Linear Algebra

Prof. J. Saxl

Michaelmas 2005

LaTeXed by Sebastian Pancratz

These notes are based on a course of lectures given by Prof. J. Saxl in Part IB of the Mathematical Tripos at the University of Cambridge in the academic year 2005–2006.

These notes have not been checked by Prof. J. Saxl and should not be regarded as official notes for the course. In particular, the responsibility for any errors is mine — please email Sebastian Pancratz (`sfp25`) with any comments or corrections.

Contents

Chapter 1

Vector spaces

Remark. Write $\mathbb{F}$ for the field $\mathbb{R}$ or $\mathbb{C}$. The crucial properties of $\mathbb{F}$ are as follows.

- $\mathbb{F}$ is an abelian group under $+$, with additive identity 0.
- $\mathbb{F} \setminus \{0\}$ is an abelian group under $\cdot$, with multiplicative identity 1.
- Multiplication is distributive over addition, $a(b+c) = ab + ac$ for all $a, b, c \in \mathbb{F}$.

Definition. The set V is a vector space over $\mathbb{F}$ if the following holds.

(A) V is an abelian group under an operation $+$.

(A0) $+ : V \to V$, $(v_1, v_2) \mapsto v_1 + v_2 \in V$;
(A1) For all $v_1, v_2, v_3 \in V$, $(v_1 + v_2) + v_3 = v_1 + (v_2 + v_3)$;
(A2) For all $v_1, v_2 \in V$, $v_1 + v_2 = v_2 + v_1$;
(A3) There exists $0 \in V$ such that $v + 0 = v$ for all $v \in V$.
(A4) For each $v \in V$ there exists $-v \in V$ such that $-v + v = 0$.

(B) There exists a multiplication $\cdot$ by scalars on V.

(B0) $\cdot : \mathbb{F} \times V \to V$, $(\lambda, v) \mapsto \lambda v$;
(B1) For all $\lambda \in \mathbb{F}, v_1, v_2 \in V$, $\lambda(v_1 + v_2) = \lambda v_1 + \lambda v_2$;
(B2) For all $\lambda_1, \lambda_2 \in \mathbb{F}, v \in V$, $(\lambda_1 + \lambda_2)v = \lambda_1 v + \lambda_2 v$;
(B3) For all $\lambda_1, \lambda_2 \in \mathbb{F}, v \in V$, $(\lambda_1 \lambda_2)v = \lambda_1(\lambda_2 v)$;
(B4) For all $v \in V$, $1v = v$.

Lemma 1.1. Let V be a vector space over $\mathbb{F}$ and let $v \in V, \lambda \in \mathbb{F}$. Then

(i) $0 \cdot v = 0$, $\lambda \cdot 0 = 0$;

(ii) $-v = (-1) \cdot v$;

(iii) if $\lambda v = 0$ then either $\lambda = 0$ or $v = 0$.

Proof. (i) $0 \cdot v = (0 + 0) \cdot v = 0 \cdot v + 0 \cdot v$, so $0 = 0 \cdot v$.

(ii) $-v + v = 0 = 0v = (-1 + 1)v = (-1)v + 1v = (-1)v + v$, so $-v = (-1)v$.

(iii) Let $\lambda v = 0$ and suppose $\lambda \neq 0$. Then λ^{-1} exists and $\lambda^{-1}(\lambda v) = \lambda^{-1}0 = 0$ but $\lambda^{-1}(\lambda v) = (\lambda^{-1}\lambda)v = 1v = v$ so $v = 0$. □

Example. (i) The space $\mathbb{F}^n$ of n-tuples with entries in $\mathbb{F}$.

(ii) Let X be any set. The set $\mathbb{F}^X$ of all functions $X \to \mathbb{F}$ is a vector space over $\mathbb{F}$ addition and multiplication defined pointwise.

Definition. Let V be a vector space over $\mathbb{F}$. A subset $U \subset V$ is a subspace of V, written $U \leq V$ if

- $0 \in U$;
- $u_1, u_2 \in U \implies u_1 + u_2 \in U$;
- $\lambda \in \mathbb{F}, u \in U \implies \lambda u \in U$.

Equivalently, $U \neq \emptyset$ and U is closed under linear combinations.

Lemma 1.2. If V is a vector space over $\mathbb{F}$ and $U \leq V$ then U is a vector space over $\mathbb{F}$ under the restriction of the operations $+$ and $\cdot$ on V to U.

Example. (i) $\mathbb{R}^{\mathbb{R}}$ is a vector space over $\mathbb{R}$. The set $C(\mathbb{R})$ of continuous real functions is a subspace, and hence a real vector space.

(ii) Similarly, one can also consider the subspaces $D(\mathbb{R})$ and $P(\mathbb{R})$) of differentiable and polynomial functions, respectively.

Remark. If $n \in \mathbb{N}_0$, $\lambda_1, \ldots, \lambda_n \in \mathbb{F}$, $v_1, \ldots, v_n \in V$ write $\sum_{i=1}^n \lambda_i v_i$ for the linear combination $\lambda_1 v_1 + \cdots + \lambda_n v_n$. By convention, $\sum_{i=1}^0 \lambda_i v_i = 0$. Also note that all linear combinations are finite. If $S \subset V$, the linear combination $\sum_{v \in S} \lambda_v v$ has only finitely many v such that $\lambda_v \neq 0$.

Definition. The vector $v_1, \ldots, v_n$ in V span (or generate) V over $\mathbb{F}$ if any $v \in V$ is a linear combination of $v_1, \ldots, v_n$. Write $V = \langle v_1, \ldots, v_n \rangle$.

More generally, if $S \subset V$ then S spans V if

$$\forall v \in V \quad \exists n \in \mathbb{N}_0 \quad \exists v_1, \ldots, v_n \in V \quad \exists \lambda_1, \ldots, \lambda_n \in \mathbb{F} \quad v = \sum_{i=1}^{n} \lambda_i v_i.$$

Example. (i) $P_2(\mathbb{R})$ is spanned by $1, x, x^2$.

(ii) $P(\mathbb{R})$ is not finite-dimensional.

Definition. The vectors $v_1, \ldots, v_n$ are linearly independent in V over $\mathbb{F}$ if whenever $\lambda_1 v_1 + \cdots \lambda_n v_n = 0$ then $\lambda_1 = \cdots = \lambda_n = 0$. Otherwise, the vectors are linearly dependent.

More generally, if $S \subset V$ then S is linearly independent precisely if every finite subset of S is linearly independent.

Remark. 0 is never contained in a linearly independent set as $1 \cdot 0 = 0$.

Remark. (i) $V = \mathbb{C}$ is a vector space over $\mathbb{R}$ and $1, i$ are linearly independent.

(ii) $V = \mathbb{C}$ is also a vector space over $\mathbb{C}$. In this case, $1, i$ are linearly dependent.

Definition. The vectors $v_1, \ldots, v_n$ form a basis of V if they both span V over $\mathbb{F}$ and are linearly independent.

Example. (i) $P_2(\mathbb{R})$ has standard basis $1, x, x^2$.

(ii) $\mathbb{F}^n$ has standard basis $e_1, \ldots, e_n$ where $e_i = (0, \ldots, 1, \ldots, 0)^T$.

(iii) $\{0\}$ has basis $\emptyset$.

Lemma 1.3. $v_1, \ldots, v_n \in V$ form a basis of V over $\mathbb{F}$ if and only if each element $v \in V$ can be written uniquely as $v = \sum_{i=1}^n \lambda_i v_i$ with $\lambda_i \in \mathbb{F}$.

Proof. Let $v \in V$. $v_1, \ldots, v_n$ span V so $v = \sum_{i=1}^n \lambda_i v_i$ for some $\lambda_i \in \mathbb{F}$. If also $v = \sum_{i=1}^n \mu_i v_i$ then $0 = \sum_{i=1}^n (\lambda_i - \mu_i) v_i$, so as $v_1, \ldots, v_n$ are linearly independent, we have $\lambda_i = \mu_i$ for $i = 1, \ldots, n$.

Conversely, since each $v \in V$ is a linear combination of $v_1, \ldots, v_n$, we have $v_1, \ldots, v_n$ span V over $\mathbb{F}$. They are linearly dependent since if $\sum_{i=1}^n \lambda_i v_i = 0 = \sum_{i=1}^n 0 v_i$ then $\lambda_i = 0$ for $i = 1, \ldots, n$ by uniqueness. □

Lemma 1.4. if $v_1, \ldots, v_n$ span V over $\mathbb{F}$, some subset of $\{v_1, \ldots, v_n\}$ is a basis of V.

Proof. If $v_1, \ldots, v_n$ are linearly independent, we are done. Otherwise, for some k, there exist $\alpha_1, \ldots, \alpha_{k-1} \in \mathbb{F}$ with $v_k = \alpha_1 v_1 + \cdots + \alpha_{k-1} v_{k-1}$. (If $\lambda_1 v_1 + \cdots + \lambda_n v_n = 0$ with not all $\lambda_i = 0$, take k maximal with $\lambda_k \neq 0$ and let $\alpha_i = -\frac{\lambda_i}{\lambda_k}$.) Then $v_1, \ldots, v_{k-1}, v_{k+1}, \ldots, v_n$ still span V since whenever $v = \sum_{i=1}^n \lambda_i v_i$ then $v = \sum_{i=1}^{k-1} (\alpha_i + \lambda_i) v_i + \sum_{i=k+1}^n \lambda_i v_i$. Continue deleting until we have a basis. □

Theorem 1.5 (Steinitz Exchange Lemma)**.** Let V be a finite dimensional vector space over $\mathbb{F}$. Let $v_1, \ldots, v_m$ be linearly independent and let $w_1, \ldots, w_n$ span V over $\mathbb{F}$. Then $m \leq n$, and reordering the w_i if necessary, the vectors $v_1, \ldots, v_m, w_{m+1}, \ldots, w_n$ span V.

Proof. Suppose we have replaced $r \geq 0$ of the w_i already, renumbering the w_i if necessary, we have $v_1, \ldots, v_r, w_{r+1}, \ldots, w_n$ span V. If $r = m$, we are done. So assume $r < m$. Then

$$v_{r+1} = \sum_{i=1}^{r} \alpha_i v_i + \sum_{i=r+1}^{n} \beta_i w_i$$

for some $\alpha_i, \beta_i \in \mathbb{F}$. Note that $\beta_i \neq 0$ for some i since $v_1, \ldots, v_r, v_{r+1}$ are linearly independent. After reordering $w_{r+1}, \ldots, w_n$ we have $\beta_{r+1} \neq 0$. Then

$$w_{r+1} = \sum_{i=1}^{r} \frac{-\alpha_i}{\beta_{r+1}} v_i + \frac{1}{\beta_{r+1}} v_{r+1} + \sum_{i=r+2} \frac{-\beta_i}{\beta_{r+1}} w_i.$$

It follows that V is spanned by $v_1, \ldots, v_r, v_{r+1}, w_{r+2}, \ldots, w_n$. After m steps, we shall have replaced m of the w_i by the v_i, retaining the spanning property. It follows that $m \leq n$. □

Theorem 1.6. If V is a finite dimensional vector space over $\mathbb{F}$, any two bases have the same size, the dimension of V over $\mathbb{F}$, $\dim_{\mathbb{F}} V$.

Proof. Let $v_1, \dots, v_m$ and $w_1, \dots, w_n$ be two bases. Then $m \leq n$ since the v_i are linearly independent and the w_i span V. Also $n \leq m$ since the w_i are linearly independent and the v_i span V. □

Example. (i) $\dim_{\mathbb{F}} \mathbb{F}^n = n$;

(ii) $\dim_{\mathbb{R}} P_2(\mathbb{R}) = 3$;

(iii) $\dim_{\mathbb{R}} \mathbb{C} = 2$;

(iv) $\dim_{\mathbb{F}} \mathbb{F} = 1$.

Lemma 1.7. If V is a finite dimensional vector space over $\mathbb{F}$, and if the vectors $v_1, \dots, v_k$ are linearly independent for some $k \geq 0$, there exists a basis $v_1, \dots, v_k, v_{k+1}, \dots, v_n$ of V.

Proof. If $v_1, \dots, v_k$ span V, we are done. Otherwise, take $v_{k+1} \in V \setminus \langle v_1, \dots, v_k \rangle$. Then $v_1, \dots, v_{k+1}$ are linearly independent. We shall obtain a basis after $\dim V - k$ steps. □

Remark. If V is finite dimensional then whenever $U \leq V$ then $\dim U \leq \dim V$ with equality if and only if $U = V$.

Lemma 1.8. Let V be a vector space over $\mathbb{F}$ with $\dim V = n$.

(i) An independent set has at most n vectors, with equality if and only if it as basis.

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

Proof. (i) This follows from Lemma 1.7 and Theorem 1.6.

(ii) This follows from Lemma 1.4 and Theorem 1.6. □

Lemma 1.9. Let $\dim_{\mathbb{F}} V = n$. The following are equivalent.

(i) $v_1, \dots, v_n$ form a basis.

(ii) $v_1, \dots, v_n$ are linearly independent.

(iii) $v_1, \dots, v_n$ span V.

Lemma 1.10. Let $S \subset V$. There is a unique smallest subspace U of V containing S, denoted by $U = \langle S \rangle$, the subspace generated by S. In fact, U consists of linear combinations of elements of S.

Proof. If we write U for the set of linear combinations of elements of S then U is a subspace of V. On the other hand, U as defined above has to be in any subspace containing S. □

Example. Let $V = \mathbb{R}^{\mathbb{R}}$ and $S = \{1, x, x^2, \dots\}$. Then $\langle S \rangle = P(\mathbb{R})$, the subspace of polynomial functions.

Remark. (i) The intersection of any collection of subspaces is a subspace.

(ii) If $U, W \leq V$ define $U + W = \{u + w : u \in U, w \in W\}$. Then $U + W \leq V$.

(iii) Note that $U \cup W$ is a subspace if and only if $U \subset W$ or $W \subset U$.

Theorem 1.11. If U, W are finite dimensional subspaces of V then $U+W$ is also finite dimensional and $\dim U + W = \dim U + \dim W - \dim U \cap W$.

Proof. Let $v_1, \ldots, v_k$ be a basis for $U \cap W$. Extend this to $v_1, \ldots, v_k, u_1, \ldots, u_l$ a basis for U and to $v_1, \ldots, v_k, w_1, \ldots, w_m$ a basis for W. We claim $v_1, \ldots, v_k, u_1, \ldots, u_l, w_1, \ldots, w_m$ is a basis for $U + W$.

- If $v \in U + W$, then $v = u + w$ for some $u \in U$, $w \in W$. Now $u = \sum \alpha_i v_i + \sum \beta_i u_i$ for some $\alpha_i, \beta_i \in \mathbb{F}$ and $w = \sum \alpha_i' v_i + \sum \gamma_i w_i$ for some $\alpha_i', \gamma_i \in \mathbb{F}$, and therefore
$$v = \sum (\alpha_i + \alpha_i') v_i + \sum \beta_i u_i + \sum \gamma_i w_i.$$
- Suppose $\sum \alpha_i v_i + \sum \beta_i u_i + \sum \gamma_i w_i = 0$. Then
$$\begin{aligned} \sum \alpha_i v_i + \sum \beta_i u_i &= -\sum \gamma_i w_i \\ &= \sum \delta_i v_i \end{aligned}$$
for some $\delta_i \in \mathbb{F}$, using that the LHS is in U and the RHS is in W, so both vectors are in $U \cap W$. Then
$$\sum (\alpha_i - \delta_i) v_i + \sum \beta_i u_i = 0,$$
and as $v_1, \ldots, v_k, u_1, \ldots, u_l$ form a basis of U, we have that all β_i are 0. But then
$$\sum \alpha_i v_i + \sum \gamma_i w_i = 0,$$
and as $v_1, \ldots, v_k, w_1, \ldots, w_m$ form a basis of W, we have that all α_i, γ_i are 0. □

Definition. Let V be a vector space over $\mathbb{F}$ and suppose $U, W \leq V$. Then
$$V = U \oplus W$$
if every element v of V can be written uniquely as $v = u + w$ with $u \in U$ and $w \in W$. If so, we say W is the complement (or complementary subspace) of U in V.

Lemma 1.12. Suppose $U, W \leq V$. Then $V = U \oplus W$ if and only if $U + W = V$ and $U \cap W = \{0\}$.

Lemma 1.13. If V is a finite dimensional vector space over $\mathbb{F}$ and $U \leq V$, then U has a complement in V. (Note this is not at all unique unless $U = \{0\}$ or $U = V$.)

Proof. Take $v_1, \ldots, v_k$ a basis for U and extend to a basis $u_1, \ldots, u_k, w_{k+1}, \ldots, w_n$ for V. Then $W = \langle w_{k+1}, \ldots, w_n \rangle$ is a complement of U in V. □

Lemma 1.14. Suppose $V_1, \ldots, V_k \leq V$ and let $\sum V_i = \{\sum v_i : v_i \in V_i\} \leq V$. The sum is direct, written as $\bigoplus V_i$, if each $v \in V$ is uniquely expressible as $\sum v_i$ with $v_i \in V_i$.

Lemma 1.15. Let $V_1, \ldots, V_k \leq V$. The following are equivalent.

(i) $\sum V_i$ is direct.

(ii) If B_i is a basis for V_i then $B = \bigcup_{i=1}^{k} B_i$ is a basis for $\sum V_i$.

(iii) For each i, $V_i \cap \sum_{j \neq i} V_j = \{0\}$.

Note that in (iii), if $k > 2$ it is not enough to assume $V_i \cap V_j = \{0\}$ for all $i \neq j$.

Proof. We show (i) $\implies$ (ii). Let B_i be a basis for V_i, let $B = \bigcup_{i=1}^{k} B_i$. If $v \in \sum V_i$ then $v = \sum_{i=1}^{k} v_i$, and v_i can be written as a linear combination of vectors in B_i, so substitute for each v_i and obtain v as a linear combination of B. If we have a linear combination of vectors in B equal to 0, collect together terms in each V_i, let v_i denote the part in V_i. Then $v_1 + \cdots + v_k = 0$. By uniqueness of expression for 0, we have $v_i = 0$. Now B_i is linear independent, so all coefficients are 0. $\square$

Chapter 2

Linear maps, matrices

2.1 Basic definitions and properties

Definition. Let V, W be vector spaces over $\mathbb{F}$. The map $\alpha : V \to W$ is linear if

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

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

Example. (i) The map $D : D(\mathbb{R}) \to F(\mathbb{R}) = \mathbb{R}^{\mathbb{R}}$, $f \mapsto \frac{\mathrm{d}f}{\mathrm{d}t}$ is linear.

(ii) The map $\int_0^x : C[0,1] \to F[0,1]$, $f \mapsto \int_0^x f(t)\,\mathrm{d}t$ is linear.

(iii) If A is an $m \times n$ matrix, the map $\alpha : \mathbb{F}^n \to \mathbb{F}^m$, $x \mapsto Ax$ is linear.

Lemma 2.1. Let U, V, W be vector spaces.

(i) $\iota_v : V \to V, v \mapsto v$ is linear.

(ii) If $U \xrightarrow{\beta} V \xrightarrow{\alpha} W$ with α, β linear, then so is $\alpha \circ \beta : U \to W$.

Lemma 2.2. Let V, W be vector spaces over $\mathbb{F}$, let B be a basis for V. If $\alpha_0 : B \to W$ is any map, there is a unique linear map $\alpha : V \to W$ which extends α_0, so $\alpha(v) = \alpha_0(v)$ for all $v \in B$.

Proof. For $v \in V$ with $v = \lambda_1 v_1 + \cdots + \lambda_n v_n$ where $v_i \in B$, $\lambda_i \in \mathbb{F}$, we must have $\alpha(v) = \sum \lambda_i \alpha_0(v_i)$. Then α is linear. □

Definition. If V, W are vector spaces over $\mathbb{F}$, the map $\alpha : V \to W$ is an isomorphism if it is linear and bijective. Write $V \simeq W$ is an isomorphism $V \to W$ exists.

Lemma 2.3. $\simeq$ is an equivalence relation on the set of vector spaces over $\mathbb{F}$.

(i) $\iota_V : V \to V$ is an isomorphism.

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

(iii) If $U \xrightarrow{\alpha} V \xrightarrow{\beta} W$ are isomorphisms then so is $\beta \circ \alpha : U \to W$.

Proof. (i) and (iii) are clear. To prove (ii), suppose α is an isomorphism so $\alpha^{-1} : W \to V$ exists and is a bijection.

$$\begin{aligned}
\alpha^{-1}(w_1 + w_2) &= \alpha^{-1}(\alpha(v_1) + \alpha(v_2)) \\
&= \alpha^{-1}(\alpha(v_1 + v_2)) \\
&= v_1 + v_2 \\
&= \alpha^{-1}(w_1) + \alpha^{-1}(w_2) \\
\alpha^{-1}(\lambda w) &= \alpha^{-1}(\lambda\alpha(v)) \\
&= \alpha^{-1}(\alpha(\lambda v)) \\
&= \lambda v \\
&= \lambda\alpha^{-1}(w)
\end{aligned}$$

So α^{-1} is linear. □

Theorem 2.4. If V is a vector space over $\mathbb{F}$ of finite dimension n, then $V \simeq \mathbb{F}^n$.

Proof. Choose a basis $v_1, \ldots, v_n$. The map $\alpha : V \to \mathbb{F}^n, \sum_{i=1}^n \lambda_i v_i \mapsto (\lambda_1, \ldots, \lambda_n)^T$ is an isomorphism. □

Theorem 2.5. The vector spaces V, W over $\mathbb{F}$ are isomorphic if and only if they have the same dimension.

Proof. If $v_1, \ldots, v_n$ and $w_1, \ldots, w_n$ are bases for V and W, respectively, then the map $\alpha : V \to W, \sum \lambda_i v_i \mapsto \sum \lambda_i w_i$ is an isomorphism.

The image of a basis of V under an isomorphism is a basis of W. Let B be a basis of V, $\alpha : V \to W$ an isomorphism. Then $\alpha(B)$ is a basis of W.

- If $w \in W$, then $w = \alpha(v)$ for some $v \in V$; write $v = \sum_{i=1}^n \lambda_i v_i$ with $v_1, \ldots, v_n \in B$, $\lambda_i \in \mathbb{F}$. Then $w = \alpha(\sum \lambda_i v_i) = \sum \lambda_i \alpha(v_i)$.
- If $\lambda_1\alpha(v_1) + \cdots + \lambda_n\alpha(v_n) = 0$ then $\alpha(\lambda_1 v_1 + \cdots + \lambda_n v_n) = 0$ so $\lambda_1 v_1 + \cdots + \lambda_n v_n = 0$ as α is injective. Hence $\lambda_i = 0$ for all i as B is independent. □

Remark. These isomorphisms are not *natural*, they depend on the choice of bases.

Definition. Let $\alpha : V \to W$ be a linear map. Write $\ker(\alpha) = \{v \in V : \alpha(v) = 0\} = N(\alpha)$, the nullspace of α, and $\mathrm{Im}(\alpha) = \{w \in W : w = \alpha(v) \text{ for some } v \in V\}$. Then $N(\alpha) \leq V$, $\mathrm{Im}(\alpha) \leq W$. Note that α is injective if and only if $N(\alpha) = \{0\}$ and surjective if and only if $\mathrm{Im}(\alpha) = W$. Define $n(\alpha) = \dim N(\alpha)$, the nullity of α, and $\mathrm{rank}(\alpha) = \dim \mathrm{Im}(\alpha)$, the rank of α.

Theorem 2.6 (Rank-nullity theorem)**.** Let V, W be vector spaces over $\mathbb{F}$ with $\dim_{\mathbb{F}} V$ finite. Let $\alpha : V \to W$ be a linear map. Then $\dim V = \mathrm{rank}(\alpha) + n(\alpha)$.

Proof. Let $v_1, \ldots, v_k$ be a basis for $N(\alpha)$, extend to $v_1, \ldots, v_k, v_{k+1}, \ldots, v_n$ a basis for V. We claim $\alpha(v_{k+1}), \ldots, \alpha(v_n)$ is a basis for $\mathrm{Im}(\alpha)$.

- If $w \in \mathrm{Im}(\alpha)$, say $w = \alpha(v)$, $v \in V$, then $v = \sum_{i=1}^n \lambda_i v_i$ for some $\lambda_i \in \mathbb{F}$. So $w = \alpha(v) = \sum_{i=1}^n \lambda_i \alpha(v_i) = \sum_{k+1}^n \lambda_i \alpha(v_i)$, since $\alpha(v_i) = 0$ for $i = 1, \ldots, k$.

- If $\sum_{i=k+1}^{n} \lambda_i \alpha(v_i) = 0$ then $\alpha(\sum_{i=k+1}^{n} \lambda_i v_i) = 0$, so $\sum_{i=k+1}^{n} \lambda_i v_i \in N(\alpha)$, so it can be written as $\sum_{i=1}^{k} \lambda_i v_i$. But $v_1, \ldots, v_n$ are linearly independent, so $\lambda_i = 0$ for $i = 1, \ldots, n$. □

Remark. The rank-nullity theorem is a linear version of the isomorphism theorem. Let V be a vector space over $\mathbb{F}$ and $N \leq V$. Then $V/N = \{v + N : v \in V\}$ is a vector space over $\mathbb{F}$ with the operations defined as follows.

$$(v_1 + N) + (v_2 + N) = (v_1 + v_2) + N$$
$$\lambda(v + N) = (\lambda v) + N$$

Write $\bar{V} = V/N$, $\bar{v} = v + N$. Choose a basis $v_1, \ldots, v_k, v_{k+1}, \ldots, v_n$ for V containing the basis $v_1, \ldots, v_k$ for $N(\alpha)$. Then $\bar{v_{k+1}}, \ldots, \bar{v_n}$ is a basis for $\bar{V}$. Hence $\dim V/N = \dim V - \dim N$. Now let $\alpha : V \to W$ be linear for some vector space W over $\mathbb{F}$. Then $V/N(\alpha) \approx \mathrm{Im}(\alpha)$. The map $v + N(\alpha) \mapsto \alpha(v)$ is linear. Hence $\dim \mathrm{Im}(\alpha) = \dim V/N(\alpha) = \dim V - \dim N(\alpha)$.

Lemma 2.7. Let V be a vector space of finite dimension over $\mathbb{F}$, let $\alpha : V \to V$ be linear. (More generally, consider $\alpha : V \to W$ with $\dim V = \dim W$.) The following are equivalent.

(i) α is an isomorphism.

(ii) α is surjective.

(iii) α is injective.

Proof. Apply the rank-nullity theorem. □

2.2 The space of linear maps

Definition. Let U, V be vector spaces over $\mathbb{F}$. $\mathcal{L}(U, V) = \{a : U \to V \mid \alpha \text{ linear}\}$ is a vector space with

$$(\alpha_1 + \alpha_1)(u) = \alpha_1(u) + \alpha_2(u)$$
$$(\lambda\alpha)(u) = \lambda\alpha(u)$$

for $u \in U$, for $\alpha, \alpha_1, \alpha_2 \in \mathcal{L}(U, V)$, $\lambda \in \mathbb{F}$. (To show this is a vector space over $\mathbb{F}$, consider the vector space of all functions $U \to V$ and show $\mathcal{L}(U, V)$ is a subspace.)

Proposition 2.8. If U, V are vector spaces over $\mathbb{F}$, then $\mathcal{L}(U, V)$ is a vector space over $\mathbb{F}$. If both U, V are finite dimensional, then so is $\mathcal{L}(U, V)$ and $\dim \mathcal{L}(U, V) = \dim U \dim V$.

Proof. It remains to check the dimension claim. Let $u_1, \ldots, u_n$ be a basis for U, let $v_1, \ldots, v_m$ be a basis for V. For $1 \leq i \leq m$, $1 \leq j \leq n$, define $\varepsilon_{ij} : u_k \mapsto \delta_{jk} v_i$ for $1 \leq k \leq n$, and extend linearly. Then $\varepsilon_{ij} \in \mathcal{L}(U, V)$, and they form a basis.

If $\sum_{i,j} \lambda_{ij} \varepsilon_{ij} = 0$, then in particular for all $1 \leq k \leq n$,

$$0 = \sum_{i,j} \lambda_{ij} \varepsilon_{ij}(u_k) = \sum_{i} \lambda_{ik} v_i.$$

Now the v_i are linearly independent, so $\lambda_{ik} = 0$ for all $1 \leq i \leq m$. This is true for all $1 \leq k \leq n$, so the ε_{ij} are linearly independent.

Let $\alpha \in \mathcal{L}(U, V)$, say $\alpha(u_k) = \sum_i a_{ik} v_i$. Then

$$\sum_{i,j} (a_{ij} \varepsilon_{ij}(u_k)) = \sum_i a_{ik} v_i = \alpha(u_k)$$

for any $1 \leq k \leq n$, so the maps are equal on a basis of U, hence they are equal by Lemma 2.2. □

2.3 Matrices

Definition. An $m \times n$ matrix over $\mathbb{F}$ is an array $A = (a_{ij})$ with m rows and n columns and $a_{ij} \in \mathbb{F}$ for $1 \leq i \leq m$, $1 \leq n \leq n$. Write $M_{m,n}(\mathbb{F})$ for the set of all $m \times n$ matrices over $\mathbb{F}$.

Proposition 2.9. $M_{m,n}(\mathbb{F})$ is a vector space under operations

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

with $\dim_{\mathbb{F}} M_{m,n}(\mathbb{F}) = mn$.

Proof. We prove the dimension claim. For $1 \leq i \leq m$, $1 \leq j \leq n$ define

$$E_{ij} = \begin{cases} e_{ij} = 1 \\ e_{i'j'} = 0 \quad \text{if } (i', j') \neq (i, j). \end{cases}$$

This is a natural basis. □

Let U, V be finite dimensional vector spaces over $\mathbb{F}$, let $\alpha : U \to V$ be linear. Fix bases $B = \{u_1, \ldots, u_n\}$ and $C = \{v_1, \ldots, v_m\}$ for U and V, respectively. Define $A = (a_{ij})$ by $\alpha(u_j) = \sum_i a_{ij} v_i$. For $u \in U$ with $u = \sum_i \lambda_i u_i$ write $[u]_B = (\lambda_1, \ldots, \lambda_n)^T$. Then

$$A = \begin{pmatrix} [\alpha(u_1)]_C & \cdots & [\alpha(u_n)]_C \end{pmatrix},$$

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

Lemma 2.10. For all $u \in U$, $[\alpha(u)]_C = [\alpha]_{B,C} [u]_B$

Proof. If $u \in U$, $u = \sum_j \lambda_j u_j$ so $[u]_B = (\lambda_1, \ldots, \lambda_n)^T$, then

$$\begin{aligned} \alpha(u) &= \sum_j \lambda_j \alpha(u_j) \\ &= \sum_j \lambda_j \sum_i a_{ij} v_i \\ &= \sum_i \sum_j a_{ij} \lambda_j v_i \\ &= \sum_i (A \cdot [u]_B)_i v_i \end{aligned}$$

so $[\alpha(u)]_C = A \cdot [u]_B$. □

Remark. Let $B = \{u_1, \ldots, u_n\}$, $C = \{v_1, \ldots, v_m\}$ be bases for U, V, respectively. Set $\varepsilon = \varepsilon_B : U \to \mathbb{F}^n, u \mapsto [u]_B$, $\phi = \phi_C : V \to \mathbb{F}^m, v \mapsto [v]_C$. We have the following commuting diagram.

$$\begin{array}{ccc} U & \xrightarrow{\alpha} & V \\ {\scriptstyle \varepsilon_B}\downarrow & & \downarrow{\scriptstyle \phi_C} \\ \mathbb{F}^n & \xrightarrow{A\cdot} & \mathbb{F}^m \end{array}$$

Remark. In fact, A as defined above is the only matrix for which $[\alpha(u)]_C = A \cdot [u]_B$ for all $u \in U$. If also $A' \cdot [u]_B = [\alpha(u)]_C$ for all $u \in U$, then this is in particular true for $u_1, \ldots, u_n$. But $[u_k]_B = e_k$ and $A' \cdot e_k$ is the kth column of A'. This is true for each $1 \leq k \leq n$, so it determines the matrix.

Proposition 2.11. If $\alpha : U \to V$ is linear and $\dim U = n$, $\dim V = m$, then $\mathcal{L}(U, V) \simeq M_{m,n}(\mathbb{F})$.

Proof. Let $B = \{u_1, \ldots, u_n\}$ and $C = \{v_1, \ldots, v_m\}$ be bases for U and V, respectively. Then $\theta : \mathcal{L}(U, V) \to M_{m,n}(\mathbb{F}), \alpha \mapsto [\alpha]_{B,C}$ is an isomorphism. □

Lemma 2.12. Let $U \xrightarrow{\alpha} V \xrightarrow{\beta} W$ be linear and choose bases B, C, D for U, V, W, respectively. Then $[\beta \circ \alpha]_{B,D} = [\beta]_{C,D} \cdot [\alpha]_{B,C}$.

Proof. Write $A = [\alpha]_{B,C}$, $B = [\beta]_{C,D}$. Then

$$\begin{aligned} \beta \circ \alpha(u_k) &= \beta \sum_j a_{jk} v_j \\ &= \sum_j a_{jk} \sum_i b_{ij} w_i \\ &= \sum_i \sum_j b_{ij} a_{jk} w_i \\ &= \sum_i (BA)_{ik} w_i \end{aligned}$$

so $[\beta \circ \alpha]_{B,D} = BA$. □

2.4 Change of bases

Let U and V be vector spaces over $\mathbb{F}$. Suppose U has bases $B = \{u_1, \ldots, u_n\}$ and $B' = \{u'_1, \ldots, u'_2\}$, V has bases $C = \{v_1, \ldots, v_m\}$ and $C' = \{v'_1, \ldots, v'_m\}$. The matrix $P = (p_{ij})$ is the change of bases matrix from B to B' if $u'_j = \sum_i p_{ij} u_i$, so

$$P = \begin{pmatrix} [u'_1]_B & \ldots & [u'_n]_B \end{pmatrix} = [\iota_U]_{B'B}.$$

Then $[u]_B = P[u]_{B'}$ for all $u \in U$. (This is clear for each u'_j since $[u'_j]_{B'} = e_j$.) Note that P is invertible, in fact, P^{-1} is the change of basis matrix from B' to B. Similarly, we obtain the change of basis matrix Q from C to C'.

Lemma 2.13. Let $\alpha : U \to V$ be a linear map, let B, B', C, C' be as above and let $A = [\alpha]_{B,C}$, $A' = [\alpha]_{B',C'}$. Then $A' = Q^{-1}AP$.

Proof. For all $u \in U$ we have

$$\begin{aligned}[\alpha(u)]_C &= A[u]_B = AP[u]_{B'} \\ &= Q[\alpha(u)]_{C'}\end{aligned}$$

so $A' = Q^{-1}AP$. □

Definition. The $m \times n$ matrices $A, A' \in M_{m,n}(\mathbb{F})$ are equivalent if there exist invertible matrices $Q \in M_{m,m}(\mathbb{F})$, $P \in M_{n,n}(\mathbb{F})$ such that $A' = QAP$. This defines an equivalence relation on $M_{m,n}(\mathbb{F})$.

Equivalent matrices arise as representing the same linear map from a space U to a space V of dimension m and n with respect to different bases.

Lemma 2.14. (i) Let U, V be vector spaces over $\mathbb{F}$ with $\dim U = n$, $\dim V = m$. Let $\alpha : U \to V$ be linear. There exist bases B of U and C of V such that $[\alpha]_{B,C} = \left(\begin{smallmatrix} I_r & 0 \\ 0 & 0 \end{smallmatrix}\right)$ for some r.

(ii) Any $m \times n$ matrix is equivalent to $\left(\begin{smallmatrix} I_r & 0 \\ 0 & 0 \end{smallmatrix}\right)$ for some r.

Proof. (i) Let $u_{r+1}, \ldots, u_n$ be a basis for $N(\alpha)$, extend this to a basis $B = \{u_1, \ldots, u_r, u_{r+1}, \ldots, u_n\}$ of U. Then $\alpha(u_1), \ldots, \alpha(r)$ is a basis of $\mathrm{Im}(\alpha)$. We can extend this to a basis $C = \{\alpha(u_1), \ldots, \alpha(u_r), v_{r+1}, \ldots, v_m\}$ of V. Then $[\alpha]_{B,C} = \left(\begin{smallmatrix} I_r & 0 \\ 0 & 0 \end{smallmatrix}\right)$.

(ii) Let $A \in M_{m,n}(\mathbb{F})$. Define $\alpha : \mathbb{F}^n \to \mathbb{F}^m, x \mapsto Ax$. With respect to the standard bases of $\mathbb{F}^n$, $\mathbb{F}^m$, α has matrix A. By part (i) and Lemma 2.13, A is equivalent to $\left(\begin{smallmatrix} I_r & 0 \\ 0 & 0 \end{smallmatrix}\right)$ for some r. □

2.5 Rank

Definition. Let $A \in M_{m,n}(\mathbb{F})$. The (column) rank of A, denoted $\mathrm{rank}(A)$ is the dimension of the column space of A. The column space is the subspace of $\mathbb{F}^n$ generated by the column vectors of A.

Lemma 2.15. Let $\alpha : U \to V$ be linear, let B be a basis for U, and C be a basis for V. Let $A = [\alpha]_{B,C}$. Then $\mathrm{rank}(\alpha) = \mathrm{rank}(A)$.

Proof. The map $\theta : \mathrm{Im}(\alpha) \to \mathrm{colsp}(A), \alpha(u) \mapsto [\alpha(u)]_C$ is an isomorphism. □

Proposition 2.16. The matrices $A, A' \in M_{m,n}(\mathbb{F})$ are equivalent if and only if $\mathrm{rank}(A) = \mathrm{rank}(A')$.

Proof. Assume $A' = Q^{-1}AP$ with Q, P invertible. Let $\alpha : \mathbb{F}^n \to \mathbb{F}^m, x \mapsto Ax$. Then α with respect to standard bases B, C has matrix A. Let B' be the set of columns of P, let C' be the set of columns of Q'. Then $[\alpha]_{B',C'} = Q^{-1}AP$ by Lemma 2.13 since P and C are the change of basis matrices from B to B' and C to C', respectively. So $\mathrm{rank}(A) = \mathrm{rank}(\alpha) = \mathrm{rank}(A')$ by Lemma 2.15.

We have that A and A' are equivalent to $\left(\begin{smallmatrix} I_r & 0 \\ 0 & 0 \end{smallmatrix}\right)$ and $\left(\begin{smallmatrix} I_{r'} & 0 \\ 0 & 0 \end{smallmatrix}\right)$ for some r and r', respectively. By the first part, $\mathrm{rank}(A) = r$, $\mathrm{rank}(A') = r'$. Since $\mathrm{rank}(A) = \mathrm{rank}(A')$ we have $r = r'$ so A and A' are equivalent by transitivity of the equivalence relation. □

Definition. For $A \in M_{m,n}(\mathbb{F})$ let $\operatorname{rowrk}(A) = \dim \operatorname{rowsp}(A) = \operatorname{rank}(A^T)$.

Theorem 2.17. For $A \in M_{m,n}(\mathbb{F})$, $\operatorname{rowrk}(A) = \operatorname{rank}(A)$.

Proof. Let $A \in M_{m,n}(\mathbb{F})$, let $r = \operatorname{rank}(A)$. Then A is equivalent to $\left(\begin{smallmatrix} I_r & 0 \\ 0 & 0 \end{smallmatrix}\right)_{m\times n}$, so $\left(\begin{smallmatrix} I_r & 0 \\ 0 & 0 \end{smallmatrix}\right)_{m\times n} = QAP$ for some invertible matrices Q, P. Consider the transpose,

$$P^T A^T Q^T = \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}_{n\times m}$$

sp A^T is equivalent to $\left(\begin{smallmatrix} I_r & 0 \\ 0 & 0 \end{smallmatrix}\right)_{n\times m}$. Visibly, $\operatorname{rank}\left(\begin{smallmatrix} I_r & 0 \\ 0 & 0 \end{smallmatrix}\right)_{n\times m} = r$, so $\operatorname{rowrk}(A) = \operatorname{rank}(A^T) = \operatorname{rank}\left(\begin{smallmatrix} I_r & 0 \\ 0 & 0 \end{smallmatrix}\right)_{n\times m} = r = \operatorname{rank}(A)$. □

2.6 Calculations

Definition. The following are the elementary column operations on an $m \times n$ matrix over $\mathbb{F}$.

(i) Swap columns i and j.

(ii) Replace column i by λ column i, where $\lambda \in \mathbb{F} \setminus \{0\}$.

(iii) Add λ column i to column j, where $i \neq j$ and $\lambda \in \mathbb{F}$.

The corresponding elementary matrices are obtained by applying these operations to I_n. Call these T_{ij}, $M_{i,\lambda}$ and $C_{i,j,\lambda}$. An elementary column operation on A can be performed by postmultiplying A with the corresponding elementary matrix. All these operations are reversible.

Definition. A an $m \times n$ matrix with the following properties is said to be in column echelon form.

(i) The highest placed non-zero entry in column j is 1 in row i_j, with $i_1 \leq i_2 \leq \cdots$.

(ii) The entry in row i_j and column k with $k < j$ is 0.

Lemma 2.18. Any matrix A can be reduced to a matrix in column echelon form by a sequence of elementary column matrices.

Remark. If A is a square $n \times n$ matrix and is invertible, the equivalent column echelon form is I_n. This can be used to find A^{-1}.

$$A \mapsto AE_1E_2\cdots E_k = I$$
$$I_n \mapsto I_nE_1E_2\cdots E_k = A^{-1}.$$

Lemma 2.19. If A is an invertible $n \times n$ matrix then A is a product of elementary matrices.

Proof. By the above remark, $A^{-1} = E_1 \cdots E_k$ is a product of elementary matrices, hence so is $A = E_k^{-1} \cdots E_1^{-1}$. □

Definition. Two $n \times n$ matrices A, A' are similar if $A' = P^{-1}AP$ for some invertible matrix P.

Chapter 3

Determinant and trace

Definition. For $A \in M_n(\mathbb{F})$ define $\operatorname{tr} A = \sum_{i=1}^n a_{ii}$. Note $\operatorname{tr} : M_n(\mathbb{F}) \to \mathbb{F}$ is linear.

Lemma 3.1. $\operatorname{tr}(AB) = \operatorname{tr}(BA)$.

Proof. $\operatorname{tr}(AB) = \sum_i \sum_j a_{ij} b_{ji} = \sum_j \sum_i b_{ji} a_{ij} = \operatorname{tr}(BA)$. □

Lemma 3.2. Similar matrices have the same trace.

Proof. $\operatorname{tr}(P^{-1}AP) = \operatorname{tr}(APP^{-1}) = \operatorname{tr}(A)$. □

Remark. For $\alpha \in \operatorname{End}(V)$ we can define $\operatorname{tr}(\alpha) = \operatorname{tr}[\alpha]_B$, where B is any basis for V.

Recall that S_n is the group of all permutations of $\{1, \dots, n\}$. The elements are permutations, the group operation is composition of permutations $(\sigma \circ \tau)(j) = \sigma(\tau(j))$. Any σ can be written as a product of transpositions.

$$\varepsilon(\sigma) = \begin{cases} +1 & \text{if \# permutations is even,} \\ -1 & \text{if \# permutations is odd.} \end{cases}$$

$\varepsilon : S_n \to \{+1, -1\}$ is a homomorphism.

Definition. For $A \in M_n(\mathbb{F})$ define

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

Writing $A^{(i)}$ for the ith column of a matrix A we have $A = (A^{(1)}, \dots, A^{(n)})$, so we can think of A is an n-tuple of columns in $\mathbb{F}^n$. Write $\{e_1, \dots, e_n\}$ for the standard basis of $\mathbb{F}^n$.

Definition. The function $d : \mathbb{F}^n \times \cdots \times \mathbb{F}^n \to \mathbb{F}$ is a volumne form on $\mathbb{F}^n$ if

(i) it is multilinear, i.e.

$$\begin{aligned} d(v_1, \dots, \lambda_i v_i, \dots, v_n) &= \lambda d(v_1, \dots, v_i, \dots, v_n) \\ d(v_1, \dots, v_i + v_i', \dots, v_n) &= d(v_1, \dots, v_i, \dots, v_n) + d(v_1, \dots, v_i', \dots, v_n) \end{aligned}$$

(ii) it is alternating, i.e. whenever $i \neq j$ and $v_i = v_j$ then $d(v_1, \dots, v_n) = 0$.

d is a determinant function if it is also normalised, i.e. $d(e_1, \dots, e_n) = 1$.

Lemma 3.3. Swapping columns in a volume form changes the sign. For $i \neq j$,

$$d(v_1, \dots, v_j, \dots, v_i, \dots, v_n) = -d(v_1, \dots, v_i, \dots, v_j, \dots, v_n).$$

Proof.

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

Corollary 3.4. If $\sigma \in S_n$ and d is a volume form on $\mathbb{F}^n$ then

$$d(v_{\sigma(1)}, \dots, v_{\sigma(n)} = \varepsilon(\sigma) d(v_1, \dots, v_n)$$

for $v_1, \dots, v_n \in \mathbb{F}^n$. In particular,

$$\begin{aligned} d(e_{\sigma(1)}, \dots, e_{\sigma(n)} &= \varepsilon(\sigma) d(e_1, \dots, e_n) \\ &= \varepsilon(\sigma) \end{aligned}$$

if d is a determinant function.

Theorem 3.5. If d is a volume form on $\mathbb{F}^n$ and $A = (a_{ij}) = (A^{(1)}, \dots, A^{(n)}) \in M_n(\mathbb{F})$ then $d(A^{(1)}, \dots, A^{(n)}) = \det A d(e_1, \dots, e_n)$.

Proof.

$$\begin{aligned} d(A^{(1)}, \dots, A^{(n)}) &= d(\sum_{j_1} a_{j_1 1} e_{j_1}, A^{(2)}, \dots, A^{(n)}) \\ &= \sum_{j_1} a_{j_1 1} d(e_{j_1}, A^{(2)}, \dots, A^{(n)}) \\ &= \sum_{j_1, j_2} a_{j_1 1} a_{j_2 2} d(e_{j_1}, e_{j_2}, A^{(3)}, \dots, A^{(n)}) \\ &= \cdots \\ &= \sum_{j_1, \dots, j_n} a_{j_1 1} \cdots a_{j_n n} d(e_{j_1}, \dots, e_{j_n}) \\ &= \sum_{\sigma \in S_n} a_{\sigma(1) 1} \cdots a_{\sigma(n) n} \varepsilon(\sigma) d(e_1, \dots, e_n) \\ &= (\det A) d(e_1, \dots, e_n). \end{aligned}$$ □

Theorem 3.6. If we define $d : \mathbb{F}^n \times \cdots \times \mathbb{F}^n \to \mathbb{F}$ by $d(A^{(1)}, \dots, A^{(n)}) = \det A$ for $A = (A^{(1)}, \dots, A^{(n)})$ then d is a determinant function.

Proof. (i) $\prod_{j=1}^n a_{\sigma(j) j}$ is multilinear, hence so is the linear combination $\det A$.

(ii) If $A^{(k)} = A^{(l)}$ with $k \neq l$ then $\det A = 0$. Write $\tau = (kl)$, a transposition in S_n.

$$\det A = \sum_{\sigma \in S_n} \varepsilon(\sigma) \prod_j a_{\sigma(j) j}$$

Reorder the sum, take each even permutation σ followed by the odd permutation $\sigma\tau$ and note $\varepsilon(\sigma) = 1$, $\varepsilon(\sigma\tau) = -1$. Thus

$$\det A = \sum_{\sigma \text{ even}} \left(\prod_j a_{\sigma(j)j} - \prod_j a_{\sigma\tau(j)j} \right) = 0$$

as each of the summands is zero since

$$\begin{aligned} &a_{\sigma(1)1} \cdots a_{\sigma(k)k} \cdots a_{\sigma(l)l} \cdots a_{\sigma(n)n} \\ -\, &a_{\sigma(1)1} \cdots a_{\sigma(l)k} \cdots a_{\sigma(k)l} \cdots a_{\sigma(n)n} = 0. \end{aligned}$$

(iii) $\det A = \sum_{\sigma \in S_n} \varepsilon(\sigma) \prod_{j=1}^n \delta_{\sigma(j)j} = \varepsilon(\iota) \cdot 1 = 1$. □

Lemma 3.7. $\det A^T = \det A$.

Proof. If $\sigma \in S_n$ then $\prod_{j=1}^n a_{\sigma(j)j} = \prod_{j=1}^n a_{j\sigma(j)}$ as they contain the same factors but in a different order. Also, as σ runs through S_n, so does σ^{-1} and $\varepsilon(\sigma^{-1}) = \varepsilon(\sigma)$. Hence

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

□

Lemma 3.8. det is the unique multilinear alternating form in rows normalised at I.

Lemma 3.9. If A is an upper triangular matrix, i.e. $a_{ij} = 0$ for all $i > j$, then $\det A = a_{11} \cdots a_{nn}$.

Proof. From the definition the determinant,

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

For a product to contribute, we must have $\sigma(i) \leq i$ for all $i = 1, \ldots, n$. Hence $\sigma(1) = 1$, $\sigma(2) = 2$, …, $\sigma(n) = n$, so $\sigma = \iota$ and hence $\det A = a_{11} \cdots a_{nn}$. □

Lemma 3.10. If E is an elementary matrix, then for any $n \times n$ matrix A,

$$\det(AE) = \det A \det E = \det(EA).$$

Performing an elementary column or row operation on A multiplies $\det A$ by the determinant of the corresponding elementary matrix.

Proof. Note that for the determinants of the elementary matrices we have $\det T_{ij} = -1$, $\det M_{i,\lambda} = \lambda$, $\det C_{i,j,\lambda} = 1$. Performing the corresponding elementary column or row operation multiplies $\det A$ by -1, λ, 1, respectively. □

Theorem 3.11. Let A be a square matrix. Then A is non-singular if and only if $\det A \neq 0$.

Proof. If A is non-singular then A can be written as a product of elementary matrices by Lemma 2.19, so $\det A$ is the product of the corresponding determinants, so $\det A \neq 0$.

If A is singular we can obtain a 0 column as a non-trivial combination of columns of A, so using elementary column operations on A we can obtain a matrix with a 0 column. Hence $\det A = 0$ by Lemma 3.10. □

Theorem 3.12. If $A, B \in M_n(\mathbb{F})$, then $\det(AB) = \det(A)\det(B)$.

Proof. Fix A; then $d_A : (B^{(1)}, \dots, B^{(n)}) \mapsto \det(AB)$ for $B = (B^{(1)}, \dots, B^{(n)})$ is a volume form on $\mathbb{F}^n$. Note that $\det(AB) = d_A(AB^{(1)}, \dots, AB^{(n)})$ an so d_A is multilinear and alternating. Hence

$$\begin{aligned}\det(AB) &= d_A(B^{(1)}, \dots, B^{(n)}) \\ &= \det B d_A(e_1, \dots, e_n) \\ &= \det(B)\det(A),\end{aligned}$$

using Theorem 3.5. □

Proof. Expand as before.

$$\begin{aligned}\det(AB) &= \det\left(\sum_{j_1} b_{j_1 1}A^{(j_1)}, \dots, \sum_{j_n} b_{j_n n}A^{(j_n)}\right) \\ &= \sum_{\sigma\in S_n}\left(\prod_{j=1}^{n} b_{\sigma(j)j}\right)\det(A^{\sigma(1)}, \dots, A^{\sigma(n)}) \\ &= \sum_{\sigma\in S_n}\left(\prod_{j=1}^{n} b_{\sigma(j)j}\right)\varepsilon(\sigma)\det A \\ &= \det(A)\det(B).\end{aligned}$$

□

Proof. If B is singular, so is AB and hence $\det B = 0 = \det(AB)$. So assume B is non-singular and write it as a product of elementary matrices, $B = E_1 \cdots E_k$ by Lemma 2.19. Using Lemma 3.10,

$$\begin{aligned}\det(AB) &= \det(AE_1 \cdots E_k) \\ &= \det A \det E_1 \cdots \det E_k \\ &= \det A \det B.\end{aligned}$$

□

Corollary 3.13. If A is invertible then $\det A^{-1} = (\det A)^{-1}$.

Proof. As A is invertible, $AA^{-1} = I$ so $(\det A)(\det A^{-1}) = \det I = 1$ and hence $\det A^{-1} = (\det A)^{-1}$. □

Corollary 3.14. Conjugate $n \times n$ matrices have the same determinant.

Proof.

$$\begin{aligned}\det(P^{-1}AP) &= \det P^{-1} \det A \det P \\ &= (\det A)(\det P)(\det P)^{-1} \\ &= \det A\end{aligned}$$

□

Definition. If $\alpha : V \to V$ is a linear endomorphism, define $\det \alpha = \det[\alpha]_B$ for any basis B of V.

Theorem 3.15. $\det : \mathrm{End}(V) \to \mathbb{F}$ has the following properties.

(i) $\det \iota = 1$.

(ii) $\det \alpha \circ \beta = \det \alpha \det \beta$.

(iii) $\det \alpha \neq 0$ if and only if α is invertible, and if α is invertible then $\det \alpha^{-1} = (\det \alpha)^{-1}$.

Remark. Consider the group $GL(V)$ of automorphisms of V and the group $GL_n(\mathbb{F})$ of invertible $n \times n$ matrices over $\mathbb{F}$. Then $GL(V) \simeq GL_n(\mathbb{F})$ and $\det : GL_n(\mathbb{F}) \to \mathbb{F}$ is a homomorphism.

Lemma 3.16. If $A \in M_m(\mathbb{F})$, $B \in M_k(\mathbb{F})$ and $C \in M_{m,k}(\mathbb{F})$. Then $\det\left(\begin{smallmatrix} A & C \\ 0 & B \end{smallmatrix}\right) = \det A \det B$.

Proof. Fix B, C. Then $d_{B,C} : A \mapsto \det\left(\begin{smallmatrix} A & C \\ 0 & B \end{smallmatrix}\right)$ is a volume form on the column space $\mathbb{F}^m$. Hence by Theorem 3.5, $d_{B,C}(A) = \det A \det\left(\begin{smallmatrix} I & C \\ 0 & B \end{smallmatrix}\right)$. Now keep C fixed. The map $B \mapsto \det\left(\begin{smallmatrix} I & C \\ 0 & B \end{smallmatrix}\right)$ is a volume form on the rowspace $\mathbb{F}^k$. Hence $\det\left(\begin{smallmatrix} I & C \\ 0 & B \end{smallmatrix}\right) = \det B \left(\begin{smallmatrix} I & C \\ 0 & I \end{smallmatrix}\right)$. Now $\det\left(\begin{smallmatrix} I & C \\ 0 & I \end{smallmatrix}\right) = 1$ as $\left(\begin{smallmatrix} I & C \\ 0 & I \end{smallmatrix}\right)$ is triangular. So $\det\left(\begin{smallmatrix} A & C \\ 0 & B \end{smallmatrix}\right) = \det A \det B$. □

Proof. Write $X = \left(\begin{smallmatrix} A & C \\ 0 & B \end{smallmatrix}\right)$ and expand the expression for the determinant.

$$\det \begin{pmatrix} A & C \\ 0 & B \end{pmatrix} = \sum_{\sigma \in \S_{m+n}} \varepsilon(\sigma) \prod_{j=1}^{m+n} x_{\sigma(j)j}$$

Note $x_{\sigma(j)j} = 0$ if $j \leq m$, $\sigma(j) > m$, so we only sum over σ with the following properties.

(i) For $j \in [1, m]$, $\sigma(j) \in [1, m]$. Here $x_{\sigma(j)j} = a_{\sigma_1(j)j}$ where $\sigma_1 \in S_m$ is the restriction of σ to $[1, m]$.

(ii) For $j \in [m+1, m+k]$, $\sigma(j) \in [m+1, m+k]$. Here, writing $l = j - m$, we have $x_{\sigma(j)j} = b_{\sigma_2(l)l}$ where $\sigma_2(l) = \sigma(m+l) - m$.

Noting also that $\varepsilon(\sigma) = \varepsilon(\sigma_1)\varepsilon(\sigma_2)$ for such σ, we obtain

$$\begin{aligned}\det\left(\begin{smallmatrix} A & C \\ 0 & B \end{smallmatrix}\right) &= \left(\sum_{\sigma_1 \in S_m} \varepsilon(\sigma_1) \prod_{j=1}^{m} a_{\sigma_1(j)j}\right) \left(\sum_{\sigma_2 \in S_k} \varepsilon(\sigma_2) \prod_{l=1}^{k} a_{\sigma_2(j)j}\right) \\ &= \det A \det B\end{aligned}$$

□

Lemma 3.17. Let $A = (a_{ij})$ be an $n \times n$ matrix. Write $A_{\widehat{ij}}$ for the $(n-1) \times (n-1)$ matrix obtained from A by deleting row i and column j.

(i) For fixed j, $\det A = \sum_{i=1}^{n} (-1)^{i+j} a_{ij} \det(A_{\widehat{ij}})$.

(ii) For fixed i, $\det A = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} \det(A_{\widehat{ij}})$.

Proof.

$$\begin{aligned}\det A &= d(A^{(1)}, \ldots, A^{(n)}) \\ &= \sum_{i=1}^{n} a_{ij} (-1)^{i+j-2} d \begin{pmatrix} 1 & * \\ 0 & A_{\widehat{ij}} \end{pmatrix} \\ &= \sum_{i=1}^{n} (-1)^{i+j} a_{ij} \det A_{\widehat{ij}},\end{aligned}$$

using Lemma 3.16. □

Remark. This can be used as a definition of determinant.

Definition. Let $A \in M_n(\mathbb{F})$. The adjugate $\operatorname{adj} A$ is the $n \times n$ matrix with (i, j) entry equal to $(-1)^{i+j} \det A_{\widehat{ij}}$.

Theorem 3.18. (i) $(\operatorname{adj} A)A = (\det A)I$.

(ii) If A is invertible then $A^{-1} = \frac{1}{\det A} \operatorname{adj} A$.

Proof. (i) By Lemma 3.17,

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

Also, for $j < k$,

$$\begin{aligned}0 &= \det(A^{(1)}, \ldots, A^{(k)}, \ldots, A^{(k)}, \ldots, A^{(n)}) \\ &= \sum_{i=1}^{n} (\operatorname{adj} A)_{ji} a_{ik} \\ &= (\operatorname{adj} A \cdot A)_{jk}.\end{aligned}$$

(ii) If A is invertible, then $\det A \neq 0$, so $\frac{1}{\det A} \operatorname{adj} A \cdot A = I$ and hence we deduce $A^{-1} = \frac{1}{\det A} \operatorname{adj} A$. □

Consider the system of linear equations $Ax = b$ with m equations and n unknowns. Here A is an $m \times n$ matrix, b is a column vector in $\mathbb{F}^m$. This has a solution if $\operatorname{rank} A = \operatorname{rank}(A|\, b)$. The solution is unique if and only if $n = \operatorname{rank} A = \operatorname{rank}(A|\, b)$, then the solution is $x = A^{-1}b$. To solve this equation, use Gaussian elimination. In the case $m = n$, there is another method.

Lemma 3.19 (Cramer's rule)**.** If $A \in M_n(\mathbb{F})$ is non-singular then $Ax = b$ has $x = (x_1, \ldots, x_n)^T$ with $x_i = \frac{1}{detA} \det A_{\widehat{ib}}$ for $i = 1, \ldots, n$ as its unique solution, where $A_{\widehat{ib}}$ is the matrix obtained from A by deleting column i and inserting b.

Proof. Assume x is the solution of $Ax = b$. Then

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

so $x_i = \frac{1}{detA} \det A_{\hat{i}b}$ for $i = 1, \ldots, n$. □

Corollary 3.20. If $A \in M_n(\mathbb{Z})$ with $\det A = \pm 1$ and if $b \in \mathbb{Z}^n$ then we can solve $Ax = b$ over $\mathbb{Z}$.

Chapter 4

Endomorphisms, matrices, eigenvectors

In this chapter, unless stated otherwise we assume V is a vector space over $\mathbb{F}$, where $\mathbb{F}$ is $\mathbb{R}$ or $\mathbb{C}$, and $\alpha : V \to V$ is a linear map.

Definition. Let $\alpha \in \mathrm{End}(V)$. α is diagonalisable if there is a basis B of V such that $[\alpha]_B$ is diagonal, i.e. whenever $i \neq j$ then $a_{ij} = 0$. α is triangulisable if there is a basis B of V such that $[\alpha]_B$ is upper triangular, i.e. whenever $i > j$ then $a_{ij} = 0$.

A square matrix is diagonalisable (resp. triangulisable) if it is conjugate to a diagonal (resp. upper triangular) matrix.

Definition. Let $\alpha \in \mathrm{End}(V)$. Then $\lambda \in \mathbb{F}$ is an eigenvalue of α if there exists a vector $v \in V$ with $v \neq 0$ and $\alpha(v) = \lambda v$. v is then called an eigenvector corresponding to λ.

Remark. λ is an eigenvalue of α if and only if $\alpha - \lambda\iota$ is singular, or $\det(\alpha - \lambda\iota) = 0$, equivalently. In particular, $\lambda = 0$ is an eigenvalue if and only if α is singular.

Definition. The polynomial $\chi_\alpha(t) = \det(\alpha - t\iota)$ is the characteristic polynomial of α. It is a polynomial of degree $n = \dim V$. If $A \in M_n(\mathbb{F})$, $\chi_\alpha(t) = \det(A - tI)$.

Remark. Eigenvalues of α are precisely the roots of the characteristic polynomial. For a matrix A, λ is an eigenvalue of A if $\chi_A(\lambda) = 0$ and $v \in \mathbb{F}^n$ is a corresponding eigenvector if $v \neq 0$ and $Av = \lambda v$.

Lemma 4.1. Conjugate matrices have the same characteristic polynomial and hence the same eigenvalues.

Proof.

$$\begin{aligned}\chi_{P^{-1}AP}(t) &= \det(P^{-1}AP - tI) \\ &= \det P^{-1} \det(A - tI) \det P \\ &= \chi_A(t).\end{aligned}$$

□

Remark. The matrix $\left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$ is not diagonalisable. The only diagonal 2×2 matrix with eigenvalues $1, 1$ is I, but this is self-conjugate.

Recall that every polynomial of degree at least 1 over $\mathbb{C}$ has a root (and hence in fact n roots, counting with multiplicities).

Lemma 4.2. If V is a finite dimensional vector space over $\mathbb{C}$ and $\alpha \in \mathrm{End}(V)$ then α has an eigenvector in V.

Theorem 4.3. Let V be a finite dimensional vector space over $\mathbb{C}$ and let $\alpha \in \mathrm{End}(V)$. There exists a basis B of V such that $[\alpha]_B$ is upper triangular. In other words, there exists a basis $B = \{v_1, \ldots, v_n\}$ with $\alpha(v_j) \in \langle v_1, \ldots, v_j \rangle$ for each $j = 1, \ldots, n$.

Proof. We prove this by induction on n. It is clear if $n = 1$, so assume $n > 1$. Since V is a vector space over $\mathbb{C}$, there exists $\lambda \in \mathbb{C}$ such that $\alpha - \lambda\iota$ is singular. Consider $U = \mathrm{Im}(\alpha - \lambda\iota) \lneq V$, where we know U is a proper subspace by the rank-nullity theorem. Then $\alpha(U) \subset U$.

$$\alpha(U) = \alpha((\alpha - \lambda\iota)(V)) = (\alpha - \lambda\iota)(\alpha V) \leq (\alpha - \lambda\iota)(V) = U.$$

Consider $\alpha' = \alpha|_U : U \to U$. This is a linear map and $\dim U < \dim V$. By induction, there exists a basis $B' = \{v_1, \ldots, v_k\}$ of U with $A' = [\alpha']_{B'}$ upper triangular. Extend this to a basis $B = \{v_1, \ldots, v_k, \ldots, v_n\}$ of V. We claim $[\alpha]_B$ is upper triangular. In fact,

$$[\alpha]_B = \begin{pmatrix} A' & * \\ 0 & \lambda I \end{pmatrix}.$$

If $1 \leq j \leq k$, $\alpha(v_j) = \alpha'(v_j) \in U$, so the entries in the first k columns are as claimed. If $j > k$, $(\alpha - \lambda\iota)(v_j) \in U$ by definition of U, so $\alpha(v_j) = \lambda v_j + u$ for some $u \in U$, so the first k entries of column j are the coordinates of u with respect to B' and the only other non-zero entry is λ in the position (j, j). □

Proof. Let v_1 be an eigenvector of α, say $\alpha(v_1) = \lambda v_1$. Let U be any complementary subspace to $\langle v_1 \rangle$ in V. For $v \in V$ we have $v = \lambda_v v_1 + u$ with $u \in U$ unique, $\lambda_v \in \mathbb{F}$. Write $\pi(v) = u$, a projection from V to U. For $u \in U$ define $\tilde{\alpha} : U \to U$ by $\tilde{\alpha}(u) = \pi(\alpha(u))$. Then $\tilde{\alpha} \in \mathrm{End}(U)$. By induction, there is a basis $v_2, \ldots, v_n$ of U with $\tilde{\alpha}(v_j) \in \langle v_2, \ldots, v_j \rangle$ for $2 \leq j \leq n$. Then $\alpha(v_j) = \lambda_{\alpha(v_j)} v_j + \tilde{\alpha}(v_j) \in \langle v_1, \ldots, v_j \rangle$ and $\alpha(v_1) \in \langle v_1 \rangle$. □

Theorem 4.4. Every square matrix over $\mathbb{C}$ is conjugate to an upper triangular matrix.

Remark. This is not true over $\mathbb{R}$, e.g. rotations other than $\pm I$ on $\mathbb{R}^2$.

Theorem 4.5. Let V be a finite dimensional vector space over a field $\mathbb{F}$ and let $\alpha \in \mathrm{End}(V)$. There exists a basis B of V such that $[\alpha]_B$ is upper triangular if and only if χ_α factorises into linear factors, i.e. if and only all roots of χ_α are in $\mathbb{F}$.

Proof. If

$$[\alpha]_B = \begin{pmatrix} a_{11} & & * \\ & \ddots & \\ 0 & & a_{nn} \end{pmatrix}$$

then $\chi_\alpha(t) = (a_{11} - t) \cdots (a_{nn} - t)$ with $a_{ij} \in \mathbb{F}$.

We prove the converse by induction on $\dim_{\mathbb{F}} V$. Let λ be an eigenvalue in $\mathbb{F}$, let $U = (\alpha - \lambda\iota)(V)$. Then $\alpha(U) \leq U \lneq V$. Let $\alpha' = \alpha|_U \in \mathrm{End}(U)$. Let B' be any basis for U and extend this to a basis B of V. Then

$$[\alpha]_B = \begin{pmatrix} [\alpha']_{B'} & * \\ 0 & \lambda I \end{pmatrix}.$$

Now $\chi_\alpha(t) = \chi_{\alpha'}(t)\chi_{\lambda I}(t)$. It follows that all roots of $\chi_{\alpha'}$ lie in $\mathbb{F}$, so we can use induction. Replace B' is necessary so that $[\alpha']_{B'}$ is upper triangular. □

Remark (See Examples Sheet 3)**.** Let $\alpha \in \mathrm{End}(V)$ and asume $U \lneq V$ with $\alpha(U) \leq U$. Let $B' = \{v_1, \ldots, v_k\}$ be a basis of U and extend this to a basis $B = \{v_1, \ldots, v_k, \ldots, v_n\}$ of V. Let $\bar{V} = V/U$, $\bar{v} = v + U$ for $v \in V$. Then $\bar{B} = \{\bar{v}_{k+1}, \ldots, \bar{v}_n\}$ is a basis of $\bar{V}$. We can define $\alpha' = \alpha|_U \in \mathrm{End}(U)$ and $\bar{\alpha} : \bar{V} \to \bar{V}, \bar{v} \mapsto \alpha(\bar{v})$, a well-defined endomorphism of $\bar{V}$. Then

$$[\alpha]_B = \begin{pmatrix} [\alpha']_{B'} & * \\ 0 & [\bar{\alpha}]_{\bar{B}} \end{pmatrix}.$$

Note that also $\chi_\alpha = \chi_{\alpha'} \cdot \chi_{\bar{\alpha}}$.

Remark. Let V be a finite dimensional vector space over $\mathbb{F}$, $\alpha \in \mathrm{End}(V)$ and B a basis for V. Then $[\alpha]_B$ is diagonal if and only if B consists of eigenvectors of α.

Lemma 4.6. If $\alpha \in \mathrm{End}(V)$ and $\lambda_1, \ldots, \lambda_k$ are distinct eigenvalues of α, put $V_j = N(\alpha - \lambda_j \iota)$, called the eigenspace of λ_j. Then the sum $V_1 + \cdots + V_k$ is direct, so if B_j is a basis for V_j then $\bigcup_{j=1}^k B_j$ is a basis of $V_1 + \cdots + V_k$. In particular, if $\sum_{j=1}^k \dim V_j = \dim V$ then $[\alpha]_B$ is diagonal and $V = V_1 \oplus \cdots \oplus V_k$.

Proof. We need to show that if $v_1 + \ldots + v_k = 0$ with $v_j \in V_j$ then $v_j = 0$ for $j = 1, \ldots, k$. Suppose not and let

$$v_1 + \cdots + v_j = 0$$

be the shortest non-trivial expression. Apply α and subtract λ_1 times the above expression,

$$\begin{aligned} &\alpha(v_1) + \cdots + \alpha(v_j) - \lambda_1 v_1 - \cdots - \lambda_1 v_j = 0 \\ \iff &(\lambda_2 - \lambda_1)v_2 \cdots + (\lambda_j - \lambda_1)v_j = 0, \end{aligned}$$

which is a shorter non-trivial expression, contradiction. So $\sum V_j = \bigoplus V_j$. The rest follows by Lemma 1.15. □

Theorem 4.7. Let V be a vector space over $\mathbb{F}$ of finite dimension. Then $\alpha \in \mathrm{End}(V)$ is diagonalisable over $\mathbb{F}$ if and only if its minimal polynomial has distinct linear factors.

Proof. If

$$[\alpha]_B = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_k \end{pmatrix}$$

with $\lambda_1, \ldots, \lambda_k$ distinct, put $p(t) = \prod_{j=1}^k (\lambda_j - t)$. If $v \in B$ then $\alpha(v) = \lambda_l v$ for some $l \leq k$. So $(\lambda_l \iota - \alpha)v = 0$, so $p(\alpha)(v) = 0$. But then $p(\alpha)$ is the 0 endomorphism since it is zero on B.

We need to show that each $v \in V$ is a sum of eigenvectors, the rest then follows from Lemma 4.6. Write $p(t) = \prod_{j=1}^k (\lambda_j - t)$ with $\lambda_1, \ldots, \lambda_k$ distinct. Put

$$\begin{aligned} p_j(t) &= (\lambda_1 - t) \cdots (\lambda_{j-1} - t)(\lambda_{j+1} - t) \cdots (\lambda_k - t) \\ h_j(t) &= \frac{p_j(t)}{p_j(\lambda_j)} \end{aligned}$$

then $h_j(\lambda_i) = \delta_{ij}$ for all $1 \leq i, j \leq k$. Hence $h(t) = \sum_{j=1}^{k} h_j(t) = 1$ as $h(t) - 1$ has degree less than k and $\lambda_1, \ldots, \lambda_k$ are k distinct roots. Let $v \in V$. Then

$$v = \iota(v) = h(\alpha)(v) = \sum_{j=1}^{k} h_j(\alpha)(v) = \sum_{j=1}^{k} v_j$$

where $v_j = h_j(\alpha)(v)$. Note that $(\alpha - \lambda_j \iota)v_j = 0$ since $p(\alpha) = 0$, so v_j is an eigenvector of α corresponding to λ_j. Hence every $v \in V$ is a sum of eigenvectors, and the assertion follows from Lemma 4.6. (The $h_j(\alpha)$ are orthogonal projections, see Examples Sheet 2 Question 8. For another proof, see Examples Sheet 2 Question 13.) □

Remark. Let $A \in M_n(\mathbb{F})$. Then $P^{-1}AP$ is diagonal for some P if and only if $p(A) = 0$ for some polynomial $p \in \mathbb{F}[t]$ with distinct linear factors. To find P, consider

$$P^{-1}AP = D = \begin{pmatrix} d_1 & & 0 \\ & \ddots & \\ 0 & & d_n \end{pmatrix}$$
$$AP = PD$$
$$AP^{(j)} = d_j P^{(j)}$$

so the jth column of P is an eigenvector of A corresponding to the eigenvalue d_j.

Theorem 4.8 (Simultaneous diagonalisation)**.** Let $\alpha_1, \alpha_2 \in \mathrm{End}(V)$. If they are both diagonalisable and if $\alpha_1\alpha_2 = \alpha_2\alpha_1$ then they are simultaneously diagonalisable, i.e. there exists a basis B of V such that $[\alpha_1]_B$, $[\alpha_2]_B$ are diagonal.

Proof. We have $V = V_1 \oplus \cdots \oplus V_k$, where each V_j is an eigenspace of α_1, say, $\alpha_1(v_j) = \lambda_j v_j$ for $v_j \in V_j$. Then $\alpha_2(V_j) \subset V_j$. If $v \in V_j$, $\alpha_1(\alpha_2(v)) = \alpha_2(\alpha_1(v)) = \alpha_2(\lambda_j v) = \lambda_j \alpha_2(v)$ so $\alpha_2(v) \in V_j$. Now $\alpha_2|_{V_j}$ is diagonalisable by Lemma 4.6, so there exists a basis B_j of V_j consisting of eigenvectors of α_2, which are also eigenvectors of α_1, of course. Putting these bases together gives a basis B consisting of eigenvectors of both α_1 and α_2. □

Remark. (i) The condition is necessary.

(ii) In fact, the statement is true for any number of commuting endomorphisms.

For a polynomial $p(t) \in \mathbb{F}[t]$,

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

where $a_i \in \mathbb{F}$ for $0 \leq i \leq n$. For $p(t), q(t) \in \mathbb{F}[t]$, addition and multiplication in $\mathbb{F}[t]$ are defined as follows.

$$p(t) = a_n t^n + \cdots + a_1 t + a_0$$
$$q(t) = b_m t^m + \cdots + b_1 t + b_0$$

Assuming $m \leq n$,

$$(p+q)(t) = a_n t^n + \cdots + (a_m + b_m)t^m + \cdots + (a_1 + b_1)t + (a_0 + b_0)$$
$$(pq)(t) = a_n b_m t^{n+m} + \cdots + (a_1 b_0 + a_0 b_1)t + a_0 b_0$$

The degree $\deg p$ of a polynomial p is the greatest l with $a_l \neq 0$ (and it is $-\infty$ if p is the 0 polynomial). Note that $\deg pq = \deg p + \deg q$.

There is a Euclidean algorithm in $\mathbb{F}[t]$. Given $a, b \in \mathbb{F}[t]$ with $b \neq 0$, there exists $q, r \in \mathbb{F}[t]$ such that $a = bq + r$ with $\deg r < \deg b$ or $r = 0$. For, if $a = a_n t^n + \cdots + a_0$, $b = b_m t^m + \cdots + b_0$ with $b_m \neq 0$, we may assume $n \geq m$ as otherwise we can take $q = 0$, $r = a$. Replace a by $a' = a - \frac{a_n}{b_m} t^{n-m} b$, then $\deg a' < \deg a$, so now we have $a' = bq' + r$ for some q', r with $\deg r < \deg b$. Now take $q = \frac{a_n}{b_m} t^{n-m} q'$ and keep repeating.

This has nice consequences, for example, $\mathbb{F}[t]$ has the unique factorisation property.

If $p \in \mathbb{F}[t]$ and $\lambda \in \mathbb{F}$ is a root, so $p(\lambda) = 0$, there exists $q \in \mathbb{F}[t]$ with $p(t) = (\lambda - t)q(t)$. λ is a root of p with multiplicity e if $(\lambda - t)^e$ divides p but $(\lambda - t)^{e+1}$ does not.

A polynomial of degree n has at most n roots counted with multiplicities.

If polynomials p_1, p_2 of degree less than n have n points in common then they are equal.

Lemma 4.9. (i) If $p, q \in \mathbb{F}[t]$, $\alpha \in \mathrm{End}(V)$ then $p(\alpha)q(\alpha) = q(\alpha)p(\alpha)$.

(ii) If $\alpha(v) = \lambda v$, $p \in \mathbb{F}[t]$ then $p(\alpha)(v) = p(\lambda)v$.

Lemma 4.10. Let V be a finite dimensional vector space over $\mathbb{F}$ with $\dim V = n$ and let $\alpha \in \mathrm{End}(V)$. There exists a non-zero polynomial p of degree at most n^2 with $p(\alpha) = 0$.

Proof. We have $\dim \mathrm{End}(V) = n^2$, so there exist $a_{n^2}, \ldots, a_1, a_0 \in \mathbb{F}$ not all zero so that

$$a_{n^2}\alpha^{n^2} + a_{n^2-1}\alpha^{n^2-1} + \cdots + a_1\alpha + a_0\iota = 0$$

as any n^2+1 endomorphisms are linearly dependent. Put $p(t) = a_{n^2}t^{n^2} + \cdots + a_1 t + a_0$. □

Definition. Let $\alpha \in \mathrm{End}(V)$. The minimal polynomial m_α of α is the monic polynomial of minimal degree such that $m_\alpha(\alpha) = 0$.

Lemma 4.11. If $\alpha \in \mathrm{End}(V)$ and $p \in \mathbb{F}[t]$ with $p(\alpha) = 0$, then m_α divides p.

Proof. $\mathbb{F}[t]$ is a Euclidean domain so we can write $p = m_\alpha q + r$ with $q, r \in \mathbb{F}[t]$ and $\deg r < \deg m_\alpha$ or $r = 0$. But $p(\alpha) = 0 = m_\alpha(\alpha)$ so $r(\alpha) = 0$, so $r = 0$ by the minimality of $\deg m_\alpha$. □

Corollary 4.12. The minimal polynomial is unique.

Theorem 4.13 (Caley-Hamilton)**.** Let V be a vector space of finite dimension, let $\alpha \in \mathrm{End}(V)$. Then $\chi_\alpha(\alpha) = 0$.

Theorem 4.14. Let $A \in M_n(\mathbb{F})$. Then $\chi_A(A) = 0$.

Proof. Let $A \in M_n(\mathbb{F})$. Then

$$(-1)^n \chi_A(t) = t^n + a_{n-1}t^{n-1} + \cdots + a_1 t + a_0 = \det(tI - A)$$

Now for any matrix B and its adjugate, $B \cdot \mathrm{adj}\, B = (\det B)I$. Hence as $\mathrm{adj}(tI - A)$ is a matrix of polynomials with matrix coefficients of degree less than n,

$$\begin{aligned}(tI - A)(B_{n-1}t^{n-1} + \cdots + B_1 t + B_0) &= (tI - A)\,\mathrm{adj}(tI - A) \\ &= (t^n + a_{n-1}t^{n-1} + \cdots + a_1 t + a_0)I\end{aligned}$$

$$= It^n + a_{n-1}It^{n-1} + \cdots + a_1It + a_0I$$

Comparing coefficients,

$$\begin{aligned} I &= B_{n-1} \\ a_{n-1}I &= B_{n-2} - AB_{n-1} \\ &\vdots \\ a_1I &= B_0 - AB_1 \\ a_0I &= -AB_0 \end{aligned}$$

Pre-multiply the jth row by A^{n+1-j} and add all rows,

$$A^n + a_{n-1}A^{n-1} + \cdots a_1A + a_0I = 0,$$

the zero matrix. □

Proof (over $\mathbb{C}$). Let $\alpha \in \mathrm{End}(V)$, let $B = \{v_1, \ldots, v_n\}$ be a basis of V with $\alpha(v_j) \in \langle v_1, \ldots, v_j\rangle = U_j$. Hence

$$[\alpha]_B = \begin{pmatrix} \lambda_1 & & * \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}$$
$$\chi_\alpha(t) = (\lambda_1 - t)\cdots(\lambda_n - t)$$

Then $(\alpha - \lambda_j\iota)U_j \subset U_{j-i}$. Hence

$$\begin{aligned} &(\alpha - \lambda_1\iota)\cdots(\alpha - \lambda_{n-1}\iota)(\alpha - \lambda_n\iota)V \\ \subset &(\alpha - \lambda_1\iota)\cdots(\alpha - \lambda_{n-1}\iota)U_{n-1} \\ \subset &\cdots \\ \subset &(\alpha - \lambda_1\iota)U_1 = 0. \end{aligned}$$

□

Corollary 4.15. m_α divides χ_α.

Lemma 4.16. Let V be a vector space over $\mathbb{C}$ with $\dim V = n$. We have

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

with $\lambda_1, \ldots, \lambda_k$ all distinct eigenvalues of α. Here a_j is the algebraic multiplicity of λ_j. Then $\sum_{j=1}^{k} a_j = n$.

Lemma 4.17. $m_\alpha(t) = \prod_{j=1}^{k}(t - \lambda_j)^{e_j}$ for some e_j with $1 \le e_j \le a_j$ for $1 \le j \le k$.

Proof. m_α divides χ_α, so $e_j \le a_j$ for $1 \le j \le k$. If λ is an eigenvalue, $\alpha(v) = \lambda v$ with $v \neq 0$, then $0 = m_\alpha(\alpha)v = m_\alpha(\lambda)v$, and as $v \neq 0$ we have $m_\alpha(\lambda) = 0$, so $(t - \lambda)$ divides $m_\alpha(t)$. □

Theorem 4.18. Let V be a vector space over $\mathbb{F}$ of finite dimension, let $\alpha \in \mathrm{End}(V)$. Suppose α has distinct eigenvalues $\lambda_1, \ldots, \lambda_k$. Then α is diagonalisable if and only if $m_\alpha(t) = \prod_{j=1}^{k}(t - \lambda_j)$.

Proof. α is diagonalisable if and only if $p(\alpha) = 0$ for some polynomial with distinct linear factors if and only if m_α is the product of distinct linear factors if and only if $m_\alpha(t) = \prod_{j=1}^{k}(t - \lambda_j)$. □

Lemma 4.19. Let V be a finite dimensional vector space, $a \in \mathrm{End}(V)$ and suppose $\lambda_1, \ldots, \lambda_k$ are the distinct eigenvalues of α. Then the λ_j-eigenspace is $N(\alpha - \lambda_j\iota)$. Define the geometric multiplicity of λ_j to be $g_j = \dim N(\alpha - \lambda_j\iota)$. Then $1 \leq g_j \leq a_j$. (Note α is diagonalisable if and only if $a_j = g_j$ for all $1 \leq j \leq k$.

Proof. Since λ_j is an eigenvalue we have $1 \leq g_j$. Let B be a basis containing $v_1, \ldots, v_{g_j}$ a basis of $N(\alpha - \lambda_j\iota)$. Then

$$[\alpha]_B = \begin{pmatrix} \lambda_j I_{g_j} & * \\ 0 & A' \end{pmatrix}$$

so $\chi_\alpha(t) = (\lambda_j - t)^{g_j}\chi_{A'}(t)$, so $g_j \leq a_j$. □

Remark. If $\chi_A(t) = (-1)^n t^n + a_{n-1}t^{n-1} + \cdots + a_0$ then $a_0 = \det A$ and $a_{n-1} = (-1)^{n-1}\operatorname{tr} A$.

Consider a finite dimensional vector space V over $\mathbb{C}$ and linear maps in $\mathrm{End}(V)$. We have seen the diagonal form, but not all matrices are conjugate to such, and the triangular form, but this is not quite sparse enough, i.e. it is not visible whether two matrices in this form are conjugate. We now describe the Jordan Normal Form, which contains eigenvalues along the diagonal, only the entries 0 or 1 just above the diagonal and entries 0 elsewhere.

Define a Jordan block $J(s, \lambda)$ as follows.

$$J(s,\lambda) = \begin{pmatrix} \lambda & 1 & & 0 \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ 0 & & & \lambda \end{pmatrix}_{s \times s}$$

Then the characteristic polynomial of $J(s, \lambda)$ is $(\lambda - t)^s$, the minimal polynomial is $(\lambda - t)^s$ and the dimension of the λ-eigenspace is 1.

Theorem 4.20. Suppose V is a vector space over $\mathbb{C}$, let $\alpha \in \mathrm{End}(V)$. With resepect to some basis B, the matrix $A = [\alpha]_B$ is in JNF, it is of block diagonal form as follows.

$$A = \begin{pmatrix} B_1 & & 0 \\ & \ddots & \\ 0 & & B_k \end{pmatrix}$$

There is one $a_j \times a_j$ block B_j for each eigenvalue λ_j where $1 \leq j \leq k$. Now fix $\lambda = \lambda_j$. The corresponding block B_j has block diagonal form

$$B_j = \begin{pmatrix} C_1 & & 0 \\ & \ddots & \\ 0 & & C_m \end{pmatrix}$$

where $m = m_j$ and each $C_l = J(n_l, \lambda)$ with

$$n_1 \geq n_1 \geq \cdots n_m > 0$$
$$n_1 + n_2 + \cdots + n_m = 0$$

a partition of a_j.

Remark. $a_j = \sum_{i=1}^{m_j} n_i$, $g_j = m_j$, $e_j = n_1$.

Theorem 4.21. Every square matrix over $\mathbb{C}$ is conjugate to a matrix in JNF, and this is unique up to rearranging the $\lambda_1, \ldots, \lambda_k$.

Proof. See *Groups, Rings & Modules.* □

Firstly, we break up V into generalised eigenspaces $W_j = N(\alpha - \lambda_j \iota)^{a_j}$. Then $V = \bigoplus_{j=1}^{k} W_j$ and taking a basis $B = \bigcup_{j=1}^{k} B_j$, where B_j is a basis for W_j, we obtain

$$[\alpha]_B = \begin{pmatrix} B_1 & & 0 \\ & \ddots & \\ 0 & & B_k \end{pmatrix}.$$

Setting $p_j(t) = (\lambda_j - t)^{-a_j} \prod_{r=1}^{k} (\lambda_r - t)^{a_r}$, there exist polynomials q_j with $\sum_{j=1}^{k} p_j q_j = 1$ and then $W_j = \mathrm{Im}(h_j(\alpha))$.

Secondly, note that $\alpha(W_j) \subset W_j$, so we may restrict to W_j and $\alpha|_{W_j} \in \mathrm{End}(W_j)$. Writing $\lambda = \lambda_j$, $V = W_j$ and $n = a_j$, we have $(\alpha - \lambda\iota)^n = 0$, so $\alpha - \lambda\iota$ is a nilpotent endomorphism. Now break V into cyclic blocks to obtain

$$\begin{pmatrix} \lambda & & & 0 \\ 1 & \ddots & & \\ & \ddots & \ddots & \\ 0 & & 1 & \lambda \end{pmatrix}.$$

The final vector of the basis is a λ-eigenvector. In fact, if $v_1, \ldots, v_m$ is the corresponding part of the basis, then under $\alpha - \lambda\iota$, $v_1 \mapsto v_2 \mapsto \ldots \mapsto v_m \mapsto 0$. (To obtain the transpose form of the Jordan block, take these vectors in the order order.)

Example. We list all possible Jordan Normal Forms in the case of $n = 3$ along with their respective characteristic and minimal polynomials.

	$\begin{pmatrix} \lambda_1 & & \\ & \lambda_2 & \\ & & \lambda_3 \end{pmatrix}$	$\begin{pmatrix} \lambda_1 & & \\ & \lambda_2 & 1 \\ & & \lambda_2 \end{pmatrix}$	$\begin{pmatrix} \lambda_1 & & \\ & \lambda_2 & \\ & & \lambda_2 \end{pmatrix}$
char.	$(\lambda_1 - t)(\lambda_2 - t)(\lambda_3 - t)$	$(\lambda_1 - t)(\lambda_2 - t)^2$	$(\lambda_1 - t)(\lambda_2 - t)^2$
min.	$(\lambda_1 - t)(\lambda_2 - t)(\lambda_3 - t)$	$(\lambda_1 - t)(\lambda_2 - t)^2$	$(\lambda_1 - t)(\lambda_2 - t)$
	$\begin{pmatrix} \lambda & & \\ & \lambda & \\ & & \lambda \end{pmatrix}$	$\begin{pmatrix} \lambda & 1 & \\ & \lambda & \\ & & \lambda \end{pmatrix}$	$\begin{pmatrix} \lambda & 1 & \\ & \lambda & 1 \\ & & \lambda \end{pmatrix}$
char.	$(\lambda - t)^3$	$(\lambda - t)^3$	$(\lambda - t)^3$
min.	$\lambda - t$	$(\lambda - t)^2$	$(\lambda - t)^3$

Example. In the case of $n = 4$, consider matrices with characteristic polynomial $(\lambda - t)^4$. The partitions of 4 are $4 = 3+1 = 2+2 = 2+1+1 = 1+1+1+1$. Hence the following matrices are possible.

$$\begin{array}{lccc} & \begin{pmatrix} \lambda & 1 & & \\ & \lambda & 1 & \\ & & \lambda & 1 \\ & & & \lambda \end{pmatrix} & \begin{pmatrix} \lambda & 1 & & \\ & \lambda & 1 & \\ & & \lambda & \\ & & & \lambda \end{pmatrix} & \begin{pmatrix} \lambda & 1 & & \\ & \lambda & & \\ & & \lambda & 1 \\ & & & \lambda \end{pmatrix} \\ \text{min.} & (\lambda - t)^4 & (\lambda - t)^3 & (\lambda - t)^2 \end{array}$$

$$\begin{array}{lcc} & \begin{pmatrix} \lambda & 1 & & \\ & \lambda & & \\ & & \lambda & \\ & & & \lambda \end{pmatrix} & \begin{pmatrix} \lambda & & & \\ & \lambda & & \\ & & \lambda & \\ & & & \lambda \end{pmatrix} \\ \text{min.} & (\lambda - t)^2 & \lambda - t \end{array}$$

In fact, $n((\alpha - \lambda\iota)^r)$ for various r will distinguish.

Example. Consider

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

then $\chi_A(t) = (2 - t)^4$ and

$$A - 2I = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 3 & 0 & 0 & -2 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 \end{pmatrix} \qquad (A - 2I)^2 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & -4 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

The minimal polynomial of A is $m_A(t) = (2 - t)^3$ and $n(A - 2I) = 2$. Hence the Jordan Normal Form of A is

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

To find a Jordan basis, take $v_3 \notin \ker(A - 2I)^2$, e.g. $v_3 = (0, 0, 1, 0)^T$. Under $A - 2I$,

$$v_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} \mapsto v_2 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 2 \end{pmatrix} \mapsto v_1 = \begin{pmatrix} 0 \\ -4 \\ 0 \\ 0 \end{pmatrix} \mapsto 0.$$

Take v_4 to be another eigenvector, e.g. $v_4 = (2, 0, 0, 3)^T$.

Chapter 5

Dual spaces

Definition. Let V be a finite dimensional vector space over $\mathbb{F}$. Then $V^* = \mathcal{L}(V, \mathbb{F})$ is the dual of V. The vectors of V^* are the linear functionals on V.

Lemma 5.1. Let V be a vector space over $\mathbb{F}$, let $B = \{e_1, \ldots, e_n\}$ be a basis of V. Then $B^* = \{\varepsilon_1, \ldots, \varepsilon_n\}$ is the basis of V^* dual to B, where $\varepsilon_j(e_k) = \delta_{jk}$.

Proof. If $\sum_{j=1}^n \lambda_j \varepsilon_j = 0$ then for all $k = 1, \ldots, n$ we have $\lambda_k = (\sum_{j=1}^n \lambda_j \varepsilon_j)(e_k) = 0$, so B^* is independent. If $\varepsilon \in V^*$ then $\varepsilon = \sum_{j=1}^n \varepsilon(e_j)\varepsilon_j$, so B^* spans V^*. □

Remark. If $\varepsilon = \sum_{j=1}^n a_j \varepsilon_j$, $v = \sum_{j=1}^n x_j e_j$, then

$$\varepsilon(v) = \sum_{j=1}^n a_j x_j = \begin{pmatrix} a_1 & \ldots & a_n \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}.$$

So we can think of the dual of $\mathbb{F}^n$ as the space pf n-rows.

Definition. If $U \leq V$ let $U^\circ = \{\varepsilon \in V^* : \varepsilon(u) = 0 \ \forall u \in U\}$. U° is the annihilator of U in V^*.

Lemma 5.2. (i) If $U \leq V$ then $U^\circ \leq V^*$.

(ii) If $U \leq V$ then $\dim U + \dim U^\circ = \dim V$.

Proof. (i) This is clear.

(ii) Let $U \leq V$, let $e_1, \ldots, e_k$ be a basis for U, let $B = \{e_1, \ldots, e_k, \ldots, e_n\}$ be a basis for V. We claim $U^\circ = \langle \varepsilon_{k+1}, \ldots, \varepsilon_n \rangle$, where $\varepsilon_1, \ldots, \varepsilon_n$ is the basis of V^* dual to B. If $i > k$ then $\varepsilon_i(e_j) = 0$ for $j \leq k$, so $\varepsilon_i \in U^\circ$. If $\varepsilon \in U^\circ$ then $\varepsilon = \sum_{j=1}^n \lambda_j \varepsilon_j$ and then for $j \leq k$ we have $\lambda_j = \varepsilon(e_j) = 0$, so $\varepsilon \in \langle \varepsilon_{k+1}, \ldots, \varepsilon_n \rangle$. □

Lemma 5.3. Let U, V be vector spaces over $\mathbb{F}$, let $U \xrightarrow{\alpha} V$ be a linear map. Then the map $V^* \xrightarrow{\alpha^*} U^*$ given by $\alpha^*(\varepsilon) = \varepsilon \circ \alpha$ for $\varepsilon \in V^*$ is linear, it is the dual of α.

Proof. Certainly $\varepsilon \circ \alpha : U \to \mathbb{F}$ is linear, so $\alpha^* \in U^*$. If $\theta_1, \theta_2 \in V^*$ then

$$\begin{aligned} \alpha^*(\theta_1 + \theta_2) &= (\theta_1 + \theta_2) \circ \alpha \\ &= \theta_1 \circ \alpha + \theta_2 \circ \alpha \end{aligned}$$

$$= \alpha^*(\theta_1) + \alpha^*(\theta_2)$$

anf if $\lambda \in \mathbb{F}$, $\theta \in V^*$ then

$$\alpha^*(\lambda\theta) = (\lambda\theta) \circ \alpha = \lambda(\theta \circ \alpha) = \lambda\alpha^*(\theta). \qquad \square$$

Proposition 5.4. Let U, V be vector spaces over $\mathbb{F}$ of finite dimension. Let B, C be their bases, let B^*, C^* be bases of U^*, V^* dual to B, C. Let $\alpha \in \mathcal{L}(U, V)$ and let $\alpha^* \in \mathcal{L}(U^*, V^*)$ be its dual. Then $[\alpha^*]_{C^*,B^*} = [\alpha]_{B,C}^T$.

Proof. Let $B = \{b_1, \ldots, b_n\}$, $C = \{c_1, \ldots, c_m\}$ and $B^* = \{\beta_1, \ldots, \beta_n\}$, $C^* = \{\gamma_1, \ldots, \gamma_m\}$. Let $A = [\alpha]_{B,C}$ so that $\alpha(b_j) = \sum_{i=1}^n a_{ij}c_i$. Then

$$\begin{aligned}
\alpha^*(\gamma_r)(b_s) &= \gamma_r(\alpha(b_s)) \\
&= \gamma_r\left(\sum_{i=1}^n a_{is}c_i\right) \\
&= \sum_{i=1}^n a_{is}\gamma_r(c_i) \\
&= \sum_{i=1}^n a_{is}\delta_{ri} \\
&= a_{rs} \\
&= \left(\sum_{i=1}^n a_{ri}\beta_i\right)(b_s)
\end{aligned}$$

for all $s = 1, \ldots, n$, so $\alpha^*(\gamma_r) = \sum_{i=1}^n a_{ri}\beta_i$ and hence $[\alpha^*]_{C^*,B^*} = A^T$. $\square$

Corollary 5.5. We have $\det \alpha^* = \det \alpha$, $\chi_{\alpha^*} = \chi_\alpha$, $m_{\alpha^*} = m_\alpha$ since $\det A^T = \det A$ and $p(A^T) = p(A)^T$ for any polynomial p.

Lemma 5.6. Let U, V be vector spaces over $\mathbb{F}$ of finite dimension. Let $\alpha \in \mathcal{L}(U, V)$, let $\alpha^* \in \mathcal{L}(V^*, U^*)$ its dual. Then $\ker \alpha^* = (\operatorname{Im} \alpha)^\circ$. In particular, α^* is injective if and only if α is surjective.

Proof. Let $\varepsilon \in V^*$. Then $\varepsilon \in \ker \alpha^*$ iff $\alpha^*(\varepsilon)$ is the zero functional on U iff $\varepsilon \circ \alpha$ is the zero functional on U iff $\varepsilon \in \operatorname{Im}(\alpha)^\circ$. In particular, α^* is injective iff $\ker \alpha^* = \{0\}$ iff $(\operatorname{Im} \alpha)^\circ = \{0\}$ iff $\operatorname{Im} \alpha = V$ iff α is surjective. $\square$

Corollary 5.7. If $\alpha \in \mathcal{L}(U, V)$, then $\operatorname{rank} \alpha = \operatorname{rank} \alpha^*$. If $A \in M_{m,n}(\mathbb{F})$, then $\operatorname{rank} A = \operatorname{rank} A^T$.

Proof.

$$\begin{aligned}
\operatorname{rank} \alpha^* &= \dim V^* - n(\alpha^*) \\
&= \dim V - \dim(\operatorname{Im} \alpha)^\circ \\
&= \dim V - (\dim V - \dim \operatorname{Im} \alpha \\
&= \operatorname{rank} \alpha
\end{aligned}$$

Note this gives an algebraic proof of the equality between the column rank and the row rank of a matrix. $\square$

Lemma 5.8. We also have $\operatorname{Im}\alpha^* = (\ker\alpha)^\circ$.

Remark. The map $V^* \times V \to \mathbb{F}, (\varepsilon, v) \mapsto \varepsilon(v)$ is bilinear. Denote this by $\langle\varepsilon|v\rangle$. If $U \xrightarrow{\alpha} V$, $V^* \xrightarrow{\alpha^*} U^*$ then

$$\langle\alpha^*(\varepsilon)|u\rangle = \langle\varepsilon|\alpha(u)\rangle$$

for all $u \in U$, $\varepsilon \in V^*$. We have a map $\hat{} : V \to V^{**}, v \mapsto \hat{v}$, $\hat{v}(\varepsilon) = \varepsilon(v)$.

Theorem 5.9. If V is finite dimensional over $\mathbb{F}$, the map $\hat{} : V \to V^{**}, v \mapsto \hat{v}$ with $\hat{v}(\varepsilon) = \varepsilon(v)$ is an isomorphism.

Proof. Since $\hat{v} : V^* \to \mathbb{F}$ is linear, $\hat{}$ is linear.

$$\begin{aligned}(\widehat{\lambda_1 v_1 + \lambda_2 v_2})(\varepsilon) &= \varepsilon(\lambda_1 v_1 + \lambda_2 v_2) &&= \lambda_1\varepsilon(v_1) + \lambda_2\varepsilon(v_2)\\ &= \lambda_1\hat{v}_1(\varepsilon) + \lambda_2\hat{v}_2(\varepsilon)\\ &= (\lambda_1\hat{v}_1 + \lambda_2\hat{v}_2)(\varepsilon)\end{aligned}$$

for all $\varepsilon \in V^*$. $\hat{}$ is injective (and hence surjective since V is finite dimensional and $\dim V = \dim V^{**}$). If $e_1 \neq 0, e_1 \in V$, let $e_1, \ldots, e_n$ be a basis for V, let $\varepsilon_1, \ldots, \varepsilon_n$ be the basis of V^* dual to this. Then $\hat{e}_1(\varepsilon_1) = \varepsilon_1(e_1) = 1$, so $\hat{e}_1 \neq 0$. □

Remark. This is a natural isomorphism, it is independent of the bases.

Remark. If $\varepsilon_1, \ldots, \varepsilon_n$ is a basis of V^*, and $E_1, \ldots, E_n$ is the basis of V^{**} dual to this, then $E_j = \hat{e}_j$ for a unique $e_j \in V$ and $\varepsilon_1, \ldots, \varepsilon_n$ is the basis of V^* dual to the basis $e_1, \ldots, e_n$ of V.

Lemma 5.10. Let V be finite dimensional, let $U \leq V$. If we identify V and V^{**}, then $U = U^{\circ\circ}$. (More precisely, $\hat{U} = U^{\circ\circ}$.)

Proof. We first show $U \leq U^{\circ\circ}$. If $u \in U$ then $\varepsilon(u) = 0$ for all $\varepsilon \in U^\circ$, and hence $\hat{u}(\varepsilon) = 0$ for all $\varepsilon \in U^\circ$, whence $\hat{u} \in U^{\circ\circ}$. As also $\dim U = \dim U^{\circ\circ}$, it follows that $U = U^{\circ\circ}$. □

Lemma 5.11. If $U_1, U_2 \leq V$ where $\dim V$ is finite then

(i) $(U_1 + U_2)^\circ = U_1^\circ \cap U_2^\circ$;

(ii) $(U_1 \cap U_2)^\circ = U_1^\circ + U_2^\circ$.

Remark. The situation is different when V is not of finite dimension. Consider $V = P(\mathbb{R})$, the space of real polynomials. Then $V^* = \mathbb{R}^{\mathbb{N}}$, the space of real sequences, which is not isomorphism to $P(\mathbb{R}) = \langle p_0, p_1, \ldots\rangle$. Any element $\varepsilon \in V^*$ can be given as $(\varepsilon(p_0), \varepsilon(p_1), \ldots)$.

Chapter 6

Bilinear forms

Definition. Let U, V be vector spaces over $\mathbb{F}$. The function $\psi : U \times V \to \mathbb{F}$ is bilinear if it is linear in each coordinate. For fixed $u \in U$, $\psi(u, v)$ is linear in v, and for fixed $v \in V$, $\psi(u, v)$ is linear in u. (Here we are mainly concerned with the case $U = V$ and bilinear forms on V.)

Example. (i) $\mathbb{F} = \mathbb{R}$, $V = \mathbb{R}^n$, $\psi(x, y) = \sum_{i=1}^n x_i y_i = x^T y$.

(ii) $V = \mathbb{F}^n$, $A \in M_n(\mathbb{F})$, $\psi(u, v) = u^T A v$.

Definition. Let V be a vector space over $\mathbb{F}$ with $\dim V = n$. Let $B = \{v_1, \ldots, v_n\}$ be a basis of V. The matrix of the bilinear form ψ on V with respect to B is $A = (\psi(v_i, v_j)) = [\psi]_B$.

Lemma 6.1. If ψ is a bilinear form on V and B is a basis for V, then $\psi(u, v) = [u]_B^T [\psi]_B [v]_B$ for all $u, v \in V$.

Proof. Let $B = \{v_1, \ldots, v_n\}$. If $u = \sum_{i=1}^n a_i v_i$, $v = \sum_{i=1}^n b_i v_i$, then

$$\begin{aligned}
\psi(u, v) &= \psi(\sum a_i v_i, \sum b_j v_j) \\
&= \sum_{i,j} a_i b_j \psi(v_i, v_j) \\
&= \begin{pmatrix} a_1 & \cdots & a_n \end{pmatrix} [\psi]_B \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix} \\
&= [u]_B^T [\psi]_B [v]_B.
\end{aligned}$$

Moreover, $[\psi]_B$ is the only matrix for which this holds for all $u, v \in V$. If $\psi(u, v) = [u]_B^T A [v]_B$ for all $u, v \in V$, apply this for $u = v_i$, $v = v_j$ to obtain $A_{ij} = \psi(v_i, v_j)$. $\square$

Let $B = \{v_1, \ldots, v_n\}, B' = \{v_1', \ldots, v_n'\}$ be bases for V and let P be the change of basis matrix from B to B'. Then

$$v_j' = \sum_{i=1}^n p_{ij} v_i \qquad\qquad [v]_B = P[v]_{B'}$$

Theorem 6.2. $[\psi]_{B'} = P^T [\psi]_B P$.

Proof. For all $u, v \in V$,

$$\begin{aligned}\psi(u,v) &= [u]_B^T [\psi]_B [v]_B \\ &= (P[u]_{B'})^T [\psi]_B (P[v]_{B'}) \\ &= [u]_{B'}^T P^T [\psi]_B P [v]_{B'},\end{aligned}$$

so $[\psi]_{B'} = P^T [\psi]_B P$ by uniqueness of $[\psi]_{B'}$. □

Definition. The real square matrices A, B are congruent if $B = P^T A P$ for some invertible matrix P (i.e. if they represent the same bilinear form with respect to different bases).

Lemma 6.3. Congruence is an equivalence relation on $M_n(\mathbb{R})$.

Definition. The rank of a bilinear form is the rank of any matrix representing it. (This is independent of the choice of basis.)

Definition. The real bilinear form ψ is symmetric if $\psi(u,v) = \psi(v,u)$ for all $u, v \in V$. This is equivalent to $A = [\psi]_B$ being symmetric, i.e. $A = A^T$.

Remark. If $P^T A P = D$ for some non-singular P then $A = A^T$.

Definition. Let V be a real vector space. The function $Q : V \to \mathbb{R}$ is a quadratic form on V if $Q(\lambda v) = \lambda^2 Q(v)$ for all $\lambda \in \mathbb{R}, v \in V$, and there exist a symmetric bilinear form ψ on V so that $Q(v) + Q(w) + 2\psi(v,w) = Q(v+w)$ for all $v, w \in V$.

We can start with a symmetric bilinear form ψ and define $Q(v) = \psi(v,v)$. Conversely, we can start with a quadratic form Q and let $\psi(v,w) = \frac{1}{2}(Q(v+w) - Q(v) - Q(w))$.

Theorem 6.4. Any real symmetric bilinear form can be represented by a diagonal matrix of the form

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

Any real symmetric matrix is congruent to a diagonal matrix of this form.

Proof. First we prove by induction on $\dim V = n$ that any symmetric bilinear form can be represented by a diagonal matrix. If $\psi(u,v) = 0$ for all $u, v \in V$ then $[\psi]_B = 0$ for any basis B. Suppose this is not the case. Then there exists $e \in V$ with $\psi(e,e) \neq 0$. (Otherwise $2\psi(u,v) = \psi(u+v, u+v) - \psi(u,u) - \psi(v,v)$ is 0 for all $u, v \in V$.) Let $W = \{v \in V : \psi(e,v) = 0\}$. Then $V = \langle e \rangle \oplus W$. (If $v \in V$ then $v = \lambda e + (v - \lambda e)$ for all $\lambda \in \mathbb{R}$. Choose λ so that $v - \lambda e \in W$, by taking $\lambda = \frac{\psi(e,v)}{\psi(e,e)}$. Observe $\langle e \rangle \cap W = \{0\}$ since $\psi(e, \lambda e) \neq 0$ for $\lambda \neq 0$.) Now the restriction ψ' of ψ to W is a symmetric bilinear form, so by induction, there exists a basis $e_2, \ldots, e_n$ of W with respect to which ψ' is diagonal, say the matrix representing ψ' is

$$\begin{pmatrix} d_2 & & 0 \\ & \ddots & \\ 0 & & d_n \end{pmatrix}.$$

Let $B_0 = \{e_1, e_2, \ldots, e_n\}$ where $e_1 = e$. Then $[\psi]_{B_0}$ is

$$\begin{pmatrix} d_1 & 0 & & 0 \\ 0 & d_2 & & \\ & & \ddots & \\ 0 & & & d_n \end{pmatrix},$$

where $d_1 = \psi(e, e)$. This completes the diagonalisation part of the proof.

Now reorder B_0 if necessary so that $d_1, \ldots, d_p > 0$, $d_{p+1}, \ldots, d_{p+q} < 0$ and $d_i = 0$ for all $i > p + q$. Now for $1 \leq i \leq p + q$ replace e_i by $\frac{1}{\sqrt{|d_i|}} e_i$ to obtain B. Then

$$[\psi]_B = \begin{pmatrix} I_p & & 0 \\ & -I_q & \\ 0 & & 0 \end{pmatrix}. \qquad \square$$

Remark. Note that $\operatorname{rank} \psi = p + q$.

Definition. The signature of ψ is $s(\psi) = p - q$.

Remark. We have $p = \frac{1}{2}(r + s)$, $q = \frac{1}{2}(r - s)$. Note that some authors call (p, q) the signature.

Theorem 6.5 (Sylvester's law of inertia)**.** If the real symmetric bilinear form ψ is represented by

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

with respect to different bases then $p = p'$, $q = q'$ and so the rank and signature are well-defined.

Proof. Let $B = \{v_1, \ldots, v_p, v_{p+1}, \ldots, v_{p+q}, v_{p+q+1}, \ldots, v_n\}$ with

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

Let $X = \langle v_1, \ldots, v_p \rangle$, $Y = \langle v_{p+1}, \ldots, v_n \rangle$. We show that ψ is positive definite on X and that p is the largest dimension of any subspace of V on which ψ is positive definite. $Q(\sum_{i=1}^{p} \lambda_i v_i) = \sum_{i=1}^{p} \lambda_i^2 \geq 0$ with equality if and only if $\sum_{i=1}^{p} \lambda_i v_i = 0$, so ψ is positive definite on X. If $X' \leq V$ with $\dim X' = p'$ and ψ is positive definite on X' then $X' \cap Y = \{0\}$ as ψ is negative semidefinite on Y. Hence $\dim X' + \dim Y \leq n$. So $p' = \dim X' \leq n - \dim Y = p$.

Similarly, if $N = \langle v_{p+1}, \ldots, v_{p+q} \rangle$ then ψ is negative definite on N and q is the largest dimension of a subspace of V on which ψ is negative definite.

Hence p and q are independent of the choice of basis. $\square$

Remark. p is determined by ψ, but there may be many subspaces like X of dimension p, on which ψ is positive definite. (Similarly for q and N.)

Remark. Let $t = \min\{p, q\}$. Then the restriction of ψ to the subspace $\langle v_1 + v_{p+1}, \ldots, v_t + v_{p+t}, v_{p+q+1}, \ldots, v_n\rangle$ is 0, and $n - \max\{p, q\}$ is the maximal dimension of any subspace on which ψ is 0. For is $\psi = 0$ on $U \leq V$ then $U \cap X = \{0\} = U \cap N$, so $\dim U \leq n - p$ and $\dim U \leq n - q$.

Definition. The kernel of ψ is $\{v \in V : \psi(v, w) = 0 \ \forall w \in V\}$. In the above, $\ker \psi = \langle v_{p+q+1}, \ldots, v_n\rangle$.

Definition. The real symmetric bilinear form ψ is non-singular if $\ker \psi = \{0\}$. Equivalently, $[\psi]_B$ with respect to any basis B is non-singular. Also equivalently, $n = p + q$.

Example. Consider the quadratic form Q on $V = \mathbb{R}^3$ where

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

The matrix of Q with respect to the standard basis is

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

(i) (Completing squares.)

$$\begin{aligned} Q(x_1, x_2, x_3) &= x_1^2 + x_2^2 + 2x_3^2 + 2x_1x_2 + 2x_1x_3 - 2x_2x_3 \\ &= (x_1 + x_2 + x_3)^2 + x_3^2 - 4x_2x_3 \\ &= (x_1 + x_2 + x_3)^2 + (x_3 - 2x_2)^2 - (2x_2)^2 \end{aligned}$$

Hence $\operatorname{rank}(Q) = 3$, $s(Q) = 2 - 1 = 1$. Also

$$P^{-1} = \begin{pmatrix} 1 & 1 & 1 \\ 0 & -2 & 1 \\ 0 & 2 & 0 \end{pmatrix} \qquad P^TAP = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}$$

(ii) (Elementary matrices.) We want P invertible such that P^TAP is diagonal. Apply elementary column operations followed immediately by the corresponding elementary row operation. Then

$$A \to E_1^TAE_1 \to \ldots \to E_k^T \cdots E_1^TAE_1 \cdots E_k = D$$

where $P = E_1 \cdots E_k$.

(iii) (From the proof of Theorem 6.4.) Choose e_1 with $Q(e_1) \neq 0$, e.g. $e_1 = (1, 0, 0)^T$, $Q(e_1) = 1$. Set $W = \{v \in V : \psi(e, v) = 0\} = \{(a, b, c)^T : a + b + c = 0\}$. Note $e_1^TA = (1, 1, 1)$. Choose $e_2 \in W$ with $Q(e_2) \neq 0$, e.g. $e_2 = (1, 0, -1)$, $Q(e_2) = 1$. Now choose $e_3 \in W$ so that $\psi(e_2, e_3) = 0$ so $e_3 = (a, b, c)^T$ with $a + b + c = 0$, $2b - c = 0$ and $e_2^TA = (0, 2, -1)$, so $e_3 = \frac{1}{2}(-3, 1, 2)^T$, $Q(e_3) = -1$.

(iv) (See Chapter 7.) We shall see that $s(Q)$ is the difference of the number of positive eigenvalues and the number of negative eigenvalues.

Now consider bilinear symmetric forms over $\mathbb{C}$. As before in Theorem 6.4, there exists a basis $e_1, \ldots, e_n$ such that the matrix representing ψ with respect to this basis is

$$\begin{pmatrix} d_1 & & 0 \\ & \ddots & \\ 0 & & d_n \end{pmatrix}.$$

Reorder this basis so that $d_1, \ldots, d_r \neq 0$ and $d_i = 0$ for all $i > r$. To normalise, replace e_j by $\frac{1}{\sqrt{d_j}} e_j$ for $1 \leq j \leq r$. Then the matrix representing ψ is

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

Lemma 6.6. Any symmetric complex matrix satisfies $P^T A P = \left(\begin{smallmatrix} I_r & 0 \\ 0 & 0 \end{smallmatrix}\right)$ for some invertible matrix P. We have $\operatorname{rank} A = r$.

Definition. Let V be a complex vector space. A complex Hermitian form on V is a function $\psi : V \times V \to \mathbb{C}$ such that

(i) for $u \in V$, the function $v \mapsto \psi(u, v)$ is linear;

(ii) for all $u, v \in V$, $\psi(u, v) = \overline{\psi(v, u)}$.

Remark. Note that ψ is not bilinear, it is sesquilinear, i.e.

$$\psi(u, \lambda_1 v_1 + \lambda_2 v_2) = \lambda_1 \psi(u, v_1) + \lambda_2 \psi(u, v_2)$$
$$\psi(\lambda_1 u_1 + \lambda_2 u_2, v) = \overline{\lambda_1} \psi(u_1, v) + \overline{\lambda_2} \psi(u_2, v)$$

and complex-symmetric, i.e. $\psi(u, v) = \overline{\psi(v, u)}$.

Example. Inner products of complex inner product spaces are complex Hermitian forms, see Chapter 7.

Remark. Given a complex Hermitian form ψ on V, we can define a complex quadratic form $Q : V \to \mathbb{C}$, $Q(v) = \psi(v, v)$. Note that $Q(v) \in \mathbb{R}$ for all $v \in V$. Here we have $Q(\lambda v) = |v|^2 Q(v)$. we can recover ψ as follows.

$$\psi(u, v) = \tfrac{1}{4}(Q(u + v) - Q(u - v) - iQ(u + iv) + iQ(u - iv))$$

If $B = \{v_1, \ldots, v_n\}$ is a basis of V, the matrix of ψ with respect to B is $[\psi]_B = (\psi(v_i, v_j)) = A$. Then $\psi(u, v) = \overline{[\psi]}_B^T [\psi]_B [v]_B$. Note that $A = \bar{A}^T$, the matrix is Hermitian. Finally, a change of basis results in the matrix $\bar{P}^T A P$ with a non-singular matrix P, the change of basis matrix.

Theorem 6.7. If ψ is a Hermitian form on a complex vector space V, there is a basis B of V with respect to which ψ has diagonal matrix

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

Moreover, p and q are uniquely determinend by ψ, independent of B. So if the basis is $v_1, \ldots, v_n$ then

$$Q(\sum_{i=1}^{n} \xi_i v_i) = |\xi_1|^2 + \cdots + |\xi_p|^2 - |\xi_{p+1}|^2 - \cdots - |\xi_{p+q}|^2.$$

Proof. We can prove this as before for Theorem 6.4 and Theorem 6.5. □

Definition. ψ is positive definite (on V) if $p = n$.

Remark. The bilinear forms we have considered were mainly symmetric. The next important class are skew-symmetric bilinear forms satisfying $\psi(u, v) = -\psi(v, u)$ for all $u, v \in V$ and hence $\psi(u, u) = 0$ for all $u \in V$. If $A = [\psi]_B$ for some basis B of V then $A^T = -A$, and A is called skew-symmetric.

Remark. Any real matrix A can be expressed as the sum of a symmetric and skew-symmetric matrix,

$$A = \frac{1}{2}(A + A^T) + \frac{1}{2}(A - A^T)$$

Theorem 6.8. If ψ is a skew-symmetric bilinear form on a real vector space V, there exists a basis $v_1, w_1, \ldots, v_m, w_m, v_{2m+1}, \ldots, v_n$ with respect to which ψ has the block diagonal matrix

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

Any skew-symmetric matrix is congruent to a matrix of this form.

Proof (Sketch). Prove this by induction on $\dim V = n$. If $\psi = 0$ the result is clear. Otherwise take v_1, w_1 so that $\psi(v_1, w_1) = 1$, so also $\psi(w_1, v_1) = -1$. Let $U = \langle v_1, w_1 \rangle$, $W = \{v \in V : \psi(v_1, v) = 0 = \psi(w_1, v)\}$. Then $V = U \oplus W$. If $v = (av_1 + bw_1) + (v - av_1 - bw_1)$ take $a = \psi(v, w_1), b = (\psi(v_1, v)$ to obtain $v - av_1 - bw_1 \in W$. Also $U \cap W = \{0\}$ since if $av_1 + bw_1 \in W$ then $\psi(av_1 + bw_1, av_1 - bw_1) = a^2 + b^2$. Now finish by working in W. □

Remark. (i) If ψ is non-singular skew-symmetric, then $n = 2m$.

(ii) Reordering the basis B as $v_1, \ldots, v_m, w_1, \ldots, w_m, v_{2m+1}, \ldots, v_n$, we obtain

$$\begin{pmatrix} 0 & I_m & 0 \\ -I_m & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Remark (*)**.** Here are some remarks on general bilinear forms on $U \times V$, where U, V are vector spaces over the same field $\mathbb{F}$. We assume $\psi : U \times V \to \mathbb{F}$ is linear in each coordinate. Examples include

(i) $U = V^*$, $\psi : V^* \times V \to \mathbb{F}, (\alpha, v) \mapsto \alpha(v)$;

(ii) $\mathbb{F} = \mathbb{C}$, define $\bar{V}$ to be the space with elements as V, addition as in V and multiplication $\cdot$ by $\lambda \cdot v = \bar{\lambda} v$. (Then sesquilinear forms are just bilinear forms on $\bar{V} \times V$.)

Suppose $\psi : U \times V \to \mathbb{F}$ is bilinear. Define

$$\psi_L : U \to V^*, u \mapsto (\psi_L(u) : v \mapsto \psi(u, v))$$
$$\psi_R : V \to U^*, v \mapsto (\psi_R(v) : u \mapsto \psi(u, v))$$

ψ is said to be non-singular if $\ker \psi_L = \{0\}$, $\ker \psi_R = \{0\}$. Note that

(i) if ψ is non-singular then $\dim U = \dim V$ for $\dim U \leq V^* = \dim V$, $\dim V \leq \dim U^* = \dim U$;

(ii) if $\dim U = \dim V$, then $\ker \psi_L = \{0\}$ if and only if $\ker \psi_R = \{0\}$.

If ψ is a bilinear non-singular form on $U \times V$, let $u_1, \ldots, u_n$ be a basis of U. Then $\psi_L(u_1), \ldots, \psi_L(u_n)$ is a basis of V^*. Let $v_1, \ldots, v_n$ be the basis of V dual to this. Then $\psi(u_i, v_j) = \delta_{ij}$.

Lemma 6.9 (*)**.** Let ψ be a non-singular bilinear form on V. For $W \leq V$, let $W^\perp = \{v \in V : \psi(w, v) = 0 \ \forall w \in W\}$. Then $W^\perp \leq V$ and $\dim W + \dim W^\perp = \dim V$.

Proof (Sketch). Let $u_1, \ldots, u_n$ be a basis of V containing a basis $u_1, \ldots, u_m$ of W. Let $v_1, \ldots, v_n$ be the basis paired to it as above. Then $W^\perp = \langle v_{m+1}, \ldots, v_n \rangle$.

Or more algebraically, consider

$$\psi_L : V \to V^*, u \mapsto (\psi_L(u) : V \to \mathbb{F}, v \mapsto \psi(u, v)).$$

This is an isomorphism as ψ is non-singular. Now $W^\perp = (\psi_L(W))^\circ$ as

$$\begin{aligned} v \in W^\perp &\iff \psi(w, v) = 0 \ \forall w \in W \\ &\iff \psi_L(w)(v) = 0 \ \forall w \in W \\ &\iff v \in (\psi_L(W))^\circ. \end{aligned}$$

Hence

$$\begin{aligned} \dim W + \dim W^\perp &= \dim W + \dim(\psi_L(W))^\circ \\ &= \dim W + \dim V - \dim \psi_L(W) \\ &= \dim V. \end{aligned}$$

□

Chapter 7

Inner product spaces

Definition. Let V be a real (resp. complex) space. An inner product on V is a positive definite symmetric bilinear (resp. Hermitian) form on V. Write $\langle v, w\rangle$ for the value of the form on $(v, w) \in V \times V$. A vector space V equipped with an inner product is a real (resp. complex) inner product space, also called a Euclidean (resp. unitary) space. So $\langle , \rangle$ is a real symmetric bilinear (resp. Hermitian) form such that $\langle v, v\rangle > 0$ for all $v \in V \setminus \{0\}$.

Definition. The length of v is defined to be $\|v\| = \sqrt{\langle v, v\rangle}$. We have $|v| > 0$ for $v \neq 0$.

Lemma 7.1 (Schwartz's inequality)**.** For all $v, w \in V$, $|\langle v, w\rangle| \leq \|v\|\|w\|$.

Proof. If $v = 0$ then the result is clear. So suppose $v \neq 0$ and consider the real and the complex case separately.

- (Real case) For all $t \in \mathbb{R}$, we have

$$0 \leq \|tv - w\|^2 = t^2\|v\|^2 - 2t\langle v, w\rangle + \|w\|^2.$$

  Set $t = \frac{\langle v,w\rangle}{\|v\|^2}$ to obtain the result.

- (Complex case) For all $t \in \mathbb{C}$, we have

$$0 \leq \|tv - w\|^2 = t\bar{t}\|v\|^2 - \bar{t}\langle v, w\rangle - t\overline{\langle v, w\rangle} + \|w\|^2.$$

  Set $t = \frac{\langle v,w\rangle}{\|v\|^2}$, so $\bar{t} = \frac{\overline{\langle v,w\rangle}}{\|v\|^2}$ to obtain the result. □

Lemma 7.2. In the Euclidean case, if $v, w \neq 0$, the angle θ between them is given by $\cos\theta = \frac{\langle v,w\rangle}{\|v\|\|w\|}$ taking $\theta \in [0, 2\pi)$.

Lemma 7.3 (Triangle inequality)**.** For all $v, w \in V$, $\|v + w\| \leq \|v\| + \|w\|$.

Proof.

$$\begin{aligned}\|v + w\|^2 &= \|v\|^2 + \langle v, w\rangle + \overline{\langle v, w\rangle} + \|w\|^2 \\ &\leq \|v\|^2 + 2\|v\|\|w\| + \|w\|^2 \\ &= (\|v\| + \|w\|)^2\end{aligned}$$

□

Remark. Defining $d(v, w) = \|v - w\|$, we obtain a metric on V.

Example. (i) Dot products on $\mathbb{R}^n$, $\mathbb{C}^n$;

(ii) $V = C[0,1]$, real or complex valued, $\langle (, f\rangle, g) = \int_0^1 \overline{f(t)} g(t)\, dt$.

Definition. A set $\{e_1, \ldots, e_k\}$ is orthogonal if $\langle e_i, e_j\rangle = 0$ whenever $i \neq j$. It is orthonormal if also $\|e_j\| = 1$ for all j. It is an orthonormal basis if it is also a basis.

Lemma 7.4. Orthonormal sets are linearly independent. In fact, if $v = \sum_{j=1}^k \lambda_j e_j$ then $\lambda_j = \langle e_j, v\rangle$.

Theorem 7.5 (Gram–Schmidt)**.** Let $v_1, \ldots, v_n$ be a basis for an inner product space V. THere exists an orthonormal basis $e_1, \ldots, e_n$ of V such that $\langle v_1, \ldots, v_k\rangle = \langle e_1, \ldots, e_k\rangle$ for all $1 \leq k \leq n$.

Proof. Let $e_1 = \frac{1}{\|v_1\|} v_1$. Suppose we have found $e_1, \ldots, e_k$. Now take $e'_{k+1} = v_{k+1} - \sum_{j=1}^k \lambda_j e_j$, with λ_j chosen such that $\langle e_j, e'_{k+1}\rangle = 0$ for $1 \leq j \leq k$. So $\lambda_j = \langle e_j, v_{k+1}\rangle$. Then $e'_{k+1} \neq 0$, as $v_1, \ldots, v_k, v_{k+1}$ are linearly independent. Let $e_{k+1} = \frac{1}{\|e'_{k+1}\|} e'_{k+1}$. Then $\langle e_j, e_{k+1}\rangle = 0$ for $1 \leq j \leq k$, and $\|e_{k+1}\| = 1$, and $\langle e_1, \ldots, e_{k+1}\rangle = \langle v_1, \ldots, v_{k+1}\rangle$. □

Remark. In calculations, it may be best to normalise only at the end. But then we have to adjust and take $\lambda_j = \frac{\langle e_j, v_{k+1}\rangle}{\langle e_j, e_j\rangle}$.

Corollary 7.6. In a finite dimensional inner product space, any orthonormal set of vectors can be extended to an orthonormal basis.

Proof. If $e_1, \ldots, e_k$ is an orthonormal set, it is linearly independent, so extend it to a basis $e_1, \ldots, e_k, v_{k+1}, \ldots, v_n$ of V. Apply the Gram–Schmidt process to obtain $e_1, \ldots, e_k, e_{k+1}, \ldots, e_n$ an orthonormal basis of V. □

Corollary 7.7. (i) Any real non-singular matrix A can be written as $A = RT$ with R orthogonal and T upper-triangular.

(ii) Any complex non-singular matrix A can be written as $A = UT$ with U unitary and T upper-triangular.

Proof. (i) Work in $\mathbb{R}^n$ with the dot product. Let $v_1, \ldots, v_n$ be the columns $A^{(1)}, \ldots, A^{(n)}$ of A. Let $e_1, \ldots, e_n$ be the orthonormal basis obtained from this y the Gram–Schmidt process. Let R be the matrix with columns $R^{(j)} = e_j$. Then $R^T R = I$. Let $v_k = \sum_{j=1}^n t_{jk} e_j$. Since $v_k \in \langle e_1, \ldots, e_k\rangle$, we see that the matrix $T = (t_{ij})$ is upper-triangular. We have $A = RT$ as $A^{(k)} = \sum_{j=1}^n t_{jk} R^(j)$.

(ii) For a complex non-singular matrix A, work over $\mathbb{C}^n$ with the dot product. Replace R by U with $\bar{U}^T U = I$ so U is unitary. □

Definition. Let V be an inner product space. If $W \leq V$, write $W^\perp = \{v \in V : v \perp w\ \forall w \in W\}$. This is the orthogonal complement for W in V.

Theorem 7.8. If V is a finite dimensional inner product space and $W \leq V$ then $V = W \oplus W^\perp$.

Proof. Let $e_1, \ldots, e_k$ be an orthonormal basis of W, extend this to $e_1, \ldots, e_k, e_{k+1}, \ldots, e_n$ an orthonormal basis of V. (Then $e_{k+1}, \ldots, e_n$ is an orthonormal basis for $W^\perp$.) Let $v \in V$. Write $v = \sum_{j=1}^{k} \lambda_j e_j + \sum_{j=k+1}^{n} \lambda_j e_j$, so $V = W + W^\perp$. And $W \cap W^\perp = \{0\}$ since if $\langle v, v \rangle = 0$ then $v = 0$ as the inner product is positive definite. □

Definition. Let V be an inner product space, let $W \leq V$. Then the endomorphism $\pi = \pi_W$ of V is an orthogonal projection onto W if $\pi^2 = \pi$, $W = \operatorname{Im} \pi$ and $W^\perp = \ker \pi$.

As above, let $e_1, \ldots, e_k$ be an orthonormal basis of W, extend this to $e_1, \ldots, e_k, e_{k+1}, \ldots, e_n$ an orthonormal basis of V. If $v = \sum_{j=1}^{n} \lambda_j e_j$ then $\pi_w(v) = \sum_{j=1}^{k} \lambda_j e_j$, that is, $\pi_w(v) = \sum_{j=1}^{k} \langle e_j, v \rangle e_j$.

Remark. Note that $\iota_V = \pi_W + \pi_{W^\perp}$, $\pi_W \pi_{W^\perp} = 0$.

Lemma 7.9. If $W \leq V$ as above and $v \in V$ then $\pi_W(v)$ is the vector in W nearest to v, that is letting $w_0 = \pi_W(v)$ we have $d(w_0, v) \leq d(w, v)$ for all $w \in W$.

Proposition 7.10. Let V be an inner product space of finite dimension, let $\alpha \in \operatorname{End}(V)$. There exists a unique endomorphism α^* of V, the adjoint of α, such that

$$\alpha^* : V \to V \qquad\qquad \langle \alpha v, w \rangle = \langle v, \alpha^* w \rangle$$

for all $v, w \in V$. Moreoever, if B is an orthonormal basis of V, then $[\alpha^*]_B = \overline{[\alpha]}_B^T$.

Proof. Let $B = \{e_1, \ldots, e_n\}$ be an orthonormal basis of V. Let $A = [\alpha]_B = (a_{ij})$. Let α^* be the endomorphism such that $[\alpha^*]_B = \bar{A}^T$. Then, writing $C = \bar{A}^T$, for $1 \leq i, j \leq n$, we have

$$\begin{aligned}
\langle \alpha e_i, e_j \rangle &= \langle \sum_{k=1}^{n} a_{ki} e_k, e_j \rangle \\
&= \sum_{k=1}^{n} \bar{a_{ki}} \langle e_k, e_j \rangle \\
&= \bar{a_{ji}} \\
&= c_{ij} \\
&= \sum_{k=1}^{n} c_{kj} \langle e_i, e_k \rangle \\
&= \langle e_i, \sum_{k=1}^{n} c_{kj} e_k \rangle \\
&= \langle e_i, \alpha^* e_j \rangle
\end{aligned}$$

Thus $\langle \alpha v, w \rangle = \langle v, \alpha^* w \rangle$ for all $v, w \in V$ since this is true on a basis. α^* is unique as $[\alpha^*]_B = \bar{A}^T$ is forced. □

Lemma 7.11. For adjoint endomorphisms, we have $(\alpha + \beta)^* = \alpha^* + \beta^*$, $(\lambda\alpha)^* = \bar{\lambda}\alpha^*$, $\alpha^{**} = \alpha$, $\iota^* = \iota$, $(\alpha\beta)^* = \beta^*\alpha^*$.

Proof. All of these follow from matrices, or directly, e.g. we have $\langle \alpha^* v, w \rangle = \langle v, \alpha^{**} w \rangle$ but also

$$\langle \alpha^* v, w \rangle = \overline{\langle w, \alpha^* v \rangle}$$

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

so $\langle v, \alpha w - \alpha^{**} w \rangle = 0$ for all $v \in V$ so $\alpha w = \alpha^{**} w$ for all $w \in V$, so $\alpha^{**} = \alpha$. □

Definition. Let A be a real (resp. complex) $n \times n$ matrix.

- A is symmetric (resp. Hermitian) if $A^T = A$ (resp. $\bar{A}^T = A$), A is self-adjoint.
- A is orthogonal (resp. unitary) if $A^T = A^{-1}$ (resp. $\bar{A}^T = A^{-1}$), A is inverse-adjoint.

Let V be an inner product space over $\mathbb{R}$ (resp. $\mathbb{C}$), let $\alpha \in \mathrm{End}(V)$.

- α is symmetric (resp. Hermitian) if $\alpha = \alpha^*$, equivalently $\langle \alpha v, w \rangle = \langle v, \alpha w \rangle$ for all $v, w \in V$.
- α is orthogonal (resp. unitary) if $\alpha^* = \alpha^{-1}$, equivalently $\langle \alpha v, \alpha w \rangle = \langle v, w \rangle$, so α is an isometry of V.
- α is normal if $\alpha\alpha^* = \alpha^*\alpha$.

Lemma 7.12. If V is an inner product space and $\alpha \in \mathrm{End}(V)$, and if B is an orthonormal basis of V, then α is Hermitian (unitary, symmetric, or orthogonal) if and only if $[\alpha]_B$ is.

Theorem 7.13 (Spectral theorem)**.** Let V be a complex inner product space, let $\alpha \in \mathrm{End}(V)$ and assume that α is Hermitian (unitary). There exists an orthonormal basis of V consisting of eigenvectors of α. Moreover, each eigenvalue is real (lies on the unit circle).

Proof. Since V is a vector space over $\mathbb{C}$, there exists an eigenvalue $\lambda \in \mathbb{C}$. Let $e \in V$ with $\alpha(e) = \lambda e$ and $\|e\| = 1$. Let $W = \langle e \rangle^{\perp}$. Then W is α-invariant. If $w \in W$ then

$$\langle \alpha(w), e \rangle = \langle w, \alpha^*(e) \rangle = \langle w, \alpha(e) \rangle = \langle w, \lambda e \rangle = \lambda \langle w, e \rangle = 0$$

and so $\alpha(w) \in W$. (If $w \in W$ then

$$\langle \alpha(w), e \rangle = \langle w, \alpha^*(e) \rangle = \langle w, \alpha^{-1}(e) \rangle = \langle w, \lambda^{-1} e \rangle = \lambda^{-1} \langle w, e \rangle = 0$$

and so $\alpha(w) \in W$.) Now $V = \langle e \rangle \oplus W$ by Theorem 7.8. Note that $\alpha|_W$ is Hermitian (unitary) and continue in W. There exists an orthonormal basis $e_2, \ldots, e_n$ of W consisting of eigenvectors of α. Then $B = \{e_1, e_2, \ldots, e_n\}$ is an orthonormal basis consisting of eigenvectors of α.

Moreover $[\alpha]_B = \overline{[\alpha]}_B^T$. Now $[\alpha]_B$ is diagonal,

$$[\alpha]_B = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix},$$

so $\bar{\lambda_j} = \lambda_j$ for all $1 \leq j \leq n$, so λ_j is real. (Or $[\alpha^{-1}]_B = \overline{[\alpha]}_B^T$, so $\bar{\lambda_j} = \lambda_j^{-1}$ for all $1 \leq j \leq n$, so $|\lambda_j| = 1$.) □

Lemma 7.14. If $\alpha \in \mathrm{End}(V)$ is Hermitian and if v_1, v_2 are eigenvectors of α corresponding to distinct eigenvalues λ_1, λ_2 then $v_1 \perp v_2$.

Proof. Consider $\langle \alpha(v_1), v_2 \rangle$.

$$\bar{\lambda_1}\langle v_1, v_2 \rangle = \langle \alpha(v_1), v_2 \rangle = \prod v_1 \alpha(v_2) = \lambda_2 \langle v_1, v_2 \rangle$$

so either $\lambda_1 = \bar{\lambda_1} = \lambda_2$ or $\langle v_1, v_2 \rangle = 0$. As $\lambda_1 \neq \lambda_2$, the result follows. □

Lemma 7.15. If α is a symmetric endomorphism of a real inner product space, then α has real eigenvalues. Also, eigenvalues corresponding to distinct eigenvalues are orthogonal.

Proof. Let B be any orthonormal basis of V. Then $[\alpha]_B$ is a real symmetric matrix, so also Hermitian (as a complex matrix). Hence the eigenvalues of $[\alpha]_B$, and so of α, are real by Theorem 7.13, and the rest follows from Lemma 7.14. □

Theorem 7.16. Let V be a real inner product space, let $\alpha \in \mathrm{End}(V)$ be symmetric. There is an orthonormal basis of V consisting of eigenvectors of α.

Proof. We prove this exactly as before for Theorem 7.13, but use Lemma 7.15 to get started. □

Remark. If V is a real inner product space and $\alpha \in \mathrm{End}(V)$ is orthogonal, there need not be a basis of eigenvectors. But Example Sheet 4 Question 14 shows that if α is orthogonal then there exists an orthonormal basis B of V such that

$$[\alpha]_B = \begin{pmatrix} 1 & 0 & & & & & \\ 0 & 1 & & & & & \\ & & -1 & 0 & & & \\ & & 0 & -1 & & & \\ & & & & \cos\theta_1 & \sin\theta_1 & \\ & & & & -\sin\theta_1 & \cos\theta_1 & \\ & & & & & & \ddots \end{pmatrix}$$

for some $\theta_i \in \mathbb{R}$.

Remark. Let A be a real symmetric (resp. complex Hermitian) matrix. Regard it as an endomorphism on the inner product space $\mathbb{R}^n$ (resp. $\mathbb{C}^n$) with the usual dot product. There exists an orthonormal basis $v_1, \ldots, v_n$ of eigenvectors of A. The matrix $P = (v_1, \ldots, v_n)$ is orthogonal (resp. unitary) and $AP = PD$ with D diagonal. Hence $P^{-1}AP = D = P^TAP$ (resp. $\bar{P}^TAP = D$).

Proposition 7.17. Let ψ be a real symmetric bilinear (resp. complex Hermitian) form on a real (resp. complex) vector space V. Let A be its matrix with respect to any basis B of V. Then the signature of ψ is the difference of the number of positive and the number of negative eigenvalues of A.

Proof. The matrix A is symmetric (resp. Hermitian), so by the previous remark there exists an orthogonal (resp. unitary) matrix P such that $P^{-1}AP = D = P^TAP$ (resp. $\bar{P}^TAP = D$). The assertian now follows. □

Remark. Let A be a real symmetric $n \times n$ matrix, let V be a real vector space of dimension n with a basis B. We can define $\alpha \in \mathrm{End}(V)$ with $[\alpha]_B = A$ and a bilinear symmetric form ψ with $[\psi]_B = A$. Changing to a different basis C will do different things to A. But let $\langle , \rangle$ be the inner product on V such that B is orthonormal. (Define $\langle e_i, e_j \rangle = \delta_{ij}$ and extend.) If C is also orthonormal with respect to $\langle , \rangle$ and P is the change of basis matrix, then $P^{-1} = P^T$ and $P^{-1}AP = P^TAP$, so $[\alpha]_C = [\psi]_C$.

Theorem 7.18 (Simultaneous diagonalisation of quadratic forms)**.** Let ψ and ϕ be symmetric bilinear (resp. Hermitian) forms on a real (resp. complex) vector space V. Assume ψ is positive definite. There exists a basis of V with respect to which both ψ and ϕ are diagonal.

Proof. Fix any basis and let A, C be the matrices representing ψ, ϕ. Now diagonalise ψ. For some non-singular matrix P, $P^TAP = I$ as ψ is positive definite. Now P^TCP is symmetric (resp. Hermitian) so for some orthogonal matrix Q, Q^TP^TCPQ is diagonal. Then $Q^TP^TAPQ = Q^TIQ = I$. Write

$$Q^TP^TCPQ = D = \begin{pmatrix} d_1 & & 0 \\ & \ddots & \\ 0 & & d_n \end{pmatrix}.$$

In fact, the diagonal entries $d_1, \ldots, d_n$ of D are the roots of the polynomial $\det(C - tA)$. They are certainly roots of $\det(D - tI)$ and

$$\det(D - tI) = \det((PQ)^2(C - tA)(PQ)) = (\det(PQ))^2 \det(C - tA),$$

so $\det(D - tI)$ and $\det(C - tA)$ have the same roots. □

We now provide a second proof of Proposition 7.10 in case of a real inner product space V.

Proof. Consider $\alpha \in \mathrm{End}(V)$, $\alpha^* \in \mathrm{End}(V)$, $\langle \alpha(v), w \rangle = \langle v, \alpha^*(w) \rangle$.

Fix $w \in V$; the map $\phi(w) : V \to \mathbb{F}, v \mapsto \overline{\langle v, w \rangle}$ is a linear functional on V. The map $\bar{V} \to V^*, w \mapsto \phi(w)$ is an isomorphism. ($\bar{V}$ is as V, but $\lambda \cdot v = \bar{\lambda} v$.) Hence any linear functional on V can be written for some unique $w' \in V$ as $v \mapsto \langle v, w' \rangle$.

Fix $w \in V$; now $v \mapsto \overline{\langle \alpha(v), w \rangle}$ is a linear functional on V, so there is a unique $w' = \alpha^*(w)$ with $\langle \alpha(v), w \rangle = \langle v, \alpha^*(w) \rangle$ for all $v, w \in V$. Finally, check that α^* is linear. □