\phi(\lambda_1) = \ker(\phi - \lambda_1 \operatorname{id})^n$ so that, by Theorem 4.8, we have

$$V = G_\phi(\lambda_1) \oplus \operatorname{im}(\phi - \lambda_1 \operatorname{id})^n.$$

Set $U := \operatorname{im}(\phi - \lambda_1 \operatorname{id})^n$ and write $\hat{\phi} = \phi_{|U}$. We claim:

1. $\hat{\phi}$ has eigenvalues $\lambda_2, \dots, \lambda_k$.
2. For $i \geq 2$, $G_{\hat{\phi}}(\lambda_i) = G_\phi(\lambda_i)$.

Given the claim, since $\dim U < n$, the induction hypothesis applies to give

$$U = \bigoplus_{i=2}^{k} G_\phi(\lambda_i)$$

whence

$$V = G_\phi(\lambda_1) \oplus U = \bigoplus_{i=1}^{k} G_\phi(\lambda_i)$$

as required. The magic of induction now proves the theorem.

It remains to prove the claim. For this, first note that if λ is an eigenvalue of $\hat{\phi}$ with eigenvector $u \in U$ then

$$\lambda u = \hat{\phi}(u) = \phi(u)$$

so that λ is an eigenvalue of ϕ.

Next, observe that

$$E_\phi(\lambda_1) \cap U \leq G_\phi(\lambda_1) \cap U = \{0\}$$

so that λ_1 is not an eigenvalue of $\hat{\phi}$.

On the other hand, for $i \geq 2$, Theorem 4.4 tells us that

$$G_\phi(\lambda_i) = \ker(\phi - \lambda_i \operatorname{id}_V)^n = (G_\phi(\lambda_i) \cap G_\phi(\lambda_1)) \oplus (G_\phi(\lambda_i) \cap U) = G_\phi(\lambda_i) \cap U,$$

where the last equality comes from Theorem 4.10. From this we learn that $G_\phi(\lambda_i) \leq U$ so that, first, λ_i is an eigenvalue of $\hat{\phi}$ and also that $G_{\hat{\phi}}(\lambda_i) = G_\phi(\lambda_i)$ (since it is always true that $G_{\hat{\phi}}(\lambda_i) = G_\phi(\lambda_i) \cap U$). This settles the claim and so the whole proof. □

Let us summarise the situation. With $V_i = G_\phi(\lambda_i)$ and $\phi_i = \phi_{|V_i}$, we have $V = V_1 \oplus \dots \oplus V_k$ and

$$\phi_i = \lambda_i \operatorname{id}_{V_i} + N_i,$$

where we have set $N_i = \phi_i - \lambda_i \operatorname{id}_{V_i} \in L(V_i)$. The key point is that $N_i^n = 0$ which prompts some terminology.

Definition. A linear operator ϕ on a vector space V is *nilpotent* if $\phi^k = 0$, for some $k \in \mathbb{N}$. or, equivalently, if $\ker \phi^k = V$.

Remark. If V is finite-dimensional, we may take $k = \dim V$ by Theorem 4.7.

Our remaining task is to understand nilpotent operators. As a useful first pass at this, we have:

Proposition 4.12. *Let $\phi \in L(V)$ be a linear operator on a finite-dimensional vector space V over $\mathbb{F}$.*

Then ϕ is nilpotent if and only if there is a basis with respect to which ϕ has a strictly upper triangular matrix A (thus $A_{ij} = 0$ whenever $i \geq j$):

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

Proof. Begin by observing that ϕ has strictly upper triangular matrix with respect to $\mathcal{B} : v_1, \dots, v_n$ if and only if $\phi(v_1) = 0$ and $\phi(v_j) \in \operatorname{span}\{v_1, \dots, v_{j-1}\}$, for $j > 1$.

Thus, if ϕ has strictly upper triangular matrix $A \in M_n(\mathbb{F})$ with respect $v_1, \dots, v_n$, we can iterate to see that ϕ^k vanishes on $v_1, \dots, v_k$ and $\phi^k(v_j) \in \operatorname{span}\{v_1, \dots, v_{j-k}\}$, for $j > k$. In particular $\phi^n = 0$. Alternatively, A^k has zeros on the first $k-1$ super-diagonals:

$$A^k = \begin{pmatrix} 0 & \cdots & 0 & & * \\ & \ddots & & \ddots & \\ & & \ddots & & 0 \\ & & & \ddots & \vdots \\ 0 & & & & 0 \end{pmatrix}.$$

In particular, $A^n = 0$ so that $\phi^n = 0$ also.

For the converse, if ϕ is nilpotent, we consider the subspaces

$$\{0\} \leq \ker \phi \leq \ker \phi^2 \leq \dots \leq \ker \phi^{\dim V} = V.$$

Note that, if $v \in \ker \phi^k$, $0 = \phi^k(v) = \phi^{k-1}(\phi(v))$ so that $\phi(v) \in \ker \phi^{k-1}$, for $k \geq 1$.

Now take a basis $v_1, \dots, v_\ell$ of $\ker \phi$, extend it successively to one of $\ker \phi^k$, for each k, until we arrive at a basis $v_1, \dots, v_n$ of V with the property that each $\phi(v_j) \in \operatorname{span}\{v_1, \dots, v_{j-1}\}$. This means precisely that the matrix of ϕ with respect to $v_1, \dots, v_n$ is strictly upper triangular. □

Apply Theorem 4.12 to each N_i to get a basis of V_i for which ϕ_i has a matrix of the form

$$\begin{pmatrix} \lambda_i & & * \\ & \ddots & \\ 0 & & \lambda_i \end{pmatrix}$$

so that, in particular, $\Delta_{\phi_i} = (\lambda_i - x)^{\dim V_i}$. In view of Theorem 4.4(4), we conclude that

$$\Delta_\phi = \prod_{i=1}^{k} \Delta_{\phi_i} = \pm \prod_{i=1} (x - \lambda_i)^{\dim V_i}.$$

Otherwise said, $\operatorname{am}(\lambda_i) = \dim V_i$ and we have proved:

Proposition 4.13. *Let $\lambda \in \mathbb{C}$ be an eigenvalue of a linear operator ϕ on a complex finite-dimensional vector space. Then*

$$\operatorname{am}(\lambda) = \dim G_\phi(\lambda).$$

Remark. Since $E_\phi(\lambda) \leq G_\phi(\lambda)$, this explains the Algebra 1B result[8]that $\operatorname{gm}(\lambda) \leq \operatorname{am}(\lambda)$.

Finally, we can say something useful about the minimal polynomial of ϕ: it is the product of the minimal polynomials of the ϕ_i:

Proposition 4.14. *Let $\phi \in L(V)$ be a linear operator on a finite-dimensional vector space over $\mathbb{C}$ with distinct eigenvalues $\lambda_1, \dots, \lambda_k$. Set $\phi_i = \phi_{|G_\phi(\lambda_i)}$. Then*

[8] Proposition 3.4.6

(1) *Each* $m_{\phi_i} = (x-\lambda_i)^{s_i}$*, for some* $s_i \leq \dim G_\phi(\lambda_i)$*.*

(2) $m_\phi = \prod_{i=1}^k m_{\phi_i} = \prod_{i=1}^k (x-\lambda_i)^{s_i}$*.*

Proof. We know from Theorem 3.13(1) that m_{ϕ_i} divides $\Delta_{\phi_i} = (\lambda_i - x)^{\dim G_\phi(\lambda_i)}$ so (1) is immediate.

For (2), let $p = \prod_{i=1}^k (x-\lambda_i)^{s_i}$. Then $p(\phi) = \bigoplus_{i=1}^k p(\phi_i) = 0$ since each $p(\phi_i) = 0$. Thus m_ϕ divides p and we see conclude that

$$m_\phi = \prod_{i=1}^k (x-\lambda_i)^{t_i},$$

with each $1 \leq t_i \leq s_i$.

On the other hand, each $m_{\phi_i} = (x-\lambda)^{s_i}$ divides m_ϕ since $m_\phi(\phi_i) = m_\phi(\phi)_{|V_i} = 0$. Thus $s_i \leq t_i$, for $1 \leq i \leq k$, and $m_\phi = p$. □

As a corollary, we get an efficient (in the sense of low powers of $(\phi - \lambda_i \operatorname{id}_V)$) expression for $G_\phi(\lambda_i)$:

Corollary 4.15. *Let* $\phi \in L(V)$ *be a linear operator with minimum polynomial* $\prod_{i=1}^k (x-\lambda_i)^{s_i}$*. Then*

$$G_\phi(\lambda_i) = \ker(\phi - \lambda_i \operatorname{id}_V)^{s_i}.$$

Proof. By definition, $\ker(\phi - \lambda_i \operatorname{id}_V)^{s_i} \leq G_\phi(\lambda_i)$. On the other hand, with $V_i = G_\phi(\lambda_i)$ and $\phi_i = \phi_{|V_i}$, we know that $0 = m_{\phi_i}(\phi_i) = (\phi_i - \lambda_i \operatorname{id}_{V_i})^{s_i}$. Otherwise said, $(\phi - \lambda_i \operatorname{id}_V)^{s_i}_{|V_i} = 0$ so that $G_\phi(\lambda_i) \leq \ker(\phi - \lambda_i \operatorname{id}_V)^{s_i}$. □

Example. Let $\phi = \phi_A \in L(\mathbb{C}^3)$ where

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

Find m_ϕ, the eigenspaces and generalised eigenspaces of ϕ.

Solution: A being upper triangular, we see at once that $\Delta_\phi = \Delta_A = (1-x)^2(2-x)$ so that m_A is either $(x-1)(x-2)$ or $(x-1)^2(x-2)$ by Theorem 3.13. We check the first possibility:

$$(A - I_3)(A - 2I_3) = \begin{pmatrix} 0 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} -1 & 1 & 1 \\ 0 & -1 & 1 \\ 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & -1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \neq 0.$$

We conclude that $m_\phi = (x-1)^2(x-2)$ and immediately deduce from Theorem 4.15 that $G_\phi(1) = \ker(\phi - \operatorname{id})^2$ while $G_\phi(2) = \ker(\phi - 2\operatorname{id}) = E_\phi(2)$.

It remains to compute these:

$$E_\phi(1) = \ker(\phi - \operatorname{id}) = \ker \begin{pmatrix} 0 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix} = \operatorname{span}\{(1,0,0)\}$$

$$G_\phi(1) = \ker(\phi - \operatorname{id})^2 = \ker \begin{pmatrix} 0 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}^2 = \ker \begin{pmatrix} 0 & 0 & 2 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix} = \operatorname{span}\{(1,0,0), (0,1,0)\}$$

$$E_\phi(2) = G_\phi(2) = \ker(\phi - 2\operatorname{id}) = \ker \begin{pmatrix} -1 & 1 & 1 \\ 0 & -1 & 1 \\ 0 & 0 & 0 \end{pmatrix} = \operatorname{span}\{(2,1,1)\}.$$

4.4 Jordan normal form

We complete our analyis of linear operators by improving on Theorem 4.12.

First we introduce the key ingredient.

4.4.1 Jordan blocks

Definition. The *Jordan block of size $n \in \mathbb{Z}_+$ and eigenvalue $\lambda \in \mathbb{F}$* is $J(\lambda, n) \in M_n(\mathbb{F})$ with λ's on the diagonal, 1's on the super-diagonal and zeros elsewhere. Thus

$$J(\lambda,n)=\begin{pmatrix}\lambda & 1 & 0 & \cdots & 0\\ & \ddots & \ddots & \ddots & \vdots\\ & & \ddots & \ddots & 0\\ & & & \ddots & 1\\ 0 & & & & \lambda\end{pmatrix}$$

Notation. Set $J_n := J(0, n)$ so that $J(\lambda, n) = \lambda I_n + J_n$.

We have:

Exercises.[9]

(1) $\ker J_n^k = \operatorname{span}\{e_1, \dots, e_k\}$. In particular, J_n is nilpotent: $J_n^n = 0$.

(2) $\operatorname{im} J_n^k = \operatorname{span}\{e_1, \dots, e_{n-k}\}$.

(3) λ is the only eigenvalue of $J(\lambda, n)$ and $E_{J(\lambda,n)}(\lambda) = \operatorname{span}\{e_1\}$, $G_{J(\lambda,n)}(\lambda) = \mathbb{F}^n$.

(4) $m_{J(\lambda,n)} = \pm\Delta_{J(\lambda,n)} = (x - \lambda)^n$.

We are going to prove that any nilpotent operator $\phi \in L(V)$ on a finite-dimensional vector space has a basis for which the matrix of ϕ is a direct sum of Jordan blocks: $J_{n_1} \oplus \cdots \oplus J_{n_k}$ with $n_1 + \cdots + n_k = \dim V$.

We start by spelling out what it means for an operator to have a Jordan block as matrix:

Lemma 4.16. *Let $v_1, \dots, v_n$ be a basis for a vector space V and $\phi \in L(V)$.*

Then the following are equivalent:

(1) *ϕ has matrix J_n with respect to $v_1, \dots, v_n$.*

(2) *$\phi(v_1) = 0$ and $\phi(v_i) = v_{i-1}$, for $2 \le i \le n$.*

(3) *$v_i = \phi^{n-i}(v_n)$, $0 \le i \le n-1$ and $\phi^n(v_n) = 0$.*

Proof. The equivalence of (1) and (2) comes straight from the definitions since $(J_n)_{i-1,i} = 1$ and all other entries in the i-th column vanish.

The equivalence of (2) and (3) is an easy exercise[10]. □

We will work with characterisation (3) and prove:

Theorem 4.17. *Let $\phi \in L(V)$ be a nilpotent operator on a finite-dimensional vector space over $\mathbb{F}$. Then there are $v_1, \dots, v_k \in V$ and $n_1, \dots, n_k \in \mathbb{Z}_+$ such that*

$$\phi^{n_1-1}(v_1), \dots, \phi(v_1), v_1, \dots, \phi^{n_k-1}(v_k), \dots, \phi(v_k), v_k$$

is a basis of V and $\phi^{n_i}(v_i) = 0$, for $1 \le i \le k$.

Using this basis and Theorem 4.16 we immediately conclude:

Corollary 4.18. *Let $\phi \in L(V)$ be a nilpotent operator on a finite-dimensional vector space over $\mathbb{F}$. Then there is a basis for which ϕ has matrix $J_{n_1} \oplus \cdots \oplus J_{n_k}$.*

[9] Exercise sheet 5, question 1.

[10] Question 2 on sheet 5.

Remark. Note that direct sums of the J_{n_i} are characterised by having 1's and zeros (at the joins of successive blocks) on the super-diagonal and zeros elsewhere.

Proof of Theorem 4.17. Once again we induct on $\dim V$.

If $\dim V = 1$, the only nilpotent operator is the zero operator and any basis v_1 will do.

For the induction step, suppose that the theorem is true when $\dim V < n$ and suppose that $\dim V = n$. We prove the theorem for V in three steps.

Step 1: apply the induction hypothesis to $\operatorname{im}\phi$. We let $r = \operatorname{rank}\phi$ and $k = n - r = \dim\ker\phi$. Since ϕ is nilpotent, $k > 0$ so that $r = \dim\operatorname{im}\phi < n$. We therefore apply the induction hypothesis to $\phi_{|\operatorname{im}\phi}$ to get $w_1, \dots, w_\ell \in \operatorname{im}\phi$, $m_1, \dots, m_\ell \in \mathbb{Z}_+$ such that

$$u_1, \dots, u_r := \phi^{m_1-1}(w_1), \dots, \phi(w_1), w_1, \dots, \phi^{m_\ell-1}(w_\ell), \dots, \phi(w_\ell), w_\ell$$

is a basis of $\operatorname{im}\phi$ and $\phi^{m_i}(w_i) = 0$, for $1 \leq i \leq \ell$. Observe that each $\phi(u_i)$ is either u_{i-1} or zero.

Step 2: Find the first ℓ of the v_i. Each $w_i \in \operatorname{im}\phi$ so choose $v_1, \dots, v_\ell$ such that $\phi(v_i) = w_i$, for $1 \leq i \leq \ell$.

We claim that $u_1, \dots, u_r, v_1, \dots, v_\ell$ are linearly independent. For this, suppose that we have a linear relation

$$\sum_{j=1}^{r} \lambda_j u_j + \sum_{i=1}^{\ell} \mu_i v_i = 0 \tag{4.4}$$

and take ϕ of this to get

$$\sum_{j=1}^{r} \lambda_j \phi(u_j) + \sum_{i=1}^{\ell} \mu_i \phi(v_i) = 0$$

which reads

$$\sum_{j|\phi(u_j)\neq 0} \lambda_j u_{j-1} + \sum_{i=1}^{\ell} \mu_i w_i = 0. \tag{4.5}$$

Since these u_{j-1} and w_i are distinct, (4.5) is still a linear relation on the linearly independent u_j and so, in particular, each $\mu_i = 0$. Now (4.4) becomes a linear relation on the u_j and so all $\lambda_j = 0$ also. This proves the claim.

Step 3: extend $u_1, \dots, u_r, v_1, \dots, v_\ell$ to a basis of V by adding elements of $\ker\phi$. Define $U \leq V$ by

$$U = \operatorname{span}\{u_1, \dots, u_r, v_1, \dots, v_\ell\} \geq \operatorname{im}\phi$$

and note that $\operatorname{im}\phi = \phi(U)$ since any $u_i = \phi^m(v_j)$, for some $1 \leq j \leq \ell$ and $1 \leq m \leq m_j$. We extend to get a basis

$$u_1, \dots, u_r, v_1, \dots, v_\ell, x_{\ell+1}, \dots, x_k$$

of V. Now, for $\ell + 1 \leq j \leq k$, there is some $y_j \in U$ such that $\phi(y_j) = \phi(x_j)$ whence $v_j := x_j - y_j \in \ker\phi$.

By construction

$$\operatorname{span}\{u_1, \dots, u_r, v_1, \dots, v_k\} = \operatorname{span}\{u_1, \dots, u_r, v_1, \dots, v_\ell, x_{\ell+1}, \dots, x_k\} = V$$

so that $u_1, \dots, u_r, v_1, \dots, v_k$ is a basis of V. Moreover, setting

$$n_i = \begin{cases} m_i + 1 & 1 \leq i \leq \ell \\ 1 & \ell + 1 \leq i \leq k \end{cases}$$

we have $\phi^{n_i}(v_i) = 0$, for all $1 \leq i \leq k$ and our basis, reordered to slot the first ℓ v_i into the right places, is

$$\phi^{n_1-1}(v_1), \dots, \phi(v_1), v_1, \dots, \phi^{n_\ell-1}(v_\ell), \dots, \phi(v_\ell), v_\ell, v_{\ell+1}, \dots, v_k,$$

which is of the required form. □

The only question left is how unique are the n_i? We already know from the proof of Theorem 4.17 that there are $k = \dim\ker\phi$ of them[11] but we can do better. For this, set $A = J_{n_1} \oplus \cdots \oplus J_{n_k}$ so that, for $s \in \mathbb{N}$, $A^s = J^s_{n_1} \oplus \cdots \oplus J^s_{n_k}$. Now

$$\dim\ker J^s_{n_i} = s,$$

for $s \leq n_i$ so that

$$\dim\ker J^s_{n_i} - \dim\ker J^{s-1}_{n_i} = \begin{cases} 1 & 1 \leq s \leq n_i \\ 0 & s > n_i. \end{cases} \tag{4.6}$$

Now $\ker A^s = \bigoplus_{i=1}^k \ker J^s_{n_i}$ so summing (4.6) over i yields:

$$\#\{i \mid n_i \geq s\} = \dim\ker A^s - \dim\ker A^{s-1}.$$

This proves:

Proposition 4.19. *Let $\phi \in L(V)$ be nilpotent with matrix $J_{n_1} \oplus \cdots \oplus J_{n_k}$ for some basis of V. Then $n_1, \ldots, n_k$ are unique up to order. Indeed,*

$$\#\{i \mid n_i \geq s\} = \dim\ker\phi^s - \dim\ker\phi^{s-1},$$

for each $s \geq 1$.

Exercise.[12] In the situation of Theorem 4.19, show that

$$\#\{i \mid n_i = s\} = 2\dim\ker\phi^s - \dim\ker\phi^{s-1} - \dim\ker\phi^{s+1}.$$

In another direction:

Proposition 4.20. *In the situation of Theorem 4.19, we have*

$$m_\phi = x^s,$$

where $s = \max\{n_1, \ldots, n_k\}$.

Proof. Exercise[13]! □

4.4.2 Jordan normal form

We put §4.4.1 together with Theorem 4.11 to prove the ultimate structure theorem for linear operators on a finite-dimensional complex vector space.

Theorem 4.21. *Let $\phi \in L(V)$ be a linear operator on a finite-dimensional vector space V over $\mathbb{C}$. Then there is a basis of V for which ϕ has as matrix a direct sum of Jordan blocks which are unique up to order.*

Such a basis is called a Jordan basis *and the direct sum of Jordan blocks is called the* Jordan normal form (JNF) *of ϕ.*

Proof. Let $\lambda_1, \ldots, \lambda_k$ be the distinct eigenvalues of ϕ. By Theorem 4.11, $V = \bigoplus V_i$, for $V_i = G_\phi(\lambda_i)$ and then $\phi_i := \phi_{|V_i}$ can be written

$$\phi_i = \lambda_i \operatorname{id}_{V_i} + N_i,$$

with N_i nilpotent. Apply Theorem 4.18 to get a basis of V_i for which N_i has matrix $J_{n_1} \oplus \cdots \oplus J_{n_\ell}$. By Theorem 4.19, the $n_1, \ldots, n_\ell$ are unique up to order. Now ϕ_i has matrix

$$J(\lambda_i, n_1) \oplus \cdots \oplus J(\lambda_i, n_\ell).$$

We then concatenate these bases to get the required Jordan basis of V. □

[11] Alternatively, if you have not read the proof: if there are k Jordan blocks J_{n_i}, we have $\dim\ker\phi = \sum_{i=1}^k \dim\ker J(n_i) = k$ since $\dim\ker J(n_i) = 1$.

[12] Question 3 on sheet 5.

[13] Question 4 on sheet 5.

From this, Theorem 4.14 and Theorem 4.20, we get a complete account of the minimum polynomial:

Corollary 4.22. *Let $\phi \in L(V)$ be a linear operator on a finite-dimensional vector space V over $\mathbb{C}$ with distinct eigenvalues $\lambda_1, \dots, \lambda_k$. Then*

$$m_\phi = \prod_{i=1}^{k} (x - \lambda_i)^{s_i}$$

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

Exercise.[14] ϕ is diagonalisable if and only if $m_\phi = \prod_{i=1}^{k}(x - \lambda_i)$ (that is, all $s_i = 1$).

We can apply all this to matrices and solve the similarity problem.

Corollary 4.23. *Any $A \in M_n(\mathbb{C})$ is similar to a direct sum of Jordan blocks, that is, there is an invertible matrix $P \in M_n(\mathbb{C})$ such that*

$$P^{-1}AP = A_1 \oplus \cdots \oplus A_r,$$

with each A_i a Jordan block.

$A_1 \oplus \cdots \oplus A_r$ is called the Jordan normal form (JNF) of *A and is unique up to the order of the A_i.*

Proof. Apply Theorem 4.21 to $\phi_A : \mathbb{C}^n \to \mathbb{C}^n$ and let P be the change of basis matrix from the standard basis to the Jordan basis of ϕ_A (so that the columns of P are the Jordan basis). □

This gives:

Theorem 4.24. *Matrices $A, B \in M_n(\mathbb{C})$ are similar if and only if they have the same Jordan normal form, up to reordering the Jordan blocks.*

4.4.3 Examples

Example. Let $\phi = \phi_A : \mathbb{C}^4 \to \mathbb{C}^4$ where

$$A = \begin{pmatrix} 2 & -4 & 2 & 2 \\ -2 & 0 & 1 & 3 \\ -2 & -2 & 3 & 3 \\ -2 & -6 & 3 & 7 \end{pmatrix}.$$

let us find the Jordan normal form of A and a Jordan basis of ϕ.

Step 1: compute Δ_A. This turns out to be $(2-x)^2(4-x)^2$ so that we have eigenvalues $2, 4$ and Theorem 4.13 tells us that

$$\dim G_\phi(2) = \dim G_\phi(4) = 2.$$

Step 2: compute m_A by trial and error. It must be $(x-2)^{s_1}(x-4)^{s_2}$ with $1 \leq s_i \leq 2$ so first try $(x-2)(x-4)$:

$$(A - 2I)(A - 4I) = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & -4 & 2 & 2 \\ 0 & -4 & 2 & 2 \\ 0 & -4 & 2 & 2 \end{pmatrix} \neq 0.$$

Next try $(x-2)(x-4)^2$:

$$(A - 2I)(A - 4I)^2 = 0 \in M_4(\mathbb{C})$$

so that $m_A = (x-2)(x-4)^2$.

Step 3: deduce the shape of the Jordan normal form using Theorem 4.22:

[14] Question 5 on sheet 5.

Since $s_1 = 1$, all Jordan blocks with eigenvalue 2 have size 1, $E_\phi(2) = G_\phi(2)$.

Since $s_2 = 2$, there is at least one Jordan block of size 2 with eigenvalue 4 and since $\dim G_\phi(4) = 2$ there is no room for any other block.

We conclude that A has JNF $J(2,1) \oplus J(2,1) \oplus J(4,2)$:

$$\begin{pmatrix} 2 & & & \\ & 2 & & \\ & & 4 & 1 \\ & & & 4 \end{pmatrix}.$$

We find a Jordan basis by finding one for each generalised eigenspace in turn. Any basis of $E_\phi(2)$ will do for the 2-generalised eigenspace so solve $(A - 2I)\mathbf{v} = 0$ to find one. I found $(2,1,0,2)$, $(0,1,2,0)$.

For the 4-generalised eigenspace, we need a basis of the form $(\phi - 4\,\mathrm{id})v, v$ with $(\phi - 4\,\mathrm{id})^2(v) = 0$. For this we work backwards:

(a) Find an eigenvector with eigenvalue 4 by solving $A\mathbf{w} = 4\mathbf{w}$. One solution is $w = (0,1,1,1)$.

(b) Find v by solving $(A - 4I)\mathbf{v} = \mathbf{w}$. One solution is $(1,0,0,1)$.

We therefore have a Jordan basis $(2,1,0,2)$, $(0,1,2,0)$, $(0,1,1,1)$, $(1,0,0,1)$.

It follows that

$$P = \begin{pmatrix} 2 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \\ 0 & 2 & 1 & 0 \\ 2 & 0 & 1 & 1 \end{pmatrix}$$

satisfies

$$P^{-1}AP = \begin{pmatrix} 2 & & & \\ & 2 & & \\ & & 4 & 1 \\ & & & 4 \end{pmatrix}.$$

Example. Let $\phi \in L(V)$ with $\Delta_\phi = (x-5)^4$ and $m_\phi = (x-5)^2$. What can be said about the JNF of ϕ?

Solution: We see from Δ_ϕ that 5 is the only eigenvalue of ϕ and that $\dim V = \deg \Delta_\phi = 4$.

From m_ϕ, we see that there must be at least one Jordan block of size 2. This gives two possibilities:

$$J(5,2) \oplus J(5,2)$$
$$J(5,2) \oplus J(5,1) \oplus J(5,1).$$

In the first case, $\dim E_\phi(5) = 2$ and, in the second, $\dim E_\phi(5) = 3$.

Example. What is the JNF of A given by

$$\begin{pmatrix} -1 & 1 & 0 \\ -1 & 1 & 0 \\ -1 & 1 & 0 \end{pmatrix}?$$

Find a Jordan basis for A.

Solution: One readily checks that $\Delta_A = x^3$, and $A^2 = 0$ whence A is nilpotent with $m_A = x^2$. Thus A has at least one $J_2 = J(0,2)$ block of size two so the JNF must be $J_2 \oplus J_1$.

A Jordan basis is v_1, v_2, v_3 with $A\mathbf{v}_2 = \mathbf{v}_1$ and $A\mathbf{v}_1 = A\mathbf{v}_3 = 0$ so we seek $v_1 \in \operatorname{im} A \cap \ker A$ and work backwards from there.

Solve linear equations to see that

$$\ker A = \{(x,x,y) \mid x,y \in \mathbb{F}\}$$
$$\operatorname{im} A = \{(x,x,x) \mid x \in \mathbb{F}\}$$

so take $v_1 = (1, 1, 1)$ and solve $A\mathbf{v}_2 = v_1$ to get, for example, $v_2 = (0, 1, 0)$. Finally take any $v_3 \in \ker A$ that is linearly independent of v_1: $(0, 0, 1)$ will do.

Thus we have arrived at the Jordan basis $(1, 1, 1)$, $(0, 1, 0)$, $(0, 0, 1)$.

Remark. We see from these computations that Jordan bases of ϕ are far from unique: many choices are made when finding one.

Chapter 5

Symmetric bilinear forms and quadratic forms

We give describe a generalisation of real inner products to vectors spaces V over an arbitrary field $\mathbb{F}$ and use this to study the simplest non-linear functions on V.

5.1 Bilinear forms and matrices

Definition. Let V be a vector space over a field $\mathbb{F}$. A map $B : V \times V \to \mathbb{F}$ is *bilinear* if it is linear in each slot separately:

$$\begin{aligned} B(\lambda v_1 + v_2, v) &= \lambda B(v_1, v) + B(v_2, v) \\ B(v, \lambda v_1 + v_2) &= \lambda B(v, v_1) + B(v, v_2), \end{aligned}$$

for all $v, v_1, v_2 \in V$, $v, v_1, v_2 \in V$ and $\lambda \in \mathbb{F}$.

A bilinear map $V \times V \to \mathbb{F}$ is called a *bilinear form on V*.

Remark. A bilinear form $B : V \times V \to \mathbb{F}$ has $B(v, 0) = B(0, v) = 0$, for all $v \in V$. Indeed,

$$B(v, 0) = B(v, 0 + 0) = B(v, 0) + B(v, 0)$$

and similarly for $B(0, v)$.

Examples.

(1) Any *real* inner product is a bilinear form (what goes wrong for complex inner products?).

(2) Let $A \in M_n(\mathbb{F})$ and define a bilinear form $B_A : \mathbb{F}^n \times \mathbb{F}^n \to \mathbb{F}$ by

$$B_A(x, y) = \mathbf{x}^T A \mathbf{y}.$$

This gives us a new use for matrices.

There is a converse to this last example:

Definition. Let V be a vector space over $\mathbb{F}$ with basis $\mathcal{B} = v_1, \ldots, v_n$ and let $B : V \times V \to \mathbb{F}$ be a bilinear form. The *matrix of B with respect to $\mathcal{B}$* is $A \in M_n(\mathbb{F})$ given by

$$A_{ij} = B(v_i, v_j),$$

for $1 \leq i, j \leq n$.

The matrix A along with $\mathcal{B}$ tells the whole story:

Proposition 5.1. *Let $B : V \times V \to \mathbb{F}$ be a bilinear form with matrix A with respect to $\mathcal{B} = v_1, \dots, v_n$. Then B is completely determined by A: if $v = \sum_{i=1}^n x_i v_i$ and $w = \sum_{j=1}^n y_j v_j$ then*

$$B(v, w) = \sum_{i,j=1}^n x_i y_j A_{ij} = \mathbf{x}^T A \mathbf{y}.$$

Proof. We simply expand out using the bilinearity of B:

$$B(v, w) = \sum_{i,j=1}^n x_i y_j B(v_i, v_j) = \sum_{i,j=1}^n x_i y_j A_{ij}.$$

□

Remark. When $V = \mathbb{F}^n$ and $\mathcal{B} : e_1, \dots, e_n$ is the standard basis, this tells us that any bilinear form on V is B_A where $A_{ij} = B(e_i, e_j)$.

How does A change when we change basis of V?

Proposition 5.2. *Let $B : V \times V \to \mathbb{F}$ be a bilinear form with matrices A and A' with respect to bases $\mathcal{B} : v_1, \dots, v_n$ and $\mathcal{B}' : v'_1, \dots, v'_n$ of V. Then*

$$A' = P^T A P$$

where P is the change of basis matrix[1]from $\mathcal{B}$ to $\mathcal{B}'$: thus $v'_j = \sum_{i=1}^n P_{ij} v_i$, for $1 \leq j \leq n$.

Proof. Using the bilinearity to expand things out, we compute:

$$\begin{aligned} A'_{ij} = B(v'_i, v'_j) = B(\sum_k P_{ki} v_k, \sum_h P_{hj} v_h) \\ = \sum_{k,h} P_{ki} B(v_k, v_h) P_{hj} = \sum_{k,h} (P^T)_{ik} A_{kh} P_{hj} = (P^T A P)_{ij}. \end{aligned}$$

□

This prompts:

Definition. We say that matrices $A, B \in M_n(\mathbb{F})$ are *congruent* if there is $P \in \mathrm{GL}(n, \mathbb{F})$ such that

$$B = P^T A P.$$

5.2 Symmetric bilinear forms

Definition. A bilinear form $B : V \times V \to \mathbb{F}$ is *symmetric* if, for all $v, w \in V$,

$$B(v, w) = B(w, v)$$

Exercise. If V is finite-dimensional, B is symmetric if and only if $B(v_i, v_j) = B(v_j, v_i)$, $1 \leq i, j \leq n$, for some basis $v_1, \dots, v_n$ of V.

Thus B is symmetric if and only if its matrix A with respect to some (and then any) basis is a symmetric matrix: $A^T = A$.

Example. A real inner product is a symmetric bilinear form. Thinking of symmetric bilinear forms as a generalisation of inner products is a good source of intuition.

[1] Algebra 1B, Definition 1.7.1.

5.2.1 Rank and radical

Definitions. Let $B : V \times V \to \mathbb{F}$ be a symmetric bilinear form.

The *radical* $\operatorname{rad} B$ *of* B is given by

$$\operatorname{rad} B := \{v \in V \mid B(v, w) = 0, \text{ for all } w \in V\}.$$

We shall shortly see that $\operatorname{rad} B \leq V$.

We say that B is *non-degenerate* if $\operatorname{rad} B = \{0\}$.

If V is finite-dimensional, the *rank* of B is $\dim V - \dim \operatorname{rad} B$ (so that B is non-degenerate if and only if $\operatorname{rank} B = \dim V$).

Remark. A real inner product B is non-degenerate since $B(v, v) > 0$ when $v \neq 0$.

Lemma 5.3. *Let* $B\colon V \times V \to \mathbb{F}$ *be a symmetric bilinear form with matrix* A *with respect to a basis* $v_1, \dots, v_n$*. Then* $v = \sum_{i=1}^n x_i v_i \in \operatorname{rad} B$ *if and only if* $A\mathbf{x} = 0$ *if and only if* $\mathbf{x}^T A = 0$.

Proof. Since the v_i span V, we see that $B(v, w) = 0$, for all $w \in V$, if and only if $B(v, v_i) = 0$ for $i \leq 1 \leq n$. Thus, $v \in \operatorname{rad} B$ if and only if $\sum_{j=1}^n x_j A_{ji} = 0$, for each i. Otherwise said, $v \in \operatorname{rad} B$ if and only if $\mathbf{x}^T A = 0$ or, taking transposes and remembering that $A^T = A$, $A\mathbf{x} = 0$. □

This enables us to compute $\operatorname{rank} B$:

Corollary 5.4. *Let* $B : V \times V \to \mathbb{F}$ *be a symmetric bilinear form on a finite-dimensional vector space* V *with matrix* A *with respect to some basis of* V*. Then*

$$\operatorname{rank} B = \operatorname{rank} A.$$

In particular, B *is non-degenerate if and only if* $\det A \neq 0$.

Proof. We have, for $n = \dim V$:

$$\operatorname{rank} B = n - \dim \operatorname{rad} B = n - \dim \ker A = \operatorname{rank} A,$$

where the last equality is rank-nullity. □

Examples. We contemplate some symmetric bilinear forms on $\mathbb{F}^3$:

(1) $B(x, y) = x_1 y_1 + x_2 y_2 - x_3 y_3$. With respect to the standard basis, we have

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

so that $\operatorname{rank} B = 3$.

(2) $B(x, y) = x_1 y_2 + x_2 y_1$. Here the matrix with respect to the standard basis is

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

so that B has rank 2 and radical $\operatorname{span}\{e_3\}$.

(3) In general, $B(x, y) = \sum_{i,j=1}^3 A_{ij} x_i y_j$ so we can read off A from the coefficients of the $x_i y_j$.

5.2.2 Classification of symmetric bilinear forms

Convention. In this section, we work with a field $\mathbb{F}$ where $1+1 \neq 0$ so that $\frac{1}{2} = (1+1)^{-1}$ makes sense. This excludes, for example, the 2-element field $\mathbb{Z}_2$.

We can always find a basis with respect to which B has a diagonal matrix. First a lemma:

Lemma 5.5. *Let $B : V \times V \to \mathbb{F}$ be a symmetric bilinear form such that $B(v,v) = 0$, for all $v \in V$. Then $B \equiv 0$.*

Proof. Let $v, w \in V$. We show that $B(v,w) = 0$. We know that $B(v+w, v+w) = 0$ and expanding out gives us

$$0 = B(v,v) + 2B(v,w) + B(w,w) = 2B(v,w).$$

Since $2 \neq 0$ in $\mathbb{F}$, $B(v,w) = 0$. □

We can now prove:

Theorem 5.6 (Diagonalisation Theorem)**.** *Let B be a symmetric bilinear form on a finite-dimensional vector space over $\mathbb{F}$. Then there is a basis $v_1, \dots, v_n$ of V with respect to which the matrix of B is diagonal:*

$$B(v_i, v_j) = 0,$$

for all $1 \leq i \neq j \leq n$. We call $v_1, \dots, v_n$ a diagonalising basis *for B.*

Proof. This is reminiscent of the spectral theorem[2] and we prove it in a similar way by inducting on $\dim V$.

So our inductive hypothesis is that such a diagonalising basis exists for symmetric bilinear forms on a vector space of dimension n.

Certainly the hypothesis holds vacuously if $\dim V = 1$. Now suppose it holds for all vector spaces of dimension at most $n-1$ and that B is a symmetric bilinear form on a vector space V with $\dim V = n$.

There are two possibilities: if $B(v,v) = 0$, for all $v \in V$, then, by Theorem 5.5, $B(v,w) = 0$, for all $v, w \in V$, and any basis is trivially diagonalising.

Otherwise, there is $v_1 \in V$ with $B(v_1, v_1) \neq 0$ and we set

$$U := \operatorname{span}\{v_1\}, \qquad W := \{v \mid B(v_1, v) = 0\} \leq V.$$

We have:

(1) $U \cap W = \{0\}$: if $\lambda v_1 \in W$ then $0 = B(v_1, \lambda v_1) = \lambda B(v_1, v_1)$ forcing $\lambda = 0$.

(2) $V = U + W$: for $v \in V$, write

$$v = \tfrac{B(v_1,v)}{B(v_1,v_1)} v_1 + (v - \tfrac{B(v_1,v)}{B(v_1,v_1)} v_1).$$

The first summand is in U while

$$B\big(v_1, v - \tfrac{B(v_1,v)}{B(v_1,v_1)} v_1\big) = B(v_1, v) - B(v_1, v) = 0$$

so the second summand is in W.

We conclude that $V = U \oplus W$. We therefore apply the inductive hypothesis to $B_{|W \times W}$ to get a basis $v_2, \dots, v_n$ of W with $B(v_i, v_j) = 0$, for $2 \leq i \neq j \leq n$.

Now $v_1, \dots, v_n$ is a basis of V and, further, since $v_j \in W$, for $j > 1$, $B(v_1, v_j) = 0$ so that

$$B(v_i, v_j) = 0,$$

for all $1 \leq i \neq j \leq n$.

Thus the inductive hypothesis holds at $\dim V = n$ and so the theorem is proved. □

[2] Theorem 5.2.11 from Algebra 1B

Remark. We can do a little better if $\mathbb{F}$ is $\mathbb{C}$ or $\mathbb{R}$: when $B(v_i, v_i) \neq 0$, either

(1) If $\mathbb{F} = \mathbb{C}$, replace v_i with $v_i/\sqrt{B(v_i, v_i)}$ to get a diagonalising basis with each $B(v_i, v_i)$ either 0 or 1.

(2) If $\mathbb{F} = \mathbb{R}$, replace v_i with $v_i/\sqrt{|B(v_i, v_i)|}$ to get a diagonalising basis with each $B(v_i, v_i)$ either 0, 1 or -1.

Corollary 5.7. *Let $A \in M_{n\times n}(\mathbb{F})$ be symmetric. Then there is an invertible matrix $P \in \mathrm{GL}(n, \mathbb{F})$ such that $P^T AP$ is diagonal.*

Proof. We apply Theorem 5.6 to B_A to get a diagonalising basis $\mathcal{B}$ and then let P be the change of basis matrix from the standard basis to $\mathcal{B}$. Now apply Theorem 5.2. □

Remark. When $\mathbb{F} = \mathbb{R}$, Theorem 5.7 also follows from the spectral theorem for real symmetric matrices[3], which assures the existence of $P \in \mathrm{O}(n)$ with $P^{-1}AP = P^T AP$ diagonal.

Theorem 5.6 also gives us a recipe for computing a diagonalising basis: find v_1 with $B(v_1, v_1) \neq 0$, compute $W = \{v \mid B(v_1, v) = 0\}$ and iterate. In more detail:

(1) Find $v_1 \in V$ with $B(v_1, v_1) \neq 0$.

(2) Suppose we already have found $v_1, \dots, v_{k-1}$. Now find non-zero $y \in V$ solving

$$B(v_1, y) = \cdots = B(v_{k-1}, y) = 0. \tag{5.1}$$

(3) If $k = \dim V$, take $v_k = y$ and we are done. Otherwise:

(4) Inspect $B(y, y)$. There are three possibilities:
   (i) If $B(y, y) \neq 0$, then set $v_k = y$, and return to step 2 to find v_{k+1}.
   (ii) If $B(y, y) = 0$ and $y \in \operatorname{rad} B$ (so that $B(y, v) = 0$ for all $vinV$), then again set $v_k = y$, and return to step 2 to find v_{k+1}.
   (iii) Otherwise reject y (it cannot be a member of a diagonalising basis[4]) and try another solution of (5.1).

Here are some examples:

Examples.

(1) Problem: find a diagonalising basis for $B = B_A : \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}$ where

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

Solution: First note that $A_{11} \neq 0$ so take $v_1 = e_1$. We seek v_2 among y such that

$$0 = B(v_1, y) = \begin{pmatrix} 1 & 0 & 0 \end{pmatrix} A\mathbf{y} = \begin{pmatrix} 1 & 2 & 1 \end{pmatrix} \mathbf{y} = y_1 + 2y_2 + y_3.$$

We try $v_2 = (1, -1, 1)$ for which

$$B(v_2, y) = \begin{pmatrix} 1 & -1 & 1 \end{pmatrix} A\mathbf{y} = \begin{pmatrix} 0 & 3 & 0 \end{pmatrix} \mathbf{y} = 3y_2$$

In particular, $B(v_2, v_2) = -3 \neq 0$ so we can carry on.
Now seek v_3 among y such that $B(v_1, y) = B(v_2, y) = 0$, that is:

$$\begin{aligned} y_1 + 2y_2 + y_3 &= 0 \\ 3y_2 &= 0. \end{aligned}$$

A solution is given by $v_3 = (1, 0, -1)$ and $B(v_3, v_3) = -1$.
We have therefore arrived at the diagonalising basis $(1, 0, 0), (1, -1, 1), (1, 0, -1)$.
Note that such bases are far from unique: starting from a different v_1 would give a different, equally correct answer.

[3] Algebra 1B, Theorem 5.2.16.
[4] See question 1 on sheet 6.

(2) The same calculation solves another problem: find $P \in \mathrm{GL}(3, \mathbb{R})$ such that $P^T AP$ is diagonal. Solution: we take our diagonalising basis as the columns of P so that

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

Exercise. Check that $P^T AP$ really is diagonal!

Remark. We could also solve this by finding an orthonormal basis of eigenvectors of A but this is way more difficult because we would have to find the eigenvalues by solving a cubic equation.

(3) Now let us take

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 3 & 6 & 9 \end{pmatrix}$$

and find a diagonalising basis for $B = B_A$.
Solution: As before, we can take $v_1 = e_1$ and seek v_2 among y with

$$0 = B(v_1, y) = y_1 + 2y_2 + 3y_3.$$

Let us try $v_2 = (3, 0, -1)$. Then

$$B(v_2, y) = \begin{pmatrix} 3 & 0 & -1 \end{pmatrix} A\mathbf{y} = 0,$$

for *all* y. Otherwise said, $v_2 \in \operatorname{rad} B$. We keep v_2 and try again with $v_3 = (0, -3, 2)$. Again we find that $v_3 \in \operatorname{rad} B$ and conclude that v_1, v_2, v_3 are a diagonalising basis with $B(v_1, v_1) = 1$ and $B(v_2, v_2) = B(v_3, v_3) = 0$.

(4) Here is a trick that can short-circuit these computations if there is a zero in an off-diagonal slot. Take

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

and seek a diagonalising basis for $B = B_A$.
We can exploit the zero in the $(1, 3)$-slot of A: observe that

$$\begin{aligned} B(e_1, e_1) &= 1 \\ B(e_3, e_3) &= -1 \\ B(e_1, e_3) &= 0 \end{aligned}$$

so we are well on the way to getting a diagonalising basis starting with e_1, e_3. To get the last basis vector, we seek $y \in \mathbb{R}^3$ with

$$\begin{aligned} 0 &= B(e_1, y) = y_1 + y_2 \\ 0 &= B(e_3, y) = y_2 - y_3. \end{aligned}$$

We solve these to get $y = (-1, 1, 1)$, for example, and so that $(1, 0, 0), (0, 0, 1), (-1, 1, 1)$ are a diagonalising basis and

$$B(y, y) = 1 - 2 + 2 - 1 = 0.$$

5.2.3 Sylvester's Theorem

Let B be a symmetric bilinear form on a real finite-dimensional vector space. We know that there is a diagonalising basis $v_1, \ldots, v_n$ with each $B(v_i, v_i) \in \{\pm 1, 0\}$ and would like to know how many of each there are. We give a complete answer.

Definitions. Let B be a symmetric bilinear form on a *real* vector space V.

Say that B is *positive definite* if $B(v,v) > 0$, for all $v \in V \setminus \{0\}$.

Say that B is *negative definite* if $-B$ is positive definite.

If V is finite-dimensional, the *signature* of B is the pair (p,q) where

$$p = \max\{\dim U \mid U \leq V \text{ with } B_{|U\times U} \text{ positive definite}\}$$
$$q = \max\{\dim W \mid W \leq V \text{ with } B_{|W\times W} \text{ negative definite}\}.$$

Remark. A symmetric bilinear form B on V is positive definite if and only if it is an inner product on V.

The signature is easy to compute:

Theorem 5.8 (Sylvester's Law of Inertia)**.** *Let B be a symmetric bilinear form of signature (p,q) on a finite-dimensional real vector space Then:*

- *$p + q = \operatorname{rank} B$;*
- *any* diagonal *matrix representing B has p positive entries and q negative entries (necessarily on the diagonal!).*

Proof. Set $K = \operatorname{rad} B$, $r = \operatorname{rank} B$ and $n = \dim V$ so that $\dim K = n - r$.

Let $U \leq V$ be a p-dimensional subspace on which B is positive definite and W a q-dimensional subspace on which B is negative definite.

First note that $U \cap K = \{0\}$ since $B(k,k) = 0$, for all $k \in K$. Thus, by the dimension formula,

$$\dim(U + K) = \dim U + \dim K = p + n - r.$$

Moreover, if $v = u + k \in U + K$, with $u \in U$ and $k \in K$, then $B(v,v) = B(u+k, u+k) = B(u,u) \geq 0$.

From this we see that $W \cap (U+K) = \{0\}$: if $w \in W \cap (U+K)$ then $B(w,w) \geq 0$ by what we just proved but also $B(w,w) \leq 0$ since $w \in W$. Thus $B(w,w) = 0$ and so, by definiteness on W, $w = 0$. Thus

$$\dim(W + (U+K)) = \dim W + \dim(U+K) = q + n + p - r \leq \dim V = n$$

so that $p + q \leq r$.

Now let $v_1, \dots, v_n$ be a diagonalising basis of B with $\hat{p}$ positive entries on the diagonal of the corresponding matrix representative A of B and $\hat{q}$ negative entries. Then B is positive definite on the $\hat{p}$-dimensional space $\operatorname{span}\{v_i \mid B(v_i,v_i) > 0\}$ (exercise[5]!). Thus $\hat{p} \leq p$. Similarly, $\hat{q} \leq q$.

However $r = \operatorname{rank} A$ is the number of non-zero entries on the diagonal, that is $r = \hat{p} + \hat{q}$. We therefore have

$$r = \hat{p} + \hat{q} \leq p + q \leq r$$

so that $p = \hat{p}$, $q = \hat{q}$ and $p + q = r$. □

Example. Find the rank and signature of $B = B_A$ where

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

Solution: we have already found a diagonalising basis $v_1 = (1,0,0), v_2 = (1,-1,1), v_3 = (1,0,-1)$ so we need only count how many $B(v_i,v_i)$ are positive and how many negative. In this case, $B(v_1,v_1) = 1 > 0$ while $B(v_2,v_2) = -3 < 0$ and $B(v_3,v_3) = -1 < 0$. Thus the signature is $(1,2)$ while $\operatorname{rank} B = 1 + 2 = 3$.

[5] Question 2 on sheet 6.

Remarks.

(1) Here is a useful sanity check: symmetric bilinear B of signature (p,q) on an n-dimensional V has $p,q,p+q \leq n$ (since p,q are dimensions of subspaces of n-dimensional V while $n-(p+q) = \dim \operatorname{rad} B \geq 0$).

(2) A symmetric bilinear form of signature $(n,0)$ on a real n-dimensional vector space is simply an inner product.

(3) In physics, the setting for Einstein's theory of special relativity is a 4-dimensional real vector space (*space-time*) equipped with a symmetric bilinear form of signature $(3,1)$.

5.3 Application: Quadratic forms

Convention. We continue working with a field $\mathbb{F}$ where $1+1 \neq 0$.

We can construct a function on V from a bilinear form B (which is a function on $V \times V$).

Definition. A *quadratic form* on a vector space V over $\mathbb{F}$ is a function $Q : V \to \mathbb{F}$ of the form

$$Q(v) = B(v,v),$$

for all $v \in V$, where $B : V \times V \to \mathbb{F}$ is a symmetric bilinear form.

Remark. For $v \in V$ and $\lambda \in \mathbb{F}$, $Q(\lambda v) = B(\lambda v, \lambda v) = \lambda^2 Q(v)$ so Q is emphatically not a linear function!

Examples. Here are two quadratic forms on $\mathbb{F}^3$:

(1) $Q(x) = x_1^2 + x_2^2 - x_3^2 = B_A(x,x)$ where

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

(2) $Q(x) = x_1 x_2 = B_A(x,x)$ where

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

We can recover the symmetric bilinear form B from its quadratic form Q:

Lemma 5.9. *Let $Q : V \to \mathbb{F}$ be a quadratic form with $Q(v) = B(v,v)$ for a symmetric bilinear form B. Then*

$$B(v,w) = \tfrac{1}{2}\big(Q(v+w) - Q(v) - Q(w)\big),$$

for all $v, w \in V$.

B is called the polarisation *of Q.*

Proof. Expand out to get

$$Q(v+w) - Q(v) - Q(w) = B(v,w) + B(w,v) = 2B(v,w).$$

□

Here is how to do polarisation in practice: any quadratic form $Q : \mathbb{F}^n \to \mathbb{F}$ is of the form

$$Q(x) = \sum_{1 \leq i \leq j \leq n} q_{ij} x_i x_j = \mathbf{x}^T \begin{pmatrix} q_{11} & & \frac{1}{2} q_{ji} \\ & \ddots & \\ \frac{1}{2} q_{ij} & & q_{nn} \end{pmatrix} \mathbf{x}$$

so that the polarisation is B_A where

$$A_{ij} = A_{ji} = \begin{cases} q_{ii} & \text{if } i = j; \\ \frac{1}{2} q_{ij} & \text{if } i < j. \end{cases}$$

Example. Let $Q : \mathbb{R}^3 \to \mathbb{R}$ be given by

$$Q(x) = x_1^2 + 2x_2^2 + 2x_1x_2 + x_1x_3.$$

Let us find the polarisation B of Q, that is, we find A so that $B = B_A$: we have $q_{11} = 1$, $q_{22} = 2$, $q_{12} = 2$ and $q_{13} = 1$ with all other q_{ij} vanishing so

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

Definitions. Let Q be a quadratic form on a finite-dimensional vector space V over $\mathbb{F}$.

The *rank* of Q is the rank of its polarisation.

If $\mathbb{F} = \mathbb{R}$, the *signature* of Q is the signature of its polarisation.

What does the diagonalisation theorem mean for a quadratic form Q? We take a practical point of view and let $Q : F^n \to \mathbb{F}$ be a quadratic form on $\mathbb{F}^n$ with polarisation B. We have a diagonalising basis $v_1, \dots, v_n$ of B and let P be the change of basis matrix from the standard basis to $v_1, \dots, v_n$. Then, with $x = \sum_i x_i e_i = \sum_j y_j v_j$, we have

$$Q(x) = \sum_{i=1}^n B(v_i, v_i) y_i^2 = \sum_{i=1}^n B(v_i, v_i) (\sum_{j=1}^n \hat{P}_{ij} x_j)^2,$$

where $\hat{P}_{ij} = (P^{-1})_{ij}$. Otherwise said, Q is a linear combination of squares of linear functions in the x_i and the linear functions have linearly independent coefficients (the rows of P^{-1}).

Let us now apply the classification results of §5.2 and summarise the situation for quadratic forms on vector spaces over our favourite fields:

Theorem 5.10. *Let Q be a quadratic form with rank r polarisation on a finite-dimensional vector space over $\mathbb{F}$.*

(1) When $\mathbb{F} = \mathbb{C}$, there is a basis $v_1, \dots, v_n$ of V such that

$$Q(\sum_{i=1}^n x_i v_i) = x_1^2 + \cdots + x_r^2.$$

(2) When $\mathbb{F} = \mathbb{R}$ and Q has signature (p, q), there is a basis $v_1, \dots, v_n$ of V such that

$$Q(\sum_{i=1}^n x_i v_i) = x_1^2 + \cdots + x_p^2 - x_{p+1}^2 - \cdots - x_r^2.$$

Example. Find the signature of $Q : \mathbb{R}^3 \to \mathbb{R}$ given by

$$Q(x) = x_1^2 + x_2^2 + x_3^2 + 2x_1x_3 + 4x_2x_3.$$

Q has polarisation $B = B_A$ with

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

Solution: exploit the zero in the $(1, 2)$-slot of A to see that $e_1, e_2, y = (-1, -2, 1)$ is a diagonalising basis and so gives us a diagonal matrix representing B with $Q(e_1) = Q(e_2) = 1 > 0$ and $Q(y) = -4 < 0$ along the diagonal. So the signature is $(2, 1)$.

Here are two alternative techniques:

(1) Orthogonal diagonalisation yields a diagonal matrix representing B with the eigenvalues of A down the diagonal so we just count how many positive and negative eigenvalues there are.
In fact, A has eigenvalues 1 and $1 \pm \sqrt{5}$. Since $\sqrt{5} > 2$, $1 - \sqrt{5} < 0$ and we again conclude that the signature is $(2, 1)$.
Danger: this method needed us to solve a cubic equation which is already difficult. For an $n \times n$ A with $n \geq 5$, this could be impossible!

(2) Finally, we could try and write Q as a linear combination of linearly independent squares and then count the number of positive and negative coefficients. In fact,

$$\begin{aligned} Q(x) &= x_1^2 + x_2^2 + x_3^2 + 2x_1x_3 + 4x_2x_3 \\ &= (x_1 + x_3)^2 + x_2^2 + 4x_2x_3 = (x_1 + x_3)^2 + (x_2 + 2x_3)^2 - 4x_3^2. \end{aligned}$$

We must check that the linear functions $x_1 + x_3, x_2 + 2x_3, x_3$ have linearly independent coefficients (that is, $(1, 0, 1)$, $(0, 1, 2)$, $(0, 0, 1)$ are linearly independent) but that is easy. Now the coefficients of these squares are $1, 1, -4$ and so, once more, we get that the signature is $(2, 1)$.