University of Cambridge

Mathematics Tripos

Part IB

Linear Algebra

Michaelmas, 2017

Lectures by

A. M. Keating

Notes by

Qiangru Kuang

Contents

1 Vector Space

Convention. Throughout this course, $\mathbb{F}$ denotes a general field. If you wish, think of it as $\mathbb{R}$ or $\mathbb{C}$.

1.1 Definitions

> **Definition** (Vector space)**.** An $\mathbb{F}$-*vector space* (or a vector space over $\mathbb{F}$) is an abelian group $(V, +)$ equipped with a function, called *scalar multiplication*:
>
> $$\begin{aligned} \mathbb{F} \times V &\to V \\ (\lambda, v) &\mapsto \lambda \cdot v \end{aligned}$$
>
> satisfying the axioms
>
> - distributive over vectors: $\lambda(v_1 + v_2) = \lambda(v_1 + v_2)$,
> - distributive over scalars: $(\lambda_1 + \lambda_2)v = \lambda_1 v + \lambda_2 v$,
> - $\lambda(\mu v) = \lambda\mu v$,
> - $1 \cdot v = v$.

The additive unit of V is denoted by $\mathbf{0}$.

Example.

1. For all $n \in \mathbb{N}, \mathbb{F}^n$ is the space of column vectors of length n with entries in $\mathbb{F}$. It is an vector space by entry-wise addition and entry-wise scalar multiplication.
2. $\mathcal{M}_{m,n}(\mathbb{F})$, the set of $m \times n$ matrices with entries in $\mathbb{F}$, with the operation defined as entry-wise addition.
3. For any set X, $\mathbb{R}^X = \{f : X \to \mathbb{R}\}$, the set of $\mathbb{R}$-valued functions on X, with addition and scalar multiplication defined pointwise. For instance, $(f_1 + f_2)(x) = f_1(x) + f_2(x)$.

Exercise.

1. Check the above examples satisfy the axioms.
2. $0 \cdot v = \mathbf{0}$ and $(-1) \cdot v = -v$ for all $v \in V$.

1.2 Vector Subspace

> **Definition** (Vector subspace)**.** Let V be an $\mathbb{F}$-vector space. A subset $U \subseteq V$ is a *subspace*, denoted $U \leq V$, if
>
> - $\mathbf{0} \in U$,
> - U is closed under addition: $\forall u_1, u_2 \in U, u_1 + u_2 \in U$,
> - U is closed under scalar multiplication: $\forall u \in U, \forall \lambda \in \mathbb{F}, \lambda u \in U$.

Exercise. If U is a subspace of V, then U is also an $\mathbb{F}$-vector space.

Example.

1. $V = \mathbb{R}^{\mathbb{R}}$, the set all functions from $\mathbb{R}$ to itself, has a (proper) subspace $C(\mathbb{R})$, the space of continuous functions on $\mathbb{R}$ as continuous functions are closed under addition and scalar multiplication. $C(\mathbb{R})$ in turn has a proper subspace $P(\mathbb{R})$, the set of all polynomials in $\mathbb{R}$.
2. $\{(x_1, x_2, x_3) \in \mathbb{R}^3 : x_1 + x_2 + x_3 = t\}$ where t is some fixed constant is a subspace of $\mathbb{R}^3$ if and only if $t = 0$.

> **Proposition 1.1.** *Let V be an $\mathbb{F}$-vector space, $U, W \leq V$. Then $U \cap W \leq V$.*

Proof.

- $\mathbf{0} \in U, \mathbf{0} \in V$ so $\mathbf{0} \in U \cap W$.
- Suppose $u, w \in U \cap W$. Fix $\lambda, \mu \in \mathbb{F}$. As $U \leq V$, $\lambda u + \mu w \in U$. As $W \leq V$, $\lambda u + \mu w \in W$ so $\lambda u + \mu w \in U \cap W$. Take $\lambda = \mu = 1$ for vector addition and $\mu = 0$ for scalar multiplication.

□

Example. $V = \mathbb{R}^3, U = \{(x, y, z) : x = 0\}, W = \{(x, y, z) : y = 0\}$, then $U \cap W = \{(x, y, z) : x = y = 0\}$.

Note. The union of a family of subspaces is *almost never* a subspace. For example, $V = \mathbb{R}^2$, U, V be x- and y-axis.

> **Definition** (Sum of vector spaces)**.** Let V be an $\mathbb{F}$-vector space, $U, W \leq V$, the *sum* of U and W is the set
>
> $$U + W = \{u + w : u \in U, w \in W\}$$

Example. Use the definition from the previous example, $U + W = V$.

> **Proposition 1.2.** $U + W \leq V$.

Proof.

- $\mathbf{0} = \mathbf{0} + \mathbf{0} \in U + W$,
- $u_1, u_1 \in U, w_1, w_2 \in W$, $(u_1 + w_2) + (u_2 + w_2) = (u_1 + u_2) + (w_1 + w_2) \in U + W$,
- similar for scalar multiplication. Left as an exercise.

□

Note. $U + W$ is the smallest subspace containing both U and W. This is because all elements of the form $u + w$ are in such a space by closure under addition.

> **Definition** (Quotient vector space)**.** Let V be an $\mathbb{F}$-vector space, $U \leq V$. The *quotient space* V/U is the abelian gropup V/U equipped with scalar multiplication
>
> $$\begin{aligned} \mathbb{F} \times V/U &\to V/U \\ (\lambda, v + U) &\mapsto \lambda v + U \end{aligned}$$

> **Proposition 1.3.** *This is well-defined and V/U is an $\mathbb{F}$-vector space.*

Proof. First check it is well-defined. Suppose $v_1 + U = v_2 + U \in V/U$. Then $v_1 - v_2 \in U$. Now use closure under scalar multiplication and distributivity, $\lambda v_1 - \lambda v_2 = \lambda(v_1 - v_2) \in U$ so $\lambda v_1 + U = \lambda v_2 + U \in V/U$. Now check vector space axioms of V/U, which will follow from the axioms for V:

- $\lambda(\mu(v + U)) = \lambda(\mu v + U) = \lambda(\mu v) + U = (\lambda\mu)v + U = \lambda\mu(v + U)$,
- other axioms are left as an exercise.

□

1.3 Span, Linear Independence & Basis

> **Definition** (Span)**.** Let V be a $\mathbb{F}$-vector space, $S \subseteq V$ be a subset. The *span* of S
>
> $$\langle S \rangle = \Big\{ \sum_{s \in S} \lambda_s s : \lambda_s \in \mathbb{F} \Big\}$$
>
> is the set of all the finite linear combinations of elements (i.e. all but finitely many of the λ are zero) of S.

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

Convention. $\langle \emptyset \rangle = \{\mathbf{0}\}$

Example.

1. $V = \mathbb{R}^3$, $S = \{(1,0,0), (0,1,2), (3,-2,-4)\}$, $\langle S \rangle = \{(a, b, 2b) : a, b \in \mathbb{R}\}$
2. For any set X, $\mathbb{R}^X$ is a vector space. For $x \in X$, define $\delta_x : X \to \mathbb{R}, \delta_x(x) = 1, \delta_x(y) = 0 \, \forall y \neq x$, then

   $$\langle \delta_x : x \in X \rangle = \{ f \in \mathbb{R}^X : f \text{ has finite support} \}$$

> **Definition** (Span)**.** S spans V if $\langle S \rangle = V$.

> **Definition** (Finite-dimensional)**.** V is *finite-dimensional* over $\mathbb{F}$ if it is spanned by a finite set.

Definition (Linear independence)**.** The vectors $v_1, \dots, v_n$ are *linearly independent* over $\mathbb{F}$ if

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

A subset $S \subseteq V$ is *linearly independent* if every finite subset of S is linearly independent.

A subset is *linearly dependent* if it is not linearly independent.

Example. In the first example above, the three vectors are not linearly independent.

Exercise. The set $\{\delta_x : x \in X\}$ is linearly independent.

Definition (Basis)**.** S is a *basis* of V if it is linearly independent and spans V.

Example.

1. $\mathbb{F}^n$ has standard basis $\{e_1, e_2, \dots, e_n\}$ where e_i is the column vector with 1 in the ith entry and 0 elsewhere.
2. $V = \mathbb{C}$ over $\mathbb{C}$ has natural basis $\{1\}$, but over $\mathbb{R}$ it has natural basis $\{1, i\}$.
3. $V = P(\mathbb{R})$, the space of real polynomials, has natural basis

   $$\{1, x, x^2, \dots\}.$$

   It is an exercise to check this carefully.

Lemma 1.4. *Let V be a $\mathbb{F}$-vector space. The vectors $v_1, \dots, v_n$ form a basis of V if and only if each vector $v \in V$ has a unique expression*

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

Proof.

- $\Rightarrow$: Fix $v \in V$. The v_i span V, so exists $\lambda_i \in \mathbb{F}$ such that $v = \sum \lambda_i v_i$. Suppose also $v = \sum \mu_i v_i$ for some $\mu_i \in \mathbb{F}$. Then the difference

  $$\sum (\mu_i - \lambda_i) v_i = \mathbf{0}.$$

  Since the v_i are linearly independent, $\mu_i - \lambda_i = 0$ for all i.
- $\Leftarrow$: The v_i span V by assumption. Suppose $\sum_{i=1}^{n} \lambda_i v_i = \mathbf{0}$. Note that $\mathbf{0} = \sum_{i=0}^{n} 0 \cdot v_i$. By appying uniqueness to $\mathbf{0}$, $\lambda_i = 0$ for all i.

□

> **Lemma 1.5.** *If $v_1, \dots, v_n$ spans V over $\mathbb{F}$, then some subset of $v_1, \dots, v_n$ is a basis of V over $\mathbb{F}$.*

Proof. If $v_1, \dots, v_n$ is linearly independent then done. Otherwise for some ℓ, there exist $\alpha_1, \dots, \alpha_{\ell-1} \in \mathbb{F}$ such that

$$v_\ell = \sum_{i=1}^{\ell-1} \alpha_i v_i.$$

(If $\sum \lambda_i v_i = 0$, not all λ_i is zero. Take ℓ maximal with $\lambda_\ell \neq 0$, then $\alpha_i = -\frac{\lambda_i}{\lambda_\ell}$.)

Now $v_1, \dots, v_{\ell-1}, v_{\ell+1}, \dots, v_n$ still span V. Continue iteratively until we have linear independence. □

> **Theorem 1.6** (Steinitz Exchange Lemma). *Let V be a finite-dimensional vector space over $\mathbb{F}$. Take $v_1, \dots, v_m$ to be linearly independent, $w_1, \dots, w_n$ to span V. Then*
>
> - *$m \leq n$, and*
> - *reordering the w_i if needed, $v_1, \dots, v_m, w_{m+1}, \dots, w_n$ spans V.*

Proof. Proceed by induction. Suppose that we have replaced $\ell \geq 0$ of the w_i. Reordering w_i if needed, $v_1, \dots, v_\ell, w_{\ell+1}, \dots, w_n$ spans V.

- If $\ell = m$, done.
- If $\ell < m$, then $v_{\ell+1} = \sum_{i=1}^{\ell} \alpha_i v_i + \sum_{i>\ell} \beta_i w_i$. As the v_i are linearly independent, $\beta_i \neq 0$ for some i. After reordering, $\beta_{\ell+1} \neq 0$,

  $$w_{\ell+1} = \frac{1}{\beta_{\ell+1}}(v_{\ell+1} - \sum_{i \leq \ell} \alpha_i v_i - \sum_{i > \ell+1} \beta_i w_i).$$

  Thus $v_1, \dots, v_\ell, v_{\ell+1}, w_{\ell+2}, \dots, w_n$ also spans V. After m steps, we will replace m of the w_i by v_i. Thus $m \leq n$.

□

1.4 Dimension

> **Theorem 1.7.** *If V is a finite-dimensional vector space over $\mathbb{F}$, then any two bases for V have the same cardinality, which is called the* dimension *of V, donoted* $\dim_{\mathbb{F}} V$.

Proof. If $v_1, \dots, v_n$ and $w_1, \dots, w_m$ are both bases, then $\{v_i\}$ is linearly independent and $\{w_i\}$ spans V so $n \leq m$. Similarly $m \leq n$. □

Example. $\dim_{\mathbb{C}} \mathbb{C} = 1$, but $\dim_{\mathbb{R}} \mathbb{C} = 2$.

> **Lemma 1.8.** *Let V be a finite-dimensional $\mathbb{F}$-vector space. If $w_1, \dots, w_\ell$ is a linearly independent set of vectors, we can extend it to a basis $w_1, \dots, w_\ell, w_{\ell+1}, \dots, w_n$.*

Proof. Apply Steinitz exchange lemma to $w_1, \dots, w_\ell$ and any basis $v_1, \dots, v_n$.

Or more direcly, if $V = \langle w_1, \dots, w_\ell \rangle$, done. Otherwise take $v_{\ell+1} \in V \setminus \langle w_1, \dots, w_\ell \rangle$. Now $w_1, \dots, w_\ell, w_{\ell+1}$ is linearly independent. Iterate. □

Corollary 1.9. *Let V be a finite-dimensional vector space of dimension n. Then*

1. *Any linearly independent set of vectors has at most n elements, with equality if and only if the set is a basis.*
2. *Any spanning set of vectors has at least n elements, with equaility if and only if the set is a basis.*

Slogan. Choose the best basis for the job.

Theorem 1.10. *Let U, W be subspaces of V. If V and W are finite-dimensional, so is $U + W$ and*

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

Proof. Pick basis $v_1, \dots, v_\ell$ of $U \cap W$. Extend it to basis $v_1, \dots, v_\ell, u_1, \dots, u_m$ of U and $v_1, \dots, v_\ell, w_1, \dots, w_n$ of W. Claim $v_1, \dots, v_\ell, u_1, \dots, u_m, w_1, \dots, w_n$ is a basis for $U + W$:

- spanning: if $u \in U$, then $u = \sum \alpha_i v_i + \sum \beta_i u_i$ and if $w \in W$, $w = \sum \gamma_i v_i + \sum \delta_i w_i$, so $u + w = \sum (\alpha_i + \gamma_i) v_i + \sum \beta_i u_i + \sum \delta_i u_i$.
- linear independence: assume $\sum \alpha_i v_i + \sum \beta_i u_i + \sum \gamma_i w_i = 0$. Rearrange, $\sum \alpha_i v_i + \sum \beta_i u_i = -\sum \gamma_i w_i \in U \cap W$ so it equals to $\sum \delta_i v_i$ for some $\delta_i \in \mathbb{F}$ because v_i is a basis for $U \cap W$. As v_i and w_i are linearly independent, $\gamma_i = \delta_i = 0$ for all i. Thus $\sum \alpha_i v_i + \sum \beta_i v_i = 0$, so $\alpha_i = \beta_i = 0$ since v_i and u_i form a basis for U.

□

Theorem 1.11. *Let V be a finite-dimensional vector space over $\mathbb{F}$ and $U \leq V$, then U and V/U are also finite-dimensional and*

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

Proof. Left as an exercise. Outline: first show U is finite-dimensional, then let $u_1, \dots, u_\ell$ be a basis for U. Extend it to a basis for V, say $u_1, \dots, u_\ell, w_{\ell+1}, \dots, w_n$ of V. Check $w_{\ell+1} + U, \dots, w_n + U$ form a basis for V/U. □

Corollary 1.12. *If U is a proper subspace of V, which is finite-dimensional, then* $\dim U < \dim V$.

Proof. $V/U \neq 0$ so $\dim V/U > 0$. □

1.5 Direct Sum

Definition (Direct sum)**.** Let V be a vector space over $\mathbb{F}$, $U, W \leq V$. Then

$$V = U \oplus W$$

if every element of V can be written as $v = u + w$ for some unique $u \in U, w \in W$. This is called the *internal direct sum*. W is a *direct complement* of U in V.

Lemma 1.13. *Suppose $U, W \leq V$, TFAE:*

1. $V = U \oplus W$*,*
2. $V = U + W$ *and* $U \cap W = 0$*,*
3. *Given* $\mathcal{B}_1$ *any basis of* U*,* $\mathcal{B}_2$ *any basis of* V*,* $\mathcal{B}_1 \cap \mathcal{B}_2 = \emptyset$ *and* $\mathcal{B} = \mathcal{B}_1 \cup \mathcal{B}_2$ *is a basis of* V*.*

Proof.

- $2 \Rightarrow 1$: any $v \in V$ is $u + w$ for some $u \in U, w \in W$. Suppose $u_1 + w_1 = u_2 + w_2$, then $u_1 - u_2 = w_2 - w_1 \in U \cap W = 0$. Thus $u_1 = u_2, w_1 = w_2$.
- $1 \Rightarrow 3$: $\mathcal{B}$ spans as any $v \in V$ is $u + w$. Write u in terms of $\mathcal{B}_1$ and w in terms of $\mathcal{B}_2$. Then $u + w$ is a linear combination of elements of $\mathcal{B}$. To show $\mathcal{B}$ is linearly independent, suppose $\sum_{v \in \mathcal{B}} \lambda_v v = \mathbf{0} = \mathbf{0}_V + \mathbf{0}_W$. Write LHS as $\sum_{v \in \mathcal{B}_1} \lambda_v v + \sum_{v \in \mathcal{B}_2} \lambda_v v$. By uniqueness of expression, $\sum_{v \in \mathcal{B}_1} \lambda_v v = \mathbf{0}_V$ and $\sum_{w \in \mathcal{B}_2} \lambda_w w = \mathbf{0}_w$. As $\mathcal{B}_1, \mathcal{B}_2$ are bases, all of the λ_v, λ_w are zero.
- $3 \Rightarrow 2$: if $v \in V, v = \sum_{x \in V} \lambda_x x = \sum_{u \in \mathcal{B}_1} \lambda_u u + \sum_{w \in \mathcal{B}_1} \lambda_w w$ so $v \in U + W$. Conversely, if $v \in U \cap W, v = \sum_{u \in \mathcal{B}_1} \lambda_u u = \sum_{w \in \mathcal{B}_2} \lambda_w w$ so all λ_u, λ_v are zero since $\mathcal{B}_1 \cup \mathcal{B}_2$ is linearly independent.

□

Lemma 1.14. *Let V be a finite-dimensional vector space over $\mathbb{F}$ and $U \leq V$. Then there exists a direct complement to U in V.*

Proof. Let $u_1, \dots, u_\ell$ be a basis for U. Extend this to a basis $u_1, \dots, u_\ell, w_{\ell+1}, \dots, w_n$ for V. Then $\langle w_{\ell+1}, \dots, w_n \rangle$ is a direct complement of U. □

Caution. Direct complements are *not* unique.

Definition (Direct sum)**.** Suppose $V_1, \dots, V_\ell \leq V$, then the sum

$$\sum_i V_i = V_1 + \dots + V_\ell = \{v_1 + \dots + v_\ell : v_i \in V_i\}.$$

is *direct* if

$$v_1 + \dots + v_\ell = v_1' + \dots + v_\ell' \Rightarrow v_i = v_i' \text{ for all } i.$$

> In which case it is denoted
> $$V = \bigoplus_{i=1}^{\ell} V_i.$$

Exercise. $V_1, \dots, V_\ell \leq V$, TFAE:

1. The sum $\sum_i V_i$ is direct,
2. $V_i \cap \sum_{j \neq i} V_j = 0$ for all i,
3. For any basis B_i of V_i, the union $B = \bigcup_{i=1}^{\ell} B_i$ is a basis for $\sum_i V_i$.

> **Definition** (Direct sum)**.** Let U, W be vector spaces over $\mathbb{F}$. The *external direct sum* is
> $$U \oplus W = \{(u, w) : u \in U, w \in W\}$$
> with pointwise addition and scalar multiplication.

2 Linear Map

2.1 Definitions

> **Definition** (Linear map)**.** V, W two $\mathbb{F}$-vector space, a map $\alpha : V \to W$ is *linear* if
>
> - $\alpha(v_1 + v_2) = \alpha(v_1) + \alpha(v_2)$,
> - $\alpha(\lambda v) = \lambda\alpha(v)$.
>
> This is equivalent to
>
> $$\alpha(\lambda_1 v_1 + \lambda_2 v_2) = \lambda_1\alpha(v_1) + \lambda_2\alpha(v_2).$$

Example.

1. Given an $n \times m$ matrix A with coefficients in $\mathbb{F}$, the map $\alpha : \mathbb{F}^m \to \mathbb{F}^n, v \to Av$.
2. Differentiation $D : P(\mathbb{R}) \to P(\mathbb{R}), f \mapsto \frac{df}{dx}$.
3. Integration $I : C[0,1] \to C[0,1], f \mapsto I(f)$ where $I(f)(x) = \int_0^x f(t)dt$.
4. Fix $x \in [0,1]$, the map $C[0,1] \to \mathbb{R}, f \mapsto f(x)$.

Note (Categoricity of $\mathbf{Vect}_\mathbb{F}$)**.** Suppose U, V, W are $\mathbb{F}$-vector spaces, then

1. $\mathrm{id} : V \to V$ is linear.
2. Given $U \xrightarrow{\alpha} V \xrightarrow{\beta} W$, if α, β are linear then so is $\beta \circ \alpha$.

> **Lemma 2.1** (Free functor $\mathbf{Set} \to \mathbf{Vect}_\mathbb{F}$)**.** *Suppose V, W are $\mathbb{F}$-vector spaces and $\mathcal{B}$ is a basis for V. If $\alpha_0 : \mathcal{B} \to W$ is* any *map, then there is a* unique *linear map $\alpha : V \to W$ extending α_o.*

Proof. Let $v \in V$. Write $v = \sum \lambda_i v_i$ in a unique way. By linearity $\alpha(v) = \alpha(\sum \lambda_i v_i) = \sum \lambda_i \alpha(v_i) = \sum \lambda_i \alpha_0(v_i)$. Uniqueness follows. □

Note.

- This is true for infinite-dimensional vector spaces as well.
- Very often, to define a linear map, define it on a basis and extend it linearly to the vector space.
- Two linear maps $\alpha_1, \alpha_2 : V \to W$ are equal if and only if they agree on a basis.

2.2 Isomorphism of Vector Spaces

> **Definition** (Isomorphism)**.** Given V, W two $\mathbb{F}$-vector spaces, the map $\alpha : V \to W$ is an *isomorphism* if it is linear and bijective, denoted $V \cong W$.

Lemma 2.2. $\cong$ *is an equivalence relation on the class of all* $\mathbb{F}$*-vector spaces.*

Proof.

- symmetric: obvious.
- reflexive: blah blah in lecture. Left as an exercise to reader.
- transitive: obvious.

□

Theorem 2.3. *If* V *is an* $\mathbb{F}$*-vector space, then* $V \cong \mathbb{F}^n$ *for some* n*.*

Proof. Choose a basis for V, say $v_1, \dots, v_n$. Define a map

$$\begin{aligned} V &\to \mathbb{F}^n \\ \sum_i \lambda_i v_i &\mapsto (\lambda_1, \dots, \lambda_n) \end{aligned}$$

which is an isomorphism. □

Remark. Choosing an isomorphism $V \cong \mathbb{F}^n$ is equivalent to choosing a basis for V, i.e. there is a bijection

$$\{\alpha \in \operatorname{Hom}(V, \mathbb{F}^n), \alpha \text{ bijective}\} \leftrightarrow \{\text{bases of } V\}.$$

Theorem 2.4. *Given two finite-dimensional* $\mathbb{F}$*-vector spaces* V, W*, they are isomorphic if and only if they have the same dimension.*

Proof.

- $\Leftarrow$: $V \cong \mathbb{F}^{\dim V} = \mathbb{F}^{\dim W} \cong W$.
- $\Rightarrow$: let $a : V \to W$ be an isomorphism and $\mathcal{B}$ be a basis for V. Claim $\alpha(\mathcal{B})$ is a basis for W: $\alpha(\mathcal{B})$ spans W due to surjectivity and $\alpha(\mathcal{B})$ is linearly independent due to injectivity.

□

Definition (Kernel & Image)**.** Given $\alpha : V \to W$,

- $N(\alpha) = \ker \alpha = \{v \in V : \alpha(v) = 0\} \leq V$,
- $\operatorname{im} \alpha = \{w \in W : \exists v \in V, \alpha(v) = w\} \leq W$.

Proposition 2.5.

- α *is injective if and only if* $N(\alpha) = 0$*,*
- α *is surjective if and only if* $\operatorname{im} \alpha = W$*.*

Proof. Easy. □

Example. Let $\alpha : C^\infty(\mathbb{R}) \to C^\infty(\mathbb{R}), \alpha(f)(t) = f''(t) + 2f'(t) + 5f(t)$. $\ker\alpha = \{f : f'' + 2f' + 5f = 0\}$ and $g \in \operatorname{im}\alpha$ if and only if there exists an f such that $f'' + 2f' + 5f = g$.

> **Theorem 2.6** (First Isomorphism Theorem)**.** *Let $\alpha : V \to W$ be a linear map. It induces an isomprhism*
>
> $$\begin{aligned}\bar\alpha : V/\ker\alpha &\to \operatorname{im}\alpha\\ v + \ker\alpha &\mapsto \alpha(v)\end{aligned}$$

Proof. Check the following:

- $\bar\alpha$ is well-defined,
- $\bar\alpha$ is linear: immediate from linearity of α,
- $\bar\alpha$ is surjective.

□

> **Definition** (Rank & Nullity)**.**
>
> - $r(\alpha) = rk(\alpha) = \dim(\operatorname{im}\alpha)$ is the *rank* of α,
> - $n(\alpha) = \dim N(\alpha)$ is the *nullity* of α.

> **Theorem 2.7** (Rank-nullity)**.** *Let U, V be $\mathbb{F}$-vector spaces,* $\dim U < \infty$. *Let $\alpha : U \to V$ be a linear map. Then*
>
> $$\dim U = r(\alpha) + n(\alpha).$$

Proof. $U/\ker\alpha \cong \operatorname{im}\alpha$ so $\dim U - \dim(\ker\alpha) = \dim(\operatorname{im}\alpha)$. Rearrange. □

> **Lemma 2.8.** *Let V, W be $\mathbb{F}$-vector spaces with equal, finite dimension. Let $\alpha : V \to W$ be linear, then TFAE:*
>
> 1. *α is injective,*
> 2. *α is surjective,*
> 3. *α is an isomorphism.*

Proof. Rank-nullity theorem. □

2.3 Linear Maps as Vector Space

Suppose V and W are $\mathbb{F}$-vector spaces. Let $L(V, W) = \{\alpha : V \to W, \alpha \text{ linear}\}$.

Proposition 2.9. *$L(V, W)$ is an $\mathbb{F}$-vector space, under operations*

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

Proof. $\alpha_1 + \alpha_2, \lambda\alpha$ as above are well-defined linear maps. The vector space axioms can be easily checked. □

Proposition 2.10. *If both V and W are finite-dimensional over $\mathbb{F}$ then so is $L(V, W)$ and $L(V, W) = \dim V \cdot \dim W$.*

Proof. See Lemma 2.15. □

2.3.1 Matrices, an Interlude

Definition (Matrix)**.** An $m \times n$ *matrix* over $\mathbb{F}$ is an array with m rows and n columns with entries in $\mathbb{F}$. We write

$$A = (a_{ij}), a_{ij} \in \mathbb{F}, 1 \leq i \leq m, 1 \leq j \leq n.$$

Definition. $\mathcal{M}_{m,n}(\mathbb{F})$ is the set of all such $m \times n$ matrices.

Proposition 2.11. *$\mathcal{M}_{m,n}(\mathbb{F})$ is an $\mathbb{F}$-vector space and $\dim \mathcal{M}_{m,n}(\mathbb{F}) = m \cdot n$.*

Proof. See the example on page 3 for the proof of vector space axioms. For the dimensional claim, a standard basis for $\mathcal{M}_{m,n}(F)$ is

$$E_{ij} = \begin{pmatrix} 0 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & 1 & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 0 \end{pmatrix}$$

with 1 in the (i, j)th entry so $a_{ij} = \sum_{i,j} a_{ij} E_{ij}$, from which span and linear independence follow. The basis has cardinality $m \cdot n$. □

2.3.2 Representation of Linear Maps by Matrices

Let V and W be finite-dimensional $\mathbb{F}$-vector space, $\alpha : V \to W$ linear. Let $\mathcal{B} = \{v_1, \dots, v_n\}$ be a basis for V, $\mathcal{C} = \{w_1, \dots, w_m\}$ be a basis for W. If $v = \sum_i \lambda_i v_i \in V$, write

$$[v]_{\mathcal{B}} = \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix} \in \mathbb{F}^n$$

which is called the *coordinate vector of v with respect to $\mathcal{B}$*. Similarly $[w]_{\mathcal{C}} \in \mathbb{F}^m$.

Definition (Matrix representation)**.** $[\alpha]_{\mathcal{B},\mathcal{C}}$ is the matrix representation of α with respect to $\mathcal{B}$ and $\mathcal{C}$ with

$$\begin{aligned}[\alpha]_{\mathcal{B},\mathcal{C}} &= \left([\alpha(v_1)]_{\mathcal{C}} \,\middle|\, [\alpha(v_2)]_{\mathcal{C}} \,\middle|\, \cdots \,\middle|\, [\alpha(v_n)]_{\mathcal{C}}\right) \\ &= (a_{ij})\end{aligned}$$

The matrix says

$$\alpha(v_j) = \sum_i a_{ij} w_i.$$

Lemma 2.12. *For any $v \in V$,*

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

where $\cdot$ is matrix multiplication.

Proof. Fix $v = \sum_{j=1}^n \lambda_j v_j \in V$, so

$$[v]_{\mathcal{B}} = \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix}$$

$$\begin{aligned}\alpha(v) &= \alpha\left(\sum_j \lambda_j v_j\right) \\ &= \sum_j \lambda_j \alpha(v_j) \\ &= \sum_j \lambda_j \left(\sum_i \alpha_{ij} w_i\right) \\ &= \sum_i \left(\sum_j a_{ij} \lambda_j\right) w_i\end{aligned}$$

so the ith entry of $\alpha(v)$ is the ith entry of $[\alpha]_{\mathcal{B},\mathcal{C}} \cdot [v]_{\mathcal{B}}$. □

Lemma 2.13. *Suppose $U \xrightarrow{\beta} V \xrightarrow{\alpha} W$ with α, β linear, with $\alpha \circ \beta : U \to W$. Let $\mathcal{A}, \mathcal{B}, \mathcal{C}$ be bases for U, V, W respectively. Then*

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

Proof.

$$\begin{aligned}(\alpha \circ \beta)(u_\ell) &= \alpha(\beta(u_\ell)),\ u_\ell \in A \\ &= \alpha\Big(\sum_j b_{jl} v_j\Big),\ v_j \in B \\ &= \sum_j b_{jl}\alpha(v_j) \\ &= \sum_j b_{jl}\sum_i a_{ij}w_i,\ w_i \in W \\ &= \sum_i \left(\sum_j a_{ij}b_{jl}\right) w_i\end{aligned}$$

□

Proposition 2.14. *Let* V *and* W *be* $\mathbb{F}$*-vector spaces with* $\dim V = n, \dim W = m$*, then*

$$L(V,W) \cong \mathcal{M}_{m,n}(\mathbb{F}).$$

Proof. Fix bases $B = \{v_1 \dots, v_n\}, C = \{w_1, \dots, w_m\}$ for V and W respectively. Claim

$$\begin{aligned}\theta : L(V,W) &\to \mathcal{M}_{m,n}(\mathbb{F}) \\ \alpha &\mapsto [\alpha]_{\mathcal{B},\mathcal{C}}\end{aligned}$$

is an isomorphism:

- linearity: $[\lambda_1\alpha_1 + \lambda_2\alpha_2]_{\mathcal{B},\mathcal{C}} = \lambda_1[\alpha_1]_{\mathcal{B},\mathcal{C}} + \lambda_2[\alpha_2]_{\mathcal{B},\mathcal{C}}$.
- surjectivity: given $A = (a_{ij})$, let $\alpha : v_j \mapsto \sum_{i=1}^m a_{ij}w_i$ and extend linearly. It follows that $\alpha \in L(V,W)$ and $\theta(\alpha) = A$.
- injectivity: $[\alpha]_{\mathcal{B},\mathcal{C}} = \mathbf{0}$ implies that α is the zero map.

□

Corollary 2.15.

$$\dim L(V,W) = \dim V \cdot \dim W.$$

Example. Suppose $\alpha : V \to W$, $Y \leq V, Z \leq W$ with $\alpha(Y) \leq Z$. Let $\mathcal{B}' = \{v_1, \dots, v_k\}$ be a basis of Y and extend to $\mathcal{B} = \{v_1, \dots, v_k, v_{k+1}, \dots, v_n\}$ a basis for V. Similarly $\mathcal{C}' = \{w_1, \dots, w_l\}$ and $\mathcal{C}$ for Z and W.

- $[\alpha]_{\mathcal{B},\mathcal{C}} = \begin{pmatrix} A & B \\ 0 & C \end{pmatrix}$ for some A, B, C because for $1 \leq j \leq k$, $\alpha(v_j)$ is a linear combination of w_i where $1 \leq i \leq l$.
- $[\alpha|_Y]_{B',C'} = A$.

- α induces a map

$$\begin{aligned}\bar{\alpha} : V/Y &\to W/Z \\ v + Y &\mapsto \alpha(v) + Z\end{aligned}$$

This is well-defined. Linearity follows from that of α. A basis for V/Y is $\mathcal{B}'' = \{v_{k+1} + Y, \dots, v_n + Y\}$ and similarly for W/Z. It is an exercise to show $[\bar{\alpha}]_{\mathcal{B}'',\mathcal{C}''} = C$.

2.3.3 Change of Bases

Throughout this section, let V and W be $\mathbb{F}$-vector spaces and suppose they have the following bases:

Vector space	V	W
Basis 1	$\mathcal{B} = \{v_1, \dots, v_n\}$	$\mathcal{C} = \{w_1, \dots, w_m\}$
Basis 2	$\mathcal{B}' = \{v'_1, \dots, v'_n\}$	$\mathcal{C}' = \{w'_1, \dots, w'_m\}$

Definition (Change-of-basis matrix)**.** The *change-of-basis matrix* from $\mathcal{B}'$ to $\mathcal{B}$ is $P = (p_{ij})$ given by

$$\begin{aligned}v'_j &= \sum_i p_{ij} v_i \\ P &= \left([v'_1]_{\mathcal{B}} \,\middle|\, [v'_2]_{\mathcal{B}} \,\middle|\, \dots \,\middle|\, [v'_n]_{\mathcal{B}} \right) = [\mathrm{id}]_{\mathcal{B}',\mathcal{B}}\end{aligned}$$

Lemma 2.16.

$$[v]_{\mathcal{B}} = P[v]_{\mathcal{B}'}.$$

Proof.

$$P[v]_{\mathcal{B}'} = [\mathrm{id}]_{\mathcal{B}',\mathcal{B}}[v]_{\mathcal{B}'} = [v]_{\mathcal{B}}.$$

□

Lemma 2.17. *P is an invertible $n \times n$ matrix and P^{-1} is the change-of-basis matrix from $\mathcal{B}$ to $\mathcal{B}'$.*

Proof.

$$\begin{aligned}[\mathrm{id}]_{\mathcal{B},\mathcal{B}'}[\mathrm{id}]_{\mathcal{B}',\mathcal{B}} &= [\mathrm{id}]_{\mathcal{B}',\mathcal{B}'} = I_n \\ [\mathrm{id}]_{\mathcal{B}',\mathcal{B}}[\mathrm{id}]_{\mathcal{B},\mathcal{B}'} &= [\mathrm{id}]_{\mathcal{B},\mathcal{B}} = I_n\end{aligned}$$

□

Let Q be the change-of-basis matrix from $\mathcal{C}'$ to $\mathcal{C}$. Then Q is an invertible $m \times m$ matrix.

Proposition 2.18. *Let $\alpha : V \to W$ be a linear map, $A = [\alpha]_{\mathcal{B},\mathcal{C}}$, $A' = [\alpha]_{\mathcal{B}',\mathcal{C}'}$, then*

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

Proof.

$$\underbrace{[\mathrm{id}]_{\mathcal{C},\mathcal{C}'}}_{Q^{-1}}[\alpha]_{\mathcal{B},\mathcal{C}}\underbrace{[\mathrm{id}]_{\mathcal{B}',\mathcal{B}}}_{P} = \underbrace{[\mathrm{id}\circ\alpha\circ\mathrm{id}]_{\mathcal{B}',\mathcal{C}'}}_{A'}$$

□

> **Definition** (Equivalence of matrices)**.** $A, A' \in \mathcal{M}_{m,n}(\mathbb{F})$ are *equivalent* if
>
> $$A' = Q^{-1}AP$$
>
> for some invertible $P \in \mathcal{M}_{n,n}(\mathbb{F})$ and $Q \in \mathcal{M}_{m,m}(\mathbb{F})$.

Note. This defines an equivalence relation on $\mathcal{M}_{m,n}(\mathbb{F})$.

> **Proposition 2.19.** *Let V, W be $\mathbb{F}$-vector spaces of dimension n and m respectively. Let $\alpha : V \to W$ be a linear map. Then there exist bases $\mathcal{B}$ of V, $\mathcal{C}$ of W, and some $r \leq m, n$ such that*
>
> $$[\alpha]_{\mathcal{B},\mathcal{C}} = \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}$$
>
> *where I_r the is $r \times r$ the identity matrix.*

Note. $r = rk(\alpha) = r(\alpha)$.

Proof. Fix r such that $\dim N(\alpha) = n - r$. Fix a basis for $N(\alpha)$, say $v_{r+1}, \dots, v_n$. Extend this to a basis $\mathcal{B}$ for V, say $v_1, \dots, v_r, v_{r+1}, \dots, v_n$. Now $\alpha(v_1), \dots, \alpha(v_r)$ is a basis for $\mathrm{im}(\alpha)$:

- span: $\alpha(v_1), \dots, \alpha(v_n)$ certainly span $\mathrm{im}(\alpha)$. Since $v_{r+1}, \dots, v_n \in \ker\alpha$, $\alpha(v_{r+1}), \dots, \alpha(v_n) = 0$ so we can remove them from the spanning set.
- linear independence: assume $\sum_{i=1}^n \lambda_i\alpha(v_i) = \mathbf{0}$. Then $\alpha\big(\sum_{i=1}^n \lambda_i v_i\big) = \mathbf{0}$. This implies that

  $$\sum_{i=1}^n \lambda_i v_i = \sum_{j=r+1}^n \mu_j v_j.$$

  As $v_1, \dots v_n$ are linearly independent, $\lambda_i = \mu_j = 0$ for all i, j.

Extend $\alpha(v_1), \dots, \alpha(v_r)$ to a basis for W, say $\mathcal{C}$. By construction,

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

□

Remark. In the proof above we didn't need to assume that $r = r(\alpha)$. This gives us another way prove Rank-nullity Theorem.

Corollary 2.20. *Any $m \times n$ matrix is equivalent to*

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

for some r.

Definition (Row and column rank)**.** Let $A \in \mathcal{M}_{m,n}(\mathbb{F})$.

- The *column rank* of A, $r(A)$ is the dimension of the subspace of $\mathbb{F}^m$ spanned by the columns of A.
- The *row rank* of A is the column rank of A^T.

Note. If α is a linear map represented by A with respect to any choice of bases, then $r(\alpha) = r(A)$.

Proposition 2.21. *Two $m \times n$ matrices A, A' are equivalent if and only if*

$$r(A) = r(A').$$

Proof.

- $\Leftarrow$: Both A and A' are equivalent to $\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}$ and matrix equivalence is transitive.
- $\Rightarrow$: Let $\alpha : \mathbb{F}^n \to \mathbb{F}^m$ be the linear map represented by A with respect to, say, the standard basis. Since $A' = Q^{-1}AP$ for some invertible P and Q, A' represents the same α with respect to another bases. $r(\alpha)$ is defined in a basis-invariant way so $r(A) = r(\alpha) = r(A')$.

□

Theorem 2.22.

$$r(A) = r(A^T).$$

Proof. $Q^{-1}AP = \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}_{m,n}$ where P and Q are invertible. Take transpose of the whole equation:

$$\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}_{n,m} = (Q^{-1}AP)^T = P^T A^T (Q^T)^{-1}$$

so A^T is equivalent to

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

□

Note a special case for change of basis: $V = W$, $\mathcal{C} = \mathcal{B}$ and $\mathcal{C}' = \mathcal{B}'$. P, the change-of-basis matrix from $\mathcal{B}'$ to $\mathcal{B}$, is given the map $\alpha \in L(V, V)$

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

Definition (Similar matrices)**.** Given $A, A' \in \mathcal{M}_{n,n}(\mathbb{F})$, A and A' are *similar*, or *conjugate* if

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

for some invertible P.

2.3.4 Elementary Matrices and Operations

Definition (Elementary column operation)**.** *Elementary column operation* on a $m \times n$ matrix A is one of the following operations:

1. swap column i and j (wlog $i \neq j$),
2. scale column i by λ ($\lambda \neq 0$),
3. add λ times column i to column j ($i \neq j, \lambda \neq 0$).

Definition (Elementary row operation)**.** Defined analoguously, replacing “column” by “row”.

Note. All of these operations are invertible.

Definition (Elementary matrix)**.** The elementary column (row, respectively) operations have corresponding elementary matrices, which are the results of performing these column (row, respectively) operations on I_n (I_m, respectively):

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

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

3. $I_n + \lambda E_{ij}$ where E_{ij} is the matrix with 1 on ijth entry and 0 elsewhere.

An elementary column (row, respectively) operation on $A \in \mathcal{M}_{m,n}(\mathbb{F})$ can be performed by multiplying A by these corresponding elementary matrices on the right (left respectively).

Example.

$$\begin{pmatrix}1 & 2\\ 3 & 4\end{pmatrix}\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix} = \begin{pmatrix}2 & 1\\ 4 & 3\end{pmatrix}$$

Given the elementary matrices, we can give a constructive proof that any $m \times n$ matrix is equivalent to $\begin{pmatrix}I_r & 0\\ 0 & 0\end{pmatrix}$ for some r:

Constructive proof of Theorem 2.22. Start with A. If all entries of A are zero then done. If not then some $a_{ij} = \lambda \neq 0$. Perform the following:

1. swap row 1 and i, swap column 1 and j so λ is in position $(1, 1)$,
2. multiply column 1 by $1/\lambda$ to get 1 in position $(1, 1)$,
3. add $(-a_{12})$ times column 1 to column 2. Do so for the other entries in row 1. Also use row operations to clear out all other entries in column 1. Now the matrix is in the form
$$\begin{pmatrix}1 & 0\\ 0 & A'\end{pmatrix}$$
4. iterate for A'. Stop when the new $A' = 0$.

The result of these operations is

$$\begin{pmatrix}I_r & 0\\ 0 & 0\end{pmatrix} = \underbrace{E'_\ell E'_{\ell-1} \dots E'_1}_{Q^{-1}} A \underbrace{E_1 E_2 \dots E_{\ell-1} E_\ell}_{P}.$$

As elementary operations are invertible, the elementary matrices are invertible so

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

□

If you only use elementary row operations, we can get the *row echelon form* of a matrix:

$$\begin{pmatrix}a & b & \dots & c\\ 0 & d & \dots & e\\ \vdots & & \ddots & \vdots\\ 0 & 0 & \dots & f\end{pmatrix}$$

Lemma 2.23. *If A is an $n \times n$ invertible matrix then we can obtain I_n by using only elementary row/column operations.*

Proof. We prove the column operation case. Use induction on n, the number of rows. Suppose we have got $\begin{pmatrix}I_k & 0\\ \star & *\end{pmatrix}$ for some $k \geq 0$. There exists $j > k$ such that $a_{k+1,j} \neq 0$, (i.e. in the $*$ block) as otherwise $(0, \dots, 1, \dots, 0)$ with 1 in $(k+1)$th position would not be in the span of the column vectors, contradicting the invertiblity. Next we carry out the following operations:

1. swap column $k+1$ and j,
2. divide column $k+1$ by $a_{k+1,k+1}$ so have 1 in $(k+1,k+1)$ position,
3. use column operation to clear other entries of $(k+1)$th row.

Proceed inductively. □

Note that the equality

$$AE_1E_2\dots E_c = I_n$$

gives

$$A^{-1} = E_1E_2\dots E_c,$$

which is one way to compute inverses.

> **Proposition 2.24.** *Any invertible matrix can be written as a product of elementary ones.*

3 Dual Space & Dual Map

3.1 Definitions

> **Definition** (Dual space)**.** Let V be an $\mathbb{F}$-vector space. The *dual space* of V is defined to be
>
> $$V^* = L(V, \mathbb{F}) = \{\alpha : V \to \mathbb{F}, \alpha \text{ linear}\}.$$

V^* is itself an $\mathbb{F}$-vector space. Its elements are sometimes called *linear functionals.*

Example.

1. $\mathbb{R}^3 \to \mathbb{R}, (a, b, c) \mapsto a - c$ is an element of V^*.
2. $\operatorname{tr} : \mathcal{M}_{n,n}(\mathbb{F}) \to \mathbb{F}, A \mapsto \sum_i A_{ii}$ is an element of $\mathcal{M}_{n,n}(\mathbb{F})^*$.

> **Lemma 3.1** (Dual basis)**.** *Let V be a finite-dimensional $\mathbb{F}$-vector space with basis $\mathcal{B} = \{e_1, \dots, e_n\}$. Then there is a basis for V^*, given by*
>
> $$\mathcal{B}^* = \{\varepsilon_1, \dots, \varepsilon_n\}$$
>
> *where*
>
> $$\varepsilon_j\Big(\sum_{i=1}^n a_i e_i\Big) = a_j$$
>
> *for $1 \le j \le m$.*
>
> *$\mathcal{B}^*$ is called the* dual basis *to $\mathcal{B}$.*

Proof.

- linear independence: suppose

  $$\sum_{j=1}^n \lambda_j \varepsilon_j = 0.$$

  Apply the relation to basis vectors,

  $$0 = \Big(\sum_{j=1}^n \lambda_j \varepsilon_j\Big) e_i = \sum_{j=1}^n \lambda_j \varepsilon_j(e_i)$$

  The last expression is

  $$\varepsilon_j(e_i) = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j \end{cases}$$

  so $\lambda_i = 0$ for all $1 \le i \le n$.

- span: if $\alpha \in V^*$, then

  $$\alpha = \sum_{i=1}^n \alpha(e_i)\varepsilon_i$$

  since linear maps are uniquely determined by the action on basis.

□

Corollary 3.2. *If V is a finite-dimensional $\mathbb{F}$-vector space then*

$$\dim V = \dim V^*.$$

Remark. Sometimes it is useful to think about $(\mathbb{F}^n)^*$ as the space of row vectors of length n over $\mathbb{F}$.

3.2 Dual Map

It turns out dual spaces have maps between them. Before studying them in detail, we introduce this concept to add richness to the theory of dual map:

Definition (Annihilator)**.** If $U \subseteq V$, the *annihilator* of U is

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

Lemma 3.3.

1. *$U^\circ \leq V^*$,*
2. *If $U \leq V$ and $\dim V = n < \infty$ then*

$$\dim V = \dim U + \dim U^\circ.$$

Proof.

1. $0 \in U^\circ$. If α and α' are in U° then

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

for all $u \in U$. Similarly $\lambda\alpha \in U^\circ$ for any $\lambda \in \mathbb{F}$.

2. Let $\mathcal{B} = \{e_1, \dots, e_k\}$ be a basis for U and extend it to a basis for V, say $e_1, \dots, e_k, e_{k+1}, \dots, e_n$. Let $\mathcal{B}^* = \{\varepsilon_1, \dots, \varepsilon_n\}$ be its dual basis. Claim $\varepsilon_{k+1}, \dots, \varepsilon_n$ is a basis for U°:
   - If $i > k, j \leq k$ then $\varepsilon_i(e_j) = 0$ so $\varepsilon_i \in U^\circ$.
   - Linear independence comes from the fact that B^* is a basis.
   - If $\alpha \in U^\circ$, $\alpha = \sum_{i=1}^n a_i\varepsilon_i$ for some $\alpha_i \in \mathbb{F}$. Then for any $j \leq k$,

$$\Big(\sum_{i=1}^n a_i\varepsilon_i\Big)(e_j) = 0$$

so $a_j = 0$. It follows that $\alpha \in \langle \varepsilon_{k+1}, \dots, \varepsilon_n \rangle$.

□

Lemma 3.4 (Dual space as a contravariant functor)**.** *Let V and W be $\mathbb{F}$-vector spaces. Let $\alpha \in L(V, W)$. Then the map*

$$\begin{aligned} \alpha^* : W^* &\to V^* \\ \varepsilon &\mapsto \varepsilon \circ \alpha \end{aligned}$$

> *is linear. α^* is called the* dual *of α.*

Proof.

- $\varepsilon \circ \alpha \in V^*$ since composition preserves linearity.
- Fix $\theta_1, \theta_2 \in W^*$,

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

- Similarly $\alpha^*(\lambda\theta) = \lambda\alpha^*(\theta)$.

□

> **Proposition 3.5.** *Let V and W be $\mathbb{F}$-vector spaces with bases $\mathcal{B}$ and $\mathcal{C}$ respectively. Let $\mathcal{B}^*$ and $\mathcal{C}^*$ be the dual bases. Consider $\alpha \in L(V, W)$ with dual α^*, then*
> $$[\alpha^*]_{\mathcal{C}^*,\mathcal{B}^*} = [\alpha]^T_{\mathcal{B},\mathcal{C}}.$$

Proof. Say $\mathcal{B} = \{b_1, \dots, b_n\}$, $\mathcal{B}^* = \{\beta_1, \dots, \beta_n\}$, $\mathcal{C} = \{c_1, \dots, c_m\}$ and $\mathcal{C}^* = \{\gamma_1, \dots, \gamma_n\}$. Further let $[\alpha]_{\mathcal{B},\mathcal{C}} = (a_{ij})$, an $m \times n$ matrix.

$$\begin{aligned}\alpha^*(\gamma_r)(b_s) &= \gamma_r \circ \alpha(b_s) \\ &= \gamma_r(\alpha(b_s)) \\ &= \gamma_r\Big(\sum_t a_{ts}c_t\Big) \\ &= \sum_t a_{ts}\gamma_r(c_t) \\ &= a_{rs} \\ &= \Big(\sum_i a_{ri}\beta_i\Big)(b_s)\end{aligned}$$

Thus

$$\alpha^*(\gamma_r) = \sum_i a_{ri}\beta_i$$

so

$$[\alpha^*]_{\mathcal{C}^*,\mathcal{B}^*} = [\alpha]^T_{\mathcal{B},\mathcal{C}}.$$

□

It follows that

> **Lemma 3.6.** *Let V be a finite-dimensional $\mathbb{F}$-vector space with bases $\mathcal{E}$ and $\mathcal{F}$. They have correponding dual bases $\mathcal{E}^*$ and $\mathcal{F}^*$. If the change-of-basis matrix from $\mathcal{F}$ to $\mathcal{E}$ is P then the change-of-basis matrix from $\mathcal{F}^*$ to $\mathcal{E}^*$ is*
> $$(P^{-1})^T.$$

Proof.

$$[\mathrm{id}]_{\mathcal{F}^*,\mathcal{E}^*} = [\mathrm{id}]_{\mathcal{E},\mathcal{F}}^T = ([\mathrm{id}]_{\mathcal{F},\mathcal{E}}^{-1})^T.$$

□

Caution. $V \cong V^*$ only if V is finite-dimensional. Let $V = P$, the space of all real polynomials. It has basis $\{p_j\}_{j\in\mathbb{N}}$ where $p(j) = t^j$. In example sheet 2 we will see

$$\begin{aligned} P^* &\cong \mathbb{R}^{\mathbb{N}} \\ \varepsilon &\mapsto (\varepsilon(p_0), \varepsilon(p_1), \dots) \end{aligned}$$

and on example sheet 1 we prove

$$P \ncong \mathbb{R}^{\mathbb{N}}$$

as the latter does not have a countable basis.

Now we move on to more discussion about annhilator.

> **Lemma 3.7.** *Let V and W be $\mathbb{F}$-vector spaces. Fix $\alpha \in L(V,W)$ and let $\alpha^* \in L(W^*,V^*)$ be its dual. Then*
>
> - $N(\alpha^*) = (\operatorname{im}\alpha)^\circ$, *so α^* is injectve if and only if α is surjective.*
> - $\operatorname{im}\alpha^* \leq N(\alpha)^\circ$, *with equality if V and W are both finite-dimensional, in which case α^* is surjective if and only if α is injective.*

Proof.

- Let $\varepsilon \in W^*$, then

$$\begin{aligned} &\varepsilon \in N(\alpha^*) \\ \Leftrightarrow &\alpha^*(\varepsilon) = 0 \\ \Leftrightarrow &\varepsilon \circ \alpha = 0 \\ \Leftrightarrow &\varepsilon(u) = 0\, \forall u \in \operatorname{im}\alpha \\ \Leftrightarrow &\varepsilon \in (\operatorname{im}\alpha)^\circ \end{aligned}$$

- Let $\varepsilon \in \operatorname{im}\alpha^*$, Then $\varepsilon = \alpha^*(\phi)$ for some $\phi \in W^*$. For any $u \in N(a)$,

$$\varepsilon(u) = (\alpha^*(\phi))(u) = (\phi \circ \alpha)(u) = \phi(\alpha(u)) = \phi(0) = 0$$

  so $\varepsilon \in N(\alpha)^\circ$.

  Now use the fact that V and W are finite-dimensional:

$$\dim \operatorname{im}(\alpha^*) = r(\alpha^*) = r(\alpha)$$

  as $r(A) = r(A^T)$. On the other hand,

$$r(\alpha) = \dim V - \dim N(\alpha) = \dim(N(\alpha))^\circ$$

  Thus they are equal.

□

3.3 Double Dual

Let V be an $\mathbb{F}$-vector space. Then $V^* = L(V, \mathbb{F})$ is its dual space. The natural (oops) next step is

> **Definition** (Double dual)**.** The *double dual* of V is
>
> $$V^{**} = V^* = L(V^*, \mathbb{F}).$$

> **Theorem 3.8** (Naturallity of double dual)**.** *If V is an $\mathbb{F}$-vector space, then the map*
>
> $$\begin{aligned} \hat{\cdot} : V &\to V^{**} \\ v &\mapsto \hat{v} \end{aligned}$$
>
> *where $\hat{v}(\varepsilon) = \varepsilon(v)$, is a natural homomorphism. In particular when V is finite-dimensional this is a natural isomorphism.*

Proof.

- For $v \in V$, the map $\hat{v} : V^* \to \mathbb{F}$ is linear so $\hat{\cdot}$ does give a map from V to V^{**}.
- Linearity: for $v_1, v_2 \in V$, $\lambda_1, \lambda_2 \in \mathbb{F}$, $\varepsilon \in V^*$, then

  $$\widehat{\lambda_1 v_1 + \lambda_2 v_2}(\varepsilon) = \varepsilon(\lambda_1 v_1 + \lambda_2 v_2) = \lambda_1 \varepsilon(v_1) + \lambda_2 \varepsilon(v_2) = \lambda_1 \hat{v}_1(\varepsilon) + \lambda_2 \hat{v}_2(\varepsilon)$$
- Injectivity: let $e \in V \setminus \{0\}$. Extend it to a basis of V, say $e, e_2, \dots, e_n$. Let $\varepsilon, \varepsilon_2, \dots, \varepsilon_n$ be its corresponding dual basis. Now

  $$\hat{e}(\varepsilon) = \varepsilon(e) = 1$$

  so $\hat{e}$ is non-zero. $\hat{\cdot}$ is injective.
- Finally, if V is finite-dimensional, $\dim V = \dim V^* = \dim V^{**}$ so $\hat{\cdot}$ is an isomorphism.

□

> **Lemma 3.9.** *Let V be an $\mathbb{F}$-vector space and $U \leq V$. Then*
>
> $$\hat{U} \leq U^{\circ\circ}.$$
>
> *If V is finite-dimensional then $\hat{U} = U^{\circ\circ}$ so $U \cong U^{\circ\circ}$.*

Proof.

- First show $\hat{U} \leq U^{\circ\circ}$: given $u \in U$, for all $\varepsilon \in U^\circ$, $\varepsilon(u) = 0$ so $\hat{u}(\varepsilon) = 0$. Thus $\hat{u} \in (U^\circ)^\circ = U^{\circ\circ}$.
- If V is finite-dimensional then

  $$\dim U^{\circ\circ} = \dim V^* - \dim U^\circ = \dim V - \dim U^\circ = \dim U$$

  so $\hat{U} = U^{\circ\circ}$.

□

> **Lemma 3.10.** *Let V be a finite-dimensional $\mathbb{F}$-vector space and $U_1, U_2 \leq V$. Then*
>
> - $(U_1 + U_2)^\circ = U_1^\circ \cap U_2^\circ$,
> - $(U_1 \cap U_2)^\circ = U_1^\circ + U_2^\circ$.

Proof.

- Let $\theta \in V^*$, $\theta \in (U_1 + U_2)^\circ$ if and only if $\theta(u_1 + u_2) = 0$ for all $u_1 \in U_1, u_2 \in U_2$, if and only if $\theta(u) = 0$ for all $u \in U_1 \cup U_2$, so $\theta \in U_1^\circ \cap U_2^\circ$.
- Apply ° to the first result and use the previous lemma.

□

4 Bilinear Form I

> **Definition** (Bilinear form)**.** Let U and V be $\mathbb{F}$-vector spaces. A map $\varphi : U \times V \to \mathbb{F}$ is *bilinear* if it is linear in both arguments, i.e.
>
> $$\begin{aligned} \forall u \in U,\ \varphi(u, -) \in V^*, \\ \forall v \in V,\ \varphi(-, v) \in U^*. \end{aligned}$$

Example.

1. $V \times V^* \to \mathbb{F}, (v, \theta) \mapsto \theta(v)$.
2. $U = V = \mathbb{R}^n, \varphi(x, y) = \sum_{i=1}^n x_i y_i$.
3. $A \in \mathcal{M}_{m,n}(\mathbb{F}), \varphi : \mathbb{F}^m \times \mathbb{F}^n \to \mathbb{F}, (u, v) \mapsto u^T A v$.
4. $U = V = C([0, 1], \mathbb{R}), (f, g) \mapsto \int_0^1 f(t)g(t)dt$

> **Definition** (Matrix of bilinear form)**.** Let $\mathcal{B} = \{e_1, \dots, e_m\}$ be a basis for U and $\mathcal{C} = \{f_1, \dots, f_n\}$ be a basis for V. Given a bilinear map $\varphi : U \times V \to \mathbb{F}$, the *matrix of* φ with respect to $\mathcal{B}$ and $\mathcal{C}$ is
>
> $$[\varphi]_{\mathcal{B},\mathcal{C}} = (\varphi(e_i, f_j))_{m \times n}\,.$$

> **Lemma 4.1.**
> $$\varphi(u, v) = [u]_{\mathcal{B}}^T [\varphi]_{\mathcal{B},\mathcal{C}} [v]_{\mathcal{C}}.$$

Proof. Let $u = \sum_i \lambda_i e_i, v = \sum_j \mu_j f_j$, then

$$\begin{aligned} \varphi(u, v) &= \varphi\left(\sum_i \lambda_i e_i, \sum_j \mu_j f_j\right) \\ &= \sum_i \lambda_i \varphi\left(e_i, \sum_j \mu_j f_j\right) \\ &= \sum_{i,j} \lambda_i \varphi(e_i, f_j) \mu_j \end{aligned}$$

□

Note.

1. $[\varphi]_{\mathcal{B},\mathcal{C}}$ is the unique matrix with this property.
2. A bilinear form $\varphi : U \times V \to \mathbb{F}$ induces linear maps

$$\begin{aligned} \varphi_L : U &\to V^* \\ u &\mapsto \varphi(u, -) \\ \varphi_R : V &\to U^* \\ v &\mapsto \varphi(-, v) \end{aligned}$$

Lemma 4.2. *Let $\mathcal{B} = \{e_1, \dots, e_m\}$ be a basis for U, $\mathcal{B}^* = \{\varepsilon_1, \dots, \varepsilon_m\}$ a basis for U^*, $\mathcal{C} = \{f_1, \dots, f_n\}, \mathcal{C}^* = \{\eta_1, \dots, \eta_n\}$ for V and V^*. If $[\varphi]_{\mathcal{B},\mathcal{C}} = A$ then*

$$\begin{aligned}
[\varphi_L]_{\mathcal{B},\mathcal{C}^*} &= A^T, \\
[\varphi_R]_{\mathcal{C},\mathcal{B}^*} &= A.
\end{aligned}$$

Proof.

$$\varphi_L(e_i)(f_j) = A_{ij} \implies \varphi_L(e_i) = \sum_j A_{ij}\eta_j$$

$$\varphi_R(f_j)(e_i) = A_{ij} \implies \varphi_R(f_j) = \sum_i A_{ij}\varepsilon_i$$

□

Definition (Left and right kernel)**.** The *left (right, respectively) kernel* of φ is $\ker \varphi_L$ ($\ker \varphi_R$, respectively).

Definition (Degeneracy)**.** φ is *non-degenerate* if $\ker \varphi_L = 0$ and $\ker \varphi_R = 0$. Otherwise, φ is *degenerate.*

Lemma 4.3. *Let U, V have bases as before, and φ, A as before. Then φ is non-degenerate if and only if A is invertible.*

Proof.

$$\begin{aligned}
&\varphi \text{ is non-degenerate} \\
\Leftrightarrow &\ker \varphi_L = 0 \text{ and } \ker \varphi_R = 0 \\
\Leftrightarrow &n(A^T) = n(A) = 0 \\
\Leftrightarrow &r(A^T) = \dim V, r(A) = \dim U \\
\Leftrightarrow &A \text{ is invertible}
\end{aligned}$$

□

Corollary 4.4. *If φ is non-degenerate and U and V are finite-dimensional then* $\dim U = \dim V$.

Corollary 4.5. *When U and V are finite-dimensional, choosing a non-degenerate bilinear form $\varphi : U \times V \to \mathbb{F}$ is equivalent to picking an homomorphism $\varphi_L : U \to V^*$.*

> **Definition.** For $T \subseteq U, S \subseteq V$,
>
> $$T^\perp = \{v \in V : \varphi(t,v) = 0 \,\forall t \in T\} \leq V$$
> $$^\perp S = \{u \in U : \varphi(u,s) = 0 \,\forall s \in S\} \leq U$$

They are generalisation of annihilators.

> **Proposition 4.6.** *Suppose U have bases $\mathcal{B}, \mathcal{B}'$ and V have bases $\mathcal{C}, \mathcal{C}'$, $P = [\mathrm{id}]_{\mathcal{B}',\mathcal{B}}, Q = [\mathrm{id}]_{\mathcal{C}',\mathcal{C}}$. Let $\varphi : U \times V \to \mathbb{F}$ be a bilinear form. Then*
>
> $$[\varphi]_{\mathcal{B}',\mathcal{C}'} = P^T [\varphi]_{\mathcal{B},\mathcal{C}} Q.$$

Proof.

$$\begin{aligned} \varphi(u,v) &= [u]_{\mathcal{B}}^T [\varphi]_{\mathcal{B},\mathcal{C}} [v]_{\mathcal{C}} \\ &= (P[u]_{\mathcal{B}'})^T [\varphi]_{\mathcal{B},\mathcal{C}} (Q[v]_{\mathcal{C}'}) \\ &= [u]_{\mathcal{B}'}^T P^T [\varphi]_{\mathcal{B},\mathcal{C}} Q [v]_{\mathcal{C}'} \end{aligned}$$

□

> **Definition** (Rank of bilinear form)**.** The *rank* of φ, $r(\varphi)$, is the rank of its matrix representation (which is well-defined by the previous proposition).

Note.

$$r(\varphi) = r(\varphi_L) = r(\varphi_R).$$

5 Determinant & Trace

5.1 Trace

Definition (Trace). For $A \in \mathcal{M}_n(\mathbb{F}) = \mathcal{M}_{n,n}(\mathbb{F})$, the *trace* of A is

$$\operatorname{tr}(A) = \sum_{i=1}^{n} A_{ii}.$$

Lemma 5.1. *For $A, B \in \mathcal{M}_n(\mathbb{F})$,*

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

Proof.

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

□

Lemma 5.2. *Similar (or conjugate) matrices have the same trace.*

Proof. Suppose A and B are conjugates, then there exists P such that $B = P^{-1}AP$ so

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

□

Definition (Trace). Let $\alpha : V \to V$ be a linear map. The *trace* of α is

$$\operatorname{tr} \alpha = \operatorname{tr}[\alpha]_{\mathcal{B}} = \operatorname{tr}[\alpha]_{\mathcal{B},\mathcal{B}}$$

with repsect to a basis $\mathcal{B}$. This is well-defined by the previous lemma.

Lemma 5.3. *Let $\alpha : V \to V$ be linear and $\alpha^* : V^* \to V^*$ be its dual. Then*

$$\operatorname{tr} \alpha = \operatorname{tr} \alpha^*.$$

Proof.

$$\operatorname{tr} \alpha = \operatorname{tr}[\alpha]_{\mathcal{B}} = \operatorname{tr}[\alpha]_{\mathcal{B}}^T = \operatorname{tr}[\alpha^*]_{\mathcal{B}^*} = \operatorname{tr} \alpha^*.$$

□

5.2 Determinant

Recall some results from IA Groups: let S_n be the permutation group of the set $\{1, 2, \dots, n\}$ and $\varepsilon : S_n \to \{1, -1\}$ be the signature of a permutation, i.e.

$$\varepsilon(\sigma) = \begin{cases} 1 & \text{if } \sigma \text{ is a product of even number of transpotitions} \\ 0 & \text{otherwise} \end{cases}$$

Definition (Determinant)**.** Suppose $A \in \mathcal{M}_n(\mathbb{F})$, $A = (a_{ij})$, the *determinant* of A is

$$\det A = \sum_{\sigma \in S_n} \varepsilon(\sigma) a_{\sigma(1),1} a_{\sigma(2),2} \cdots a_{\sigma(n),n}.$$

There are $n!$ terms in this summation and each is the signed product of n elements (one from each row and each column).

Example. For $n = 2$,

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

Lemma 5.4. *If $A = (a_{ij})$ is an upper-triangular matrix (i.e. $a_{ij} = 0$ for all $i > j$) then*

$$\det A = a_{11}a_{22} \dots a_{nn}.$$

Similar for lower-trianglular matrices.

Proof. In the summation

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

for a summand to be non-zero, we need $\sigma(j) \leq j$ for all j. Thus $\sigma = \mathrm{id}$. □

Lemma 5.5.

$$\det A = \det A^T$$

Proof.

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

□

Definition (Volume form)**.** A *volume form* on $\mathbb{F}^n$ is a function

$$d : \underbrace{\mathbb{F}^n \times \mathbb{F}^n \times \cdots \times \mathbb{F}^n}_{n \text{ copies}} \to \mathbb{F}$$

which is:

- multilinear: for any i and $v_1, \dots, v_{i-1}, v_i, v_{i+1}, \dots, v_n \in \mathbb{F}^n$,
$$d(v_1, \dots, v_{i-1}, -, v_{i+1}, \dots, v_n) \in (\mathbb{F}^n)^*.$$
- alternating: if $v_i = v_j$ for $i \neq j$, $d(v_1, \dots, v_n) = 0$.

Notation. Given $A = (a_{ij})$, write A in column form
$$\left(A^{(1)} | \cdots | A^{(n)}\right).$$

For example, if $\{e_i\}$ is a standard basis for $\mathbb{F}^n$ then
$$I = (e_1 | \cdots | e_n).$$

Lemma 5.6.
$$\begin{aligned} \det : \mathbb{F}^n \times \cdots \times \mathbb{F}^n &\to \mathbb{F} \\ (A^{(1)}, \dots, A^{(n)}) &\mapsto \det A \end{aligned}$$
is a volume form.

Proof.

- Multilinear: for any fixed $\sigma \in S_n$, $\prod_{i=1}^n a_{\sigma(i),i}$ contains exactly one term from each column so it is multilinear. Multilinearity is preserved under addition.
- Alternating: suppose $A^{(k)} = A^{(l)}$ for some $l \neq k$. Let $\tau = (kl)$. Then $a_{ij} = a_{i\tau(j)}$ for all i, j. Also S_n can be expressed as a union of two disjoint cosets A_n and τA_n so
$$\det A = \sum_{\sigma \in A_n} \varepsilon(\sigma) \prod_{i=1}^n a_{i,\sigma(i)} - \sum_{\sigma \in A_n} \varepsilon(\sigma) \prod_{i=1}^n a_{i,\tau\sigma(i)}$$
since ε is a homomorphism
$$\begin{aligned} &= \sum_{\sigma \in A_n} \varepsilon(\sigma) \prod_{i=1}^n a_{i,\sigma(i)} - \sum_{\sigma \in A_n} \varepsilon(\sigma) \prod_{i=1}^n a_{i,\sigma(i)} \\ &= 0 \end{aligned}$$

□

In the rest of the section we are going to prove that the converse is also true, i.e. all volume forms are determinant up to a scaling constant.

Lemma 5.7. *Let d be a volume form. Then swapping two entries changes the sign:*
$$d(v_1, \dots, v_i, \dots, v_j, \dots, v_n) = -d(v_1, \dots, v_j, \dots, v_i, \dots, v_n).$$

Proof.

$$\begin{aligned}0 &= d(v_1,\dots,v_{i-1},v_i+v_j,v_{i+1},\dots,v_{j-1},v_i+v_j,v_{j+1},\dots,v_n)\\ &= \underbrace{d(v_1,\dots,v_i,\dots,v_i,\dots,v_n)}_{=0} + d(v_1,\dots,v_j,\dots,v_i,\dots,v_n)\\ &+ d(v_1,\dots,v_i,\dots,v_j,\dots,v_n) + \underbrace{d(v_1,\dots,v_j,\dots,v_j,\dots,v_n)}_{=0}\end{aligned}$$

Rearrange. □

Corollary 5.8. *If $\sigma \in S_n$,*

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

Theorem 5.9. *Let d be a volume form on $\mathbb{F}^n$, $A = (A^{(1)}|\dots|A^{(n)})$, then*

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

Proof.

$$\begin{aligned}d(A^{(1)},\dots,A^{(n)}) &= d\left(\sum_{i=1}^n a_{i1}e_i, A^{(2)},\dots,A^{(n)}\right)\\ &= \sum_{i=1}^n a_{i_1} d(e_i, A^{(2)},\dots,A^{(n)})\\ &= \sum_i\sum_j a_{i1}a_{j2} d(e_i,e_j,\dots,A^{(n)})\\ &= \sum_{i_1,i_2,\dots,i_n}\prod_{k=1}^n a_{i_k,k} d(e_{i_1},\dots e_{i_n})\end{aligned}$$

The last term is 0 unless all of i_k are distinct, i.e. exists $\sigma \in S_n$ such that $i_k = \sigma(k)$. Thus

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

□

Corollary 5.10. det *is the unique volume form d such that $d(e_1,\dots,e_n) =$* 1.

Proposition 5.11. *Suppose $A, B \in \mathcal{M}_n(\mathbb{F})$, then*

$$\det AB = \det A \det B.$$

Proof. Define

$$\begin{aligned} d_A : \mathbb{F}^n \times \cdots \times \mathbb{F}^n &\to \mathbb{F} \\ (v_1, \dots, v_n) &\mapsto \det(Av_1|\cdots|Av_n) \end{aligned}$$

which is a volume form:

- multilinear: $v_i \mapsto Av_i$ is linear and det is multilinear.
- alternating: $v_i = v_j$ implies $Av_i = Av_j$ and det is alternating.

It follows that

$$\begin{aligned} d_A(Be_1, \dots, Be_n) &= \det B \cdot d_A(e_1, \dots, e_n) = \det B \det A \\ &= \det(ABe_1|\cdots|ABe_n) = \det AB \end{aligned}$$

□

> **Definition** (Singular)**.** $A \in \mathcal{M}_n(\mathbb{F})$ is *singular* if $\det A = 0$. Otherwise it is *non-singular*.

> **Lemma 5.12.** *If A is invertible then it is non-singular and*
>
> $$\det A^{-1} = \frac{1}{\det A}.$$

Proof.

$$1 = \det I_n = \det(AA^{-1}) = \det A \det A^{-1}$$

□

> **Theorem 5.13.** *Suppose $A \in \mathcal{M}_n(\mathbb{F})$ then TFAE:*
>
> 1. *A is invertible,*
> 2. *A is non-singular,*
> 3. *$r(A) = n$.*

Proof.

- $1 \Rightarrow 2$: done.
- $2 \Rightarrow 3$: suppose that $r(A) < n$. By rank-nullity $n(A) > 0$ so $\exists \lambda \in \mathbb{F}^n \setminus \{0\}$ such that $A\lambda = 0$. Say $\lambda = (\lambda_i)$ and $\lambda_k \neq 0$. Have $\sum_{i=1}^n A^{(i)}\lambda_i = 0$. Let

  $$B = (e_1|e_2|\cdots|e_{k-1}|\lambda|e_{k+1}|\cdots|e_n)$$

  It follows that AB has kth column zero so

  $$0 = \det AB = \det A \det B = \lambda_k \det A.$$

  So $\det A = 0$.
- $3 \Rightarrow 1$: by rank-nullity.

□

5.3 Determinant of Linear Maps

Lemma 5.14. *Conjugate matrices have the same determinant.*

Proof. Let $B = P^{-1}AP$. Then

$$\det B = \det(P^{-1}AP) = \det P^{-1} \det A \det P = \det(P^{-1}P) \det A = \det A.$$

□

Definition (Determinant)**.** Let $\alpha : V \to V$ where V is a finite-dimensional vector space. The *determinant* of α is

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

where $\mathcal{B}$ is any basis for V.

This is well-defined by the previous lemma.

Theorem 5.15. $\det : L(V, V) \to \mathbb{F}$ *satisfies*

1. $\det \mathrm{id} = 1$,
2. $\det \alpha \circ \beta = \det \alpha \det \beta$,
3. $\det \alpha \neq 0$ *if and only if* α *is invertible, in which case* $\det(\alpha^{-1}) = \frac{1}{\det \alpha}$.

Proof. Restatement of previous results. □

5.4 Determinant of Block-triangular Matrices

Lemma 5.16. *Suppose* $A \in \mathcal{M}_k(\mathbb{F}), B \in \mathcal{M}_\ell(\mathbb{F})$ *and* $C \in \mathcal{M}_{k,\ell}(\mathbb{F})$*, then*

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

Proof. Let $n = k + \ell$ and call the block matrix $X = (x_{ij})$, which is an element of $\mathcal{M}_n(\mathbb{F})$. Then

$$\det X = \sum_{\sigma \in S_n} \varepsilon(\sigma) \prod_{i=1}^{n} x_{\sigma(i),i}$$

Note that $x_{\sigma(i),i} = 0$ if $i \leq k$ and $\sigma(i) > k$. Thus we are only concerned about σ under which these are in different orbits, i.e. $\sigma = \sigma_1\sigma_2$ where $\sigma_1 \in \mathrm{Sym}_{\{1,\dots,k\}}$ and $\sigma_2 \in \mathrm{Sym}_{\{k+1,\dots,n\}}$.

$$\begin{aligned} &= \sum_{\sigma_1 \in \mathrm{Sym}_{\{1,\dots,k\}}} \varepsilon(\sigma_1) \prod_{j=1}^{k} a_{\sigma_1(j),j} \\ &\times \sum_{\sigma_2 \in \mathrm{Sym}_{\{k+1,\dots,n\}}} \varepsilon(\sigma_2) \prod_{j=k+1}^{n} a_{\sigma_2(j),j} \\ &= \det A \det B \end{aligned}$$

□

Corollary 5.17. *For a sequence of matrices $A_1, \dots, A_k$,*

$$\det \begin{pmatrix} A_1 & & & & \\ & A_2 & & * & \\ & & A_3 & & \\ & 0 & & \ddots & \\ & & & & A_k \end{pmatrix} = \prod_{i=1}^k \det A_i$$

Proof. Apply the previous lemma inductively. □

Caution. In general,

$$\det \begin{pmatrix} A & B \\ C & D \end{pmatrix} \neq \det A \det D - \det B \det C.$$

5.5 Volume Interpretation of Determinant

In $\mathbb{R}^2$, the determinant of a matrix

$$\det \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$$

can be intepreted as the signed area of the parallelogram spanned by the column vectors $\binom{a_{11}}{a_{21}}$ and $\binom{a_{12}}{a_{22}}$ of the matrix.

Similarly in $\mathbb{R}^3$ the determinant of a matrix is the signed volume of the parallelepiped spanned by the column vectors of the matrix.

For higher dimensions, although difficult to visualise, the same interpretation still works: consier a hypercube $H = [0,1]^n \subseteq \mathbb{R}^n$. Then a map $A \in \mathcal{M}_n(\mathbb{F})$ sends

$$\begin{aligned} H &\to A(H) \\ \sum_{i=1}^n t_i e_i &\mapsto \sum_{i=1}^n t_i A^{(i)} \end{aligned}$$

and the generalised signed volume of RHS is $\det A$.

5.6 Determinant of Elementary Operation

Consider the determinants of elementary column operation matrices:

- E_1 swaps two columns so $\det E_1 = -1$,
- E_2 multiplies a column by $\lambda \neq 0$ so $\det E_2 = \lambda$,
- E_3 adds λ times of a column to another column so $\det E_3 = 1$.

One could prove properties of det by decomposing any matrix into elementary matrices.

5.7 Column Expansion & Adjugate Matrices

Lemma 5.18. *Suppose $A \in \mathcal{M}_n(\mathbb{F})$, $A = (a_{ij})$. Define $A_{\widehat{ij}} \in \mathcal{M}_{n-1}(\mathbb{F})$ by deleting row i and column j from A. Then $\det A$ can be calculated by*

1. *expansion in column j: for a fixed j,*
$$\det A = \sum_{i=1}^{n} (-1)^{i+j} a_{ij} \det A_{\widehat{ij}}.$$
2. *expansion in row i: for a fixed i,*
$$\det A = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} \det A_{\widehat{ij}}.$$

Remark. It is possible to use one of the expressions above to define determinant inductively, with base case $\det a = a$ for $n = 1$.

Example.
$$\begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} = a \begin{vmatrix} e & f \\ h & i \end{vmatrix} - b \begin{vmatrix} d & f \\ g & i \end{vmatrix} + c \begin{vmatrix} d & e \\ g & h \end{vmatrix}$$

Proof. We prove 1:

$$\begin{aligned} \det A &= \det \left(A^{(1)} | \cdots | \sum_{i=1}^{n} a_{ij} e_i | \cdots | A^{(n)} \right) \\ &= \sum_{i=1}^{n} a_{ij} \det(A^{(1)} | \cdots | e_i | \cdots | A^{(n)}) \end{aligned}$$

use row and column operations to move the entry to top left corner,

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

□

Definition (Adjugate)**.** Let $A \in \mathcal{M}_n(\mathbb{F})$. The *adjugate matrix* of A, $\operatorname{adj} A$, is the $n \times n$ matrix
$$(\operatorname{adj} A)_{ij} = (-1)^{i+j} \det A_{\widehat{ji}}.$$

Notice the transposition of indices.

Theorem 5.19.

1. $(\operatorname{adj} A)A = \det A \cdot I$,

2. *If A is invertible then*

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

Proof.

1. For a fixed j, $\det A = \sum_i (\operatorname{adj} A)_{ji} a_{ij} = (\operatorname{adj} A \cdot A)_{jj}$. For $j \neq k$, replace the jth column with the kth:

$$\begin{aligned} 0 &= \det(A^{(1)} | \cdots | A^{(k)} | \cdots | A^{(k)} | \cdots | A^{(n)}) \\ &= \sum_i (\operatorname{adj} A)_{ji} a_{ik} \\ &= (\operatorname{adj} A \cdot A)_{jk} \end{aligned}$$

2. If A is invertible then $\det A \neq 0$ so

$$I = \frac{\operatorname{adj} A}{\det A} A.$$

□

5.8 Application: System of Linear Equations

A system of linear equations can be written as

$$A\mathbf{x} = \mathbf{b}$$

where $A \in \mathcal{M}_{m,n}(\mathbb{F})$ and $\mathbf{b} \in \mathcal{M}_{m,1}(\mathbb{F})$ are known and $\mathbf{x} \in \mathcal{M}_{n,1}(\mathbb{F})$ is unknown.

The system has a solution if and only if $r(A) = r(A|\mathbf{b})$ where the matrix on RHS is the *augmented matrix* by adding $\mathbf{b}$ as a column to A since this happens if and only if $\mathbf{b}$ is a linear combination of columns of A.

The solution is unique if and only if $r(A) = n$.

In particular, if $m = n$, if A is non-singular then there is a unique solution

$$\mathbf{x} = A^{-1}\mathbf{b}.$$

Although in theory we could invert the matrix to solve the system of equations, it is terribly inefficient. Instead, we use

Proposition 5.20 (Cramer's rule)**.** *If $A \in \mathcal{M}_n(\mathbb{F})$ is invertible then the system*

$$A\mathbf{x} = \mathbf{b}$$

has unique solution $\mathbf{x} = (x_i)$ where

$$x_i = \frac{\det(A_{\hat{i}\mathbf{b}})}{\det A}$$

where $A_{\hat{i}\mathbf{b}}$ is obtained from A by deleting ith column and replacing it with $\mathbf{b}$.

Proof. Assume $\mathbf{x}$ is a solution of the system.

$$\begin{aligned}\det(A_{\hat{i}\mathbf{b}}) &= \det(A^{(1)}|\cdots|\mathbf{b}|\cdots|A^{(n)}) \\ &= \det(A^{(1)}|\cdots|A\mathbf{x}|\cdots|A^{(n)}) \\ &= \sum_{j=1}^{n} x_j \det(A^{(1)}|\cdots|A^{(j)}|\cdots|A^{(n)})\end{aligned}$$

$A^{(j)}$ is one of the other columns unless $j = i$ so

$$= x_i \det A$$

□

Corollary 5.21. *If* $A \in \mathcal{M}_n(\mathbb{Z})$ *with* $\det A = \pm 1$*, then*

1. $A^{-1} \in \mathcal{M}_n(\mathbb{Z})$.
2. *Given* $\mathbf{b} \in \mathbb{Z}^n$*,* $A\mathbf{x} = \mathbf{b}$ *has an integer solution.*

6 Endomorphism

6.1 Definitions

Let V be an $\mathbb{F}$-vector space with $\dim V = n < \infty$. Let $\mathcal{B} = \{v_1, \dots, v_n\}$ be a basis and $\alpha \in L(V) = L(V, V)$. The general problem studied in this chapter is to choose a basis $\mathcal{B}$ such that $[\alpha]_\mathcal{B}$ has "nice forms", for example, to be amenable to det and tr.

Suppose there is another basis $\mathcal{B}'$ with change-of-basis matrix P. Recall that

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

The above problem is thus euqivalent to the following: given $A \in \mathcal{M}_n(\mathbb{F})$, find A' conjugate to A and in a "nice form".

What are the nice forms that we desire? The best we can have is

> **Definition** (Diagonalisable). $\alpha \in L(V)$ is *diagonalisable* if there exists $\mathcal{B}$ such that $[\alpha]_\mathcal{B}$ is diagonal.

A slightly weaker, albeit still "nice" enough form is

> **Definition** (Triangulable). $\alpha \in L(V)$ is *triangulable* if there exists $\mathcal{B}$ such that $[\alpha]_\mathcal{B}$ is upper triangular.

Equivalent, rephrasing using languages of matrices, $A \in \mathcal{M}_n(\mathbb{F})$ is diagonalisable (triangulable, respectively) if it is conjugate to a diagonal (upper triangle, respectively) matrix.

> **Definition** (Eigenvalue, eigenvector, eigenspace).
>
> 1. $\lambda \in \mathbb{F}$ is an *eigenvalue* of α if there exists some $v \in V \setminus \{0\}$ such that $\alpha(v) = \lambda v$.
> 2. $v \in V$ is an *eigenvector* of α if $\alpha(v) = \lambda v$ for some eigenvalue λ.
> 3. $V_\lambda = \{v \in V : \alpha(v) = \lambda v\}$ is the λ-*eigenspace* of α.

Remark. It is easy to check that $V_\lambda \leq V$.

Remark.

1. $V_\lambda = \ker(\alpha - \lambda\iota)$ and

$$\begin{aligned} &\lambda \text{ is an eigenvalue} \\ \Leftrightarrow &\alpha - \lambda\iota \text{ is singular} \\ \Leftrightarrow &\det(\alpha - \lambda\iota) = 0 \end{aligned}$$

2. If $\alpha(v_j) = \lambda v_j$ then the jth column of $[\alpha]_\mathcal{B}$ is $(0, \dots, \lambda, \dots, 0)^T$.
3. $[\alpha]_\mathcal{B}$ is diagonal if and only if $\mathcal{B}$ consists of eigenvectors. $[\alpha]_\mathcal{B}$ is upper triangular if and only if $\alpha(v_j) \in \langle v_1, \dots, v_j\rangle$ for all j. In particular, v_1 is an eigenvector.

6.2 Polynomial Ring, an Aside

Before discussing polynomials associated with a linear map, we need some background knowledge about the ambient polynomial space that we will be working with. The following results should be self-evident and proofs are omitted. Most of them will be studied in detail in IB Groups, Rings and Modules and a proof the Fundamental Theorem of Algebra can be found in IB Complex Analysis.

Let

$$\mathbb{F}[t] = \{\text{polynomials with coefficients in } \mathbb{F}\}$$

and $\deg f$ be the degree of f in $\mathbb{F}[t]$. In addition for the convenience of stating the following properties we let $\deg 0 = -\infty$. We have the following properties:

1. $\deg(f+g) \leq \max(\deg f, \deg g), \deg(fg) = \deg f + \deg g$.
2. If $\lambda \in \mathbb{F}$ is a root of some $f \in \mathbb{F}[t]$, i.e. $f(\lambda) = 0$ then $(t-\lambda) \mid f$. In other words, $f(t) = (t-\lambda)g(t)$ for some $g(t) \in \mathbb{F}[t]$ and $\deg g = \deg f - 1$.
3. We say λ is a root of $f \in \mathbb{F}[t]$ with *multiplicity* $e \in \mathbb{N}$ if $(t-\lambda)^e \mid f$ but $(t-\lambda)^{e+1} \nmid f$.
4. A polynomial of degree n has at most n roots, counted with multiplicity.
5. Fundamental Theorem of Algebra: any $f \in \mathbb{C}[t]$ of positive degree has a root (hence $\deg f$ roots).

6.3 Characteristic Polynomial of Endormorphism

> **Definition** (Characteristic polynomial)**.** The *characteristic polynomial* of $\alpha \in L(V)$ is
>
> $$\chi_\alpha(t) = \det(\alpha - t\iota).$$
>
> The *characteristic polynomial* of $A \in \mathcal{M}_n(\mathbb{F})$ is
>
> $$\chi_A(t) = \det(A - tI).$$

Conjugate matrices have the same characteristic polynomial.

> **Theorem 6.1.** *A linear map α is triangulable if and only if $\chi_\alpha(t)$ can be written as a product of linear factors over $\mathbb{F}$.*

Proof.

- $\Rightarrow$: suppose α is triangulable and is represented by

  $$\begin{pmatrix} a_1 & \cdots & * \\ & \ddots & \vdots \\ 0 & & a_n \end{pmatrix}$$

  with respect to some basis. Then

  $$\chi_\alpha(t) = \det \begin{pmatrix} a_1 - t & \cdots & * \\ & \ddots & \vdots \\ 0 & & a_n - t \end{pmatrix} = \prod_{i=1}^n (a_i - t)$$

- $\Leftarrow$: induction of $n = \dim V$: if $n = 1$ then done. Suppose $n > 1$ and the theorem holds for all endomorphisms of spaces of smaller dimensions. By hypothesis $\chi_\alpha(t)$ has a root in $\mathbb{F}$, say λ. Let $U = V_\lambda \neq 0$, then $\alpha(U) \leq U$ so α induces $\overline{\alpha} : V/U \to V/U$. Pick basis $v_1, \dots, v_k$ for U and extend it to a basis $\mathcal{B} = \{v_1, \dots, v_n\}$ for V. With respect to $\mathcal{B}$, α has representation

  $$\begin{pmatrix} \lambda I_k & * \\ 0 & C \end{pmatrix}$$

  so

  $$\chi_\alpha(t) = \det(\alpha - t\iota) = (\lambda - t)^k \chi_{\overline{\alpha}}(t).$$

  Thus $\chi_{\overline{\alpha}}(t)$ is also a product of linear factors. Since $\chi_{\overline{\alpha}}(t)$ acts on a linear space of strictly smaller dimension, by induction hypothesis there is a basis $w_{k+1} + U, \dots, w_n + U$ for V/U with respect to which $\overline{\alpha}$ has an upper-triangular matrix representation, say T. Then with respect to basis $v_1, \dots, v_k, w_{k+1}, \dots, w_n$, α has matrix representation

  $$\begin{pmatrix} \lambda I_k & * \\ 0 & T \end{pmatrix}$$

□

Example. Let $\mathbb{F} = \mathbb{R}$, $V = \mathbb{R}^2$ and α be a rotation. Then with respect to the standard basis α has representation

$$\begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}$$

and thus $\chi_\alpha(t) = t^2 - 2\cos\theta t + 1$, which is irreducible in general. Thus α is not triangulable over $\mathbb{R}$.

> **Lemma 6.2.** *Let V be an n-dimensional $\mathbb{F}$-vector space and $\alpha \in L(V)$ with $\chi_\alpha(t) = (-1)^n t^n + c_{n-1}t^{n-1} + \dots c_0$. Then*
>
> - $c_0 = \det\alpha$,
> - $c_{n-1} = (-1)^{n-1} \operatorname{tr}\alpha$ *for* $\mathbb{F} = \mathbb{R}$ *or* $\mathbb{C}$.

Proof.

- $c_0 = \chi_\alpha(0) = \det(\alpha - 0) = \det\alpha$.
- If $\mathbb{F} = \mathbb{R}$ then there is an extension of scalars $\mathcal{M}_n(\mathbb{R}) \otimes \mathbb{C} \hookrightarrow \mathcal{M}_n(\mathbb{C})$ induced by $\mathbb{R} \hookrightarrow \mathbb{C}$ (i.e. complexification). For $\mathbb{F} = \mathbb{C}$, use Fundamental Theorem of Algebra to write

  $$\chi_\alpha(t) = \det \begin{pmatrix} a_0 - t & \cdots & * \\ & \ddots & \vdots \\ 0 & & a_n - t \end{pmatrix} = \prod_{i=1}^n (a_i - t)$$

  where $\sum_{i=1}^n a_i = \operatorname{tr}\alpha$.

□

Notation. Let $p(t)$ be a polynomial over $\mathbb{F}$,

$$p(t) = a_n t^n + \cdots + a_0 \in \mathbb{F}[t].$$

For $A \in \mathcal{M}_n(\mathbb{F})$, define

$$p(A) = a_n A^n + \cdots + a_0 I \in \mathcal{M}_n(\mathbb{F}).$$

For $\alpha \in L(V)$, define

$$p(\alpha) = a_n \alpha^n + \cdots + a_0 \iota \in L(V).$$

> **Theorem 6.3.** *Let V be a finite-dimensional $\mathbb{F}$-vector space. Let $\alpha \in L(V)$. Then α is diagonalisable if and only if $p(\alpha) = 0$ for some $p \in \mathbb{F}[t]$ which is the product of distinct linear factors.*

Proof.

1. $\Rightarrow$: Suppose α is diagonalisable with distinct eigenvalues $\lambda_1, \dots, \lambda_k$. Let

   $$p(t) = (t - \lambda_1) \cdots (t - \lambda_k).$$

   Let $\mathcal{B}$ be a basis of eigenvectors. For $v \in \mathcal{B}$, $\alpha(v) = \lambda_i v$ for some i. Thus

   $$0 = (\alpha - \lambda_i \iota) v \Rightarrow p(\alpha)(v) = 0.$$

   As this holds for all $v \in \mathcal{B}$, $p(\alpha) = 0$.

2. $\Leftarrow$: Suppose $p(\alpha) = 0$ for this p, which is monic wlog. Claim that

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

   Proof. For $j = 1, \dots, k$, let

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

   and $q(t) = \sum_{j=1}^k q_j(t)$. $q(t)$ has degree at most $k-1$ and $q(\lambda_i) = 1$ for all $i = 1, \dots, k$ so $q(t) = 1$.

   Let $\pi_j = q_j(\alpha) : V \to V$. By construction

   $$\sum_{j=1}^k \pi_j = q(\alpha) = \iota \in L(V)$$

   so given $v \in V$,

   $$v = q(\alpha)(v) = \sum_{j=1}^k \pi_j(v).$$

Also

$$(\alpha - \lambda_j \iota)(\pi_j(v)) = (\alpha - \lambda_j \iota)(q_j(\alpha)(v)) = \frac{1}{\prod_{i \neq j}(\lambda_j - \lambda_i)} \underbrace{p(\alpha)}_{=0}(v) = 0.$$

We have thus shown that

$$\operatorname{im} \pi_j \leq \ker(\alpha - \lambda_j \iota) = V_{\lambda_j}.$$

Thus $V = \sum_{j=1}^k V_{\lambda_j}$.

To prove the sum is direct, suppose $v \in V_{\lambda_j} \cap (\sum_{i \neq j}^k V_{\lambda_i})$ and apply π_j to v:

$$\begin{aligned} v \in V_{\lambda_j} &\Rightarrow \pi_j(v) = \prod_{i=1, i \neq j}^k \frac{\lambda_j - \lambda_i}{\lambda_j - \lambda_i} v = v \\ v \in \sum_{i \neq j}^k V_{\lambda_i} &\Rightarrow \pi_j(v) = 0 \end{aligned}$$

so $v = 0$ and the sum is direct.

Now take the union of bases for V_{λ_i} as a basis for V. □

□

Remark.

1. π_j is the projection from V to V_{λ_j}.
2. The proof shows that for k distinct eigenvalues $\lambda_1, \dots \lambda_k$ of α, the sum $\sum_j V_{\lambda_j}$ is direct. The only way for diagonalisation to fail is if $\sum_j V_{\lambda_j} \lneq V$.

Corollary 6.4. *Suppose $A \in \mathcal{M}_n(\mathbb{C})$ has finite order then A is diagonalisable.*

Proof. $p(A) = 0$ for $p(t) = t^m - 1$ where m is the order of A. This factorises as $\prod_{i=0}^{m-1}(t - \xi^i)$ where ξ is a primitive mth root of unity. □

Theorem 6.5 (Simultaneous diagonalisation)**.** *Let $\alpha, \beta \in L(V)$ be diagonalisable. Then α and β are* simultaneous diagonalisable *(there exists a basis with respect to which they are both diagonal) if and only if α and β commute.*

Proof.

- $\Rightarrow$: Suppose there is a basis $\mathcal{B}$ such that $A = [\alpha]_{\mathcal{B}}$ and $B = [\beta]_{\mathcal{B}}$ are diagonal. Any two diagonal matrices commute so $AB = BA$, $\alpha\beta = \beta\alpha$.

- $\Leftarrow$: Suppose α and β commute and both are diagonalisable. We have

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

  where $V_i = \ker(\alpha - \lambda_i \iota)$. Claim that $\beta(V_j) \leq V_j$: suppose $v \in V_j$,

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

  As β is diagonalisable, there is a polynomial p with distinct linear factors such that $p(\beta) = 0$. Now

  $$p(\beta|_{V_i}) = p(\beta)|_{V_i} = 0$$

  so $\beta|_{V_i} \in L(V_i)$ is diagonal. Pick a basis $\mathcal{B}_i$ of V_i combining its eigenvectors for β. By construction these are also eigenvectors for α. With respect to $\mathcal{B} = \bigcup_i \mathcal{B}_i$ both α and β are diagonal.

□

Lemma 6.6 ($\mathbb{F}[t]$ as a Euclidean domain)**.** *Given $a, b \in \mathbb{F}[t]$ with $b \neq 0$, there eixst $q, r \in \mathbb{F}[t]$ with $\deg r < \deg b$ and $a = qb + r$.*

Proof. IB Groups, Rings and Modules. □

Definition (Minimal polynomial)**.** Suppose $\alpha \in L(V)$ and V is finite-dimensional. The *minimal polynomial* of α, m_α, is the monic non-zero polynomial of smallest degree such that

$$m_\alpha(\alpha) = 0.$$

Remark. Let $\dim V = n < \infty$, $\dim L(V) = n^2$ so

$$\iota, \alpha, \alpha^2, \dots, \alpha^{n^2} \in L(V)$$

must be linearly dependent so there is a non-trivial relation. Thus minimal polynomial exists.

Lemma 6.7. *Let $\alpha \in L(V)$, $p \in \mathbb{F}[t]$. Then $p(\alpha) = 0$ if and only if $m_\alpha(t) \mid p(t)$.*

Proof. By Euclidean algorithm there exist $q, r \in \mathbb{F}[t]$ such that

$$p(t) = m_\alpha(t) q(t) + r(t)$$

where $\deg r < \deg m_\alpha$. Then

$$0 = p(\alpha) = m_\alpha(\alpha) q(\alpha) + r(\alpha)$$

so $r(\alpha) = 0$. By the minimality of the degree of m_α, $r = 0$. □

Corollary 6.8. *m_α is uniquely defined.*

Proof. Suppose m_1 and m_2 are both minimal polynomials of α. Then by the previous lemma $m_1 \mid m_2$ and vice versa. By assumption both of them are monic so $m_1 = m_2$. □

Theorem 6.9 (Cayley-Hamilton Theorem)**.** *Let V be a finite-dimensional $\mathbb{F}$-vector space. Let $\alpha \in L(V)$. Then*

$$\chi_\alpha(\alpha) = 0.$$

First we give a proof for $\mathbb{F} = \mathbb{C}$:

Proof. For some basis $\mathcal{B} = \{v_1, \dots, v_n\}$, α has matrix representation

$$[\alpha]_\mathcal{B} = \begin{pmatrix} a_1 & \cdots & * \\ & \ddots & \vdots \\ 0 & & a_n \end{pmatrix}$$

Let $U_j = \langle v_1, \dots, v_j \rangle$. Then $(\alpha - a_j\iota)U_j \leq U_{j-1}$ so

$$\underbrace{(\alpha - a_1\iota)(\alpha - a_2\iota)\cdots\underbrace{(\alpha - a_{n-1}\iota)\underbrace{(\alpha - a_n\iota)V}_{\leq U_{n-1}}}_{\leq U_{n-2}}}_{\leq (\alpha - a_1\iota)U_1 = 0} = 0$$

so

$$\chi_\alpha(\alpha) = 0.$$

□

However, this proof is unsatisfactory in that it relies on the fact of $\mathbb{C}$ being algebraically closed, which is partially, to say the least, an analytical and not an algebraic property[1]. In actuality, Cayley-Hamilton is a general result that applies to all fields. Thus we give an alternative algebraic proof:

Proof. (Non-examinable) Let $A \in \mathcal{M}_n(\mathbb{F})$, then

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

For any matrix B, we have

$$B \operatorname{adj} B = \det B \cdot I.$$

Let $B = tI - A$. Then $\operatorname{adj} B$ is a matrix whose entries are polynomials in t of degree smaller than n, i.e. polynomials in t with coefficients in $\mathcal{M}_n(\mathbb{F})$.

Thus

$$(tI - A)\underbrace{(B_{n-1}t^{n-1} + \cdots + B_1t + B_0)}_{\operatorname{adj} B} = \underbrace{(t^n + a_{n-1}t^{n-1} + \cdots + a_0)}_{\det B} I$$

[1] While being closed is an algebraic property, the construction of $\mathbb{C}$ via $\mathbb{R}$ from $\mathbb{Q}$ is not. The point here is that Caylay-Hamilton holds for all fields, not just closed ones.

Equating the coefficients of each power of t,

$$\begin{aligned} I &= B_{n-1} \\ a_{n-1}I &= B_{n-2} - AB_{n-1} \\ &\vdots \\ a_0 I &= -AB_0 \end{aligned}$$

multiply by A^{n-i+1} for the ith row

$$\begin{aligned} A^n &= A^n B_{n-1} \\ a_{n-1}A^{n-1} &= A^{n-1}B_{n-2} - A^n B_{n-1} \\ &\vdots \\ a_0 I &= -AB_0 \end{aligned}$$

and add them up,

$$A^n + a_{n-1}A^{n-1} + \dots a_1 A + A_0 I = 0.$$

□

Definition (Algebraic multiplicity)**.** Let λ be an eigenvalue of $\alpha \in L(V)$ where V is a finite-dimensional $\mathbb{F}$-vector space. Write

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

for some $q(t) \in \mathbb{F}[t]$ and $(t-\lambda) \nmid q(t)$. a_λ is the *algebraic multiplicity* of λ as an eigenvalue of α.

Definition (Geometric multiplicity)**.** $g_\lambda = n(\alpha - \lambda\iota)$ is the *geometric multiplicity* of α.

Lemma 6.10. *If λ is an eigenvalue then*

$$1 \le g_\lambda \le a_\lambda.$$

Proof. $1 \le g_\lambda$ since $\alpha - \lambda\iota$ is singular.

Let $\mathcal{B} = \{v_1, \dots v_n\}$ be a basis of V with $\{v_1, \dots, v_g\}$ a basis of $\ker(\alpha - \lambda\iota)$. Let $g = g_\lambda$. Then

$$[\alpha]_\mathcal{B} = \begin{pmatrix} \lambda I_g & * \\ 0 & A_1 \end{pmatrix}$$

where $A_1 \in \mathcal{M}_{n-g}(\mathbb{F})$. Thus

$$\chi_\alpha(t) = (t-\lambda)^g \alpha_{A_1}(t)$$

and $g_\lambda \le a_\lambda$. □

> **Lemma 6.11.** *Let λ be an eigenvalue. Let c_λ be the multiplicity of λ as a root of m_α. Then*
> $$1 \leq c_\lambda \leq a_\lambda.$$

Proof. As $m_\alpha \mid \chi_\alpha$, $c_\lambda \leq a_\lambda$.

As λ is an eigenvalue, $\alpha(v) = \lambda v$ for some $v \neq 0$. Now given $p \in \mathbb{F}[t]$, $p(\alpha)(v) = p(\lambda)(v)$. Apply this to m_α,

$$0 = m_\alpha(\alpha)(v) = m_\alpha(\lambda)(v)$$

so $m_\alpha(\lambda) = 0$. □

Example.

1. Let
$$A = \begin{pmatrix} 1 & 0 & -2 \\ 0 & 1 & 1 \\ 0 & 0 & 2 \end{pmatrix}$$
then
$$\chi_A(t) = \det(A - tI) = (2-t)(1-t)^2.$$
There are two candidates for the minimal polynomial:
   - $(t-2)(t-1)^2$,
   - $(t-2)(t-1)$.

   We can check that $(A-I)(A-2I) = 0$ so the second one is the minimal polynomial. It follows that A is diagonalisable.
2. Let
$$A = \begin{pmatrix} \lambda & 1 & & \\ & \lambda & 1 & \\ & & \ddots & 1 \\ & & & \lambda \end{pmatrix}$$
which has $g_\lambda = 1, a_\lambda = n, c_\lambda = n$.

> **Lemma 6.12.** *Let $\alpha \in L(V)$, then TFAE:*
>
> 1. *α is diagonalisable,*
> 2. *$a_\lambda = g_\lambda$ for all eigenvalues λ,*
> 3. *if $\mathbb{F} = \mathbb{C}$, $c_\lambda = 1$ for all eigenvalues λ.*

Proof.

- $1 \Leftrightarrow 2$: Let $\lambda_1, \dots, \lambda_k$ be eigenvalues of α. Then α is diagonalisable if and only if $V = \bigoplus_i V_{\lambda_i}$. Take dimension of both sides,
$$\dim V = n = \deg \chi_\alpha = a_1 + \dots + a_k$$
$$\dim \bigoplus_i V_{\lambda_i} = g_1 + \dots + g_k$$
But $g_i \leq a_i$ for all i so α is diagonalisable if and only if $g_i = \alpha_i$ for all i.

- 2 $\Leftrightarrow$ 3: By the Fundamental Theorem of Algebra, m_α is a product of linear factors. α is diagonalisable if and only if these are all distinct, i.e. $c_\lambda = 1$ for all eigenvalues λ.

□

Remark. Over $\mathbb{C}$,

$$\begin{aligned}\chi_\alpha(t) &= (\lambda_1 - t)^{a_1} \cdots (\lambda_k - t)^{a_k} \\ m_\alpha(t) &= (t - \lambda_1)^{c_1} \cdots (t - \lambda_k)^{c_k}\end{aligned}$$

with $1 \leq c_i \leq a_i$.

Definition (Jordan normal form)**.** $A \in \mathcal{M}_n(\mathbb{F})$ is in *Jordan normal form* if it is a block diagonal matrix

$$A = \begin{pmatrix} J_{n_1}(\lambda_1) & & & \\ & J_{n_2}(\lambda_2) & & \\ & & \ddots & \\ & & & J_{n_k}(\lambda_k) \end{pmatrix}$$

where $k \geq 1$, $n_1, \dots, n_k \in \mathbb{N}$ with $\sum_i n_i = n$, $\lambda_i \in \mathbb{F}$ not necessarily distinct and

$$J_m(\lambda) = \begin{pmatrix} \lambda & 1 & & \\ & \lambda & 1 & \\ & & \ddots & 1 \\ & & & \lambda \end{pmatrix} \in M_m(\mathbb{F})$$

is a *Jordan block*.

Theorem 6.13. *Every $A \in \mathcal{M}_n(\mathbb{C})$ is similar to a matrix in Jordan normal form, unique up to reordering the Jordan blocks.*

Proof. (Non-examinable) It is a consequence of a main theorem on modules in IB Groups, Rings and Modules. □

In the rest of this section assume $\mathbb{F} = \mathbb{C}$ unless stated otherwise.

Example.

1. Classification of Jordan normal forms for $\mathcal{M}_2(\mathbb{C})$:

$\begin{pmatrix} \lambda_1 & \\ & \lambda_2 \end{pmatrix}$	$\begin{pmatrix} \lambda & \\ & \lambda \end{pmatrix}$	$\begin{pmatrix} \lambda & 1 \\ & \lambda \end{pmatrix}$
$(t - \lambda_1)(t - \lambda_2)$	$t - \lambda$	$(t - \lambda)^2$

2. Classification of Jordan normal forms for $\mathcal{M}_3(\mathbb{C})$:

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

> **Theorem 6.14** (Generalised Eigenspace Decomposition)**.** *Let V be a finite-dimensional $\mathbb{C}$-vector space and $\alpha \in L(V)$. Suppose that*
>
> $$m_\alpha(t) = (t - \lambda_1)^{c_1} \cdots (t - \lambda_k)^{c_k}$$
>
> *where λ_i's are distinct. Then*
>
> $$V = \bigoplus_j V_j$$
>
> *where*
>
> $$V_j = N(\alpha - \lambda_j \iota)^{c_j}$$
>
> *is the* generalised eigenspace.

Sketch of proof. Let

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

The p_j have no common factor so by Euclidean algorithm we can find $q_1, \dots, q_k \in \mathbb{C}[t]$ such that

$$\sum_j p_j(t) q_j(t) = 1.$$

Let $\pi_j = q_j(\alpha) p_j(\alpha) \in L(V)$. Note $\sum_{j=1}^k \pi_j = \iota$.

1. As $m_\alpha(\alpha) = 0$, $(\alpha - \lambda_j \iota)^{c_j} \pi_j = 0$ so $\operatorname{im} \pi_j \leq V_j$.
2. Suppose $v \in V$, $v = \iota(v) = \sum \pi_j(v)$ so $V = \sum_j V_j$.
3. To show the sum is direct, $\pi_i \pi_j = 0$ for $i \neq j$ so

   $$\pi_i = \pi_i \left(\sum_{j=1}^k \pi_j \right) = \pi_i^2$$

   i.e. π_i is a projection. Then

   $$\pi_i|_{V_j} = \begin{cases} \text{id} & i = j \\ 0 & i \neq j \end{cases}$$

   Directness follows.

□

Remark.

1. We can use Generalised Eigenspace Decomposition to reduce the proof of Theorem 6.13 to a single eigenvalue.
2. Considering $\alpha - \lambda \iota$ can reduce to the case of eigenvalue 0.

Lemma 6.15. *Let $\alpha \in L(V)$ with Jordan normal form $A \in \mathcal{M}_n(\mathbb{C})$. Then the number of Jordan blocks $J_\ell(\lambda)$ of A with $\ell \geq 1$ is*

$$n((\alpha - \lambda\iota)^\ell) - n((\alpha - \lambda\iota)^{\ell-1}).$$

Proof. Work blockwise, for each $s \times s$ block,

$$\begin{array}{c|c|c} \begin{pmatrix} \lambda & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & \lambda \end{pmatrix} & \begin{pmatrix} 0 & 1 & & \\ & & \ddots & \\ & & & 1 \\ & & & 0 \end{pmatrix} & \begin{pmatrix} 0 & 0 & 1 & \\ & & \ddots & 1 \\ & & & 0 \\ & & & 0 \end{pmatrix} \\ \hline J_s(\lambda) & J_s(\lambda) - \lambda I & (J_s(\lambda) - \lambda I)^2 \end{array}$$

so

$$n((J_s(\lambda) - \lambda I)^k) = \begin{cases} k & k \leq s \\ s & k \geq s \end{cases}$$

□

Example. Let

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

we want to find a basis $\mathcal{B} = \{v_1, v_2\}$ with respect to which A is in Jordan normal form.

1. $$\chi_A(t) = \begin{vmatrix} -t & -1 \\ 1 & 2-t \end{vmatrix} = t^2 - 2t + 1 = (t-1)^2$$

   There are two possibilities:

   (a) $m_A(t) = t - 1$. Then the Jordan normal form is

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

   (b) $m_A = (t-1)^2$. Then the Jordan normal form is

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

   A trick here is to note that if A is conjugate to I then $A = I$. Thus (b) holds.

2. The eigenspace is spanned by $v_1 = \binom{1}{-1}$.

3. v_2 satisfies $(A - I)v_2 = v_1$ so

   $$\begin{pmatrix} -1 & -1 \\ 1 & 1 \end{pmatrix} v_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

   so $v_2 = \binom{-1}{0}$. Note that v_2 is not unique.

4. Finally,
$$A = \begin{pmatrix} 1 & -1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & -1 \\ -1 & 0 \end{pmatrix}^{-1}$$

We can use the diagonalisation to calculate powers of matrices:
$$A^n = (P^{-1}JP)^n = P^{-1}J^nP = P^{-1} \begin{pmatrix} 1 & n \\ 0 & 1 \end{pmatrix} P$$

Remark. In Jordan normal form,

- a_λ is the total number of times that λ appears in the diagonal.
- g_λ is the number of λ-Jordan blocks.
- c_λ is the size the largest λ-Jordan block.

7 Bilinear Form II

7.1 Symmetric Bilinear Forms

In this chapter we are going to studying a special bilinear form (and variants whereof) in detail. Let $\varphi : V \times V \to \mathbb{F}$ be a bilinear form on V and assume we take the same basis for both factors of V, say $\mathcal{B}$. Therefore if $\dim V < \infty$, φ has matrix representation

$$[\varphi]_{\mathcal{B}} = [\varphi]_{\mathcal{B},\mathcal{B}}.$$

Recall that

> **Lemma 7.1.** *Let $\varepsilon : V \times V \to \mathbb{F}, \dim V < \infty$ and $\mathcal{B}$ and $\mathcal{B}'$ are two bases for V. Let $P = [\mathrm{id}]_{\mathcal{B}',\mathcal{B}}$. Then*
>
> $$[\varphi]_{\mathcal{B}'} = P^T [\varphi]_{\mathcal{B}} P.$$

Proof. Special case of Proposition 4.6. □

This motivates us to define a relation on $\mathcal{M}_n(\mathbb{F})$

> **Definition** (Congruency)**.** $A, B \in \mathcal{M}_n(\mathbb{F})$ are *congruent* if
>
> $$A = P^T B P$$
>
> for some invertible P.

Note. This is an equivalence relation.

Naturally, we want to find nice forms to which a general bilinear form is congruent. Certainly the nicest form we can have is diagonal matrix. It turns out the property we require a bilinear form to be "diagonalisable" is

> **Definition** (Symmetric)**.** A bilinear form φ on V is *symmetric* if
>
> $$\varphi(u, v) = \varphi(v, u)$$
>
> for all $u, v \in V$.

Note.

- $A \in \mathcal{M}_n(\mathbb{F})$ is symmetric if $A = A^T$. Then φ is symmetric if and only if $[\varphi]_{\mathcal{B}}$ is symmetric for any basis $\mathcal{B}$, if and only if $[\varphi]_{\mathcal{B}}$ is symmetric for one $\mathcal{B}$.
- To be able to be represented by a diagonal matrix, φ needs to be symmetric:
$$[\varphi]_{\mathcal{B}} = P^T A P = D \Rightarrow D^T = D = P^T A^T P \Rightarrow A = A^T$$

> **Definition** (Quadratic form)**.** A map $Q : V \to \mathbb{F}$ is a *quadratic form* if

> there is a bilinear form φ on V such that
>
> $$Q(v) = \varphi(v, v)$$
>
> for all $v \in V$.

Example. Let $V = \mathbb{R}^2$. A general quadratic form is

$$\begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = ax^2 + (b+c)xy + dy^2$$

Remark. A quadratic form does not change under $A \mapsto \frac{1}{2}(A + A^T)$ where A is a representation of the inducing bilinear form.

> **Proposition 7.2.** *Assume* $\operatorname{char} \mathbb{F} \neq 2$. *If $Q : V \to \mathbb{F}$ is a quadratic form then there exists a unique symmetric bilinear form φ on V such that*
>
> $$Q(v) = \varphi(v, v)$$
>
> *for all $v \in V$.*

Proof. First we prove the existence. Let ψ be a bilinear form on V such that $Q(v) = \psi(v, v)$ for all $v \in V$. We construct a symmetric bilinear form by adding ψ and its transpose. Let

$$\varphi(u, v) = \frac{1}{2}(\psi(u, v) + \psi(v, u))$$

(this is where we require $\operatorname{char} \mathbb{F} \neq 2$) then it is bilinear and symmetric and

$$\varphi(v, v) = \psi(v, v) = Q(u).$$

To show the uniqueness, suppose φ is such a symmetric bilinear form. Consider

$$\begin{aligned} Q(u + v) &= \varphi(u + v, u + v) \\ &= \varphi(u, u) + \varphi(u, v) + \varphi(v, u) + \varphi(v, v) \\ &= Q(u) + 2\varphi(u, v) + Q(v) \end{aligned}$$

Rearrange,

$$\varphi(u, v) = \frac{1}{2}(Q(u + v) - Q(u) - Q(v))$$

which is uniquely determined. □

Remark. The last identity

$$\varphi(u, v) = \frac{1}{2}(Q(u + v) - Q(u) - Q(v))$$

is called the *polarisation identity* and will appear later.

The note after the definition of symmetric bilinear form shows that being symmetric is a necessary condition for a bilinear form to be “diagonalisable”. The following theorem says that it is also sufficient:

> **Theorem 7.3.** *Let φ be a symmetric bilinear form on V, an $\mathbb{F}$-vector space and assume* $\operatorname{char}\mathbb{F} \neq 2$ *and* $\dim V < \infty$. *Then there is a basis $\mathcal{B}$ of V such that $[\varphi]_{\mathcal{B}}$ is diagonal.*

Proof. Induction on $n = \dim V$. If $n = 0$ or 1 then obviously true.

Suppose the theorem holds for all spaces of dimension smaller than n. There are two cases to consider:

1. if $\varphi(u, u) = 0$ for all u then by the polarisation identity $\varphi = 0$ so diagonal.
2. otherwise choose $e_1 \in V$ such that $\varphi(e_1, e_1) \neq 0$. Let

$$U = \langle e_1 \rangle^{\perp} = \{u \in V : \varphi(e_1, u) = 0\} = \ker(\varphi(e_1, -) : V \to \mathbb{F})$$

which has dimension $n - 1$ by rank-nullity. Moreover, $V = \langle e_1 \rangle \oplus U$ since $\langle e_1 \rangle \cap U = 0$ and $\dim(\langle e_1 \rangle \oplus U) = n$. Consider $\varphi|_U : U \times U \to \mathbb{F}$ which is also symmetric bilinear. By induction hypothesis there is a basis $e_2, \dots e_n$ of U with respect to which $\varphi|_U$ is diagonal. Now φ is diagonal with respect to $e_1, \dots, e_n$.

□

Notation. In $V = \mathbb{R}^n$ with standard basis $e_1, \dots, e_n$, write

$$Q(x_1, x_2, \dots, x_n) = Q\left(\sum_{i=1}^{n} x_i e_i\right).$$

Example. Let $V = \mathbb{R}^3$ with standard basis e_1, e_2, e_3 and

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

We want a basis f_1, f_2, f_3 of $\mathbb{R}^3$ such that

$$Q(af_1 + bf_2 + cf_3) = \lambda a^2 + \mu b^2 + \nu c^2$$

for some $\lambda, \mu, \nu \in \mathbb{R}$, which are the diagonal entries.

The martix representation of Q with repect to e_1, e_2, e_3 is

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

We could use the algorithm as outlined in the induction proof above but choose to do it differently by completing the square:

$$\begin{aligned} Q(x_1, x_2, x_3) &= \underbrace{(x_1 + x_2 + x_3)^2}_{\text{used all terms in } x_1} + x_3^2 - 2x_2x_3 - 2x_2x_3 \\ &= (x_1 + x_2 + x_3)^2 + \underbrace{(x_3 - 2x_2)^2}_{\text{used all terms in } x_3} - (2x_2)^2 \end{aligned}$$

From here we can read off the diagonal matrix and the basis: for some P,

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

To find P, note that

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

Corollary 7.4. *Let φ be a symmetric bilinear form on V, a finite-dimensional $\mathbb{C}$-vector space. Then there is a basis $\mathcal{B} = \{v_1, \dots v_n\}$ of V such that*

$$[\varphi]_{\mathcal{B}} = \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}$$

where $r = r(\varphi)$.

Proof. Pick basis $\mathcal{E} = \{e_1, \dots e_n\}$ such that

$$[\varphi]_{\mathcal{E}} = \begin{pmatrix} a_1 & & \\ & \ddots & \\ & & a_n \end{pmatrix}$$

Reorder the e_i's such that

$$\begin{cases} a_i \neq 0 & 1 \leq i \leq r \\ a_i = 0 & i > r \end{cases}$$

For $i \leq r$, pick a complex square root of a_i, say $\sqrt{a_i}$. Now let

$$v_i = \begin{cases} \frac{e_i}{\sqrt{a_i}} & 1 \leq i \leq r \\ e_i & i > r \end{cases}$$

□

Corollary 7.5. *Every symmetric matrix $A \in \mathcal{M}_n(\mathbb{C})$ is congruent to a unique matrix of the form $\left(\begin{smallmatrix} I_r & 0 \\ 0 & 0 \end{smallmatrix}\right)$.*

Equivalently,

$$Q\left(\sum_{i=1}^{n} \lambda_i v_i\right) = \sum_{i=1}^{r} \lambda_i^2.$$

We have derived a corollary for our favourite field $\mathbb{C}$, and there is another one corresponding to our second favourite field, $\mathbb{R}$:

Corollary 7.6. *Let φ be a symmetric bilinear form on V, a finite-dimensional $\mathbb{R}$-vector space. There is a basis $\mathcal{B} = \{v_1, \dots, v_n\}$ such that*

$$[\varphi]_{\mathcal{B}} = \begin{pmatrix} I_p & & \\ & -I_q & \\ & & 0 \end{pmatrix}$$

> *where $p, q \geq 0$ and $p + q = r(\varphi)$.*

Proof. The proof is the same as for $\mathbb{C}$ up to the point of exhibiting a basis with respect to which φ is diagonal. Note that we *cannot* choose a square root for all the entries. Instead, reorder the indices such that

$$\begin{cases} a_i > 0 & 1 \leq i \leq p \\ a_i < 0 & p+1 \leq i \leq p+q \\ a_i = 0 & i > p+q \end{cases}$$

and let

$$v_i = \begin{cases} \frac{e_i}{\sqrt{a_i}} & 1 \leq i \leq p \\ \frac{e_i}{\sqrt{-a_i}} & p+1 \leq i \leq p+q \\ e_i & i > p+1 \end{cases}$$

□

Equivalently,

$$Q\left(\sum_{i=1}^{n} \lambda_i v_i\right) = \sum_{i=1}^{p} \lambda_i^2 - \sum_{i=p+1}^{q+p} \lambda_i^2.$$

> **Definition** (Positive/Negative (semi-)definiteness)**.** A symmetric bilinear form φ on a real vector space V is
>
> - *positive definite* if $\varphi(u, u) > 0$ for all $u \in V \setminus \{0\}$.
> - *positive semi-definite* if $\varphi(u, u) \geq 0$ for all $u \in V \setminus \{0\}$.
> - *negative definite* if $\varphi(u, u) < 0$ for all $u \in V \setminus \{0\}$.
> - *negative semi-definite* if $\varphi(u, u) \leq 0$ for all $u \in V \setminus \{0\}$.
> - *indefinite* if none of the above.

The same terminologies apply to quadratic forms.

Example. A bilinear form on $\mathbb{R}^n$ represented by $\left(\begin{smallmatrix} I_p & 0 \\ 0 & 0 \end{smallmatrix}\right) \in \mathcal{M}_n(\mathbb{R})$ is positive definite if $p = n$ and positive semi-definite if $p < n$.

> **Definition** (Signature)**.** The *signature* of a real symmetric bilinear form φ is
>
> $$s(\varphi) = p - q.$$

Again, this applies to quadratic forms as well.

However, we have not even checked whether this is well-defined. Thus we need

> **Theorem 7.7** (Sylvester's Law of Inertia)**.** *If a real symmetric bilinear bilinear form φ has with respect to basis $\mathcal{B}$ and $\mathcal{B}'$*
>
> $$[\varphi]_{\mathcal{B}} = \begin{pmatrix} I_p & & \\ & -I_q & \\ & & 0 \end{pmatrix} \quad [\varphi]_{\mathcal{B}'} = \begin{pmatrix} I_p' & & \\ & -I_q' & \\ & & 0 \end{pmatrix}$$

then

$$\begin{aligned} p &= p', \\ q &= q'. \end{aligned}$$

It is then immediate that

Corollary 7.8. *Signature is well-defined.*

Proof. For uniqueness of p, show that p is the largest dimension of a subspace on which φ is positive definite. This suffices as it is a basis invariant characterisation.

Let $\mathcal{B} = \{v_1, \dots, v_n\}$. Let $X = \langle v_1, \dots, v_p \rangle$ and $Y = \langle v_{p+1}, \dots, v_n \rangle$. $\dim X = p$ and φ is positive definite on X:

$$Q(v) = Q\left(\sum_{i=1}^{p} \lambda_i v_i\right) = \sum_{i=1}^{p} \lambda_i^2 > 0$$

for all $v \neq 0$. Similarly φ is negative semi-definite on Y.

Suppose is φ is positive definite on some other subspace X'. Then $X' \cap Y = 0$ since Q is positive definite on X' and negative semi-definite on Y. Therefore

$$\dim(Y + X') = \dim Y \oplus X' = \dim Y + \dim X' \leq n$$

but since $\dim Y = n - p$ we have $\dim X' \leq p$.

For q, we can either run the same argument with negative definite spaces, or use the fact that $q = r(\varphi) - q$ is invariant. □

The zero diagonal block is not very interesting but it does get a speical name:

Definition (Kernel of symmetric bilinear form)**.** The *kernel* of a symmetric bilinear form is

$$K = \{v \in V : \varphi(u, v) = 0 \text{ for all } u \in V\}.$$

Note.

$$\dim K = n - r(\varphi).$$

In our previous notation, the kernel is simply

$$K = \langle v_{p+q+1}, \dots, v_n \rangle.$$

Caution. There is a subspace T of dimension $n - (p + q) + \min(p, q)$ such that $\varphi|_T = 0$: say $p \geq q$,

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

Exercise. Check that T above is the largest possible such space.

7.2 Sesquilinear Form

Let $\mathbb{F} = \mathbb{C}$ throughout this section.

The dot product on a real vector space comes naturally as a bilinear form. However, its generalisation to complex vector space, the standard inner product defined by

$$\langle x, y\rangle = \sum_{i=1}^{n} x_i \overline{y}_i$$

is not bilinear: the second coordinate transforms by "conjugate-linearity" instead of linearity. Among many other examples of the same spirit, this serves as a motivation to modify the definition of bilinear forms for $\mathbb{C}$-vector spaces:

> **Definition** (Sesquilinear form)**.** Let V and W be $\mathbb{C}$-vector spaces. A *sesquilinear form* is a function $\varphi : V \times W \to \mathbb{C}$ such that
>
> $$\varphi(\lambda_1 v_1 + \lambda_2 v_2, w) = \lambda_1 \varphi(v_1, w) + \lambda_2 \varphi(v_2, w)$$
> $$\varphi(v, \mu_1 w_1 + \mu_2 w_2) = \overline{\mu}_1 \varphi(v, w_1) + \overline{\mu}_2 \varphi(v, w_2)$$
>
> for all $\lambda_1, \mu_1 \in \mathbb{C}$ and $v, v_1, v_2 \in V$, $w, w_1, w_2 \in W$.

Naturally we would expect a sesquilinear form, just like a bilinear form, to have a matrix representation which behaves and transforms accordingly under change-of-basis:

> **Definition** (Matrix of sesquilinear form)**.** Same notation as above. Let $\mathcal{B} = \{v_1, \dots, v_m\}$ be a basis for V and $\mathcal{C} = \{w_1, \dots, w_n\}$ be a basis for W. Then the *matrix* of φ with respect to $\mathcal{B}$ and $\mathcal{C}$ is
>
> $$[\varphi]_{\mathcal{B},\mathcal{C}} = (\varphi(v_i, w_j))_{i,j}$$

> **Lemma 7.9.**
> $$\varphi(u, v) = [u]_{\mathcal{B}}^T [\varphi]_{\mathcal{B},\mathcal{C}} \overline{[v]}_{\mathcal{C}}.$$

Proof. Easy. □

> **Lemma 7.10.** *Let $\mathcal{B}, \mathcal{B}'$ be bases for V, $P = [\mathrm{id}]_{\mathcal{B}',\mathcal{B}}$ and $\mathcal{C}, \mathcal{C}'$ be bases for W, $Q = [\mathrm{id}]_{\mathcal{C}',\mathcal{C}}$. Then*
>
> $$[\varphi]_{\mathcal{B}',\mathcal{C}'} = P^T [\varphi]_{\mathcal{B},\mathcal{C}} \overline{Q}.$$

Proof. Ditto. □

7.3 Hermitian Form

We have special bilinear forms that are symmetric. The analogue for sesquilinear form is

Definition (Hermitian form)**.** A sesquilinear form $\varphi : V \times V \to \mathbb{C}$ is *Hermitian* if

$$\varphi(u,v) = \overline{\varphi(v,u)}.$$

Note.

1. For φ Hermitian, $\varphi(u,u) \in \mathbb{R}$ and $\varphi(\lambda u, \lambda u) = |\lambda|^2 \varphi(u,u)$ so we can still talk about positive/negative (semi-)definite Hermitian forms.
2. For a Hermitian form $\varphi : V \times V \to \mathbb{C}$, let $\mathcal{B}$ be a basis for V. Then we write
$$[\varphi]_{\mathcal{B}} = [\varphi]_{\mathcal{B},\mathcal{B}}.$$

Lemma 7.11. *A sesquilinear form $\varphi : V \times V \to \mathbb{C}$ is Hermitian if and only if for any basis $\mathcal{B}$,*

$$[\varphi]_{\mathcal{B}} = \overline{[\varphi]}_{\mathcal{B}}^T.$$

As before, if and only if this holds for one basis.

Proof. Let $\mathcal{B} = \{v_1, \dots, v_n\}$ and $A = [\varphi]_{\mathcal{B}} = (a_{ij})$.

- $\Rightarrow$:
$$\begin{aligned} a_{ij} &= \varphi(v_i, v_j) \\ &= \overline{\varphi(v_j, v_i)} \\ &= \overline{a_{ji}} \end{aligned}$$
- $\Leftarrow$:
$$\begin{aligned} \varphi\left(\sum \lambda_i v_i, \sum \mu_j v_j\right) &= \lambda^T A \overline{\mu} \\ &= \lambda^T \overline{A}^T \overline{\mu} \\ &= \overline{\mu}^T \overline{A} \lambda \text{ taking transpose of a scalar} \\ &= \overline{\mu^T A \overline{\lambda}} \\ &= \overline{\varphi\left(\sum \mu_j v_j, \sum \lambda_i v_i\right)} \end{aligned}$$

□

Similarly we have polarisation identity for sesquilinear form: a Hermitian form φ on a $\mathbb{C}$-vector space V is determined by

$$\begin{aligned} Q : V &\to \mathbb{R} \\ v &\mapsto \varphi(v,v) \end{aligned}$$

via the formula

$$\varphi(u,v) = \frac{1}{4}\left(Q(u+v) - Q(u-v) + iQ(u+iv) - iQ(u-iv)\right).$$

Proof. Exercise. □

Lastly,

Theorem 7.12 (Diagonalisation of Hermitian Form and Sylvester's Law)**.** *Let V be a finite-dimensional $\mathbb{C}$-vector space and $\varphi : V \times V \to \mathbb{C}$ be a Hermitian form. There is a basis $\mathcal{B} = \{v_1, \dots, v_n\}$ of V with respect to which*

$$[\varphi]_\mathcal{B} = \begin{pmatrix} I_p & & \\ & -I_q & \\ & & 0 \end{pmatrix}$$

where p and q are invariants of φ.

Proof. This is nearly as identical to the symmetric case so we only give a sketch here.

For existence, if $\varphi(u, u) = 0$ for all u then by polarisation identity $\varphi(u, v) = 0$ so done. Assume not. There exists e_1 such that $\varphi(e_1, e_1) \neq 0$. Rescale to have

$$v_1 = \frac{e_1}{\sqrt{|\varphi(e_1, e_1)|}}$$

so $\varphi(v_1, v_1) = \pm 1$. Note that we used the fact that $\varphi(e_1, e_1) \in \mathbb{R}$.

Consider the complementary space

$$W = \langle v_1 \rangle^\perp = \{w \in V : \varphi(v_1, w) = 0\}$$

Check that $V = \langle v_1 \rangle \oplus W$. Now proceed by induction on W.

For uniqueness part (Sylvester's law), note p is the maximal dimension of a subspace of V on which φ is positive definite. □

7.4 Alternating Form

Definition (Alternating form)**.** A bilinear form $\varphi : V \times V \to \mathbb{F}$ is *alternating* or *skew-symmetric* if

$$\varphi(u, v) = -\varphi(v, u)$$

for all $u, v \in V$.

As a consequence $\varphi(u, u) = 0$ for all $u \in V$ and for any basis $\mathcal{B}$, $[\varphi]_\mathcal{B} = -[\varphi]_\mathcal{B}^T$.

Remark. Alternating form is useful since for any $A \in \mathcal{M}_n(\mathbb{F})$ with $\operatorname{char} \mathbb{F} \neq 2$,

$$A = \underbrace{\frac{1}{2}(A + A^T)}_{\text{symmetric}} + \underbrace{\frac{1}{2}(A - A^T)}_{\text{skew-symmetric}} .$$

Theorem 7.13. *If φ is skew-symmetric, there exists a basis*

$$\mathcal{B} = \{v_1, w_1, v_2, w_2, \dots, v_m, w_m, v_{2m+1}, \dots, v_n\}$$

> *such that*
>
> $$[\varphi]_{\mathcal{B}} = \begin{pmatrix} 0 & 1 &&&&&&&& \\ -1 & 0 &&&&&&&& \\ && 0 & 1 &&&&&& \\ && -1 & 0 &&&&&& \\ &&&& \ddots &&&&& \\ &&&&& 0 & 1 &&& \\ &&&&& -1 & 0 &&& \\ &&&&&&& 0 && \\ &&&&&&&& \ddots & \\ &&&&&&&&& 0 \end{pmatrix}$$
>
> *where there are m blocks of* $\left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right)$.

Remark. By reordering the basis, with respect to

$$\{v_1, \dots, v_m, w_1, \dots, w_m, v_{2m+1}, \dots v_n\}$$

it has matrix representation

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

Remark. Skew-symmetric matrices have even rank.

Sketch of proof. Induction on $\dim V$: If $\varphi = 0$ then done. Assume not. Then there exists v_1, w_1 such that $\varphi(v_1, w_1) \neq 0$. In particular v_1 and w_1 are linearly independent. Scale v_1 to get $\varphi(v_1, w_1) = 1 = -\varphi(w_1, v_1)$ and let

$$\begin{aligned} U &= \langle v_1, w_1 \rangle \\ W &= {}^{\perp}U = \{v \in V : \varphi(v, v_1) = \varphi(v, w_1) = 0\} \end{aligned}$$

Check $V = U \oplus W$ by dimension argument. Now apply the induction hypothesis to $\varphi|_W$. □

8 Inner Product Space

8.1 Definitions

Let $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$ in this chapter.

> **Definition** (Inner product)**.** Let V be a vector space over $\mathbb{R}$ ($\mathbb{C}$, repsectively). An *inner product* on V is a positive definite symmetric bilinear form (Hermitian form, respectively) φ on V.
>
> V is called a real (complex, respectively) *inner product space*, or a *Euclidean* (*unitary*, repsectively) space.

Notation. Write $\langle u, v\rangle$ for $\varphi(u, v)$. Note that it is the same as our notation for span so we will spell out span whenever we use it in this chapter.

Example.

- Dot product on $\mathbb{R}^n$ or $\mathbb{C}^n$.
- $V = C([0, 1], \mathbb{C})$, $\langle f, g\rangle = \int_0^1 f(t)\overline{g(t)}dt$.
- This can be generalised. Given $w : [0, 1] \to \mathbb{R}_{>0}$ continuous, think of it as a weight function, we can define an inner product

$$\langle f, g\rangle = \int_0^1 w(t)f(t)\overline{g(t)}dt.$$

Remark. An inner product induces a distance function, i.e. a norm on V by

$$\|v\| = \sqrt{\langle v, v\rangle}$$

whose axioms will be checked later.

Conversely, $\|\cdot\|$ determines the inner product because of the polarisation identity.

> **Lemma 8.1** (Cauchy-Schwarz Inequality)**.**
>
> $$|\langle u, v\rangle| \leq \|u\| \cdot \|v\|$$
>
> *for all $u, v \in V$.*

Proof. Wlog $u \neq 0$. For all $t \in \mathbb{F}$,

$$\begin{aligned} 0 &\leq \|tu - v\|^2 \\ &= \langle tu - v, tu - v\rangle \\ &= \|tu\|^2 - t\langle u, v\rangle - \overline{t\langle u, v\rangle} + \|v\|^2 \end{aligned}$$

by setting $t = \overline{\langle u, v\rangle}/\|u\|^2$,

$$\leq -\frac{|\langle u, v\rangle|^2}{\|u\|^2} + \|v\|^2$$

Rearrange. □

Note. We only used polarisation identity and did not assume any of the norm properties of $\|\cdot\|$, which we will prove now.

Corollary 8.2 (Triangle Inequaility)**.**

$$\|u+v\| \leq \|u\| + \|v\|$$

for all $u, v \in V$.

Proof.

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

□

Corollary 8.3. *$\|\cdot\|$ is a norm.*

Remark. For $\mathbb{F} = \mathbb{R}$, the angle θ between two non-zero vectors u and v satisfies (or defined by, actually)

$$\cos\theta = \frac{\langle u, v\rangle}{\|u\| \cdot \|v\|}.$$

8.2 Orthonomal Basis

Definition (Orthogonality)**.** A set $\{e_1, \dots, e_k\}$ of vectors in V is *orthogonal* if

$$\langle e_i, e_j\rangle = 0$$

for $i \neq j$.

Definition (Orthonormality)**.** A set $\{e_1, \dots, e_k\}$ of vectors in V is *orthonormal* if

$$\langle e_i, e_j\rangle = \delta_{ij}$$

for all i, j.

Lemma 8.4. *If $\{e_1, \dots, e_k\}$ is orthogonal and non-zero then they are linearly independent.*

Moreover if $v = \sum_{j=1}^k \lambda_j e_j$,

$$\lambda_j = \frac{\langle v, e_j\rangle}{\langle e_j, e_j\rangle}.$$

Proof.

$$\langle v, e_j\rangle = \left\langle \sum_{i=1}^k \lambda_i e_i, e_j \right\rangle = \lambda_j \langle e_j, e_j\rangle$$

and the results follow. □

Lemma 8.5 (Parseval's Identity). *Let V be a finite-dimensional inner product space with an orthonormal basis $e_1, \dots, e_n$. Then*

$$\langle u, v\rangle = \sum_{i=1}^{n} \langle u, e_i\rangle \overline{\langle v, e_i\rangle}.$$

Proof. Follows immediately from the orthonormal basis expansion formula in the previous lemma:

$$\langle u, v\rangle = \left\langle \sum_{i=1}^{n} \langle u, e_i\rangle e_i, \sum_{j=1}^{n} \langle v, e_j\rangle e_j \right\rangle.$$

□

Theorem 8.6 (Gram-Schmidt Orthonormalisation Process). *Let V be an inner product space and $\{v_1, v_2, \dots\}$ be a countable set of linearly indpendent vectors in V. Then there exists a sequence $e_1, e_2, \dots$ orthonormal such that*

$$\mathit{span}\{v_1, \dots, v_k\} = \mathit{span}\{e_1, \dots, e_k\}$$

for all k.

Proof. We see the word "countable" and instinctly use induction on k. If $k = 1$ then done. Suppose we have found $e_1, \dots, e_k$. Inspired by the orthonormal basis expansion formula, let

$$e'_{k+1} = v_{k+1} - \underbrace{\sum_{i=1}^{k} \langle v_{k+1}, e_i\rangle e_i}_{\text{linear combination of } v_1, \dots, v_k}$$

which is non-zero by linear independence of $\{v_1, \dots, v_{k+1}\}$. Also $\langle e'_{k+1}, e_i\rangle = 0$ for $1 \le i \le k$ by construction. Finally,

$$\operatorname{span}\{v_1, \dots, v_k, v_{k+1}\} = \operatorname{span}\{e_1, \dots, e_k, e'_{k+1}\}.$$

Finally normalise it by

$$e_{k+1} = \frac{e'_{k+1}}{\|e'_{k+1}\|}.$$

□

Corollary 8.7. *Let V be a finite-dimensional inner product space. Any orthonormal set of vectors can be extended to an orthonormal basis.*

Proof. Say $\{e_1, \dots, e_k\}$ are orthonormal. They are linearly independent so we can extend to a basis $\{e_1, \dots, e_k, v_{k+1}, \dots, v_n\}$ of V.

Now apply Gram-Schmidt to this set. As the first k vectors are already orthonormal it has no effect on them. □

Note. $A \in \mathcal{M}_{m,n}(\mathbb{R})$ has orthonormal columns if $A^T A = I$ and $A \in \mathcal{M}_{m,n}(\mathbb{C})$ has orthonormal columns if $A^T \overline{A} = I$.

We give special names to them:

Definition (Orthogonal matrix)**.** $A \in \mathcal{M}_n(\mathbb{R})$ is *orthogonal* if $A^TA = I$.

Equivalently $A^{-1} = A^T$.

Definition (Unitary matrix)**.** $A \in \mathcal{M}_n(\mathbb{R})$ is *unitary* if $A^T\overline{A} = I$.

Equivalently $A^{-1} = \overline{A}^T$.

Given these terminologies, Gram-Schmidt may be equivalently formulated as follow:

Proposition 8.8. *$A \in \mathcal{M}_n(\mathbb{R})$ ($\mathcal{M}_n(\mathbb{C})$, respectively) non-singular can be written as $A = RT$ where*

- *T is upper triangular,*
- *R is orthogonal (unitary, respectively).*

Proof. Apply Gram-Schmidt to columns of A. The details are left as an exercise. □

8.3 Orthogonal Complements & Projections

Definition (Orthogonal direct sum)**.** Let V be an inner product space and $V_1, V_2 \leq V$. V is the *orthogonal direct sum* of V_1 and V_2 if

1. $V = V_1 \oplus V_2$,
2. $\langle v_1, v_2 \rangle = 0$ for all $v_1 \in V_1, v_2 \in V_2$.

Write $V = V_1 \perp V_2$.

Note. The first condition is actually redundant: $V_1 \cap V_2 = 0$ because of positive definiteness of inner product.

Definition (Orthogonal complement)**.** Let $W \leq V$. The *orthogonal complement* of W in V is

$$W^\perp = \{v \in V : \langle v, w \rangle = 0 \text{ for all } v \in W\}.$$

Lemma 8.9. *Let V be a finite-dimensional inner product space and $W \leq V$. Then*

$$V = W \perp W^\perp.$$

Proof. If $w \in W, u \in W^\perp$ then $\langle w, u \rangle = 0$ so it remains to show that $V = W + W^\perp$.

Let $\{e_1, \dots, e_k\}$ be an orthonormal basis for W. By a previous lemma we can extend it to an orthonormal basis $\{e_1, \dots, e_n\}$ of V. Note that $e_{k+1}, \dots, e_n \in W^\perp$. □

Note. Complementary space is, in general, not unique but orthogonal complement is.

A concept closely related to orthogonal complement is

Definition (Projection)**.** Suppose $V = U \oplus W$, a *projection* from V to W is a map

$$\begin{aligned} \pi : V &\to W \\ u + w &\mapsto w \end{aligned}$$

where $u \in U, w \in W$. This a well-defined linear map and is idempotent, i.e. $\pi^2 = \pi$.

If the direct sum is orthogonal, however, there is a unique projection:

Definition (Orthogonal projection)**.** If $U = W^\perp$ above then π is the *orthgonal projection* from V to W.

Note. $\pi' = \iota - \pi$ is the orthogonal projection from V to $W^\perp$.

So far we have discussed orthogonal complement only at an abstract level and we don't know yet how to find one, although you should have a good intuition of how. The following lemma tells us how it works:

Lemma 8.10. *Let V be an inner product space, $W \leq V$ with orthonormal basis $e_1, \dots, e_k$ and π is the orthogonal projection onto W. Then*

1. *For all $v \in V$,*
$$\pi(v) = \sum_{i=1}^k \langle v, e_i \rangle e_i$$
2. *$\|v - \pi(v)\| \leq \|v - w\|$ for all $v \in V$, $w \in W$ with equality if and only if $\pi(v) = w$. Equivalently, $\pi(v)$ is the closest point in W to v.*

Proof.

1. We need
$$v - \sum_{i=1}^k \langle v, e_1 \rangle e_i \in W^\perp.$$
But
$$\left\langle v - \sum_{i=1}^k \langle v, e_i \rangle e_i, e_j \right\rangle = \langle v, e_j \rangle - \langle v, e_j \rangle = 0.$$
2.
$$\begin{aligned} \|v - w\|^2 &= \|\underbrace{v - \pi(v)}_{\in W^\perp} + \underbrace{\pi(v) - w}_{\in W}\|^2 \\ &= \|v - \pi(v)\|^2 + \|\pi(v) - w\|^2 \\ &\leq \|v - \pi(v)\|^2 \end{aligned}$$
with equality if and only if $\pi(v) = w$.

□

Remark. We met internal and external direct sum before. There is an analogous distinction for orthogonal direct sum.

Given V_1, V_2 two inner product spaces over $\mathbb{F}$, we can define the *external orthogonal* direct sum $V_1 \perp V_2$ by equipping $V_1 \oplus V_2$ the inner product

$$\langle (u_1, u_2), (v_1, v_2) \rangle = \langle u_1, v_1 \rangle + \langle u_2, v_2 \rangle.$$

In practice we often suppress the distinction between internal and external (orthogonal) direct sums.

8.4 Adjoints

Proposition 8.11. *Let V and W be finite-dimensional inner product spaces and $\alpha \in L(V, W)$. Then there exists a unique linear map $\alpha^* : W \to V$ such that for all $v \in V, w \in W$,*

$$\langle \alpha v, w \rangle = \langle v, \alpha^* w \rangle.$$

If $\mathcal{B}$ is an orthonormal basis for V and $\mathcal{C}$ is an orthonormal basis for W,

$$[\alpha^*]_{\mathcal{C},\mathcal{B}} = \overline{[\alpha]}^T_{\mathcal{B},\mathcal{C}}.$$

Definition (Adjoint)**.** α^* as above is the *adjoint* of α.

Proof. Let $\mathcal{B} = \{v_1, \dots, v_n\}, \mathcal{C} = \{w_1, \dots, w_m\}$ and as usual, let

$$\begin{aligned} A &= [\alpha]_{\mathcal{B},\mathcal{C}} = (a_{ij}) \\ \overline{A}^T &= C = (c_{ij}) \end{aligned}$$

where $c_{ij} = \overline{a_{ij}}$.

Since the formula for the adjoint map is given, we might as well just verify it. Consider the linear map β such that $[\beta]_{\mathcal{C},\mathcal{B}} = C$. Then

$$\left\langle \alpha \left(\sum_i \lambda_i v_i \right), \sum_j \mu_j w_j \right\rangle = \left\langle \sum_{i,k} \lambda_i a_{ki} w_k, \sum_j \mu_j w_j \right\rangle = \sum_{i,j} \lambda_i a_{ji} \overline{\mu_j}$$

while on the other hand

$$\left\langle \sum_i \lambda_i v_i, \beta \left(\sum_j \mu_j w_j \right) \right\rangle = \left\langle \sum_i \lambda_i v_i, \sum_{j,k} \mu_j c_{kj} v_k \right\rangle = \sum_{i,j} \lambda_i \overline{c_{ij} \mu_j}$$

and we see that they are equal since $c_{ij} = \overline{a_{ij}}$. Thus we have proved the existence of adjoint.

By specialising the above calculation to basis elements uniqueness follows. □

Notation. Denote the Hermitian conjugate by

$$A^\dagger = \overline{A}^T.$$

Caution. We use the same notation α^* for the adjoint and the dual of α. Hopefully the context should be clear which one is in use.

Remark. This usage of notation is not entirely coincidental. In fact, let V and W be finite-dimensional real inner product spaces and $\alpha \in L(V, W)$. In finite dimension there is an isomorphism

$$\begin{aligned} \psi_{R,V} : V &\to V^* \\ v &\mapsto \langle -, v \rangle \end{aligned}$$

and similarly $\psi_{R,W} : W \to W^*$ such that the following diagram commutes:

$$\begin{tikzcd} W \arrow[r, "\text{adjoint of } \alpha"] \arrow[d, "\psi_{R,W}"'] & V \arrow[d, "\psi_{R,V}"] \\ W^* \arrow[r, "\text{dual of } \alpha"'] & V^* \end{tikzcd}$$

In the fancy language of category theory, this essentially says that adjoint and dual are naturally isomorphic as contravariant functors, with components $\psi_{R,-}$.

8.5 Self-adjoint Maps & Isomoetries

Let $V = W$ throughout this section.

Definition (Self-adjoint)**.** Let V be an inner product space and $\alpha \in L(V)$. Let α^* be the adjoint of α. α is *self-adjoint* if it satisfies one of the equivalent properties below:

- For all $u, v \in V$, $\langle \alpha u, v \rangle = \langle u, \alpha v \rangle$,
- $\alpha = \alpha^*$.

α is said to be symmetric (Hermitian, respectively) if the vector space is real (complex, respectively).

Definition (Isometry)**.** Let V be an inner product space and $\alpha \in L(V)$. Let α^* be the adjoint of α. α is an *isometry* if it satisfies one of the equivalent properties below:

- For all $u, v \in V$, $\langle \alpha u, \alpha v \rangle = \langle u, v \rangle$,
- $\alpha^{-1} = \alpha^*$.

α is said to be orthogonal (unitary, respectively) if the vector space is real (complex, respectively).

The equivalences should be quite obvious and in case you don't find it so,

Proof of 2nd equivalence.

- $\Rightarrow$: $\|\alpha v\|^2 = \|v\|^2$ so α is injective and α^{-1} exists. For all $u, v \in V$,

$$\langle u, \alpha^* v \rangle = \langle \alpha u, v \rangle = \langle u, \alpha^{-1} v \rangle$$

so $\alpha^{-1} = \alpha^*$.

- $\Leftarrow$:
$$\langle \alpha u, \alpha v\rangle = \langle u, \alpha^* \alpha v\rangle = \langle u, v\rangle$$
for all $u, v \in V$.

□

Remark. By the polarisation identity there is yet another equivalent definition of isometry: α is an isometry if

$$\|\alpha v\| = \|v\|$$

for all $v \in V$, which might be closer to the intuition of an "isometry".

Lemma 8.12. *Let V be a finite-dimensional real (complex, respectively) inner product space and $\alpha \in L(V)$. Then*

- *α is self-adjoint if and only if for all* orthonormal basis *$\mathcal{B}$, $[\alpha]_{\mathcal{B}}$ is symmetric (Hermitian, repsectively).*
- *α is an isometry if and only if for all* orthonormal basis *$\mathcal{B}$, $[\alpha]_{\mathcal{B}}$ is orthogonal (unitary, respectively).*

Proof. There is very little to do actually. For any orthonormal basis $\mathcal{B}$,

$$[\alpha^*]_{\mathcal{B}} = \overline{[\alpha]}_{\mathcal{B}}^T$$

and the two cases follow. □

It turns out all the isometries on an inner product space form a group:

Definition (Orthogonal/Unitary group)**.**

- If $\mathbb{F} = \mathbb{R}$, the *orthogonal group* of V is
$$O(V) = \{\alpha \in L(V) : \alpha \text{ isometry}\}.$$
- If $\mathbb{F} = \mathbb{C}$, the *unitary group* of V is
$$U(V) = \{\alpha \in L(V) : \alpha \text{ isometry}\}.$$

Lemma 8.13. *Let V be an inner product space with orthonormal basis $e_1, \dots, e_n$. Then*

- *if $\mathbb{F} = \mathbb{R}$, there is a correspondence*
$$\begin{aligned} O(V) &\leftrightarrow \{\textit{orthonormal basis of } V\} \\ \alpha &\leftrightarrow (\alpha(e_1), \dots, \alpha(e_n)) \end{aligned}$$

> - *if $\mathbb{F} = \mathbb{C}$, there is a correspondence*
>
> $$\begin{aligned} U(V) &\leftrightarrow \{\textit{orthonormal basis of } V\} \\ \alpha &\leftrightarrow (\alpha(e_1), \dots, \alpha(e_n)) \end{aligned}$$

8.5.1 Spectral Theory for Self-adjoint Maps

Spectral theory is the study of eigenvalues and eigenvectors of linear operators, particularly those on infinite dimensional spaces. They have enormous importance in many areas of mathematics and physics, including for example functional analysis, harmonic analysis and quantum mechanics. In this course, spectral simply refers the the collection of all eigenvalues of an endomorphism on a finite-dimensional vector space.

> **Lemma 8.14.** *Let V be an inner product space. If $\alpha \in L(V)$ is self-adjoint then*
>
> *1. α has real eigenvalues.*
>
> *2. eigenvectors of α for different eigenvalues are orthogonal.*

Note that this true for any inner product space, regardless of dimension.

Proof.

1. Suppose $\alpha v = \lambda v$ for some non-zero $v \in V$ and $\lambda \in \mathbb{C}$. Then

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

so $\lambda = \overline{\lambda} \in \mathbb{R}$.

2. Suppose $\alpha v = \lambda v, \alpha w = \mu w$ where $\lambda \neq \mu \in \mathbb{R}$. Use the similar idea,

$$\lambda\langle v, w\rangle = \langle \lambda v, w\rangle = \langle \alpha v, w\rangle = \langle v, \alpha w\rangle = \mu\langle v, w\rangle$$

so $\langle v, w\rangle = 0$.

□

For infinite dimensional space, we may not have any eigenvalues (although in which case the above is vacuously true). However, in finite-dimesional case we have

> **Theorem 8.15.** *Let V be a finite-dimensional inner product space and $\alpha \in L(V)$ is self-adjoint. Then V has an orthonormal basis of eigenvectors, whose eigenvalues are real by the previous lemma.*

Proof. Let $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$. Induction on $n = \dim V$. If $n = 0$ this is vacuously true. If $n = 1$ then also true by the previous lemma. Suppose $n > 1$. Say

$$[\alpha]_{\mathcal{B}} = A$$

where $\mathcal{B}$ is the standard basis. Passing to $\mathbb{C}$ in the same way we did in the proof of Lemma 6.2. By Fundamental Theorem of Algebra, $\chi_A(t) \in \mathbb{C}[t]$ has a root so $A \in \mathcal{M}_n(\mathbb{C})$ has an eigenvalue. Note that the argument so far applies to all maps, not just self-adjoint operators.

Now using self-adjointness, this eigenvalue is real. So $\chi_A(t)$ has a (real) root. Thus for both fields α has a real eigenvalue, say λ. Pick $v_1 \in V \setminus \{0\}$ such that $\alpha v_1 = \lambda v_1$. Now use the old trick of passing to a space of strictly smaller dimension, in this case the orthogonal complement

$$U = \langle v_1 \rangle^\perp \leq V$$

where $\langle v_1 \rangle$ is the subspace spanned by v_1. Check conditions for induction: if $u \in V$,

$$\langle \alpha u, v_1 \rangle = \langle u, \alpha v_1 \rangle = \langle u, \lambda v_1 \rangle = \lambda \langle u, v_1 \rangle = 0$$

so α is U-stable. $\alpha|_U \in L(U)$ is obviously self-adjoint so by induction hypothesis there is an orthonormal basis $v_2, \dots, v_n$ of U which are eigenvectors for $\alpha|_U$. Adjoining $\frac{v_1}{\|v_1\|}$ gives an orthonormal basis of eigenvectors of α. □

> **Corollary 8.16.** *Let V be a finite-dimensional inner product space. If $\alpha \in L(V)$ is self-adjoint, V is the orthogonal direct sum of all the eigenspaces of α.*

Proof. Immediate. □

One reason self-adjoint operators are important is that many physical systems can be described by self-adjoint operators. By the theorem we can decompose such a space into orthogonal direct sum of eigenspaces, and when in orthonormal basis, the action of a self-adjoint operator is simply "scaling".

8.5.2 Spectral Theory for Unitary Maps

The other important map is isometry. Let $\mathbb{F} = \mathbb{C}$ throughout this subsection.

> **Lemma 8.17.** *Let V be a complex inner product space and $\alpha \in L(V)$ unitary. Then*
>
> 1. *all eigenvalues lie on the unit circle.*
> 2. *eigenvectors corresponding to different eigenvalues are orthogonal.*

Proof. This involves similar ideas as the lemma in the last subsection.

1. Suppose $\alpha v = \lambda v$ for non-zero $v \in V$. In addition $\lambda \neq 0$ as α is invertible. Then

$$\lambda \langle v, v \rangle = \langle \lambda v, v \rangle = \langle \alpha v, v \rangle = \langle v, \alpha^{-1} v \rangle = \langle v, \lambda^{-1} v \rangle = \overline{\lambda}^{-1} \langle v, v \rangle$$

so $\lambda = \overline{\lambda}^{-1}$, $|\lambda|^2 = 1$.

2. Suppose $\alpha v = \lambda v, \alpha w = \mu w$ where $\lambda \neq \mu$. Then

$$\lambda \langle v, w \rangle = \langle \alpha v, w \rangle = \langle v, \alpha^{-1} w \rangle = \overline{\mu}^{-1} \langle v, w \rangle = \mu \langle v, w \rangle$$

so $\langle v, w \rangle = 0$.

□

> **Theorem 8.18.** *Let V be a finite-dimensional complex inner product space and $\alpha \in L(V)$. Then V has an orthonormal basis of eigenvectors.*

Proof. By Fundamental Theorem of Algebra, α has an eigenvalue, say $\lambda \in \mathbb{C}$. Fix non-zero $v_1 \in V$ such that $\alpha v_1 = \lambda v_1$ and further assume $\|v_1\| = 1$. Let

$$U = \langle v_1 \rangle^\perp \leq V.$$

For all $u \in U$,

$$\langle \alpha u, v_1 \rangle = \langle u, \alpha^{-1} v_1 \rangle = \overline{\lambda}^{-1} \langle u, v_1 \rangle = 0$$

so α is U-stable. By induction on dimension, U has an orthonormal basis of eigenvectors of $\alpha|_U$, say $v_2, \dots, v_n$. Thus $v_1, \dots, v_n$ is an orthonormal basis of eigenvectors of α. □

Remark.

1. Self-adjoint operators and isometries have different physical properties. However spectral theory says that they have similar properties and the proofs of which are essentially identical. This is because they are both examples of a more general type of operators called *normal* maps, which are defined to be those satisfying

   $$\alpha \alpha^* = \alpha^* \alpha.$$

   Other examples of normal maps include skew-Hermitian maps. We will meet more in example sheet.

2. Note that unlike the previous subsection, we only discuss unitary operators (i.e. complex isometries). An orthogonal matrix $A \in \mathcal{M}_n(\mathbb{R})$ cannot, in general, be diagonalised over $\mathbb{R}$. For example, rotation of $\mathbb{R}^2$. See example on page 44.

   However, we can still get orthonormal basis with respect to which it is *block-diagonal* with each blocks of size 1 or 2. See also example sheet. In a sense, this is the "worst" can happen to an isometry.

8.5.3 Application to Bilinear Forms

Recall spectral theorem for self-adjoint maps and isometries.

> **Corollary 8.19.** *Let $A \in \mathcal{M}_n(\mathbb{R})$ ($\mathcal{M}_n(\mathbb{C})$ respectively) be a symmetric (Hermitian respectively) matrix. Then there is an orthogonal (unitary respectively) matrix P such that $P^\dagger AP$ is diagonal with real entries.*

Proof. Let $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$ and equip $\mathbb{F}^n$ with the standard inner product. Then $A \in L(\mathbb{F}^n)$ is self-adjoint. Thus there is an orthonormal basis of $\mathbb{F}^n$ of eigenvectors of A (with real eigenvalues), say $v_1, \dots, v_n$. Let

$$P = (v_1 \mid \cdots \mid v_n)$$

then $P^{-1}AP$ is diagonal with real entries. Now use $P^{-1} = P^\dagger$. □

Note. For an orthogonal change-of-basis matrix P,

$$P^{-1}AP = P^\dagger AP$$

where LHS is change-of-basis of A as a linear map while RHS is change-of-basis of A as a bilinear (sesquilinear respectively) form. This interpretation can be exploited to tell us more about the structure of it.

> **Corollary 8.20.** *Let V be a finite-dimensional real (complex respectively) inner product space and $\varphi : V \times V \to \mathbb{F}$ be a symmetric bilinear (Hermitian respectively) form. Then there is an orthonormal basis of V with respect to which φ is represented by a diagonal matrix with real entries.*

Recall that previously we have shown that in general a symmetric bilinear (Hermitian respectively) form is diagonalisable using perpendicular space. However, if we equip the same space with an inner product, this corollary not only tells us that the bilinear form is diagonalisable, but also gives us an orthonormal basis with respect to which this holds.

Proof. Let $\mathcal{B} = \{v_1, \dots, v_n\}$ be any orthonormal basis and $A = [\varphi]_\mathcal{B}$. $A = A^\dagger$ and there is an orthogonal (unitary respectively) matrix P such that $D = P^\dagger AP$ is diagonal. Let w_i be the ith column of P, then $\mathcal{B}' = \{w_1, \dots, w_n\}$ is an orthonormal basis of V and $[\varphi]_{\mathcal{B}'} = D$. □

Remark. The diagonal entries of $P^\dagger AP$ are the eigenvalues of A and thus the signature could be equivalently defined as

$$s(\varphi) = \#\text{positive eigenvalues of } A - \#\text{negative eigenvalues of } A.$$

> **Corollary 8.21** (Simultaneous diagonalisation of bilinear forms)**.** *Let V be a finite-dimensional real (complex respectively) vector space. Let φ, ψ be symmetric (Hermitian respectively) bilinear forms on V. Assume φ is positive definite. There is a basis $\mathcal{B} = \{v_1, \dots, v_n\}$ of V such that $[\varphi]_\mathcal{B}$ and $[\psi]_\mathcal{B}$ are both diagonal.*

Proof. First note that V equipped with φ is an inner product space. Thus there exists an orthonormal (with respect to this inner product) basis with respect to which ψ is represented by a diagonal matrix. By definition φ is represented by the identity matrix, which is unchanged under any change of basis. The result follows. □

Caution. The positive definite assumption is necessary. See example sheet for counterexamples when this assumption is not satisfied.

And we have a version correpsonding to matrices:

> **Corollary 8.22.** *Let $A, B \in \mathcal{M}_n(\mathbb{R})$ ($\mathcal{M}_n(\mathbb{C})$ respectively) be symmetric (Hermitian respectively). Suppose $\overline{x}^T Ax > 0$ for all non-zero x. Then there exists $Q \in \mathcal{M}_n(\mathbb{R})$ ($\mathcal{M}_n(\mathbb{C})$ respectively) invertible such that $Q^T AQ$ and $Q^T BQ$ ($Q^T A\overline{Q}$ and $Q^T B\overline{Q}$ respectively) are both diagonal.*

Proof. Easy. □

Index