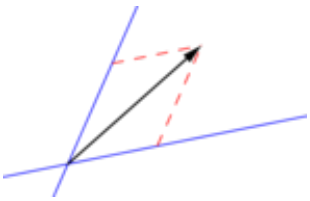


MA20216: Algebra 2A

Notes by Fran Burstall

Corrections by:

Callum Kemp
Carlos Galeano Rios
Kate Powell
Tobias Beith
Krunoslav Lehman Pavasovic
Dan Corbie
Phaidra Anastasiadou
Louise Hannon
Vlad Brebeanu
Lauren Godfrey
Elizabeth Crowley
James Green
Reuben Russell
Ross Trigoll
Emerald Dilworth
George Milton
Caitlin Ray
Liberty Curtis
Harry Todd
Daniel Ng
Papageorgiou Dimosthenis
Kerry Finch
Daniel Dodd
Solla Tapping
Alex Walker
Charlie Hadfield
Sam Cortinhas
Rob Brown

Contents

Chapter 1

Linear algebra: concepts and examples

Let us warm up by revising some of the key ideas from Algebra 1B. Along the way, we will see some new examples and prove a couple of new results.

1.1 Vector spaces

Recall from Algebra 1B, §2.1:

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

addition $V \times V \to V : (v, w) \mapsto v + w$ with respect to which V is an abelian group:

- $v + w = w + v$, 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 $v + (-v) = 0 = (-v) + v$.

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:

$$(\lambda_1, \ldots, \lambda_n) + (\mu_1, \ldots, \mu_n) := (\lambda_1 + \mu_1, \ldots, \lambda_n + \mu_n)$$
$$\lambda(\lambda_1, \ldots, \lambda_n) := (\lambda\lambda_1, \ldots, \lambda\lambda_n).$$

3. Let $M_{m\times n}(\mathbb{F})$ denotes 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 $x = (x_1, \ldots, x_n) \in \mathbb{F}^n$ with the column vector
$$\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}.$$

4. Here is a very general example: let $\mathcal{I}$ be any set and V a vector space. Recall that $V^{\mathcal{I}}$ denotes the set $\{f : \mathcal{I} \to V\}$ of all maps from $\mathcal{I}$ to V.
I claim that $V^{\mathcal{I}}$ is a vector space under pointwise addition and scalar multiplication. That is, for $f, g : \mathcal{I} \to V$ and $\lambda \in \mathbb{F}$, we define
$$(f+g)(i) := f(i) + g(i)$$
$$(\lambda f)(i) := \lambda(f(i)),$$
for all $i \in \mathcal{I}$.
The zero element is just the constant zero function:
$$0(i) := 0,$$
and the additive inverses are defined pointwise also:
$$(-f)(i) := -(f(i)).$$
Exercise.[1] Prove the claim! That is, show that $V^{\mathcal{I}}$ is a vector space under pointwise addition and scalar multiplication.

Remark. For suitable $\mathcal{I}$, this last example captures many familiar vector spaces. For example:

- We identify F^n with $\mathbb{F}^{\{1,\ldots,n\}}$ by associating $(x_1, \ldots, x_n) \in \mathbb{F}^n$ with the map $(i \mapsto x_i)$.
- Similarly, we identify $M_{m\times n}(\mathbb{F})$ with $\mathbb{F}^{\{1,\ldots,m\}\times\{1,\ldots,n\}}$ by associating the matrix A with the map $(i,j) \mapsto A_{ij}$.
- $\mathbb{R}^{\mathbb{N}}$ is the set of real sequences $\{(a_n)_{n\in\mathbb{N}} : a_n \in \mathbb{R}\}$ that played such a starring role in Analysis 1.

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.[2] $U \subseteq V$ 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] Question 4 on sheet 1.
[2] Question 1 on sheet 1.

Examples. A good way to see that something is a vector space is to see that it is a subspace of some $V^{\mathcal{I}}$. That way, there is no need to verify all the tedious axioms (associativity, distributivity and so on).

1. The set $c := \{\text{real convergent sequences}\} \leq \mathbb{R}^{\mathbb{N}}$ and so is a vector space. This is part of the content of the Algebra of Limits Theorem in Analysis 1.
2. Let $[a, b] \subseteq \mathbb{R}$ be an interval and set
$$C^0[a, b] := \{f : [a, b] \to \mathbb{R} \mid f \text{ is continuous}\},$$
the set of continuous functions.
Then $C^0[a, b] \leq \mathbb{R}^{[a,b]}$. This is most of the Algebra of Continuous Functions Theorem from Analysis 1.

1.3 Bases

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

1. The *span* of $v_1, \ldots, v_n$ is
$$\operatorname{span}\{v_1, \ldots, v_n\} := \{\lambda_1 v_1 + \cdots + \lambda_n v_n \mid \lambda_i \in \mathbb{F}, 1 \leq i \leq n\} \leq V.$$
2. $v_1, \ldots, v_n$ *span* V (or *are a spanning list for* V) if $\operatorname{span}\{v_1, \ldots, v_n\} = V$.
3. $v_1, \ldots, v_n$ are *linearly independent* if, whenever $\lambda_1 v_1 + \cdots + \lambda_n v_n = 0$, then each $\lambda_i = 0$, $1 \leq i \leq n$, and *linearly dependent* otherwise.
4. $v_1, \ldots, v_n$ is a *basis* for V if they are linearly independent and span V.

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 $v_1, \ldots, v_n$ be a list of vectors.

1. A vector of the form $\lambda_1 v_1 + \cdots + \lambda_n v_n$ is called a *linear combination of the* v_i.
2. An equation of the form $\lambda_1 v_1 + \cdots + \lambda_n v_n = 0$ is called a *linear relation on the* v_i.

Recall:

Proposition 1.1 (Algebra 1B, Chapter 2, Proposition 4)**.** *$v_1, \ldots, v_n$ is a basis for V if and only if any $v \in V$ can be written in the form*
$$v = \lambda_1 v_1 + \cdots + \lambda_n v_n \tag{1.1}$$
for unique *$\lambda_1, \ldots, \lambda_n \in \mathbb{F}$. In this case, $(\lambda_1, \ldots, \lambda_n)$ is called the* coordinate vector *of v with respect to $v_1, \ldots, v_n$.*

1.3.1 Standard bases

In general, finite-dimensional vector spaces have many bases and there is no good reason to prefer any particular one. However, some lucky vector spaces come equipped with a natural basis.

Proposition 1.2. *For $\mathcal{I}$ a set and $i \in \mathcal{I}$, define $e_i \in \mathbb{F}^{\mathcal{I}}$ by*
$$e_i(j) = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j, \end{cases}$$
for all $j \in \mathcal{I}$.

If $\mathcal{I}$ is finite then $(e_i)_{i \in \mathcal{I}}$ is a basis, called the standard basis*, of $\mathbb{F}^{\mathcal{I}}$.*

In particular, $\dim \mathbb{F}^{\mathcal{I}} = |\mathcal{I}|$.

Proof. For $f \in \mathbb{F}^{\mathcal{I}}$, we observe that

$$f = \sum_{i \in \mathcal{I}} f(i) e_i.$$

Indeed, for $j \in \mathcal{I}$,

$$\big(\sum_{i \in \mathcal{I}} f(i) e_i\big)(j) = \sum_{i \in \mathcal{I}} f(i) e_i(j) = \sum_{i \neq j} f(i)0 + f(j)1 = f(j).$$

In particular, $(e_i)_{i \in \mathcal{I}}$ span.

For linear independence, suppose that $\sum_{i \in \mathcal{I}} \lambda_i e_i = 0$ and evaluate both sides at $j \in \mathcal{I}$ to get

$$\lambda_j = 0.$$

□

Examples.

- Identify $\mathbb{F}^n$ with $\mathbb{F}^{\{1,\dots,n\}}$ and then $e_i = (0, \dots, 1, \dots, 0)$ with a single 1 in the i-th place.
- Similarly, the vector space of column vectors has a standard basis with $\mathbf{e}_i$, the column vector with a single 1 in the i-th row:

$$\mathbf{e}_i = \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix}.$$

- Finally, identifying $M_{m \times n}(\mathbb{F})$ with $\mathbb{F}^{\{1,\dots,m\} \times \{1,\dots,n\}}$ yields the standard basis $(e_{(i,j)})_{i,j}$ of $M_{m \times n}(\mathbb{F})$ where $e_{(i,j)}$ differs from the zero matrix by a single 1 in the i-th row and j-th column.

1.3.2 Useful facts

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

Proposition 1.3 (Algebra 1B, Chapter 3, Theorem 6(b))**.** *Any linearly independent list of vectors in a finite-dimensional vector space can be extended to a basis.*

Here is another helpful result :

Lemma 1.4 (Algebra 1B, Chapter 3, Theorem 5)**.** *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 $\phi : V \to W$ of vector spaces over $\mathbb{F}$ is a *linear map* (or, in older books, *linear transformation*) if

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

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 $\psi \circ \phi : V \to U$.
4. Recall that c is the vector space of convergent sequences.
   The map $\lim_{n\to\infty} : (a_n)_{n\in\mathbb{N}} \mapsto \lim_{n\to\infty} a_n : c \to \mathbb{R}$ is linear thanks to the Algebra of Limits Theorem in Analysis 1.
5. $\int_a^b : f \mapsto \int_a^b f : C^0[a, b] \to \mathbb{R}$ is also linear.

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 1.5. *$\phi : V \to W$ is an isomorphism if and only if ϕ is a linear bijection (and then $\psi = \phi^{-1}$).*

1.4.1 Vector spaces of linear maps

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 1.6 (Linearity is a linear condition)**.** *$L(V, W)$ is a vector space under pointwise addition and scalar multiplication. Otherwise said, $L(V, W) \leq W^V$.*

Proof. It is enough to show that $L(V, W)$ is a vector subspace of W^V, that is, is non-empty and closed under addition and scalar multiplication.

First observe that the zero map $0 : v \mapsto 0 \in W$ is linear:

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

In particular, $L(V, W)$ is non-empty.

Now let $\phi, \psi \in L(V, W)$ and show that $\phi + \psi$ is linear:

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

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. □

1.4.2 Linear maps and matrices

Recall from Algebra 1B §2.5:

Definition. Let V, W be finite-dimensional vector spaces over $\mathbb{F}$ with bases $\mathcal{B} : v_1, \ldots, v_n$ and $\mathcal{B}' : w_1, \ldots, w_m$ respectively. Let $\phi \in L(V, W)$. The *matrix of ϕ with respect to $\mathcal{B}, \mathcal{B}'$* is the matrix $A = (A_{ij}) \in M_{m \times n}(\mathbb{F})$ defined by:

$$\phi(v_j) = \sum_{i=1}^{m} A_{ij} w_i, \tag{1.2}$$

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

In the special case where $V = W$ *and* $\mathcal{B} = \mathcal{B}'$, we call A the *matrix of ϕ with respect to $\mathcal{B}$.*

Thus the recipe for computing A is: *expand $\phi(v_j)$ in terms of $w_1, \ldots, w_m$ to get the j-th column of A.*

Equivalently, $\phi(x_1 v_1 + \cdots + x_n v_n) = y_1 w_1 + \cdots + y_m w_m$ where

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

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

$$\phi_{\mathcal{B}}(\lambda_1, \ldots, \lambda_n) = \sum_{i=1}^{n} \lambda_i v_i. \tag{1.3}$$

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:

$$\begin{array}{ccc} V & \xrightarrow{\phi} & W \\ {\scriptstyle \phi_{\mathcal{B}}}\uparrow & & \uparrow{\scriptstyle \phi_{\mathcal{B}'}} \\ \mathbb{F}^n & \xrightarrow{\phi_A} & \mathbb{F}^m \end{array}$$

(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!)

The map $\phi \mapsto A$ is a linear isomorphism $L(V, W) \cong M_{m \times n}(\mathbb{F})$ which also plays well with composition and matrix multiplication: if U is a third vector space with basis $\mathcal{B}''$ and $\psi \in L(W, U)$ has matrix B with respect to $\mathcal{B}', \mathcal{B}''$ then $\psi \circ \phi$ has matrix BA with respect to $\mathcal{B}, \mathcal{B}''$. This gives us a compelling dictionary between linear maps and matrices.

1.4.3 Extension by linearity

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

Proposition 1.7 (Extension by linearity)**.** *Let V, W be vector spaces over $\mathbb{F}$. Let $v_1, \ldots, v_n$ be a basis of V and $w_1, \ldots, w_n$* any *vectors in W.*

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

$$\phi(v_i) = w_i, \qquad 1 \leq i \leq n. \tag{1.4}$$

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

Let $v \in V$. By Proposition 1.1, 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

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

so that (1.4) 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 (1.4). 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 (1.4) for first ϕ and then ϕ'. We conclude that $\phi = \phi'$ and we are done. □

Remark. In the context of Proposition 1.7, ϕ is an isomorphism if and only if $w_1, \dots, w_n$ is a basis for W (exercise[3]!).

1.4.4 The rank-nullity theorem

Easily the most important result in Algebra 1B is the famous Rank-nullity theorem:

Theorem 1.8 (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 $\ker \phi = \{0\}$, we saw in Algebra 1B:

Proposition 1.9. *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:

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

1. *ϕ is injective.*
2. *ϕ is surjective.*
3. *ϕ is an isomorphism.*

Remark. Proposition 1.9 is flat-out false for infinite-dimensional V, W. For example: let $S : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}^{\mathbb{N}}$ be the *shift* operator:

$$S((a_0, a_1, \dots)) := (a_1, \dots).$$

We readily check that:

- S is linear;
- S surjects;
- S is not injective. For example: $S((1, 0, 0, \dots)) = 0$.

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 $V_1, \ldots, V_k \leq V$. The *sum* $V_1 + \cdots + V_k$ is the set

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

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

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

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

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

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

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

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

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

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 $V_1, \ldots, V_k \leq V$. Any $v \in V_1 + \cdots + V_k$ can be written

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

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

Definition. Let $V_1, \ldots, V_k \leq V$. 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 $V_1 + \cdots + V_k$.

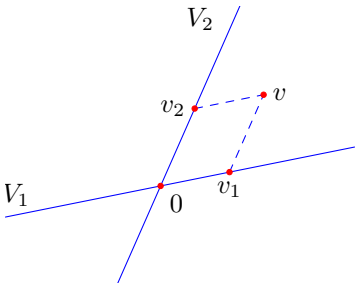


Figure 2.1: $\mathbb{R}^2 = V_1 \oplus V_2$

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

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

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

When is a sum direct? We consider the case of two summands first where there is a very simple answer.

Proposition 2.2. *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. First suppose that $V_1 + V_2$ is direct and let $v \in V_1 \cap V_2$. Then we can write v in two ways:

$$\begin{aligned} v &= v_1 + 0 \\ &= 0 + v_2, \end{aligned}$$

with $v = v_1 = v_2$. The uniqueness of the decomposition now forces $v = 0$.

For the converse, suppose that $V_1 \cap V_2 = \{0\}$ and that $v \in V_1 + V_2$ can be written

$$v = v_1 + v_2 = w_1 + w_2$$

with $v_i, w_i \in V_i$, $i = 1, 2$. Then

$$(v_1 - w_1) = (w_2 - v_2)$$

with the left hand in V_1, the right in V_2 and so both in $V_1 \cap V_2$ from which we immediately get $v_i = w_i$, $i = 1, 2$ so that $V_1 + V_2$ is direct. □

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 Proposition 2.2, 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.

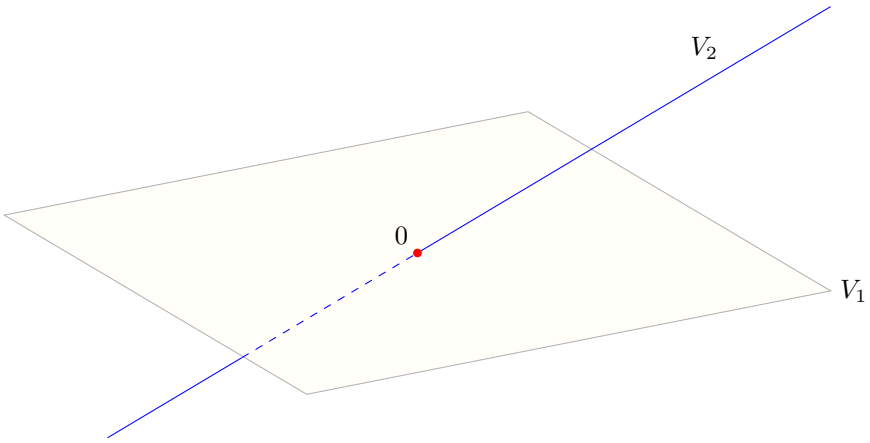


Figure 2.2: $\mathbb{R}^3$ as a direct sum of a line and a plane

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

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

Proposition 2.3. *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 Proposition 2.2. □

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 Direct sums and projections

Definition. Let V be a vector space. A linear map $\pi : V \to V$ is a *projection* if $\pi \circ \pi = \pi$.

Exercise.[1] If π is a projection, $V = \ker \pi \oplus \operatorname{im} \pi$.

In fact, all direct sums arise this way:

Proposition 2.4. *Let $V_1, V_2 \leq V$ with $V = V_1 \oplus V_2$. Then there are projections $\pi_1, \pi_2 : V \to V$ such that:*

(a) $\operatorname{im} \pi_i = V_i$, $i = 1, 2$;
(b) $\ker \pi_1 = V_2$, $\ker \pi_2 = V_1$;
(c) $v = \pi_1(v) + \pi_2(v)$, *for all* $v \in V$. *Otherwise said,* $\mathrm{id}_V = \pi_1 + \pi_2$.

Proof. Item (c) tells us how to define the π_i so we do this first: for $v \in V$, we have that $v = v_1 + v_2$ for unique $v_i \in V_i$, $i = 1, 2$. So we define $\pi_i(v)$ to be

$$\pi_i(v) := v_i,$$

for $i = 1, 2$.

Our first task is to prove that the π_i are linear: for $v, w \in V$ and $\lambda \in \mathbb{F}$, we have

$$\begin{aligned} v &= \pi_1(v) + \pi_2(v) \\ w &= \pi_1(w) + \pi_2(w) \\ v + \lambda w &= \pi_1(v + \lambda w) + \pi_2(v + \lambda w). \end{aligned}$$

[1] Question 3 on sheet 2.

However, the first two equalities also give

$$v + \lambda w = (\pi_1(v) + \lambda\pi_1(w)) + (\pi_2(v) + \lambda\pi_2(w))$$

so the uniqueness in the third equality gives

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

$i = 1, 2$, so that both π_i are linear.

By definition, $\operatorname{im}\pi_i \leq V_i$. For the converse, note that, for $v_1 \in V_1$, we have $v_1 = v_1 + 0$, with $0 \in V_2$, so that $\pi_1(v_1) = v_1$. In particular, $V_1 \leq \operatorname{im}\pi_1$ so that $\operatorname{im}\pi_1 = V_1$. Moreover, taking $v_1 = \pi_1(v)$, we get $\pi_1(\pi_1(v)) = \pi_1(v)$, for any $v \in V$ so that π_1 is a projection. Similarly π_2 is a projection and (a) holds.

Finally, $v = \pi_1(v) + \pi_2(v) \in \ker\pi_1$ if and only if $v = \pi_2(v)$, or, as we have just seen, $v \in V_2$. Thus $\ker\pi_1 = V_2$ and similarly $\ker\pi_2 = V_1$ settling (b). □

As a corollary, we see that dimensions add in direct sums:

Proposition 2.5. *Let $V = V_1 \oplus V_2$ with V finite-dimensional. Then*

$$\dim V = \dim V_1 + \dim V_2.$$

Proof. We apply the rank-nullity theorem to π_1:

$$\begin{aligned}\dim V &= \dim\operatorname{im}\pi_1 + \dim\ker\pi_1 \\ &= \dim V_1 + \dim V_2.\end{aligned}$$

□

2.2.2 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 2.6. *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. Then any $v \in V_1 + \cdots + V_{k-1}$ can be written

$$v = v_1 + \cdots + v_{k-1} + 0$$

for unique $v_i \in V_i$, $1 \leq i \leq k-1$ so that $V_1 + \cdots + V_{k-1}$ is direct. Moreover, any $v \in (V_1 + \cdots + V_{k-1}) + V_k = V_1 + \cdots + V_k$ can be written

$$v = v_1 + \cdots + v_k = (v_1 + \cdots + v_{k-1}) + v_k$$

with, in particular, unique $v_k \in V_k$ so that $(V_1 + \cdots + V_{k-1}) + V_k$ is direct.

For the converse, suppose that $V_1 + \cdots + V_{k-1}$ and $(V_1 + \cdots + V_{k-1}) + V_k$ are both direct. Then any $v \in V_1 + \cdots + V_k$ can be written $v = w + v_k$ for unique $w \in V_1 + \cdots + V_{k-1}$ and $v_k \in V_k$. Also, there is a unique way to write w as

$$w = v_1 + \cdots + v_{k-1}$$

with $v_i \in V_i$, $1 \leq i \leq k-1$. Putting this together, we get

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

for unique $v_i \in V_i$, $1 \leq i \leq k$ so that $V_1 + \cdots + V_k$ is direct. □

Here is a sample application:

Corollary 2.7. *Let $V_1, \ldots, V_k \leq V$ be subspaces of a finite-dimensional vector space V with $V_1 + \cdots + V_k$ direct. Then*

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

Proof. We induct on k using Proposition 2.5 and Lemma 2.6 in the induction step. In more detail: the induction hypothesis is that the formula holds for k summands. The base case reads

$$\dim V_1 = \dim V_1$$

which trivially holds. For the induction step, suppose that the formula holds for any $k-1$ summands. Then, if $V_1 + \cdots + V_k$ is direct, Lemma 2.6 says that $V_1 + \cdots + V_{k-1}$ is direct and then the induction hypothesis says that $\dim V_1 \oplus \cdots \oplus V_{k-1} = \dim V_1 + \cdots + \dim V_{k-1}$. Now Lemma 2.6 says that

$$V_1 \oplus \cdots \oplus V_k = (V_1 \oplus \cdots \oplus V_{k-1}) \oplus V_k$$

so that Proposition 2.5 applies to give

$$\dim V_1 \oplus \cdots \oplus V_k = (\dim V_1 + \cdots + \dim V_{k-1}) + \dim V_k.$$

□

2.2.3 Direct sums and bases

Proposition 2.5 suggests that there is a relation between the bases of V_1, V_2 and the basis of $V_1 \oplus V_2$. This is indeed the case:

Proposition 2.8. *Let $V_1, V_2 \leq V$ be finite-dimensional subspaces with bases $\mathcal{B}_1 : v_1, \ldots, v_k$ and $\mathcal{B}_2 : w_1, \ldots, w_l$. Then $V_1 + V_2$ is direct if and only if the* concatenation[2] *$\mathcal{B}_1\mathcal{B}_2 : v_1, \ldots, v_k, w_1, \ldots, w_l$ 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

$$\sum_{i=1}^k \lambda_i v_i = -\sum_{j=1}^l \mu_j w_j \in V_1 \cap V_2$$

which last is the zero subspace by Proposition 2.2. Thus

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

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 $v \in V_1 \cap V_2$, we can write v in two ways: $v = \sum_{i=1}^k \lambda_i v_i$, for some $\lambda_1, \ldots, \lambda_k \in \mathbb{F}$, since $v \in V_1$ and, similarly, $v = \sum_{j=1}^l \mu_j w_j$. We therefore have a linear relation $\sum_{i=1}^k \lambda_i v_i - \sum_{j=1}^l \mu_j w_j = 0$ and so, by linear independent of $\mathcal{B}_1\mathcal{B}_2$, all λ_i, μ_j vanish so that $v = 0$. Thus $V_1 \cap V_2 = \{0\}$ and $V_1 + V_2$ is direct by Proposition 2.2. □

Again, this along with Lemma 2.6 and induction on k yields the many-summand version:

Corollary 2.9. *Let $V_1, \ldots, V_k \leq V$ be finite-dimensional subspaces with $\mathcal{B}_i$ a basis of V_i, $1 \leq i \leq k$. Then $V_1 + \cdots + V_k$ is direct if and only if the concatenation $\mathcal{B}_1 \ldots \mathcal{B}_k$ is a basis for $V_1 + \cdots + V_k$.*

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

2.2.4 Complements

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

Proposition 2.10 (Complements exist)**.** *Let $U \leq V$, a finite-dimensional vector space. Then there is a complement to U.*

Proof. Let $\mathcal{B}_1 : v_1, \ldots, v_k$ be a basis for U and so a linearly independent list of vectors in V. By Proposition 1.3, we can extend the list to get a basis $\mathcal{B} : v_1, \ldots, v_n$ of V. Set $W = \operatorname{span}\{v_{k+1}, \ldots, v_n\} \leq V$: this is a complement to U.

Indeed, $\mathcal{B}_2 : v_{k+1}, \ldots, v_n$ is a basis for W and $\mathcal{B} = \mathcal{B}_1\mathcal{B}_2$ so that $V = U \oplus W$ by Proposition 2.8. □

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

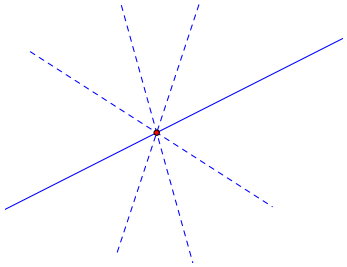


Figure 2.3: Each dashed line is a complement to the undashed subspace.

Here is an application:

Proposition 2.11 (Extension of linear maps)**.** *Let V, W be vector spaces with V finite-dimensional. Let $U \leq V$ be a subspace and $\phi : U \to W$ a linear map. Then there is a linear map $\Phi : V \to W$ such that the restriction[3] of Φ to U is ϕ: $\Phi_{|U} = \phi$. Otherwise said: for all $u \in U$*

$$\Phi(u) = \phi(u).$$

Proof. By Proposition 2.10, U has a complement and so, by Proposition 2.4, there is a projection $\pi : V \to V$ with image U.

Set $\Phi = \phi \circ \pi : V \to W$. This is a linear map and

$$\Phi_{|U} = \phi \circ \pi_{|U} = \phi$$

since, for $u = \pi(v) \in \operatorname{im} \pi = U$, $\pi(u) = \pi(\pi(v)) = \pi(v) = u$. □

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.

[3] Recall that if $f : X \to Y$ is a map of sets and $A \subseteq X$ then the *restriction* of f to A is the map $f_{|A} : A \to Y$ given by $f_{|A}(a) = f(a)$, for all $a \in A$.

Lemma 2.12. *Congruence modulo U is an equivalence relation.*

Proof. Exercise[4]! □

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 $v + U$.

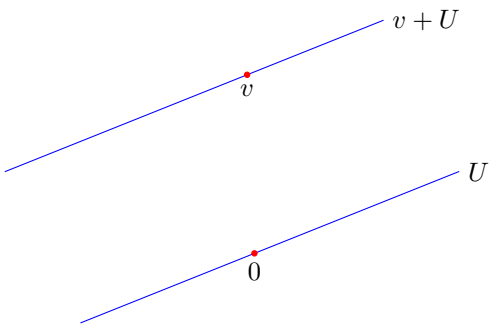


Figure 2.4: A subspace $U \leq \mathbb{R}^2$ and a coset $v + U$.

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).

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*[5] $\mathcal{P}(V)$ of V.

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

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

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

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

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

$$(v + U) + (w + U) := (v + w) + U$$
$$\lambda(v + U) := (\lambda v) + U$$

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 $v \in V$.

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

$$q(v) + q(w) := q(v + w)$$
$$\lambda q(v) := q(\lambda v)$$

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

$$(v + w) - (v' + w') = (v - v') + (w - w') \in U$$
$$\lambda v - \lambda v' = \lambda(v - v') \in U,$$

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 $\ker q = U$. □

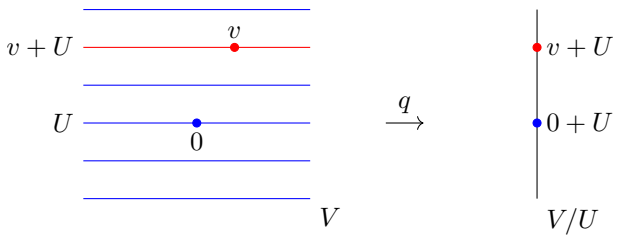


Figure 2.5: The quotient map q.

Corollary 2.14. *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 2.13 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 2.13, from this perspective, is simply that quotients exist!

Theorem 2.15 (First Isomorphism Theorem)**.** *Let $\phi : V \to W$ be a linear map of vector spaces.*

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.

In particular, $V/\ker\phi \cong \operatorname{im}\phi$.

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 (see MA30237) and ring theory (MA20217).

Chapter 3

Inner product spaces

In this chapter, we equip real or complex vector spaces with extra structure that generalises the familiar dot product.

Convention. In this chapter, we take the field $\mathbb{F}$ of scalars to be either $\mathbb{R}$ or $\mathbb{C}$.

3.1 Inner products

3.1.1 Definition and examples

Recall the dot (or scalar) product on $\mathbb{R}^n$: for $x = (x_1, \ldots, x_n), y = (y_1, \ldots, y_n) \in \mathbb{R}^n$,

$$x \cdot y := x_1 y_1 + \cdots + x_n y_n = \mathbf{x}^T \mathbf{y}.$$

Using this we define:

- the *length* of x: $\|x\| := \sqrt{x \cdot x}$;
- the *angle* θ between x and y: $x \cdot y = \|x\| \|y\| \cos\theta$.

There is also a dot product on $\mathbb{C}^n$: for $x, y \in \mathbb{C}^n$,

$$x \cdot y = \bar{x}_1 y_1 + \cdots + \bar{x}_n y_n = \mathbf{x}^\dagger \mathbf{y},$$

where $\mathbf{x}^\dagger$ (pronounced "x-dagger") is the conjugate transpose $\bar{\mathbf{x}}^T$ of $\mathbf{x}$. We then have that

$$x \cdot x = \sum_{i=1}^{n} \bar{x}_i x_i = \sum_{i=1}^{n} |x_i|^2$$

is real, non-negative and vanishes exactly when $x = 0$.

We abstract the key properties of the dot product into the following:

Definition. Let V be a vector space of $\mathbb{F}$ (which is $\mathbb{R}$ or $\mathbb{C}$).

An *inner product* on V is a map $V \times V \to \mathbb{F} : (v, w) \mapsto \langle v, w \rangle$ which is:

(1) *(conjugate) symmetric*: $\langle w, v \rangle = \overline{\langle v, w \rangle}$, for all $v, w \in V$. In particular $\langle v, v \rangle = \overline{\langle v, v \rangle}$ and so is real.
(2) *linear in the second slot*:

$$\langle u, v + w \rangle = \langle u, v \rangle + \langle u, w \rangle$$
$$\langle u, \lambda v \rangle = \lambda \langle u, v \rangle,$$

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

(3) *positive definite:* For all $v \in V$, $\langle v, v\rangle \geq 0$ with equality if and only if $v = 0$.

A vector space with an inner product is called an *inner product space.*

Remark. Any subspace U of an inner product space V is also an inner product space: just restrict $\langle\ ,\ \rangle$ to $U \times U$.

Let us spell out the implications of this definition in the real and complex cases.

Suppose first that $\mathbb{F} = \mathbb{R}$. Then the conjugate symmetry is just symmetry: $\langle v, w\rangle = \langle w, v\rangle$ and it follows that we also have linearity in the first slot:

$$\begin{aligned}\langle v + w, u\rangle &= \langle v, u\rangle + \langle w, u\rangle \\ \langle \lambda v, u\rangle &= \lambda\langle v, u\rangle.\end{aligned}$$

We summarise the situation by saying that a real inner product is a *positive definite, symmetric, bilinear form.* We shall have more to say about bilinear forms later in chapter 6.

Now let us turn to the case $\mathbb{F} = \mathbb{C}$. Now it is not the case that an inner product is linear in the first slot.

Definition. A map $\phi : V \to W$ of complex vector spaces is *conjugate linear* (or *anti-linear*) if

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

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

We see from properties (1) and (2) that a complex inner product has

$$\begin{aligned}\langle v + w, u\rangle &= \langle v, u\rangle + \langle w, u\rangle \\ \langle \lambda v, u\rangle &= \bar{\lambda}\langle v, u\rangle\end{aligned}$$

and so is conjugate linear in the first slot and linear in the second. Such a function is said to be *sesquilinear* (from the Latin *sesqui* which means one-and-a-half). Thus an inner product on a complex vector spaces is a *positive definite, conjugate symmetric, sesquilinear form.*

Definition. Let V be an inner product space.

1. The *norm* of $v \in V$ is $\|v\| := \sqrt{\langle v, v\rangle} \geq 0$.
2. Say $v, w \in V$ are *orthogonal* or *perpendicular* if $\langle v, w\rangle = 0$. In this case, we write $v \perp w$.

Remarks.

1. The norm allows us to define the *distance* between v and w by $\|v - w\|$. We can now do analysis on V: this is one of the Big Ideas in MA20218.
2. **Warning**: There is another convention for complex inner products which is prevalent in Analysis: there they ask that $\langle\ ,\ \rangle$ be linear in the *first* slot and conjugate linear in the second. There are good reasons for either choice.
3. Physicists often write $\langle v|w\rangle$ for $\langle v, w\rangle$. Inner product spaces (especially infinite-dimensional ones) are the setting for quantum mechanics.

Examples.

1. The dot product on $\mathbb{R}^n$ or $\mathbb{C}^n$ is an inner product.
2. Let $[a, b] \subseteq \mathbb{R}$ be a closed, bounded interval. Define a real inner product on $C^0[a, b]$ by

$$\langle f, g\rangle = \int_a^b fg.$$

This is clearly symmetric, bilinear and non-negative. To see that it is definite, one must show that if $\int_a^b f^2 = 0$ then $f = 0$. This is an exercise in Analysis using the inertia property of continuous functions (see MA20218).

3. The set of *square summable sequences* $\ell_2 \subseteq \mathbb{R}^{\mathbb{N}}$ is given by

$$\ell_2 := \{(a_n)_{n\in\mathbb{N}} \mid \sum_{n\in\mathbb{N}} a_n^2 < \infty\}.$$

Exercises.[1]

(a) $\ell_2 \leq \mathbb{R}^{\mathbb{N}}$.

(b) If $a, b \in \ell_2$ then $\sum_{n\in\mathbb{N}} a_n b_n$ is absolutely convergent and then

$$\langle a, b\rangle := \sum_{n\in\mathbb{N}} a_n b_n$$

defines an inner product on ℓ_2.

Hint: for $x, y \in \mathbb{R}$, rearrange $0 \leq (|x| - |y|)^2$ to get

$$2|x||y| \leq x^2 + y^2 \tag{3.1a}$$

and then deduce

$$(x + y)^2 \leq 2(x^2 + y^2). \tag{3.1b}$$

Judicious use of equations (3.1) and the comparison theorem from MA10207 will bake the cake.

Remark. Perhaps surprisingly, ℓ_2 and $C^0[a, b]$ are closely related: this is what Fourier series are about: see MA20223.

3.1.2 Cauchy–Schwarz inequality

Here is one of the most important and ubiquitous inequalities in all of mathematics:

Theorem 3.1 (Cauchy–Schwarz inequality)**.** *Let V be an inner product space. For $v, w \in V$,*

$$|\langle v, w\rangle| \leq \|v\|\|w\| \tag{3.2}$$

with equality if and only if v, w are linearly dependent, that is, either $v = 0$ or $w = \lambda v$, for some $\lambda \in \mathbb{F}$.

Proof. The idea of the proof is to write $w = \lambda v + u$ where $u \perp v$ (see Figure 3.1) and then use the fact that $\|u\|^2 \geq 0$.

In detail, first note that if $v = 0$ then both sides of the inequality vanish and there is nothing to prove. Otherwise, let us seek $\lambda \in \mathbb{F}$ so that $u := w - \lambda v \perp v$. We therefore need

$$0 = \langle v, w - \lambda v\rangle = \langle v, w\rangle - \lambda\langle v, v\rangle$$

so that

$$\lambda = \frac{\langle v, w\rangle}{\|v\|^2}.$$

The situation is shown in Figure 3.1.

With λ and then u so defined we have

$$\begin{aligned} 0 \leq \|u\|^2 &= \langle w - \lambda v, u\rangle = \langle w, u\rangle = \langle w, w - \lambda v\rangle \\ &= \langle w, w\rangle - \lambda\langle w, v\rangle \\ &= \|w\|^2 - \frac{|\langle v, w\rangle|^2}{\|v\|^2}, \end{aligned}$$

where we used the sesquilinearity of the inner product to reach the second line and the conjugate symmetry to reach the third. Rearranging this yields

$$|\langle v, w\rangle|^2 \leq \|v\|^2\|w\|^2$$

and taking a square root gives us the Cauchy–Schwarz inequality.

Finally, we have equality if and only if $\|u\| = 0$ or, equivalently, $u = 0$, that is, $w = \lambda v$. □

[1] Question 7 on sheet 4.

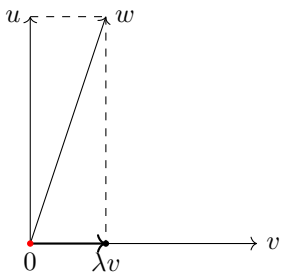


Figure 3.1: Construction of u.

Examples.

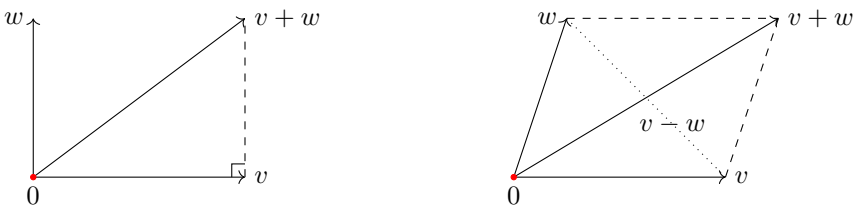
1. Let (Ω, P) be a finite probability space. Then the space $\mathbb{R}^\Omega$ of real random variables is an inner product space with
$$\langle f, g\rangle = \mathbb{E}(fg) = \sum_{x\in\Omega} f(x)g(x)P(x),$$
so long as $P(x) > 0$ for each $x \in \Omega$ (we need this for positive-definiteness). Now the (square of) the Cauchy–Schwarz inequality reads
$$\mathbb{E}(fg)^2 \le \mathbb{E}(f^2)\mathbb{E}(g^2).$$

2. For $a, b \in \ell_2$, the Cauchy–Schwarz inequality reads:
$$\left|\sum_{n\in\mathbb{N}} a_n b_n\right| \le \left(\sum_{n\in\mathbb{N}} a_n^2\right)^{1/2}\left(\sum_{n\in\mathbb{N}} b_n^2\right)^{1/2}.$$

The Cauchy–Schwarz inequality is an essentially 2-dimensional result about the inner product space $\operatorname{span}\{v, w\}$. Here are some more that are almost as fundamental:

Proposition 3.2. *Let V be an inner product space and $v, w \in V$.*

1. ***Pythagoras Theorem**: If $v \perp w$ then*
$$\|v + w\|^2 = \|v\|^2 + \|w\|^2. \tag{3.3}$$
2. ***Triangle inequality**: $\|v + w\| \le \|v\| + \|w\|$ with equality if and only if $v = 0$ or $w = \lambda v$ with $\lambda \ge 0$.*
3. ***Parallelogram identity**: $\|v + w\|^2 + \|v - w\|^2 = 2(\|v\|^2 + \|w\|^2)$.*



(a) Pythagoras' Theorem (b) Parallelogram identity

Figure 3.2: The identities of Proposition 3.2

Proof.

1. Exercise[2]: expand out $\|v+w\|^2 = \langle v+w, v+w\rangle$.
2. We prove $\|v+w\|^2 \leq (\|v\|+\|w\|)^2$. We have
$$\|v+w\|^2 = \|v\|^2 + 2\mathrm{Re}\langle v, w\rangle + \|w\|^2.$$
Now,
$$\mathrm{Re}\langle v, w\rangle \leq |\langle v, w\rangle| \leq \|v\|\|w\|$$
by Cauchy–Schwarz so that
$$\|v+w\|^2 \leq \|v\|^2 + 2\|v\|\|w\| + \|w\|^2 = (\|v\|+\|w\|)^2$$
with equality if and only if $\mathrm{Re}\langle v, w\rangle = |\langle v, w\rangle| = \|v\|\|w\|$ in which case we first get $w = \lambda v$, for some $\lambda \in \mathbb{F}$, and then that $\mathrm{Re}\lambda = |\lambda|$ so that $\lambda \geq 0$.
3. Exercise[3]!

□

3.2 Orthogonality

3.2.1 Orthonormal bases

Definition. A list of vectors $u_1, \ldots, u_k$ in an inner product space V is *orthonormal* if, for all $1 \leq i, j \leq k$,
$$\langle u_i, u_j\rangle = \delta_{ij} := \begin{cases} 1 & \text{if } i = j; \\ 0 & \text{if } i \neq j. \end{cases}$$
If $u_1, \ldots, u_k$ is also a basis, we call it an *orthonormal basis*.

Example. The standard basis $e_1, \ldots, e_n$ of $\mathbb{F}^n$ is orthonormal for the dot product.

Orthonormal bases are very cool. Here is why: if $u_1, \ldots, u_k$ is orthonormal and $v \in \mathrm{span}\{u_1, \ldots, u_k\}$ then we can write
$$v = \lambda_1 u_1 + \cdots + \lambda_k u_k.$$
How can we compute the coordinates λ_i? In general, this amounts to solving a system of linear equations and so involves something tedious and lengthy like Gaussian elimination. However, in our case, things are *much* easier. Observe:
$$\langle u_i, v\rangle = \langle u_i, \sum_j \lambda_j u_j\rangle = \sum_j \lambda_j \langle u_i, u_j\rangle = \sum_j \lambda_j \delta_{ij} = \lambda_i.$$
Thus
$$\lambda_i = \langle u_i, v\rangle \tag{3.4}$$
which is very easy to compute.

Let us enshrine this analysis into the following lemma:

Lemma 3.3. *Let V be an inner product space with orthonormal basis $u_1, \ldots, u_n$ and let $v \in V$. Then*
$$v = \sum_{i=1}^{n} \langle u_i, v\rangle u_i.$$

As an immediate consequence of (3.4):

[2] Question 2(a) on sheet 4.
[3] Question 2(c) on sheet 4.

Lemma 3.4. *Any orthonormal list of vectors $u_1, \ldots, u_k$ is linearly independent.*

Proof. If $\lambda_1 u_1 + \cdots + \lambda_k u_k = 0$ then (3.4) gives $\lambda_i = \langle u_i, 0\rangle = 0$. □

What is more, these coordinates λ_i are all you need to compute inner products.

Proposition 3.5. *Let $u_1, \ldots, u_n$ be an orthonormal basis of an inner product space V.*

Let $v = x_1 u_1 + \cdots + x_n u_n$ and $w = y_1 u_1 + \cdots + y_n u_n$. Then

$$\langle v, w\rangle = \sum_{i=1}^{n} \bar{x}_i y_i = x \cdot y.$$

Thus the inner product of two vectors is the dot product of their coordinates with respect to an orthonormal basis.

Proof. We simply expand out $\langle v, w\rangle$ by sesquilinearity:

$$\langle v, w\rangle = \langle \sum_i x_i u_i, \sum_j y_j u_j\rangle = \sum_{i,j} \bar{x}_i y_j \langle u_i, u_j\rangle = \sum_{i,j} \bar{x}_i y_j \delta_{ij} = \sum_i \bar{x}_i y_i = x \cdot y.$$

□

To put it another way:

Proposition 3.6. *Let $u_1, \ldots, u_n$ be an orthonormal basis of an inner product space V and $v, w \in V$. Then:*

(1) ***Parseval's identity****: $\langle v, w\rangle = \sum_{i=1}^{n} \langle v, u_i\rangle\langle u_i, w\rangle$.*

(2) ***Bessel's equality****: $\|v\|^2 = \sum_{i=1}^{n} |\langle v, u_i\rangle|^2$.*

Proof.

(1) This comes straight from Proposition 3.5, using conjugate symmetry of the inner product to get $\bar{x}_i = \overline{\langle u_i, v\rangle} = \langle v, u_i\rangle$.

(2) Put $v = w$ in (1).

□

All of this should make us eager to get our hands on orthonormal bases and so we would like to know if they always exist. To see that they do, we need the following construction:

Theorem 3.7 (Gram–Schmidt orthogonalisation)**.** *Let $v_1, \ldots, v_m$ be linearly independent vectors in an inner product space V.*

Then there is an orthonormal list $u_1, \ldots, u_m$ such that

$$\operatorname{span}\{u_1, \ldots, u_k\} = \operatorname{span}\{v_1, \ldots, v_k\},$$

for all $1 \leq k \leq m$, defined inductively by:

$$u_k := w_k / \|w_k\|$$

where,

$$w_1 := v_1$$

and, for $k > 1$,

$$w_k := v_k - \sum_{j=1}^{k-1} \langle u_j, v_k\rangle u_j = v_k - \sum_{j=1}^{k-1} \frac{\langle w_j, v_k\rangle}{\|w_j\|^2} w_j.$$

Proof. We induct with inductive hypothesis at k that $u_1, \dots, u_k$ is orthonormal and that, for $1 \le \ell \le k$, $\operatorname{span}\{u_1, \dots, u_\ell\} = \operatorname{span}\{v_1, \dots, v_\ell\}$.

At $k = 1$, this reads $\|u_1\| = 1$ and $\operatorname{span}\{u_1\} = \operatorname{span}\{v_1\}$ which is certainly true.

Now assume the hypothesis is true at $k-1$ so that $u_1, \dots, u_{k-1}$ is orthonormal and $\operatorname{span}\{u_1, \dots, u_{k-1}\} = \operatorname{span}\{v_1, \dots, v_{k-1}\}$. Then

$$\operatorname{span}\{u_1, \dots, u_k\} = \operatorname{span}\{v_1, \dots, v_{k-1}, u_k\} = \operatorname{span}\{v_1, \dots, v_{k-1}, w_k\} = \operatorname{span}\{v_1, \dots, v_k\},$$

where the first equality comes from the induction hypothesis and the last from the definition of w_k. Moreover, for any $i < k$,

$$\langle u_i, w_k\rangle = \langle u_i, v_k\rangle - \sum_{j<k}\langle u_j, v_k\rangle\langle u_i, u_j\rangle = \langle u_i, v_k\rangle - \sum_{j<k}\langle u_j, v_k\rangle\delta_{ij} = \langle u_i, v_k\rangle - \langle u_i, v_k\rangle = 0$$

Thus $w_k \perp u_1, \dots, u_{k-1}$ so that u_k is also whence $u_1, \dots, u_k$ is orthonormal. Thus the inductive hypothesis is true at k and so at m by induction. □

Corollary 3.8. *Any finite-dimensional inner product space V has an orthonormal basis.*

Proof. Let $v_1, \dots, v_n$ be any basis of V and apply Theorem 3.7 to get an orthonormal (and so linearly independent by Lemma 3.4) list $u_1, \dots, u_n$ with

$$\operatorname{span}\{u_1, \dots, u_n\} = \operatorname{span}\{v_1, \dots, v_n\} = V.$$

Thus the $u_1, \dots, u_n$ span also and so are an orthonormal basis. □

Remark. For practical purposes such as computations, it is easiest to phrase the Gram–Schmidt algorithm in terms of the w_k: we set

$$\begin{aligned} w_1 &:= v_1 \\ w_k &:= v_k - \sum_{j=1}^{k-1} \frac{\langle w_j, v_k\rangle}{\|w_j\|^2} w_j \end{aligned}$$

(which can be computed without introducing square roots) and then finally set $u_k = w_k/\|w_k\|$.

Example. Let $U \le \mathbb{R}^3$ be given by $\{x \in \mathbb{R}^3 \mid x_1 + x_2 + x_3 = 0\}$. Let us find an orthonormal basis for U.

First we need a basis of U to start with: $\dim U = 2$ (why?) with basis $v_1 = (1, 0, -1)$, $v_2 = (0, 1, -1)$ (of course, there are many other bases).

First, $\|w_1\|^2 = \|v_1\|^2 = 2$ while $\langle w_1, v_2\rangle = \langle v_1, v_2\rangle = 1$ so that

$$\begin{aligned} w_2 &= v_2 - \frac{\langle w_1, v_2\rangle}{\langle w_1, w_1\rangle} w_1 \\ &= (0, 1, -1) - \tfrac{1}{2}(1, 0, -1) = (-\tfrac{1}{2}, 1, -\tfrac{1}{2}). \end{aligned}$$

This means that $\|w_2\|^2 = 1/4 + 1 + 1/4 = 3/2$ so that

$$\begin{aligned} u_1 &= w_1/\sqrt{2} = (\tfrac{1}{\sqrt{2}}, 0, -\tfrac{1}{\sqrt{2}}) \\ u_2 &= \sqrt{\tfrac{2}{3}} w_2 = \sqrt{\tfrac{2}{3}}(-\tfrac{1}{2}, 1, -\tfrac{1}{2}) = (-\tfrac{1}{\sqrt{6}}, \sqrt{\tfrac{2}{3}}, \tfrac{-1}{\sqrt{6}}). \end{aligned}$$

Let us conclude our discussion of orthonormal bases with an application of Gram–Schmidt which has uses in Statistics (see MA20227) and elsewhere. First, a definition:

Definition. A matrix $Q \in M_{n\times n}(\mathbb{R})$ is *orthogonal* if

$$Q^T Q = \mathrm{I}_n,$$

or, equivalently, Q has orthonormal columns with respect to the dot product. Here I_n is the $n \times n$ identity matrix.

Remark. The two conditions in this definition are indeed equivalent: if $\mathbf{q}_i$ is the i-th column of Q then

$$(Q^TQ)_{ij} = \mathbf{q}_i^T\mathbf{q}_j.$$

Theorem 3.9 (QR decomposition)**.** *Let $A \in M_{n\times n}(\mathbb{R})$ be an invertible matrix. Then we can write*

$$A = QR,$$

where Q is orthogonal and R is upper triangular ($R_{ij} = 0$ if $i > j$) with positive entries on the diagonal.

Proof. We apply Theorem 3.7 to the columns of A to get the columns of Q.

So let $\mathbf{v}_1, \ldots, \mathbf{v}_n$ be the columns of A. Since A is invertible, these are a basis so we can apply Theorem 3.7 to get an orthonormal basis $\mathbf{u}_1, \ldots, \mathbf{u}_n$. Let Q be the orthogonal matrix whose columns are the $\mathbf{u}_i$.

Unravelling the formulae of the Gram–Schmidt procedure, we have

$$\begin{aligned}\mathbf{v}_1 &= \|\mathbf{v}_1\|\mathbf{u}_1\\ \mathbf{v}_2 &= \|\mathbf{w}_2\|\mathbf{u}_2 + \langle\mathbf{u}_1, \mathbf{v}_2\rangle\mathbf{u}_1\end{aligned}$$

and, more generally,

$$\mathbf{v}_k = \|\mathbf{w}_k\|\mathbf{u}_k + \sum_{j<k}\langle\mathbf{u}_j, \mathbf{v}_k\rangle\mathbf{u}_j.$$

Otherwise said, $A = QR$ where $R_{kk} = \|\mathbf{w}_k\|$, $R_{jk} = \langle\mathbf{u}_j, \mathbf{v}_k\rangle$, for $j < k$, and $R_{ij} = 0$ if $i > j$. □

To compute Q and R in practice, first do Gram–Schmidt orthogonalisation on the columns of A to get Q and then note that $Q^TA = Q^TQR = \mathrm{I}_nR = R$ so that

$$R = Q^TA$$

which is probably easier to compute than keeping track of intermediate coefficients in the orthogonalisation!

Remarks.

1. In pure mathematics, the QR decomposition is a special case of the *Iwasawa decomposition*.
2. We shall have more to say about orthogonal matrices in the next chapter, see §4.1.3.

3.2.2 Orthogonal complements and orthogonal projection

Definition. Let V be an inner product space and $U \leq V$. The *orthogonal complement $U^\perp$ of U (in V)* is given by

$$U^\perp := \{v \in V \mid \langle u, v\rangle = 0, \text{ for all } u \in U\}.$$

Proposition 3.10. *Let V be an inner product space and $U \leq V$. Then*

(1) $U^\perp \leq V$;
(2) $U \cap U^\perp = \{0\}$;
(3) $U \leq (U^\perp)^\perp$.

Proof. (1) This is a straightforward exercise using the second slot linearity of the inner product.
(2) If $u \in U \cap U^\perp$, $\langle u, u\rangle = 0$ so that $u = 0$ by positive-definiteness of the inner product.
(3) If $u \in U$ and $w \in U^\perp$ then

$$\langle w, u\rangle = \overline{\langle u, w\rangle} = 0$$

so that $u \in (U^\perp)^\perp$.

□

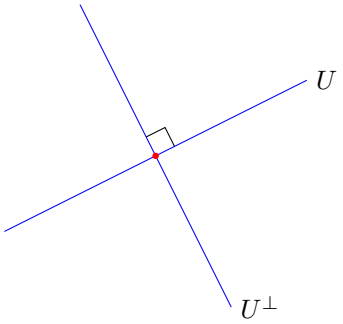


Figure 3.3: Orthogonal complements in $\mathbb{R}^2$

If U is finite-dimensional then $U^\perp$ is a complement to U in the sense of §2.2 (even if V is infinite-dimensional!):

Theorem 3.11. *Let U be a finite-dimensional subspace of an inner product space V. Then V is an internal direct sum:*

$$V = U \oplus U^\perp.$$

Proof.

By Proposition 2.2 and Proposition 3.10(2), we just need to prove that $V = U + U^\perp$. For this, let $u_1, \dots, u_k$ be an orthonormal basis of U and let $v \in V$. We write

$$v = \Big(\sum_{i=1}^{k} \langle u_i, v\rangle u_i\Big) + \Big(v - \sum_{i=1}^{k} \langle u_i, v\rangle u_i\Big) =: v_1 + v_2.$$

Now $v_1 \in U$ being in the span of the u_i while, for $1 \le j \le k$,

$$\begin{aligned}\langle u_j, v_2\rangle &= \langle u_j, v\rangle - \sum_{i=1}^{k} \langle u_i, v\rangle \langle u_j, u_i\rangle \\ &= \langle u_j, v\rangle - \langle u_j, v\rangle = 0.\end{aligned}$$

Thus, for $u = \lambda_1 u_1 + \cdots + \lambda_k u_k \in U$,

$$\langle u, v_2\rangle = \sum_{j=1}^{k} \bar{\lambda}_j \langle u_j, v_2\rangle = 0$$

so that $v_2 \in U^\perp$. □

Corollary 3.12. *Let V be a finite-dimensional inner product space and $U \le V$. Then*

(1) $\dim U^\perp = \dim V - \dim U$.

(2) $U = (U^\perp)^\perp$.

Proof.

(1) This is immediate from Proposition 2.5.

(2) We give two proofs of this.

- Two applications of (1) give

$$\dim(U^\perp)^\perp = \dim V - \dim U^\perp = \dim U$$

while Proposition 3.10(3) gives $U \le (U^\perp)^\perp$. We conclude that we have equality by Lemma 1.4.

- Alternatively, let $v \in (U^\perp)^\perp$ and write $v = v_1 + v_2$ with $v_1 \in U$ and $v_2 \in U^\perp$. Then
$$0 = \langle v_2, v\rangle = \langle v_2, v_1\rangle + \langle v_2, v_2\rangle = \|v_2\|^2$$
so that $v_2 = 0$ giving $v = v_1 \in U$. Thus $(U^\perp)^\perp \leq U$ which, with Proposition 3.10(3) yields $U = (U^\perp)^\perp$. This argument works when V, and even U, are infinite-dimensional so long as $V = U \oplus U^\perp$.

□

Remark. Both Theorem 3.11 and Corollary 3.12(2) can fail when U is infinite-dimensional.

When $V = U \oplus U^\perp$, we can apply the results of §2.2.1 to get projections onto U and $U^\perp$.

Definition. Let V be an inner product space and $U \leq V$ such that $V = U \oplus U^\perp$. The projection $\pi_U : V \to V$ with image U and kernel $U^\perp$ is called the *orthogonal projection onto* U.

Remark. $\pi_{U^\perp} = \mathrm{id}_V - \pi_U$. The situation is illustrated in Figure 3.4.

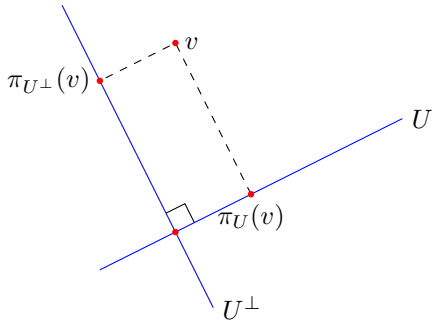


Figure 3.4: Orthogonal projections

Lemma 3.13. *Let V be an inner product space and $U \leq V$ a finite-dimensional subspace with orthonormal basis $u_1, \ldots, u_k$ then, for all $v \in V$,*
$$\pi_U(v) = \sum_{i=1}^{k} \langle u_i, v\rangle u_i.$$

Proof. This is the proof of Theorem 3.11. □

Let us conclude this chapter with an application to a minimisation problem that, among other things, underlies much of Fourier analysis (see MA20223).

Theorem 3.14. *Let V be an inner product space and $U \leq V$ such that $V = U \oplus U^\perp$.*

For $v \in V$, $\pi_U(v)$ is the nearest point of U to v: for all $u \in U$,
$$\|v - \pi_U(v)\| \leq \|v - u\|.$$

Proof. As we see in Figure 3.5, this is just the Pythagoras theorem (Proposition 3.2). Indeed, for $u \in U$, note that $\pi_U(v) - u \in U$ while $v - \pi_U(v) = \pi_{U^\perp}(v) \in U^\perp$. Thus,
$$\begin{aligned}\|v - u\|^2 &= \|v - \pi_U(v) + \pi_U(v) - u\|^2 = \|v - \pi_U(v)\|^2 + \|\pi_U(v) - u\|^2 \\ &\geq \|v - \pi_U(v)\|^2.\end{aligned}$$
Now take square roots! □

Exercise. Read Example 6.58 on pages 199–200 of Axler's *Linear Algebra Done Right* to see a beautiful application of this result. He takes $V = C^0[-\pi, \pi]$ and U to be the space of polynomials of degree at most 5 to get an astonishingly accurate polynomial approximation to sin.

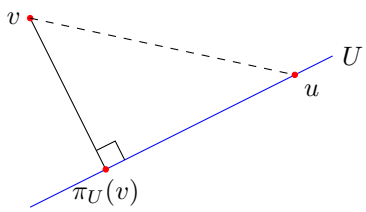


Figure 3.5: The orthogonal projection minimises distance to U

Chapter 4

Linear operators on inner product spaces

Convention. In this chapter, we once again take the field $\mathbb{F}$ of scalars to be either $\mathbb{R}$ or $\mathbb{C}$.

4.1 Linear operators and their adjoints

4.1.1 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 $L(V, V)$).

A special case of the analysis of §1.4.2 tell us that linear operators in the presence of a basis are closely related to square matrices: if V is a finite-dimensional vector space over $\mathbb{F}$ with basis $\mathcal{B} = v_1, \ldots, v_n$ and $\phi \in L(V)$ then the matrix of ϕ with respect to $\mathcal{B}$ is the square matrix $A = (A_{ij}) \in M_{n\times n}(\mathbb{F})$ with

$$\phi(v_j) = \sum_{i=1}^{n} A_{ij} v_i, \tag{4.1}$$

for $1 \leq j \leq n$.

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

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

4.1.2 Adjoints

First a preliminary lemma:

Lemma 4.1 (Nondegeneracy Lemma)**.** *Let V be an inner product space and $v \in V$. Then $\langle v, w\rangle = 0$, for all $w \in V$, if and only if $v = 0$.*

Proof. For the forward implication, take $v = w$ to get $\langle v, v\rangle = 0$ and so $v = 0$ by positive-definiteness of inner product.

Conversely, if $v = 0$, $\langle v, w\rangle = 0$, for any $w \in V$, since the inner product is anti-linear in the first slot[1]. □

[1] To spell it out: $\langle 0, w\rangle = \langle 0 + 0, w\rangle = \langle 0, w\rangle + \langle 0, w\rangle$

Remark. To put this another way: $V^\perp = \{0\}$.

Definition. Let V be an inner product space and $\phi \in L(V)$. An *adjoint to* ϕ is a linear operator $\phi^* \in L(V)$ such that, for all $v, w \in V$, we have

$$\langle \phi^*(v), w \rangle = \langle v, \phi(w) \rangle$$

or, equivalently, by conjugate symmetry,

$$\langle w, \phi^*(v) \rangle = \langle \phi(w), v \rangle.$$

Adjoints are well-behaved under most linear map constructions:

Proposition 4.2. *Let V be an inner product space and suppose $\phi, \psi \in L(V)$ have adjoints. Then $\phi \circ \psi$; $\phi + \lambda\psi$, $\lambda \in \mathbb{F}$; ϕ^* and id_V all have adjoints given by:*

(1) $(\phi \circ \psi)^* = \psi^* \circ \phi^*$ *(note the change of order here!).*

(2) $(\phi + \lambda\psi)^* = \phi^* + \bar{\lambda}\psi^*$.

(3) $(\phi^*)^* = \phi$.

(4) $\mathrm{id}_V^* = \mathrm{id}_V$.

Proof. These are all easy exercises[2]. □

When V is finite-dimensional, any $\phi \in L(V)$ has a unique adjoint:

Proposition 4.3. *Let V be a* finite-dimensional *inner product space and $\phi \in L(V)$ a linear operator. Then*

(1) ϕ has a unique adjoint ϕ^*.

(2) Let $u_1, \ldots, u_n$ be an orthonormal basis of V with respect to which ϕ has matrix A. Then ϕ^* has matrix $A^\dagger := \overline{A}^T$ (which is A^T when $\mathbb{F} = \mathbb{R}$).

Proof. For (1), let $u_1, \ldots, u_n$ be an orthonormal basis of V. Lemma 3.3 tells us that any $u \in V$ can be written

$$u = \sum_{i=1}^{n} \langle u_i, u \rangle u_i.$$

In particular, if ϕ^* existed, we would have

$$\phi^*(v) = \sum_{i=1}^{n} \langle u_i, \phi^*(v) \rangle u_i = \sum_{i=1}^{n} \langle \phi(u_i), v \rangle u_i.$$

Inspired by this, we *define* $\phi^* : V \to V$ by

$$\phi^*(v) = \sum_{i=1}^{n} \langle \phi(u_i), v \rangle u_i. \tag{4.2}$$

We note that ϕ^* so defined is linear since the inner product is linear in the second slot. Moreover, for $w \in V$, we have

$$w = \sum_{i=1}^{n} \langle u_i, w \rangle u_i$$

whence

$$\phi(w) = \sum_{i=1}^{n} \langle u_i, w \rangle \phi(u_i).$$

[2] Question 1 on sheet 6.

Thus,

$$\langle \phi^*(v), w\rangle = \sum_{i=1}^{n} \overline{\langle \phi(u_i), v\rangle}\langle u_i, w\rangle = \sum_{i=1}^{n} \langle v, \phi(u_i)\rangle\langle u_i, w\rangle$$
$$= \langle v, \sum_{i=1}^{n} \langle u_i, w\rangle \phi(u_i)\rangle = \langle v, \phi(w)\rangle.$$

This establishes the existence of ϕ^*.

For uniqueness, either observe that (4.2) uniquely determines ϕ^*, or better (because the argument works for infinite-dimensional V), suppose that ϕ has adjoints ϕ_1^* and ϕ_2^*. Then we have

$$\langle \phi_1^*(v), w\rangle = \langle v, \phi(w)\rangle = \langle \phi_2^*(v), w\rangle$$

so that $\langle \phi_1^*(v) - \phi_2^*(v), w\rangle = 0$ and $\phi_1^*(v) = \phi_2^*(v)$ by the Nondegeneracy Lemma 4.1.

For (2), note that if $\psi \in L(V)$ has matrix B with respect to $u_1, \ldots, u_n$ so that

$$\psi(u_j) = \sum_{i=1}^{n} B_{ij} u_i,$$

then

$$\langle u_k, \psi(u_j)\rangle = \sum_{i=1}^{n} B_{ij}\langle u_k, u_i\rangle = B_{kj}.$$

Taking ψ to be ϕ^* and then ϕ in this, we see that if C is the matrix of ϕ^* then

$$C_{kj} = \langle u_k, \phi^*(u_j)\rangle = \langle \phi(u_k), u_j\rangle = \overline{\langle u_j, \phi(u_k)\rangle} = \overline{A_{jk}}.$$

Thus $C = A^\dagger$. □

Remarks.

1. Proposition 4.3 along with Proposition 4.2(2) tells us that when V is finite-dimensional, $\phi \mapsto \phi^* : L(V) \to L(V)$ is an anti-linear map (and so a linear map when $\mathbb{F} = \mathbb{R}$).
2. The uniqueness part of Proposition 4.3 holds even for infinite-dimensional V: if $\phi \in L(V)$ has an adjoint at all, it has only one.

Example. Let $A \in M_{n\times n}(\mathbb{R})$. Let us show that $(\phi_A)^* = \phi_{A^T}$. Indeed:

$$\phi_{A^T}(y) \cdot x = (A^T\mathbf{y})^T\mathbf{x} = \mathbf{y}^T A\mathbf{x} = y \cdot \phi_A(x)$$

where we have used the familiar identities $(AB)^T = B^T A^T$, for $B \in M_{n\times m}$, and $(A^T)^T = A$.

Similarly, when $\mathbb{F} = \mathbb{C}$, $(\phi_A)^* = \phi_{A^\dagger}$.

Definitions.

1. Let V be an inner product space and $\phi \in L(V)$.
   Say that ϕ is *self-adjoint* if $\phi^* = \phi$, or, equivalently, for all $v, w \in V$,

   $$\langle \phi(v), w\rangle = \langle v, \phi(w)\rangle.$$

   Say ϕ is *skew-adjoint* if $\phi^* = -\phi$ or, equivalently, for all $v, w \in V$,

   $$\langle \phi(v), w\rangle = -\langle v, \phi(w)\rangle.$$

2. Let $A \in M_{n\times n}(\mathbb{F})$.
   (a) If $\mathbb{F} = \mathbb{C}$, say that A is *Hermitian* if $A^\dagger = A$ and *skew-Hermitian* if $A^\dagger = -A$.
   (b) If $\mathbb{F} = \mathbb{R}$, say that A is *symmetric* if $A^T = A$ and *skew-symmetric* if $A^T = -A$.

Remark. If ϕ has matrix A with respect to an orthonormal basis of V then Proposition 4.3(2) tells us that ϕ is self-adjoint if and only if A is Hermitian or symmetric (according to whether $\mathbb{F}$ is $\mathbb{C}$ or $\mathbb{R}$). Similarly, ϕ is skew-adjoint if and only if A is skew-Hermitian or skew-symmetric.

Examples. Using Proposition 4.2, we get

(i) $\mathrm{id}_V^* = \mathrm{id}_V$ so id_V is self-adjoint.

(ii) $\phi^* \circ \phi$ is self-adjoint: $(\phi^* \circ \phi)^* = \phi^* \circ (\phi^*)^* = \phi^* \circ \phi$. Replacing ϕ by ϕ^* shows that $\phi \circ \phi^*$ is self-adjoint also.

(iii) Exercise[3]: when $\mathbb{F} = \mathbb{C}$, ϕ is self-adjoint if and only if $i\phi$ is skew-adjoint (here $i = \sqrt{-1}$!).

4.1.3 Linear isometries

Definition. Let V, W be inner product spaces with inner products $\langle\ ,\ \rangle_V$ and $\langle\ ,\ \rangle_W$ respectively. A linear map $\phi : V \to W$ is a *linear isometry* if, for all $v_1, v_2 \in V$,

$$\langle \phi(v_1), \phi(v_2) \rangle_W = \langle v_1, v_2 \rangle_V .$$

In this case we have that:

- ϕ is *norm-preserving*: $\|\phi(v)\|_W = \|v\|_V$, for all $v \in V$.
- ϕ is *distance-preserving*: $\|\phi(v_1) - \phi(v_2)\|_W = \|v_1 - v_2\|_V$, for all $v_1, v_2 \in V$, (since both sides equal $\|\phi(v_1 - v_2)\|_W$).

Example. Let $\mathcal{B} = u_1, \ldots, u_n$ be an orthonormal basis of an inner product space V. Then $\phi_{\mathcal{B}} : \mathbb{F}^n \to V$ is a linear isometry when $\langle\ ,\ \rangle_{\mathbb{F}^n}$ is dot product. This is the content of Proposition 3.5.

Let us now focus on the case where $V = W$ and is finite-dimensional.

Proposition 4.4. *Let V be a finite-dimensional inner product space and $\phi \in L(V)$. Then ϕ is a linear isometry if and only if ϕ is an isomorphism with $\phi^{-1} = \phi^*$ (equivalently, $\phi^* \circ \phi = \mathrm{id}_V = \phi \circ \phi^*$).*

Proof. We prove the reverse implication first: if $\phi^{-1} = \phi^*$, then $\phi^* \circ \phi = \mathrm{id}_V$ so that, for all $v, w \in V$,

$$\langle v, w \rangle = \langle v, \phi^*(\phi(w)) \rangle = \langle \phi(v), \phi(w) \rangle .$$

Thus ϕ is a linear isometry.

Conversely, if ϕ is a linear isometry then, for all $v, w \in V$,

$$\langle v, w \rangle = \langle \phi(v), \phi(w) \rangle = \langle \phi^* \circ \phi(v), w \rangle$$

so that $\langle v - \phi^* \circ \phi(v), w \rangle = 0$, for all $w \in V$. Thus by the nondegeneracy Lemma 4.1, we get $v = \phi^* \circ \phi = \mathrm{id}_V$, that is:

$$\phi^* \circ \phi = \mathrm{id}_V . \tag{4.3}$$

Since id_V injects, ϕ injects also and so, by Proposition 1.9, is an isomorphism[4]. Now apply ϕ^{-1} on the right of both sides of (4.3) to yield $\phi^* = \phi^{-1}$. □

Remark. If V is infinite-dimensional, linear isometries $V \to V$ need not surject. For example: $Z : \ell_2 \to \ell_2$ given by $(a_0, a_1, \ldots) \mapsto (0, a_0, a_1, \ldots)$ is a non-surjective isometry (exercise[5]!).

Proposition 4.4 prompts some terminology:

Definitions. Let V be an inner product space over $\mathbb{F}$ and $\phi \in L(V)$. If ϕ is an isomorphism with $\phi^{-1} = \phi^*$, then say ϕ is:

[3] Question 3 on sheet 6.
[4] This is the part where we use that V is finite-dimensional.
[5] Question 7 on sheet 6.

- an *orthogonal transformation* if $\mathbb{F} = \mathbb{R}$;
- a *unitary transformation* if $\mathbb{F} = \mathbb{C}$.

The set of all orthogonal, resp. unitary transformations is denoted $\mathrm{O}(V)$, resp. $\mathrm{U}(V)$.

Let $A \in M_{n\times n}(\mathbb{F})$.

- A is *orthogonal* if $\mathbb{F} = \mathbb{R}$ and $A^T A = \mathrm{I}$;
- A is *unitary* if $\mathbb{F} = \mathbb{C}$ and $A^\dagger A = \mathrm{I}$.

The set of all $n \times n$ orthogonal, resp. unitary matrices is denoted $\mathrm{O}(n)$, resp. $\mathrm{U}(n)$.

Remarks.

1. The argument of Proposition 4.4 shows that A is orthogonal if and only if A is invertible with $A^{-1} = A^T$. Similarly A is unitary if and only if A is invertible with $A^{-1} = A^\dagger$.
2. If ϕ has matrix A with respect to an *orthonormal* basis of V then ϕ is an orthogonal or unitary transformation if and only if A is an orthogonal or unitary matrix.

What sort of an object is $\mathrm{O}(n)$ or $\mathrm{U}(n)$? They are *groups*! Recall from Algebra 1A:

A *group* is a set G with a binary operation $G \times G \to G$, written $(g, h) \mapsto gh$, such that:

- $(gh)k = g(hk)$, for all $g, h, k \in G$;
- there is an element $1 \in G$ such that $g1 = 1g = g$, for all $g \in G$. 1 is the *neutral* or *identity element* of G.
- Each $g \in G$ has an *inverse* g^{-1} such that $g^{-1}g = 1 = gg^{-1}$.

A *subgroup* H of a group G is a non-empty subset $H \subset G$ which is closed under multiplication and inverses: whenever $h, k \in H$ then $hk \in H$ and $h^{-1} \in H$. We know that a subgroup is a group in its own right using the multiplication of G.

With our minds refreshed on these matters, we turn to some groups of matrices.

Definitions. Let V be a vector space. The *general linear group of* V, denoted $\mathrm{GL}(V)$, is:

$$\mathrm{GL}(V) := \{\phi \in L(V) \mid \phi \text{ is an isomorphism}\}.$$

Similarly, the *general linear group of* $n \times n$ *matrices over* $\mathbb{F}$, denoted $\mathrm{GL}(n, \mathbb{F})$, is:

$$\mathrm{GL}(n, \mathbb{F}) := \{A \in M_{n\times n}(\mathbb{F}) \mid A \text{ is invertible}\}.$$

Proposition 4.5.

(1) Let V be a vector space. Then $\mathrm{GL}(V)$ is a group under composition: $\psi\phi := \psi \circ \phi$.

(2) If V is an inner product space, then $\mathrm{O}(V)$, resp. $\mathrm{U}(V)$, is a subgroup of $\mathrm{GL}(V)$, when $\mathbb{F} = \mathbb{R}$, resp. $\mathbb{C}$.

Proof. (1) We check the axioms for a group, recalling that $\phi \in L(V)$ is an isomorphism if and only if it has an inverse.

- The multiplication is well-defined: that is, if $\psi, \phi \in \mathrm{GL}(V)$ then so is $\psi \circ \phi$ which has inverse $\phi^{-1} \circ \psi^{-1}$.
- Composition of maps is always associative.
- $\mathrm{id}_V \in \mathrm{GL}(V)$ since $\mathrm{id}_V^{-1} = \mathrm{id}_V$ and is the identity element for composition.
- If $\phi \in \mathrm{GL}(V)$, then its inverse $\phi^{-1} \in \mathrm{GL}(V)$ also since $(\phi^{-1})^{-1} = \phi$.

(2) Denote $\mathrm{O}(V)$ or $\mathrm{U}(V)$, depending on $\mathbb{F}$, by K. Thus

$$K = \{\phi \in L(V) \mid \phi^{-1} = \phi^*\}.$$

Then

- K is closed under composition: if $\psi^{-1} = \psi^*$ and $\phi^{-1} = \phi^*$ then, by Proposition 4.2(1),
$$(\psi \circ \phi)^{-1} = \phi^{-1} \circ \psi^{-1} = \phi^* \circ \psi^* = (\psi \circ \phi)^*$$
so that $\psi \circ \phi \in K$.
- K is non-empty: Proposition 4.2(4) tells us that $\mathrm{id}_V^* = \mathrm{id}_V = \mathrm{id}_V^{-1}$ so $\mathrm{id}_V \in K$.
- K is closed under inversion: using Proposition 4.2(3), we have $(\phi^{-1})^* = (\phi^*)^* = \phi = (\phi^{-1})^{-1}$.

□

Remark. The same argument shows that $\mathrm{GL}(n, \mathbb{F})$ is a group under matrix multiplication with subgroup $\mathrm{O}(n)$ or $\mathrm{U}(n)$, depending on $\mathbb{F}$.

As an application of these ideas, let us characterise the distance-preserving transformations of a real inner product space.

Theorem 4.6 (Classification of rigid motions)**.** *Let V be a* real *inner product space. Recall that the* distance *between $v, w \in V$ is $d(v, w) := \|v - w\|$.*

A map $f : V \to V$ (not necessarily linear) is distance-preserving *or a* rigid motion *if $d(f(v), f(w)) = d(v, w)$, for all $v, w \in V$.*

f is distance-preserving if and only if there is a $v_0 \in V$ and $\phi \in L(V)$ a linear isometry such that
$$f(v) = \phi(v) + v_0, \tag{4.4}$$
for all $v \in V$.

Proof. Suppose that f is distance-preserving and that (4.4) held. Then we would have $f(0) = \phi(0) + v_0 = 0 + v_0$. So, inspired by this, we define
$$\begin{aligned} v_0 &:= f(0) \\ \phi(v) &:= f(v) - v_0, \end{aligned}$$
for all $v \in V$, and observe:
$$\phi(0) = v_0 - v_0 = 0 \tag{4.5a}$$
$$\|\phi(v) - \phi(w)\| = \|v - w\|. \tag{4.5b}$$
We show that ϕ is a linear isometry in four steps:

1. ϕ is norm-preserving: using (4.5), we have
$$\|\phi(v)\| = \|\phi(v) - \phi(0)\| = \|v - 0\| = \|v\|,$$
for all $v \in V$.
2. $\langle \phi(v), \phi(w) \rangle = \langle v, w \rangle$: we know from (4.5b) that $\|\phi(v) - \phi(w)\|^2 = \|v - w\|^2$. Now expand this out, using the fact that the (real) inner product is symmetric, to get
$$\|\phi(v)\|^2 - 2\langle \phi(v), \phi(w) \rangle + \|\phi(w)\|^2 = \|v\|^2 - 2\langle v, w \rangle + \|w\|^2.$$
But, from step 1, we know that $\|\phi(v)\|^2 = \|v\|^2$ and $\|\phi(w)\|^2 = \|w\|^2$ and the result follows.
3. ϕ preserves addition: we compute:
$$\begin{aligned} \|\phi(v + w) - (\phi(v) + \phi(w))\|^2 &= \|\phi(v + w)\|^2 - 2\langle \phi(v + w), \phi(v) + \phi(w) \rangle + \|\phi(v) + \phi(w)\|^2 \\ &= \|\phi(v + w)\|^2 - 2\langle \phi(v + w), \phi(v) \rangle - 2\langle \phi(v + w), \phi(w) \rangle \\ &\quad + \|\phi(v)\|^2 + 2\langle \phi(v), \phi(w) \rangle + \|\phi(w)\|^2 \\ &= \|v + w\|^2 - 2\langle v + w, v \rangle - 2\langle v + w, w \rangle + \|v\|^2 + 2\langle v, w \rangle + \|w\|^2 \\ &= \|v + w\|^2 - 2\|v + w\|^2 + \|v + w\|^2 = 0. \end{aligned}$$
Here we use steps 1 and 2 to reach the third line from the second. We now conclude from the definiteness of the inner product that $\phi(v + w) = \phi(v) + \phi(w)$, for all $v, w \in V$.

4. ϕ preserves scalar multiplication: let $\lambda \in \mathbb{R}$ and compute:

$$\begin{aligned}\|\phi(\lambda v) - \lambda\phi(v)\|^2 &= \|\phi(\lambda v)\|^2 - 2\langle\phi(\lambda v), \lambda\phi(v)\rangle + \|\lambda\phi(v)\|^2 \\ &= \|\phi(\lambda v)\|^2 - 2\lambda\langle\phi(\lambda v), \phi(v)\rangle + \lambda^2\|\phi(v)\|^2 \\ &= \|\lambda v\|^2 - 2\lambda\langle\lambda v, v\rangle + \lambda^2\|v\|^2 \\ &= \lambda^2\|v\|^2 - 2\lambda^2\|v\|^2 + \lambda^2\|v\|^2 = 0,\end{aligned}$$

where, again, we use steps 1 and 2 for the third equality. Definiteness now yields $\phi(\lambda v) = \lambda\phi(v)$, for all $v \in V$, $\lambda \in \mathbb{R}$, so that ϕ is a linear isometry.
Conversely, any f of the form (4.4) is clearly distance-preserving.

□

Remark. Unitary groups play a fundamental role in theoretical physics. At the heart of the Standard Model that unifies electromagnetism with the sub-atomic strong and weak forces is the group $\mathrm{U}(1) \times \mathrm{SU}(2) \times \mathrm{SU}(3)$ where $\mathrm{SU}(n) = \{A \in \mathrm{U}(n) \mid \det A = 1\}$ and I leave it to you to work out how one might make a Cartesian product of groups into a group.

4.2 The spectral theorem

4.2.1 Eigenvalues and eigenvectors

Recall from Chapter 5 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

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

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

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

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

We know that $\Delta_\phi(\lambda)$ is a polynomial in λ with coefficients in $\mathbb{F}$ and $\deg\Delta_\phi = \dim V$. It is important to us because:

Lemma 4.7. *A scalar $\lambda \in \mathbb{F}$ is an eigenvalue of ϕ if and only if $\Delta_\phi(\lambda) = 0$, that is, λ is a root of Δ_ϕ.*

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

Corollary 4.8. *Let ϕ be a linear operator on a finite-dimensional* complex *vector space V. Then ϕ has an eigenvalue.*

Remark. When $\mathbb{F} = \mathbb{R}$ things are more problematic: consider $\phi : \mathbb{R}^2 \to \mathbb{R}^2$ given by $\phi(x_1, x_2) = (-x_2, x_1)$. Thus $\phi = \phi_A$ for

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

Then $\Delta_\phi(\lambda) = \lambda^2 + 1$ which has no real roots at all!

4.2.2 Invariant subspaces and adjoints

Definition. Let V be a vector space and $\phi \in L(V)$.

A subspace $U \leq V$ is *ϕ-invariant* if $\phi(U) \leq U$, that is, $\phi(u) \in U$, for all $u \in U$.

Note that, in this case, ϕ restricts to give a linear operator on U: $\phi_{|U} \in L(U)$.

Example. Any eigenspace of ϕ is ϕ-invariant: if $v \in E_\phi(\lambda)$, then $\phi(v) = \lambda v \in E_\phi(\lambda)$, since $E_\phi(\lambda)$ is a subspace and so closed under scalar multiplication.

Here are two ways to find invariant subspaces.

Lemma 4.9. *Let $\phi, \psi \in L(V)$ and suppose that*

- *$\psi \circ \phi = \phi \circ \psi$ (say that ϕ and ψ* commute*).*
- *$U = E_\phi(\lambda)$ is an eigenspace of ϕ.*

Then U is ψ-invariant.

Proof. Let $u \in U$ so that $\phi(u) = \lambda u$. Take ψ of both sides to get $\psi(\phi(u)) = \lambda\psi(u)$. But $\psi(\phi(u)) = \phi(\psi(u))$ so this reads:

$$\phi(\psi(u)) = \lambda\psi(u).$$

Thus $\psi(u) \in E_\phi(\lambda)$ as required. □

Lemma 4.10. *Let V be a finite-dimensional[6] inner product space and $\phi \in L(V)$.*

Let $U \leq V$ be a ϕ-invariant subspace. Then $U^\perp$ is ϕ^*-invariant.

Proof. Let $v \in U^\perp$ so that $\langle u, v\rangle = 0$, for all $u \in U$. Then

$$\langle u, \phi^*(v)\rangle = \langle \phi(u), v\rangle = 0,$$

for all $u \in U$, since $\phi(u) \in U$ also. Thus $\phi^*(v) \in U^\perp$ as required. □

These two constructions come together for an interesting class of operators:

Definition. Let V be a finite-dimensional inner product space. A linear operator $\phi \in L(V)$ is *normal* if it commutes with its adjoint: $\phi^* \circ \phi = \phi \circ \phi^*$.

Examples.

- Self/skew-adjoint operators are normal: here $\phi^* = \pm\phi$ so
$$\phi^* \circ \phi = \pm\phi^2 = \phi \circ \phi^*.$$
- Unitary/orthogonal transformations are normal: here $\phi^* = \phi^{-1}$ so
$$\phi^* \circ \phi = \mathrm{id}_V = \phi \circ \phi^*.$$

We now have:

Proposition 4.11. *Let V be a finite-dimensional inner product space and $\phi \in L(V)$.*

Suppose that:

- *ϕ is normal;*
- *$U \leq V$ is an eigenspace of ϕ.*

Then $U^\perp$ is ϕ-invariant.

Proof. Since ϕ is normal, ϕ and ϕ^* commute whence U is ϕ^*-invariant by Lemma 4.9. But now Lemma 4.10 tells us that $U^\perp$ is $(\phi^*)^* = \phi$-invariant. □

[6] We only need this hypothesis to ensure that ϕ^* exists.

4.2.3 The spectral theorem for normal operators

Recall from Algebra 1B:

Definition. Let V be a finite-dimensional vector space. A linear operator $\phi \in L(V)$ is *diagonalisable* if V has a basis of eigenvectors of ϕ.

This means that we have a basis $v_1, \ldots, v_n$ of V and scalars $\lambda_1, \ldots, \lambda_n \in \mathbb{F}$ with

$$\phi(v_i) = \lambda_i v_i,$$

for $1 \leq i \leq n$. Equivalently, the matrix A of ϕ with respect to $v_1, \ldots, v_n$ is *diagonal*

$$A = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}, \tag{4.6}$$

with the eigenvalues as diagonal entries.

Exercise.[7] Let $\phi \in L(V)$ be diagonalisable and $v_1, \ldots, v_n$ be a basis of eigenvectors with eigenvalues $\lambda_1, \ldots, \lambda_n$. If λ is an eigenvalue of ϕ then

$$E_\phi(\lambda) = \operatorname{span}\{v_i \mid \lambda_i = \lambda\}.$$

Thus λ appears $\dim E_\phi(\lambda)$ times in the list $\lambda_1, \ldots, \lambda_n$.

Definition. Let V be a finite-dimensional inner product space. A linear operator $\phi \in L(V)$ is *orthogonally diagonalisable* if V has an *orthonormal* basis of eigenvectors.

The Big Question we are going to answer is:

> *When is a linear operator orthogonally diagonalisable?*

We begin by observing that there is a necessary condition:

Proposition 4.12. *Let V be a finite-dimensional inner product space over $\mathbb{F}$ and $\phi \in L(V)$ an orthogonally diagonalisable linear operator. Then:*

(1) If $\mathbb{F} = \mathbb{C}$, ϕ is normal.

(2) If $\mathbb{F} = \mathbb{R}$, ϕ is self-adjoint.

Proof. Let $u_1, \ldots, u_n$ is an orthonormal basis of eigenvectors of ϕ. The matrix A of ϕ with respect to $u_1, \ldots, u_n$ has the form (4.6) so that, by Proposition 4.3(2), ϕ^* has matrix

$$A^\dagger = \begin{pmatrix} \overline{\lambda_1} & & \\ & \ddots & \\ & & \overline{\lambda_n} \end{pmatrix}$$

which is also diagonal. In particular, if $F = \mathbb{R}$ so that each $\lambda_i = \overline{\lambda_i}$, $A = A^T$ whence $\phi = \phi^*$. Otherwise said, ϕ is self-adjoint.

On the other hand, when $\mathbb{F} = \mathbb{C}$, we have $A^\dagger A = AA^\dagger$ (diagonal matrices commute) so that $\phi^* \circ \phi = \phi \circ \phi^*$, that is, ϕ is normal. □

In both cases, the converse is also true. We do the complex case first.

[7] Question 2 on sheet 7.

Theorem 4.13 (Spectral theorem for normal operators)**.** *Let V be a finite-dimensional* complex *inner product space and $\phi \in L(V)$ a linear operator. Then ϕ is orthogonally diagonalisable if and only if ϕ is normal.*

Proof. The forward implication is Proposition 4.12.

We prove the reverse implication by induction on $\dim V$. Thus our inductive hypothesis at n is that the theorem holds when $\dim V \leq n$.

First suppose that $\dim V = 1$ and let $u \in V$ have unit length so that u is an orthonormal basis all on its own. Then any $\phi \in L(V)$ has $\phi(u) \in V$ so that $\phi(u) = \lambda u$, for some $\lambda \in \mathbb{F}$. Thus u is an eigenvector of ϕ and the inductive hypothesis holds at $n = 1$.

Now assume that the hypothesis holds at $n-1$ and suppose that $\dim V = n$ and ϕ is normal. We make three observations:

(1) Since $\mathbb{F} = \mathbb{C}$, ϕ has an eigenvalue λ by Corollary 4.8. We let $U \leq V$ be the λ-eigenspace.

(2) Let $u_1, \ldots, u_k$ be an orthonormal basis of U, which exists by Corollary 3.8. Of course, each u_i is an eigenvector of ϕ with eigenvalue λ.

(3) Since ϕ is normal, Proposition 4.11 tells us that $U^\perp$ is ϕ-invariant. Moreover, $\dim U^\perp = \dim V - \dim U < n$ so we can try to apply the inductive hypothesis to $\phi_{|U^\perp} \in L(U^\perp)$. For this, we need $\phi_{|U^\perp}$ to be normal. However, $U^\perp$ is ϕ^*-invariant by Lemma 4.10 so that $\phi^*_{|U^\perp} \in L(U^\perp)$. Moreover, for $w_1, w_2 \in U^\perp$,
$$\langle \phi^*_{|U^\perp}(w_2), w_1 \rangle = \langle \phi^*(w_2), w_1 \rangle = \langle w_2, \phi(w_1) \rangle = \langle w_2, \phi_{|U^\perp}(w_1) \rangle$$
so that $(\phi_{|U^\perp})^* = \phi^*_{|U^\perp}$. Since ϕ and ϕ^* commute, so do their restrictions to $U^\perp$ and we conclude that $\phi_{|U^\perp}$ is indeed normal. We can therefore apply the inductive hypothesis to get an orthonormal basis $u_{k+1}, \ldots, u_n$ of $U^\perp$ consisting of eigenvectors of $\phi_{|U^\perp}$:
$$\phi_{|U^\perp}(u_j) = \lambda_j u_j,$$
or, equivalently,
$$\phi(u_j) = \lambda_j u_j,$$
for $k+1 \leq j \leq n$.

We conclude that $u_1, \ldots, u_n$ is an orthonormal basis of V with each u_i an eigenvector of ϕ. Thus the inductive hypothesis is true at n and so always. □

Remark. The only place in the argument where we used $\mathbb{F} = \mathbb{C}$ was at step (1) to see that ϕ has an eigenvalue at all. In the next section, we will show that real self-adjoint operators also have eigenvalues and so prove the spectral theorem in that case also.

4.2.4 The spectral theorem for real self-adjoint operators

To prove the spectral theorem in the real case, we need the following lemma which is interesting in its own right.

Lemma 4.14. *Let V be an inner product space*[8] *and $\phi \in L(V)$ be self-adjoint.*

(1) Any eigenvalue of ϕ is real*.*

(2) If $v, w \in V$ are eigenvectors of ϕ with eigenvalues $\lambda \neq \mu$ then $v \perp w$.

Proof. Let $v, w \in V$ be eigenvectors of ϕ with eigenvalues λ, μ. Since ϕ is self-adjoint, we have
$$\langle \phi(v), w \rangle = \langle v, \phi(w) \rangle. \tag{4.7}$$

[8] We do *not* demand that V be finite-dimensional.

However,

$$\langle \phi(v), w\rangle = \langle \lambda v, w\rangle = \bar{\lambda}\langle v, w\rangle$$
$$\langle v, \phi(w)\rangle = \langle v, \mu w\rangle = \mu\langle v, w\rangle$$

so that (4.7) reads

$$(\bar{\lambda} - \mu)\langle v, w\rangle = 0.$$

Now, for (1), take $v = w$ so that $\lambda = \mu$ to get $(\bar{\lambda} - \lambda)\|v\|^2 = 0$ so that $\lambda = \bar{\lambda}$: thus $\lambda \in \mathbb{R}$.

Now, if $\lambda \neq \mu$, we get $(\lambda - \mu)\langle v, w\rangle = 0$ giving $\langle v, w\rangle = 0$, settling (2). □

We apply this to find the missing ingredient for the real spectral theorem.

Proposition 4.15. *A self-adjoint operator ϕ on a real, finite-dimensional inner product space V has an eigenvalue.*

Proof. Choose an orthonormal basis $u_1, \ldots, u_n$ of V and let A be the matrix of ϕ with respect to $u_1, \ldots, u_n$. Then, by Proposition 4.3(2), $A = A^T$

Now view A as a *complex* matrix (with real entries) and let $\phi_A^{\mathbb{C}} : \mathbb{C}^n \to \mathbb{C}^n$ be the complex linear map given by matrix multiplication with A. Observe:

1. $A = A^T = A^\dagger$ so that $\phi_A^{\mathbb{C}}$ is self-adjoint for the dot inner product on $\mathbb{C}^n$.
2. Thus $\phi_A^{\mathbb{C}}$ has an eigenvalue by Proposition 4.8 and, further, that eigenvalue λ is real by Lemma 4.14.
3. Otherwise said, λ is a root of the characteristic polynomial $\Delta_{\phi_A^{\mathbb{C}}}$ of $\phi_A^{\mathbb{C}}$. But

$$\Delta_{\phi_A^{\mathbb{C}}}(\lambda) = \det(A - \lambda \mathrm{I}) = \Delta_\phi(\lambda)$$

so that λ is an eigenvalue of ϕ.

□

We now have:

Theorem 4.16 (Spectral theorem for real self-adjoint operators)**.** *Let V be a real, finite-dimensional inner product space and $\phi \in L(V)$ a linear operator. Then ϕ is orthogonally diagonalisable if and only if ϕ is self-adjoint.*

Proof. The proof is exactly the same as that of the complex spectral Theorem 4.13 except that we use Proposition 4.15 as a replacement for Corollary 4.8 in step 1. □

Remark. A serious application of these ideas is to Quantum Mechanics. Here a physical system is modelled by an inner product space V (usually infinite-dimensional) and the *observables* (things we can measure like position, momentum or spin) by self-adjoint operators. This is a probabilistic theory where all we can hope to find out is the *expected value* of an observable when the system is in a state $v \in V$. However, there are *pure states* where we can measure ϕ with certainty: these are the eigenvectors of ϕ and the corresponding eigenvalue is what we measure in that state.

Let $u_1, \ldots, u_n$ be an orthonormal basis of eigenvectors of ϕ with eigenvalues $\lambda_1, \ldots, \lambda_n$. Then the probability that we will get a measurement of λ_i when in state v is

$$\frac{|\langle u_i, v\rangle|^2}{\|v\|^2}.$$

Note that this lies between 0 and 1 by Cauchy–Schwarz (3.2) and the different probabilities sum to 1 by the Bessel equality (Proposition 3.6(2)).

It follows (exercise!) that the expected value of a measurement of ϕ when in state v

$$\frac{\langle \phi(v), v\rangle}{\|v\|^2}.$$

We need the theory of *Hilbert spaces* (see MA40256) to make sense of all this in the infinite-dimensional case.

4.2.5 The spectral theorem for symmetric and Hermitian matrices

The spectral theorem can be reformulated as a result about symmetric and Hermitian matrices:

Theorem 4.17 (Spectral theorem for symmetric/hermitian matrices)**.**

(1) Let $A \in M_{n\times n}(\mathbb{R})$ be symmetric. Then there is an orthogonal matrix $P \in \mathrm{O}(n)$ such that $P^{-1}AP$ is diagonal.

(2) Let $A \in M_{n\times n}(\mathbb{C})$ be Hermitian. Then there is an unitary matrix $P \in \mathrm{U}(n)$ such that $P^{-1}AP$ is diagonal.

Proof. In either case, $\phi_A : \mathbb{F}^n \to \mathbb{F}^n$ is self-adjoint with respect to the dot inner product and so, by the relevant spectral theorem, orthogonally diagonalisable. So let $\mathcal{B} = u_1, \ldots, u_n$ be an orthonormal basis of eigenvectors of ϕ_A with eigenvalues $\lambda_1, \ldots, \lambda_n$.

Contemplate $\phi_{\mathcal{B}} : \mathbb{F}^n \to \mathbb{F}^n$ for which $\phi_{\mathcal{B}}(e_i) = u_i$. Then $\phi_{\mathcal{B}} = \phi_P$ where $P \in M_{n\times n}(\mathbb{F})$ is the matrix whose columns are $\mathbf{u}_1, \ldots, \mathbf{u}_n$.

Observe:

- The columns of P are orthonormal so that P is orthogonal/unitary.
- $\phi_{P^{-1}AP} = \phi_P^{-1} \circ \phi_A \circ \phi_P$ has eigenvectors $e_1, \ldots, e_n$ with eigenvalues $\lambda_1, \ldots, \lambda_n$:

$$\phi_P^{-1} \circ \phi_A \circ \phi_P(e_i) = \phi_P^{-1}(\phi_A(u_i)) = \lambda_i \phi_P^{-1}(u_i) = \lambda_i e_i.$$

Otherwise said, $P^{-1}AP$ is the diagonal matrix

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

□

Example. Let A be the Hermitian matrix

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

Problem: Find a unitary matrix P such that $P^{-1}AP$ is diagonal.

Solution: P will have for columns an orthonormal basis of eigenvectors of A. Then $P^{-1}AP$ will be diagonal with the corresponding eigenvalues as entries.

First we find the eigenvalues:

$$\det(A - \lambda \mathrm{I}) = \begin{vmatrix} 1-\lambda & i \\ -i & 1-\lambda \end{vmatrix} = \lambda^2 - 2\lambda = \lambda(\lambda - 2)$$

so that the eigenvalues are 0 and 2.

Corresponding eigenvectors are guaranteed to be orthogonal by Lemma 4.14 so all we have to do is find unit length eigenvectors to get an orthonormal basis.

For eigenvalue 0, we must solve $A\mathbf{x} = 0$ which we readily do to get

$$\mathbf{x} = \mu \begin{pmatrix} 1 \\ i \end{pmatrix}$$

and taking $\mu = 1/\sqrt{2}$ we get a unit eigenvector

$$\mathbf{u}_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ i \end{pmatrix}$$

For the eigenvalue 2, we can proceed in two ways: either solve $A\mathbf{y} = 2\mathbf{y}$ or simply seek $\mathbf{y}$ such that $\mathbf{y} \perp \mathbf{x}$. Either way, we get

$$\mathbf{u}_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} i \\ 1 \end{pmatrix}$$

so that

$$P = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & i \\ i & 1 \end{pmatrix}$$

and

$$P^{-1}AP = \begin{pmatrix} 0 & 0 \\ 0 & 2 \end{pmatrix}.$$

We got all this by following the proof of Theorem 4.17 but we can check our answer directly. First we check that P really is unitary:

$$P^\dagger P = \frac{1}{2} \begin{pmatrix} 1 & -i \\ -i & 1 \end{pmatrix} \begin{pmatrix} 1 & i \\ i & 1 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} = \mathrm{I}.$$

Second we check that $P^{-1}AP$ really is the diagonal matrix we say it is:

$$\begin{aligned} P^{-1}AP = P^\dagger AP &= \tfrac{1}{2} \begin{pmatrix} 1 & -i \\ -i & 1 \end{pmatrix} \begin{pmatrix} 1 & i \\ -i & 1 \end{pmatrix} \begin{pmatrix} 1 & i \\ i & 1 \end{pmatrix} \\ &= \tfrac{1}{2} \begin{pmatrix} 0 & 0 \\ -2i & 2 \end{pmatrix} \begin{pmatrix} 1 & i \\ i & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 2 \end{pmatrix}. \end{aligned}$$

4.2.6 Singular value decomposition

Here is an application of the spectral theorem to general linear operators.

Let V be a finite-dimensional inner product space and $\phi \in L(V)$ a linear operator.

Let us first consider the situation where ϕ is self-adjoint. By the spectral theorem, V has an orthonormal basis $u_1, \ldots, u_n$ of eigenvectors with *real* eigenvalues $\lambda_1, \ldots, \lambda_n$ where each distinct eigenvalue λ appears $\dim E_\phi(\lambda)$ times. For $v \in V$, Lemma 3.3 says

$$v = \sum_{i=1}^{n} \langle u_i, v \rangle u_i$$

and taking ϕ of this and using $\phi(u_i) = \lambda_i u_i$ yields:

$$\phi(v) = \sum_{i=1}^{n} \lambda_i \langle u_i, v \rangle u_i. \tag{4.8}$$

But what can be said if ϕ is not self-adjoint or even diagonalisable? One answer is to consider $\phi^* \circ \phi$ instead! Recall from §4.1.2 that $\phi^* \circ \phi$ is always self-adjoint. Moreover

Lemma 4.18. *Let V be a finite-dimensional inner product space and $\phi \in L(V)$. Then:*

(1) All eigenvalues of $\phi^* \circ \phi$ are non-negative.

(2) $\ker(\phi^* \circ \phi) = \ker \phi$.

Proof. If $\phi^*(\phi(u)) = \lambda u$ with $u \neq 0$ (thus u is an eigenvector of $\phi^* \circ \phi$ with eigenvalue λ), then

$$\lambda\langle u, u\rangle = \langle u, \lambda u\rangle = \langle u, \phi^*(\phi(u))\rangle = \langle \phi(u), \phi(u)\rangle$$

so that

$$\lambda = \frac{\|\phi(u)\|^2}{\|u\|^2} \geq 0.$$

This settles (1). For (2), this formula says that $\lambda = 0$ if and only if $\phi(u) = 0$ so that the 0-eigenspace of $\phi^* \circ \phi$ (that is, $\ker \phi^* \circ \phi$) coincides with $\ker \phi$. □

Definition. Let V be a finite-dimensional inner product space and $\phi \in L(V)$. The *singular values of* ϕ are $\sigma_1, \ldots, \sigma_n$ where $\sigma_i = \sqrt{\mu_i} \geq 0$ and $\mu_1, \ldots, \mu_n$ are the eigenvalues of $\phi^* \circ \phi$ listed with multiplicity (thus each distinct μ appears $\dim E_{\phi^* \circ \phi}(\mu)$ times).

Exercise.[9] If ϕ is self-adjoint with eigenvalues λ_i then the singular values of ϕ are the $|\lambda_i|$.

We use these singular values to get a nice analogue of (4.8) for general ϕ at the cost of employing *two* orthonormal bases of V.

Theorem 4.19 (Singular value decomposition)**.** *Let V be a finite-dimensional inner product space and $\phi \in L(V)$ a linear operator with singular values $\sigma_1, \ldots, \sigma_n$.*

Then there are orthonormal bases $u_1, \ldots, u_n$ and $w_1, \ldots, w_n$ of V such that

$$\phi(v) = \sum_{i=1}^{n} \sigma_i \langle u_i, v\rangle w_i, \tag{4.9}$$

for all $v \in V$.

Proof. Let $u_1, \ldots, u_n$ be an orthonormal basis of eigenvectors of $\phi^* \circ \phi$ ordered so that $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n$ and let k be the largest index for which $\sigma_k \neq 0$. Thus, $\sigma_j = 0$ for $k < j \leq n$ so that, by Lemma 4.18(2), $\phi(u_j) = 0$.

For $1 \leq i \leq k$, set $w_i := \phi(u_i)/\sigma_i$. We show that $w_1, \ldots, w_k$ is orthonormal:

$$\langle w_i, w_j\rangle = \frac{1}{\sigma_i \sigma_j}\langle \phi(u_i), \phi(u_j)\rangle = \frac{1}{\sigma_i \sigma_j}\langle u_i, \phi^* \circ \phi(u_j)\rangle = \frac{\sigma_j^2}{\sigma_i \sigma_j}\delta_{ij} = \delta_{ij},$$

for $1 \leq i, j \leq k$. Now extend $w_1, \ldots, w_k$ to an orthonormal basis of V by choosing an orthonormal basis $w_{k+1}, \ldots, w_n$ of $\operatorname{span}\{w_1, \ldots, w_k\}^\perp$. Remark that for $1 \leq i \leq n$ we have

$$\phi(u_i) = \sigma_i w_i.$$

Indeed, for $i \leq k$, this is how we defined w_i while, for $i > k$, both $\sigma_i = 0$ and $\phi(u_i) = 0$.

Thus, for $v \in V$,

$$\phi(v) = \phi\Big(\sum_{i=1}^{n}\langle u_i, v\rangle u_i\Big) = \sum_{i=1}^{n}\langle u_i, v\rangle \phi(u_i) = \sum_{i=1}^{n}\sigma_i\langle u_i, v\rangle w_i$$

as required. □

Remark. This result has real life applications to, among other things, image compression. The idea is that you can approximate a complicated matrix by the terms in the singular value decomposition with the largest singular values.

[9] Question 2 on sheet 8.

Chapter 5

Duality

5.1 Dual spaces

Recall from Theorem 1.6: if V and W are vector spaces over $\mathbb{F}$ then the set $L(V, W)$ of linear maps from V to W is also a vector space under pointwise addition and scalar multiplication. In this chapter we will study the special case where $W = \mathbb{F}$ the field of scalars.

Definition. Let V be a vector space over $\mathbb{F}$. The *dual space* V^* *of* V is

$$V^* := L(V, \mathbb{F}) = \{\alpha : V \to \mathbb{F} \mid \alpha \text{ is linear}\}.$$

Elements of V^* are called *linear functionals* or (less often) *linear forms*.

Let us spell this out. An element $\alpha \in V^*$ is a function $\alpha : V \to \mathbb{F}$ which is linear:

$$\alpha(v_1 + \lambda v_2) = \alpha(v_1) + \lambda\alpha(v_2),$$

for all $v_1, v_2 \in V$ and $\lambda \in \mathbb{F}$. The addition and scalar multiplication on the right are the field addition and multiplication in $\mathbb{F}$.

The dual space V^* is a vector space (indeed a subspace of $\mathbb{F}^V$) under pointwise addition and scalar multiplication. Thus:

$$\begin{aligned}(\alpha_1 + \alpha_2)(v) &:= \alpha_1(v) + \alpha_2(v)\\ (\lambda\alpha)(v) &:= \lambda(\alpha(v)),\end{aligned}$$

for all $\alpha, \alpha_1, \alpha_2 \in V^*$, $v \in V$ and $\lambda \in F$. Again, the algebraic operations on the right hand side of these formulae are those of the field $\mathbb{F}$.

Examples.

1. Fix $\alpha_1, \ldots, \alpha_n \in \mathbb{F}$ and define $\alpha : \mathbb{F}^n \to \mathbb{F}$ by
$$\alpha(x_1, \ldots, x_n) = \alpha_1 x_1 + \cdots + \alpha_n x_n.$$
2. Let $P := \mathbb{R}[t]$ be the vector space of polynomials on $\mathbb{R}$. Here are some linear functionals on P:
   (a) integration over an interval $[a, b]$: $p \mapsto \int_a^b p$.
   (b) Evaluation at a point: for example, $p \mapsto p(\sqrt{2})$.
   (c) Evaluation of a derivative at a point: for example $p \mapsto p'''(\pi)$.
3. Let V be an inner product space and $w \in V$. Define $\alpha_w \in V^*$ by $\alpha_w(v) = \langle w, v\rangle$.
   Note that α_w is indeed linear since the inner product is linear in the second slot.
   Theorem 5.3 will show that when V is finite-dimensional, any element of V^* is of this form.

When V is finite-dimensional, so is V^*. Indeed:

Proposition 5.1. *Let V be a finite-dimensional vector space with basis $v_1, \dots, v_n$.*

Define $v_1^*, \dots, v_n^* \in V^*$ by setting

$$v_i^*(v_j) = \delta_{ij} \in \mathbb{F}$$

and extending by linearity (thus applying Proposition 1.7).

Then $v_1^*, \dots, v_n^*$ is a basis of V^* called the dual basis to $v_1, \dots, v_n$.

Proof. Here is the key computation: if $\sum_{i=1}^n \lambda_i v_i^* \in V^*$ is a linear combination of the v_i^* then evaluating on v_j gives

$$\sum_{i=1}^n \lambda_i v_i^*(v_j) = \sum_{i=1}^n \lambda_i \delta_{ij} = \lambda_j.$$

In particular, if $\sum_{i=1}^n \lambda_i v_i^* = 0$ then each $\lambda_j = 0(v_j) = 0$ and $v_1^*, \dots, v_n^*$ are linearly independent.

Now let $\alpha \in V^*$ and set $\lambda_i = \alpha(v_i)$, for $1 \le i \le n$. Then α and $\sum_{i=1}^n \lambda_i v_i^*$ agree on each v_j and so everywhere:

$$\alpha = \sum_{i=1}^n \alpha(v_i) v_i^*.$$

Thus $v_1^*, \dots, v_n^*$ span. □

Remark. We have met these v_i^* before, perhaps without realising it. Write $v \in V$ in terms of the $v_1, \dots, v_n$: $v = \sum_{j=1}^n \lambda_j v_j$. Then

$$v_i^*(v) = \sum_{j=1}^n \lambda_j v_i^*(v_j) = \lambda_i.$$

Thus v_i^* is the *i-th coordinate function on V with respect to $v_1, \dots, v_n$*.

Corollary 5.2. *If V is finite-dimensional then* $\dim V = \dim V^*$.

When V is an inner product space also, we get a complete understanding of V^*:

Theorem 5.3 (Riesz Representation Theorem)**.** *Let V be a finite-dimensional inner product space and $\alpha \in V^*$. Then there is a unique $w \in V$ such that*

$$\alpha(v) = \langle w, v \rangle,$$

for all $v \in V$. Thus $\alpha = \alpha_w$.

Proof. Let $u_1, \dots, u_n$ be an orthonormal basis of V with dual basis $u_1^*, \dots, u_n^*$. Then

$$u_i^*(u_j) = \delta_{ij} = \langle u_i, u_j \rangle.$$

The uniqueness part of Proposition 1.7 now tells us that

$$u_i^*(v) = \langle u_i, v \rangle, \tag{5.1}$$

for all $v \in V$.

Now write $\alpha = \lambda_1 u_1^* + \cdots + \lambda_n u_n^*$ and use (5.1) to get

$$\alpha(v) = \sum_{i=1}^n \lambda_i u_i^*(v) = \sum_{i=1}^n \lambda_i \langle u_i, v \rangle = \langle \sum_{i=1}^n \bar{\lambda}_i u_i, v \rangle.$$

Thus, setting $w := \sum_{i=1}^n \bar{\lambda}_i u_i$, we get

$$\alpha(v) = \langle w, v \rangle,$$

for all $v \in V$.

For uniqueness, if w' had the same property, we would have $\langle w - w', v \rangle = \alpha(v) - \alpha(v) = 0$, for all $v \in V$, so that $w = w'$ by the Nondegeneracy Lemma 4.1. □

Example. Fix $d \in \mathbb{N}$ and let $V \leq C^0[-1,1]$ be given by

$$V = \{p \in C^0[-1,1] \mid p \text{ is a polynomial of degree no more than } d\}.$$

Then V is a finite-dimensional inner product space with inner product

$$\langle p, q \rangle = \int_{-1}^{1} pq.$$

Define $\alpha : V \to \mathbb{R}$ by $\alpha(p) = p(0) + p'(1/\sqrt{2})$. We have already noted that $\alpha \in V^*$.

So Theorem 5.3 tells us that there is a $q \in V$ such that $\alpha(p) = \langle q, p \rangle$, for all $p \in V$. Otherwise said, there is a clever polynomial q of degree d such that

$$\int_{-1}^{1} qp = p(0) + p'(1/\sqrt{2}),$$

for all polynomials of degree no more than d!

A basic question is how big is V^*: are there enough linear functionals to detect all elements of V? The answer is yes and the key is the following theorem:

Theorem 5.4 (Sufficiency principle)**.** *Let V be a vector space and $v \in V$. Then $\alpha(v) = 0$, for all $\alpha \in V^*$, if and only if $v = 0$.*

Proof. A complete proof requires a tool from set theory called Zorn's Lemma, equivalent to the Axiom of Choice, which has the faintly controversial property that it is logically independent from the usual axioms of set theory (so you can choose to believe it or not without running into a contradiction). Rather than get involved in all that we simply prove the result in two cases of interest.

(1) If V is an inner product space and $v \in V$ is non-zero then $\alpha_v(v) = \langle v, v \rangle \neq 0$, where $\alpha_v \in V^*$ is given by $\alpha_v(w) = \langle v, w \rangle$.

(2) If V is finite-dimensional, choose a basis $v_1, \ldots, v_n$. For $v \in V$, write $v = \lambda_1 v_1 + \cdots + \lambda_n v_n$. If $\alpha(v) = 0$ for all $\alpha \in V^*$ then, in particular, for each i, $0 = v_i^*(v) = \lambda_i$ so that $v = 0$.

□

Exercise.[1] Let $v \in V$ and $U \leq V$ with $v \notin U$. Show that there is $\alpha \in V^*$ such that $\alpha(v) \neq 0$ while $\alpha_{|U} = 0$.

Hint: apply Theorem 5.4 to V/U.

We apply Theorem 5.4 to get a converse to Proposition 5.1:

Proposition 5.5. *Let V be a finite-dimensional vector space and $\alpha_1, \ldots, \alpha_n$ a basis of V^*. Then there is a basis $v_1, \ldots, v_n$ of V such that*

$$\alpha_i(v_j) = \delta_{ij}.$$

Thus $\alpha_i = v_i^*$, for $1 \leq i \leq n$.

Proof. Define a linear map $\phi : V \to \mathbb{F}^n$ by

$$\phi(v) = (\alpha_1(v), \ldots, \alpha_n(v))$$

and observe that $v \in \ker \phi$ if and only if $\alpha_i(v) = 0$, for $1 \leq i \leq n$, whence, since any $\alpha \in V^*$ is a linear combination of the α_i, $\alpha(v) = 0$, for all $\alpha \in V^*$. We deduce from Theorem 5.4 that $v = 0$ so that $\ker \phi = \{0\}$ and ϕ is injective. On the other hand, $\dim V = \dim V^* = n = \dim \mathbb{F}^n$ so that ϕ is an isomorphism.

[1] Question 5 on sheet 8.

Now set $v_i = \phi^{-1}(e_i)$, $1 \leq i \leq n$, to get a basis of V since $e_1, \ldots, e_n$ is a basis of $\mathbb{F}^n$. Then

$$\phi(v_j) = (\alpha_1(v_j), \ldots, \alpha_n(v_j)) = e_j = (0, \ldots, 1, \ldots, 0),$$

where the 1 is in the j-th slot. Otherwise said, $\alpha_i(v_j) = \delta_{ij}$ as required. □

Since the dual space V^* is a vector space, we can contemplate its dual space $V^{**} := (V^*)^*$, the *double dual of* V. This is closely related to V itself. Indeed, each $v \in V$ defines a linear map $\mathrm{ev}(v) : V^* \to \mathbb{F}$ by evaluation at v:

$$\mathrm{ev}(v)(\alpha) := \alpha(v) \in \mathbb{F}.$$

Exercises.[2]

1. $\mathrm{ev}(v)$ is indeed linear: for $\alpha, \beta \in V^*$ and $\lambda \in \mathbb{F}$,

   $$\mathrm{ev}(v)(\alpha + \lambda\beta) = \mathrm{ev}(v)(\alpha) + \lambda\,\mathrm{ev}(v)(\beta).$$

   Thus $\mathrm{ev}(v) \in V^{**}$.
2. We therefore have a map $\mathrm{ev} : V \to V^{**}$. Show that ev is linear: that is,

   $$\mathrm{ev}(v + \lambda w) = \mathrm{ev}(v) + \lambda\,\mathrm{ev}(w),$$

   for all $v, w \in V$, $\lambda \in \mathbb{F}$. To spell it out even more, this means

   $$\mathrm{ev}(v + \lambda w)(\alpha) = \mathrm{ev}(v)(\alpha) + \lambda\,\mathrm{ev}(w)(\alpha),$$

   for all $\alpha \in V^*$.
3. ev is injective (use Theorem 5.4) and so, when V is finite-dimensional, an isomorphism since $\dim V = \dim V^* = \dim V^{**}$.

Thus:

Theorem 5.6. *If V is a finite-dimensional vector space then $\mathrm{ev} : V \to V^{**}$ is an isomorphism.*

Remark. In general, a vector space for which $\mathrm{ev} : V \to V^{**}$ is an isomorphism is said to be *reflexive*.

5.2 Solution sets and annihilators

Here is one way to think about V^*: consider the equation

$$\alpha(v) = 0, \tag{5.2}$$

for some $\alpha \in V^*$ and $v \in V$. If we choose dual bases $v_1, \ldots, v_n$ and $v_1^*, \ldots, v_n^*$ of V and V^*, (5.2) reads

$$\alpha_1 x_1 + \cdots + \alpha_n x_n = 0$$

where we have written $\alpha = \alpha_1 v_1^* + \cdots + \alpha_n v_n^*$ and $v = x_1 v_1 + \cdots + x_n v_n$. This is a single homogeneous linear equation.

This gives us the idea of viewing V^* as the set of linear equations on V. From this point of view, a subspace $E \leq V^*$ should be viewed as a system of linear equations and so we should be interested in the solutions of that system:

Definition. Let $E \leq V^*$. The *solution set of* E is

$$\mathrm{sol}\, E := \{v \in V \mid \alpha(v) = 0, \text{ for all } \alpha \in E\} = \bigcap_{\alpha \in E} \ker \alpha \leq V.$$

[2] Question 7 on sheet 8.

Exercise.[3] If $\alpha_1, \ldots, \alpha_k$ span E then

$$\operatorname{sol} E = \bigcap_{i=1}^{k} \ker \alpha_i.$$

For finite-dimensional V, one might expect each equation in a linear system to reduce the dimension of the solution set by one and this is exactly what happens:

Proposition 5.7. *If V is finite-dimensional and $E \leq V^*$ then*

$$\dim \operatorname{sol} E = \dim V - \dim E.$$

We say that E and $\operatorname{sol} E$ have complementary dimension.

Proof. Let $v_1^*, \ldots, v_k^*$ be a basis of E and extend to a basis $v_1^*, \ldots, v_n^*$ of V^*. Let $v_1, \ldots, v_n$ be the dual basis of V provided by Proposition 5.5.

Now $E = \operatorname{span}\{v_1^*, \ldots, v_k^*\}$ so that $\operatorname{sol} E = \bigcap_{i=1}^{k} \ker v_i^*$. Thus $v = \sum_{j=1}^{n} \lambda_j v_j$ lies in $\operatorname{sol} E$ if and only if $\lambda_i = v_i^*(v) = 0$, for $1 \leq i \leq k$. Otherwise said,

$$\operatorname{sol} E = \operatorname{span}\{v_{k+1}, \ldots, v_n\}$$

so that

$$\dim \operatorname{sol} E = n - k = \dim V - \dim E.$$

□

Remark. Here is a slicker argument. Let $\operatorname{ev}^E : V \to E^*$ be the linear map given by

$$\operatorname{ev}^E(v)(\alpha) = \alpha(v).$$

1. $\operatorname{im} \operatorname{ev}^E = E^*$: for this, you use Theorem 5.6 along with the fact that restriction to E is a surjection from V^{**} to E^* thanks to Proposition 2.11.
2. $\ker \operatorname{ev}^E = \{v \in V \mid \alpha(v) = 0, \text{ for all } \alpha \in E\} = \operatorname{sol} E$.

So rank-nullity tells us that

$$\dim \operatorname{sol} E + \dim E^* = \dim V$$

and, since $\dim E = \dim E^*$, we are done.

Corollary 5.8. *Let V have dimension n and suppose that $\alpha_1, \ldots, \alpha_n \in V^*$ are such that*

$$\bigcap_{i=1}^{n} \ker \alpha_i = \{0\}.$$

Then $\alpha_1, \ldots, \alpha_n$ is a basis of V^*.

Proof. Let $E := \operatorname{span}\{\alpha_1, \ldots, \alpha_n\}$. The hypothesis says that $\operatorname{sol} E = \{0\}$ so, by Proposition 5.7, $\dim E = n$ whence $E = V^*$. Thus $\alpha_1, \ldots, \alpha_n$ span V^* and so are a basis. □

Here is an application:

Example. Let P_2 be the vector space of polynomials of degree at most 2. Thus $\dim P_2 = 3$.

Define $\alpha_i : P_2 \to \mathbb{R}$, $i = 1, 2, 3$, by

$$\begin{aligned} \alpha_1(p) &= p(1) \\ \alpha_2(p) &= p(\sqrt{2}) \\ \alpha_3(p) &= p(\pi), \end{aligned}$$

[3] Question 1 on sheet 9.

for all $p \in P_2$. These are all linear maps so that $\alpha_1, \alpha_2, \alpha_3 \in P_2^*$. We apply Corollary 5.8 so see that $\alpha_1, \alpha_2, \alpha_3$ are a basis of P_2^*. Indeed, if $p \in \bigcap_{i=1}^3 \ker \alpha_i$ then $p(1) = p(\sqrt{2}) = p(\pi) = 0$ so that p has three distinct roots and so must vanish since it has degree no more than 2.

Thus any $\alpha \in P_2^*$ is a linear combination of the α_i. For example, define α by

$$\alpha(p) = \int_0^1 p.$$

Then there are $\lambda_1, \lambda_2, \lambda_3 \in \mathbb{R}$ such that $\alpha = \lambda_1\alpha_1 + \lambda_2\alpha_2 + \lambda_3\alpha_3$. Otherwise said, we have found clever λ_i such that, for all $p \in P_2$,

$$\int_0^1 p = \lambda_1 p(1) + \lambda_2 p(\sqrt{2}) + \lambda_3 p(\pi).$$

Solution sets behave somewhat like orthogonal complements (except that E and $\operatorname{sol} E$ live in entirely different vector spaces):

Proposition 5.9. *Let $E, F \leq V^*$. Then*

(1) If $E \leq F$ then $\operatorname{sol} F \leq \operatorname{sol} E$.

(2) sol *swaps sums and intersections:*

$$\operatorname{sol}(E + F) = (\operatorname{sol} E) \cap (\operatorname{sol} F)$$
$$(\operatorname{sol} E) + (\operatorname{sol} F) \leq \operatorname{sol}(E \cap F)$$

with equality if V is finite-dimensional.

Proof.

(1) Let $E \leq F$ and $v \in \operatorname{sol} F$. Then $\alpha(v) = 0$, for all $\alpha \in F$ and so, in particular, for all $\alpha \in E$. Thus $v \in \operatorname{sol} E$.

(2) This is an exercise[4] similar to one you have already done about orthogonal complements[5].

□

Still thinking of V^* as the linear equations on V, we can turn things around and ask which equations the elements of a subspace $U \leq V$ satisfy:

Definition. Let $U \leq V$. The *annihilator of U, denoted* $\operatorname{ann} U$ *or* U°, is given by:

$$\operatorname{ann} U := \{\alpha \in V^* \mid \alpha_{|U} = 0\} = \{\alpha \in V^* \mid \alpha(u) = 0, \text{ for all } u \in U\ \}.$$

Exercise.[6] Show that $\operatorname{ann} U \leq V^*$.

Annihilators have very similar properties to solution sets. They also have complementary dimension:

Proposition 5.10. *Let V be finite-dimensional and $U \leq V$. Then*

$$\dim \operatorname{ann} U = \dim V - \dim U.$$

Proof. This is an exercise[7] in imitating the proof of Proposition 5.7: start with a basis $v_1, \ldots, v_k$ of U, extend to a basis $v_1, \ldots, v_n$ of V and see that $\operatorname{ann} U = \operatorname{span}\{v_{k+1}^*, \ldots, v_n^*\}$. Can you find a slick argument? □

Again annihilators swap the order of inclusions and sums with intersections:

[4] Question 4(a) on sheet 9.
[5] Question 4 on sheet 6.
[6] Question 2 on sheet 9.
[7] Question 3 on sheet 9.

Proposition 5.11. *Let $U, W \leq V$. Then*

(1) If $U \leq W$ then $\operatorname{ann} W \leq \operatorname{ann} U$.

(2)
$$\operatorname{ann}(U + W) = (\operatorname{ann} U) \cap (\operatorname{ann} W)$$
$$(\operatorname{ann} U) + (\operatorname{ann} W) \leq \operatorname{ann}(U \cap W)$$

with equality if V is finite-dimensional.

Proof. This is an exercise[8] in imitating the proof of Proposition 5.9. □

What is the relation between annihilators and solution sets?

Lemma 5.12. *Let $U \leq V$ and $E \leq V^*$ then $U \leq \operatorname{sol} E$ if and only if $E \leq \operatorname{ann} U$.*

Proof. Both inclusions amount to saying $\alpha(u) = 0$, for all $u \in U$ and $\alpha \in E$. □

With this in hand, we have:

Theorem 5.13. *Let $U \leq V$ and $E \leq V^*$. Then*

$$U \leq \operatorname{sol}(\operatorname{ann} U)$$
$$E \leq \operatorname{ann}(\operatorname{sol} E),$$

with equality if V is finite-dimensional.

Proof. Clearly $\operatorname{ann} U \leq \operatorname{ann} U$ so putting $E = \operatorname{ann} U$ in Lemma 5.12 gives

$$U \leq \operatorname{sol}(\operatorname{ann} U).$$

Similarly, $\operatorname{sol} E \leq \operatorname{sol} E$ so Lemma 5.12 gives

$$E \leq \operatorname{ann}(\operatorname{sol} E).$$

If V is finite-dimensional,

$$\dim \operatorname{sol}(\operatorname{ann} U) = \dim V - \dim \operatorname{ann} U = \dim U$$

so that $U = \operatorname{sol}(\operatorname{ann} U)$. Similarly, $E = \operatorname{ann}(\operatorname{sol} E)$. □

Remark. We can view ann and sol as maps:

$$\operatorname{ann} : \{\text{subspaces of } V\} \to \{\text{subspaces of } V^*\}$$
$$\operatorname{sol} : \{\text{subspaces of } V^*\} \to \{\text{subspaces of } V\}.$$

When V is finite-dimensional, Theorem 5.13 is telling us that these maps are mutually inverse bijections. This has a beautiful application to geometry that you can see in MA30231.

5.3 Transposes

There is a duality construction for linear maps also: let V, W be vector spaces, $\phi \in L(V, W)$ and $\alpha \in W^*$. Then $\alpha \circ \phi : V \to \mathbb{F}$ is also linear, so that $\alpha \circ \phi \in V^*$. This prompts:

Definition. Let $\phi \in L(V, W)$ be a linear map of vector spaces. The *transpose* ϕ^T *of* ϕ is the map $\phi^T : W^* \to V^*$ given by

$$\phi^T(\alpha) := \alpha \circ \phi,$$

for all $\alpha \in W^*$.

[8] Question 4(b) on sheet 9.

Lemma 5.14. *$\phi^T : W^* \to V^*$ is also a linear map.*

Proof. Let $\alpha, \beta \in W^*$ and $\lambda \in \mathbb{F}$. We must show that

$$\phi^T(\alpha + \lambda\beta) = \phi^T(\alpha) + \lambda\phi^T(\beta).$$

Unravelling the definition, this means

$$(\alpha + \lambda\beta) \circ \phi = \alpha \circ \phi + \lambda\beta \circ \phi.$$

This is an equality of functions and so holds exactly when

$$(\alpha + \lambda\beta)(\phi(v)) = \alpha(\phi(v)) + \lambda(\beta(\phi(v))),$$

for all $v \in V$. However, this last is true by the very definition of addition and scalar multiplication in W^*. □

Examples.

1. $\mathrm{id}_V^T = \mathrm{id}_{V^*}$. Indeed, $\mathrm{id}_V^T(\alpha) = \alpha \circ \mathrm{id}_V = \alpha$, for all $\alpha \in V^*$.
2. $(\psi \circ \phi)^T = \phi^T \circ \psi^T$. Indeed, $(\psi \circ \phi)^T(\alpha) = \alpha \circ \psi \circ \phi = \phi^T(\alpha \circ \psi) = \phi^T(\psi^T(\alpha))$.

Here is why ϕ^T is called the transpose of ϕ:

Proposition 5.15. *Let V, W be finite-dimensional vector spaces and $\phi \in L(V, W)$ with matrix $A \in M_{m\times n}(\mathbb{F})$ with respect to bases $v_1, \ldots, v_n$ and $w_1, \ldots, w_m$ of V and W.*

Then ϕ^T has matrix A^T with respect to the dual bases $w_1^*, \ldots, w_m^*$ and $v_1^*, \ldots, v_n^*$ of W^* and V^*.

Proof. Let ϕ^T have matrix B so that

$$\phi^T(w_j^*) = \sum_{i=1}^{n} B_{ij} v_i^*.$$

Evaluate both sides of this at v_k to get

$$\phi^T(w_j^*)(v_k) = B_{kj}$$

or, unravelling the definition of ϕ^T,

$$w_j^*(\phi(v_k)) = B_{kj}.$$

Now

$$\phi(v_k) = \sum_{i=1}^{m} A_{ik} w_i$$

so that we also get

$$w_j^*(\phi(v_k)) = A_{jk}.$$

Comparing these we get $B_{kj} = A_{jk}$ whence $B = A^T$. □

The kernels and images of ϕ and ϕ^T are intimately related via the annihilators and solution sets of §5.2:

Theorem 5.16. *Let $\phi \in L(V, W)$ be a linear map of vector spaces. Then*

(1)
$$\ker \phi = \mathrm{sol}(\mathrm{im}\, \phi^T)$$
$$\mathrm{im}\, \phi \leq \mathrm{sol}(\ker \phi^T)$$

with equality if V, W are finite-dimensional.

(2)
$$\ker \phi^T = \operatorname{ann}(\operatorname{im} \phi)$$
$$\operatorname{im} \phi^T \leq \operatorname{ann}(\ker \phi)$$

with equality if V, W are finite-dimensional.

Proof. We will prove (1) and leave (2) as an exercise[9].

For the first equality, observe that $v \in \ker \phi$ if and only if $\phi(v) = 0$ or, equivalently, by Theorem 5.4, $\alpha(\phi(v)) = 0$, for all $\alpha \in W^*$, which is the same as $\phi^T(\alpha)(v) = 0$, for all $\alpha \in W^*$, that is, $v \in \operatorname{sol}(\operatorname{im} \phi^T)$.

If V, W are finite-dimensional we now use this, along with rank-nullity and Proposition 5.7, to get

$$\dim V - \dim \operatorname{im} \phi = \dim \ker \phi = \dim \operatorname{sol}(\operatorname{im} \phi^T) = \dim V - \dim \operatorname{im} \phi^T$$

so that

$$\operatorname{rank} \phi = \dim \operatorname{im} \phi = \dim \operatorname{im} \phi^T = \operatorname{rank} \phi^T. \tag{5.3}$$

For $\operatorname{im} \phi \leq \operatorname{sol}(\ker \phi^T)$, let $w \in \operatorname{im} \phi$ and $\alpha \in \ker \phi^T$ so that $\alpha \circ \phi = 0$ and $w = \phi(v)$, for some $v \in V$. Then $\alpha(w) = \alpha(\phi(v)) = (\alpha \circ \phi)(v) = 0$ so that $w \in \operatorname{sol}(\ker \phi^T)$. Thus $\operatorname{im} \phi \leq \operatorname{sol}(\ker \phi^T)$.

Moreover, if V, W are finite-dimensional, use (5.3), rank-nullity and Proposition 5.7 to get

$$\dim \operatorname{im} \phi = \dim \operatorname{im} \phi^T = \dim W - \dim \ker \phi^T = \dim \operatorname{sol}(\ker \phi^T).$$

We conclude that $\operatorname{im} \phi$ and $\dim \operatorname{sol}(\ker \phi^T)$ have the same dimension and so coincide. □

Along the way, we got (5.3):

Corollary 5.17. *Let $\phi \in L(V, W)$ be a linear map of finite-dimensional vector spaces. Then*

$$\operatorname{rank} \phi = \operatorname{rank} \phi^T.$$

Remark. This gives us a new take on an old result from Algebra 1B. Let $A \in M_{m \times n}(\mathbb{F})$ be the matrix of ϕ with respect to bases of V and W so that, by Proposition 5.15, A^T is the matrix of ϕ^T with respect to the dual bases. Then the rank of ϕ is the column rank of A while the rank of ϕ^T is the column rank of A^T which is the row rank of A. Thus row rank and column rank coincide.

The punchline of Theorem 5.16 is that ϕ and ϕ^T have "opposite" properties. For example:

Proposition 5.18. *Let $\phi \in L(V, W)$ be a linear map of finite-dimensional vector spaces. Then*

(1) ϕ injects if and only if ϕ^T surjects.

(2) ϕ^T injects if and only if ϕ surjects.

Proof. For (1), ϕ injects if and only if $\ker \phi = \{0\}$ while ϕ^T surjects if and only if $\dim \operatorname{im} \phi^T = \dim V$. By Theorem 5.16, the first happens if and only if $\operatorname{sol}(\operatorname{im} \phi^T) = \{0\}$ but, by Proposition 5.7, this is equivalent to the $\dim \operatorname{im} \phi^T = \dim V$.

Item (2) is similar. □

Remarks.

1. This result is useful as it is sometimes easier to prove injectivity than surjectivity.
2. With a bit more effort, we can do better than Proposition 5.18: for example, using Theorem 5.4, we can prove that Proposition 5.18(2) holds even in infinite dimensions.

[9] Question 5 on sheet 9.

Chapter 6

Bilinearity

We give an introduction to a general theory of "multiplication" of vectors.

6.1 Bilinear maps

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

$$B(\lambda u_1 + u_2, v) = \lambda B(u_1, v) + B(u_2, v)$$
$$B(u, \lambda v_1 + v_2) = \lambda B(u, v_1) + B(u, v_2),$$

for all $u, u_1, u_2 \in U$, $v, v_1, v_2 \in V$ and $\lambda \in \mathbb{F}$.

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

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

Remark. A bilinear map $B : U \times V \to W$ has $B(u, 0) = B(0, v) = 0$, for all $u \in U$ and $v \in V$. Indeed,

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

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

Examples.

1. Matrix multiplication is bilinear:
$$(A, B) \mapsto AB : M_{m\times n}(\mathbb{F}) \times M_{n\times k}(\mathbb{F}) \to M_{m\times k}(\mathbb{F}).$$
2. Composition of maps is bilinear:
$$(\psi, \phi) \mapsto \psi \circ \phi : L(U, W) \times L(V, U) \to L(V, W).$$
3. Evaluation $(\alpha, v) \mapsto \alpha(v) : V^* \times V \to \mathbb{F}$ is a bilinear pairing.
4. Any *real* inner product is a bilinear form (what goes wrong for complex inner products?).
5. Let $A \in M_{m\times n}(\mathbb{F})$ and define a bilinear pairing $B_A : \mathbb{F}^m \times \mathbb{F}^n \to \mathbb{F}$ by
$$B_A(x, y) = \mathbf{x}^T A\mathbf{y}.$$
This gives us a new use for matrices.

Notation. We let $\mathrm{Bil}(U, V; W)$ denote the set of bilinear maps $U \times V \to W$.

Exercise. Show that $\mathrm{Bil}(U, V; W) \leq W^{U\times V}$. Otherwise said, $\mathrm{Bil}(U, V; W)$ is a vector space under pointwise addition and scalar multiplication.

6.2 Bilinear forms and quadratic forms

We focus on the simplest case: bilinear forms $B : V \times V \to \mathbb{F}$.

6.2.1 Bilinear forms and matrices

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 $A \in M_{n\times n}(\mathbb{F})$ 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 6.1. *Let $B : V \times V \to \mathbb{F}$ be a bilinear form with matrix A with respect to $\mathcal{B} = v_1, \ldots, v_n$. 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*

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

or, equivalently, for all $x, y \in \mathbb{F}^n$,

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

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

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

□

Remarks.

1. When $V = \mathbb{F}^n$ and $\mathcal{B}$ is the standard basis (so that $\phi_{\mathcal{B}} = \mathrm{id}_{\mathbb{F}^n}$), this tells us that any bilinear form on V is B_A for some matrix $A \in M_{n\times n}(\mathbb{F})$.
2. There is a similar analysis for any bilinear map $B : U \times V \to W$. In this case, B is determined by $B(u_i, v_j) \in W$ for $u_1, \ldots, u_m$ a basis of U and $v_1, \ldots, v_n$ a basis of V.

How does A change when we change basis of V?

Proposition 6.2. *Let $B : V \times V \to \mathbb{F}$ be a bilinear form with matrices A and A' with respect to bases $\mathcal{B}$ and $\mathcal{B}'$ of V. Then*

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

where P is the change of basis matrix[1] *from $\mathcal{B}$ to $\mathcal{B}'$: thus $\phi_P = \phi_{\mathcal{B}}^{-1} \circ \phi_{\mathcal{B}'}$.*

Proof. Since $\phi_{\mathcal{B}'} = \phi_{\mathcal{B}} \circ \phi_P$, we have

$$\begin{aligned}\mathbf{x}^T A' \mathbf{y} = B(\phi_{\mathcal{B}'}(x), \phi_{\mathcal{B}'}(y)) &= B(\phi_{\mathcal{B}}(\phi_P(x)), \phi_{\mathcal{B}}(\phi_P(y))) \\ &= B_A(\phi_P(x), \phi_P(y)) = (P\mathbf{x})^T A (P\mathbf{y}) = \mathbf{x}^T (P^T A P) \mathbf{y},\end{aligned}$$

for all $x, y \in \mathbb{F}^n$. Taking $x = e_i$ and $y = e_j$, this gives $A'_{ij} = (P^T A P)_{ij}$ so that $A' = P^T A P$. □

This prompts:

[1] See Definition 5 in Section 2.5 of Algebra 1B.

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

$$B = P^T AP$$

Remark. The congruence relation should remind us of orthogonal diagonalisation but here there is no demand that $P^{-1} = P^T$.

6.2.2 Symmetric bilinear forms

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

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

Remark. If V is finite-dimensional, B is symmetric if and only if $B(v_i, v_j) = B(v_j, v_i)$, $1 \le i, j \le n$, for some basis $v_1, \ldots, v_n$ of V (you should think through the reverse implication). Thus B is symmetric if and only if its matrix with respect to some (and then any) basis is symmetric.

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 $\operatorname{rad} B \le V$.

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 is non-degenerate thanks to Lemma 4.1.

Here is how to understand both the rank and the radical of B. We use B to define a map $\beta : V \to V^*$ by

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

for $v, w \in V$. Then:

- $\beta(v) \in V^*$ since B is linear in the second slot.
- $\beta : V \to V^*$ is linear since B is linear in the first slot.
- $\ker \beta = \{v \in V \mid \beta(v) = 0\} = \{v \in V \mid B(v, w) = 0 \text{ for all } w \in V\} = \operatorname{rad} B$. Thus $\operatorname{rad} B \le V$ and $\operatorname{rank} B = \operatorname{rank} \beta$ when V is finite-dimensional.
  Moreover B is non-degenerate if and only if β injects or, when V is finite-dimensional, is an isomorphism.
- Let B have matrix A with respect to a basis $v_1, \ldots, v_n$ of V. Then

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

  where we used the symmetry of A in the last equality. It follows that

  $$\beta(v_j) = \sum_{i=1}^{n} A_{ij} v_i^*$$

  so that A is the matrix of β with respect to the dual bases $v_1, \ldots, v_n$ and $v_1^*, \ldots, v_n^*$ of V and V^*.

We learn from this how to compute the rank of B:

Lemma 6.3. *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 $\det A \neq 0$.

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

1. $B(x,y) = x_1y_1 + x_2y_2 - x_3y_3$. 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. $B(x,y) = x_1y_2 + x_2y_1$. 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, $B(x,y) = \sum_{i,j=1}^{3} A_{ij}x_iy_j$ so we can read off A from the coefficients of the x_iy_j.

6.2.3 Quadratic 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 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. $Q(x) = x_1x_2 = B_A(x,x)$ 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 6.4. *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

$$Q(x) = \sum_{1 \le i \le j \le n} q_{ij} x_i x_j = \mathbf{x}^T \begin{pmatrix} q_{11} & & \frac{1}{2}q_{ji} \\ & \ddots & \\ \frac{1}{2}q_{ij} & & q_{nn} \end{pmatrix} \mathbf{x}$$

so that the polarisation is B_A where

$$A_{ij} = A_{ji} = \begin{cases} q_{ii} & \text{if } i = j; \\ \frac{1}{2}q_{ij} & \text{if } i < j. \end{cases}$$

Otherwise said, the polarisation is given by

$$B_A(x, y) = \sum_{1 \le i \le j \le n} \tfrac{1}{2} q_{ij}(x_i y_j + x_j y_i).$$

Note that, when $i = j$, the summand on the right reduces to $q_{ii} x_i y_i$. Thus we get the formula for the polarisation $B(x, y)$ from that of $Q(x)$ by replacing $x_i x_j$ with $\frac{1}{2}(x_i y_j + x_j y_i)$ (of course, the latter reduces to the former when $x = y$ as it should!).

Example. Let $Q : \mathbb{R}^3 \to \mathbb{R}$ be given by

$$Q(x) = x_1^2 + 2x_2^2 + 2x_1 x_2 + x_1 x_3.$$

Let us find the polarisation B of Q. Two approaches:

1. 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
$$A = \begin{pmatrix} 1 & 1 & \frac{1}{2} \\ 1 & 2 & 0 \\ \frac{1}{2} & 0 & 0 \end{pmatrix}.$$
2. Compute $B(x, y)$ by replacing each $x_i x_j$ in the formula for Q with $\frac{1}{2}(x_i y_j + x_j y_i)$:
$$\begin{aligned} B(x, y) &= x_1 y_1 + 2x_2 y_2 + 2 \cdot \tfrac{1}{2}(x_1 y_2 + x_2 y_1) + \tfrac{1}{2}(x_1 y_3 + x_3 y_1) \\ &= x_1 y_1 + 2x_2 y_2 + x_1 y_2 + x_2 y_1 + \tfrac{1}{2} x_1 y_3 + \tfrac{1}{2} x_3 y_1. \end{aligned}$$

6.2.4 Classification of symmetric bilinear and quadratic forms

Convention. We retain our assumption that $1 + 1 \neq 0$ in $\mathbb{F}$.

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

Theorem 6.5 (Diagonalisation Theorem)**.** *Let B be a symmetric bilinear form on a finite-dimensional vector space over $\mathbb{F}$. Then there is a basis $v_1, \ldots, v_n$ of V with respect to which the matrix of B is diagonal:*

$$B(v_i, v_j) = 0,$$

for all $1 \le i \neq j \le n$. We call $v_1, \ldots, v_n$ a diagonalising basis *for B.*

Proof. This is reminiscent of the spectral theorem 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 Lemma 6.4, $B(v,w) = 0$, 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 $v_2, \ldots, v_n$ of W with $B(v_i, v_j) = 0$, for $2 \leq i \neq j \leq n$.

Now $v_1, \ldots, v_n$ 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. □

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.

Remark. For real symmetric bilinear forms B, we can a slightly different approach: let A be the matrix of B with respect to some basis and apply the spectral theorem (Theorem 4.17) to A to get $P \in \mathrm{O}(n)$ with $P^T A P$ diagonal. Then, by Proposition 6.2, P is the change of basis matrix from our given basis to a diagonalising one.

If V is a real inner product space and our original basis was orthonormal, then the diagonalising basis we find this way is also orthonormal so we have found a basis that simultaneously diagonalises both B and the inner product. In general, the question of whether two symmetric bilinear forms can be simultaneously diagonalised is a subtle one.

What does the diagonalisation theorem mean for a quadratic form Q? Observe:

- Any $\alpha \in V^*$ can be squared to give a quadratic form: $\alpha^2 : V \to \mathbb{F}$ given by $\alpha^2(v) = \alpha(v)^2$. Note that this is indeed a quadratic form with polarisation $B(v,w) = \alpha(v)\alpha(w)$.
- If $v_1, \ldots, v_n$ diagonalises the polarisation B of Q then $Q(\sum_i \lambda_i v_i) = \sum_i B(v_i, v_i)\lambda_i^2$ so that

$$Q = \sum_{i=1}^{n} B(v_i, v_i)(v_i^*)^2.$$

  That is, we have written Q as a linear combination of n squares.

In general, there are three methods to find a diagonalising basis:

1. Find v_1 with $B(v_1, v_1) \neq 0$, compute $W = \{v \in V \mid B(v_1, v) = 0\}$ and then iterate the procedure on W.

2. When $\mathbb{F} = \mathbb{R}$ and A is the matrix of B with respect to some basis, find an orthonormal basis of eigenvectors of A to get an orthogonal matrix P with $P^{-1}AP = P^TAP$ diagonal.
3. By inspection or by completing squares, we try to write Q as a linear combination of squares of *linearly independent* elements of V^*. Care is required to ensure linear independence here.

Example. Diagonalise $Q : \mathbb{R}^3 \to \mathbb{R}$ where

$$Q(x) = x_1^2 + 2x_1x_2 + 2x_2x_3 - x_3^2.$$

Thus we are being asked to find a basis v_1, v_2, v_3 for which $Q(\sum_i \lambda_i v_i) = a_1\lambda_1^2 + a_2\lambda_2^2 + a_3\lambda_3^2$, for some $a_1, a_2, a_3 \in \mathbb{R}$.

The polarisation B of Q is given by

$$B(x, y) = x_1y_1 + x_1y_2 + x_2y_1 + x_2y_3 + x_3y_2 - x_3y_3$$

with matrix A with respect to the standard basis where

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

There are now three ways to proceed:

1. Orthogonally diagonalise A: A has eigenvalues $\lambda_1, \lambda_2, \lambda_3 = 0, \sqrt{3}, -\sqrt{3}$ with eigenvectors

   $$v_1 = (-1, 1, 1), \quad v_2 = (1, -1 + \sqrt{3}, 2 - \sqrt{3}), \quad v_3 = (-1, -1 - \sqrt{3}, 2 + \sqrt{3}).$$

   We don't need to normalise these to unit length since scales of a diagonalising basis are also diagonalising but the corresponding diagonal matrix will have diagonal entries $\lambda_i\|v_i\|^2$.
   The downside of this approach is that we have to do a long computation and get a fairly messy answer.
2. We can exploit the zero in the $(1, 3)$-slot of A: observe that

   $$\begin{aligned} Q(e_1) &= 1 \\ Q(e_3) &= -1 \\ B(e_1, e_3) &= 0 \end{aligned}$$

   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

   $$B(y, y) = 1 - 2 + 2 - 1 = 0.$$

   This approach has the virtue of avoiding the computation of characteristic polynomials and finding eigenvectors.
3. Finally we can just attempt to write Q as a linear combination of squares by "eye-ball". In this case, we complete the square in the x_1, x_2 and x_2, x_3 variables:

   $$x_1^2 + 2x_1x_2 + 2x_2x_3 - x_3^2 = (x_1^2 + 2x_1x_2 + x_2^2) - (x_2^2 - 2x_2x_3 + x_3^2) = (x_1 + x_2)^2 - (x_2 - x_3)^2.$$

   Moreover $x_1 + x_2$ and $x_2 - x_3$ are linearly independent elements of $(\mathbb{R}^3)^*$ since they are not scalar multiples of each other.
   Strictly, in this case, we have only found two out of three dual basis vectors here but, in an exam situation, that would suffice.

Remarks.

1. The first two methods produce bases with respect to which B has rather different (diagonal) matrices: the first has $0, \pm 12(\sqrt{3} \mp 1)$ down the diagonal and the second has $1, -1, 0$. Observe however that these matrices have the same number of positive, negative and zero entries. This is no accident was we are about to see.
2. When A is diagonal, $\operatorname{rank} B = \operatorname{rank} A$ is the number of non-zero entries on the diagonal.

Definitions. Let Q be a quadratic form on a *real* vector space V.

Say that Q is *positive definite* if $Q(v) \geq 0$, for all $v \in V$, with equality if and only if $v = 0$.

Say that Q is *negative definite* if $-Q$ is positive definite.

If V is finite-dimensional, the *signature* of Q (or its polarisation B) is the pair (p, q) where

$$p = \max\{\dim U \mid U \leq V \text{ with } Q_{|U} \text{ positive definite}\}$$
$$q = \max\{\dim W \mid W \leq V \text{ with } Q_{|W} \text{ negative definite}\}.$$

Remark. A quadratic form $Q : V \to \mathbb{R}$ is positive definite on V if and only if its polarisation B is an inner product on V.

The signature is easy to compute:

Theorem 6.6 (Sylvester's Law of Inertia)**.** *Let Q be a quadratic form of signature (p, q) on a finite-dimensional real vector space and B its polarisation. Then:*

- $p + q = \operatorname{rank} B$*;*
- *any* diagonal *matrix representing B has p positive entries on the diagonal and q negative entries.*

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 Q is positive definite and W a q-dimensional subspace on which Q is negative definite.

First note that $U \cap K = \{0\}$ since $Q_{|K} = 0$. 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 $Q(v) = B(u + k, u + k) = B(u, u) \geq 0$.

From this we see that $W \cap (U + K) = \{0\}$: if $w \in W \cap (U + K)$ then $Q(w) \geq 0$ by what we just proved but also $Q(w) \leq 0$ since $w \in W$. Thus $Q(w) = 0$ 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 $v_1, \ldots, v_n$ be a diagonalising basis of B with $\hat{p}$ positive entries on the diagonal of the corresponding matrix representative of B and $\hat{q}$ negative entries. Then Q is positive definite on the $\hat{p}$-dimensional space $\operatorname{span}\{v_i \mid Q(v_i) > 0\}$. Thus $\hat{p} \leq p$. Similarly, $\hat{q} \leq q$.

However r is the number of non-zero entries on the diagonal, that is $r = \hat{p} + \hat{q}$. We therefore have

$$r = \hat{p} + \hat{q} \leq p + q = r$$

so that $p = \hat{p}$, $q = \hat{q}$ and $p + q = r$. □

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

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

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 conclude that the signature is $(2, 1)$.

Alternatively, if you can't be doing with finding roots of the characteristic polynomial, 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 again the signature is $(2, 1)$.

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,

$$\begin{aligned} Q(x) &= x_1^2 + x_2^2 + x_3^2 + 2x_1x_3 + 4x_2x_3 \\ &= (x_1 + x_3)^2 + x_2^2 + 4x_2x_3 = (x_1 + x_3)^2 + (x_2 + 2x_3)^2 - 4x_3^2. \end{aligned}$$

But now we need to check that x_1+x_3, x_2+2x_3, x_3 are linearly independent linear functionals on $\mathbb{R}^3$. Here Corollary 5.8 comes to the rescue and says we only need show that $(\ker x_1+x_2) \cap (\ker x_2+2x_3) \cap (\ker x_3) = \{0\}$. But $x_3 = 0 = x_1 + x_3 = x_2 + 2x_3$ rapidly implies that each $x_i = 0$ and we are done. The coefficients of these squares are $1, 1, -4$ and so, once more, we get that the signature is $(2, 1)$.

Let us pull all this together and summarise the situation for quadratic forms on vector spaces over our favourite fields:

Theorem 6.7. *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 $v_1, \ldots, v_n$ 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 $v_1, \ldots, v_n$ of V such that

$$Q(\sum_{i=1}^{n} x_i v_i) = x_1^2 + \cdots + x_p^2 - x_{p+1}^2 - \cdots - x_r^2.$$

Remarks.

1. A symmetric bilinear form of signature $(n, 0)$ on a real n-dimensional vector space is simply an inner product.
2. 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 $(3, 1)$.