Linear Algebra

Cambridge University Mathematical Tripos: Part IB

21st May 2024

Contents

1 Vector spaces and linear dependence

1.1 Vector spaces

Definition. Let F be an arbitrary field. An *F-vector space* is an abelian group $(V, +)$ equipped with a function

$$F \times V \to V; \quad (\lambda, v) \mapsto \lambda v$$

such that
(i) $\lambda(v_1 + v_2) = \lambda v_1 + \lambda v_2$
(ii) $(\lambda_1 + \lambda_2)v = \lambda_1 v + \lambda_2 v$
(iii) $\lambda(\mu v) = (\lambda\mu)v$
(iv) $1v = v$
Such a vector space may also be called a *vector space over F*.

Example. Let X be a set, and define $\mathbb{R}^X = \{f : X \to \mathbb{R}\}$. Then $\mathbb{R}^X$ is an $\mathbb{R}$-vector space, where $(f_1 + f_2)(x) = f_1(x) + f_2(x)$.

Example. Define $M_{n,m}(F)$ to be the set of $n \times m$ F-valued matrices. This is an F-vector space, where the sum of matrices is computed elementwise.

Remark. The axioms of scalar multiplication imply that $\forall v \in V, 0_F v = 0_V$.

1.2 Subspaces

Definition. Let V be an F-vector space. The subset $U \subseteq V$ is a vector subspace of V, denoted $U \leq V$, if
(i) $0_V \in U$
(ii) $u_1, u_2 \in U \implies u_1 + u_2 \in U$
(iii) $(\lambda, u) \in F \times U \implies \lambda u \in U$
Conditions (ii) and (iii) are equivalent to

$$\forall \lambda_1, \lambda_2 \in F, \forall u_1, u_2 \in U, \lambda_1 u_1 + \lambda_2 u_2 \in U$$

This means that U is *stable* by addition and scalar multiplication.

Proposition. If V is an F-vector space, and $U \leq V$, then U is an F-vector space.

Example. Let $V = \mathbb{R}^\mathbb{R}$ be the space of functions $\mathbb{R} \to \mathbb{R}$. The set $C(\mathbb{R})$ of continuous real functions is a subspace of V. The set $\mathbb{P}$ of polynomials is a subspace of $C(\mathbb{R})$.

Example. Consider the subset of $\mathbb{R}^3$ such that $x_1 + x_2 + x_3 = t$ for some real t. This is a subspace for $t = 0$ only, since no other t values yield the origin as a member of the subset.

Proposition. Let V be an F-vector space. Let $U, W \leq V$. Then $U \cap W$ is a subspace of V.

Proof. First, note $0_V \in U, 0_V \in W \implies 0_V \in U \cap W$. Now, consider stability:

$$\lambda_1, \lambda_2 \in F, v_1, v_2 \in U \cap W \implies \lambda_1 v_1 + \lambda_2 v_2 \in U, \lambda_1 v_1 \lambda_2 v_2 \in W$$

Hence stability holds. □

1.3 Sum of subspaces

Remark. The union of two subspaces is not, in general, a subspace. For instance, consider $\mathbb{R}, i\mathbb{R} \subset \mathbb{C}$. Their union does not span the space; for example, $1 + i \notin \mathbb{R} \cup i\mathbb{R}$.

Definition. Let V be an F-vector space. Let $U, W \leq V$. The sum $U + W$ is defined to be the set

$$U + W = \{u + w : u \in U, w \in W\}$$

Proposition. $U + W$ is a subspace of V.

Proof. First, note $0_{U+W} = 0_U + 0_W = 0_V$. Then, for $\lambda_1, \lambda_2 \in F$, and $u \in U, w \in W$,

$$\lambda_1 u + \lambda_2 w = u' + w' \in U + W$$

since $u' \in U, w' \in W$. We can decompose a vector from $U+W$ into its U and W components. Adding these components independently (noting that V is abelian) yields the requirements of a subspace. □

Proposition. The sum $U + W$ is the smallest subspace of V that contains both U and W.

1.4 Quotients

Definition. Let V be an F-vector space. Let $U \leq V$. The quotient space V/U is the abelian group V/U equipped with the scalar multiplication function

$$F \times V/U \to V/U; \quad (\lambda, v + U) \mapsto \lambda v + U$$

Proposition. V/U is an F-vector space.

Proof. We must check that the multiplication operation is well-defined. Indeed, suppose $v_1 + U = v_2 + U$. Then,

$$v_1 - v_2 \in U \implies \lambda(v_1 - v_2) \in U \implies \lambda v_1 + U = \lambda v_2 + U \in V/U$$

□

1.5 Span

Definition. Let V be an F-vector space. Let $S \subset V$. We define the span of S, written $\langle S \rangle$, as the set of finite linear combinations of elements of S. In particular,

$$\langle S \rangle = \left\{ \sum_{s \in S} \lambda_s v_s : \lambda_s \in F, v_s \in S, \text{only finitely many nonzero } \lambda_s \right\}$$

By convention, we specify

$$\langle \varnothing \rangle = \{0\}$$

so that all spans are subspaces.

Remark. $\langle S \rangle$ is the smallest vector subspace of V containing S.

Example. Let $V = \mathbb{R}^3$, and

$$S = \left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix} \right\}, \begin{pmatrix} 3 \\ -2 \\ -4 \end{pmatrix}$$

Then we can check that

$$\langle S \rangle = \left\{ \begin{pmatrix} a \\ b \\ 2b \end{pmatrix} : (a, b) \in \mathbb{R} \right\}$$

Example. Let $V = \mathbb{R}^n$. We define

$$e_i = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

where the 1 is in the ith position. Then $V = \left\langle (e_i)_{1 \leq i \leq n} \right\rangle$.

Example. Let X be a set, and $\mathbb{R}^X = \{f : X \to \mathbb{R}\}$. Then let $S_x : X \to \mathbb{R}$ be defined by

$$S_x(y) = \begin{cases} 1 & y = x \\ 0 & \text{otherwise} \end{cases}$$

Then, $\langle (S_x)_{x \in X} \rangle = \left\{ f \in \mathbb{R}^X : f \text{ has finite support} \right\}$, where the support of f is defined to be $\{x : f(x) \neq 0\}$.

1.6 Dimensionality

Definition. Let V be an F-vector space. Let $S \subset V$. We say that S spans V if $\langle S \rangle = V$. If S spans V, we say that S is a generating family of V.

Definition. Let V be an F-vector space. V is finite-dimensional if it is spanned by a finite set.

Example. Consider the set $V = \mathbb{P}[x]$ which is the set of polynomials on $\mathbb{R}$. Further, consider $V_n = \mathbb{P}_n[x]$ which is the subspace with degree less than or equal to n. Then V_n is spanned by $\{1, x, x^2, \dots, x^n\}$, so V_n is finite-dimensional. Conversely, V is infinite-dimensional; there is no finite set S such that $\langle S \rangle = V$.

1.7 Linear independence

Definition. We say that $v_1, \dots, v_n \in V$ are linearly independent if, for $\lambda_i \in F$,

$$\sum_{i=1}^{n} \lambda_i v_i = 0 \implies \forall i, \lambda_i = 0$$

Definition. Similarly, $v_1, \dots, v_n \in V$ are linearly dependent if

$$\exists \lambda \in F^n, \sum_{i=1}^{n} \lambda_i v_i = 0, \exists i, \lambda_i \neq 0$$

Equivalently, one of the vectors can be written as a linear combination of the remaining ones.

Remark. If $(v_i)_{1 \leq i \leq n}$ are linearly independent, then

$$\forall i \in \{1, \dots, n\}, v_i \neq 0$$

1.8 Bases

Definition. $S \subset V$ is a basis of V if

(i) $\langle S \rangle = V$

(ii) S is a linearly independent set

So, a basis is a linearly independent (also known as *free*) generating family.

Example. Let $V = \mathbb{R}^n$. The *canonical basis* (e_i) is a basis since we can show that they are free and span V.

Example. Let $V = \mathbb{C}$, considered as a $\mathbb{C}$-vector space. Then $\{1\}$ is a basis. If V is a $\mathbb{R}$-vector space, $\{1, i\}$ is a basis.

Example. Consider again $\mathbb{P}[x]$. Then $S = \{x^n : n \in \mathbb{N}\}$ is a basis of $\mathbb{P}$.

Lemma. Let V be an F-vector space. Then, $(v_1, \dots, v_n)$ is a basis of V if and only if any vector $v \in V$ has a unique decomposition

$$v = \sum_{i=1}^{n} \lambda_i v_i, \forall i, \lambda_i \in F$$

In the above definition, we call $(\lambda_1, \dots, \lambda_n)$ the *coordinates* of v in the basis $(v_1, \dots, v_n)$.

Proof. Suppose $(v_1, \dots, v_n)$ is a basis of V. Then $\forall v \in V$ there exists $\lambda_1, \dots, \lambda_n \in F$ such that

$$v = \sum_{i=1}^{n} \lambda_i v_i$$

So there exists a tuple of λ values. Suppose two such λ tuples exist. Then

$$v = \sum_{i=1}^{n} \lambda_i v_i = \sum_{i=1}^{n} \lambda_i' v_i \implies \sum_{i=1}^{n} (\lambda_i - \lambda_i') v_i = 0 \implies \lambda_i = \lambda_i'$$

The converse is left as an exercise. □

Lemma. If $\langle \{v_1, \dots, v_n\} \rangle = V$, then some subset of this set is a basis of V.

Proof. If $(v_1, \dots, v_n)$ are linearly independent, this is a basis. Otherwise, one of the vectors can be written as a linear combination of the others. So, up to reordering,

$$v_n \in \langle \{v_1, \dots, v_{n-1}\} \rangle = V$$

So we have removed a vector from this set and preserved the span. By induction, we will eventually reach a basis. □

1.9 Steinitz exchange lemma

Theorem. Let V be a finite dimensional F-vector space. Let $(v_1, \dots, v_m)$ be linearly independent, and $(w_1, \dots, w_n)$ which spans V. Then,
(i) $m \leq n$; and
(ii) up to reordering, $(v_1, \dots, v_m, w_{m+1}, \dots w_n)$ spans V.

Proof. Suppose that we have replaced $\ell \geq 0$ of the w_i.

$$\langle v_1, \dots, v_\ell, w_{\ell+1}, \dots w_n \rangle = V$$

If $m = \ell$, we are done. Otherwise, $\ell < m$. Then, $v_{\ell+1} \in V = \langle v_1, \dots, v_\ell, w_{\ell+1}, \dots w_n \rangle$ Hence $v_{\ell+1}$ can be expressed as a linear combination of the generating set. Since the $(v_i)_{1 \leq i \leq m}$ are linearly independent (free), one of the coefficients on the w_i are nonzero. In particular, up to reordering we can express $w_{\ell+1}$ as a linear combination of $v_1, \dots, v_{\ell+1}, w_{\ell+2}, \dots, w_n$. Inductively, we may replace m of the w terms with v terms. Since we have replaced m vectors, necessarily $m \leq n$. □

1.10 Consequences of Steinitz exchange lemma

Corollary. Let V be a finite-dimensional F-vector space. Then, any two bases of V have the same number of vectors. This number is called the dimension of V, $\dim_F V$.

Proof. Suppose the two bases are $(v_1, \dots, v_n)$ and $(w_1, \dots, w_m)$. Then, $(v_1, \dots, v_n)$ is free and $(w_1, \dots, w_m)$ is generating, so the Steinitz exchange lemma shows that $n \leq m$. Vice versa, $m \leq n$. Hence $m = n$. □

Corollary. Let V be an F-vector space with finite dimension n. Then,

(i) Any independent set of vectors has at most n elements, with equality if and only if it is a basis.

(ii) Any spanning set of vectors has at least n elements, with equality if and only if it is a basis.

Proof. Exercise. □

1.11 Dimensionality of sums

Proposition. Let V be an F-vector space. Let U, W be subspaces of V. If U, W are finite-dimensional, then so is $U + W$, with

$$\dim_F(U + W) = \dim_F U + \dim_F W - \dim_F(U \cap W)$$

Proof. Consider a basis $(v_1, \dots, v_n)$ of the intersection. Extend this basis to a basis

$$(v_1, \dots, v_n, u_1, \dots, u_m) \text{ of } U; \quad (v_1, \dots, v_n, w_1, \dots, w_k) \text{ of } W$$

Then, we will show that $(v_1, \dots, v_n, u_1, \dots, u_m, w_1, \dots, w_k)$ is a basis of $\dim_F(U + W)$, which will conclude the proof. Indeed, since any component of $U + W$ can be decomposed as a sum of some element of U and some element of W, we can add their decompositions together. Now we must show that this new basis is free.

$$\sum_{i=1}^{n} \alpha_i v_i + \sum_{i=1}^{m} \beta_i u_i + \sum_{i=1}^{k} \gamma_i w_i = 0$$
$$\underbrace{\sum_{i=1}^{n} \alpha_i v_i + \sum_{i=1}^{m} \beta_i u_i}_{\in U} = \underbrace{\sum_{i=1}^{k} \gamma_i w_i}_{\in W}$$
$$\sum_{i=1}^{k} \gamma_i w_i \in U \cap W$$
$$\sum_{i=1}^{k} \gamma_i w_i = \sum_{i=1}^{n} \delta_i v_i$$
$$\sum_{i=1}^{n} (\alpha_i + \delta_i) v_i + \sum_{i=1}^{m} \beta_i u_i = 0$$
$$\beta_i = 0, \alpha_i = -\delta_i$$
$$\sum_{i=1}^{n} \alpha_i v_i + \sum_{i=1}^{k} \gamma_i w_i = 0$$
$$\alpha_i = 0, \gamma_i = 0$$

□

Proposition. If V is a finite-dimensional F-vector space, and $U \leq V$, then U and V/U are also finite-dimensional. In particular, $\dim_F V = \dim_F U + \dim_F (V/U)$.

Proof. Let $(u_1, \dots, u_\ell)$ be a basis of U. We extend this basis to a basis of V, giving

$$(u_1, \dots, u_\ell, w_{\ell+1}, \dots, w_n)$$

We claim that $(w_{\ell+1} + U, \dots, w_n + U)$ is a basis of the vector space V/U. □

Remark. If V is an F-vector space, and $U \leq V$, then we say U is a proper subspace if $U \neq V$. Then if U is proper, then $\dim_F U < \dim_F V$ and $\dim_F (V/U) > 0$ because $(V/U) \neq \varnothing$.

1.12 Direct sums

Definition. Let V be an F-vector space and U, W be subspaces of V. We say that $V = U \oplus W$, read as the direct sum of U and W, if $\forall v \in V, \exists! u \in U, \exists! w \in W, u + w = v$. We say that W is *a* direct complement of U in V; there is no uniqueness of such a complement.

Lemma. Let V be an F-vector space, and $U, W \leq V$. Then the following statements are equivalent.

(i) $V = U \oplus W$

(ii) $V = U + W$ and $U \cap W = \{0\}$
(iii) For any basis B_1 of U and B_2 of W, $B_1 \cup B_2$ is a basis of V

Proof. First, we show that (ii) implies (i). If $V = U + W$, then certainly $\forall v \in V, \exists u \in U, \exists w \in W, v = u + w$, so it suffices to show uniqueness. Note, $u_1 + w_1 = u_2 + w_2 \implies u_1 - u_2 = w_2 - w_1$. The left hand side is an element of U and the right hand side is an element of W, so they must be the zero vector; $u_1 = u_2, w_1 = w_2$.

Now, we show (i) implies (iii). Suppose B_1 is a basis of U and B_2 is a basis of W. Let $B = B_1 \cup B_2$. First, note that B is a generating family of $U + W$. Now we must show that B is free.

$$\underbrace{\sum_{u \in B_1} \lambda_u u}_{\in U} + \underbrace{\sum_{w \in B_2} \lambda_w w}_{\in W} = 0$$

Hence both sums must be zero. Since B_1, B_2 are bases, all λ are zero, so B is free and hence a basis.

Now it remains to show that (iii) implies (ii). We must show that $V = U + W$ and $U \cap W = \{0\}$. Now, suppose $v \in V$. Then, $v = \sum_{u \in B_1} \lambda_u u + \sum w \in B_2 \lambda_w w$. In particular, $V = U + W$, since the λ_u, λ_w are arbitrary. Now, let $v \in U \cap W$. Then

$$v = \sum_{u \in B_1} \lambda_u u = \sum_{w \in B_2} \lambda_w w \implies \lambda_u = \lambda_w = 0$$

□

Definition. Let V be an F-vector space, with subspaces $V_1, \dots, V_p \leq V$. Then

$$\sum_{i=1}^{p} V_i = \{v_1, \dots, v_\ell, v_i \in V_i, 1 \leq i \leq \ell\}$$

We say the sum is direct, written

$$\bigoplus_{i=1}^{p} V_i$$

if the decomposition is unique. Equivalently,

$$V = \bigoplus_{i=1}^{p} V_i \iff \exists! v_1 \in V_1, \dots, v_n \in V_n, v = \sum_{i=1}^{n} v_i$$

Lemma. The following are equivalent:
(i) $\sum_{i=1}^{p} V_i = \bigoplus_{i=1}^{p} V_i$
(ii) $\forall 1 \leq i \leq l, V_i \cap \left(\sum_{j \neq i} V_j\right) = \{0\}$
(iii) For any basis B_i of V_i, $B = \bigcup_{i=1}^{n} B_i$ is a basis of $\sum_{i=1}^{n} V_i$.

Proof. Exercise. □

2 Linear maps

2.1 Linear maps

Definition. If V, W are F-vector spaces, a map $\alpha : V \to W$ is *linear* if

$$\forall \lambda_1, \lambda_2 \in F, \forall v_1, v_2 \in V, \alpha(\lambda_1 v_1 + \lambda_2 v_2) = \lambda_1 \alpha(v_1) + \lambda_2 \alpha(v_2)$$

Example. Let M be a matrix with n rows and m columns. Then the map $\alpha : \mathbb{R}^m \to \mathbb{R}^n$ defined by $x \mapsto Mx$ is a linear map.

Example. Let $\alpha : \mathcal{C}([0,1], \mathbb{R}) \to \mathcal{C}([0,1], \mathbb{R})$ defined by $f \mapsto a(f)(x) = \int_0^x f(t)\,\mathrm{d}t$. This is linear.

Example. Let $x \in [a,b]$. Then $\alpha : \mathcal{C}([a,b], \mathbb{R}) \to \mathbb{R}$ defined by $f \mapsto f(x)$ is a linear map.

Remark. Let U, V, W be F-vector spaces. Then,

(i) The identity function $i_V : V \to V$ defined by $x \mapsto x$ is linear.

(ii) If $\alpha : U \to V$ and $\beta : V \to W$ are linear, then $\beta \circ \alpha$ is linear.

Lemma. Let V, W be F-vector spaces. Let B be a basis for V. If $\alpha_0 : B \to V$ is *any* map (not necessarily linear), then there exists a unique linear map $\alpha : V \to W$ extending α_0: $\forall v \in B, \alpha_0(v) = \alpha(v)$.

Proof. Let $v \in V$. Then, given $B = (v_1, \dots, v_n)$.

$$v = \sum_{i=1}^n \lambda_i v_i$$

By linearity,

$$\alpha(v) = \alpha\left(\sum_{i=1}^n \lambda_i v_i\right) = \sum_{i=1}^n \alpha(\lambda_i v_i) = \sum_{i=1}^n \alpha_0(\lambda_i v_i)$$

□

Remark. This lemma is also true in infinite-dimensional vector spaces. Often, to define a linear map, we instead define its action on the basis vectors, and then we 'extend by linearity' to construct the entire map.

Remark. If $\alpha_1, \alpha_2 : V \to W$ are linear maps, then if they agree on any basis of V then they are equal.

2.2 Isomorphism

Definition. Let V, W be F-vector spaces. A map $\alpha : V \to W$ is an *isomorphism* if and only if

(i) α is linear

(ii) α is bijective

If such an α exists, we say that V and W are isomorphic, written $V \simeq W$.

Remark. If α in the above definition is an isomorphism, then $\alpha^{-1} : W \to V$ is linear. Indeed, if $w_1, w_2 \in W$ with $w_1 = \alpha(v_1)$ and $w_2 = \alpha(v_2)$,

$$\alpha^{-1}(w_1 + w_2) = \alpha^{-1}(\alpha(v_1) + \alpha(v_2)) = \alpha^{-1}\alpha(v_1 + v_2) = v_1 + v_2 = \alpha^{-1}(w_1) + \alpha^{-1}(w_2)$$

Similarly, for $\lambda \in F, w \in W$,

$$\alpha^{-1}(\lambda w) = \lambda\alpha^{-1}(w)$$

Lemma. Isomorphism is an equivalence relation on the class of all vector spaces over F.

Proof.

(i) $i_V : V \to V$ is an isomorphism

(ii) If $\alpha : V \to W$ is an isomorphism, $\alpha^{-1} : W \to V$ is an isomorphism.

(iii) If $\beta : U \to V, \alpha : V \to W$ are isomorphisms, then $\alpha \circ \beta : U \to W$ is an isomorphism.

The proofs of each part are left as an exercise. □

Theorem. If V is an F-vector space of dimension n, then $V \simeq F^n$.

Proof. Let $B = (v_1, \dots, v_n)$ be a basis for V. Then, consider $\alpha : V \to F^n$ defined by

$$v = \sum_{i=1}^{n} \lambda_i v_i \mapsto \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix}$$

We claim that this is an isomorphism. This is left as an exercise. □

Remark. Choosing a basis for V is analogous to choosing an isomorphism from V to F^n.

Theorem. Let V, W be F-vector spaces with finite dimensions n, m. Then,

$$V \simeq W \iff n = m$$

Proof. If $\dim V = \dim W = n$, then there exist isomorphisms from both V and W to F^n. By transitivity, therefore, there exists an isomorphism between V and W.

Conversely, if $V \simeq W$ then let $\alpha : V \to W$ be an isomorphism. Let B be a basis of V, then we claim that $\alpha(B)$ is a basis of W. Indeed, $\alpha(B)$ spans W from the surjectivity of α, and $\alpha(B)$ is free due to injectivity. □

2.3 Kernel and image

Definition. Let V, W be F-vector spaces. Let $\alpha : V \to W$ be a linear map. We define the kernel and image as follows.

$$N(\alpha) = \ker\alpha = \{v \in V : \alpha(v) = 0\}$$

$$\operatorname{Im}(\alpha) = \{w \in W : \exists v \in V, w = \alpha(v)\}$$

Lemma. $\ker \alpha$ is a subspace of V, and $\operatorname{Im} \alpha$ is a subspace of W.

Proof. Let $\lambda_1, \lambda_2 \in F$ and $v_1, v_2 \in \ker \alpha$. Then

$$\alpha(\lambda_1 v_1 + \lambda_2 v_2) = \lambda_1 \alpha(v_1) + \lambda_2 \alpha(v_2) = 0$$

Hence $\lambda_1 v_1 + \lambda_2 v_2 \in \ker \alpha$.

Now, let $\lambda_1, \lambda_2 \in F$, $v_1, v_2 \in V$, and $w_1 = \alpha(v_1), w_2 = \alpha(v_2)$. Then

$$\lambda_1 w_1 + \lambda_2 w_2 = \lambda_1 \alpha(v_1) + \lambda_2 \alpha(v_2) = \alpha(\lambda_1 v_1 + \lambda_2 v_2) \in \operatorname{Im} \alpha$$

□

Remark. $\alpha \colon V \to W$ is injective if and only if $\ker \alpha = \{0\}$. Further, $\alpha \colon V \to W$ is surjective if and only if $\operatorname{Im} \alpha = W$.

Theorem. Let V, W be F-vector spaces. Let $\alpha \colon V \to W$ be a linear map. Then $\overline{\alpha} \colon V / \ker \alpha \to \operatorname{Im} \alpha$ defined by

$$\overline{\alpha}(v + \ker \alpha) = \alpha(v)$$

is an isomorphism. *This is the isomorphism theorem from IA Groups.*

Proof. First, note that $\overline{\alpha}$ is well defined. Suppose $v + \ker \alpha = v' + \ker \alpha$. Then $v - v' \in \ker \alpha$, hence

$$\alpha(v - v') = 0 \implies \alpha(v) - \alpha(v') = 0$$

so $\overline{\alpha}$ is indeed well defined.

Now, we show $\overline{\alpha}$ is injective.

$$\overline{\alpha}(v + \ker \alpha) = 0 \implies \alpha(v) = 0 \implies v \in \ker \alpha$$

Hence, $v + \ker \alpha = 0 + \ker \alpha$.

Further, $\overline{\alpha}$ is surjective. This follows from the definition the image. □

2.4 Rank and nullity

Definition. The *rank* of α is

$$r(\alpha) = \dim \operatorname{Im} \alpha$$

The *nullity* of α is

$$n(\alpha) = \dim \ker \alpha$$

Theorem (Rank-nullity theorem)**.** Let U, V be F-vector spaces such that the dimension of U is finite. Let $\alpha \colon U \to V$ be a linear map. Then,

$$\dim U = r(\alpha) + n(\alpha)$$

Proof. We have proven that $U/\ker\alpha \simeq \operatorname{Im}\alpha$. Hence, the dimensions on the left and right match: $\dim(U/\ker\alpha) = \dim\operatorname{Im}\alpha$.

$$\dim U - \dim\ker\alpha = \dim\operatorname{Im}\alpha$$

and the result follows. □

Lemma (Characterisation of isomorphisms)**.** Let V, W be F-vector spaces with equal, finite dimension. Let $\alpha\colon V \to W$ be a linear map. Then, the following are equivalent.

(i) α is injective.
(ii) α is surjective.
(iii) α is an isomorphism.

Proof. Clearly, (iii) follows from (i) and (ii) and vice versa. The rest of the proof is left as an exercise, which follows from the rank-nullity theorem. □

2.5 Space of linear maps

Let V and W be F-vector spaces. Consider the space of linear maps from V to W. Then $L(V, W) = \{\alpha\colon V \to W \text{ linear}\}$.

Proposition. $L(V, W)$ is an F-vector space under the operation

$$(\alpha_1 + \alpha_2)(v) = \alpha_1(v) + \alpha_2(v);$$

$$(\lambda\alpha)(v) = \lambda(\alpha(v))$$

Further, if V and W are finite-dimensional, then so is $L(V, W)$ with

$$\dim_F L(V, W) = \dim_F V \dim_F W$$

Proof. Proving that $L(V, W)$ is a vector space is left as an exercise. The dimensionality part is proven later. □

2.6 Matrices

Definition. An $m \times n$ matrix over F is an array of m rows and n columns, with entries in F.

We write $M_{m\times n}(F)$ for the set of $m \times n$ matrices over F.

Proposition. $M_{m\times n}(F)$ is an F-vector space under

$$((a_{ij}) + (b_{ij})) = (a_{ij} + b_{ij});$$

$$\lambda(a_{ij}) = (\lambda a_{ij})$$

Proposition. $\dim_F M_{m,n}(F) = mn$.

Proof. Consider the basis defined by, the 'elementary matrix' for all i, j:

$$e_{pq} = \delta_{ip}\delta_{jq}$$

Then (e_{ij}) is a basis of $M_{m\times n}(F)$, since it spans $M_{m\times n}(F)$ and we can show that it is free. $\square$

2.7 Linear maps as matrices

Consider bases B of V and C of W:

$$B = (v_1, \dots, v_n); C = (w_1, \dots, w_n)$$

Then let $v \in V$. We have

$$v = \sum_{j=1}^{n} \lambda_j v_j \equiv [v]_B = \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix} \in F^n$$

where the vector given is the coordinates in basis B. We can equivalently find $[w]_C$, the coordinates of w in basis C. We can now define a matrix of some linear map α in the B, C basis.

Definition.

$$[\alpha]_{B,C} = \big([\alpha(v_1)]_C, \dots, [\alpha(v_n)]_C\big) \in M_{m\times n}(F)$$

Note that if $[\alpha]_{BC} = (a_{ij})$, then by definition

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

Lemma. For all $v \in V$,

$$[\alpha(v)]_C = [\alpha]_{BC} \cdot [v]_B$$

Proof. We have

$$v = \sum_{i=1}^{n} \lambda_j v_j$$

Hence

$$\alpha\left(\sum_{i=1}^{n} \lambda_j v_j\right) = \sum_{j=1}^{n} \lambda_j \alpha(v_j) = \sum_{j=1}^{n} \lambda_i \sum_{i=1}^{m} a_{ij} w_i = \sum_{i=1}^{m} \left(\sum_{j=1}^{n} a_{ij}\lambda_j\right) w_i$$

$\square$

Lemma. Let $\beta : U \to V$ and $\alpha : V \to W$ be linear maps. Then, if A, B, C are bases of U, V, W respectively, then

$$[\alpha \circ \beta]_{A,C} = [\alpha]_{B,C} \cdot [\beta]_{A,B}$$

Proof. Consider $u \in A$. Then
$$(\alpha \circ \beta)(u) = \alpha(\beta(u))$$
giving
$$\alpha\left(\sum_j b_{jp} v_i\right) = \sum_j b_{jp}\alpha(v_j) = \sum_j b_{jp} \sum_i a_{ij} w_i = \sum_i (\sum_j a_{ij} b_{jp}) w_i$$
where $a_{ij} p_{jp}$ is the (i, j) element of AB by the definition of the product of matrices. □

Proposition. If V, W are F-vector spaces, and $\dim V = n, \dim W = m$, then
$$L(V, W) \simeq M_{m\times n}(F)$$
which implies the dimensionality of $L(V, W)$ in F is $m \times n$.

Proof. Consider two bases B, C of V, W. We claim that
$$\theta : L(V, W) \to M_{m\times n}(F)$$
defined by $\theta(\alpha) = [\alpha]_{B,C}$. is an isomorphism. First, note that θ is linear. Then, θ is surjective; consider any matrix $A = (a_{ij})$ and consider $\alpha : v_j \mapsto \sum_{i=1}^m a_{ij} w_i$. Then this is certainly a linear map which extends uniquely by linearity to A, giving $[\alpha]_{B,C} = (a_{ij}) = A$. Now, θ is injective since $[\alpha]_{B,C} = 0 \implies \alpha = 0$. □

Remark. If B, C are bases of V, W respectively, and $\varepsilon_B : V \to F^n$ is defined by $v \mapsto [v]_B$, and analogously for ε_C, then
$$[\alpha]_{B,C} \circ \varepsilon_B = \varepsilon_C \circ \alpha$$
so the operations commute.

Example. Let $\alpha : V \to W$ be a linear map and $Y \le V$, where V, W are finite-dimensional. Then let $\alpha(Y) = Z \le W$. Consider a basis B of V, such that $B' = (v_1, \dots, v_k)$ is a basis of Y completed by $B'' = (v_{k+1}, \dots, v_n)$ into $B = B' \cup B''$. Then let C be a basis of W, such that $C' = (w_1, \dots, w_\ell)$ is a basis of Z completed by $C'' = (w_{\ell+1}, \dots, w_m)$ into $C = C' \cup C''$. Then
$$[\alpha]_{B,C} = \begin{pmatrix} \alpha(v_1) & \dots & \alpha(v_k) & \alpha(v_{k+1}) & \dots & \alpha(v_n) \end{pmatrix}$$
For $1 \le i \le k$, $\alpha(v_i) \in Z$ since $v_i \in Y, \alpha(Y) = Z$. So the matrix has an upper-left $\ell \times k$ block A which is $\alpha : Y \to Z$ on the basis B', C'. We can show further that α induces a map $\overline{\alpha} : V/Y \to W/Z$ by $v + Y \mapsto \alpha(v) + Z$. This is well-defined; $v_1 + Y = v_2 + Y$ implies $v_1 - v_2 \in Y$ hence $\alpha(v_1 - v_2) \in Z$ as required. The bottom-right block is $[\overline{\alpha}]_{B'',C''}$.

2.8 Change of basis

Suppose we have two bases $B = \{v_1, \dots, v_n\}, B' = \{v'_1, \dots, v'_n\}$ of V and corresponding C, C' for W. If we have a linear map $[\alpha]_{B,C}$, we are interested in finding the components of this linear map in another basis, that is,
$$[\alpha]_{B,C} \mapsto [\alpha]_{B',C'}$$

Definition. The *change of basis* matrix P from B' to B is

$$P = \begin{pmatrix} [v_1']_B & \cdots & [v_n']_B \end{pmatrix}$$

which is the identity map in B', written

$$P = [I]_{B',B}$$

Lemma. For a vector v,

$$[v]_B = P[v]_{B'}$$

Proof. We have

$$[\alpha(v)]_C = [\alpha]_{B,C} \cdot [v]_C$$

Since $P = [I]_{B',B}$,

$$[I(v)]_B = [I]_{B',B} \cdot [v]_{B'} \implies [v]_B = P[v]_{B'}$$

as required. □

Remark. P is an invertible $n \times n$ square matrix. In particular,

$$P^{-1} = [I]_{B,B'}$$

Indeed,

$$I_n = [I \cdot I]_{B,B} = [I]_{B',B} \cdot [I]_{B',B}$$

where I_n is the $n \times n$ identity matrix.

Proposition. If α is a linear map from V to W, and $P = [I]_{B',B}$, $Q = [I]_{C',C}$, we have

$$A' = [\alpha]_{B',C'} = [I]_{C,C'}[\alpha]_{B,C}[I]_{B,'B} = Q^{-1}AP$$

where $A = [\alpha]_{B,C}, A' = [\alpha]_{B',C'}$.

Proof.

$$\begin{aligned}
[\alpha(v)]_C &= Q[\alpha(v)]_{C'} \\
&= Q[\alpha]_{B',C'}[v]_{B'} \\
[\alpha(v)]_C &= [\alpha]_{B,C}[v]_B \\
&= AP[v]_{B'} \\
\therefore \forall v,\ QA[v]_{B'} &= AP[v]_{B'} \\
\therefore QA &= AP
\end{aligned}$$

as required. □

2.9 Equivalent matrices

Definition. Matrices A, A' are called *equivalent* if

$$A' = Q^{-1}AP$$

for some invertible $m \times m, n \times n$ matrices Q, P.

Remark. This defines an equivalence relation on $M_{m,n}(F)$.

- $A = I_m^{-1}AI_n$;
- $A' = Q^{-1}AP \implies A = QA'P^{-1}$;
- $A' = Q^{-1}AP, A'' = (Q')^{-1}A'P' \implies A'' = (QQ')^{-1}A(PP')$.

Proposition. Let $\alpha : V \to W$ be a linear map. Then there exists a basis B of V and a basis C of W such that

$$[\alpha]_{B,C} = \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}$$

so the components of the matrix are exactly the identity matrix of size r in the top-left corner, and zeroes everywhere else.

Proof. We first fix $r \in \mathbb{N}$ such that $\dim \ker \alpha = n - r$. Then we will construct a basis $\{v_{r+1}, \dots, v_n\}$ of the kernel. We extend this to a basis of the entirety of V, that is, $\{v_1, \dots, v_n\}$. Then, we want to show that

$$\{\alpha(v_1), \dots, \alpha(v_r)\}$$

is a basis of $\operatorname{Im}\alpha$. Indeed, it is a generating family:

$$\begin{aligned} v &= \sum_{i=1}^{n} \lambda_i v_i \\ \alpha(v) &= \sum_{i=1}^{n} \lambda_i \alpha(v_i) \\ &= \sum_{i=1}^{r} \lambda_i \alpha(v_i) \end{aligned}$$

Then if $y \in \operatorname{Im} \alpha$, there exists v such that $\alpha(v) = y$. Further, it is a free family:

$$\begin{aligned}
\sum_{i=1}^r \lambda_i \alpha(v_i) &= 0 \\
\alpha\left(\sum_{i=1}^r \lambda_i v_i\right) &= 0 \\
\sum_{i=1}^r \lambda_i v_i &\in \ker \alpha \\
\sum_{i=1}^r \lambda_i v_i &= \sum_{i=r+1}^n \lambda_i v_i \\
\sum_{i=1}^r \lambda_i v_i - \sum_{i=r+1}^n \lambda_i v_i &= 0
\end{aligned}$$

But since $\{v_1, \dots, v_n\}$ is a basis, $\lambda_i = 0$ for all i. Hence $\{\alpha(v_i)\}$ is a basis of $\operatorname{Im} \alpha$. Now, we wish to extend this basis to the whole of W to form

$$\{\alpha(v_1), \dots, \alpha(v_r), w_{r+1}, \dots, w_n\}$$

Now,

$$\begin{aligned}
[\alpha]_{BC} &= \begin{pmatrix} \alpha(v_1) & \cdots & \alpha(v_r) & \alpha(v_{r+1}) & \cdots & \alpha(v_n) \end{pmatrix} \\
&= \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}
\end{aligned}$$

□

Remark. This also proves the rank-nullity theorem:

$$\operatorname{rank} \alpha + \operatorname{null} \alpha = n$$

Corollary. Any $m \times n$ matrix A is equivalent to a matrix of the form

$$\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}$$

where $r = \operatorname{rank} A$.

2.10 Column rank and row rank

Definition. Let $A \in M_{m,n}(F)$. Then, the *column rank* of A, here denoted $r_c(A)$, is the dimension of the subspace of F^n spanned by the column vectors.

$$r_c(A) = \dim \operatorname{span} \{c_1, \dots, c_n\}$$

Remark. If α is a linear map, represented in bases B, C by the matrix A, then

$$\operatorname{rank} \alpha = r_c(A)$$

Proposition. Two matrices are equivalent if they have the same column rank:

$$r_c(A) = r_c(A')$$

Proof. If the matrices are equivalent, then $A = [\alpha]_{BC}, A' = [\alpha]_{B',C'}$. Then

$$r_c(A) = r_c(\alpha) = r_c(A')$$

Conversely, if $r_c(A) = r_c(A') = r$, then A, A' are equivalent to

$$\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}$$

By transitivity, A, A' are equivalent. □

Theorem. Column rank $r_c(A)$ and row rank $r_c(A^\intercal)$ are equivalent.

Proof. Let $r = r_C(A)$. Then,

$$Q^{-1}AP = \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}_{m \times n}$$

Then, consider

$$P^\intercal A^\intercal \left(Q^{-1}\right)^\intercal = (Q^{-1}AP)^\intercal = \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}_{m \times n}^\intercal = \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}_{n \times m}$$

Note that we can swap the transpose and inverse on Q because

$$\begin{aligned} (AB)^\intercal &= B^\intercal A^\intercal \\ \left(QQ^{-1}\right)^\intercal &= Q^\intercal \left(Q^{-1}\right)_\intercal \\ I &= Q^\intercal \left(Q^{-1}\right)_\intercal \\ \left(Q^\intercal\right)^{-1} &= \left(Q^{-1}\right)_\intercal \end{aligned}$$

Then $r_c(A) = \operatorname{rank}(A) = \operatorname{rank}(A^\intercal) = r_c(A^\intercal)$. □

So we can drop the concepts of column and row rank, and just talk about rank as a whole.

2.11 Conjugation and similarity

Consider the following special case of changing basis. If $\alpha : V \to V$ is linear, α is called an *endomorphism*. If $B = C, B' = C'$ then the special case of the change of basis formula is

$$[\alpha]_{B',B'} = P^{-1}[\alpha]_{B,B}P$$

Then, we say square matrices A, A' are *similar* or *conjugate* if there exists P such that $A' = P^{-1}AP$.

2.12 Elementary operations

Definition. An *elementary column operation* is

(i) swap columns i, j

(ii) replace column i by λ multiplied by the column

(iii) add λ multiplied by column i to column j

We define analogously the elementary row operations. Note that these elementary operations are invertible (for $\lambda \neq 0$). These operations can be realised through the action of elementary matrices. For instance, the column swap operation can be realised using

$$T_{ij} = \begin{pmatrix} I_n & 0 & 0 \\ 0 & A & 0 \\ 0 & 0 & I_m \end{pmatrix}; \quad A = \begin{pmatrix} 0 & 0 & 1 \\ 0 & I_k & 0 \\ 1 & 0 & 1 \end{pmatrix}$$

To multiply a column by λ,

$$n_{i,\lambda} = \begin{pmatrix} I_n & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & I_m \end{pmatrix}$$

To add a multiple of a column,

$$c_{ij,\lambda} = I + \lambda E_{ij}$$

where E_{ij} is the matrix defined by elements $(e_{ij})_{pq} = \delta_{ip}\delta_{jq}$. An elementary column (or row) operation can be performed by multiplying A by the corresponding elementary matrix from the right (on the left for row operations). This will essentially provide a constructive proof that any $n \times n$ matrix is equivalent to

$$\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}$$

We will start with a matrix A. If all entries are zero, we are done. So we will pick $a_{ij} = \lambda \neq 0$, and swap rows $i, 1$ and columns $j, 0$. This ensures that $a_{11} = \lambda \neq 0$. Now we multiply column 1 by $\frac{1}{\lambda}$. Finally, we can clear out row 1 and column 1 by subtracting multiples of the first row or column. Then we can perform similar operations on the $(n-1)\times(n-1)$ matrix in the bottom right block and inductively finish this process.

2.13 Gauss' pivot algorithm

If only row operations are used, we can reach the 'row echelon' form of the matrix, a specific case of an upper triangular matrix. On each row, there are a number of zeroes until there is a one, called the pivot. First, we assume that $a_{ij} \neq 0$. We swap rows $i, 1$. Then divide the first row by $\lambda = a_{i1}$ to get a one in the top left. We can use this one to clear the rest of the first column. Then, we can repeat on the next column, and iterate. This is a technique for solving a linear system of equations.

2.14 Representation of square invertible matrices

Lemma. If A is an $n \times n$ square invertible matrix, then we can obtain I_n using only row elementary operations, or only column elementary operations.

Proof. We show an algorithm that constructs this I_n. This is exactly going to invert the matrix, since the resultant operations can be combined to get the inverse matrix. We will show here the proof for column operations. We argue by induction on the number of rows. Suppose we can make the form

$$\begin{pmatrix} I_k & 0 \\ A & B \end{pmatrix}$$

We want to obtain the same structure with $k+1$ rows. We claim that there exists $j > k$ such that $a_{k+1,j} \neq 0$. Indeed, otherwise we can show that the vector

$$\begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix} = \delta_{k+1,i}$$

is not in the span of the column vectors of A. This contradicts the invertibility of the matrix. Now, we will swap columns $k+1, j$ and divide this column by λ. We can now use this 1 to clear the rest of the $k+1$ row.

Inductively, we have found $AE_1 \dots E_n = I_n$ where E_n are elementary. Thus, we can find A^{-1}. □

Proposition. Any invertible square matrix is a product of elementary matrices.

The proof is exactly the proof of the lemma above.

3 Dual spaces

3.1 Dual spaces

Definition. Let V be an F-vector space. Then $V^\star$ is the *dual* of V, defined by

$$V^\star = L(V, F) = \{\alpha \colon V \to F\}$$

where the α are linear. If $\alpha \colon V \to F$ is linear, then we say α is a linear form. So the dual of V is the set of linear forms on V.

Example. For instance, the trace $\operatorname{tr} \colon M_{n,n}(F) \to F$ is a linear form on $M_{n,n}(F)$.

Example. Consider functions $[0,1] \to \mathbb{R}$. We can define $T_f \colon \mathcal{C}^\infty([0,1], \mathbb{R}) \to \mathbb{R}$ such that $\phi \mapsto \int_0^1 f(x)\phi(x)\,\mathrm{d}x$. Then T_f is a linear form on $\mathcal{C}^\infty([0,1], \mathbb{R})$. We can then reconstruct f given T_f. This mathematical formulation is called distribution.

Lemma. Let V be an F-vector space with a finite basis $B = \{e_1, \dots, e_n\}$. Then there exists a basis $B^\star$ for $V^\star$ given by

$$B^\star = \{\varepsilon_1, \dots, \varepsilon_n\}; \quad \varepsilon_j\left(\sum_{i=1}^n a_i e_i\right) = a_j$$

We call $B^\star$ the *dual basis* for B.

Proof. We know

$$\varepsilon_j\left(\sum_{i=1}^n a_i e_i\right) = a_j$$

Equivalently,

$$\varepsilon_j(e_i) = \delta_{ij}$$

First, we will show that the set of linear forms as defined is free. For all i,

$$\begin{aligned} \sum_{j=1}^n \lambda_j \varepsilon_j &= 0 \\ \therefore \left(\sum_{j=1}^n \lambda_j \varepsilon_j\right) e_i &= 0 \\ \sum_{j=1}^n \lambda_j \varepsilon_j(e_i) &= 0 \\ \lambda_i &= 0 \end{aligned}$$

Now we show that the set spans $V^\star$. Suppose $\alpha \in V^\star$, $x \in V$.

$$\begin{aligned} \alpha(x) &= \alpha\left(\sum_{j=1}^n \lambda_j e_j\right) \\ &= \sum_{i=1}^n \lambda_j \alpha(e_j) \end{aligned}$$

Conversely, we can write

$$\sum_{i=1}^n \alpha(e_j)\varepsilon(j) \in V^\star$$

Thus,

$$\begin{aligned} \left(\sum_{i=1}^n \alpha(e_j)\varepsilon_j\right)(x) &= \sum_{j=1}^n \alpha(e_j)\varepsilon_j\left(\sum_{k=1}^n \lambda_k e_k\right) \\ &= \sum_{j=1}^n \alpha(e_j) \sum_{k=1}^n \lambda_k \varepsilon_j(e_k) \\ &= \sum_{j=1}^n \alpha(e_j) \sum_{k=1}^n \lambda_k \delta_{jk} \\ &= \sum_{j=1}^n \alpha(e_j)\lambda_j \\ &= \alpha(x) \end{aligned}$$

We have then shown that

$$\alpha = \sum_{j=1}^{n} \alpha(e_j)\varepsilon_j$$

as required. □

Corollary. If V is finite-dimensional, $V^\star$ has the same dimension.

Remark. It is sometimes convenient to think of $V^\star$ as the spaces of row vectors of length $\dim V$ over F. For instance, consider the basis $B = (e_1, \dots, e_n)$, so $x = \sum_{i=1}^n x_i e_i$. Then we can pick $(\varepsilon_1, \dots, \varepsilon_n)$ a basis of $V^\star$, so $\alpha = \sum_{i=1}^n \alpha_i \varepsilon_i$. Then

$$\alpha(x) = \sum_{i=1}^{n} \alpha_i \varepsilon_i(x) = \sum_{i=1}^{n} \alpha_i \varepsilon \left(\sum_{j=1}^{n} x_j e_j \right) = \sum_{i=1}^{n} \alpha_i x_i$$

This is exactly

$$\begin{pmatrix} \alpha_1 & \cdots & \alpha_n \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

which essentially defines a scalar product between the two spaces.

3.2 Annihilators

Definition. Let $U \subseteq V$. Then the annihilator of U is

$$U^0 = \{\alpha \in V^\star : \forall u \in U, \alpha(u) = 0\}$$

Lemma. (i) $U^0 \leq V^\star$;
(ii) If $U \leq V$ and $\dim V < \infty$, then $\dim V = \dim U + \dim U^0$.

Proof. (i) First, note that $0 \in U^0$ since $\alpha(0) = 0$ by linearity. If $\alpha, \alpha' \in U^0$, then for all $u \in U$,

$$(\alpha + \alpha')(u) = \alpha(u) + \alpha'(u) = 0$$

Further, for all $\lambda \in F$,

$$(\lambda\alpha)(u) = \lambda\alpha(u) = 0$$

Hence $U^0 \leq V^\star$.

(ii) Let $(e_1, \dots, e_k)$ be a basis of U, completed into a basis $B = (e_1, \dots, e_k, e_{k+1}, \dots, e_n)$ of V. Let $(\varepsilon_1, \dots, \varepsilon_n)$ be the dual basis $B^\star$. We then will prove that

$$U^0 = \langle \varepsilon_{k+1}, \dots, \varepsilon_n \rangle$$

If $i > k$, then $\varepsilon_i(e_k) = \delta_{ik} = 0$. Hence $\varepsilon_i \in U^0$. Thus $\langle \varepsilon_{k+1}, \dots, \varepsilon_n \rangle \subset U^0$. Conversely, let $\alpha \in U^0$. Then $\alpha = \sum_{i=1}^n \alpha_i \varepsilon_i$. For $i \leq k$, $\alpha \in U^0$ hence $\alpha(e_i) = 0$. Hence,

$$\alpha = \sum_{i=k+1}^{n} \alpha_i \varepsilon_i$$

Thus
$$\alpha \in \langle \varepsilon_{k+1}, \dots, \varepsilon_n \rangle$$
as required.

□

3.3 Dual maps

Lemma. Let V, W be F-vector spaces. Let $\alpha \in L(V, W)$. Then there exists a unique $\alpha^\star \in L(W^\star, V^\star)$ such that
$$\varepsilon \mapsto \varepsilon \circ \alpha$$
called the dual map.

Proof. First, note $\varepsilon(\alpha) : V \to F$ is a linear map. Hence, $\varepsilon \circ \alpha \in V^\star$. Now we must show $\alpha^\star$ is linear.
$$\alpha^\star(\theta_1 + \theta_2) = (\theta_1 + \theta_2)(\alpha) = \theta_1 \circ \alpha + \theta_2 \circ \alpha = \alpha^\star(\theta_1) + \alpha^\star(\theta_2)$$
Similarly, we can show
$$\alpha^\star(\lambda\theta) = \lambda\alpha^\star(\theta)$$
as required. Hence $\alpha^\star \in L(W^\star, V^\star)$. □

Proposition. Let V, W be finite-dimensional F-vector spaces with bases B, C respectively. Then
$$[\alpha^\star]_{C^\star, B^\star} = [\alpha]_{B,C}^\intercal$$
Thus, we can think of the dual map as the *adjoint* of α.

Proof. This follows from the definition of the dual map. Let $B = (b_1, \dots, b_n)$, $C = (c_1, \dots, c_m)$, $B^\star = (\beta_1, \dots, \beta_n)$, $C^\star = (\gamma_1, \dots, \gamma_m)$. Let $[\alpha]_{B,C} = (a_{ij})$. Then, we compute
$$\begin{aligned}
\alpha^\star(\gamma_r)(b_s) &= \gamma_r \circ \alpha(b_s) \\
&= \gamma_r\left(\sum_t a_{ts} c_t\right) \\
&= \sum_t a_{ts}\gamma_r(c_t) \\
&= \sum_t a_{ts}\delta_{tr} \\
&= a_{rs}
\end{aligned}$$

We can conversely write $[\alpha^\star]_{C^\star,B^\star} = (m_{ij})$ and

$$\begin{aligned}
\alpha^\star(\gamma_r) &= \sum_{i=1}^n m_{ir}\beta_i \\
\alpha^\star(\gamma_r)(b_s) &= \sum_{i=1}^n m_{ir}\beta_i(b_s) \\
&= \sum_{i=1}^n m_{ir}\delta_{is} \\
&= m_{sr}
\end{aligned}$$

Thus,

$$a_{rs} = m_{sr}$$

as required. □

3.4 Properties of dual map

Let $\alpha \in L(V, W)$, and $\alpha^\star \in L(W^\star, V^\star)$. Let B and C be bases of V, W respectively, and $B^\star, C^\star$ be their duals. We have proven that

$$[\alpha]_{B,C} = [\alpha^\star]^\intercal_{B,C}$$

Lemma. Suppose that $E = (e_1, \dots, e_n)$ and $F = (f_1, \dots, f_n)$ are bases of V. Let $P = [I]_{F,E}$ be a change of basis matrix from F to E. The bases $E^\star = (\varepsilon_1, \dots, \varepsilon_n)$, $F^\star = (\eta_1, \dots, \eta_n)$ are the corresponding dual bases. Then, the change of basis matrix from $F^\star$ to $E^\star$ is

$$\left(P^{-1}\right)^\intercal$$

Proof. Consider

$$[I]_{F^\star,E^\star} = [I]^\intercal_{E,F} = \left([I]^{-1}_{F,E}\right)^\intercal = \left(P^{-1}\right)^\intercal$$

□

Lemma. Let V, W be F-vector spaces. Let $\alpha \in L(V, W)$. Let $\alpha^\star$ be the corresponding dual map. Then, denoting $N(\alpha)$ for the kernel of α,

(i) $N(\alpha^\star) = (\operatorname{Im}\alpha)^0$, so $\alpha^\star$ is injective if and only if α is surjective.

(ii) $\operatorname{Im}\alpha^\star \leq (N(\alpha))^0$, with equality if V, W are finite-dimensional. In this finite-dimensional case, $\alpha^\star$ is surjective if and only if α is injective.

Remark. In many applications, it is often simpler to understand the dual map $\alpha^\star$ than it is to understand α.

Proof. First, we prove (i). Let $\varepsilon \in W^\star$. Then, $\varepsilon \in N(\alpha^\star)$ means $\alpha^\star(\varepsilon) = 0$. Hence, $\alpha^\star(\varepsilon) = \varepsilon \circ \alpha = 0$ So for any $v \in V$, $\varepsilon(\alpha(v)) = 0$. Equivalently, ε is an element of the annihilator of $\operatorname{Im}\alpha$.

Now, we will show (ii). Let $\varepsilon \in \operatorname{Im}\alpha^\star$. Then $\alpha^\star(\phi) = \varepsilon$ for some $\phi \in W^\star$. Then, for all $u \in N(\alpha)$, $\varepsilon(u) = (\alpha^\star(\phi))(u) = \phi \circ \alpha(u) = \phi(\alpha(u)) = 0$. Certainly then $\varepsilon \in (N(\alpha))^0$. Then, $\operatorname{Im}\alpha^\star \leq (N(\alpha))^0$.

In the finite-dimensional case, we can compare the dimension of these two spaces.

$$\dim \operatorname{Im} \alpha^\star = r(\alpha^\star) = r\big([\alpha^\star]_{C^\star,B^\star}\big) = r\big([\alpha]_{B,C}^\intercal\big) = r\big([\alpha]_{B,C}\big) = r(\alpha) = \dim \operatorname{Im} \alpha$$

Due to the rank-nullity theorem, $\dim \operatorname{Im} \alpha^\star = \dim V - \dim N(\alpha) = \dim \left[(N(\alpha))^0\right]$. Hence,

$$\operatorname{Im} \alpha^\star \leq (N(\alpha))^0; \quad \dim \operatorname{Im} \alpha^\star = \dim(N(\alpha))^0$$

The dimensions are equal, and one is a subspace of the other, hence the spaces are equal. □

3.5 Double duals

Definition. Let V be an F-vector space. Let $V^\star$ be the dual of V. The *double dual* or bidual of V is

$$V^{\star\star} = L(V^\star, F) = (V^\star)^\star$$

Remark. In general, there is no obvious relation between V and $V^\star$. However, the following useful facts hold about V and $V^{\star\star}$.

(i) There is a *canonical embedding* from V to $V^{\star\star}$. In particular, there exists i in $L(V, V^{\star\star})$ which is injective.

(ii) There are examples of infinite-dimensional spaces where $V \simeq V^{\star\star}$. These are called reflexive spaces. Such spaces are investigated in the study of Banach spaces.

Theorem. V embeds into $V^{\star\star}$.

Proof. Choose a vector $v \in V$ and define the linear form $\hat{v} \in L(V^\star, F)$ such that

$$\hat{v}(\varepsilon) = \varepsilon(v)$$

So clearly $\hat{v}$ is linear. We want to show $\hat{v} \in V^{\star\star}$. If $\varepsilon \in V^\star, \varepsilon(v) \in F$. Further, $\lambda_1, \lambda_2 \in F$ and $\varepsilon_1, \varepsilon_2 \in V^\star$ give

$$\hat{v}(\lambda_1\varepsilon_1 + \lambda_2\varepsilon_2) = (\lambda_1\varepsilon_1 + \lambda_2\varepsilon_2)(v) = \lambda_1\varepsilon_1(v) + \lambda_2\varepsilon_2(v) = \lambda_1\hat{v}(\varepsilon_1) + \lambda_2\hat{v}(\varepsilon_2)$$

□

Theorem. If V is finite-dimensional, then $i : V \to V^{\star\star}$ given by $i(v) = \hat{v}$ is an isomorphism.

Proof. We will show i is linear. If $v_1, v_2 \in V, \lambda_1, \lambda_2 \in F$, then

$$i(\lambda_1 v_1 + \lambda_2 v_2)(\varepsilon) = \varepsilon(\lambda_1 v_1 + \lambda_2 v_2) = \lambda_1\varepsilon(v_1) + \lambda_2\varepsilon(v_2) = \lambda_1\hat{v}_1(\varepsilon) + \lambda_2\hat{v}_2(\varepsilon)$$

Now, we will show that i is injective for finite-dimensional V. Let $e \in V \setminus \{0\}$. We will show that $e \notin \ker i$. We extend e into a basis $(e, e_2, \dots, e_n)$ of V. Now, let $(\varepsilon, \varepsilon_2, \dots, \varepsilon_n)$ be the dual basis. Then $\hat{e}(\varepsilon) = \varepsilon(e) = 1$. In particular, $\hat{e} \neq 0$. Hence $\ker i = \{0\}$, so it is injective.

We now show that i is an isomorphism. We need to simply compute the dimension of the image under i. Certainly, $\dim V = \dim V^\star = \dim(V^\star)^\star = \dim V^{\star\star}$. Since i is injective, $\dim V = \dim V^{\star\star}$. So i is surjective as required. □

Lemma. Let V be a finite-dimensional F-vector space. Let $U \leq V$. Then,

$$\hat{U} = U^{00}$$

After identifying V and $V^{\star\star}$, we typically say

$$U = U^{00}$$

although this is is incorrect notation and not an equality.

Proof. We will show that $\hat{U} \leq U^{00}$. Indeed, let $u \in U$, then by definition

$$\forall \varepsilon \in U^0, \varepsilon(u) = 0 \implies \hat{u}(\varepsilon) = 0$$

Hence $\hat{u} \in U^{00}$ and so $\hat{U} \leq U^{00}$.

Now, we will compute dimension: $\dim U^{00} = \dim V - \dim U^0 = \dim U$. Since $\hat{U} \simeq U$, their dimensions are the same, so $U^{00} = \hat{U}$. □

Remark. Due to this identification of $V^{\star\star}$ and V, we can define

$$T \leq V^\star, T^0 = \{v \in V : \forall \theta \in T, \theta(v) = 0\}$$

Lemma. Let V be a finite-dimensional F-vector space. Let U_1, U_2 be subspaces of V. Then
(i) $(U_1 + U_2)^0 = U_1^0 \cap U_2^0$;
(ii) $(U_1 \cap U_2)^0 = U_1^0 + U_2^0$

Proof. Let $\theta \in V^\star$. Then $\theta \in (U_1 + U_2)^0 \iff \forall u_1 \in U_1, u_2 \in U_2, \theta(u_1 + u_2) = 0$. Hence $\theta(u) = 0$ for all $u \in U_1 \cup U_2$ by linearity. Hence $\theta \in U_1^0 \cap U_2^0$. Now, take the annihilator of (i) and $U^{00} = U$ to complete part (ii). □

4 Bilinear forms

4.1 Introduction

Definition. Let U, V be F-vector spaces. Then $\phi : U \times V \to F$ is a *bilinear form* if it is linear in both components. For example, ϕ at a fixed $u \in U$ is a linear form $V \to F$ and an element of $V^\star$.

Example. Consider the map $V \times V^\star \to F$ given by

$$(v, \theta) \mapsto \theta(v)$$

Example. The scalar product on $U = V = \mathbb{R}^n$ is given by

$$\psi(x, y) = \sum_{i=1}^{n} x_i y_i$$

Example. Let $U = V = C([0,1], \mathbb{R})$ and consider

$$\phi(f,g) = \int_0^1 f(t)g(t)\,\mathrm{d}t$$

Definition. If $B = (e_1, \dots, e_m)$ is a basis of U and $C = (f_1, \dots, f_n)$ is a basis of V, and $\phi : U \times V \to F$ is a bilinear form, then the matrix of the bilinear form in this basis is

$$[\phi]_{B,C} = \big(\phi(e_i, f_j)\big)_{1 \le i \le m, 1 \le j \le n}$$

Lemma. We can link ϕ with its matrix in a given basis as follows.

$$\phi(u,v) = [u]_B^\intercal [\phi]_{B,C} [v]_C$$

Proof. Let $u = \sum_{i=1}^m \lambda_i u_i$ and $v = \sum_{j=1}^n \mu_j v_j$. Then

$$\phi(u,v) = \phi\left(\sum_{i=1}^m \lambda_i u_i, \sum_{j=1}^n \mu_j v_j\right) = \sum_{i=1}^m \sum_{j=1}^n \lambda_i \mu_j \phi(u_i, v_j) = [u]_B^\intercal [\phi]_{B,C} [v]_C$$

□

Remark. Note that $[\phi]_{B,C}$ is the only matrix such that $\phi(u,v) = [u]_B^\intercal [\phi]_{B,C} [v]_C$.

Definition. Let $\phi : U \times V \to F$ be a bilinear form. Then ϕ induces two linear maps given by the partial application of a single parameter to the function.

$$\phi_L : U \to V^\star; \quad \phi_L(u) : V \to F; \quad v \mapsto \phi(u,v)$$

$$\phi_R : V \to U^\star; \quad \phi_R(v) : U \to F; \quad u \mapsto \phi(u,v)$$

In particular,

$$\phi_L(u)(v) = \phi(u,v) = \phi_R(v)(u)$$

Lemma. Let $B = (e_1, \dots, e_m)$ be a basis of U, and let $B^\star = (\varepsilon_1, \dots, \varepsilon_m)$ be its dual; and let $C = (f_1, \dots, f_n)$ be a basis of V, and let $C^\star = (\eta_1, \dots, \eta_n)$ be its dual. Let $A = [\phi]_{B,C}$. Then

$$[\phi_R]_{C,B^\star} = A; \quad [\phi_L]_{B,C^\star} = A^\intercal$$

Proof.

$$\phi_L(e_i)(f_j) = \phi(e_i, f_j) = A_{ij}$$

Since η_j is the dual of f_j,

$$\phi_L(e_i) = \sum_i A_{ij} \eta_j$$

Further,

$$\phi_R(f_j)(e_i) = \phi(e_i, f_j) = A_{ij}$$

and then similarly

$$\phi_R(f_j) = \sum_i A_{ij}\varepsilon_i$$

$\square$

Definition. $\ker \phi_L$ is called the *left kernel* of ϕ. $\ker \phi_R$ is the *right kernel* of ϕ.

Definition. We say that ϕ is *non-degenerate* if $\ker \phi_L = \ker \phi_R = \{0\}$. Otherwise, ϕ is *degenerate*.

Theorem. Let B be a basis of U, and let C be a basis of V, where U, V are finite-dimensional. Let $\phi : U \times V \to F$ be a bilinear form. Let $A = [\phi]_{B,C}$. Then, ϕ is non-degenerate if and only if A is invertible.

Corollary. If ϕ is non-degenerate, then $\dim U = \dim V$.

Proof. Suppose ϕ is non-degenerate. Then $\ker \phi_L = \ker \phi_R = \{0\}$. This is equivalent to saying that $n(\phi_L) = n(\phi_R) = 0$. We can use the rank-nullity theorem to state that $r(A^\intercal) = \dim V$ and $r(A) = \dim V$. This is equivalent to saying that A is invertible. Note that this forces $\dim U = \dim V$. $\square$

Remark. The canonical example of a non-degenerate bilinear form is the scalar product $\mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ represented by the identity matrix in the standard basis.

Corollary. If U and V are finite-dimensional with $\dim U = \dim V$, then choosing a non-degenerate bilinear form $\phi : U \times V \to F$ is equivalent to choosing an isomorphism $\phi_L : U \simeq V^\star$.

Definition. If $T \subset U$, then we define

$$T^\perp = \{v \in V : \forall t \in T, \phi(t, v) = 0\}$$

Further, if $S \subset V$, we define

$$^\perp S = \{u \in U : \forall s \in S, \phi(u, s) = 0\}$$

These are called the *orthogonals* of T and S.

4.2 Change of basis for bilinear forms

Proposition. Let B, B' be bases of U and $P = [I]_{B',B}$, let C, C' be bases of V and $Q = [I]_{C',C}$, and finally let $\phi : U \times V \to F$ be a bilinear form. Then

$$[\phi]_{B',C'} = P^\intercal [\phi]_{B,C} Q$$

Proof. We have $\phi(u, v) = [u]_B^\intercal [\phi]_{B,C} [v]_C$. Changing coordinates, we have

$$\phi(u, v) = (P[u]_{B'})^\intercal [\phi]_{B,C} (Q[v]_{C'}) = [u]_{B'}^\intercal (P^\intercal [\phi]_{B,C} Q)[v]_{C'}$$

□

Lemma. The *rank* of a bilinear form ϕ, denoted $r(\phi)$ is the rank of any matrix representing ϕ. This quantity is well-defined.

Remark. $r(\phi) = r(\phi_R) = r(\phi_L)$, since $r(A) = r(A^\intercal)$.

Proof. For any invertible matrices P, Q, $r(P^\intercal A Q) = r(A)$. □

5 Trace and determinant

5.1 Trace

Definition. The *trace* of a square matrix $A \in M_{n,n}(F) \equiv M_n(F)$ is defined by

$$\operatorname{tr} A = \sum_{i=1}^{n} a_{ii}$$

The trace is a linear form.

Lemma. $\operatorname{tr}(AB) = \operatorname{tr}(BA)$ for any matrices $A, B \in M_n(F)$.

Proof. We have

$$\operatorname{tr}(AB) = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} b_{ji} = \sum_{j=1}^{n} \sum_{i=1}^{n} b_{ji} a_{ij} = \operatorname{tr}(BA)$$

□

Corollary. Similar matrices have the same trace.

Proof.

$$\operatorname{tr}\big(P^{-1} A P\big) = \operatorname{tr}\big(A P^{-1} P\big) = \operatorname{tr} A$$

□

Definition. If $\alpha : V \to V$ is linear, we can define the trace of α as

$$\operatorname{tr} \alpha = \operatorname{tr}[\alpha]_B$$

for any basis B. This is well-defined by the corollary above.

Lemma. If $\alpha : V \to V$ is linear, $\alpha^\star : V^\star \to V^\star$ satisfies

$$\operatorname{tr} \alpha = \operatorname{tr} \alpha^\star$$

Proof.

$$\operatorname{tr} \alpha = \operatorname{tr}[\alpha]_B = \operatorname{tr}[\alpha]_B^\intercal = \operatorname{tr}[\alpha^\star]_{B^\star} = \operatorname{tr} \alpha^\star$$

□

5.2 Permutations and transpositions

Recall the following facts about permutations and transpositions. S_n is the group of permutations of the set $\{1, \dots, n\}$; the group of bijections $\sigma : \{1, \dots, n\} \to \{1, \dots, n\}$. A transposition $\tau_{k\ell} = (k, \ell)$ is defined by $k \mapsto \ell, \ell \mapsto k, x \mapsto x$ for $x \neq k, \ell$. Any permutation σ can be decomposed as a product of transpositions. This decomposition is not necessarily unique, but the parity of the number of transpositions is well-defined. We say that the signature of a permutation, denoted $\varepsilon : S_n \to \{-1, 1\}$, is 1 if the decomposition has even parity and -1 if it has odd parity. We can then show that ε is a homomorphism.

5.3 Determinant

Definition. Let $A \in M_n(F)$. We define

$$\det A = \sum_{\sigma \in S_n} \varepsilon(\sigma) A_{\sigma(1)1} \dots A_{\sigma(n)n}$$

Example. Let $n = 2$. Then,

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \implies \det A = a_{11}a_{22} - a_{12}a_{21}$$

Lemma. If $A = (a_{ij})$ is an upper (or lower) triangular matrix (with zeroes on the diagonal), then $\det A = 0$.

Proof. Let $(a_{ij}) = 0$ for $i > j$. Then

$$\det A = \sum_{\sigma \in S_n} \varepsilon(\sigma) a_{\sigma(1)1} \dots a_{\sigma(n)n}$$

For the summand to be nonzero, $\sigma(j) \leq j$ for all j. Thus,

$$\det A = a_{11} \dots a_{nn} = 0$$

□

Lemma. Let $A \in M_n(F)$. Then, $\det A = \det A^\intercal$.

Proof.

$$\begin{aligned}
\det A &= \sum_{\sigma \in S_n} \varepsilon(\sigma) a_{\sigma(1)1} \dots a_{\sigma(n)n} \\
&= \sum_{\sigma^{-1} \in S_n} \varepsilon(\sigma) a_{\sigma(1)1} \dots a_{\sigma(n)n} \\
&= \sum_{\sigma \in S_n} \varepsilon(\sigma^{-1}) a_{1\sigma(1)} \dots a_{n\sigma(n)} \\
&= \sum_{\sigma \in S_n} \varepsilon(\sigma) a_{1\sigma(1)} \dots a_{n\sigma(n)} \\
&= \det A^\intercal
\end{aligned}$$

□

5.4 Volume forms

Definition. A volume form d on F^n is a function $d : \underbrace{F^n \times \dots \times F^n}_{n \text{ times}} \to F$ satisfying

(i) d is multilinear: for all $i \in \{1, \dots, n\}$ and for all $v_1, \dots, v_{i-1}, v_{i+1}, \dots, v_n \in F^n$, the map from F^n to F defined by

$$v \mapsto (v_1, \dots, v_{i-1}, v, v_{i+1}, \dots, v_n)$$

is linear. In other words, this map is an element of $(F^n)^\star$.

(ii) d is alternating: for $v_i = v_j$ for some $i \neq j$, $d = 0$.

So an alternating multilinear form is a volume form. We want to show that, up to multiplication by a scalar, the determinant is the only volume form.

Lemma. The map $(F^n)^n \to F$ defined by $(A^{(1)}, \dots, A^{(n)}) \mapsto \det A$ is a volume form. This map is the determinant of A, but thought of as acting on the column vectors of A.

Proof. We first show that this map is multilinear. Fix $\sigma \in S_n$, and consider $\prod_{i=1}^n a_{\sigma(i)i}$. This product contains exactly one term in each column of A. Thus, the map $(A^{(1)}, \dots, A^{(n)}) \mapsto \prod_{i=1}^n a_{\sigma(i)i}$ is multilinear. This then clearly implies that the determinant, a sum of such multilinear maps, is itself multilinear.

Now, we show that the determinant is alternating. Let $k \neq \ell$, and $A^{(k)} = A^{(\ell)}$. Let $\tau = (k\ell)$ be the transposition exchanging k and ℓ. Then, for all $i, j \in \{1, \dots, n\}$, $a_{ij} = a_{i\tau(j)}$. We can decompose permutations into two disjoint sets: $S_n = A_n \cup \tau A_n$, where A_n is the alternating group of order n. Now, note that $\prod_{i=1}^n a_{\sigma(i)i} + \prod_{i=1}^n a_{(\tau\circ\sigma)(i)i} = 0$. So the sum over all $\sigma \in A_n$ gives zero. So the determinant is alternating, and hence a volume form. □

Lemma. Let d be a volume form. Then, swapping two entries changes the sign.

Proof. Take the sum of these two results:

$$\begin{aligned} d(v_1, \dots, v_i, \dots, v_j, \dots, v_n) + d(v_1, \dots, v_j, \dots, v_i, \dots, v_n) &= d(v_1, \dots, v_i, \dots, v_j, \dots, v_n) \\ &+ d(v_1, \dots, v_j, \dots, v_i, \dots, v_n) \\ &+ d(v_1, \dots, v_i, \dots, v_i, \dots, v_n) \\ &+ d(v_1, \dots, v_j, \dots, v_j, v_n) \\ &= 2d(v_1, \dots, v_i + v_j, \dots, v_i + v_j, \dots, v_n) \\ &= 0 \end{aligned}$$

as required. □

Corollary. If $\sigma \in S_n$ and d is a volume form, $d(v_{\sigma(1)}, \dots, v_{\sigma(n)}) = \varepsilon(\sigma) d(v_1, \dots, v_n)$.

Proof. We can decompose σ as a product of transpositions $\prod_{i=1}^{n_\sigma} e_i$. □

Theorem. Let d be a volume form on F^n. Let A be a matrix whose columns are $A^{(i)}$. Then

$$d(A^{(1)}, \dots, A^{(n)}) = \det A \cdot d(e_1, \dots, e_n)$$

So there is a unique volume form up to a constant multiple. We can then see that $\det A$ is the only volume form such that $d(e_1, \dots, e_n) = 1$.

Proof.

$$d(A^{(1)}, \dots, A^{(n)}) = d\left(\sum_{i=1}^n a_{i1} e_i, A^{(2)}, \dots, A^{(n)}\right)$$

Since d is multilinear,

$$d(A^{(1)}, \dots, A^{(n)}) = \sum_{i=1}^n a_{i1} d(e_i, A^{(2)}, \dots, A^{(n)})$$

Inductively on all columns,

$$d(A^{(1)}, \dots, A^{(n)}) = \sum_{i=1}^n \sum_{j=1}^n a_{i1} a_{j2} d(e_i, e_j, A^{(3)}, \dots, A^{(n)}) = \dots = \sum_{1 \le i_1, \le \dots \le n} \prod_{k=1}^n a_{i_\ell k} d(e_{i_1}, \dots e_{i_n})$$

Since d is alternating, we know that for $d(e_{i_1}, \dots, e_{i_n})$ to be nonzero, the i_k must be different, so this corresponds to a permutation $\sigma \in S_n$.

$$d(A^{(1)}, \dots, A^{(n)}) = \sum_{\sigma \in S_n} \prod_{k=1}^n a_{\sigma(k)k} \varepsilon(\sigma) d(e_1, \dots, e_n)$$

which is exactly the determinant up to a constant multiple. □

5.5 Multiplicative property of determinant

Lemma. Let $A, B \in M_n(F)$. Then $\det(AB) = \det(A)\det(B)$.

Proof. Given A, we define the volume form $d_A : (F^n)^n \to F$ by

$$d_A(v_1, \dots, v_n) \mapsto \det(Av_1, \dots, Av_n)$$

$v_i \mapsto Av_i$ is linear, and the determinant is multilinear, so d_A is multilinear. If $i \neq j$ and $v_i = v_j$, then $\det\big(\dots, Av_i, \dots, Av_j, \dots\big) = 0$ so d_A is alternating. Hence d_A is a volume form. Hence there exists a constant C_A such that $d_A(v_1, \dots, v_n) = C_A \det(v_1, \dots, v_n)$. We can compute C_A by considering the basis vectors; $Ae_i = A_i$ where A_i is the ith column vector of A. Then,

$$C_A = d_A(e_1, \dots, e_n) = \det(Ae_1, \dots, Ae_n) = \det A$$

Hence,

$$\det(AB) = d_A(B) = \det A \det B$$

□

5.6 Singular and non-singular matrices

Definition. Let $A \in M_n(F)$. We say that
(i) A is *singular* if $\det A = 0$;
(ii) A is *non-singular* if $\det A \neq 0$.

Lemma. If A is invertible, it is non-singular.

Proof. If A is invertible, there exists A^{-1}. Then, since the determinant is a homomorphism,

$$\det\big(AA^{-1}\big) = \det I = 1$$

Thus $\det A \det A^{-1} = 1$ and hence neither of these determinants can be zero. □

Theorem. Let $A \in M_n(F)$. The following are equivalent.
(i) A is invertible;
(ii) A is non-singular;
(iii) $r(A) = n$.

Proof. We have already shown that (i) implies (ii). We have also shown that (i) and (iii) are equivalent by the rank-nullity theorem. So it suffices to show that (ii) implies (iii).

Suppose $r(A) < n$. Then we will show A is singular. We have $\dim \operatorname{span}(A_1, \dots, A_n) < n$. Therefore, since there are n vectors, $(A_1, \dots, A_n)$ is not free. So there exist scalars λ_i not all zero such that $\sum_i \lambda_i A_i = 0$. Choose j such that $\lambda_j \neq 0$. Then,

$$A_j = -\frac{1}{\lambda_j} \sum_{i \neq j} \lambda_i A_i$$

So we can compute the determinant of A by

$$\det A = \det\left(A_1, \dots, -\frac{1}{\lambda_j}\sum_{i\neq j}\lambda_i A_i, \dots, A_n\right)$$

Since the determinant is alternating and linear in the jth entry, its value is zero. So A is singular as required. □

Remark. The above theorem gives necessary and sufficient conditions for invertibility of a set of n linear equations with n unknowns.

5.7 Determinants of linear maps

Lemma. Similar matrices have the same determinant.

Proof.

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

□

Definition. If α is an endomorphism, then we define

$$\det \alpha = \det[\alpha]_{B,B}$$

where B is any basis of the vector space. This is well-defined, since this value does not depend on the choice of basis.

Theorem. $\det : L(V, V) \to F$ satisfies the following properties.

(i) $\det I = 1$;
(ii) $\det(\alpha\beta) = \det \alpha \det \beta$;
(iii) $\det \alpha \neq 0$ if and only if α is invertible, and in this case, $\det\left(\alpha^{-1}\right)\det \alpha = 1$.

This is simply a reformulation of the previous theorem for matrices. The proof is simple, and relies on the invariance of the determinant under a change of basis.

5.8 Determinant of block-triangular matrices

Lemma. Let $A \in M_k(F)$, $B \in M_\ell(F)$, $C \in M_{k,\ell}(F)$. Consider the matrix

$$M = \begin{pmatrix} A & C \\ 0 & B \end{pmatrix}$$

Then $\det M = \det A \det B$.

Proof. Let $n = k + \ell$, so $M \in M_n(F)$. Let $M = (m_{ij})$. We must compute

$$\det M = \sum_{\sigma \in S_n} \varepsilon(\sigma) \prod_{i=1}^{n} m_{\sigma(i)i}$$

Observe that $m_{\sigma(i)i} = 0$ if $i \leq k$ and $\sigma(i) > k$. Then, we need only sum over $\sigma \in S_n$ such that for all $j \leq k$, we have $\sigma(j) \leq k$. Thus, for all $j \in \{k+1, \dots, n\}$, we have $\sigma(j) \in \{k+1, \dots, n\}$. We can then uniquely decompose σ into two permutations $\sigma = \sigma_1\sigma_2$, where σ_1 is restricted to $\{1, \dots, k\}$ and σ_2 is restricted to $\{k+1, \dots, n\}$. Hence,

$$\begin{aligned}
\det M &= \sum_{\sigma_1 \in S_k} \sum_{\sigma_2 \in S_{n-k}} \varepsilon(\sigma) \prod_{i=1}^{n} m_{\sigma(i)i} \\
&= \sum_{\sigma_1 \in S_k} \sum_{\sigma_2 \in S_{n-k}} \varepsilon(\sigma_1)\varepsilon(\sigma_2) \prod_{i=1}^{k} m_{\sigma(i)i} \prod_{i=k+1}^{n} m_{\sigma(i)i} \\
&= \sum_{\sigma_1 \in S_k} \varepsilon(\sigma_1) \prod_{i=1}^{k} m_{\sigma(i)i} \sum_{\sigma_2 \in S_{n-k}} \varepsilon(\sigma_2) \prod_{i=k+1}^{n} m_{\sigma(i)i} \\
&= \det A \det B
\end{aligned}$$

□

Corollary. We need not restrict ourselves to just two blocks, since we can apply the above lemma inductively. In particular, this implies that an upper-triangular matrix with diagonal elements λ_i has determinant $\prod_i \lambda_i$.

6 Adjugate matrices

6.1 Column and row expansions

Let $A \in M_n(F)$ with column vectors $A^{(i)}$. We know that

$$\det\left(A^{(1)}, \dots, A^{(j)}, \dots, A^{(k)}, \dots, A^{(n)}\right) = -\det\left(A^{(1)}, \dots, A^{(k)}, \dots, A^{(j)}, \dots, A^{(n)}\right)$$

Using the fact that $\det A = \det A^\intercal$ we can similarly see that swapping two rows will invert the sign of the determinant.

Remark. We could have proven all of the properties of the determinant above by using the decomposition of A into elementary matrices.

Definition. Let $A \in M_n(F)$. Let $i, j \in \{1, \dots, n\}$. We define the *minor* $A_{\hat{i}\hat{j}} \in M_{n-1}(F)$ to be the matrix obtained by removing the ith row and the jth column.

Lemma. Let $A \in M_n(F)$.

(i) Let $j \in \{1, \dots, n\}$. The determinant of A is given by the *column expansion with respect* to the *jth column*:

$$\det A = \sum_{i=1}^{n} (-1)^{i+j} a_{ij} \det A_{\hat{i}\hat{j}}$$

(ii) Let $i \in \{1, \dots, n\}$. The same determinant is also given by the *row expansion with respect to the ith row*:

$$\det A = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} \det A_{\hat{i}\hat{j}}$$

This is a process of reducing the computation of $n \times n$ determinants to $(n-1) \times (n-1)$ determinants.

Proof. We will prove case (i), the column expansion with respect to the jth column. Then (ii) will follow from the transpose of the matrix. Let $j \in \{1, \dots, n\}$. We can write $A^{(j)} = \sum_{i=1}^{n} a_{ij} e_i$ where the e_i are the canonical basis. Then, by swapping rows and columns,

$$\begin{aligned}
\det A &= \det\left(A^{(1)}, \dots, \sum_{i=1}^{n} a_{ij} e_i, \dots, A^{(n)}\right) \\
&= \sum_{i=1}^{n} a_{ij} \det\left(A^{(1)}, \dots, e_i, \dots, A^{(n)}\right) \\
&= \sum_{i=1}^{n} a_{ij} (-1)^{j-1} \det\left(e_i, A^{(1)}, \dots, A^{(n)}\right) \\
&= \sum_{i=1}^{n} a_{ij} (-1)^{j-1} (-1)^{i-1} \det\left(e_1, \overline{A}^{(1)}, \dots, \overline{A}^{(n)}\right)
\end{aligned}$$

This has brought the matrix into block form, where there is an element of value 1 in the top left, and the matrix $A_{\hat{i}\hat{j}}$ in the bottom right. The bottom left block is entirely zeroes. Hence,

$$\det A = \sum_{i=1}^{n} (-1)^{i+j} a_{ij} \det A_{\hat{i}\hat{j}}$$

as required. □

Remark. We have proven that

$$\det\left(A^{(1)}, \dots, e_i, \dots, A^{(n)}\right) = (-1)^{i+j} \det A_{\hat{i}\hat{j}}$$

6.2 Adjugates

Definition. Let $A \in M_n(F)$. The *adjugate matrix* of A, denoted $\operatorname{adj} A$, is the $n \times n$ matrix given by

$$(\operatorname{adj} A)_{ij} = (-1)^{i+j} \det A_{\hat{j}\hat{i}}$$

Hence,

$$\det\left(A^{(1)}, \dots, e_i, \dots, A^{(n)}\right) = (\operatorname{adj} A)_{ji}$$

Theorem. Let $A \in M_n(F)$. Then

$$(\operatorname{adj} A)A = (\det A)I$$

In particular, when A is invertible,

$$A^{-1} = \frac{\operatorname{adj} A}{\det A}$$

Proof. We have

$$\det A = \sum_{i=1}^{n} (-1)^{i+j} a_{ij} \det A_{\hat{i}\hat{j}}$$

Hence,

$$\det A = \sum_{i=1}^{n} (\operatorname{adj} A)_{ji} a_{ij} = ((\operatorname{adj} A)A)_{jj}$$

So the diagonal terms match. Off the diagonal,

$$0 = \det\left(A^{(1)}, \dots, \underbrace{A^{(k)}}_{j\text{th position}}, \dots, A^{(k)}, \dots, A^{(n)}\right)$$

By linearity,

$$\begin{aligned}
0 &= \det\left(A^{(1)}, \dots, \underbrace{\sum_{i=1}^{n} a_{ik} e_i}_{j\text{th position}}, \dots, A^{(k)}, \dots, A^{(n)}\right) \\
&= \sum_{i=1}^{n} a_{ik} \det\left(A^{(1)}, \dots, \underbrace{e_i}_{j\text{th position}}, \dots, A^{(k)}, \dots, A^{(n)}\right) \\
&= \sum_{i=1}^{n} a_{ik} (\operatorname{adj} A)_{ji} \\
&= ((\operatorname{adj} A)A)_{jk}
\end{aligned}$$

□

6.3 Cramer's rule

Proposition. Let A be an invertible square matrix of dimension n. Let $b \in F^n$. Then the unique solution to $Ax = b$ is given by

$$x_i = \frac{1}{\det A} \det\left(A_{\hat{i}\hat{b}}\right)$$

where $A_{\hat{i}\hat{b}}$ is obtained by replacing the ith column of A by b. This is an algorithm to compute x, avoiding the computation of A^{-1}.

Proof. Let A be invertible. Then there exists a unique $x \in F^n$ such that $Ax = b$. Then, since the determinant is alternating,

$$\begin{aligned}
\det(A_{\hat{ib}}) &= \det(A^{(1)}, \dots, A^{(i-1)}, b, A^{(i+1)}, \dots, A^{(n)}) \\
&= \det\left(A^{(1)}, \dots, A^{(i-1)}, \sum_{j=1}^{n} x_j A^{(j)}, A^{(i+1)}, \dots, A^{(n)}\right) \\
&= \det\left(A^{(1)}, \dots, A^{(i-1)}, x_i A^{(i)}, A^{(i+1)}, \dots, A^{(n)}\right) \\
&= x_i \det A
\end{aligned}$$

So the formula works. □

7 Eigenvectors and eigenvalues

7.1 Eigenvalues

Let V be an F-vector space. Let $\dim V = n < \infty$, and let α be an endomorphism of V. We wish to find a basis B of V such that, in this basis, $[\alpha]_B \equiv [\alpha]_{B,B}$ has a simple (e.g. diagonal, triangular) form. Recall that if B' is another basis and P is the change of basis matrix, $[\alpha]_{B'} = P^{-1}[\alpha]_B P$. Equivalently, given a square matrix $A \in M_n(F)$ we want to conjugate it by a matrix P such that the result is 'simpler'.

Definition. Let $\alpha \in L(V)$ be an endomorphism. We say that α is *diagonalisable* if there exists a basis B of V such that the matrix $[\alpha]_B$ is diagonal. We say that α is *triangulable* if there exists a basis B of V such that $[\alpha]_B$ is triangular.

Remark. We can express this equivalently in terms of conjugation of matrices.

Definition. A scalar $\lambda \in F$ is an *eigenvalue* of an endomorphism α if and only if there exists a vector $v \in V \setminus \{0\}$ such that $\alpha(v) = \lambda v$. Such a vector is an *eigenvector* with eigenvalue λ. $V_\lambda = \{v \in V : \alpha(v) = \lambda v\} \leq V$ is the *eigenspace* associated to λ.

Lemma. λ is an eigenvalue if and only if $\det(\alpha - \lambda I) = 0$.

Proof. If λ is an eigenvalue, there exists a nonzero vector v such that $\alpha(v) = \lambda v$, so $(\alpha - \lambda)(v) = 0$. So the kernel is non-trivial. So $\alpha - \lambda I$ is not injective, so it is not surjective by the rank-nullity theorem. Hence this matrix is not invertible, so it has zero determinant. □

Remark. If $\alpha(v_j) = \lambda v_j$ for $j \in \{1, \dots, m\}$, we can complete the family v_j into a basis $(v_1, \dots, v_n)$ of V. Then in this basis, the first m columns of the matrix α has diagonal entries λ_j.

7.2 Polynomials

Recall the following facts about polynomials on a field, for instance

$$f(t) = a_n t^n + \dots + a_1 t + a_0$$

We say that the degree of f, written $\deg f$ is n. The degree of $f + g$ is at most the maximum degree of f and g. $\deg(fg) = \deg f + \deg g$. Let $F[t]$ be the vector space of polynomials with coefficients in F. If λ is a root of f, then $(t - \lambda)$ divides F.

Proof.

$$f(t) = a_n t^n + \cdots + a_1 t + a_0$$

Hence,

$$f(\lambda) = a_n \lambda^n + \cdots + a_1 \lambda + a_0 = 0$$

which implies that

$$f(t) = f(t) - f(\lambda) = a_n(t^n - \lambda^n) + \cdots + a_1(t - \lambda)$$

But note that, for all n,

$$t^n - \lambda^n = (1 - \lambda)(t^{n-1} + \lambda t^{n-2} + \cdots + \lambda^{n-2} t + \lambda^{n-1})$$

□

Remark. We say that λ is a root of *multiplicity* k if $(t - \lambda)^k$ divides f but $(t - \lambda)^{k+1}$ does not.

Corollary. A nonzero polynomial of degree n has at most n roots, counted with multiplicity.

Corollary. If f_1, f_2 are two polynomials of degree less than n such that $f_1(t_i) = f_2(t_i)$ for $i \in \{1, \dots, n\}$ and t_i distinct, then $f_1 \equiv f_2$.

Proof. $f_1 - f_2$ has degree less than n, but has n roots. Hence it is zero. □

Theorem. Any polynomial $f \in \mathbb{C}[t]$ of positive degree has a complex root. When counted with multiplicity, f has a number of roots equal to its degree.

Corollary. Any polynomial $f \in \mathbb{C}[t]$ can be factorised into an amount of linear factors equal to its degree.

7.3 Characteristic polynomials

Definition. Let α be an endomorphism. The *characteristic polynomial* of α is

$$\chi_\alpha(\lambda) = \det(\alpha - \lambda I)$$

Remark. χ_α is a polynomial because the determinant is defined as a polynomial in the terms of the matrix. Note further that conjugate matrices have the same characteristic polynomial, so the above definition is well defined in any basis. Indeed, $\det\big(P^{-1}\alpha P - \lambda I\big) = \det\big(P^{-1}(\alpha - \lambda I)P\big) = \det(\alpha - \lambda I)$.

Theorem. Let $\alpha \in L(V)$. α is triangulable if and only if χ_α can be written as a product of linear factors over F. In particular, all complex matrices are triangulable.

Proof. Suppose α is triangulable. Then for a basis B, $[\alpha]_B$ is triangulable with diagonal entries a_i. Then

$$\chi_\alpha(t) = (a_1 - t)(a_2 - t) \cdots (a_n - t)$$

Conversely, let $\chi_\alpha(t)$ be the characteristic polynomial of α with a root λ. Then, $\chi_\alpha(\lambda) = 0$ implies λ is an eigenvalue. Let V_λ be the corresponding eigenspace. Let $(v_1, \dots, v_k)$ be the basis of this eigenspace, completed to a basis $(v_1, \dots, v_n)$ of V. Let $W = \operatorname{span}\{v_{k+1}, \dots, v_n\}$, and then $V = V_\lambda \oplus W$. Then

$$[\alpha]_B = \begin{pmatrix} \lambda I & \star \\ 0 & C \end{pmatrix}$$

where $\star$ is arbitrary, and C is a block of size $(n-k) \times (n-k)$. Then α induces an endomorphism $\overline{\alpha}\colon V/U \to V/U$ with respect to the basis $(v_{k+1}, \dots, v_n)$, where $U = V_\lambda$. By induction on the dimension, we can find a basis $(w_{k+1}, \dots, w_n)$ for which C has a triangular form. Then the basis $(v_1, \dots, v_k, w_{k+1}, \dots, w_n)$ is a basis for which α is triangular. □

Lemma. Let $n = \dim V$, and V be a vector space over $\mathbb{R}$ or $\mathbb{C}$. Let α be an endomorphism on V. Then

$$\chi_\alpha(t) = (-1)^n t^n + c_{n-1} t^{n-1} + \dots + c_0$$

with

$$c_0 = \det A; \quad c_{n-1} = (-1)^{n-1} \operatorname{tr} A$$

Proof.

$$\chi_\alpha(t) = \det(\alpha - tI) \implies \chi_\alpha(0) = \det(\alpha)$$

Further, for $\mathbb{R}, \mathbb{C}$ we know that α is triangulable over $\mathbb{C}$. Hence $\chi_\alpha(t)$ is the determinant of a triangular matrix;

$$\chi_\alpha(t) = \prod_{i=1}^n (a_i - t)$$

Hence

$$c_{n-1} = (-1)^{n-1} a_i$$

Since the trace is invariant under a change of basis, this is exactly the trace as required. □

7.4 Polynomials for matrices and endomorphisms

Let $p(t)$ be a polynomial over F. We will write

$$p(t) = a_n t^n + \dots + a_0$$

For a matrix $A \in M_n(F)$, we write

$$p(A) = a_n A^n + \dots + a_0 \in M_n(F)$$

For an endomorphism $\alpha \in L(V)$,

$$p(\alpha) = a_n \alpha^n + \dots + a_0 I \in L(V); \quad \alpha^k \equiv \underbrace{\alpha \circ \dots \circ \alpha}_{k \text{ times}}$$

7.5 Sharp criterion of diagonalisability

Theorem. Let V be a vector space over F of finite dimension n. Let α be an endomorphism of V. Then α is diagonalisable if and only if there exists a polynomial p which is a product of *distinct* linear factors, such that $p(\alpha) = 0$. In other words, there exist distinct $\lambda_1, \dots, \lambda_k$ such that

$$p(t) = \prod_{i=1}^{n} (t - \lambda_i) \implies p(\alpha) = 0$$

Proof. Suppose α is diagonalisable in a basis B. Let $\lambda_1, \dots, \lambda_k$ be the $k \leq n$ *distinct* eigenvalues. Let

$$p(t) = \prod_{i=1}^{k} (t - \lambda_i)$$

Let $v \in B$. Then $\alpha(v) = \lambda_i v$ for some i. Then, since the terms in the following product commute,

$$(\alpha - \lambda_i I)(v) = 0 \implies p(\alpha)(v) = \left[\prod_{i=1}^{k} (\alpha - \lambda_i I) \right] (v) = 0$$

So for all basis vectors, $p(\alpha)(v)$. By linearity, $p(\alpha) = 0$.

Conversely, suppose that $p(\alpha) = 0$ for some polynomial $p(t) = \prod_{i=1}^{k} (t - \lambda_i)$ with distinct λ_i. Let $V_{\lambda_i} = \ker(\alpha - \lambda_i I)$. We claim that

$$V = \bigoplus_{i=1}^{k} V_{\lambda_i}$$

Consider the polynomials

$$q_j(t) = \prod_{i=1, i \neq j}^{k} \frac{t - \lambda_i}{\lambda_j - \lambda_i}$$

These polynomials evaluate to one at λ_j and zero at λ_i for $i \neq j$. Hence $q_j(\lambda_i) = \delta_{ij}$. We now define the polynomial

$$q = q_1 + \dots + q_k$$

The degree of q is at most $(k - 1)$. Note, $q(\lambda_i) = 1$ for all $i \in \{1, \dots, k\}$. The only polynomial that evaluates to one at k points with degree at most $(k - 1)$ is exactly given by $q(t) = 1$. Consider the endomorphism

$$\pi_j = q_j(\alpha) \in L(V)$$

These are called the 'projection operators'. By construction,

$$\sum_{j=1}^{k} \pi_j = \sum_{j=1}^{k} q_j(\alpha) = I$$

So the sum of the π_j is the identity. Hence, for all $v \in V$,

$$I(v) = v = \sum_{j=1}^{k} \pi_j(v) = \sum_{j=1}^{k} q_j(\alpha)(v)$$

So we can decompose any vector as a sum of its projections $\pi_j(v)$. Now, by definition of q_j and p,

$$\begin{aligned}(\alpha - \lambda_j I)q_j(\alpha)(v) &= \frac{1}{\prod_{i\neq j}(\lambda_j - \lambda_i)}(\alpha - \lambda_j I)\left[\prod_{i\neq j}(t - \lambda_i)\right](\alpha) \\ &= \frac{1}{\prod_{i\neq j}(\lambda_j - \lambda_i)}\prod_{i=1}^{k}(\alpha - \lambda_i I)(v) \\ &= \frac{1}{\prod_{i\neq j}(\lambda_j - \lambda_i)}p(\alpha)(v)\end{aligned}$$

By assumption, this is zero. For all v, we have $(\alpha - \lambda_j I)q_j(\alpha)(v)$. Hence,

$$(\alpha - \lambda_j I)\pi_j(v) = 0 \implies \pi_j(v) \in \ker(\alpha - \lambda_j I) = v_j$$

We have then proven that, for all $v \in V$,

$$v = \sum_{j=1}^{k} \underbrace{\pi_j(v)}_{\in V_j}$$

Hence,

$$V = \sum_{j=1}^{k} V_j$$

It remains to show that the sum is direct. Indeed, let

$$v \in V_{\lambda_j} \cap \left(\sum_{i\neq j} V_{\lambda_i}\right)$$

We must show $v = 0$. Applying π_j,

$$\pi_j(v) = q_j(\alpha)(v) = \prod_{i\neq j}\frac{(\alpha - \lambda_i I)(v)}{\lambda_j - \lambda_i}$$

Since $\alpha(v) = \lambda_j v$,

$$\pi_j(v) = \prod_{i\neq j}\frac{(\lambda_j - \lambda_i)v}{\lambda_j - \lambda_i} = v$$

Hence π_j really projects onto V_{λ_j}. However, we also know $v \in \sum_{i\neq j} V_{\lambda_i}$. So we can write $v = \sum_{i\neq j} w_i$ for $w \in V_{\lambda_i}$. Thus,

$$\pi_j(w_i) = \prod_{m\neq j}\frac{(\alpha - \lambda_m I)(v)}{\lambda_m - \lambda_j}$$

Since $\alpha(w_i) = \lambda_i w_i$, one of the factors will vanish, hence

$$\pi_j(w_i) = 0$$

So

$$v = \sum_{i\neq j} w_i \implies \pi_j(v) = \sum_{i\neq j}\pi_j(w_i) = 0$$

But $v = \pi_j(v)$ hence $v = 0$. So the sum is direct. Hence, $B = (B_1, \dots, B_k)$ is a basis of V, where the B_i are bases of V_{λ_i}. Then $[\alpha]_B$ is diagonal. □

Remark. We have shown further that if $\lambda_1, \dots, \lambda_k$ are distinct eigenvalues of α, then

$$\sum_{i=1}^{k} V_{\lambda_i} = \bigoplus_{i=1}^{k} V_{\lambda_i}$$

Therefore, the only way that diagonalisation fails is when this sum is not direct, so

$$\sum_{i=1}^{k} V_{\lambda_i} < V$$

Example. Let $F = \mathbb{C}$. Let $A \in M_n(F)$ such that A has finite order; there exists $m \in \mathbb{N}$ such that $A^m = I$. Then A is diagonalisable. This is because

$$t^m - 1 = p(t) = \prod_{j=1}^{m} (t - \xi_m^j); \quad \xi_m = e^{2\pi i/m}$$

and $p(A) = 0$.

7.6 Simultaneous diagonalisation

Theorem. Let α, β be endomorphisms of V which are diagonalisable. Then α, β are *simultaneously diagonalisable* (there exists a basis B of V such that $[\alpha]_B, [\beta]_B$ are diagonal) if and only if α and β commute.

Proof. Two diagonal matrices commute. If such a basis exists, $\alpha\beta = \beta\alpha$ in this basis. So this holds in any basis. Conversely, suppose $\alpha\beta = \beta\alpha$. We have

$$V = \bigoplus_{i=1}^{k} V_{\lambda_i}$$

where $\lambda_i, \dots, \lambda_k$ are the k distinct eigenvalues of α. We claim that $\beta\left(V_{\lambda_j}\right) \leq V_{\lambda_j}$. Indeed, for $v \in V_{\lambda_j}$,

$$\alpha\beta(v) = \beta\alpha(v) = \beta(\lambda_j v) = \lambda_j \beta(v) \implies \alpha(\beta(v)) = \lambda_j \beta(v)$$

Hence, $\beta(v) \in V_{\lambda_j}$. By assumption, β is diagonalisable. Hence, there exists a polynomial p with distinct linear factors such that $p(\beta) = 0$. Now, $\beta\left(V_{\lambda_j}\right) \leq V_{\lambda_j}$ so we can consider $\beta|_{V_{\lambda_j}}$. This is an endomorphism of V_{λ_j}. We can compute

$$p\left(\beta\Big|_{V_{\lambda_j}}\right) = 0$$

Hence, $\beta|_{V_{\lambda_j}}$ is diagonalisable. Let B_i be the basis of V_{λ_i} in which $\beta|_{V_{\lambda_j}}$ is diagonal. Since $V = \bigoplus V_{\lambda_i}$, $B = (B_1, \dots, B_k)$ is a basis of V. Then the matrices of α and β in V are diagonal. $\square$

7.7 Minimal polynomials

Recall from IB Groups, Rings and Modules the Euclidean algorithm for dividing polynomials. Given a, b polynomials over F with b nonzero, there exist polynomials q, r over F with $\deg r < \deg b$ and $a = qb + r$.

Definition. Let V be a finite dimensional F-vector space. Let α be an endomorphism on V. The *minimal polynomial* m_α of α is the nonzero polynomial with smallest degree such that $m_\alpha(\alpha) = 0$.

Remark. If $\dim V = n < \infty$, then $\dim L(V) = n^2$. In particular, the family $\left\{I, \alpha, \dots, \alpha^{n^2}\right\}$ cannot be free since it has $n^2 + 1$ entries. This generates a polynomial in α which evaluates to zero. Hence, a minimal polynomial always exists.

Lemma. Let $\alpha \in L(V)$ and $p \in F[t]$ be a polynomial. Then $p(\alpha) = 0$ if and only if m_α is a factor of p. In particular, m_α is well-defined and unique up to a constant multiple.

Proof. Let $p \in F[t]$ such that $p(\alpha) = 0$. If $m_\alpha(\alpha) = 0$ and $\deg m_\alpha < \deg p$, we can perform the division $p = m_\alpha q + r$ for $\deg r < \deg m_\alpha$. Then $p(\alpha) = m_\alpha(\alpha)q(\alpha) + r(\alpha)$. But $m_\alpha(\alpha) = 0$. But $\deg r < \deg m_\alpha$ and m_α is the smallest degree polynomial which evaluates to zero for α, so $r \equiv 0$ so $p = m_\alpha q$. In particular, if m_1, m_2 are both minimal polynomials that evaluate to zero for α, we have m_1 divides m_2 and m_2 divides m_1. Hence they are equivalent up to a constant. $\square$

Example. Let $V = F^2$ and

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

We can check $p(t) = (t-1)^2$ gives $p(A) = p(B) = 0$. So the minimal polynomial of A or B must be either $(t-1)$ or $(t-1)^2$. For A, we can find the minimal polynomial is $(t-1)$, and for B we require $(t-1)^2$. So B is not diagonalisable, since its minimal polynomial is not a product of distinct linear factors.

7.8 Cayley–Hamilton theorem

Theorem. Let V be a finite dimensional F-vector space. Let $\alpha \in L(V)$ with characteristic polynomial $\chi_\alpha(t) = \det(\alpha - tI)$. Then $\chi_\alpha(\alpha) = 0$.

Two proofs will provided; one more physical and based on $F = \mathbb{C}$ and one more algebraic.

Proof. Let $B = \{v_1, \dots, v_n\}$ be a basis of V such that $[\alpha]_B$ is triangular. This can be done when $F = \mathbb{C}$. Note, if the diagonal entries in this basis are a_i,

$$\chi_\alpha(t) = \prod_{i=1}^n (a_i - t) \implies \chi_\alpha(\alpha) = (\alpha - a_1 I) \dots (\alpha - a_n I)$$

We want to show that this expansion evaluates to zero. Let $U_j = \operatorname{span}\{v_1, \dots, v_j\}$. Let $v \in V = U_n$. We want to compute $\chi_\alpha(\alpha)(v)$. Note, by construction of the triangular matrix.

$$\begin{aligned}
\chi_\alpha(\alpha)(v) &= (\alpha - a_1 I) \dots \underbrace{(\alpha - a_n I)(v)}_{\in U_{n-1}} \\
&= (\alpha - a_1 I) \dots \underbrace{(\alpha - a_{n-1} I)(\alpha - a_n I)(v)}_{\in U_{n-2}} \\
&= \dots \\
&\in U_0
\end{aligned}$$

Hence this evaluates to zero. □

The following proof works for any field where we can equate coefficients, but is much less intuitive.

Proof. We will write

$$\det(tI - \alpha) = (-1)^n \chi_\alpha(t) = t^n + a_{n-1} t^{n-1} + \dots + a_0$$

For any matrix B, we have proven $B \operatorname{adj} B = (\det B) I$. We apply this relation to the matrix $B = tI - A$. We can check that

$$\operatorname{adj} B = \operatorname{adj}(tI - A) = B_{n-1} t^{n-1} + \dots + B_1 t + B_0$$

since adjugate matrices are degree $(n-1)$ polynomials for each element. Then, by applying $B \operatorname{adj} B = (\det B) I$,

$$(tI - A)[B_{n-1} t^{n-1} + \dots + B_1 t + B_0] = (\det B) I = (t^n + \dots + a_0) I$$

Since this is true for all t, we can equate coefficients. This gives

$$\begin{aligned}
t^n &: & I &= B_{n-1} \\
t^{n-1} &: & a_{n-1} I &= B_{n-2} - A B_{n-1} \\
&\vdots & &\vdots \\
t^0 &: & a_0 I &= -A B_1
\end{aligned}$$

Then, substituting A for t in each relation will give, for example, $A^n I = A^n B_{n-1}$. Computing the sum of all of these identities, we recover the original polynomial in terms of A instead of in terms of t. Many terms will cancel since the sum telescopes, yielding

$$A^n + a_{n-1} A^{n-1} + \dots + a_0 I = 0$$

□

7.9 Algebraic and geometric multiplicity

Definition. Let V be a finite dimensional F-vector space. Let $\alpha \in L(V)$ and let λ be an eigenvalue of α. Then

$$\chi_\alpha(t) = (t - \lambda)^{a_\lambda} q(t)$$

where $q(t)$ is a polynomial over F such that $(t - \lambda)$ does not divide q. a_λ is known as the *algebraic multiplicity* of the eigenvalue λ. We define the *geometric multiplicity* g_λ of λ to be the dimension of the eigenspace associated with λ, so $g_\lambda = \dim\ker(\alpha - \lambda I)$.

Lemma. If λ is an eigenvalue of $\alpha \in L(V)$, then $1 \le g_\lambda \le a_\lambda$.

Proof. We have $g_\lambda = \dim\ker(\alpha - \lambda I)$. There exists a nontrivial vector $v \in V$ such that $v \in \ker(\alpha - \lambda I)$ since λ is an eigenvalue. Hence $g_\lambda \ge 1$. We will show that $g_\lambda \le a_\lambda$. Indeed, let $v_1, \dots, v_{g_\lambda}$ be a basis of $V_\lambda \equiv \ker(\alpha - \lambda I)$. We complete this into a basis $B \equiv \left(v_1, \dots, v_{g_\lambda}, v_{g_\lambda+1}, \dots, v_n\right)$ of V. Then note that

$$[\alpha]_B = \begin{pmatrix} \lambda I_{g_\lambda} & \star \\ 0 & A_1 \end{pmatrix}$$

for some matrix A_1. Now,

$$\det(\alpha - tI) = \det\begin{pmatrix} (\lambda - t)I_{g_\lambda} & \star \\ 0 & A_1 - tI \end{pmatrix}$$

By the formula for determinants of block matrices with a zero block on the off diagonal,

$$\det(\alpha - tI) = (\lambda - t)^{g_\lambda} \det(A_1 - tI)$$

Hence $g_\lambda \le a_\lambda$ since the determinant is a polynomial that could have more factors of the same form. □

Lemma. Let V be a finite dimensional F-vector space. Let $\alpha \in L(V)$ and let λ be an eigenvalue of α. Let c_λ be the multiplicity of λ as a root of the minimal polynomial of α. Then $1 \le c_\lambda \le a_\lambda$.

Proof. By the Cayley–Hamilton theorem, $\chi_\alpha(\alpha) = 0$. Since m_α is linear, m_α divides χ_α. Hence $c_\lambda \le a_\lambda$. Now we show $c_\lambda \ge 1$. Indeed, λ is an eigenvalue hence there exists a nonzero $v \in V$ such that $\alpha(v) = \lambda v$. For such an eigenvector, $\alpha^P(v) = \lambda^P v$ for $P \in \mathbb{N}$. Hence for $p \in F[t]$, $p(\alpha)(v) = [p(\lambda)](v)$. Hence $m_\alpha(\alpha)(v) = [m_\alpha(\lambda)](v)$. Since the left hand side is zero, $m_\alpha(\lambda) = 0$. So $c_\lambda \ge 1$. □

Example. Let

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

The minimal polynomial can be computed by considering the characteristic polynomial

$$\chi_A(t) = (t-1)^2(t-2)$$

So the minimal polynomial is either $(t-1)^2(t-2)$ or $(t-1)(t-2)$ We check $(t-1)(t-2)$. $(A-I)(A-2I)$ can be found to be zero. So $m_A(t) = (t-1)(t-2)$. Since this is a product of distinct linear factors, A is diagonalisable.

Example. Let A be a Jordan block of size $n \ge 2$. Then $g_\lambda = 1$, $a_\lambda = n$, and $c_\lambda = n$.

7.10 Characterisation of diagonalisable complex endomorphisms

Lemma. Let $F = \mathbb{C}$. Let V be a finite-dimensional $\mathbb{C}$-vector space. Let α be an endomorphism of V. Then the following are equivalent.

(i) α is diagonalisable;

(ii) for all λ eigenvalues of α, we have $a_\lambda = g_\lambda$;

(iii) for all λ eigenvalues of α, $c_\lambda = 1$.

Proof. First, the fact that (i) is true if and only if (iii) is true has already been proven. Now let us show that (i) is equivalent to (ii). Let $\lambda_1, \dots, \lambda_k$ be the distinct eigenvalues of α. We have already found that α is diagonalisable if and only if $V = \bigoplus V_{\lambda_i}$. The sum was found to be always direct, regardless of diagonalisability. We will compute the dimension of V in two ways;

$$n = \dim V = \deg \chi_\alpha; \quad n = \dim V = \sum_{i=1}^k a_{\lambda_i}$$

since χ_α is a product of $(t - \lambda_i)$ factors as $F = \mathbb{C}$. Since the sum is direct,

$$\dim \left(\bigoplus_{i=1}^k V_{\lambda_i} \right) = \sum_{i=1}^k g_{\lambda_i}$$

α is diagonalisable if and only if the dimensions are equal, so

$$\sum_{i=1}^k g_{\lambda_i} = \sum_{i=1}^k a_{\lambda_i}$$

Conversely, we have proven that for all eigenvalues λ_i, we have $g_{\lambda_i} \leq a_{\lambda_i}$. Hence, $\sum_{i=1}^k g_{\lambda_i} = \sum_{i=1}^k a_{\lambda_i}$ holds if and only if $g_{\lambda_i} = a_{\lambda_i}$ for all i. □

8 Jordan normal form

For this section, let $F = \mathbb{C}$.

8.1 Definition

Definition. Let $A \in M_n(\mathbb{C})$. We say that A is in *Jordan normal form* if it is a block diagonal matrix, where each block is of the form

$$J_{n_i}(\lambda) = \begin{pmatrix} \lambda & 1 & 0 & \cdots & 0 \\ 0 & \lambda & 1 & \cdots & 0 \\ 0 & 0 & \lambda & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda \end{pmatrix}$$

We say that $J_{n_i}(\lambda) \in M_{n_i}(\mathbb{C})$ are *Jordan blocks*. The $\lambda_i \in \mathbb{C}$ need not be distinct.

Remark. In three dimensions,

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

is in Jordan normal form, with three one-dimensional Jordan blocks with the same λ value.

8.2 Similarity to Jordan normal form

Theorem. Any complex matrix $A \in M_n(\mathbb{C})$ is similar to a matrix in Jordan normal form, which is unique up to reordering the Jordan blocks.

The proof is non-examinable. This follows from IB Groups, Rings and Modules.

Example. Let $\dim V = 2$. Then any matrix is similar to one of

$$\begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}; \quad \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix}; \quad \begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix}$$

The minimal polynomials are

$$(t-\lambda_1)(t-\lambda_2); \quad (t-\lambda); \quad (t-\lambda)^2$$

8.3 Direct sum of eigenspaces

Theorem. Let V be a $\mathbb{C}$-vector space. Let $\dim V = n < \infty$. Then, the minimal polynomial $m_\alpha(t)$ of an endomorphism $\alpha \in L(V)$ satisfies

$$V = \bigoplus_{j=1}^{k} V_j$$

where $V_j = \ker[(\alpha - \lambda_j I)^{c_j}]$, and where

$$m_\alpha(t) = \prod_{i=1}^{k} (t-\lambda_i)^{c_i}$$

V_j is called a *generalised eigenspace* associated with λ_j.

Remark. Note that V_j is stable by α, that is, $\alpha(V_j) = V_j$. Note further that $(\alpha - \lambda_j I)\big|_{V_j} = \mu_j$ gives that μ_j is a nilpotent endomorphism; $\mu_j^{c_j} = 0$. So the Jordan normal form theorem is a statement about nilpotent matrices.

Note, when α is diagonalisable, $c_j = 1$ and hence we recover $V_j = \ker(\alpha - \lambda_j I)$ and $V = \bigoplus V_j$.

Proof. The key to this proof is that the projectors onto V_j are 'explicit'. First, recall

$$m_\alpha(t) = \prod_{j=1}^{k} (t-\lambda_j)^{c_j}$$

Then, let

$$p_j(t) = \prod_{i \neq j} (t - \lambda_i)^{c_i}$$

Then p_j have by definition no common factor. So by Euclid's algorithm, we can find polynomials q_i such that

$$\sum_{i=1}^{k} q_i p_i = 1$$

We define the projector $\pi_j = q_j p_j(\alpha)$, which is an endomorphism. By construction, for all $v \in V$, we have

$$\sum_{j=1}^{k} \pi_j(v) = \sum_{j=1}^{k} a_j p_j(\alpha(v)) = I(v) = v$$

Hence,

$$v = \sum_{i=1}^{k} \pi_i(v)$$

Observe further that $\pi_j(v) \in V_j$. Indeed,

$$(\alpha - \lambda_j I)^{c_j} \pi_j(v) = (\alpha - \lambda_j I)^{c_j} q_j p_j(\alpha(v)) = q_j m_\alpha(\alpha(v)) = 0$$

Hence $\pi_j(v) \in V_j$. In particular, $V = \sum_{j=1}^{k} V_j$. We need to show that this sum is direct. Note, for $i \neq j$, $\pi_i \pi_j = 0$ from the definition of π. Hence, observe that

$$\pi_i = \pi_i \left(\sum_{j=1}^{k} \pi_j \right) \implies \pi_i = \pi_i \pi_i$$

Thus, π is a projector. In particular, this implies that $\pi_i|_{V_j}$ is the identity if $i = j$ and zero if $i \neq j$. This immediately implies that th sum is direct;

$$V = \bigoplus_{j=1}^{k} V_j$$

Indeed, suppose

$$\sum_{j=1}^{k} \alpha_j v_j = 0; \quad v_j \in V_j; \quad \alpha_1 = 0$$

Then

$$v_1 = -\frac{1}{\alpha_1} \sum_{j=2}^{k} \alpha_j v_j$$

Applying π_1,

$$v_1 = -\frac{1}{\alpha_1} \sum_{j=2}^{k} \alpha_j \pi_1(v_j) = 0$$

Iterating, we find $v = 0$. □

Remark. We can compute the quantities $a_\lambda, g_\lambda, c_\lambda$ on the Jordan normal form of a matrix. Indeed, let $m \geq 2$ and consider a Jordan block $J_m(\lambda)$. Then $J_m(\lambda) - \lambda I$ is the zero matrix with ones on the off-diagonal. $(J_m(\lambda) - \lambda I)^k$ pushes the ones onto the next line iteratively, so

$$(J_m(\lambda) - \lambda I)^k = \begin{pmatrix} 0 & I_{m-k} \\ 0 & 0 \end{pmatrix}$$

Hence J is nilpotent of order exactly m. In Jordan normal form,

(i) a_λ is the sum of sizes of blocks with eigenvalue λ. This is the amount of times λ is seen on the diagonal.

(ii) g_λ is the amount of blocks with eigenvalue λ, since each block represents one eigenvector.

(iii) c_λ is the size of the largest block with eigenvalue λ.

Example. Let

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

We wish to convert this matrix into Jordan normal form; so we seek a basis for which this matrix becomes Jordan normal form.

$$\chi_A(t) = (t-1)^2$$

Hence there exists only one eigenvalue, $\lambda = 1$. $A - I \neq 0$ hence $m_\alpha(t) = (t-1)^2$. Thus, the Jordan normal form of A is of the form

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

Now,

$$\ker(A - I) = \langle v_1 \rangle; \quad v_1 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

Further, we seek a v_2 such that

$$(A - I)v_2 = v_1 \implies v_2 = \begin{pmatrix} -1 \\ 0 \end{pmatrix}$$

Such a v_2 is not unique. Now,

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

9 Properties of bilinear forms

9.1 Changing basis

Let $\phi \colon V \times V \to \mathbb{F}$ be a bilinear form. Let V be a finite-dimensional F-vector space. Let B be a basis of V and let $[\phi]_B = [\phi]_{BB}$ be the matrix with entries $\phi(e_i, e_j)$.

Lemma. Let ϕ be a bilinear form $V \times V \to F$. Then if B, B' are bases for V, and $P = [I]_{B',B}$ we have

$$[\phi]_{B'} = P^\intercal [\phi]_B P$$

Proof. This is a special case of the general change of basis formula. □

Definition. Let $A, B \in M_n(F)$ be square matrices. We say that A, B are *congruent* if there exists $P \in M_n(F)$ such that $A = P^\intercal B P$.

Remark. Congruence is an equivalence relation.

Definition. A bilinear form ϕ on V is *symmetric* if, for all $u, v \in V$, we have

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

Remark. If A is a square matrix, we say A is symmetric if $A = A^\intercal$. Equivalently, $A_{ij} = A_{ji}$ for all i, j. So ϕ is symmetric if and only if $[\phi]_B$ is symmetric for any basis B. Note further that to represent ϕ by a diagonal matrix in some basis B, it must necessarily be symmetric, since

$$P^\intercal A P = D \implies D = D^\intercal = (P^\intercal A P)^\intercal = P^\intercal A^\intercal P \implies A = A^\intercal$$

9.2 Quadratic forms

Definition. A map $Q\colon V \to F$ is a *quadratic form* if there exists a bilinear form $\phi\colon V \times V \to F$ such that, for all $u \in V$,

$$Q(u) = \phi(u, u)$$

So a quadratic form is the restriction of a bilinear form to the diagonal.

Remark. Let $B = (e_i)$ be a basis of V. Let $A = [\phi]_B = (\phi(e_i, e_j)) = (a_{ij})$. Then, for $u = \sum_i x_i e_i \in V$,

$$Q(u) = \phi(u, u) = \phi\left(\sum_i x_i e_i, \sum_j x_j e_j\right) = \sum_i \sum_j x_i x_j \phi(e_i, e_j) = \sum_i \sum_j x_i x_j a_{ij}$$

We can check that this is equal to

$$Q(u) = x^\intercal A x$$

where $[u]_B = x$. Note further that

$$x^\intercal A x = \sum_i \sum_j a_{ij} x_i x_j = \sum_i \sum_j a_{ji} x_i x_j = \sum_i \sum_j \frac{a_{ij} + a_{ji}}{2} x_i x_j = x^\intercal \left(\underbrace{\frac{A + A^\intercal}{2}}_{\text{symmetric}} \right) x$$

So we can always express the quadratic form as a symmetric matrix in any basis.

Proposition. If $Q\colon V \to F$ is a quadratic form, then there exists a unique symmetric bilinear form $\phi\colon V \times V \to F$ such that $Q(u) = \phi(u, u)$.

Proof. Let ψ be a bilinear form on V such that for all $u \in V$, we have $Q(u) = \psi(u, u)$. Then, let

$$\phi(u, v) = \frac{1}{2}[\psi(u, v) + \psi(v, u)]$$

Certainly ϕ is a bilinear form and symmetric. Further, $\phi(u,u) = \psi(u,u) = Q(u)$. So there exists a symmetric bilinear form ϕ such that $Q(u) = \phi(u,u)$, so it suffices to prove uniqueness. Let ϕ be a symmetric bilinear form such that for all $u \in V$ we have $Q(u) = \phi(u,u)$. Then, we can find

$$Q(u+v) = \phi(u+v,u+v) = \phi(u,u) + \phi(v,v) + 2\phi(u,v)$$

Thus $\phi(u,v)$ is defined uniquely by Q, since

$$2\phi(u,v) = Q(u+v) - Q(u) - Q(v)$$

So ϕ is unique (when 2 is invertible in F). This identity for $\phi(u,v)$ is known as the polarisation identity. □

9.3 Diagonalisation of symmetric bilinear forms

Theorem. Let $\phi : V \times V \to F$ be a symmetric bilinear form, where V is finite-dimensional. Then there exists a basis B of V such that $[\phi]_B$ is diagonal.

Proof. By induction on the dimension, suppose the theorem holds for all dimensions less than n for $n \geq 2$. If $\phi(u,u) = 0$ for all $u \in V$, then $\phi = 0$ by the polarisation identity, which is diagonal. Otherwise $\phi(e_1,e_1) \neq 0$ for some $e_1 \in V$. Let

$$U = (\langle e_1 \rangle)^{\perp} = \{v \in V : \phi(e_1,v) = 0\}$$

This is a vector subspace of V, which is in particular

$$\ker\{\phi(e_1,\cdot) : V \to F\}$$

By the rank-nullity theorem, $\dim U = n-1$. We now claim that $U + \langle e_1 \rangle$ is a direct sum. Indeed, for $v = \langle e_1 \rangle \cap U$, we have $v = \lambda e_1$ and $\phi(e_1,v) = 0$. Hence $\lambda = 0$, since by assumption $\phi(e_1,e_1) \neq 0$. So we find a basis $B' = (e_2, \dots, e_n)$ of U, which we extend by e_1 to $B = (e_1, e_2, \dots, e_n)$. Since $U \oplus \langle e_1 \rangle$ has dimension n, this is a basis of V. Under this basis, we find

$$[\phi]_B = \begin{pmatrix} \phi(e_1,e_1) & 0 \\ 0 & [\phi|_U]_{B'} \end{pmatrix}$$

because

$$\phi(e_1,e_j) = \phi(e_j,e_1) = 0$$

for all $j \geq 2$. By the inductive hypothesis we can take a basis B' such that the restricted ϕ to be diagonal, so $[\phi]_B$ is diagonal in this basis. □

Example. Let $V = \mathbb{R}^3$ and choose the canonical basis (e_i). Let

$$Q(x_1,x_2,x_3) = x_1^2 + x_2^2 + 2x_3^2 + 2x_1x_2 + 2x_1x_3 - 2x_2x_3$$

Then, if $Q(x_1,x_2,x_3) = x^\intercal A x$, we have

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

Note that the off-diagonal terms are halved from their coefficients since in the expansion of $x^\intercal Ax$ they are included twice. Then, we can find a basis in which A is diagonal. We could use the above algorithm to find a basis, or complete the square in each component. We can write

$$Q(x_1, x_2, x_3) = (x_1 + x_2 + x_3)^2 + x_3^2 - 4x_2x_3 = (x_1 + x_2 + x_3)^2 + (x_3 - 2x_2)^2 - (2x_2)^2$$

This yields a new coordinate basis x_1', x_2', x_3'. Then $P^{-1}AP$ is diagonal. P is given by

$$\begin{pmatrix} x_1' \\ x_2' \\ x_3' \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & 1 & 1 \\ 0 & -2 & 1 \\ 0 & -2 & 0 \end{pmatrix}}_{P^{-1}} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

9.4 Sylvester's law

Corollary. If $F = \mathbb{C}$, for any symmetric bilinear form ϕ there exists a basis of V such that $[\phi]_B$ is

$$\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}$$

Proof. Since any symmetric bilinear form ϕ in a finite-dimensional F-vector space V can be diagonalised, let $E = (e_1, \dots, e_n)$ such that $[\phi]_E$ is diagonal with diagonal entries a_i. Order the a_i such that a_i is nonzero for $1 \leq i \leq r$, and the remaining values (if any) are zero. For $i \leq r$, let $\sqrt{a_i}$ be a choice of a complex root for a_i. Then $v_i = \frac{e_i}{\sqrt{a_i}}$ for $i \leq r$ and $v_i = e_i$ for $i > r$ gives the basis B as required. □

Corollary. Every symmetric matrix of $M_n(\mathbb{C})$ is congruent to a unique matrix of the form

$$\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}$$

where r is the rank of the matrix.

Corollary. Let $F = \mathbb{R}$, and let V be a finite-dimensional $\mathbb{R}$-vector space. Let ϕ be a symmetric bilinear form on V. Then there exists a basis $B = (v_1, \dots, v_n)$ of V such that

$$[\phi]_B = \begin{pmatrix} I_p & 0 & 0 \\ 0 & -I_q & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

for some integers p, q.

Proof. Since square roots do not necessarily exist in $\mathbb{R}$, we cannot use the form above. We first diagonalise the bilinear form in some basis E. Then, reorder and group the a_i into a positive group of size

p, a negative group of size q, and a zero group. Then,

$$v_i = \begin{cases} \frac{e_i}{\sqrt{a_i}} & i \in \{1, \dots, p\} \\ \frac{e_i}{\sqrt{-a_i}} & i \in \{p+1, \dots, p+q\} \\ e_i & i \in \{p+q+1, \dots, n\} \end{cases}$$

This gives a new basis as required. □

Definition. Let $F = \mathbb{R}$. The *signature* of a bilinear form ϕ is

$$s(\phi) = p - q$$

where p and q are defined as in the corollary above.

Theorem. Let $F = \mathbb{R}$. Let V be a finite-dimensional $\mathbb{R}$-vector space. If a real symmetric bilinear form is represented by some matrix

$$\begin{pmatrix} I_p & 0 & 0 \\ 0 & -I_q & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

in some basis B, and some other matrix

$$\begin{pmatrix} I_{p'} & 0 & 0 \\ 0 & -I_{q'} & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

in another basis B', then $p = p'$ and $q = q'$. Thus, the signature of the matrix is well defined.

Definition. Let ϕ be a symmetric bilinear form on a real vector space V. We say that

(i) ϕ is *positive definite* if $\phi(u, u) > 0$ for all nonzero $u \in V$;
(ii) ϕ is *positive semidefinite* if $\phi(u, u) \geq 0$ for all $u \in V$;
(iii) ϕ is *negative definite* or *negative semidefinite* if $\phi(u, u) < 0$ or $\phi(u, u) \leq 0$ respectively for all nonzero $u \in V$.

Example. The matrix

$$\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}$$

is positive definite for $r = n$, and positive semidefinite for $r < n$.

We now prove Sylvester's law.

Proof. In order to prove uniqueness of p, we will characterise the matrix in a way that does not depend on the basis. In particular, we will show that p is the largest dimension of a vector subspace of V such that the restriction of ϕ on this subspace is positive definite. Suppose we have $B = (v_1, \dots, v_n)$ and

$$[\phi]_B = \begin{pmatrix} I_p & 0 & 0 \\ 0 & -I_q & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

We consider

$$X = \langle v_1, \dots, v_p \rangle$$

Then we can easily compute that $\phi|_X$ is positive definite. Let

$$Y = \langle v_{p+1}, \dots, v_n \rangle$$

Then, as above, $\phi|_Y$ is negative semidefinite. Suppose that ϕ is positive definite on another subspace X'. In this case, $Y \cap X' = \{0\}$, since if $y \in Y \cap X'$ we must have $Q(y) \leq 0$, but since $y \in X'$ we have $y = 0$. Thus, $Y + X' = Y \oplus X'$, so $n = \dim V \geq \dim Y + \dim X'$. But $\dim Y = n - p$, so $\dim X' \leq p$. The same argument can be executed for q, hence both p and q are independent of basis. □

9.5 Kernels of bilinear forms

Definition. Let $K = \{v \in V : \forall u \in V, \phi(u, v) = 0\}$. This is the *kernel* of the bilinear form.

Remark. By the rank-nullity theorem,

$$\dim K + \operatorname{rank} \phi = n$$

Using the above notation, we can show that there exists a subspace T of dimension $n - (p + q) + \min\{p, q\}$ such that $\phi|_T = 0$. Indeed, let $B = (v_1, \dots, v_n)$ such that

$$[\phi]_B = \begin{pmatrix} I_p & 0 & 0 \\ 0 & -I_q & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

The quadratic form has a zero subspace of dimension $n - (p + q)$ in the bottom right. But by setting

$$T = \{v_1 + v_{p+1}, \dots, v_q + v_{p+q}, v_{p+q+1}, \dots, v_n\}$$

we can combine the positive and negative blocks (assuming here that $p \geq q$) to produce more linearly independent elements of the kernel. In particular, $\dim T$ is the largest possible dimension of a subspace T' of V such that $\phi|_{T'} = 0$.

9.6 Sesquilinear forms

Let $F = \mathbb{C}$. The standard inner product on $\mathbb{C}^n$ is defined to be

$$\left\langle \begin{pmatrix} x_1 \\ \vdots \\ v_n \end{pmatrix}, \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \right\rangle = \sum_{i=1}^n x_i \overline{y}_i$$

This is not a bilinear form on $\mathbb{C}$ due to the complex conjugate, it is linear in the first entry.

Definition. Let V, W be $\mathbb{C}$-vector spaces. A form $\phi : V \times W \to \mathbb{C}$ is called *sesquilinear* if it is linear in the first entry, and

$$\phi(v, \lambda_1 w_1 + \lambda_2 w_2) = \overline{\lambda_1}\phi(v, w_1) + \overline{\lambda_2}\phi(v, w_2)$$

so it is antilinear with respect to the second entry.

Lemma. Let $B = (v_1, \dots, v_m)$ be a basis of V and $C = (w_1, \dots, w_n)$ be a basis of W. Let $[\phi]_{B,C} = \left(\phi(v_i, w_j)\right)$. Then,

$$\phi(v, w) = [v]_B^\intercal [\phi]_{B,C} \overline{[w]_C}$$

Proof. Let B, B' be bases of V and C, C' be bases of W. Let $P = [I]_{B',B}$ and $Q = [I]_{C',C}$. Then

$$[\phi]_{B',C'} = P^\intercal [\phi]_{B,C} \overline{Q}$$

□

9.7 Hermitian forms

Definition. Let V be a finite-dimensional $\mathbb{C}$-vector space. Let ϕ be a sesquilinear form on V. Then ϕ is *Hermitian* if, for all $u, v \in V$,

$$\phi(u, v) = \overline{\phi(v, u)}$$

Remark. If ϕ is Hermitian, then $\phi(u, u) = \overline{\phi(u, u)} \in \mathbb{R}$. Further, $\phi(\lambda u, \lambda u) = |\lambda|^2 \phi(u, u)$. This allows us to define positive and negative definite Hermitian forms.

Lemma. A sesquilinear form $\phi : V \times V \to \mathbb{C}$ is Hermitian if and only if, for any basis B of V,

$$[\phi]_B = [\phi]_B^\dagger$$

Proof. Let $A = [\phi]_B = (a_{ij})$. Then $a_{ij} = \phi(e_i, e_j)$, and $a_{ji} = \phi(e_j, e_i) = \overline{\phi(e_i, e_j)} = \overline{a_{ij}}$. So $\overline{A}^\intercal = A$. Conversely suppose that $[\phi]_B = A = \overline{A}^\intercal$. Now let

$$u = \sum_{i=1}^n \lambda_i e_i; \quad v = \sum_{i=1}^n \mu_i e_i$$

Then,

$$\phi(u, v) = \phi\left(\sum_{i=1}^n \lambda_i e_i, \sum_{i=1}^n \mu_i e_i\right) = \sum_{i=1}^n \sum_{j=1}^n \lambda_i \overline{\mu_j} a_{ij}$$

Further,

$$\overline{\phi(v, u)} = \overline{\phi\left(\sum_{i=1}^n \mu_i e_i, \sum_{i=1}^n \lambda_i e_i\right)} = \sum_{i=1}^n \sum_{j=1}^n \overline{\mu_j \overline{\lambda_i}}\, \overline{a_{ij}}$$

which is equivalent. Hence ϕ is Hermitian. □

9.8 Polarisation identity

A Hermitian form ϕ on a complex vector space V is entirely determined by a quadratic form $Q : V \to \mathbb{R}$ such that $v \mapsto \phi(v, v)$ by the formula

$$\phi(u, v) = \frac{1}{4}[Q(u + v) - Q(u - v) + iQ(u + iv) - iQ(u - iv)]$$

9.9 Hermitian formulation of Sylvester's law

Theorem. Let V be a finite-dimensional $\mathbb{C}$-vector space. Let $\phi : V \times V \to \mathbb{C}$ be a Hermitian form on V. Then there exists a basis $B = (v_1, \dots, v_n)$ of V such that

$$[\phi]_B = \begin{pmatrix} I_p & 0 & 0 \\ 0 & -I_q & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

where p, q depend only on ϕ and not B.

Proof. The following is a sketch proof; it is nearly identical to the case of real symmetric bilinear forms. If $\phi = 0$, existence is trivial. Otherwise, using the polarisation identity there exists $e_1 \neq 0$ such that $\phi(e_1, e_1) \neq 0$. Let

$$v_1 = \frac{e_1}{\sqrt{|\phi(e_1, e_1)|}} \implies \phi(v_1, v_1) = \pm 1$$

Consider the orthogonal space $W = \{w \in V : \phi(v_1, w) = 0\}$. We can check, arguing analogously to the real case, that $V = \langle v_1 \rangle \oplus W$. Hence, we can inductively diagonalise ϕ.

p, q are unique. Indeed, we can prove that p is the maximal dimension of a subspace on which ϕ is positive definite (which is well-defined since $\phi(u, u) \in \mathbb{R}$). The geometric interpretation of q is similar. □

9.10 Skew-symmetric forms

Definition. Let V be a finite-dimensional $\mathbb{R}$-vector space. Let ϕ be a bilinear form on V. Then ϕ is *skew-symmetric* if, for all $u, v \in V$,

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

Remark. $\phi(u, u) = -\phi(u, u) = 0$. Also, in any basis B of V, we have $[\phi]_B = -[\phi]_B^\intercal$. Any real matrix can be decomposed as the sum

$$A = \frac{1}{2}(A + A^\intercal) + \frac{1}{2}(A - A^\intercal)$$

where the first summand is symmetric and the second is skew-symmetric.

9.11 Skew-symmetric formulation of Sylvester's law

Theorem. Let V be a finite-dimensional $\mathbb{R}$-vector space. Let $\phi\colon V \times V \to \mathbb{R}$ be a skew-symmetric form on V. Then there exists a basis

$$B = (v_1, w_1, v_2, w_2, \dots, v_m, w_m, v_{2m+1}, v_{2m+2}, \dots, v_n)$$

of V such that

$$[\phi]_B = \begin{pmatrix} 0 & 1 & & & & \\ -1 & 0 & & & & \\ & & 0 & 1 & & \\ & & -1 & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{pmatrix}$$

Corollary. Skew-symmetric matrices have an even rank.

Proof. This is again very similar to the previous case. We will perform an inductive step on the dimension of V. If $\phi \neq 0$, there exist v_1, w_1 such that $\phi_1(v_1, w_1) \neq 0$. After scaling one of the vectors, we can assume $\phi(v_1, w_1) = 1$. Since ϕ is skew-symmetric, $\phi(w_1, v_1) = -1$. Then v_1, w_1 are linearly independent; if they were linearly dependent we would have $\phi(v_1, w_1) = \phi(v_1, \lambda v_1) = 0$. Let $U = \langle v_1, w_1 \rangle$ and let $W = \{v \in V : \phi(v_1, v) = \phi(w_1, v) = 0\}$ and we can show $V = U \oplus W$. Then induction gives the required result. $\square$

10 Inner product spaces

10.1 Definition

Definition. Let V be a vector space over $\mathbb{R}$ or $\mathbb{C}$. A *scalar product* or *inner product* is a positive-definite symmetric (respectively Hermitian) bilinear form ϕ on V. We write

$$\phi(u, v) = \langle u, v \rangle$$

V, when equipped with this inner product, is called a real (respectively complex) *inner product space*.

Example. In $\mathbb{C}^n$, we define

$$\langle x, y \rangle = \sum_{i=1}^n x_i \overline{y}_i$$

Example. Let $V = C^0([0,1], \mathbb{C})$. Then we can define

$$\langle f, g \rangle = \int_0^1 f(t)\overline{g}(t)\,\mathrm{d}t$$

This is the L^2 scalar product.

Example. Let $\omega : [0,1] : \mathbb{R}_+^\star$ where $\mathbb{R}_+^\star = \mathbb{R}_+ \setminus \{0\}$ and define

$$\langle f, g \rangle = \int_0^1 f(t)\overline{g}(t)w(t)\,\mathrm{d}t$$

Remark. Typically it suffices to check $\langle u, u \rangle = 0 \implies u = 0$ since linearity and positivity are usually trivial.

Definition. Let V be an inner product space. Then for $v \in V$, the *norm* of v induced by the inner product is defined by

$$\|v\| = (\langle v, v \rangle)^{1/2}$$

This is real, and positive if $v \neq 0$.

10.2 Cauchy–Schwarz inequality

Lemma. For an inner product space,

$$|\langle u, v \rangle| \leq \|a\| \cdot \|b\|$$

Proof. Let $t \in F$. Then,

$$0 \leq \|tu - v\| = \langle tu - v, tu - v \rangle = t\bar{t} \langle u, u \rangle - u \langle u, v \rangle - \bar{t} \langle v, u \rangle + \|v\|^2$$

Since the inner product is Hermitian,

$$0 \leq |t|^2 \|u\|^2 + \|v\|^2 - 2\,\mathrm{Re}(t \langle u, v \rangle)$$

By choosing

$$t = \frac{\overline{\langle u, v \rangle}}{\|u\|^2}$$

we have

$$0 \leq \frac{|\langle u, v \rangle|^2}{\|u\|^2} + \|v\|^2 - 2\,\mathrm{Re}\left(\frac{|\langle u, v \rangle|^2}{\|u\|^2}\right)$$

Since the term under the real part operator is real, the result holds. □

Note that equality implies collinearity in the Cauchy–Schwarz inequality.

Corollary (triangle inequality). In an inner product space,

$$\|u + v\| \leq \|u\| + \|v\|$$

Proof. We have

$$\|u + v\|^2 = \langle u + v, u + v \rangle = \left\|u^2\right\| + 2\,\mathrm{Re}(\langle u, v \rangle) + \|v\|^2 \leq \left\|u^2\right\| + \|v\|^2 + 2\|u\| \cdot \|v\| = (\|u\| + \|v\|)^2$$

□

Remark. Any inner product induces a norm, but not all norms derive from scalar products.

10.3 Orthogonal and orthonormal sets

Definition. A set $(e_1, \dots, e_k)$ of vectors of V is said to be *orthogonal* if $\langle e_i, e_j \rangle = 0$ for all $i \neq j$. The set is said to be *orthonormal* if it is orthogonal and $\|e_i\| = 1$ for all i. In this case, $\langle e_i, e_j \rangle = \delta_{ij}$.

Lemma. If $(e_1, \dots, e_k)$ are orthogonal and nonzero, then they are linearly independent. Further, let $v \in \langle \{e_i\} \rangle$. Then,

$$v = \sum_{j=1}^{k} \lambda_j e_j \implies \lambda_j = \frac{\langle v, e_j \rangle}{\left\| e_j \right\|^2}$$

Proof. Suppose

$$\sum_{i=1}^{k} \lambda_i e_i = 0$$

Then,

$$0 = \left\langle \sum_{i=1}^{k} \lambda_i, e_j \right\rangle \implies 0 = \sum_{i=1}^{k} \lambda_i \langle e_i, e_j \rangle$$

Thus $\lambda_j = 0$ for all j. Further, for v in the span of these vectors,

$$\langle v, e_j \rangle = \sum_{i=1}^{k} \lambda_i \langle e_i, e_j \rangle = \lambda_j \left\| e_j \right\|^2$$

□

10.4 Parseval's identity

Corollary. Let V be a finite-dimensional inner product space. Let $(e_1, \dots, e_n)$ be an orthonormal basis. Then, for any vectors $u, v \in V$, we have

$$\langle u, v \rangle = \sum_{i=1}^{n} \langle u, e_i \rangle \overline{\langle v, e_i \rangle}$$

Hence,

$$\|u\|^2 = \sum_{i=1}^{n} |\langle u, e_i \rangle|^2$$

Proof. By orthonormality,

$$u = \sum_{i=1}^{n} \langle u, e_i \rangle e_i; \quad v = \sum_{i=1}^{n} \langle v, e_i \rangle e_i$$

Hence, by sesquilinearity,

$$\langle u, v \rangle = \sum_{i=1}^{n} \langle u, e_i \rangle \overline{\langle v, e_i \rangle}$$

By taking $u = v$ we find

$$\|u\|^2 = \langle u, u \rangle = \sum_{i=1}^{n} |\langle u, e_i \rangle|^2$$

□

10.5 Gram–Schmidt orthogonalisation process

Theorem. Let V be an inner product space. Let $(v_i)_{i \in I}$ be a linearly independent family of vectors such that I is countable. Then there exists a family $(e_i)_{i \in I}$ of orthonormal vectors such that for all $k \geq 1$,

$$\langle v_1, \dots, v_k \rangle = \langle e_1, \dots, e_k \rangle$$

Proof. This proof is an explicit algorithm to compute the family (e_i), which will be computed by induction on k. For $k = 1$, take $e_1 = \frac{v_1}{\|v_1\|}$. Inductively, suppose $(e_1, \dots, e_k)$ satisfy the conditions as above. Then we will find a valid e_{k+1}. We define

$$e'_{k+1} = v_{k+1} - \sum_{i=1}^{k} \langle v_{k+1}, e_i \rangle e_i$$

This ensures that the inner product between e'_{k+1} and any basis vector e_j is zero, while maintaining the same span. Suppose $e'_{k+1} = 0$. Then, $v_{k+1} \in \langle e_1, \dots, e_k \rangle = \langle v_1, \dots, v_k \rangle$ which contradicts the fact that the family is free. Thus,

$$e_{k+1} = \frac{e'_{k+1}}{\left\| e'_{k+1} \right\|}$$

satisfies the requirements. □

Corollary. In finite-dimensional inner product spaces, there always exists an orthonormal basis. In particular, any orthonormal set of vectors can be extended into an orthonormal basis.

Remark. Let $A \in M_n(\mathbb{R})$ be a real-valued (or complex-valued) matrix. Then, the column vectors of A are orthogonal if $A^\intercal A = I$ (or $A^\intercal \overline{A} = I$ in the complex-valued case).

10.6 Orthogonality of matrices

Definition. A matrix $A \in M_n(\mathbb{R})$ is *orthogonal* if $A^\intercal A = I$, hence $A^\intercal = A^{-1}$. A matrix $A \in M_n(\mathbb{C})$ is *unitary* if $A^\intercal \overline{A} = I$, hence $A^\dagger = A^{-1}$.

Proposition. Let A be a square, non-singular, real-valued (or complex-valued) matrix. Then A can be written as $A = RT$ where T is upper triangular and R is orthogonal (or respectively unitary).

Proof. We apply the Gram–Schmidt process to the column vectors of the matrix. This gives us an orthonormal set of vectors, which gives an upper triangular matrix in this new basis. □

10.7 Orthogonal complement and projection

Definition. Let V be an inner product space. Let $V_1, V_2 \leq V$. Then we say that V is the *orthogonal direct sum* of V_1 and V_2 if $V = V_1 \oplus V_2$ and for all vectors $v_1 \in V_1, v_2 \in V_2$ we have $\langle v_1, v_2 \rangle = 0$. When this holds, we write $V = V_1 \overset{\perp}{\oplus} V_2$.

Remark. If for all vectors v_1, v_2 we have $\langle v_1, v_2 \rangle = 0$, then $v \in V_1 \cap V_2 \implies \|v\|^2 = 0 \implies v = 0$. Hence the sum is always direct if the subspaces are orthogonal.

Definition. Let V be an inner product space and let $W \leq V$. We define the *orthogonal* of W to be

$$W^\perp = \{v \in V : \forall w \in W, \langle v, w \rangle = 0\}$$

Lemma. For any inner product space V and any subspace $W \leq V$, we have $V = W \overset{\perp}{\oplus} W^\perp$.

Proof. First note that $W^\perp \leq V$. Then, if $w \in W, w \in W^\perp$, we have

$$\|w\|^2 = \langle w, w \rangle = 0$$

since they are orthogonal, so the vector subspaces intersect only in the zero vector. Now, we need to show $V = W + W^\perp$. Let $(e_1, \dots, e_k)$ be an orthonormal basis of W and extend it into $(e_1, \dots, e_k, e_{k+1}, \dots, e_n)$ which can be made orthonormal. Then, $(e_{k+1}, \dots, e_n)$ are elements of $W^\perp$ and form a basis. □

10.8 Projection maps

Definition. Suppose $V = U \oplus W$, so U is a complement of W in V. Then, we define $\pi : V \to W$ which maps $v = u + w$ to w. This is well defined, since the sum is direct. π is linear, and $\pi^2 = \pi$. We say that π is the *projection* operator onto W.

Remark. The map $\iota - \pi$ is the projection onto U, where ι is the identity map.

Lemma. Let V be an inner product space. Let $W \leq V$ be a finite-dimensional subspace. Let $(e_1, \dots, e_k)$ be an orthonormal basis for W. Then,

(i) $\pi(v) = \sum_{i=1}^k \langle v, e_i \rangle e_i$; and
(ii) for all $v \in V, w \in W$, $\|v - \pi(v)\| \leq \|v - w\|$ with equality if and only if $w = \pi(v)$, hence

$\pi(v)$ is the point in W closest to v.

Proof. We define $\pi(v) = \sum_{i=1}^k \langle v, e_i \rangle e_i$. Since $W = \langle \{e_k\} \rangle$, $\pi(v) \in W$ for all $v \in V$. Then, $v = (v - \pi(v)) + \pi(v)$ has a term in W. We claim that the remaining term is in the orthogonal; $v - \pi(v) \in W^\perp$. Indeed, we must show $\langle v - \pi(v), w \rangle = 0$ for all $w \in W$. Equivalently, $\langle v - \pi(v), e_i \rangle = 0$ for all basis vectors e_i of W. We can explicitly compute

$$\left\langle v - \pi(v), e_j \right\rangle = \left\langle v, e_j \right\rangle - \left\langle \sum_{i=1}^k \langle v, e_i \rangle e_i, e_j \right\rangle = \left\langle v, e_j \right\rangle - \sum_{i=1}^k \langle v, e_i \rangle \left\langle e_i, e_j \right\rangle = \left\langle v, e_j \right\rangle - \left\langle v, e_j \right\rangle = 0$$

Hence, $v = (v - \pi(v)) + \pi(v)$ is a decomposition into W and $W^\perp$. Since $W \cap W^\perp = \{0\}$, we have $V = W \overset{\perp}{\oplus} W^\perp$. For the second part, let $v \in V$, $w \in W$, and we compute

$$\|v - w\|^2 = \left\| \underbrace{v - \pi(v)}_{\in W^\perp} + \underbrace{\pi(v) - w}_{\in W} \right\|^2 = \|v - \pi(v)\|^2 + \|\pi(v) - w\|^2 \geq \|v - \pi(v)\|^2$$

with equality if and only if $w = \pi(v)$. □

10.9 Adjoint maps

Definition. Let V, W be finite-dimensional inner product spaces. Let $\alpha \in L(V, W)$. Then there exists a unique linear map $\alpha^\star : W \to V$ such that for all $v, w \in V, W$,

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

Moreover, if B is an orthonormal basis of V, and C is an orthonormal basis of W, then

$$[\alpha^\star]_{C,B} = \left(\overline{[\alpha]_{B,C}} \right)^\intercal$$

Proof. Let $B = (v_1, \dots, v_n)$ and $C = (w_1, \dots, w_m)$ and $A = [\alpha]_{B,C} = (a_{ij})$. To check existence, we define $[\alpha^\star]_{C,B} = \overline{A}^\intercal = (c_{ij})$ and explicitly check the definition. By orthogonality,

$$\left\langle \alpha\left(\sum \lambda_i v_i\right), \sum \mu_j w_j \right\rangle = \left\langle \sum_{i,k} \lambda_i a_{ki} w_k, \sum_j \mu_j w_j \right\rangle = \sum_{i,j} \lambda_i a_{ji} \overline{\mu_j}$$

Then,

$$\left\langle \sum \lambda_i v_i, \alpha^\star\left(\sum \mu_j w_j\right) \right\rangle = \left\langle \sum_i \lambda_i v_i, \sum_{j,k} \mu_j c_{kj} v_k \right\rangle = \sum_{i,j} \lambda_i \overline{c_{ij} \mu_j}$$

So equality requires $\overline{c_{ij}} = a_{ji}$. Uniqueness follows from the above; the expansions are equivalent for any vector if and only if $\overline{c_{ij}} = a_{ji}$. □

Remark. The same notation, $\alpha^\star$, is used for the adjoint as just defined, and the dual map as defined before. If V, W are real product inner spaces and $\alpha \in L(V, W)$, we define $\psi : V \to V^\star$ such that $\psi(v)(x) = \langle x, v \rangle$ and similarly for W. Then we can check that the adjoint for α is given by the composition of ψ from $V \to V^\star$, then applying the dual, then applying the inverse of ψ for W.

10.10 Self-adjoint and isometric maps

Definition. Let V be a finite-dimensional inner product space, and α be an endomorphism of V. Let $\alpha^\star \in L(V)$ be the adjoint map. Then,

(i) the condition $\langle \alpha v, w\rangle = \langle v, \alpha w\rangle$ is equivalent to the condition $\alpha = \alpha^\star$, and such an α is called *self-adjoint* (for $\mathbb{R}$ we call such endomorphisms *symmetric*, and for $\mathbb{C}$ we call such endomorphisms Hermitian);

(ii) the condition $\langle \alpha v, \alpha w\rangle = \langle v, w\rangle$ is equivalent to the condition $\alpha^\star = \alpha^{-1}$, and such an α is called an *isometry* (for $\mathbb{R}$ it is called *orthogonal*, and for $\mathbb{C}$ it is called *unitary*).

Proposition. The conditions for isometries defined as above are equivalent.

Proof. Suppose $\langle \alpha v, \alpha w\rangle = \langle v, w\rangle$. Then for $v = w$, we find $\|\alpha v\|^2 = \|v\|^2$, so α preserves the norm. In particular, this implies $\ker \alpha = \{0\}$. Since α is an endomorphism and V is finite-dimensional, α is bijective. Then for all $v, w \in V$,

$$\langle v, \alpha^\star(w)\rangle = \langle \alpha v, w\rangle = \left\langle \alpha v, \alpha\left(\alpha^{-1}(w)\right)\right\rangle = \left\langle v, \alpha^{-1}(w)\right\rangle$$

Hence $\alpha^\star = \alpha^{-1}$. Conversely, if $\alpha^\star = \alpha^{-1}$ we have

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

as required. □

Remark. Using the polarisation identity, we can show that α is isometric if and only if for all $v \in V$, $\|\alpha(v)\| = \|v\|$.

Lemma. Let V be a finite-dimensional real (or complex) inner product space. Then for $\alpha \in L(V)$,

(i) α is self-adjoint if and only if for all orthonormal bases B of V, we have $[\alpha]_B$ is symmetric (or Hermitian);

(ii) α is an isometry if and only if for all orthonormal bases B of V, we have $[\alpha]_B$ is orthogonal (or unitary).

Proof. Let B be an orthonormal basis for V. Then we know $[\alpha^\star]_B = [\alpha]_B^\dagger$. We can then check that $[\alpha]_B^\dagger = [\alpha]_B$ and $[\alpha]_B^\dagger = [\alpha]_B^{-1}$ respectively. □

Definition. For $F = \mathbb{R}$, we define the *orthogonal group* of V by

$$O(V) = \{\alpha \in L(V) : \alpha \text{ is an isometry}\}$$

Note that $O(V)$ is bijective with the set of orthogonal bases of V. For $F = \mathbb{C}$, we define the *unitary group* of V by

$$U(V) = \{\alpha \in L(V) : \alpha \text{ is an isometry}\}$$

Again, note that $U(V)$ is bijective with the set of orthogonal bases of V.

10.11 Spectral theory for self-adjoint maps

Spectral theory is the study of the spectrum of operators. Recall that in finite-dimensional inner product spaces V, W, $\alpha \in L(V, W)$ yields the adjoint $\alpha^\star \in L(W, V)$ such that for all $v \in V, w \in W$, we have $\langle \alpha(v), w \rangle = \langle v, \alpha^\star(w) \rangle$.

Lemma. Let V be a finite-dimensional inner product space. Let $\alpha \in L(V)$ be a self-adjoint endomorphism. Then α has real eigenvalues, and eigenvectors of α with respect to different eigenvalues are orthogonal.

Proof. Suppose $\lambda \in \mathbb{C}$, $v \in V$ nonzero such that $\alpha(v) = \lambda v$. Then, $\langle \lambda v, v \rangle = \lambda \|v\|^2$ and also

$$\langle \alpha v, v \rangle = \langle v, \alpha v \rangle = \langle v, \lambda v \rangle = \overline{\lambda} \|v\|^2$$

Hence $\lambda = \overline{\lambda}$ since $v \neq 0$. Now, suppose $\mu \neq \lambda$ and $w \in V$ nonzero such that $\alpha(w) = \mu w$. Then,

$$\lambda \langle v, w \rangle = \langle \alpha v, w \rangle = \langle v, \alpha w \rangle = \overline{\mu} \langle v, w \rangle = \mu \langle v, w \rangle$$

So if $\lambda \neq \mu$ we must have $\langle v, w \rangle = 0$. □

Theorem (spectral theorem for self-adjoint maps)**.** Let V be a finite-dimensional inner product space. Let $\alpha \in L(V)$ be self-adjoint. Then V has an orthonormal basis of eigenvectors of α. Hence α is diagonalisable in an orthonormal basis.

Proof. We will consider induction on the dimension of V. Suppose $A = [\alpha]_B$ with respect to the fundamental basis B. By the fundamental theorem of algebra, we know that $\chi_A(\lambda)$ has a (complex) root. But since λ is an eigenvalue of α and α is self-adjoint, $\lambda \in \mathbb{R}$. Now, we choose an eigenvector $v_1 = V \setminus \{0\}$ such that $\alpha(v_1) = \lambda v_1$. We can set $\|v_1\| = 1$ by linearity. Let $U = \langle v_1 \rangle^\perp \leq V$. We then observe that U is stable by α; $\alpha(U) \leq U$. Indeed, let $u \in U$. Then $\langle \alpha(u), v_1 \rangle = \langle u, \alpha(v_1) \rangle = \lambda \langle u, v_1 \rangle = 0$ by orthogonality. Hence $\alpha(u) \in U$. We can then restrict α to the domain U, and by induction we can then choose an orthonormal basis of eigenvectors for U. Since $V = \langle v_1 \rangle \overset{\perp}{\oplus} U$ we have an orthonormal basis of eigenvectors for V when including v_1. □

Corollary. Let V be a finite-dimensional inner product space. Let $\alpha \in L(V)$ be self-adjoint. Then V is the orthogonal direct sum of the eigenspaces of α.

10.12 Spectral theory for unitary maps

Lemma. Let V be a complex inner product space. Let α be unitary, so $\alpha^\star = \alpha^{-1}$. Then all eigenvalues of α have unit norm. Eigenvectors corresponding to different eigenvalues are orthogonal.

Proof. Let $\lambda \in \mathbb{C}, v \in V \setminus \{0\}$ such that $\alpha(v) = \lambda v$. First, $\lambda \neq 0$ since α is invertible, and in particular $\ker \alpha = \{0\}$. Since $v = \lambda\alpha^{-1}(v)$, we can compute

$$\lambda \langle v, v \rangle = \langle \lambda v, v \rangle = \langle \alpha v, v \rangle = \langle v, \alpha^{-1} v \rangle = \left\langle v, \frac{1}{\lambda} v \right\rangle = \frac{1}{\overline{\lambda}} \langle v, v \rangle$$

Hence $(\lambda\overline{\lambda} - 1)\|v\|^2 = 0$ giving $|\lambda| = 1$. Further, suppose $\mu \in \mathbb{C}$ and $w \in V \setminus \{0\}$ such that $\alpha(w) = \mu w, \lambda \neq \mu$. Then

$$\lambda \langle v, w \rangle = \langle \lambda v, w \rangle = \langle \alpha v, w \rangle = \langle v, \alpha^{-1} w \rangle = \left\langle v, \frac{1}{\mu} w \right\rangle = \frac{1}{\overline{\mu}} \langle v, w \rangle = \mu \langle v, w \rangle$$

since $\mu\overline{\mu} = 1$. $\square$

Theorem (spectral theorem for unitary maps)**.** Let V be a finite-dimensional complex inner product space. Let $\alpha \in L(V)$ be unitary. Then V has an orthonormal basis of eigenvectors of α. Hence α is diagonalisable in an orthonormal basis.

Proof. Let $A = [\alpha]_B$ where B is an orthonormal basis. Then $\chi_A(\lambda)$ has a complex root λ. As before, let $v_1 \neq 0$ such that $\alpha(v_1) = \lambda v_1$ and $\|v_1\| = 1$. Let $U = \langle v_1 \rangle^{\perp}$, and we claim that $\alpha(U) = U$. Indeed, let $u \in U$, and we find

$$\langle \alpha(u), v_1 \rangle = \langle u, \alpha^{-1}(v_1) \rangle = \left\langle u, \frac{1}{\lambda} v_1 \right\rangle = \frac{1}{\overline{\lambda}} \langle u, v_1 \rangle$$

Since $\langle u, v_1 \rangle = 0$, we have $\alpha(u) \in U$. Hence, α restricted to U is a unitary endomorphism of U. By induction we have an orthonormal basis of eigenvectors of α for U and hence for V. $\square$

Remark. We used the fact that the field is complex to find an eigenvalue. In general, a real-valued orthonormal matrix A giving $AA^\intercal = I$ cannot be diagonalised over $\mathbb{R}$. For example, consider

$$A = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

This is orthogonal and normalised. However, $\chi_A(\lambda) = 1 + 2\lambda\cos\theta + \lambda^2$ hence $\lambda = e^{\pm i\theta}$ which are complex in the general case.

10.13 Application to bilinear forms

We wish to extend the previous statements about spectral theory into statements about bilinear forms.

Corollary. Let $A \in M_n(\mathbb{R})$ (or $M_n(\mathbb{C})$) be a symmetric (or respectively Hermitian) matrix. Then there exists an orthonormal (respectively unitary) matrix P such that $P^\intercal AP$ (or $P^\dagger AP$) is diagonal with real-valued entries.

Proof. Using the standard inner product, $A \in L(F^n)$ is self-adjoint and hence there exists an orthonormal basis B of F^n such that A is diagonal in this basis. Let $P = (v_1, \dots, v_n)$ be the matrix of this basis. Since B is orthonormal, P is orthogonal (or unitary). The result follows from the fact that $P^{-1}AP$ is diagonal. The eigenvalues are real, hence the diagonal matrix is real. $\square$

Corollary. Let V be a finite-dimensional real (or complex) inner product space. Let $\phi : V \times V \to F$ be a symmetric (or Hermitian) bilinear form. Then, there exists an orthonormal basis B of V such that $[\phi]_B$ is diagonal.

Proof. $A^\intercal = A$ (or respectively $A^\dagger = A$), hence there exists an orthogonal (respectively unitary) matrix P such that $P^{-1}AP$ is diagonal. Let (v_i) be the ith row of $P^{-1} = P^\intercal$ (or $P^\dagger$). Then $(v_1, \dots, v_n)$ is an orthonormal basis B of V such that $[\phi]_V$ is this diagonal matrix. □

Remark. The diagonal entries of $P^{-1}AP$ are the eigenvalues of A. Moreover, we can define the signature $s(\phi)$ to be the difference between the number of positive eigenvalues of A and the number of negative eigenvalues of A.

10.14 Simultaneous diagonalisation

Corollary. Let V be a finite-dimensional real (or complex) vector space. Let ϕ, ψ be symmetric (or Hermitian) bilinear forms on V. Let ϕ be positive definite. Then there exists a basis $(v_1, \dots, v_n)$ of V with respect to which ϕ and ψ are represented with a diagonal matrix.

Proof. Since ϕ is positive definite, V equipped with ϕ is a finite-dimensional inner product space where $\langle u, v \rangle = \phi(u, v)$. Hence, there exists a basis of V in which ψ is represented by a diagonal matrix, which is orthonormal with respect to the inner product defined by ϕ. Then, ϕ in this basis is represented by the identity matrix given by $\phi(v_i, v_j) = \langle v_i, v_j \rangle = \delta_{ij}$, which is diagonal. □

Corollary. Let $A, B \in M_n(\mathbb{R})$ (or $\mathbb{C}$) which are symmetric (or Hermitian). Suppose for all $x \neq 0$ we have $x^\dagger A x > 0$, so A is positive definite. Then there exists an invertible matrix $Q \in M_n(\mathbb{R})$ (or $\mathbb{C}$) such that $Q^\intercal AQ$ (or $Q^\intercal A\overline{Q}$) and $Q^\intercal BQ$ (or $Q^\intercal B\overline{Q}$) are diagonal.

Proof. A induces a quadratic form $Q(x) = x^\dagger A x$ which is positive definite by assumption. Similarly, $\widetilde{Q}(x) = x^\dagger B x$ is induced by B. Then we can apply the previous corollary and change basis. □