Part A Linear Algebra

Alan Lauder*

December 3, 2018

1 Vector Spaces and Linear maps

We recall from Prelims some basic results on vector spaces.

1.1 Vector spaces

Before we can define vector spaces we need the notion of a field (discussed in Prelims Analysis I).

Definition 1.1. *A set $\mathbb{F}$ with two binary operations $+$ and $\times$ is a* **field** *if both $(\mathbb{F}, +, 0)$ and $(\mathbb{F} \setminus \{0\}, \times, 1)$ are abelian groups and the distribution law holds:*

$$(a + b)c = ac + bc, \text{ for all } a, b, c \in \mathbb{F}.$$

The smallest integer p such that

$$1 + 1 + \cdots + 1 \ (p \text{ times}) = 0$$

is called the **characteristic** *of $\mathbb{F}$. If no such p exists, the characteristic of $\mathbb{F}$ is defined to be zero.*

If such a p exists, it is necessarily prime. (Why?)

Example 1.2 The following are examples of fields ($\mathbb{F}_p$ and number fields like $\mathbb{Q}[i]$ are discussed in the exercises).

Characteristic 0 : $\mathbb{Q}$, $\mathbb{Q}[i]$, $\mathbb{R}$, $\mathbb{C}$.

Characteristic p : $\mathbb{F}_p = \{0, 1, \cdots, p-1\}$ with arithmetic modulo p.

*These notes are a revision of ones kindly provided by Ulrike Tillmann.

Definition 1.3. *A* **vector space** V *over a field* $\mathbb{F}$ *is an abelian group* $(V, +, 0)$ *together with a scalar multiplication* $\mathbb{F} \times V \to V$ *such that for all* $a, b \in \mathbb{F}, v, w \in V$ *:*

(1) $a(v + w) = av + aw$
(2) $(a + b)v = av + bv$
(3) $(ab)v = a(bv)$
(4) $1.v = v$

Let V be a vector space over $\mathbb{F}$.

Definition 1.4. *(1) A set* $S \subseteq V$ *is* **linearly independent** *if whenever* $a_1, \cdots, a_n \in \mathbb{F}$*, and* $s_1, \cdots, s_n \in S$*,*

$$a_1 s_1 + \cdots + a_n s_n = 0 \Rightarrow \ a_1 = \cdots = a_n = 0.$$

(2) A set $S \subseteq V$ *is* **spanning** *if for all* $v \in V$ *there exists* $a_1, \cdots, a_n \in \mathbb{F}$ *and* $s_1, \cdots, s_n \in S$ *with*

$$v = a_1 s_1 + \cdots + a_n s_n.$$

(3) A set $\mathcal{B} \subseteq V$ *is a* **basis** *of* V *if* $\mathcal{B}$ *is spanning and linearly independent. The size of* $\mathcal{B}$ *is the* **dimension** *of* V*.*

You saw in Prelims Linear Algebra I that every vector space with a finite spanning set has a basis and that the dimension of such vector spaces is well-defined.

Example 1.5

(1) $V = \mathbb{F}^n$ with standard basis $\{(1, 0, \cdots, 0), \cdots, (0, \cdots, 0, 1)\}$;

(2) $V = \mathbb{F}[x]$ with standard basis $\{1, x, x^2, \cdots\}$;

(3) Let

$$V = \mathbb{R}^{\mathbb{N}} = \{(a_1, a_2, a_3, \cdots) |\ a_i \in \mathbb{R}\},$$

Then $S = \{e_1, e_2, \cdots\}$ where $e_1 = (1, 0, 0, \cdots), \cdots$, is linearly independent but its span W is a proper subset of V. (What is an example of an element in V but not in W?)

1.2 Linear maps.

Next we consider linear maps and their relation to matrices.

Definition 1.6. *Suppose V and W are vector spaces over $\mathbb{F}$. A map $T : V \to W$ is a* **linear transformation** *(or just* **linear map***) if for all $a \in \mathbb{F}$, $v, v' \in V$,*

$$T(av + v') = aT(v) + T(v').$$

A bijective linear map is called an **isomorphism** *of vector spaces.*

Example 1.7

(1) The linear map $T : \mathbb{R}[x] \to \mathbb{R}[x]$ given by $f(x) \mapsto xf(x)$ is an injection; it defines an isomorphism from $\mathbb{R}[x]$ to its image $x\mathbb{R}[x]$

(2) The linear map $T : W \subseteq \mathbb{R}^{\mathbb{N}} \to \mathbb{R}[x]$ given by $e_n = (0, \cdots, 1, 0, \cdots) \mapsto x^{n-1}$ defines an isomorphism.

(3) Let $\mathrm{Hom}(V, W)$ be the set of linear maps from V to W. For $a \in \mathbb{F}, v \in V$, and $S, T \in \mathrm{Hom}(V, W)$ define:

$$(aT)(v) := a(T(v))$$
$$(T + S)(v) := T(v) + S(v)$$

With these definitions $\mathrm{Hom}(V, W)$ is a vector space over $\mathbb{F}$.

Now assume that V and W are finite dimensional.

Every linear map $T : V \to W$ is determined by its values on a basis $\mathcal{B}$ for V (as $\mathcal{B}$ is spanning). Vice versa, given any map $T : \mathcal{B} \to W$ it can be extended to a linear map $T : V \to W$ (as $\mathcal{B}$ is linearly independent).

Let $\mathcal{B} = \{e_1, \cdots, e_n\}$ and $\mathcal{B}' = \{e'_1, \cdots, e'_m\}$ be bases for V and W respectively. Let ${}_{\mathcal{B}'}[T]_{\mathcal{B}}$ be the matrix with (i, j)-entry a_{ij} such that:

$$T(e_j) = a_{1j}e'_1 + \cdots + a_{mj}e'_m.$$

(We call $\mathcal{B}$ the initial basis and $\mathcal{B}'$ the final basis.[1]) Note that ${}_{\mathcal{B}'}[aT]_{\mathcal{B}} = a({}_{\mathcal{B}'}[T]_{\mathcal{B}})$ and ${}_{\mathcal{B}'}[T + S]_{\mathcal{B}} = {}_{\mathcal{B}'}[T]_{\mathcal{B}} + {}_{\mathcal{B}'}[S]_{\mathcal{B}}$.

Furthermore, if $S \in \mathrm{Hom}(W, U)$ for some finite dimensional vector space U with basis $\mathcal{B}''$, then:

$${}_{\mathcal{B}''}[S \circ T]_{\mathcal{B}} = {}_{\mathcal{B}''}[S]_{\mathcal{B}'}\ {}_{\mathcal{B}'}[T]_{\mathcal{B}}$$

We can summarise the above in the following theorem.

[1] In Prelims Linear Algebra II, I used the notation $M^{\mathcal{B}}_{\mathcal{B}'}(T)$, but I think ${}_{\mathcal{B}'}[T]_{\mathcal{B}}$ is better as it helps you remember which is the initial basis and which the final one.

Theorem 1.8. *The assignment* $T \mapsto_{\mathcal{B}'} [T]_{\mathcal{B}}$ *is an isomorphism of vector spaces from* $\mathrm{Hom}(V, W)$ *to the space of* $(m \times n)$*-matrices over* $\mathbb{F}$*. It takes composition of maps to multiplication of matrices.*

In particular, if $T : V \to V$ and $\mathcal{B}$ and $\mathcal{B}'$ are two different bases with ${}_{\mathcal{B}}[Id]_{\mathcal{B}'}$ the change of basis matrix then:

${}_{\mathcal{B}'}[T]_{\mathcal{B}'} = {}_{\mathcal{B}'}[Id]_{\mathcal{B}}\, {}_{\mathcal{B}}[T]_{\mathcal{B}}\, {}_{\mathcal{B}}[Id]_{\mathcal{B}'}$ with ${}_{\mathcal{B}}[Id]_{\mathcal{B}'}\, {}_{\mathcal{B}'}[Id]_{\mathcal{B}} = {}_{\mathcal{B}}[Id]_{\mathcal{B}} = I$, the identity matrix.

2 Rings and polynomials

The study of vector spaces and linear maps between them naturally leads us to the study of rings; in particular, the ring of polynomials $\mathbb{F}[x]$ and the ring of $(n \times n)$-matrices $M_n(\mathbb{F})$.

2.1 Rings

Definition 2.1. *A non-empty set* R *with two binary operations* $+$ *and* $\times$ *is a* **ring** *if* $(R, +, 0)$ *is an abelian group, the multiplication* $\times$ *is associative and the distribution laws hold: for all* $a, b, c \in R$*,*

$$(a + b)c = ac + bc \;\text{ and }\; a(b + c) = ab + ac.$$

The ring R *is called* **commutative** *if for all* $a, b \in R$ *we have* $ab = ba$*.*

Example 2.2

(1) Any field is a commutative ring.

(2) $\mathbb{Z}$, $\mathbb{Z}[i]$, $\mathbb{F}[x]$ are commutative rings.

(3) $M_n(\mathbb{F})$, the set of $(n \times n)$-matrices over $\mathbb{F}$, and $\mathrm{Hom}(V, V)$, the set of linear self-maps for any vector space V, are non-commutative rings when $n > 1$ or $\dim(V) > 1$.

(4) For $A \in M_n(\mathbb{F})$ the set of polynomials evaluated on A, denoted $\mathbb{F}[A]$, forms a commutative subring of $M_n(\mathbb{F})$.

Definition 2.3. *A map* $\phi : R \to S$ *between two rings is a* **ring homomorphism** *if for all* $r, r' \in R$*:*

$$\phi(r + r') = \phi(r) + \phi(r') \;\text{ and }\; \phi(rr') = \phi(r)\phi(r').$$

A bijective ring homomorphism is called a ring isomorphism.

Example 2.4 When $W = V$ and $\mathcal{B}' = \mathcal{B}$, we can reinterpret Theorem 1.8 to say that $T \mapsto {}_{\mathcal{B}}[T]_{\mathcal{B}}$ defines an isomorphism of rings from $\mathrm{Hom}(V, V)$ to $M_n(\mathbb{F})$ where n is the dimension of V.

Definition 2.5. *A non-empty subset I of a ring R is an* **ideal** *if for all $s, t \in I$ and $r \in R$ we have*

$$s - t \in I \text{ and } sr, rs \in I.$$

Warning: Some books insist on rings having a multiplicative identity 1 and on ring homomorphisms taking 1 to 1. If we do not insist on rings having 1's, then any ideal is a subring. (Note that in a ring with an identity 1, any ideal that contains 1 is the whole ring.)

Example 2.6

(1) $m\mathbb{Z}$ is an ideal in $\mathbb{Z}$. Indeed, every ideal in $\mathbb{Z}$ is of this form. [To prove this, let m be the smallest non-zero integer in the ideal I and prove that $I = m\mathbb{Z}$.]

(2) The set of diagonal matrices in $M_n(\mathbb{R})$ is closed under addition and multiplication (i.e. it is a subring) but for $n > 1$ is **not** an ideal.

Ideals are to rings what normal subgroups are to groups in the sense that the set of additive cosets R/I inherit a ring structure from R if I is an ideal. For $r, r' \in R$ define

$$(r + I) + (r' + I) := (r + r') + I \text{ and } (r + I)(r' + I) := rr' + I.$$

We leave it as an exercise to check that these operations are well-defined.

Theorem 2.7 (First Isomorphism Theorem)**.** *The kernel* $\mathrm{Ker}(\phi) := \phi^{-1}(0)$ *of a ring homomorphism $\phi : R \to S$ is an ideal, its image* $\mathrm{Im}(\phi)$ *is a subring of S, and ϕ induces an isomorphisms of rings*

$$R/\mathrm{Ker}(\phi) \cong \mathrm{Im}(\phi).$$

Proof. Exercise. [Show that the underlying isomorphisms of abelian groups is compatible with the multiplication, i.e. is a ring homomorphism.] □

2.2 Polynomial rings

We will discuss polynomials over a field $\mathbb{F}$ in more detail. The first goal is to show that there is a division algorithm for polynomial rings. With the help of this we will able to show the important property that every ideal in $\mathbb{F}[x]$ is generated by one element.

Theorem 2.8. *["Division algorithm" for polynomials] Let $f(x),\ g(x) \in \mathbb{F}[x]$ be two polynomials with $g(x) \neq 0$. Then there exists $q(x), r(x) \in \mathbb{F}[x]$ such that*

$$f(x) = q(x)g(x) + r(x) \text{ and } \deg r(x) < \deg g(x).$$

Proof. If $\deg f(x) < \deg g(x)$, put $q(x) = 0,\ r(x) = f(x)$. Assume now that $\deg f(x) \geq \deg g(x)$ and let

$$f(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_0$$
$$g(x) = b_k x^k + b_{k-1}x^{k-1} + \cdots + b_0,\ (b_k \neq 0)$$

Then

$$\deg\left(f(x) - \frac{a_n}{b_k}x^{n-k}g(x)\right) < n$$

By induction on $\deg f - \deg g$, there exist $s(x),\ t(x)$ such that

$$f(x) - \frac{a_n}{b_k}x^{n-k}g(x) = s(x)g(x) + t(x) \text{ and } \deg g(x) > \deg t(x).$$

Hence put $q(x) = \frac{a_n}{b_k}x^{n-k} + s(x)$ and $r(x) = t(x)$. □

Corollary 2.9. *For all $f(x) \in \mathbb{F}[x]$ and $a \in \mathbb{F}$,*

$$f(a) = 0 \Rightarrow (x - a)|f(x).$$

Proof. By Theorem 2.8 there exist $q(x),\ r(x)$ such that

$$f(x) = q(x)(x - a) + r(x)$$

where $r(x)$ is constant (as $\deg r(x) < 1$). Evaluating at a gives

$$f(a) = 0 = q(a)(a - a) + r = r$$

and hence $r = 0$. □

Corollary 2.10. *Assume $f \neq 0$. If $\deg f \leq n$ then f has at most n roots.*

Proof. This follows from Corollary 2.9 and induction. □

Let $a(x), b(x) \in \mathbb{F}[x]$ be two polynomials. Let $c(x)$ be a monic polynomial of highest degree dividing both $a(x)$ and $b(x)$ and write $c = \gcd(a, b)$ (also wrote less commonly $\mathrm{hcf}(a, b)$).

Proposition 2.11. *Let $a, b \in \mathbb{F}[x]$ be non-zero polynomials and let $\gcd(a, b) = c$. Then there exist $s, t \in \mathbb{F}[x]$ such that:*

$$a(x)s(x) + b(x)t(x) = c(x).$$

Proof. If $c \neq 1$, divide a and b by c. We may thus assume $\deg(a) \geq \deg(b)$ and $\gcd(a,b) = 1$, and will proceed by induction on $\deg(a) + \deg(b)$.

By the Division Algorithm there exist $q, r \in \mathbb{F}[x]$ such that

$$a = qb + r \text{ with } \deg(b) > \deg(r).$$

Then $\deg(a) + \deg(b) > \deg(b) + \deg(r)$ and $\gcd(b, r) = 1$.

If $r = 0$ then $b(x) = \lambda$ is constant since $\gcd(a, b) = 1$. Hence

$$a(x) + b(x)(1/\lambda)(1 - a(x)) = 1.$$

Assume $r \neq 0$. Then by the induction hypothesis, there exist $s', t' \in \mathbb{F}[x]$ such that

$$bs' + rt' = 1.$$

Hence,

$$bs' + (a - qb)t' = 1 \text{ and } at' + b(s' - qt') = 1$$

So, we may put $t = t'$ and $s = s' - qt'$. □

Exercise: Prove that every ideal $I \subseteq \mathbb{F}[x]$ is generated by one element. In other words, given an ideal I there exists a polynomial $c(x)$ such that

$$I = \{f(x)c(x) \mid f(x) \in \mathbb{F}[x]\}.$$

2.3 Evaluating polynomials on matrices.

We will now turn to evaluating polynomials at $(n \times n)$-matrices. Given a matrix we can associate two special polynomials to it, its characteristic and its minimal polynomial. As we will see these will encode much of the information of interest.

Let $A \in M_n(\mathbb{F})$ and $f(x) = a_k x^k + \cdots + a_0 \in \mathbb{F}[x]$. Then

$$f(A) := a_k A^k + \cdots + a_0 I \ \in M_n(\mathbb{F}).$$

Since $A^p A^q = A^q A^p$ and $\lambda A = A\lambda$ for $p, q \geq 0$ and $\lambda \in \mathbb{F}$, then for all $f(x),\ g(x) \in \mathbb{F}[x]$ we have that

$$f(A)g(A) = g(A)f(A);$$
$$Av = \lambda v \ \Rightarrow \ f(A)v = f(\lambda)v.$$

Lemma 2.12. *For all $A \in M_n(\mathbb{F})$, there exists a non-zero polynomial $f(x) \in \mathbb{F}[x]$ such that $f(A) = 0$.*

Proof. Note that the dimension $\dim M_n(\mathbb{F}) = n \times n$ is finite. Hence $\{I, A, A^2, \cdots, A^k\}$ as a subset of $M_n(\mathbb{F})$ is linearly dependent for $k \geq n^2$. So there exist scalars $a_i \in \mathbb{F}$ such that

$$a_k A^k + \cdots + a_0 I = 0,$$

and $f(x) = a_k x^k + \cdots + a_0$ is an annihilating polynomial. □

We can express much of the above in terms of ring theory as follows. For any $(n \times n)$-matrix A, the assignment $f(x) \mapsto f(A)$ defines a ring homomorphism

$$E_A : \mathbb{F}[x] \to M_n(\mathbb{F}).$$

Lemma 2.12 tells us the kernel is non-zero, and moreover as $\mathbb{F}[x]$ is commutative so is the image of E_A; that is, $f(A)g(A) = g(A)f(A)$ for all polynomials f and g.

Our next step is to determine the unique monic polynomial generating the kernel of E_A.

2.4 Minimal and characteristic polynomials

Definition 2.13. *The* **minimal polynomial** *of A, denoted by $m_A(x)$, is the monic polynomial $p(x)$ of least degree such that $p(A) = 0$.*

Theorem 2.14. *If $f(A) = 0$ then $m_A | f$. Furthermore, m_A is unique (hence showing that m_A is well-defined).*

Proof. By the division algorithm, Theorem 2.8, there exist polynomials $q,\ r$ with $\deg r < \deg m_A$ such that

$$f = q m_A + r.$$

Evaluating both sides at A gives $r(A) = 0$. By the minimality property of m_A,

$$r = 0$$

and m_A divides f. To show uniqueness, let m be another monic polynomial of minimal degree and $m(A) = 0$. Then by the above $m_A | m$. Also m and m_A must have the same degree, and so $m = a m_A$ for some $a \in \mathbb{F}$. Since both polynomials are monic it follows that $a = 1$ and $m = m_A$. □

Definition 2.15. *The* **characteristic polynomial** *of A is defined as*

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

Lemma 2.16. $\chi_A(x) = (-1)^n x^n + (-1)^{n-1} \operatorname{tr}(A) x^{n-1} + \cdots + \det A.$

Proof. This is proved as Proposition 9 in Prelims Linear Algebra II. □

Definition 2.17. *Recall λ is an* **eigenvalue** *of A if there exists a non-zero $v \in \mathbb{F}^n$ such that $Av = \lambda v$, and we call v the* **eigenvector**.

Theorem 2.18.

$$\begin{aligned} &\lambda \text{ is an eigenvalue of } A \\ &\qquad \Leftrightarrow \lambda \text{ is a root of } \chi_A(x) \\ &\qquad \Leftrightarrow \lambda \text{ is a root of } m_A(x) \end{aligned}$$

Proof.

$$\begin{aligned} \chi_A(\lambda) = 0 &\Leftrightarrow \det(A - \lambda I) = 0 \\ &\Leftrightarrow A - \lambda I \text{ is singular} \\ &\Leftrightarrow \exists\ v \neq 0 : (A - \lambda I)v = 0 \\ &\Leftrightarrow \exists\ v \neq 0 : Av = \lambda I \\ &\Rightarrow m_A(\lambda)v = m_A(A)v = 0 \\ &\Rightarrow m_A(\lambda) = 0 \text{ (as } v \neq 0) \end{aligned}$$

Conversely, assume λ is a root of m_A. Then $m_A(x) = g(x)(x - \lambda)$ for some polynomial g. By minimality of m_A, we have $g(A) \neq 0$. Hence there exists $w \in \mathbb{F}^n$ such that $g(A)w \neq 0$. Put $v = g(A)w$ then

$$(A - \lambda I)v = m_A(A)w = 0,$$

and v is a λ-eigenvector for A. □

One of our next goals is to prove that χ_A annihilates A and hence that $m_A | \chi_A$.

We finish this section by recording how to translate back what we have learnt about matrices to the world of linear maps. In particular we will show that it makes sense to speak of a minimal and characteristic polynomial of a linear transformation.

Let C, P, A be $(n \times n)$-matrices such that $C = P^{-1}AP$. Then $m_C(x) = m_A(x)$ for:

$$f(C) = f(P^{-1}AP) = P^{-1}f(A)P$$

for all polynomials f. Thus

$$m_C(A) = m_C(C) = 0 \text{ and } m_A(C) = m_A(A) = 0$$

so that $m_C | m_A$ and $m_A | m_C$ and therefore $m_A = m_C$ as both are monic.

Definition 2.19. *Let V be a finite dimensional vector space and $T : V \to V$ a linear transformation. Define the* **minimal polynomial** *of T as*

$$m_T(x) = m_A(x)$$

where $A = {}_{\mathcal{B}}[T]_{\mathcal{B}}$ with respect to some basis $\mathcal{B}$ of V. As $m_A(x) = m_{P^{-1}AP}(x)$ the definition of $m_T(x)$ is independent of the choice of basis.

Definition 2.20. *For a linear transformation $T : V \to V$ define its* **characteristic polynomial** *as*

$$\chi_T(x) = \chi_A(x)$$

where $A = {}_{\mathcal{B}}[T]_{\mathcal{B}}$ with respect to some basis $\mathcal{B}$ of V. As $\chi_A(x) = \chi_{P^{-1}AP}(x)$ the definition of $\chi_T(x)$ is independent of the choice of basis.

Appendix: Algebraically closed fields.

It will be convenient and illuminating to be able to refer to algebraically closed fields.

Definition 2.21. *A field $\mathbb{F}$ is* **algebraically closed** *if every non-constant polynomial in $\mathbb{F}[x]$ has a root in $\mathbb{F}$.*

Theorem 2.22 (Fundamental Theorem of Algebra)**.** *The field of complex numbers $\mathbb{C}$ is algebraically closed.*

We will not be able to show this in this course. However, you should be able to prove it using complex analysis by the end of this term. (Consider $g(z) = 1/f(z)$. If $f(z)$ has no roots in $\mathbb{C}$, it is holomorphic and bounded on $\mathbb{C}$, which leads to a contradiction.)

Definition 2.23. *An algebraically closed field $\bar{\mathbb{F}}$ containing $\mathbb{F}$ with the property that there does not exist a smaller algebraically closed field L with*

$$\bar{\mathbb{F}} \supseteq L \supseteq \mathbb{F}$$

is called an algebraic closure of $\mathbb{F}$.

Theorem 2.24. *Every field $\mathbb{F}$ has an algebraic closure $\bar{\mathbb{F}}$.*

The proof is beyond this course but it will be convenient to appeal to this result.

Challenge: Prove that no finite field is algebraically closed. [Hint: imitate the standard proof that there are infinitely many primes.]

3 Quotient Spaces

In the theory of groups and rings the notion of a quotient is an important and natural concept. Recall that the image of a group or ring homomorphism is best understood as a quotient of the source by the kernel of the homomorphism. Similarly, for vector spaces it is natural to consider quotient spaces.

Let V be a vector space over a field $\mathbb{F}$ and let U be a subspace.

Lemma 3.1. *The set of cosets*

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

with the operations

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

for $v, w \in V$ and $a \in \mathbb{F}$ is a vector space, called the **quotient space**.

Proof. We need to check that the operations are well-defined. Assume $v + U = v' + U$ and $w + U = w' + U$. Then $v = v' + u, w = w' + \tilde{u}$ for $u, \tilde{u} \in U$. Hence:

$$\begin{aligned}(v + U) + (w + U) &= v + w + U \\ &= v' + u + w' + \tilde{u} + U \qquad \text{as } u + \tilde{u} \in U \\ &= v' + w' + U \\ &= (v' + U) + (w' + U).\end{aligned}$$

Similarly,

$$\begin{aligned}a(v + U) &= av + U \\ &= av' + au + U \qquad \text{as } au \in U \\ &= av' + U \\ &= a(v' + U)\end{aligned}$$

That these operations satisfy the vector space axioms follows immediately from the fact that the operations in V satisfy them. □

Often in the literature the quotient construction is avoided in the context of vector spaces. This is because any quotient V/U of a vector space V by a subspace U can be "realised" itself as a subspace of V itself [2]. That is, by extending a basis for U to one for V, we can choose a subspace W such that $V = U \oplus W$. Then each $v \in V$ can be written uniquely as $u + w$ for some $u \in U$ and $w \in W$, and this allows us to define an isomorphism $V/U \to W$ by $v + U \mapsto w$. However, such an isomorphism involves a choice of W and it is often easier to avoid having to make this choice (and thus avoid showing that further constructions and results are independent of it).

Let $\mathcal{E}$ be a basis of U, and extend $\mathcal{E}$ to a basis $\mathcal{B}$ of V (we assume this is possible, which we certainly know to be the case at least for V finite dimensional).

Define

$$\overline{\mathcal{B}} := \{e + U \mid e \in \mathcal{B} \backslash \mathcal{E}\} \subseteq V/U.$$

Here $\mathcal{B}\backslash\mathcal{E}$ just means the elements in $\mathcal{B}$ which are not in $\mathcal{E}$.

[2] This is in contrast to the world of groups and rings. For example $\mathbb{Z}/2\mathbb{Z}$ is a quotient group (and ring) of $\mathbb{Z}$ but cannot be realised as a subgroup of $\mathbb{Z}$.

Proposition 3.2. *The set $\overline{\mathcal{B}}$ is a basis for V/U.*

Proof. Let $v+U \in V/U$. Then there exists $e_1, \cdots, e_k \in \mathcal{E}$, $e_{k+1}, \cdots, e_n \in \mathcal{B}\backslash\mathcal{E}$ and $a_1, \cdots, a_n \in \mathbb{F}$ such that

$$v = a_1e_1 + \cdots + a_ke_k + a_{k+1}e_{k+1} + \cdots + a_ne_n,$$

as $\mathcal{B}$ is spanning. Hence

$$\begin{aligned} v + U &= a_{k+1}e_{k+1} + \cdots + a_ne_n + U \\ &= a_{k+1}(e_{k+1} + U) + \cdots + a_n(e_n + U), \end{aligned}$$

and hence $\overline{\mathcal{B}}$ is spanning.

To show independence, assume for some $a_1, \cdots, a_r \in \mathbb{F}$ and $e_1, \cdots, e_r \in \mathcal{B}\backslash\mathcal{E}$, that

$$a_1(e_1 + U) + \cdots + a_r(e_r + U) = U.$$

Then $a_1e_1 + \cdots + a_re_r \in U$ and hence

$$a_1e_1 + \cdots + a_re_r = b_1e'_1 + \cdots + b_se'_s$$

for some $e'_1, \cdots, e'_s \in \mathcal{E}$ and $b_1, \cdots, b_s \in \mathbb{F}$ as $\mathcal{E}$ spans U. But then $a_1 = \cdots = a_r = -b_s = \cdots = -b_s = 0$ as $\mathcal{B}$ is linearly independent, and thus $\bar{\mathcal{B}}$ is linearly independent. □

By a similar argument we get the "converse" statement.

Proposition 3.3. *Let $U \subset V$ be vector spaces, with $\mathcal{E}$ a basis for U, and $\mathcal{F} \subset V$ a set of vectors such that*

$$\{v + U : v \in \mathcal{F}\}$$

is a basis for the quotient V/U. Then the union

$$\mathcal{E} \cup \mathcal{F}$$

is a basis for V.

Proof. Exercise. □

Example 3.4

$$\begin{aligned} V &= \mathbb{F}[x] & \mathcal{B} &= \{1, x, x^2, \cdots\} \\ U &= \text{even polynomials} & \mathcal{E} &= \{1, x^2, x^4, \cdots\} \\ V/U &\simeq \text{odd polynomials} & \overline{\mathcal{B}} &= \{x + U, x^3 + U, \cdots\} \end{aligned}$$

Corollary 3.5. *If V is finite dimensional then*

$$\dim(V) = \dim(U) + \dim(V/U).$$

Theorem 3.6 (First Isomorphism Theorem)**.** *Let $T : V \to W$ be a linear map of vector spaces over $\mathbb{F}$. Then*

$$\begin{aligned} \overline{T} : V/\mathrm{Ker}(T) &\to \mathrm{Im}(T) \\ v + \mathrm{Ker}(T) &\mapsto T(v) \end{aligned}$$

is an isomorphism of vector spaces.

Proof. It follows from the first isomorphism theorem for groups that $\overline{T}$ is an isomorphism of (abelian) groups. $\overline{T}$ is also compatible with scalar multiplication. Thus $\overline{T}$ is a linear isomorphism. □

Detailed working:

$$\begin{aligned} \underline{\text{Well-Defined:}}\ & v + \mathrm{Ker}(T) = v' + \mathrm{Ker}(T) \\ & \Rightarrow v = v' + u, \qquad \text{for some } u \in \mathrm{Ker}(T) \\ & \Rightarrow \overline{T}(v + \mathrm{Ker}(T)) := T(v) \\ & = T(v' + u) = T(v') =: \overline{T}(v' + \mathrm{Ker}(T)) \end{aligned}$$

$$\begin{aligned} \underline{\text{Linear:}}\ & \overline{T}(a(v + \mathrm{Ker}(T)) + (v' + \mathrm{Ker}(T))) \\ & = \overline{T}(av + v' + \mathrm{Ker}(T)) \\ & := T(av + v') \\ & = aT(v) + T(v') \\ & =: a\overline{T}(v + \mathrm{Ker}(T)) + \overline{T}(v' + \mathrm{Ker}(T)) \end{aligned}$$

$$\begin{aligned} \underline{\text{Surjective:}}\ & w \in \mathrm{Im}(T) \\ & \Rightarrow \exists v \in V : T(v) = w \\ & \Rightarrow \overline{T}(v + \mathrm{Ker}(T)) = T(v) = w \\ & \Rightarrow w \in \mathrm{Im}(\overline{T}) \end{aligned}$$

$$\begin{aligned} \underline{\text{Injective:}}\ & v + \mathrm{Ker}(T) \in \mathrm{Ker}(\overline{T}) \\ & \Rightarrow \overline{T}(v + \mathrm{Ker}(T)) = T(v) = 0 \\ & \Rightarrow v \ \in \mathrm{Ker}(T) \\ & \Rightarrow v + \mathrm{Ker}(T) = 0 + \mathrm{Ker}(T). \end{aligned}$$

□

Theorem 3.7 (Rank-Nullity Theorem)**.** *If $T : V \to W$ is a linear transformation and V is finite dimensional, then*

$$\dim(V) = \dim(\mathrm{Ker}(T)) + \dim(\mathrm{Im}(T)).$$

Proof. We apply Corollary 3.5 to $U = \ker(T)$. Then

$$\dim(V) = \dim(\mathrm{Ker}(T)) + \dim(V/\mathrm{Ker}(T)).$$

By the First Isomorphism Theorem also:

$$\dim(V/\mathrm{Ker}(T)) = \dim(\mathrm{Im}(T)).$$

□

Let $T : V \to W$ be a linear map and let $A \subseteq V, B \subseteq W$ be subspaces.

Lemma 3.8. *The formula $\overline{T}(v + A) := T(v) + B$ gives a well-defined linear map of quotients $\overline{T} : V/A \to W/B$ if and only if $T(A) \subseteq B$.*

Proof. Assume $T(A) \subseteq B$. Now $\overline{T}$ will be linear if it is well-defined. Assume $v + A = v' + A$. Then $v = v' + a$ for some $a \in A$. So

$$\begin{aligned}
\overline{T}(v + A) &= T(v) + B && \text{by definition} \\
&= T(v' + a) + B \\
&= T(v') + T(a) + B && \text{as } T \text{ is linear} \\
&= T(v') + B && \text{as } T(A) \subseteq B \\
&= \overline{T}(v' + B).
\end{aligned}$$

Hence $\bar{T}$ is well-defined. Conversely, assume that $\overline{T}$ is well-defined and let $a \in A$. Then

$$\begin{aligned}
B = 0 + B = \overline{T}(A) &= \overline{T}(a + A) \\
&= T(a) + B.
\end{aligned}$$

Thus $T(a) \in B$, and so $T(A) \subseteq B$. □

Assume now that V and W are finite dimensional. Let $\mathcal{B} = \{e_1, \cdots, e_n\}$ be a basis for V with $\mathcal{E} = \{e_1, \cdots, e_k\}$ a basis for a subspace $A \subseteq V$ (so $k \leq n$). Let $\mathcal{B}' = \{e'_1, \cdots, e'_m\}$ be a basis for W with $\mathcal{E}' = \{e'_1, \cdots, e'_\ell\}$ a basis for a subspace $B \subseteq W$. The induced bases for V/A and W/B are given by

$$\begin{aligned}
\overline{\mathcal{B}} &= e_{k+1} + A, \cdots, e_n + A \quad \text{and} \\
\overline{\mathcal{B}'} &= e'_{\ell+1} + B, \cdots, e'_m + B.
\end{aligned}$$

Let $T : V \to W$ be a linear map such that $T(A) \subseteq B$. Then T induces a map $\bar{T}$ on quotients by Lemma 3.7 and restricts to a linear map

$$T|_A : A \to B \text{ with } T|_A(v) = T(v) \text{ for } v \in A.$$

Recall the notation ${}_{\mathcal{B}'}[T]_{\mathcal{B}} = (a_{ij})_{1 \leq i \leq m,\ 1 \leq j \leq n}$ where

$$T(e_j) = a_{1j}e'_1 + \cdots + a_{mj}e'_m.$$

Theorem 3.9. *There is a block matrix decomposition*

$$ {}_{\mathcal{B}'}[T]_{\mathcal{B}} = \left[\begin{array}{c|c} {}_{\mathcal{E}'}[T|_A]_{\mathcal{E}} & * \\ \hline 0 & {}_{\overline{\mathcal{B}}'}[\overline{T}]_{\overline{\mathcal{B}}} \end{array}\right], $$

where ${}_{\overline{\mathcal{B}}'}[\overline{T}]_{\overline{\mathcal{B}}} = (a_{ij})_{l+1\leq i\leq m,\ k+1\leq j\leq n}$.

Proof. For $j \leq k$, $T(e_j) \in B$ and hence $a_{ij} = 0$ for $i > \ell$ and a_{ij} is equal to the (i,j)-entry of ${}_{\mathcal{E}'}[T|_A]_{\mathcal{E}}$ for $i \leq \ell$. To identify the bottom right corner of the matrix, note that

$$\begin{aligned} \overline{T}(e_j + A) =& T(e_j) + B \\ =& a_{1j}e'_1 + \cdots + a_{mj}e'_m + B \\ =& a_{\ell+1,j}(e'_{\ell+1} + B) + \cdots + a_{mj}(e'_m + B). \end{aligned}$$

□

4 Triangular Form and the Cayley-Hamilton Theorem

The goal of this chapter is to prove that the characteristic polynomial of an $(n \times n)$-matrix is annihilating; that is, the polynomial vanishes when evaluated at the matrix. This will also give us control on the minimal polynomial.

Let $T : V \to V$ be a linear transformation.

Definition 4.1. *A subspace $U \subseteq V$ is called* T**-invariant** *if $T(U) \subseteq U$.*

By the result of the previous section, such a T induces a map $\overline{T} : V/U \to V/U$. Let $S : V \to V$ be another linear map.

If U is T- and S-invariant, then U is also invariant under the following maps:

1. the zero map
2. the identity map
3. $aT, \quad \forall a \in \mathbb{F}$
4. $S + T$
5. $S \circ T$

Hence, U is invariant under any polynomial $p(x)$ evaluated at T. Furthermore, $p(T)$ induces a map of quotients

$$\overline{p(T)} : V/U \to V/U.$$

Example 4.2 Let $V_\lambda := \ker(T - \lambda I)$ be the λ-eigenspace of T. Then V_λ is T-invariant. Let $W := \ker(g(T))$ be the kernel of $g(T)$ for some $g(x) \in \mathbb{F}[x]$. Then W is T-invariant as $g(T)T = Tg(T)$.

Proposition 4.3. *Let $T : V \to V$ be a linear transformation and assume $U \subseteq V$ is T-invariant. Then*

$$\chi_T(x) = \chi_{T|_U}(x) \ \times \ \chi_{\overline{T}}(x).$$

Proof. Extend a basis $\mathcal{E}$ for U to a basis $\mathcal{B}$ of V. Let $\overline{\mathcal{B}}$ be the associated basis for V/U. By Theorem 3.9

$${}_{\mathcal{B}}[T]_{\mathcal{B}} = \left(\begin{array}{c|c} {}_{\mathcal{E}}[T|_u]_{\mathcal{E}} & * \\ \hline 0 & {}_{\overline{\mathcal{B}}}[\overline{T}]_{\overline{\mathcal{B}}} \end{array}\right).$$

The determinant of such an upper triangular block matrix is the product of the determinants of the diagonal blocks. □

Note that this formula does not hold for the minimal polynomial (the identity map yielding a counterexample in dimension ≥ 2).

Definition 4.4. *$A = (a_{ij})$ an $n \times n$ matrix is* **upper triangular** *if $a_{ij} = 0$ for all $i > j$.*

Theorem 4.5. *Let V be a finite-dimensional vector space, and let $T : V \to V$ be a linear map such that its characteristic polynomial is a product of linear factors. Then, there exists a basis $\mathcal{B}$ of V such that ${}_{\mathcal{B}}[T]_{\mathcal{B}}$ is upper triangular.*

Note: If $\mathbb{F}$ is an algebraically closed field, such as $\mathbb{C}$, then the characteristic polynomial always satisfies the hypothesis.

Proof. By induction on the dimension of V. Note when V is one dimensional, there is nothing more to prove. In general, by assumption χ_T has a root λ and hence there exists a $v_1 \neq 0$ such that $Tv_1 = \lambda v_1$. Put $U = \langle v_1 \rangle$, the line spanned by v_1. As v_1 is an eigenvector, U is T-invariant. Thus we may consider the induced map on quotients

$$\overline{T} := V/U \to V/U.$$

By Proposition 4.3,

$$\chi_{\overline{T}}(x) = \chi_T(x)/(x - \lambda)$$

and hence is also a product of linear factors and furthermore $\dim V/U = \dim(V) - 1$. Hence, by the induction hypothesis, there exists $\overline{\mathcal{B}} = \{v_2 + U, \ldots, v_n + U\}$ such that ${}_{\overline{\mathcal{B}}}[\overline{T}]_{\overline{\mathcal{B}}}$ is upper triangular. Put $\mathcal{B} = \{v_1, v_2, \ldots, v_n\}$. Then $\mathcal{B}$ is a basis for V, by Proposition 3.3, and

$${}_{\mathcal{B}}[T]_{\mathcal{B}} = \left(\begin{array}{c|c} \lambda & * \\ \hline 0 & {}_{\overline{\mathcal{B}}}[\overline{T}]_{\overline{\mathcal{B}}} \end{array}\right).$$

is upper triangular. □

Corollary 4.6. *If A is an $n \times n$ matrix with a characteristic polynomial that is a product of linear factors, then there exists an $(n \times n)$-matrix P such that $P^{-1}AP$ is upper triangular.*

Proposition 4.7. *Let A be an upper triangular $(n \times n)$-matrix with diagonal entries $\lambda_1, \ldots, \lambda_n$. Then*

$$\prod_{i=1}^{n}(A - \lambda_i I) = 0$$

Proof. Let $e_1, \ldots, e_n$ be the standard basis vectors for $\mathbb{F}^n$. Then

$$(A - \lambda_n I)v \in \langle e_1, \ldots, e_{n-1}\rangle \quad \text{for all } v \in \mathbb{F}^n$$

and more generally

$$(A - \lambda_i I)w \in \langle e_1, \ldots, e_{i-1}\rangle \quad \text{for all } w \in \langle e_1, \ldots, e_i\rangle.$$

Hence, since

$$\mathrm{Im}(A - \lambda_n I) \subseteq \langle e_1, \ldots, e_{n-1}\rangle$$
$$\mathrm{Im}(A - \lambda_{n-1} I)(A - \lambda_n I) \subseteq \langle e_1, \ldots, e_{n-2}\rangle$$

and so on, we have that

$$\prod_{i=1}^{n}(A - \lambda_i I) = 0$$

as required.

□

Theorem 4.8 (Cayley-Hamilton)**.** *If $T : V \to V$ is a linear transformation and V is a finite dimensional vector space, then $\chi_T(T) = 0$. Hence, in particular, $m_T(x) \mid \chi_T(x)$.*

Proof. Let A be the matrix of T with respect to some basis for V. We will work over the algebraic closure $\overline{\mathbb{F}} \supseteq \mathbb{F}$. [3] In $\overline{\mathbb{F}}[x]$, every polynomial factors into linear terms. Thus, by Corollary 4.6, there exists a matrix $P \in M_n(\overline{\mathbb{F}})$ such that $P^{-1}AP$ is upper triangular with diagonal entries $\lambda_1, \ldots, \lambda_n$. Thus,

$$\chi_{P^{-1}AP}(x) = (-1)^{\dim(V)} \prod_{k=1}^{n}(x - \lambda_k)$$

and by Proposition 4.7, we have $\chi_{P^{-1}AP}(P^{-1}AP) = 0$. As

$$\chi_T(x) := \chi_{P^{-1}AP}(x)$$

we have that also $\chi_T(T) = 0$. The minimal polynomial divides annihilating polynomials by Theorem 2.14, and so $m_T(x) \mid \chi_T(x)$. □

[3] It would be enough to work over the finite extension of fields $L = \mathbb{F}[\lambda_1, \ldots, \lambda_n]$.

What is wrong with the following "proof" of the Cayley-Hamilton theorem? "$\chi_A(x) := \det(A - xI)$ and hence $\chi_A(A) = \det(A - A \cdot I) = \det(0) = 0$". (This is *not* a proof; come to the lectures to find out why.)

Example 4.9

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

$$\chi_A(x) = \det \begin{pmatrix} 1-x & 2 \\ -1 & -x \end{pmatrix} = x^2 - x + 2,$$

$$\chi_A(A) = A^2 - A + 2I = \begin{pmatrix} -1 & 2 \\ -1 & -2 \end{pmatrix} - \begin{pmatrix} 1 & 2 \\ -1 & 0 \end{pmatrix} + \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} = 0.$$

As $A - \lambda I \neq 0$ for any choice of λ, the minimal polynomial cannot be of degree one. As $m_A \mid \chi_A$ we must have $m_A = \chi_A$. (Alternatively, since $\chi_A(x)$ has non-zero discriminant, it has two distinct roots — since the minimal and characteristic polynomials have the same roots, not counting multipliicity, and $m_A | \chi_A$ we must here have $m_A = \chi_A$.)

Example 4.10

$$A = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \end{pmatrix}, \quad \chi_A(x) = (1-x)^2(2-x)^2.$$

Possible minimal polynomials:

$(x-1)(x-2)$ — No: not annihilating — $(A - I)(A - 2I) \neq 0$
$(x-1)(x-2)^2$ — No: not annihilating;
$(x-1)^2(x-2)$ — Yes: annihilating and minimal;
$(x-1)^2(x-2)^2$ — No: annihilating but not minimal.

5 The Primary Decomposition Theorem

Our goal is to use the Cayley-Hamilton Theorem and Proposition 2.11 to decompose V into T-invariant subspaces. We start with some remarks on direct sum decompositions.

Let V be a vector space. Recall that V is the direct sum

$$V = W_1 \oplus \cdots \oplus W_r$$

of subspaces $W_1, \cdots, W_r$ if every vector $v \in V$ can be written uniquely as a sum

$$v = w_1 + \cdots + w_r \text{ with } w_i \in W_i.$$

For each i, let $\mathcal{B}_i$ be a basis for W_i. Then

$$\mathcal{B} = \bigcup_i \mathcal{B}_i \text{ is a basis for } V.$$

Assume from now on that V is finite dimensional. If $T : V \to V$ is a linear map such that each W_i is T-invariant, then the matrix of T with respect to the basis $\mathcal{B}$ is block diagonal

$${}_{\mathcal{B}}[T]_{\mathcal{B}} = \begin{bmatrix} A_1 & & \\ & \ddots & \\ & & A_r \end{bmatrix} \text{ with } A_i =_{\mathcal{B}_i} [T|_{W_i}]_{\mathcal{B}_i}$$

and $\chi_T(x) = \chi_{T|_{W_1}}(x) \cdots \chi_{T|_{W_r}}(x)$.

Proposition 5.1. *Assume $f(x) = a(x)b(x)$ with $\gcd(a, b) = 1$ and $f(T) = 0$. Then*

$$V = \mathrm{Ker}(a(T)) \oplus \mathrm{Ker}(b(T))$$

is a T-invariant direct sum decomposition. Furthermore, if $f = m_T$ is the minimal polynomial of T and a and b are monic, then

$$m_{T|_{\mathrm{Ker}(a(T))}}(x) = a(x) \text{ and } m_{T|_{\mathrm{Ker}(b(T))}}(x) = b(x).$$

Proof. By Proposition 2.11, there exist s, t such that $as + bt = 1$. But then $a(T)s(T) + b(T)t(T) = Id_V$ and for all $v \in V$

$$(*) \qquad a(T)s(T)v + b(T)t(T)v = v.$$

As f is annihilating

$$a(T)(b(T)t(T)v) = f(T)t(T)v = 0 \text{ and } b(T)(a(T)s(T)v) = 0.$$

This shows that $V = \mathrm{Ker}(a(T)) + \mathrm{Ker}(b(T))$.

To show that this is a direct sum decomposition, assume that $v \in \mathrm{Ker}(a(T)) \cap \mathrm{Ker}(b(T))$. But then by equation $(*)$ we have $v = 0 + 0 = 0$. Thus

$$V = \mathrm{Ker}(a(T)) \oplus \mathrm{Ker}(b(T)).$$

To see that both factors are T-invariant note that for $v \in \mathrm{Ker}(a(T))$

$$a(T)(T(v)) = T(a(T)v) = T(0) = 0$$

and similarly $b(T)T(v) = 0$ for $v \in \mathrm{Ker}(b(T))$.

Assume now that $f = m_T$ is the minimal polynomial of T and let

$$m_1 = m_{T|_{\mathrm{Ker}(a(T))}} \text{ and } m_2 = m_{T|_{\mathrm{Ker}(b(T))}}.$$

Then $m_1 | a$ as

$$0 = a(T)|_{\mathrm{Ker}(a(T))} = a(T|_{\mathrm{Ker}(a(T))})$$

and similarly $m_2 | b$. But also

$$m_T = ab \,|\, m_1 m_2$$

as $m_1(T)m_2(T) = 0$ for: Any $v \in V$ can be written as $v = w_1 + w_2$ with $w_1 \in \mathrm{Ker}(a(T))$, $w_2 \in \mathrm{Ker}(b(T))$ and thus

$$\begin{aligned} &m_1(T)m_2(T)v = \\ &m_2(T)(m_1(T)w_1) + m_1(T)(m_2(T)w_2) = \\ &0 + 0 = 0. \end{aligned}$$

Hence, for degree reasons and because all these polynomials are monic, we see $m_1 = a$, $m_2 = b$. □

Theorem 5.2. *[Primary Decomposition Theorem] Let m_T be the minimal polynomial and write it in the form*

$$m_T(x) = f_1^{q_1}(x) \cdots f_r^{q_r}(x)$$

where the f_i are distinct monic irreducible polynomials. Put $W_i := \mathrm{Ker}(f_i^{q_i}(T))$. Then

$$\begin{aligned} &(i)\, V = W_1 \oplus \cdots \oplus W_r; \\ &(ii)\, W_i \text{ is } T\text{-invariant}; \\ &(iii)\, m_{T|_{W_i}} = f_i^{q_i}. \end{aligned}$$

Proof. Put $a = f_1^{q_1} \cdots f_{r-1}^{q_{r-1}}$, and $b = f_r^{q_r}$ and proceed by induction on r applying Proposition 5.1. □

Proposition 5.3. *There exists unique distinct irreducible monic polynomials $f_1, \cdots, f_r \in \mathbb{F}[x]$ and positive integers $n_i \geq q_i > 0$ $(1 \leq i \leq r)$ such that*

$$m_T(x) = f_1^{q_1} \cdots f_r^{q_r} \text{ and } \chi_T = \pm f_1^{n_1} \cdots f_r^{n_r}.$$

Proof. Factor $m_T = f_1^{q_1} \cdots f_r^{q_r}$ into distinct monic irreducibles over $\mathbb{F}[x]$ (this is unique, since factorisation in $\mathbb{F}[x]$ is unique). By Cayley-Hamilton as $m_T | \chi_T$ we see

$$\chi_T = f_1^{n_1} \cdots f_r^{n_r} \cdot b(x)$$

for some $n_i \geq q_i$ and $b(x) \in \mathbb{F}[x]$ with $b(x)$ coprime to $f_1^{n_1} \cdots f_r^{n_r}$. Since χ_T and m_T have the same roots over $\overline{\mathbb{F}}$ (Theorem 2.18) we see $b(x)$ has no roots and so must be constant; indeed comparing leading coefficients $b(x) = (-1)^n$ where $n = \dim(V)$. □

Note:

T is triangularisable (over a given field)
$\iff$ χ_T factors as a product of linear polynomials (over that field)
$\iff$ each f_i is linear
$\iff$ m_T factors as a product of linear polynomials

Theorem 5.4. *T is diagonalisable $\iff$ m_T factors as a product of distinct linear polynomials.*

Proof. If T is diagonisable then there exists a basis $\mathcal{B}$ of eigenvectors for V such that ${}_{\mathcal{B}}[T]_{\mathcal{B}}$ is diagonal with entries from a list of distinct eigenvalues $\lambda_1, \ldots, \lambda_r$. Then

$$m(x) = (x - \lambda_1) \ldots (x - \lambda_r)$$

is annihilating as $m(T)v = 0$ for any element $v \in \mathcal{B}$ and hence for any $v \in V$. It is also minimal as every eigenvalue is a root of the minimal polynomial by Theorem 2.18.

Conversely, assume that $m_T(x) = (x - \lambda_1) \ldots (x - \lambda_r)$. Then by the Primary Decomposition Theorem

$$V = E_{\lambda_1} \oplus \cdots \oplus E_{\lambda_r},$$

with $E_{\lambda_i} := \mathrm{Ker}(T - \lambda_i I)$, is a direct sum decomposition of V into eigenspaces. Taking $\mathcal{B} = \bigcup_i \mathcal{B}_i$ with each $\mathcal{B}_i$ a basis for E_{λ_i} gives a basis of eigenvectors with respect to which T is diagonal. $\square$

Example 5.5

$$\begin{aligned} P \text{ is a projection} &\iff P^2 = P \\ &\iff P(P - I) = 0. \end{aligned}$$

$$m_P(x) = \begin{cases} x & \Rightarrow P = 0 \\ (x-1) & \Rightarrow P = I \\ x(x-1) & \Rightarrow V = E_0 \oplus E_1, \end{cases} \qquad {}_{\mathcal{B}}[P]_{\mathcal{B}} = \begin{pmatrix} 0 & 0 \\ 0 & I \end{pmatrix}.$$

Example 5.6

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

$$\chi_A(x) = (1 - x)(1 - x) + 1 = x^2 - 2x + 2.$$

$$\begin{aligned}
\mathbb{F} = \mathbb{R} &\Rightarrow m_A(x) = \chi_A(x) \text{ has no roots,} \\
&\Rightarrow A \text{ is not triangularisable, nor diagonalisable;} \\
\mathbb{F} = \mathbb{C} &\Rightarrow m_A(x) = \chi_A(x) = (x-(1+i))(x-(1-i)) \\
&\Rightarrow A \text{ is diagonalisable;} \\
\mathbb{F} = \mathbb{F}_5 &\Rightarrow m_A(x) = \chi_A(x) = (x-3)(x-4) \\
&\Rightarrow A \text{ is diagonalisable.}
\end{aligned}$$

6 Jordan Normal Form

The goal of this chapter is to give a good description of linear transformations when restricted to the invariant subspaces that occur in the Primary Decomposition Theorem.

Let V be finite dimensional and $T : V \to V$ be a linear transformation. If $T^n = 0$ for some $n > 0$ then T is called *nilpotent*.

Theorem 6.1. *If T is nilpotent, then its minimal polynomial has the form $m_T(x) = x^m$ for some m and there exists a basis $\mathcal{B}$ of V such that:*

$$ {}_{\mathcal{B}}[T]_{\mathcal{B}} = \begin{pmatrix} 0 & * & & 0 \\ & \ddots & \ddots & \\ & & \ddots & * \\ 0 & & & 0 \end{pmatrix} \text{ with each } * = 0 \text{ or } 1. $$

The proof of this theorem is rather intricate, and best read in parallel with the illustrative example which follows it.

Proof. As T is nilpotent, $T^n = 0$ for some n, and hence $m_T(x) | x^n$. Thus $m_T(x) = x^m$ for some m.

We have

$$\{0\} \subsetneqq \operatorname{Ker}(T) \subsetneqq \operatorname{Ker}(T^2) \subsetneqq \ldots \subsetneqq \operatorname{Ker}(T^{m-1}) \subsetneqq \operatorname{Ker}(T^m) = V.$$

By the minimality of m these inclusions are indeed strict as $\operatorname{Ker}(T^k) = \operatorname{Ker}(T^{k+1})$ implies that $\operatorname{Ker}(T^k) = \operatorname{Ker}(T^{k+s})$ for all $s \geq 0$. (An easy exercise.)

For $1 \leq i \leq m$, let $\mathcal{B}_i \subset \operatorname{Ker}(T^i)$ be such that:

$$\{w + \operatorname{Ker}(T^{i-1}) \,|\, w \in \mathcal{B}_i\} \text{ is a basis for } \operatorname{Ker}(T^i)/\operatorname{Ker}(T^{i-1}).$$

Note that $|\mathcal{B}_i|$ must then be $\dim(\operatorname{Ker}(T^i)) - \dim(\operatorname{Ker}(T^{i-1}))$. (We shall shortly make a particular choice of these sets.)

By Proposition 3.3 and induction, we see that

$$\mathcal{B} = \bigcup_{i=1}^{m} \mathcal{B}_i$$

is a basis for V. More explicitly, considering $T|_{\mathrm{Ker}(T^{m-1})}$ and by induction we find that

$$\bigcup_{i=1}^{m-1} \mathcal{B}_i$$

is a basis for $\mathrm{Ker}(T^{m-1})$. Now $\overline{\mathcal{B}_m}$ is a basis for the quotient $V/\mathrm{Ker}(T^{m-1})$ and so applying Proposition 3.3 we find that

$$\bigcup_{i=1}^{m-1} \mathcal{B}_i \cup \mathcal{B}_m$$

is a basis for V.

Next we make the key observation that for $1 \leq i \leq m-1$ the set

$$\left\{T(w) + \mathrm{Ker}(T^{i-1}) \,|\, w \in \mathcal{B}_{i+1}\right\}$$

is linearly independent in $\mathrm{Ker}(T^i)/\mathrm{Ker}(T^{i-1})$. (Note here that $\mathcal{B}_{i+1} \subset \mathrm{Ker}(T^{i+1})$.) To see why, write $\mathcal{B}_{i+1} = \{w_1, \cdots, w_t\}$. Suppose there exists $a_1, \cdots, a_t \in \mathbb{R}$ with

$$\sum_{s=1}^{t} a_s \left(T(w_s) + \mathrm{Ker}(T^{i-1})\right) = 0_{\mathrm{Ker}(T^i)/\mathrm{Ker}(T^{i-1})}.$$

Then

$$T\left(\sum_{s=1}^{t} a_s w_s\right) \in \mathrm{Ker}(T^{i-1})$$

and so

$$\sum_{s=1}^{t} a_s w_s \in \mathrm{Ker}(T^i).$$

Hence

$$\sum_{s=1}^{t} a_s \left(w_s + \mathrm{Ker}(T^i)\right) = 0_{\mathrm{Ker}T^{i+1}/\mathrm{Ker}(T^i)}$$

which contradicts our choice of $\mathcal{B}_{i+1}$, unless all the coefficients $a_1, \cdots, a_t = 0$.

Note that since $|\mathcal{B}_{i+1}|$ has size $\dim(\mathrm{Ker}(T^{i+1})) - \dim(\mathrm{Ker}(T^i))$ by our key observation we must have

$$|\mathcal{B}_{i+1}| = \dim(\mathrm{Ker}(T^{i+1})) - \dim(\mathrm{Ker}(T^i)) \leq \dim(\mathrm{Ker}(T^i)) - \dim(\mathrm{Ker}(T^{i-1})) = |\mathcal{B}_i|.$$

for $1 \leq i \leq m-1$.

We are now ready to construct the desired basis $\mathcal{B}$, in an inductive manner.

We begin with $i = m$. We take $\mathcal{B}_m$ to be *any* set such that

$$\left\{w + \operatorname{Ker}(T^{m-1}) \,|\, w \in \mathcal{B}_m\right\} \text{ is a basis for } \operatorname{Ker}(T^m)/\operatorname{Ker}(T^{m-1}).$$

By our key observation above, the set

$$\left\{T(w) + \operatorname{Ker}(T^{m-2}) \,|\, w \in \mathcal{B}_m\right\}$$

is linearly independent in

$$\operatorname{Ker}(T^{m-1}) \,/\, \operatorname{Ker}(T^{m-2}).$$

Thus we can extend that set to a basis for the quotient $\operatorname{Ker}(T^{m-1}) \,/\, \operatorname{Ker}(T^{m-2})$; put another way, working in $\operatorname{Ker}(T^{m-1})$ itself we extend the set $T(\mathcal{B}_m)$ to a set

$$\mathcal{B}_{m-1} := T(\mathcal{B}_m) \cup \mathcal{E}_{m-1} \subset \operatorname{Ker}(T^{m-1})$$

whose image in the quotient $\operatorname{Ker}(T^{m-1})/\operatorname{Ker}(T^{m-2})$ is a basis.

We now repeat this process of, for $i = m-1, m-2, \cdots, 2$, considering the image of $T(\mathcal{B}_i)$ in the quotient $\operatorname{Ker}(T^{i-1}) \,/\, \operatorname{Ker}(T^{i-2})$ (which is linearly independent), and extending $T(\mathcal{B}_i)$ to a set

$$\mathcal{B}_{i-1} := T(\mathcal{B}_i) \cup \mathcal{E}_{i-1} \subset \operatorname{Ker}(T^{i-1})$$

whose image in $\operatorname{Ker}(T^{i-1})/\operatorname{Ker}(T^{i-2})$ is a basis.

We end up with a basis for V,

$$\mathcal{B} := \bigcup_{i=1}^{m} \mathcal{B}_i.$$

Defining $\mathcal{E}_m = \mathcal{B}_m$ this can be reordered as

$$\bigcup_{i=m}^{1} \left(\bigcup_{v \in \mathcal{E}_i} \left\{T^{i-1}(v), T^{i-2}(v), \cdots, T(v), v\right\} \right).$$

With respect to this basis, ordered in this way, we get a block diagonal matrix

$${}_{\mathcal{B}}[T]_{\mathcal{B}} = \begin{pmatrix} A_m & & \\ & \ddots & \\ & & A_1 \end{pmatrix}.$$

with each A_i ($m \geq i \geq 1$) itself a block diagonal matrix consisting of $|\mathcal{E}_i|$ many Jordan blocks J_i of size $i \times i$; here

$$J_i := \begin{pmatrix} 0 & 1 & & 0 \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ 0 & & & 0 \end{pmatrix}.$$

□

Example 6.2 Let $T : \mathbb{R}^3 \to \mathbb{R}^3$ be given by

$$A = \begin{pmatrix} -2 & -1 & 1 \\ 14 & 7 & -7 \\ 10 & 5 & -5 \end{pmatrix}.$$

Then $A^2 = 0$ and so $m_A(x) = x^2$ and $\chi_A(x) = -x^3$.

We have

$$\{0\} \subsetneqq \mathrm{Ker}(T) \subsetneqq \mathrm{Ker}(T^2) = \mathbb{R}^3$$

with

$$\mathrm{Ker}(T) = \langle (1,0,2)^t, (0,1,1)^t \rangle$$

and

$$\mathrm{Ker}(T^2) \,/\, \mathrm{Ker}(T) = \langle (1,0,0)^t + \mathrm{Ker}(T) \rangle.$$

Note the dimension jumps here are 2 and 1. So we may choose

$$\begin{aligned} \mathcal{B}_2 &= \{(1,0,0)^t\} \,(= \mathcal{E}_2) \\ \mathcal{B}_1 = T(\mathcal{B}_2) \cup \mathcal{E}_1 &= \{(-2,14,10)^t, (0,1,1)^t\} \\ \text{and } \mathcal{B} &= \mathcal{B}_1 \cup \mathcal{B}_2 = \bigcup_{v \in \mathcal{E}_2} \{T(v), v\} \cup \bigcup_{v \in \mathcal{E}_1} \{v\} = \{(-2,14,10)^t, (1,0,0)^t, (0,1,1)^t\} \end{aligned}$$

Hence

$${}_{\mathcal{B}}[T]_{\mathcal{B}} = \begin{pmatrix} \boxed{\begin{matrix} 0 & 1 \\ 0 & 0 \end{matrix}} & \begin{matrix} 0 \\ 0 \end{matrix} \\ \begin{matrix} 0 & 0 \end{matrix} & \boxed{0} \end{pmatrix}.$$

Corollary 6.3. *Let V be finite dimensional and $T : V \to V$ be a linear transformation. Assume $m_T(x) = (x-\lambda)^m$ for some m. Then, there exists a basis $\mathcal{B}$ of V such that ${}_{\mathcal{B}}[T]_{\mathcal{B}}$ is block diagonal with blocks of the form:*

$$J_i(\lambda) := \lambda I_i + J_i = \begin{pmatrix} \lambda & 1 & & 0 \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ 0 & & & \lambda \end{pmatrix} \text{ and } 1 \le i \le m.$$

Proof. $T - \lambda I$ is nilpotent with minimal polynomial x^m. We may apply Theorem 6.1. So there exists a basis $\mathcal{B}$ such that ${}_{\mathcal{B}}[T - \lambda I]_{\mathcal{B}}$ is block diagonal with blocks J_i and hence

$${}_{\mathcal{B}}[T]_{\mathcal{B}} = \lambda I +_{\mathcal{B}} [T - \lambda I]_{\mathcal{B}}$$

is of the desired form. □

Theorem 6.4. *Let V be finite dimensional and let $T : V \to V$ be a linear map with minimal polynomial*

$$m_T(x) = (x - \lambda_1)^{m_1} \cdots (x - \lambda_r)^{m_r}.$$

Then there exists a basis $\mathcal{B}$ of V such that ${}_{\mathcal{B}}[T]_{\mathcal{B}}$ is block diagonal and each diagonal block is of the form $J_i(\lambda_j)$ for some $1 \le i \le m_j$ and $1 \le j \le r$.

Note: (1) If $\mathbb{F}$ is an algebraically closed field, such as $\mathbb{C}$, then the minimal polynomial will always split into a product such as in the theorem.
(2) There could be several $J_i(\lambda_j)$ for each pair (i, j) (or none, but there is at least one block for $i = m_j$ for each $1 \leq j \leq r$, that is, one of maximal size for each eigenvalue). For each $1 \leq j \leq r$, the number of Jordan blocks $J_i(\lambda_j)$ for $1 \leq i \leq m_j$ is determined by, and determines, the sequence of dimensions $\dim \mathrm{Ker}((T - \lambda_j I)^i)$ for $1 \leq i \leq m_j$. As this sequence of dimensions depends only upon T, it follow that the Jordan form is unique, up to the ordering of the blocks.

Proof. By the Primary Decomposition Theorem 5.3,

$$V = \mathrm{Ker}(T - \lambda_1 I)^{m_1} \oplus \cdots \oplus \mathrm{Ker}(T - \lambda_r I)^{m_r}.$$

Furthermore, T restricted to the j-th summand has minimal polynomial $(x - \lambda_j)^{m_j}$ and hence Corollary 6.2 applies to give the desired result. □

Example 6.5 Let $T : \mathbb{R}^3 \to \mathbb{R}^3$ be given by

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

Then $\chi_T(x) = \det(A - xI) = \cdots = (2 - x)^3$ and $m_T(x) = (x - 2)^3$.

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

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

We have

$$\{0\} \subsetneqq \mathrm{Ker}(A - 2I) \subsetneqq \mathrm{Ker}((A - 2I)^2) \subsetneqq \mathrm{Ker}((A - 2I)^3) = \mathbb{R}^3$$

Note the dimensions increase by exactly one at each step. We choose

$$\begin{aligned} \mathcal{B}_3 &= \{(1, 0, 0)^t\} \text{ as } (A - 2I)^2(1, 0, 0)^t \neq 0 \\ \mathcal{B}_2 &= (A - 2I)(\mathcal{B}_3) = \{(1, -1, 0)^t\} \\ \mathcal{B}_1 &= (A - 2I)(\mathcal{B}_2) = \{(1, 0, -1)^t\} \end{aligned}$$

Here after choosing $\mathcal{B}_3$ we may make no further choices (note $\mathcal{E}_3 = \mathcal{B}_3$ and $\mathcal{E}_2$, $\mathcal{E}_1$ are empty). So we have

$$\mathcal{B} = \mathcal{B}_1 \cup \mathcal{B}_2 \cup \mathcal{B}_3 = \bigcup_{v \in \mathcal{E}_3} \{(T - 2I)^2(v), (T - 2I)(v), v\} = \{(1, 0, -1)^t, (1, -1, 0)^t, (1, 0, 0)^t\}$$

Put

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

Then

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

7 Dual Spaces

Linear maps from a vector space to the ground field play a special role. They have a special name, linear functional, and the collection of all of them form the dual space.

Definition 7.1. *Let V be a vector space over $\mathbb{F}$. Its* **dual** *V' is the vector space of linear maps from V to $\mathbb{F}$, i.e $V' = \mathrm{Hom}(V, \mathbb{F})$. Its elements are called* **linear functionals.**

Example 7.2

(1) Let $V = \mathcal{C}([0,1])$ be the vector space of continuous functions on $[0,1]$. Then, $\int : V \to \mathbb{R}$, which sends f to $\int_0^1 f(t)\,\mathrm{d}t$ is a linear functional:

$$\int_0^1 (\lambda f + g)(t)\,\mathrm{d}t = \int_0^1 (\lambda f(t) + g(t))\,\mathrm{d}t = \lambda \int_0^1 f(t)\,\mathrm{d}t + \int_0^1 g(t)\,\mathrm{d}t \text{ for all } f, g \in V, \lambda \in \mathbb{R}$$

(2) Let V be the vector space of finite sequences, that is, the space

$$\{\ (a_1, a_2, \dots) \mid \text{only finitely many } a_i \neq 0\}.$$

Let $\underline{b} = (b_1, b_2, \dots)$ be any infinite sequence. Then, $\underline{b}((a_1, a_2, \dots)) := \sum a_i b_i \in \mathbb{F}$ is well-defined and linear. Hence, $\underline{b} \in V'$.

Theorem 7.3. *Let V be finite dimensional and let $\mathcal{B} = \{e_1, \dots, e_n\}$ be a basis for V. Define the* **dual** *e_i' of e_i (relative to $\mathcal{B}$) by*

$$e_i'(e_j) = \delta_{ij} = \begin{cases} 1 \text{ if } i = j; \\ 0 \text{ if } i \neq j. \end{cases}$$

Then $\mathcal{B}' := \{e_1', \dots, e_n'\}$ is a basis for V', the **dual basis**. *In particular, the assignment $e_i \mapsto e_i'$ defines an isomorphism of vector spaces. In particular,* $\dim V = \dim V'$.

Proof. Assume for some $a_i \in \mathbb{F}$, we have $\sum a_i e_i' = 0$. Then for all j,

$$0 = \sum a_i e_i'(e_j) = a_j$$

and thus $\mathcal{B}'$ is linearly independent.

Let $f \in V'$ and put $a_i := f(e_i)$. Then $f = \sum a_j e'_j$ since they evaluate both to a_i on e_i and any linear map is determined entirely by its values on any basis. Hence $\mathcal{B}'$ is spanning. □

Note though that for $v \in V$ the symbol "v'" on its own has no meaning.

Example 7.4

(1) If $V = \mathbb{R}^n$ (column vectors) then we may "naturally" identify V' with the space $(\mathbb{R}^n)^t$ of row vectors. The dual basis of the standard basis of $\mathbb{R}^n$ is given by the row vectors $e'_i = (0, \ldots, 0, 1, 0 \ldots, 0)$ with the 1 at the i-th place. (This identification is "natural" in the sense that then $(e'_i(e_j)) = e'_i e_j$, and more generally to evaluate a linear functional in V' on an element of V we take the product of the $1 \times n$ vector and $n \times 1$ representing them with respect to the standard basis and its dual.)

(2) If V is the set of finite sequences then V' is the set of infinite sequences $\underline{b}$ as any $f \in V'$ is determined uniquely by its values on a basis. Note that, in this case, V is *not* isomorphic to V', which shows that the condition on the dimension in Theorem 7.3 is necessary.

Theorem 7.5. *Let V be a finite dimensional vector space. Then, $V \to (V')' =: V''$ defined by $v \mapsto E_v$ is a natural linear isomorphism; here E_v is the evaluation map at v defined by*

$$E_v(f) := f(v) \text{ for } v \in V'.$$

"Natural" here means independent of a choice of basis. In contrast, the isomorphism $V \cong V'$ of Theorem 7.3 is dependent on the choice of a basis for V.

Proof. First E_v is a linear map from V' to $\mathbb{F}$, since

$$E_v(f + \lambda g) := (f + \lambda g)(v) := f(v) + \lambda g(v) =: E_v(f) + \lambda E_v(g) \text{ for all } f, g \in V',\ v \in V,\ \lambda \in \mathbb{F}.$$

Hence the assignment $v \mapsto E_v$ is well-defined.

The map $v \mapsto E_v$ itself is linear, as $E_{\lambda v + w} = \lambda E_v + E_w$ for all $v, w \in V$ and $\lambda \in \mathbb{F}$, since each functional f is linear.

This map is also injective, since if $E_v = 0$, then $E_v(f) = f(v) = 0$ for all $f \in V'$. If $v \neq 0$, then we can extend $\{e_1 = v\}$ to a basis $\mathcal{B}$ of V. For $f = e'_1$ we then have $E_v(e'_1) = e'_1(e_1) = 1$ which contradicts the fact $E_v(e_1) = 0$. Hence $v = 0$ which proves that the assignment $v \mapsto E_v$ is injective.

By Theorem 7.3,

$$\dim(V) = \dim(V') = \dim(V')'.$$

Thus it follows from the injectivity and the Rank-Nullity Theorem that the assignment is also surjective. □

Note: When V has dimension n, the kernel of a non-zero linear functional $f : V \to \mathbb{F}$ is of dimension $n-1$. The preimage $f^{-1}(\{c\})$ for a constant $c \in \mathbb{F}$ is a called **hyperplane** (not necessarily containing zero) of dimension $n-1$. When $V = \mathbb{F}^n$ (column vectors) every hyperplane is defined by an equation

$$a_1 b_1 + \cdots + a_n b_n = c$$

for a fixed scalar c and fixed $\underline{b} = (b_1, \ldots, b_n) \in (\mathbb{F}^n)^t$ (row vectors).

Also note that when $c = 0$ different choices of $\underline{b}$ can define the same hyperplane (that is, scaling $\underline{b}$ does not change the hyperplane). So different functionals can have the same kernel.

7.1 Annihilators

Definition 7.6. *Let $U \subseteq V$ be a subspace of V. Define the annihilator of U to be:*

$$U^0 = \{f \in V' : \ f(u) = 0 \text{ for all } u \in U\}.$$

Thus $f \in V'$ lies in U^0 if and only if $f|_U = 0$.

Proposition 7.7. *We have that U^0 is a subspace of V'.*

Proof. Let $f, g \in U^0$ and $\lambda \in \mathbb{F}$. Then for all $u \in U$

$$(f + \lambda g)(u) = f(u) + \lambda g(u) = 0 + 0 = 0.$$

So $f + \lambda g \in U^0$. Also, $0 \in U^0$ and $U^0 \neq \emptyset$. □

Theorem 7.8. *Let V be finite dimensional and $U \subseteq V$ be a subspace. Then*

$$\dim(U^0) = \dim(V) - \dim(U).$$

Proof. Let $\{e_1, \cdots, e_m\}$ be a basis for U and extend it to a basis $\{e_1, \cdots, e_n\}$ for V. Let $\{e'_1, \cdots, e'_n\}$ be the dual basis. As $e'_j(e_i) = 0$ for $j = m+1, \cdots, n$ and $i = 1, \cdots, m$,

$$e'_j \in U^0 \text{ for } j = m+1, \cdots, n.$$

Hence $\langle e'_{m+1}, \cdots, e'_n \rangle \subseteq U^0$.

Conversely let $f \in U^0$. Then there exist $a_i \in \mathbb{F}$ such that $f = \sum_{i=1}^{n} a_i e_i'$. As $e_i \in U$ for $i = 1, \cdots, m$,

$$f(e_i) = 0 \text{ and hence } a_i = 0.$$

So $f \in \langle e_{m+1}', \cdots, e_n' \rangle$.

Thus $U^0 = \langle e_{m+1}', \cdots, e_n' \rangle$, and as this set of spanning vectors is a subset of the dual basis, it is linear independent. Thus

$$\dim(U^0) = n - m = \dim(V) - \dim(U).$$

□

Theorem 7.9. *Let U, W be subspaces of V. Then*

(1) $U \subseteq W \Rightarrow W^0 \subseteq U^0$;
(2) $(U + W)^0 = U^0 \cap W^0$;
(3) $U^0 + W^0 \subseteq (U \cap W)^0$ *and equal if* $\dim(V)$ *is finite.*

Proof.

$$\begin{aligned}
(1)\ f \in W^0 &\Rightarrow \forall w \in W :\ f(w) = 0 \\
&\Rightarrow \forall u \in U \subseteq W :\ f(u) = 0 \\
&\Rightarrow f \in U^0. \\
(2)\ f \in (U+W)^0 &\iff \forall u \in U, \forall w \in W : f(u+w) = 0 \\
&\iff \forall u \in U :\ f(u) = 0 \text{ and } \forall w \in W :\ f(w) = 0 \\
&\iff f \in U^0 \cap W^0. \\
(3)\ f \in U^0 + W^0 &\Rightarrow \exists\ g \in U^0 \text{ and } h \in W^0\ :\ f = g + h \\
&\Rightarrow \forall x \in U \cap W :\ f(x) = g(x) + h(x) = 0 \\
&\Rightarrow f \in (U \cap W)^0.
\end{aligned}$$

It follows that $U^0 + W^0 \subseteq (U \cap W)^0$. If V is finite dimensional, we show that the two spaces have the same dimension and thus are equal:

$$\begin{aligned}
&\dim(U^0 + W^0) = \dim(U^0) + \dim(W^0) - \dim(U^0 \cap W^0) \\
&= \dim(U^0) + \dim(W^0) - \dim(U+W)^0 \\
&= (\dim(V) - \dim(U)) + (\dim(V) - \dim(W)) - (\dim(V) - \dim(U+W)) \\
&= \dim(V) - \dim(U) - \dim(W) + \dim(U) + \dim(W) - \dim(U \cap W) \\
&= \dim(V) - \dim(U \cap W) \\
&= \dim((U \cap W)^0).
\end{aligned}$$

□

Note: For the last part of the proof we also used the formula familiar from Prelims: For U and W finite dimensional,

$$\dim(U + W) = \dim(U) + \dim(W) - \dim(U \cap W).$$

Theorem 7.10. *Let U be a subspace of a finite dimensional vector space V. Under the natural map $V \to V''(:= (V')')$ given by $v \mapsto E_v$, U is mapped isomorphically to $U^{00}(:= (U^0)^0)$.*

Proof. For $v \in V$, the functional E_v is in U^{00} if and only if for all $f \in U^0$ we have $E_v(f)(= f(v)) = 0$. Hence, if $v \in U$ then $E_v \in U^{00}$ and thus

$$U \subseteq U^{00}.$$

When V is finite dimensional, by Theorem 7.8 we also have that

$$\begin{aligned} \dim(U^{00}) &= \dim(V') - \dim(U^0) \\ &= \dim(V) - (\dim(V) - \dim(U)) \\ &= \dim(U), \end{aligned}$$

and thus $U = U^{00}$. □

Theorem 7.11. *Let $U \subseteq V$ be a subspace. Then there exists a natural isomorphism*

$$U^0 \simeq (V/U)'$$

given by $f \mapsto \bar{f}$, where $\bar{f}(v + U) := f(v)$ for $v \in V$.

Proof. Let $f \in U^0$. Note that $\bar{f}$ is well-defined because $f|_U = 0$. The map $f \mapsto \bar{f}$ is linear as $\overline{\lambda f + h} = \lambda \bar{f} + \bar{h}$ for all $f, h \in U^0$ and $\lambda \in \mathbb{F}$. The map is also injective because:

$$\bar{f} = 0 \Rightarrow \ \bar{f}(v + U) = f(v) = 0 \text{ for all } v \in V \ \Rightarrow f = 0.$$

In the finite dimensional case, considering dimensions of both sides gives the result.

In general, we can construct an inverse for $f \mapsto \bar{f}$ as follows. Let $g \in (V/U)'$. Define $\widehat{g} \in V'$ by $\widehat{g}(v) := g(v + U)$. Then $\widehat{g}$ is linear in v. Furthermore, the map $g \mapsto \widehat{g}$ is linear in g, with image in U^0. Finally, one checks that $\widehat{\bar{f}} = f$ and $\overline{\widehat{g}} = g$. □

7.2 Dual maps

The assignment "$V \mapsto V'$" is *functorial* in the sense that a map between two spaces gives a map between the dual spaces (but in the opposite direction).

Definition 7.12. *Let $T : V \to W$ be a linear map of vector spaces. Define the dual map by*

$$T' : W' \to V', \quad f \mapsto f \circ T$$

Note that $f \circ T : V \to W \to \mathbb{F}$ is linear, and hence $f \circ T \in V'$.

This definition is best illustrated by drawing a little triangular diagram (come to the lectures or draw it yourself).

Proposition 7.13. *We have that T' is a linear map.*

Proof. Let $f, g \in W', \lambda \in \mathbb{F}$. We need to show $T'(f + \lambda g) = T'(f) + \lambda T'(g)$ (an identity of functionals on V). So let $v \in V$. Then,

$$\begin{aligned} T'(f + \lambda g)(v) &= ((f + \lambda g) \circ T)(v) \\ &= (f + \lambda g)(Tv) \\ &= f(Tv) + \lambda g(Tv) \\ &= T'(f)(v) + \lambda T'(g)(v) \\ &= (T'(f) + \lambda T'(g))(v), \end{aligned}$$

as required. □

Theorem 7.14. *Let V and W be two finite dimensional vector spaces. The assignment $T \mapsto T'$ defines a natural isomorphism from $\hom(V, W)$ to $\hom(W', V')$.*

Proof. (a little tedious) We first check that the assignment $T \mapsto T'$ is linear in T. Let $T, S \in \hom(V, W), \lambda \in \mathbb{F}$. We need to show $(T + \lambda S)' = T' + \lambda S'$, an identity of maps from W' to V'. So let $f \in W'$. We now need to show $(T + \lambda S)'(f) = (T' + \lambda S')(f)$, an identify of functionals on V. So (finally!) let $v \in V$. Then we have

$$\begin{aligned} ((T + \lambda S)'(f))(v) &= f((T + \lambda S)(v)) \text{ (definition of dual map)} \\ &= f(T(v) + \lambda S(v)) \text{ (definition of the sum of two maps)} \\ &= f(T(v)) + \lambda f(S(v)) \text{ (as } f \text{ is linear)} \\ &= T'(f)(v) + \lambda S'(f)(v) \text{ (definition of dual map)} \\ &= (T'(f) + \lambda S'(f))(v) \text{ (definition of the sum of two maps)} \\ &= ((T' + \lambda S')(f))(v) \text{ (definition of the sum of two (dual) maps),} \end{aligned}$$

and so $(T + \lambda S)' = T' + \lambda S'$.

To prove injectivity, assume $T' = 0$. Then, for all $f \in W'$ we have $T'(f) = 0$, an identity of functionals on V, i.e., for all $f \in W'$ and for all $v \in V$ we have $T'(f)(v) = 0$. Now

$$T'(f)(v) := f(Tv) =: E_{Tv}(f) = 0.$$

But then $E_{Tv} = 0$, hence $Tv = 0$ by Theorem 7.5 (applied to W). Since this is true for all $v \in V$, we have $T = 0$. Thus, the map defined by $T \mapsto T'$ is injective.

When the vector spaces are finite dimensional, we have

$$\dim(\hom(V, W)) = \dim V \dim W = \dim W' \dim V' = \dim(\hom(W', V'))$$

and hence the map is also surjective. □

Theorem 7.15. *Let V and W be finite dimensional, and let $\mathcal{B}_W$ and $\mathcal{B}_V$ be bases for W and V. Then, for any linear map $T : V \to W$,*

$$({}_{\mathcal{B}_W}[T]_{\mathcal{B}_V})^t = {}_{\mathcal{B}'_V}[T']_{\mathcal{B}'_W}$$

where $\mathcal{B}'_W$ and $\mathcal{B}'_V$ are the dual bases.

Proof. Let $\mathcal{B}_V = \{e_1, \ldots, e_n\}$, $\mathcal{B}_W = \{x_1, \ldots, x_m\}$, and

$$_{\mathcal{B}_W}[T]_{\mathcal{B}_V} = A = (a_{ij}).$$

Then

$$T(e_j) = \sum_{i=1}^{m} a_{ij} x_i \quad \text{and hence} \quad x_i'(T(e_j)) = a_{ij}.$$

Let

$$_{\mathcal{B}_V'}[T']_{\mathcal{B}_W'} = B = (b_{ij}).$$

Then

$$T'(x_i') = \sum_{j=1}^{n} b_{ji} e_j' \quad \text{and hence} \quad (T'(x_i'))(e_j) = b_{ji}.$$

By definition we also have $(T'(x_i'))(e_j) = x_i'(T(e_j))$, and hence $a_{ij} = b_{ji}$ and $A^t = B$. □

Notice, in finite dimension, by Theorems 7.5 and 7.11 that $(U^0)'$ is naturally isomorphic to the quotient space V/U. So if you don't like quotient spaces, you can work instead with duals of annihilators (!). Challenge: prove the triangular form (Theorem 4.5) using duals of annihilators instead of quotient spaces. (It is easier in fact just to work with annihilators, and simultaneously prove a matrix has both an upper and lower triangular form by induction, but the challenge is a good work-out.) Another good challenge is to figure out the natural isomorphism from V/U to $(U^0)'$.

8 Inner Product Spaces

Recall from Prelims Geometry and Linear Algebra that the "dot product" on $\mathbb{R}^n$ (column vectors)

$$\langle v, w \rangle := v^t w$$

is an inner product. There is also a related "dot product" on $\mathbb{C}^n$

$$\langle v, w \rangle := \overline{v}^t w.$$

Note here we conjugate the first vector (whereas we followed the other convention in Prelims and conjugated the second vector). We'll call these the *usual* inner products on $\mathbb{R}^n$ and $\mathbb{C}^n$. They endow these spaces with a notion of length and distance (and angles for $\mathbb{R}^n$), and we will study linear maps which behave in certain ways with respect to these notions, e.g., maps which preserve distance.

Before that though, let us recall some more definitions

Definition 8.1. *Let V be a vector space over a field $\mathbb{F}$. A* **bilinear form** *on V is a map*

$$F : V \times V \to \mathbb{F}$$

such that for all $u, v, w \in V,\ \lambda \in \mathbb{F}$*:*

$$(i)\ F(u+v,w) = F(u,w) + F(v,w)$$
$$(ii)\ F(u,v+w) = F(u,w) + F(v,w)$$
$$(iii)\ F(\lambda v, w) = \lambda F(v,w) = F(v, \lambda w).$$

We say,
F *is* **symmetric** *if:* $F(v,w) = F(w,v)$ *for all* $v, w \in V$*.*
F *is* **non-degenerate** *if:* $F(v,w) = 0$ *for all* $v \in V$ *implies* $w = 0$*.*

Only the last definition is new. When $\mathbb{F} = \mathbb{R}$ we'll say F is **positive definite** if for all $v \neq 0 \in V:\ F(v,v) > 0$. Note that a positive definite form is always non-degenerate (since $F(v,v)$ cannot be 0 for $v \neq 0$).

Example 8.2

(1) (Minkowski space) Let $V = \mathbb{R}^4$ and

$$F((x,y,z,t),(\tilde{x},\tilde{y},\tilde{z},\tilde{t})) = x\tilde{x} + y\tilde{y} + z\tilde{z} - ct\tilde{t}$$

where c is the speed of light (that's fast). Then F is bilinear, symmetric, non-degenerate, but not positive definite. For example, $v = (\sqrt{c}, 0, 0, 1) \neq 0$ but $F(v,v) = 0$.

If F is positive definite then it is non-degenerate, but this example shows the converse does not hold.

(2) The dot product on $\mathbb{R}^n$ is bilinear, symmetric and positive definite.

(3) Let $V = \mathcal{C}[0,1]$, the space of continuous functions on $[0,1]$. Then

$$F(f,g) = \int_0^1 f(t)g(t)dt$$

is bilinear, symmetric and positive definite.

A real vector space V endowed with a bilinear, symmetric positive definite form $F(\cdot,\cdot)$ is (as you know) called an **inner product space**. We usual write the form as $\langle \cdot, \cdot \rangle$.

There is a similar notion for complex vector spaces.

Definition 8.3. *Let* V *be a vector space over* $\mathbb{C}$*. A* **sesquilinear form** *on* V *is a map*

$$F : V \times V \to \mathbb{C}$$

such that for all $u, v, w \in V,\ \lambda \in \mathbb{C}$:

$$(i)\ F(u+v,w) = F(u,w) + F(v,w)$$
$$(ii)\ F(u,v+w) = F(u,v) + F(u,w)$$
$$(iii)\ F(\bar{\lambda}v,w) = \lambda F(v,w) = F(v,\lambda w).$$

We say F *is* **conjugate symmetric** *if*

$$F(v,w) = \overline{F(w,v)} \text{ for all } v, w \in V,$$

and, if so, $F(v,v) \in \mathbb{R}$ *as* $F(v,v) = \overline{F(v,v)}$. *We call a conjugate symmetric form* F **positive definite** *if for all* $v \neq 0 \in V :\ F(v,v) > 0$ *(note* $F(v,v)$ *is necessarily real).*

Example 8.4 Let $V = \mathbb{C}^n$ and

$$F(v,w) = \bar{v}^t A w$$

for some $A \in M_{n\times n}(\mathbb{C})$. Then F is a sesquilinear form, and F is conjugate symmetric if and only if $A = \bar{A}^t$ as for all $i, j = 1, \cdots, n$ we have

$$F(e_i, e_j) = \bar{e}_i^t A e_j = a_{ij} \text{ and } \overline{F(e_j, e_i)} = \overline{\bar{e}_j^t A e_i} = \bar{a}_{ji}.$$

F is non-degenerate if and only if A is non-singular as:

$$\begin{aligned} &\text{A is singular} \\ &\iff \exists w \in V : Aw = 0 \\ &\iff \exists w \in V\ \ \forall v \in V :\ \bar{v}^t A w = 0 \\ &\iff F \text{ is degenerate (i.e. not non-degenerate).} \end{aligned}$$

So to understand all sesquilinear, conjugate symmetric (and non-degenerate) forms on $\mathbb{C}^n$ we need to understand (non-singular) matrices which are conjugate symmetric.

A complex vector space V with a sesquilinear, conjugate symmetric, positive definite form $F = \langle \cdot, \cdot \rangle$ is (as you know) called a (complex) **inner product space**.

Given a real or complex inner product space, we say $\{w_1, \cdots, w_n\}$ are mutually **orthogonal** if $\langle w_i, w_j \rangle = 0$ for all $i \neq j$, and they are **orthonormal** if they are mutually orthogonal and $\langle w_i, w_i \rangle = 1$ for each i.

Proposition 8.5. *Let* V *be an inner product space over* $\mathbb{K}$ *(equal* $\mathbb{R}$ *or* $\mathbb{C}$*) and* $w_1, \cdots, w_n \subset V$ *be orthogonal with* $w_i \neq 0$ *for all* i. *Then* $w_1, \cdots, w_n$ *are linearly independent.*

Proof. Assume $\sum_i \lambda_i w_i = 0$ for some $\lambda_i \in \mathbb{K}$. Then for all j, $\langle w_j, \sum_i \lambda_i w_i \rangle = 0$. But

$$\langle w_j, \sum_i \lambda_i w_i \rangle = \sum_i \lambda_i \langle w_j, w_i \rangle = \lambda_j \langle w_j, w_j \rangle.$$

Cancelling we see $\lambda_j = 0$ for all j. □

8.1 Gram-Schmidt orthonormalisation process

Let $\mathcal{B} = \{v_1, \cdots, v_n\}$ be a basis of the inner product space V over $\mathbb{K} = \mathbb{R}, \mathbb{C}$. Put

$$\begin{aligned} w_1 &= v_1 \\ w_2 &= v_2 - \frac{\langle w_1, v_2\rangle}{\langle w_1, w_1\rangle} w_1 \\ &\vdots \\ w_k &= v_k - \sum_{i=1}^{k-1} \frac{\langle w_i, v_k\rangle}{\langle w_i, w_i\rangle} w_i \qquad (*) \end{aligned}$$

Assuming that $\langle w_1, \cdots, w_{k-1}\rangle = \langle v_1, \cdots, v_{k-1}\rangle$, the identity $(*)$ shows that

$$\langle w_1, \cdots, w_k\rangle = \langle w_1, \cdots, w_{k-1}, v_k\rangle = \langle v_1, \cdots, v_k\rangle.$$

Assuming that $\{w_1, \cdots, w_{k-1}\}$ are orthogonal, we have for $j < k$

$$\langle w_j, w_k\rangle = \langle w_j, v_k\rangle - \frac{\langle w_j, v_k\rangle}{\langle w_j, w_j\rangle}\langle w_j, w_j\rangle = 0.$$

Then by induction $\mathcal{D} = \{w_1, \cdots, w_n\}$ is an orthogonal, spanning set and hence, by Proposition 8.5, an orthogonal basis.

Put

$$u_i = \frac{w_i}{\|w_i\|} \text{ where } \|w_i\| = \sqrt{\langle w_i, w_i\rangle}.$$

Then $\mathcal{E} = \{u_i, \cdots, u_n\}$ is an orthonormal basis.

Corollary 8.6. *Every finite dimensional inner product space V over $\mathbb{K} = \mathbb{R}, \mathbb{C}$ has an orthonormal basis.*

When we identify a finite dimensional inner product space V with $\mathbb{R}^n$ or $\mathbb{C}^n$ by choosing an orthonormal basis, the inner product may be identified with the usual (dot) inner product on $\mathbb{R}^n$ or $\mathbb{C}^n$ (check this). Thus finite dimensional inner product spaces are just an (orthonormal) basis-free way of thinking about $\mathbb{R}^n$ or $\mathbb{C}^n$ with the usual inner product.

Note that the Gram-Schmidt process tells us that given such a V and a subspace U, then any *orthonormal* basis for U may be extend to one for V (think about why). This is very important.

8.2 Orthogonal complements and duals of inner product spaces

Let V be an inner product space over $\mathbb{K} = \mathbb{R}, \mathbb{C}$. Then for all $v \in V$,

$$\begin{aligned} \langle v, \cdot \rangle : \ & V \to \mathbb{K} \\ & w \mapsto \langle v, w \rangle \end{aligned}$$

is a linear functional, as $\langle \ , \ \rangle$ is linear in the second co-ordinate.

Theorem 8.7. *The map defined by $v \mapsto \langle v, \cdot \rangle$ is a natural injective $\mathbb{R}$-linear map $\phi : V \to V'$, which is an isomorphism when V is finite dimensional.*

Note here that every complex vector space V is in particular a real vector space, and if it is finite dimensional then

$$2 \dim_{\mathbb{C}} V = \dim_{\mathbb{R}} V.$$

Proof. Note $\phi : v \mapsto \langle v, \cdot \rangle$, so and we must first show $\phi(v + \lambda w) = \phi(v) + \lambda\phi(w)$ for all $v, w \in V, \lambda \in \mathbb{R}$, i.e.

$$\langle v + \lambda w, \cdot \rangle = \langle v, \cdot \rangle + \lambda \langle w, \cdot \rangle.$$

And this is true. So ϕ is $\mathbb{R}$-linear. (Note it is conjugate linear for $\lambda \in \mathbb{C}$.) As $\langle \cdot, \cdot \rangle$ is non-degenerate, $\langle v, \cdot \rangle = \overline{\langle \cdot, v \rangle}$ is not the zero functional unless $v = 0$. Hence, ϕ is injective. If V is finite dimensional, then $\dim_{\mathbb{R}} V = \dim_{\mathbb{R}} V'$, and hence $\operatorname{Im} \phi = V'$. Thus, ϕ is surjective and hence an $\mathbb{R}$-linear isomorphism. □

Definition 8.8. *Let $U \subseteq V$ be a subspace of an inner product space V. The* **orthogonal complement** *is defined as follows:*

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

Note we might equally have said "$\langle v, u \rangle = 0$ for all $u \in U$".

Proposition 8.9. *We have that $U^{\perp}$ is a subspace of V.*

Proof. First $0 \in U^{\perp}$. Now let $v, w \in U^{\perp}$ and $\lambda \in \mathbb{K}$. Then, for all $u \in U$,

$$\langle u, v + \lambda w \rangle = \langle u, v \rangle + \lambda \langle u, w \rangle = 0 + 0 = 0.$$

□

Proposition 8.10. *1. $U \cap U^{\perp} = \{0\}$*

2. $U \oplus U^{\perp} = V$ if V is finite dimensional (and so $\dim U^{\perp} = \dim V - \dim U$)

3. $(U + W)^{\perp} = U^{\perp} \cap W^{\perp}$

4. $(U \cap W)^{\perp} \supseteq U^{\perp} + W^{\perp}$ (with equality if $\dim V < \infty$)

5. $U \subseteq (U^\perp)^\perp$ *(with equality if V is finite dimensional)*

Outside of finite dimension, part *2.* may fail, and the inclusions in *4.* and *5.* may be strict (you'll see this for *2.* and *5.* on the problem sheets).

Proof. 1. If $u \in U$ and $u \in U^\perp$ then $\langle u, u\rangle = 0$. As $\langle\ ,\ \rangle$ is positive definite, this implies that $u = 0$.

2. If V is finite-dimensional, then there exists an orthonormal basis
$$\{e_1, \ldots, e_n\}$$
of V such that $\{e_1, \ldots, e_k\}$ is a basis for U.
Now, assume
$$v = \sum_{i=1}^{n} a_i e_i \in U^\perp.$$
Then $\langle e_i, v\rangle = a_i = 0$ for $i = 1, \ldots, k$. Hence,
$$v \in \langle e_{k+1}, \ldots, e_n\rangle.$$
Conversely, note that $e_j \in U^\perp$ for $j = k+1, \ldots, n$, and hence
$$U^\perp = \langle e_{k+1}, \ldots, e_n\rangle \quad \text{and} \quad U \oplus U^\perp = V.$$

3. Exercise.

4. Exercise.

5. Let $u \in U$. Then, for all $w \in U^\perp$,
$$\langle u, w\rangle = \overline{\langle w, u\rangle} = 0$$
and hence $\langle w, u\rangle = 0$ and $u \in (U^\perp)^\perp$. If V is finite dimensional, then
$$\dim((U^\perp)^\perp) = \dim V - \dim U^\perp = \dim U$$
and so $U = (U^\perp)^\perp$.

□

Proposition 8.11. *Let V be finite dimensional. Then, under the $\mathbb{R}$-linear isomorphism $\phi : V \to V'$ given by $v \mapsto \langle v, \cdot\rangle$, the space $U^\perp$ maps isomorphically to U^0 (considered as $\mathbb{R}$ vector spaces).*

Proof. Let $v \in U^\perp$. Then for all $u \in U$,
$$\langle u, v\rangle = \overline{\langle v, u\rangle} = 0,$$
and hence $\langle v, \cdot\,\rangle \in U^0$. Hence $\mathrm{Im}(\phi|_{U^\perp}) \subseteq U^0$. We also have that
$$\dim U^\perp = \dim V - \dim U = \dim U^0,$$
and so $\phi(U^\perp) = U^0$. (Note that ϕ has trivial kernel so we need only check equality of dimensions.)
□

Example 8.12 Let V be the vector space of real polynomials with degree at most two. Define

$$\langle f, g\rangle := f(1)g(1) + f(2)g(2) + f(3)g(3).$$

Then, $\langle\ ,\ \rangle$ is bilinear, symmetric and positive definite for:

$$\begin{aligned}\langle f, f\rangle = 0 &\Rightarrow f(1) = f(2) = f(3) = 0\\ &\Rightarrow f \text{ is a polynomial of degree } \geq 3 \text{ or } f = 0.\end{aligned}$$

Since f has degree at most two, $f = 0$.

Now, let $U = \langle 1, t\rangle$. We want to find $f \in U, g \in U^\perp$ such that

$$t^2 = f + g.$$

For any orthonormal basis $\{u_1, u_2\}$ of U define

$$f = \langle t^2, u_1\rangle u_1 + \langle t^2, u_2\rangle u_2$$

Then by construction $t^2 - f$ lies in $U^\perp$, and so we may take $g = t^2 - f$.

We will apply the Gram-Schmidt process to obtain an orthogonal basis for U starting with the standard basis $\{1, t\}$. Put

$$\begin{aligned}u_1 &= \frac{1}{\sqrt{3}} \quad \text{(Note: } \langle 1, 1\rangle = 3)\\ w_2 &= t - \langle t, u_1\rangle u_1\\ &= t - \frac{1}{\sqrt{3}}(1 + 2 + 3)\frac{1}{\sqrt{3}} = t - 2\\ u_2 &= \frac{w_2}{\|w_2\|}\\ &= \frac{t-2}{((-1)^2 + 0 + 1^2)^{\frac{1}{2}}}\\ &= \frac{t-2}{\sqrt{2}}\end{aligned}$$

Hence we can take

$$\begin{aligned}f &= \langle t^2, \frac{1}{\sqrt{3}}\rangle\frac{1}{\sqrt{3}} + \langle t^2, \frac{t-2}{\sqrt{2}}\rangle\frac{t-2}{\sqrt{2}}\\ &= \frac{1}{3}(1 + 4 + 9) + \frac{1}{2}(-1 + 0 + 9)(t - 2)\\ &= 4t - \frac{10}{3}.\end{aligned}$$

8.3 Adjoint maps

Let V be an inner product space over $\mathbb{K}$ where $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$.

Definition 8.13. *Given a linear map $T : V \to V$, a linear map $T^* : V \to V$ is its* **adjoint** *if for all $v, w \in V$,*

$$\langle v, T(w)\rangle \;=\; \langle T^*(v), w\rangle. \qquad (*)$$

Lemma 8.14. *If T^* exists it is unique.*

Proof. Let $\tilde{T}$ be another map satisfying $(*)$. Then for all $v, w \in V$

$$\begin{aligned}\langle T^*(v) - \tilde{T}(v), w\rangle &= \langle T^*(v), w\rangle - \langle \tilde{T}(v), w\rangle \\ &= \langle v, T(w)\rangle - \langle v, T(w)\rangle \\ &= 0.\end{aligned}$$

But $\langle\ ,\ \rangle$ is non-degenerate and hence for all $v \in V$

$$T^*(v) - \tilde{T}(v) = 0,$$

and so $T^* = \tilde{T}$. □

Theorem 8.15. *Let $T : V \to V$ be linear where V is finite dimensional. Then the adjoint exists and is linear.*

Proof. Let $v \in V$ and consider the map $V \to \mathbb{K}$ given by

$$w \mapsto \langle v, T(w)\rangle.$$

Then $\langle v, T(\cdot)\rangle$ is a linear functional as T is linear and as $\langle\ ,\ \rangle$ is linear in the second coordinate. As V is finite dimensional, $\phi : V \to V'$ given by $\phi(u) = \langle u, \cdot\rangle$ is an $\mathbb{R}$-linear isomorphism, and in particular a surjective map. Thus there exists $u \in V$ such that

$$\langle v, T(\cdot)\rangle = \langle u, \cdot\,\rangle$$

Defining $T^*(v) := u$ we therefore have

$$\langle v, T(\cdot)\rangle = \langle T^*(v), \cdot\rangle, \text{ i.e., } \langle v, T(w)\rangle = \langle T^*(v), w\rangle \text{ for all } w \in V.$$

To see that T^* is linear, note that for all $v_1, v_2, w \in V,\ \lambda \in \mathbb{K}$,

$$\begin{aligned}\langle T^*(v_1 + \lambda v_2), w\rangle &= \langle v_1 + \lambda v_2, T(w)\rangle \\ &= \langle v_1, T(w)\rangle + \bar{\lambda}\langle v_2, T(w)\rangle \\ &= \langle T^*(v_1), w\rangle + \bar{\lambda}\langle T^*(v_2), w\rangle \\ &= \langle T^*(v_1) + \lambda T^*(v_2), w\rangle\end{aligned}$$

(These equalities have nothing to do with our actual definition of T^*, but just follow from the fact that by construction it satisfies $\langle v, T(w)\rangle = \langle T^*(v), w\rangle$ for all $v, w \in V$.) As $\langle\ ,\ \rangle$ is non-degenerate (equivalently, as ϕ is injective)

$$T^*(v_1 + \lambda v_2) = T^*(v_1) + \lambda T^*(v_2).$$

□

Proposition 8.16. *Let $T : V \to V$ be linear and let $\mathcal{B} = \{e_1, \cdots, e_n\}$ be an orthonormal basis for V. Then*

$$_{\mathcal{B}}[T^*]_{\mathcal{B}} = (\overline{_{\mathcal{B}}[T]_{\mathcal{B}}})^t.$$

Proof. Let $A = {}_{\mathcal{B}}[T]_{\mathcal{B}}$. Then

$$a_{ij} = \langle e_i, T(e_j)\rangle.$$

Let $B = {}_{\mathcal{B}}[T^*]_{\mathcal{B}}$. Then

$$b_{ij} = \langle e_i, T^*(e_j)\rangle = \overline{\langle T^*(e_j), e_i\rangle} = \overline{\langle e_j, T(e_i)\rangle} = \overline{a_{ji}},$$

and hence, $B = \bar{A}^t$. □

Note that

(1) Theorem 8.15 is false if V is not finite dimensional (the inner product defines a metric on V, and you need assumptions like the map being continuous with respect to this).

(2) Proposition 8.16 is false if $\mathcal{B}$ is not orthonormal.

(3) For $\mathbb{K} = \mathbb{R}$ and in finite dimension, under the isomorphism $\phi : V \to V'$, $v \mapsto \langle v, \cdot\rangle$, the adjoint T^* is identified with the dual map T', and an orthonormal basis $\mathcal{B}$ of V with its dual basis so that:

$$_{\mathcal{B}'}[T']_{\mathcal{B}'} = (_{\mathcal{B}}[T]_{\mathcal{B}})^t = {}_{\mathcal{B}}[T^*]_{\mathcal{B}}.$$

Proposition 8.17. *Let $S, T : V \to V$ be linear, V finite dimensional and $\lambda \in \mathbb{K}$. Then:*

(1) $(S+T)^* = S^* + T^*$
(2) $(\lambda T)^* = \bar{\lambda} T^*$
(3) $(ST)^* = T^* S^*$
(4) $(T^*)^* = T$
(5) *If m_T is the minimal polynomial of T then $m_{T^*} = \overline{m_T}$.*

Proof. Exercise. □

Definition 8.18. *A linear map $T : V \to V$ is* **self-adjoint** *if $T = T^*$.*

Lemma 8.19. *If λ is an eigenvalue of a self-adjoint linear operator then $\lambda \in \mathbb{R}$.*

Proof. Assume $w \neq 0$ and $T(w) = \lambda w$ for some $\lambda \in \mathbb{C}$. Then

$$\begin{aligned}\lambda\langle w, w\rangle &= \langle w, \lambda w\rangle = \langle w, T(w)\rangle = \langle T^*(w), w\rangle \\ &= \langle T(w), w\rangle = \langle \lambda w, w\rangle = \bar{\lambda}\langle w, w\rangle.\end{aligned}$$

Hence, as $\langle w, w\rangle \neq 0$, $\lambda = \bar{\lambda}$ and $\lambda \in \mathbb{R}$. □

Lemma 8.20. *If T is self-adjoint and $U \subseteq V$ is T-invariant, then so is $U^\perp$.*

Proof. Let $w \in U^\perp$. Then for all $u \in U$,

$$\langle u, T(w)\rangle = \langle T^*(u), w\rangle = \langle T(u), w\rangle = 0,$$

as $T(u) \in U$ and $w \in U^\perp$. Hence, $T(w) \in U^\perp$. □

Theorem 8.21. *If $T : V \to V$ is self-adjoint and V is finite dimensional, then there exists an orthonormal basis of eigenvectors for T.*

Proof. By Lemma 8.19 there exists an eigenvalue $\lambda \in \mathbb{R}$ and $v \neq 0$ such that $T(v) = \lambda v$. Consider $U = \langle v\rangle$. Then U is T-invariant and by Lemma 8.20 the restriction

$$T|_{\langle v\rangle^\perp} : \langle v\rangle^\perp \to \langle v\rangle^\perp$$

is well-defined; it is still self-adjoint. So by induction on $n := \dim(V)$, we may assume that there exists an orthonormal basis $\{e_2, \cdots, e_n\}$ of eigenvectors for $T|_{\langle v\rangle^\perp}$. Put $e_1 = \frac{v}{\|v\|}$. Then $\{e_1, \cdots, e_n\}$ is an orthonormal basis of eigenvectors for T. □

8.4 Orthogonal and unitary transformations

Let $\{e_1, \ldots, e_n\}$ be an orthonormal basis in $\mathbb{K}^n$, where $\mathbb{K} = \mathbb{R}, \mathbb{C}$, with the usual inner product. Let A be the matrix with columns e_j:

$$A = [e_1, \ldots, e_n].$$

Then

$$A\bar{A}^t = \bar{A}^t A = I \text{ and } A^{-1} = \bar{A}^t,$$

that is to say A is **orthogonal** if $\mathbb{K} = \mathbb{R}$ and **unitary** if $\mathbb{K} = \mathbb{C}$. More generally:

Definition 8.22. *Let V be a finite dimensional inner product space and $T : V \to V$ be a linear transformation. If $T^* = T^{-1}$ then T is called*

orthogonal	*when* $\mathbb{K} = \mathbb{R}$;
unitary	*when* $\mathbb{K} = \mathbb{C}$.

Let $\mathcal{B}$ be an orthonormal basis for V and let T be an orthogonal/unitary transformation of V. Then ${}_{\mathcal{B}}[T]_{\mathcal{B}}$ is an orthogonal/unitary matrix, by Proposition 8.16.

Theorem 8.23. *The following are equivalent:*

(1) $T^* = T^{-1}$;
(2) T *preserves inner products:* $\langle v, w\rangle = \langle Tv, Tw\rangle$ *for all* $v, w \in V$;
(3) T *preserves lengths:* $\|v\| = \|Tv\|$ *for all* $v \in V$.

Proof.

$$
\begin{aligned}
(1) \Rightarrow (2) \quad \langle v, w\rangle &= \langle \mathrm{Id}\, v, w\rangle \\
&= \langle T^*Tv, w\rangle \\
&= \langle Tv, Tw\rangle,\ \forall v, w \in V \\
(2) \Rightarrow (3) \quad \|v\|^2 &= \langle v, v\rangle \\
&= \langle Tv, Tv\rangle \\
&= \|Tv\|^2, \forall v \in V \\
(2) \Rightarrow (1) \quad \langle v, w\rangle &= \langle Tv, Tw\rangle \\
&= \langle T^*Tv, w\rangle,\ \forall v, w \in V \\
&\Rightarrow T^*Tv = v \quad \text{by non-deg of } \langle\ ,\ \rangle, \forall v \in V \\
&\Rightarrow T^*T = \mathrm{Id}
\end{aligned}
$$

(3) $\Rightarrow$ (2) by the (equations in the proof of the) proposition below.

□

Proposition 8.24. *The length function determines the inner product: Given two inner products* $\langle\ ,\ \rangle_1$ *and* $\langle\ ,\ \rangle_2$*,*

$$\langle v, v\rangle_1 = \langle v, v\rangle_2\ \forall\ v \in V \Leftrightarrow \langle v, w\rangle_1 = \langle v, w\rangle_2\ \forall\ v, w \in V.$$

Proof. The implication $\Leftarrow$ is trivial. For the implication $\Rightarrow$ note that

$$\langle v + w, v + w\rangle = \langle v, v\rangle + \langle v, w\rangle + \overline{\langle v, w\rangle} + \langle w, w\rangle$$

and

$$\langle v + iw, v + iw\rangle = \langle v, v\rangle + i\langle v, w\rangle - i\overline{\langle v, w\rangle} + \langle w, w\rangle.$$

Hence,

$$
\begin{aligned}
\mathrm{Re}\langle v, w\rangle &= \frac{1}{2}(\|v + w\|^2 - \|v\|^2 - \|w\|^2) \\
\mathrm{Im}\langle v, w\rangle &= -\frac{1}{2}(\|v + iw\|^2 - \|v\|^2 - \|w\|^2).
\end{aligned}
$$

Thus the inner product is given in terms of the length function. □

Note that inner product spaces are metric spaces with $d(v, w) = \|v - w\|$ and orthogonal/unitary linear transformations are isometries, so we have another equivalence:

$$(4)\ d(v, w) = \|v - w\| = \|Tv - Tw\| = d(Tv, Tw) \text{ for all } v, w \in V.$$

We define the following groups of matrices.

Definition 8.25. *Let*

$$\begin{aligned} O(n) &= \{A \in M_{n\times n}(\mathbb{R}) | A^t A = \mathrm{Id}\}, && \textit{the } \textbf{orthogonal group} \\ SO(n) &= \{A \in O(n) | \det A = 1\}, && \textit{the } \textbf{special orthogonal group} \\ U(n) &= \{A \in M_{n\times n}(\mathbb{C}) | \bar{A}^t A = \mathrm{Id}\}, && \textit{the } \textbf{unitary group} \\ SU(n) &= \{A \in U(n) | \det A = 1\}, && \textit{the } \textbf{special unitary group}. \end{aligned}$$

Lemma 8.26. *If λ is an eigenvalue of an orthogonal/unitary linear transformation $T : V \to V$, then $|\lambda| = 1$.*

Proof. Let $v \neq 0$ be a λ-eigenvector. Then

$$\langle v, v\rangle = \langle Tv, Tv\rangle$$

by Theorem 8.23 which equals

$$\langle \lambda v, \lambda v\rangle = \bar{\lambda}\lambda\langle v, v\rangle$$

and so $1 = \bar{\lambda}\lambda$, that is $|\lambda| = 1$. □

Corollary 8.27. *If A is an orthogonal/unitary $n \times n$-matrix then*

$$|\det A| = 1.$$

Proof. Working over $\mathbb{C}$ we know that $\det A$ is the product of all eigenvalues (with repetitions). Hence,

$$|\det A| = |\lambda_1\lambda_2\cdots\lambda_n| = |\lambda_1||\lambda_2|\cdots|\lambda_n| = 1.$$

□

Lemma 8.28. *Assume that V is finite dimensional and $T : V \to V$ with $T^*T = \mathrm{Id}$. Then if U is T-invariant so is $U^\perp$.*

Proof. Let $w \in U^\perp$. Then for all $u \in U$,

$$\langle u, Tw\rangle = \langle T^*u, w\rangle = \langle T^{-1}u, w\rangle.$$

As U is invariant under T it must be invariant under T^{-1}. (This follows since writing $m_T = x^m + a_{m-1}x^{m-1} + \cdots + a_1x + a_0$ we see that $T(T^{m-1} + a_{m-2}T^{m-2} + \cdots + a_1) = -a_0 I$ and since $a_0 \neq 0$ we get that T^{-1} is a polynomial in T.) Hence, $T^*u \in U$ and $\langle T^*u, w\rangle = 0$ for all $u \in U$. Thus $Tw \in U^\perp$. □

Theorem 8.29. *Assume V is finite dimensional and $T : V \to V$ is unitary. Then there exists an orthonormal basis of eigenvectors.*

Proof. As $\mathbb{K} = \mathbb{C}$ is algebraically closed, there exists a λ and $v \neq 0 \in V$ such that $Tv = \lambda v$. Then $U = \langle v\rangle$ is T-invariant and so is its complement $U^\perp$. Therefore the restriction $T|_{U^\perp}$ is a map of $U^\perp$ to itself which satisfies the hypothesis of the theorem. Working by induction on the

dimension $n := \dim(V)$ and noting that $\dim U^\perp = n-1$, we may assume that there exists an orthonormal basis $\{e_2, \cdots, e_n\}$ of $U^\perp$. Put $e_1 = \frac{v}{\|v\|}$. Then $\{e_1, e_2, \cdots, e_n\}$ is an orthonormal basis of eigenvectors for V. □

Corollary 8.30. *Let $A \in U(n)$. Then there exists $P \in U(n)$ such that $P^{-1}AP$ is diagonal.*

Note that if $A \in O(n)$ then $A \in U(n)$ but A may not be diagonalisable over the reals!

Example 8.31 Let $A \in O(2)$ and let

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

Then $A^tA = I$ and hence

$$a^2 + c^2 = b^2 + d^2 = 1, \quad ab + cd = 0 \text{ and } ad - bc = \pm 1.$$

So

$$A = R_\theta = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix} \text{ or } S_{\theta/2} = \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ \sin(\theta) & -\cos(\theta) \end{pmatrix}.$$

(See Prelims Geometry Example 67.) Note that A is a rotation and $\det(A) = \cos(\theta)^2 + \sin(\theta)^2 = 1$.

$$\begin{aligned} \chi_{R_\theta}(x) &= x^2 - (2\cos(\theta))x + (\cos^2(\theta) + \sin^2(\theta)) \\ &= (x-\lambda)(x-\bar{\lambda}) \qquad \text{for } \lambda = \cos(\theta) + i\sin(\theta) = e^{i\theta}. \end{aligned}$$

Thus R_θ has real eigenvalues if and only if $\theta = 0, \pi$. Also,

$$\begin{aligned} \chi_{S_{\theta/2}}(x) &= x^2 - \cos^2(\theta) - \sin^2(\theta) \\ &= x^2 - 1 = (x-1)(x+1) \end{aligned}$$

So $S_{\theta/2}$ is diagonisable; it is a reflection in the line generated by an eigenvector for $\lambda = 1$ (which is the line $y = x\tan(\theta/2)$ according to Geometry Example 67).

Theorem 8.32. *Let $T : V \to V$ be orthogonal and V be a finite dimensional real vector space. Then there exists an orthonormal basis $\mathcal{B}$ such that:*

$$_{\mathcal{B}}[T]_{\mathcal{B}} = \begin{bmatrix} I & & & & \\ & -I & & & \\ & & R_{\theta_1} & & \\ & & & \ddots & \\ & & & & R_{\theta_\ell} \end{bmatrix} \qquad \theta_i \neq 0, \pi.$$

Proof. [4] Let $S = T + T^{-1} = T + T^*$. Then $S^* = (T+T^*)^* = T^* + T = S$. So S is self-adjoint and has a basis of orthonormal eigenvectors by Theorem 8.21 and thus

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

[4] One can also deduce this theorem just from our spectral theorem for unitary matrices (Theorem 8.29), by grouping the non-real eigenvalues in complex conjugate pairs λ and $\overline{\lambda}$ and taking an orthonormal basis of "real vectors" for each of the two dimensional spaces you get by choosing an eigenvector for λ and the conjugate one for $\overline{\lambda}$.

decomposes into orthogonal eigenspaces of S with distinct eigenvalues $\lambda_1, \ldots, \lambda_k$. Note that each V_i is also T-invariant as for $v \in V_i$

$$S(T(v)) = T(S(v)) = \lambda_i T(v) \text{ and so } T(v) \in V_i.$$

So we may restrict ourselves to $T|_{V_i}$.

By definition of V_i, for all $v \in V_i$ we have $(T + T^{-1})v = \lambda_i v$ and hence $T^2 - \lambda_i T + I = 0$. Thus the minimal polynomial of $T|_{V_i}$ divides $x^2 - \lambda_i x + 1$ and any eigenvalue of $T|_{V_i}$ is a root of it.

If $\lambda_i = \pm 2$, then $(T + \lambda_i I)^2 = 0$ or $(T - \lambda_i I)^2 = 0$. Thus the only eigenvalue of $T|_{V_i}$ is -1 or $+1$, respectively. Since we know $T|_{V_i}$ may be diagonalised over $\mathbb{C}$ (Theorem 8.29), we must have $T|_{V_i} = +I$ or $-I$.

If $\lambda_i \neq \pm 2$ then $T|_{V_i}$ does not have any real eigenvalues as they would have to be ± 1 by Lemma 8.26 (with product $+1$) forcing $\lambda_i = \pm 2$. So $\{v, T(v)\}$ are linearly independent over the reals for $v \neq 0 \in V_i$. Consider the plane $W = \langle v, T(v) \rangle$ spanned by v and Tv. Then W is T-invariant as

$$v \mapsto T(v),\ T(v) \mapsto T^2(v) = \lambda_i T(v) - v.$$

Hence $W^\perp$ is also T-invariant by Lemma 8.28. Repeating the argument for $T|_{W^\perp}$ if necessary, we see that V_i splits into 2-dimensional T-invariant subspaces. By our Example 8.31, with respect to some orthonormal basis of W

$$T|_W = R_\theta = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}$$

for some $\theta \neq 0, \pi$. (Note, the fact that T_W does not have any real eigenvalues implies that T_W is not a reflection and $\theta \neq 0, \pi$.) □

Appendix on non-finite dimensional spaces:

Let V be a vector space and U and W subspaces. Recall that

$$U^0 + W^0 \subseteq (U \cap W)^0.$$

When V is an inner product space we have the similar looking inclusion

$$U^\perp + W^\perp \subseteq (U \cap W)^\perp.$$

This latter inclusion may be strict outside of finite dimension.[5] So what about the former inclusion?

Let's try and prove the reverse inclusion $(U \cap W)^0 \subseteq U^0 + W^0$ directly in the finite dimensional case (rather than appealing to a dimension argument).

So let $f \in (U \cap W)^0$, that is, $f : V \to \mathbb{F}$ with $f|_{U \cap W} = 0$. We need to find $g \in U^0$ and $h \in W^0$ with $f = g + h$.

[5] Thank you to David Seifert for an example, and helpful discussions around this appendix.

First we find a subspace $X \subseteq U$ with

$$(U \cap W) \oplus X = U.$$

We can do this by taking a basis for $U \cap W$ and extending it to one for U. We call this a *direct complement* for $U \cap W$ in U. Likewise we find $Y \subseteq W$ with

$$(U \cap W) \oplus Y = W.$$

One checks that

$$(U \cap W) \oplus X \oplus Y = U + W.$$

So finally we find a subspace $Z \subseteq V$ with

$$((U \cap W) \oplus X \oplus Y) \oplus Z = V.$$

Now we define $g, h : V \to \mathbb{F}$ by giving them on each summand in this direct sum decomposition:

	$U \cap W$	X	Y	Z
g	$f/2$	0	f	$f/2$
h	$f/2$	f	0	$f/2$

Then indeed $g + h = f$ and $g|_U = 0$ and $h|_W = 0$ (note if 2 is not invertible in $\mathbb{F}$ this "symmetric" construction can be easily modified).

Our proof does not mention dimensions. But we do use finite dimensionality, extending bases for a subspace to the whole space (to show every subspace has a direct complement). Can this be done outside of finite dimension too? Well yes, if we assume something called Zorn's Lemma: this is an axiom in mathematics which is (probably) not necessary for most of, for example, my own subject number theory (and one which many mathematicians try to avoid). But it seems to be unavoidable in certain contexts.