Linear Algebra *

Zhiyuan Bai

Compiled on March 12, 2023

This document serves as a set of revision materials for the Cambridge Mathematical Tripos Part IB course *Linear Algebra* in Michaelmas 2020. However, despite its primary focus, readers should note that it is NOT a verbatim recall of the lectures, since the author might have made further amendments in the content. Therefore, there should always be provisions for errors and typos while this material is being used.

Contents

*Based on the lectures under the same name taught by Prof. P. Raphael in Michaelmas 2020.

1 Vector Spaces and Subspaces

Let F be an arbitrary field.

Definition 1.1. An F-vector space (or a vector space over F) is an abelian group $(V,+)$ equipped with a function $F \times V \to V, (\lambda, v) \mapsto \lambda v$ such that for any $v, v_1, v_2 \in V, \lambda, \mu \in F$:
1. $\lambda(v_1 + v_2) = \lambda v_1 + \lambda v_2$.
2. $(\lambda_1 + \lambda_2)v = \lambda_1 v + \lambda_2 v$.
3. $\lambda(\mu v) = (\lambda \mu v)$.
4. $1v = v$.
This function is often called scalar multiplication of the vector space.

Example 1.1. 1. Take $n \in \mathbb{N}$. Then the set of n-tuples in F, denoted F^n, is a vector space under the operations

$$(x_1, \ldots, x_n) + (y_1, \ldots, y_n) = (x_1 + y_1, \ldots, x_n + y_n)$$
$$\lambda(x_1, \ldots, x_n) = (\lambda x_1, \ldots, \lambda x_n)$$

2. For any set X, write $\mathbb{R}^X = \{f : X \to \mathbb{R}\}$. It is a vector space over $\mathbb{R}$ via

$$(f_1 + f_2)(x) = f_1(x) + f_2(x), (\lambda f)(x) = \lambda f(x)$$

3. The set $M_{n,m}(F)$ consisting of F-valued $n \times m$ matrices is a vector space by interpreting it as $F^{n \times m}$.

Remark. The axioms of scalar multiplication imply that $0v = 0$ for any $v \in V$, as one can check.

Definition 1.2 (Subspace)**.** Let V be a vector space over F. A subset $U \subset V$ is a subspace of V (or $U \leq V$ as vector spaces) iff $U \leq V$ as subgroups and $\forall \lambda \in F, u \in U, \lambda u \in U$.

So basically, a subgroup U is a subspace if we can properly restrict the original scalar product to make it a vector space over F as well. One can also check oneself that a subspace of a subspace is also a subspace of the original space.

Example 1.2. Take $V = \mathbb{R}^{\mathbb{R}}$. $C(\mathbb{R}) \leq V$ (the set of continuous functions $\mathbb{R} \to \mathbb{R}$) is a subspace of V, and the set of polynomials is a subspace of $C(\mathbb{R})$. Take $V = \mathbb{R}^3$, then a line is a subspace iff it passes through the origin. A plane is a subspace iff it passes through the origin as well.

Proposition 1.1. *Let V be an F-vector space and $U, W \leq V$ as vector spaces, then $U \cap W \leq V$ as vector spaces.*

Proof. Just check. □

However, the union of two subspaces is generally not a subspace unless one is contained in the other already.

Example 1.3. Take $V = \mathbb{R}^2$ and U, W two axes, then $(1,0) + (0,1) = (1,1)$ is already not in the union.

Definition 1.3. Let V be an F-vector space. Let $U, W \leq V$. The sum of U and W is the set $U + W = \{u + w : u \in U, w \in W\}$.

Example 1.4. Take $V = \mathbb{R}^2$ and U, W two axes again, then $U + W = V$.

Proposition 1.2. *The sum of two subspaces is a subspace.*

Proof. Obvious. □

Just to mention, one can easily check that $U + W$ is the smallest subspace of V containing U and W.

Definition 1.4. Let V be an F-vector space and $U \leq V$. The quotient space V/U is the quotient group equipped with the scalar multiplication $\lambda(v + U) = \lambda v + U$.

Proposition 1.3. *This scalar multiplication is well-defined and makes V/U a vector space.*

Proof. Just check. □

2 Spans, Linear Independence and Steinitz Exchange Lemma

In this section, we shall characterise the properties of the dimension and basis of a vector space.

Definition 2.1 (Span of a Family of Vectors)**.** Let V be a vector space over F and $S \subset V$. We define the span of S to be

$$\langle S \rangle = \operatorname{span}(S) = \left\{ \sum_{i=1}^{n} \lambda_i s_i : n \in \mathbb{N}, \lambda_i \in F, s_i \in S \right\}$$

That is, $\langle S \rangle$ consists of all possible (finite) linear combination of elements of S. By convention, we say $\langle \varnothing \rangle = \{0\}$. Note also that the span of S is essentially the minimal subspace of V containing S.

Example 2.1. 1. Take $V = \mathbb{R}^3$ and

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

then

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

2. Take $V = \mathbb{R}^n$ and let e_i be the vector in V that only has 1 at the i^{th} entry and zero elsewhere, then $\langle \{e_i\}_{i=1}^n \rangle = V$.
3. Let $V = \mathbb{R}^X$ and $S_x : X \to \mathbb{R}$ be such that $S_x(y) = 1_{x=y}$. Then $\langle \{S_x\}_{x \in \mathbb{R}} \rangle$ are the set of functions $f \in \mathbb{R}^X$ that has finite support.

Definition 2.2. Let V be a vector space over F and $S \subset V$. We say S spans V if $\langle S \rangle = V$.

Example 2.2. Take $V = \mathbb{R}^2$, then any set of two non-parallel vectors would span V.

Definition 2.3. A vector space V over a field F is finite dimensional if there is a finite $S \subset V$ that spans V.

Example 2.3. The space $V = \mathbb{P}[x]$ be the set of polynomials in $\mathbb{R}$ and $V_n = \mathbb{P}_n[x]$ be the set of real polynomials with degree at most n. Then $V_n = \langle \{1, x, \dots, x^n\} \rangle$ is finite dimensional, but V is not finite dimensional as any finite set of polynomials must be contained in V_n where n is the maximal degree of polynomials in that set.

As $\mathbb{N}$ is well-ordered, there must be a minimum number of vectors that can possibly span V. We then focus on how to capture this minimality.

Definition 2.4 ((Linear) Independence)**.** Let V be a vector space over F. We say $\{v_1, \dots, v_n\} \subset V$ are (linearly) independent (or is a free family) if for any $\lambda_1, \dots, \lambda_n \in F$

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

On the other hand, this set is not linearly independent if there exists $\lambda_1, \dots, \lambda_n \in F$ not all zero such that $\sum_{i=1}^{n} \lambda_i v_i = 0$.

Example 2.4. Let $V = \mathbb{R}^3$ and

$$v_1 = (1, 0, 0)^\top, v_2 = (0, 1, 0)^\top, v_3 = (1, 1, 0)^\top, v_4 = (0, 1, 1)^\top$$

Then $\{v_1, v_2\}$ is linearly independent. Note that $v_3 \in \langle \{v_1, v_2\} \rangle$, so $\{v_1, v_2, v_3\}$ is not linearly independent. On the other hand, $v_4 \notin \langle \{v_1, v_2\} \rangle$, which as one can verify means that $\{v_1, v_2, v_4\}$ is linearly independent.

Remark. If the family $\{v_i\}_{1 \le i \le n}$ is linearly independent, then none of v_i is zero.

Definition 2.5 (Basis)**.** A subset $S \subset V$ is a basis if it is linearly independent and $\langle S \rangle = V$.

Remark. When S spans V, we say that S is a generating family of V. So a basis is just a linearly independent generating family.

Example 2.5. 1. Take $V = \mathbb{R}^n$, then the family $\{e_i\}_{1\leq i\leq n}$ where e_i is the vector having 1 at i^{th} entry and zero otherwise is a basis.
2. Take $V = \mathbb{C}$ over $\mathbb{C}$, then $\{a\}$ is a basis for any $a \neq 0$.
3. Take also $V = \mathbb{C}$ but over $\mathbb{R}$, then $\{1, i\}$ is a basis.
4. Take $V = \mathbb{P}[x]$ be the set of polynomials in $\mathbb{R}$ and $S = \{x^n : n \geq 0\}$. Then S is a basis. Worth noting that $|S| = \infty$ in this case.

Lemma 2.1. *If V is a vector space over F, then $\{v_1, \dots, v_n\}$ is a basis of V if and only if for any vector $v \in V$, there is a unique decomposition*

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

Remark. If the conditions are true, then the tuple $(\lambda_1, \dots, \lambda_n)$ (ordered via the ordering one chose on v_i) is called the coordinate of v in the basis (v_i).

Proof. Trivial. □

Lemma 2.2. *If S is a finite set that spans V, then a subset of S is a basis of V.*

Proof. If S is independent, then we are done. Otherwise, there is some $\lambda \neq 0$ and λ_w such that there is $v \in S$ with

$$\lambda v + \sum_{w\in S\setminus\{v\}} \lambda_w w = 0 \implies v = \frac{1}{\lambda} \sum_{w\in S\setminus\{v\}} \lambda_w w \in \langle S \setminus \{v\}\rangle$$

Therefore $S \setminus \{v\}$ also spans V. We can repeat this process and, by the well-ordering of $\mathbb{N}$, will reach a basis. □

Theorem 2.3 (Steinitz Exchange Lemma)**.** *Let V be a finite dimensional vector space over F, $\{v_1, \dots, v_m\} \subset V$ linearly independent, $\{w_1, \dots, w_n\} \subset V$ a generating set, then:*
1. $m \leq n$.
2. Up to relabeling, $\{v_1, \dots, v_m, w_{m+1}, \dots, w_n\}$ spans V.

Proof. Suppose $\{v_1, \dots, v_l, w_{l+1}, \dots, w_n\}$ spans V for some $l < m$, then

$$\exists \alpha_i, \beta_i \in F, v_{l+1} = \sum_{i\leq l} \alpha_i v_i + \sum_{i>l} \beta_i w_i$$

But $\{v_i\}$ is linearly independent, so one of the β_i is nonzero. By relabelling $\beta_{l+1} \neq 0$, then $w_{l+1} \in \langle\{v_1, \dots, v_l, v_{l+1}, w_{l+2} \dots, w_n\}\rangle$, therefore the set of vectors $\{v_1, \dots, v_l, v_{l+1}, w_{l+2} \dots, w_n\}$ also spans V. The theorem is then obvious by induction. □

Corollary 2.4. *Let V be a finite dimensional vector space, then any two bases of V have the same cardinality.*

Proof. Immediate. □

This corollary allows us to give a proper definition of the dimension of a vector space. Before we step right into that, another corollary of Theorem 2.3 can help us to capture important properties of a finite dimensional vector space that will come in handy in further discussions of basis.

Corollary 2.5. *Let V be a vector space with $\dim V = n$, then:*
1. Any independent set of vectors has size at most n. The size is exactly n iff this set is a basis.
2. Any spanning set has size at least n. The size is exactly n iff this set is a basis.

Proof. Obvious. □

3 Basis, Dimension and Direct Sum

Definition 3.1. This cardinality is called the dimension $\dim V$ of V.

This is well-defined due to Corollary 2.4.

Proposition 3.1. *Let U, W be subspaces of V. If they are finite dimensional, then so is $U + W$ and*

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

Proof. Pick a basis $v_1, \ldots, v_l$ of $U \cap W$. Extend it to a basis $v_1, \ldots, v_l, u_1, \ldots, u_m$ of U and a basis $v_1, \ldots, v_l, w_1, \ldots, w_n$ of W, then $\{v_i\} \cup \{u_i\} \cup \{w_i\}$ is easily a basis of $U + W$. The equality follows. □

Proposition 3.2. *If V is a finite dimensional vector space and $U \leq V$, then $U, V/U$ are both finite dimensional and $\dim V = \dim U + \dim V/U$. Furthermore,*

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

Proof. It is obvious that U is finite dimensional. Choose a basis $u_1, \ldots, u_l$ and extend it to a basis $u_1, \ldots, u_l, w_{l+1}, \ldots, w_n$ of V, then $w_{l+1} + U, \ldots, w_n + U$ can be verified to be a basis of V/U. The statement is immediate. □

Remark. If U is a proper subspace of V, written $U < V$ (meaning that $U \leq V$ and $U \neq V$), then the proposition gives us $\dim V/U \neq 0$, so $\dim U < \dim V$.

Definition 3.2. Let V be a vector space and $U, W \leq V$. We say V is the direct sum of U, W, written $V = U \oplus W$, if every element $v \in V$ can be written uniquely as $v = u + w$ for $u \in U, w \in W$.
If this happens, then we say W is a direct complement of U in V.

Note that direct complement is not unique in general.

Example 3.1. Take $U = \mathbb{R} \times \{0\}$, then both $W = \{0\} \times \mathbb{R}$ and $W' = \langle \{(1, 1)^\top\} \rangle$ are direct complements of U.

Lemma 3.3. *Let $U, W \leq V$, then the followings are equivalent:*
1. $V = U \oplus W$.
2., $V = U + W$ and $U \cap W = \{0\}$.
3. For any basis B_1 of U and B_2 of W, the union $B = B_1 \cup B_2$ is a basis of V.

Proof. Trivial. ☐

Definition 3.3. Let $V_1, \ldots, V_l \leq V$, then we define

$$\sum_{i=1}^{l} V_i = \{v_1 + \cdots + v_l : v_j \in V_j, 1 \leq j \leq l\}$$

The sum is direct, i.e.

$$\sum_{i=1}^{l} V_i = \bigoplus_{i=1}^{l} V_i$$

iff $v_1 + \cdots + v_l = v_1' + \cdots + v_l'$ implies $v_j = v_j'$ for any $1 \leq j \leq l$ and $v_j \in V_j$. Equivalently,

$$V = \bigoplus_{i=1}^{l} V_i \iff \forall v \in V, \exists!(v_1, \ldots, v_l) \in V_1 \times \cdots \times V_l, v = \sum_{i=1}^{l} v_i$$

Proposition 3.4. *The followings are equivalent:*
1.

$$\sum_{i=1}^{l} V_i = \bigoplus_{i=1}^{l} V_i$$

2. For any i*,*

$$V_i \cap \left(\sum_{j \neq i} V_j \right) = \{0\}$$

3. For any bases B_i *of* V_i*, the union* $\bigcup_i B_i$ *is a basis of* $\sum_i V_i$*.*

Proof. Trivial. ☐

4 Linear Maps, Isomorphisms and the Rank-Nullity Theorem

Definition 4.1 (Linear Map)**.** Let V, W are vector spaces over F, a function $\alpha : V \to W$ is linear if for any $\lambda_1, \lambda_2 \in F$ and $v_1, v_2 \in V$,

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

Example 4.1. 1. Let M be an $m \times n$ matrix, then $\alpha : \mathbb{R}^n \to \mathbb{R}^m$ via $x \mapsto Mx$ is a linear map.
2. The functional $\alpha : C([0,1]) \to C^1([0,1])$ via

$$\alpha(f)(x) = \int_0^x f(t)\,\mathrm{d}t$$

is a linear map.
3. Fix $x \in [a,b]$, then the evaluation map $\alpha : C([a,b]) \to \mathbb{R}$ via $f \mapsto f(x)$ is a linear map.

Remark. The identity map is a linear map. Composition of linear maps is also a linear map.

Lemma 4.1. *Let V, W be vector spaces over F and B a basis for V. Let $\alpha_0 : B \to W$ be a function, then there is a unique linear map $\alpha : V \to W$ that extends α_0.*

Proof. For any $(b_i) \in B$, necessarily $\alpha\left(\sum_i \lambda_i b_i\right) = \sum_i \lambda \alpha_0(b_i)$. This is sufficient. $\square$

Remark. This lemma is true for infinite dimensional vector spaces as well. Often, to define linear map, we often just define its values on a basis and extend it by this lemma.

Corollary 4.2. *Two linear maps that agree on a basis are the same.*

Proof. This is just the uniqueness statement. $\square$

Definition 4.2. Let V, W be vector spaces over F. A linear bijection $\alpha : V \to W$ is an isomorphism (of vector spaces). If such a map exists, then we say V, W are isomorphic (as vector spaces), written as $V \cong W$.

Remark. If α is an isomorphism, so is α^{-1}.

Lemma 4.3. $\cong$ *is an equivalence relation on the class of all vector spaces over F.*

Proof. Just check. $\square$

Theorem 4.4. *If V is a vector space over F of dimension n, then $V \cong F^n$.*

Proof. Take a basis $\{b_1, \dots, b_n\}$ of V, then

$$\alpha(x_1 b_1 + \cdots + x_n b_n) = (x_1, \dots, x_n)$$

is an isomorphism. $\square$

Remark. Choosing a basis of V is then just equivalent to choosing an isomorphism from V to F^n.

Theorem 4.5. *Let V, W be finite dimensional vector spaces over F. Then $V \cong W$ iff $\dim V = \dim W$*

Proof. Any basis of V induces a basis of W via the isomorphism, so they have the same dimension. Therefore are both isomorphic to F^n where $n = \dim V = \dim W$. $\square$

Definition 4.3. Let $\alpha : V \to W$ be a linear map. We define the kernel of α to be $\ker \alpha = \{v \in V : \alpha(v) = 0\}$ and the image to be $\operatorname{Im} \alpha = \alpha(V) = \{w \in W : \exists v \in V, \alpha(v) = w\}$.

Lemma 4.6. $\ker \alpha \le V, \operatorname{Im} \alpha \le W$.

Proof. Obvious. $\square$

Example 4.2. Take $\alpha : C^\infty(\mathbb{R}) \to C^\infty(\mathbb{R})$ by $\alpha(f)(t) = f''(t) + f(t)$. Then $\ker \alpha$ is spanned by $t \mapsto e^t$ and $t \mapsto e^{-t}$ and $\operatorname{Im} \alpha = C^\infty(\mathbb{R})$.

Theorem 4.7. *Let V, W be vector spaces over F and $\alpha : V \to W$ be linear, then $V/\ker \alpha \cong \operatorname{Im}(\alpha)$ via $v + \ker \alpha \mapsto \alpha(v)$.*

Proof. Just check. □

Definition 4.4. The rank of $\alpha : V \to W$ is $r(\alpha) = \dim \operatorname{Im} \alpha$ and nullity is $n(\alpha) = \dim \ker \alpha$.

Hence in the finite dimensional case, we can rewrite the preceding theorem to get

Theorem 4.8 (Rank-Nullity Theorem)**.** *Let $\alpha : V \to W$ be linear where V is finite dimensional. Then $\dim V = r(\alpha) + n(\alpha)$.*

Proof. Follows from the preceding theorem. □

Corollary 4.9 (Classification of Isomorphism)**.** *Let V, W be finite dimensional vector spaces with $\dim V = \dim W$ and $\alpha : V \to W$ be linear, then the followings are equivalent:*
1. α is injective.
2. α is surjective.
3. α is an isomorphism.

Proof. Follows immediately by considering dimensions. □

Example 4.3. Consider

$$V = \left\{ \begin{pmatrix} x \\ y \\ z \end{pmatrix} \in \mathbb{R}^3 : x + y + z = 0 \right\}$$

We want to compute $\dim V$. Consider $\alpha : \mathbb{R}^3 \to \mathbb{R}$ via $(x, y, z)^\top \mapsto x + y + z$, then $r(\alpha) = 1$ and $n(\alpha) = V$, so $\dim V = 3 - 1 = 2$. Geometrically, V is just a plane with normal $(1, 1, 1)^\top$.

5 Linear Maps and Matrices

The set of linear maps from V to W is a vector space over the same field F.

Definition 5.1. Let V, W be vector spaces over F. We define

$$L(V, W) = \{\alpha : V \to W, \text{ linear}\}$$

to be the vector space of linear maps from V to W under the operations $(\alpha + \beta)(v) = \alpha(v) + \beta(v)$ and $(\lambda\alpha)(v) = \lambda\alpha(v)$.

It is easy to verify that this is indeed a vector space. There is a very important theorem

Proposition 5.1. *If V, W are finite dimensional, so is $L(V, W)$ and we have $\dim L(V, W) = (\dim V)(\dim W)$.*

Definition 5.2. An $m \times n$ matrix over F is an array with m rows and n columns with entries in F. As a convention, for a matrix A we write $A = (a_{ij})_{1 \le i \le m, 1 \le j \le n} = (A_{ij})_{1 \le i \le m, 1 \le j \le n}$ where i refers to the row number and j the column number. We write $M_{m,n}(F)$ to denote the set of $m \times n$ matrices over F.

Note that we can (and often) identify vectors in $\mathbb{R}^m$ as $m \times 1$ matrices.

Proposition 5.2. *$M_{m,n}(F)$ is a vector space over F under $(a_{ij}) + (b_{ij}) = (a_{ij} + b_{ij})$ and $\lambda(a_{ij}) = (\lambda a_{ij})$.*

Proof. Trivial. ☐

Proposition 5.3. $\dim M_{m,n}(F) = mn$.

Proof. The set $\{(\delta_{ia}\delta_{jb}) : 1 \leq a \leq m, 1 \leq b \leq n\}$ with size mn is a basis of $M_{m,n}(F)$. ☐

We want to represent linear maps by matrices. For V, W vector spaces over F of finite dimensions n, m, we choose ordered bases $B = (b_1, \ldots, b_n), C = (c_1, \ldots, c_m)$ for V, W respectively. For $v \in V$, its coordinate under B is $[v]_B = (v_1, \ldots, v_n)$ where $v = \sum_i v_i b_i$ is the unique decomposition of v in this basis. Similarly, the coordinate of $w \in W$ under C is $[w]_C = (w_1, \ldots, w_m)$ where $w = \sum_i w_i c_i$ is the unique decomposition of w under C.

Definition 5.3. The matrix of $\alpha : V \to W$ in the bases B, C is the matrix

$$([\alpha]_{B,C})_{ij} = ([\alpha(b_j)]_C)_i \in M_{m\times n}(F)$$

If the bases are understood, we often write $([\alpha]_{B,C})_{ij}$ as α_{ij}.

Definition 5.4. For matrices $M \in M_{m\times n}(F)$ and $N \in M_{n\times l}(F)$, their matrix product is the $m \times l$ matrix defined by $(MN)_{ij} = (\sum_k M_{ik}N_{kj})_{ij}$.

The particular case where N is a column vector exhibits how $m \times n$ matrices induce linear maps $F^n \to F^m$.

Lemma 5.4. *For any $v \in V$,*

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

Proof. Linearity. ☐

We already know that composition of linear maps is linear, what's more is

Lemma 5.5. *Let U, V, W be finite dimensional vector spaces over F with chosen bases A, B, C respectively. Let $\beta : U \to V$ and $\alpha : V \to W$ be linear maps, then*

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

Proof. Linearity again. ☐

Proposition 5.6. *We have $L(V, W) \cong M_{m\times n}(F)$ via $\alpha \mapsto [\alpha]_{B,C}$ where B is a basis of V and C is a basis of W.*

Proof. Just check. ☐

Proof of Proposition 5.1. Simple corollary of the preceding proposition. ☐

Remark. Let $\alpha : V \to W$ be linear. The diagram

$$\begin{array}{ccc} V & \xrightarrow{\alpha} & W \\ {\scriptstyle v\mapsto[v]_B}\downarrow & & \downarrow{\scriptstyle w\mapsto[w]_C} \\ F^n & \xrightarrow[{[\alpha]_{B,C}}]{} & F^m \end{array}$$

commutes.

Example 5.1. Let $\alpha : V \to W$ be linear and $Y \leq V$ a subspace. Let B be a basis of V extending a basis B' of Y and C a basis of W extending a basis C' of $Z \geq \alpha(Y)$. Then the entries of $[\alpha]_{B,C}$ that are relevant to $B' \subset B, C' \subset C$ are exactly $[\alpha|_{Y\to Z}]_{B',C'}$. So if we rearrange the bases so that B' are in the front of B and C' in the front of C, then the matrix will look like

$$[\alpha]_{B,C} = \left(\begin{array}{c|c} [\alpha|_{Y\to Z}]_{B',C'} & * \\ \hline 0 & * \end{array}\right)$$

Proposition 5.7. *Let $\alpha : V \to W$ be linear and $\alpha(Y) \leq Z \leq W$, then α induces a map $\bar{\alpha} : V/Y \to W/Z$ via $v + Y \mapsto \alpha(v) + Z$.*

Proof. Trivial. □

6 Change of Basis and Equivalent Matrices

Consider vector spaces V, W and $B = \{v_1, \dots, v_n\}, B' = \{v'_1, \dots, v'_n\}$ bases of V, $C = \{w_1, \dots, w_n\}, C' = \{w'_1, \dots, w'_n\}$. Let $\alpha : V \to W$ be a linear map. We want to study the relationship between $[\alpha]_{B,C}$ and $[\alpha]_{B',C'}$.

Definition 6.1. For a vector space V with bases $B = \{v_1, \dots, v_n\}, B' = \{v'_1, \dots, v'_n\}$, the change of basis matrix from B' to B is $P = (p_{ij})_{1\leq i,j\leq n}$ given by $p_{ij} = ([v'_j]_B)_i$.

Indeed $P = [\mathrm{id}_V]_{B',B}$.

Lemma 6.1. $[v]_B = P[v]_{B'}$.

Proof. $P[v]_{B'} = [\mathrm{id}_V]_{B',B}[v]_{B'} = [\mathrm{id}_V(v)]_B = [v]_B$. □

Remark. Let P be the change of basis matrix from B' to B, then P is invertible and P^{-1} is the change of basis matrix from B' to B.

So for the problem we stated at the start of this section, we write $P = [\mathrm{id}_V]_{B',B}$ and $Q = [\mathrm{id}_W]_{C',C}$, then

Proposition 6.2. *Let $A = [\alpha]_{B,C}$, $A' = [\alpha]_{B',C'}$ and P, Q be as above, then $A' = Q^{-1}AP$.*

Proof. For any $v \in V$, we evaluate

$$AP[v]_{B'} = [\alpha]_{B,C}[v]_B = [\alpha(v)]_C = Q[\alpha(v)]_{C'} = Q[\alpha]_{B',C'}[v]_{B'} = QA'[v]_{B'}$$

So $AP = QA'$, which means $A' = Q^{-1}AP$. □

Definition 6.2 (Equivalent Matrices)**.** Two matrices $A, A' \in M_{m,n}(F)$ are equivalent if $A' = Q^{-1}AP$ for some $Q \in M_{m,m}, P \in M_{n,n}$ invertible.

Remark. As one can check, this defines an equivalent relation.

Proposition 6.3. *Let V, W be vector spaces over F and $\dim V = n, \dim W = m$. Let $\alpha : V \to W$ be linear. Then there exists bases B of V and C of W such that*

$$[\alpha]_{B,C} = \left(\begin{array}{c|c} I_r & 0 \\ \hline 0 & 0 \end{array}\right)$$

where I_r is the identity matrix of dimension $r = n - n(\alpha)$.

Proof. Fix a basis $v_{r+1}, \ldots, v_n$ of $\ker \alpha$ and extend it to a basis $B = \{v_1, \ldots, v_r\}$. It is easy to see that $\alpha(v_1), \ldots, \alpha(v_r)$ gives a basis of $\operatorname{Im} \alpha$ as it is a spanning set that has the right size and that

$$\sum_{i=1}^{r} \lambda_i \alpha(v_i) = 0 \implies \alpha\left(\sum_{i=1}^{r} \lambda_i v_i\right) = 0 \implies \sum_{i=1}^{r} \lambda_i v_i \in \ker \alpha \implies \forall i, \lambda_i = 0$$

Extend this to a basis C of W, then $[\alpha]_{B,C}$ can be easily seen to have the right form. □

Remark. This provides another proof of the Rank-Nullity Theorem.

Corollary 6.4. *Any $m \times n$ matrix is equivalent to a matrix in the form illustrated in the preceding proposition.*

Proof. Immediate. □

Definition 6.3. Let $A \in M_{m \times n}(F)$. The column rank $r(A)$ of A is the dimension of the subspace spanned by the columns of A in F^n. Similarly, the row rank of A is the column rank of A^T.

Remark. If α is a linear map represented by A with respect to some basis, then $r(\alpha) = r(A)$.

Proposition 6.5. *Two matrices A, A' of the same dimension are equivalent iff $r(A) = r(A')$.*

Proof. Direct consequence of Proposition 6.3 and the preceding remark. □

Theorem 6.6. $r(A) = r(A^T)$.

Proof. Let $r = r(A)$, so there are some invertible Q, P of the right sizes such that

$$Q^{-1}AP = \left(\begin{array}{c|c} I_r & 0 \\ \hline 0 & 0 \end{array}\right)$$

which is an $m \times n$ matrix. But then

$$P^\top A^\top (Q^\top)^{-1} = P^\top A^\top (Q^{-1})^\top = (Q^{-1}AP)^\top = \left(\begin{array}{c|c} I_r & 0 \\ \hline 0 & 0 \end{array}\right)$$

as an $n \times m$ matrix. So $r(A^\top) = r = r(A)$. □

In the case where $\alpha : V \to V$ is an endomorphism, the change of basis formula becomes $A' = P^{-1}AP$ where P is the change of basis matrix and A', A are the matrices of α in the two different bases. This induces the following definition.

Definition 6.4. Let A, A' be square matrices. We say A, A' are similar (or conjugate) if there is an $n \times n$ invertible square matrix P such that $A' = P^{-1}AP$

This notion is central to the study of diagonalisation and spectral theory.

7 Elementary Equations and Elementary Matrices

Definition 7.1. An elementary column operation on an $m \times n$ matrix A are one of the followings:
(i) Swap columns i, j with $i \neq j$.
(ii) Multiply the entire column i by $\lambda \in F \setminus \{0\}$.
(iii) Add λ times column i to column j where $\lambda \in F$.

We can do row operations in a analogous (transposed) way. Something remarkable is that these operations are invertible. Instead of find the inverses one-by-one, we realise these operations via the action of elementary matrices. Let E_{ij} be the matrix with 1 on the i, j entry and zero anywhere else, we have:

Definition 7.2 (Elementary Matrices)**.** The elementary matrices are $T_{ij} = I - E_{ii} - E_{jj} + E_{ij} + E_{ji}$ for $i \neq j$, $M_{i,\lambda} = I + (\lambda - 1)E_{ij}$ for $\lambda \neq 0$ and $C_{i,j,\lambda} = I + \lambda E_{ij}$.

Then, we easily see that T_{ij} corresponds to column (row) operation (i), $M_{i,\lambda}$ to operation (ii) and $C_{i,j,\lambda}$ to operation (iii) via the operation of multiplying A with the correspondinng matrix from the right (left).

Example 7.1.

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

Constructive Proof of Proposition 6.3. It suffices to show that we can get from any matrix to

$$\left(\begin{array}{c|c} I_r & 0 \\ \hline 0 & 0 \end{array}\right)$$

via elementary column and row operations. Start with a matrix A. If $A = 0$ then we are done. Otherwise pick $a_{ij} = \lambda \neq 0$ and swap rows i and 1 and then columns j and 1, after which λ is at position $1, 1$. Then multiply column 1 by λ^{-1}. So we get 1 at position $1, 1$. Now we clean up row 1 and column 1 via operation (iii) (both row and column). Afterwards we can perform the same procedure on the submatrix by removing the first row and column. By induction we can get the desired form at the end. □

There are a few variations on these row and column operations. The first one is Gauss' pivot algorithm. If one use only row operations, then one will reach the "row echelon form" (which we will define later) in the following way:

Assume $a_{i1} \neq 0$ for some i. Otherwise just move on by deleting the first zero columns. Then swap rows i and 1 and divide row 1 by $\lambda = a_{i1}$ to get 1 at position $1, 1$ and use (iii) to clean up the first column. And one can move on with the same method to the submatrix removing the first row and first column. Do this repeatedly and at the end we can get a matrix satisfying:
1. For any i here exists $k(i) \geq i$ such that $a_{ij} = 0$ for any $j < k(i)$.
2. $k(i)$ is increasing in i.
3. Row $k(i)$ equals $e_{k(i)}$.
And matrices satisfying these conditions are called matrices in row echelon form. Note that the operations above is exactly what we will get when solving a linear system of equations. So this can be an algorithm of doing that, which is now known as Gauss' pivot algorithm (or Gaussian elimination).
Another variation is the following:

Lemma 7.1. *We can obtain the identity matrix from any invertible square matrix via column operations only.*

By transpose, we can replace column operations by row operations.

Proof. We argue by induction on the k where we can guarantee to transform A to the form

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

The initial case is obvious. Suppose we can do this for some k, we shall show that we can do this for $k+1$. Now there must be some $j > k$ such that $a_{k+1,j} = \lambda \neq 0$. Otherwise the vector e_{k+1} is not in the span of the column vectors of A, contradiction. So we swap columns $k+1$ and j and then divide column $k+1$ by λ. This gets us 1 at position $k+1, k+1$. Then we can clean up the row $k+1$ by this 1, which completes the induction process. □

This immediately provides an algorithm (a quite cost-effective one) for computing the inverse. As one can see, this algorithm is analogous to the algorithm of solving a nonsingular linear system. Also,

Proposition 7.2. *Any invertible matrix is a product of elementary matrices.*

Proof. Writing the operations in the preceding lemma as a product of elementary matrices representing the operations gives the inverse of that matrix. But any invertible matrix is the inverse of its own inverse. □

8 Dual Space and Dual Maps

Definition 8.1. Let V be a vector space over F, we define the dual space V^* of V to be the set of linear maps from V to F, i.e. $V^* = L(V, F)$.
We call linear maps $V \to F$ as linear forms.

Example 8.1. 1. The map $\mathrm{tr} : M_{n,n}(F) \to F$ via $A = (a_{ij}) \mapsto \sum_i a_{ii}$ is a linear form, so $\mathrm{tr} \in (M_{n,n}(F))^*$.
2. For a function $f :\in C^\infty([0,1], \mathbb{R})$, we can define the map $T_f : C^\infty([0,1], \mathbb{R})$ via

$$T_f(\phi) = \int_0^1 f(x)\phi(x)\,\mathrm{d}x$$

An interesting thing is, if we are given all information about T_f, can we recover f? The answer is yes and is left as an exercise. This idea comes from quantum mechanics, where you only get the information about T_f by physical measurements but you want information about f.

There is a natural way of finding a basis for the dual space.

Lemma 8.1. *Let V be a vector space over F with a finite basis $B = \{e_1, \dots, e_n\}$, then there exists a basis for V^* given by $B^* = \{\epsilon_1, \dots, \epsilon_n\}$ where $\epsilon_j(\sum_i a_i e_i) = a_j$ for $1 \leq j \leq n$.*

Definition 8.2. The basis B^* in the preceding lemma is called the dual basis.

Remark. If we introduce the Kronecker delta

$$\delta_{ij} = \begin{cases} 1, \text{ if } i = j \\ 0, \text{ otherwise} \end{cases}$$

then we can define ϵ_j by extending $\epsilon_j(e_i) = \delta_{ij}$.

Proof. If there is some $\lambda_j \in F$ such that $\sum_j \lambda_j \epsilon_j = 0$, then in particular the evaluation of the left hand side at e_i is zero for each i. But $\epsilon_j(e_i) = 0$, so $\lambda_i = 0$ for each i.
To see this is spanning, one just observe that $\alpha = \sum_j \alpha(e_j)\epsilon_j$ for any linear form α. □

Corollary 8.2. *If V is finite dimensional, then $\dim V^* = \dim V$.*

Remark. It is convenient to think of V^* as the space of row vectors of length n over F. Indeed, if we have a basis $\{e_i\}$ of V and the corresponding dual basis $\{\epsilon_i\}$ for V^*, then via calculation we can obtain $\alpha(x) = \sum_i \alpha_i x_i$ where $\alpha = \sum_i \alpha_i \epsilon_i$ and $x = \sum_i x_i e_i$.

Definition 8.3. If $U \subset V$ is a subset of the vector space V, then the annihilator of U is

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

Lemma 8.3. *$U^\circ \leq V^*$ and if $U \leq V$ and $\dim V$ is finite, then $\dim V = \dim U + \dim U^\circ$.*

Proof. Quite obvious that $U^\circ \leq V^*$. To see the identity, write $n = \dim V$ and extend a basis $\{e_1, \dots, e_k\}$ of U to a basis $\{e_1, \dots, e_n\}$ of V. Let $\{\epsilon_1, \dots, \epsilon_n\}$ be the dual basis. Then it becomes obvious that $U^\circ = \langle\{\epsilon_{k+1}, \dots, \epsilon_n\}\rangle$, which gives the identity. □

Lemma 8.4. *Let V, W be vector spaces over F and $\alpha \in L(V, W)$. Then the map $\alpha^* : W^* \to V^*$ sending ϵ to $\epsilon \circ \alpha$ is linear.*

Proof. Obvious. □

Definition 8.4. This map α^* is called the dual map of α.

Proposition 8.5. *Let V, W be finite dimensional vector spaces over F with bases B, C. Let B^*, C^* be the dual bases of V^*, W^*, then $[\alpha^*]_{C^*,B^*} = [\alpha]_{B,C}^\top$.*

Proof. Just write it down. □

Lemma 8.6. *Let E, F be bases of V and $P = [\mathrm{id}]_{F,E}$ be the change-of-basis matrix from F to E. Let E^*, F^* be the corresponding dual bases, then the change-of-basis matrix from F^* to E^* is $(P^{-1})^\top$.*

Proof. We have

$$[\mathrm{id}]_{F^*,E^*} = [\mathrm{id}]_{E,F}^\top = (P^{-1})^\top$$

as desired. □

9 Properties of the Dual Map and Double Dual

Lemma 9.1. *Let V, W be vector spaces over F and $\alpha \in L(V, W)$. Let $\alpha^* \in L(W^*, V^*)$ be the dual map, then:*
1. $\ker \alpha^* = (\operatorname{Im} \alpha)^\circ$*, so α^* is injective iff α is surjective.*
2. $\operatorname{Im} \alpha^* \leq (\ker \alpha)^\circ$ *with equality if V, W are finite dimensional, in which case it implies that α^* is surjective iff α is injective.*

This is very important as it shows how we can understand α from α^*, which is often simpler.

Proof. 1. Pick $\epsilon \in W^*$, then $\epsilon \in \ker \alpha^*$ iff $\alpha^*(\epsilon) = 0$ iff $\epsilon \circ \alpha = 0$ iff $\epsilon \in (\operatorname{Im} \alpha)^\circ$.
2. We first show that $\operatorname{Im} \alpha^* \leq (\ker \alpha)^\circ$. Indeed, for any $\epsilon \in \operatorname{Im} \alpha^*$, we have $\epsilon = \alpha^*(\phi) = \phi \circ \alpha$ for some $\phi \in W^*$. But then for any $u \in \ker \alpha$ we have $\epsilon(u) = \phi \circ \alpha(u) = \phi(0) = 0$, which means $\epsilon \in (\ker \alpha)^\circ$. In finite dimension, pick bases B, C of V, W and we get

$$\begin{aligned}
\dim \operatorname{Im} \alpha^* &= r(\alpha^*) = r([\alpha^*]_{C^*,B^*}) = r([\alpha]_{B,C}^\top) \\
&= r([\alpha]_{B,C}) = r(\alpha) \\
&= \dim V - \dim \ker \alpha \\
&= \dim(\ker \alpha)^\circ
\end{aligned}$$

So they have the same dimension, hence equal. □

We now turn to a very important concept known as double dual. V^* is a vector space too, so we can also construct its dual

$$V^{**} = L(V^*, F) = (V^*)^*$$

Why is it important? Well, not much in finite dimensions, but in infinite dimensonal spaces, it is very hard to find obvious relations between V and V^*. However, there is a canonical embedding of V into V^{**}. Indeed, pick $v \in V$, consider $\hat{v} : V^* \to F$ via $\epsilon \mapsto \epsilon(v)$, which is a well-defined element of V^{**}. Quite ironically, our first theorem on this topic is about finite-dimensional spaces.

Theorem 9.2. *If V is finite dimensional, then this operation $\hat{}: V \to V^{**}$ we just described is an isomorphism of vector spaces.*

So we can just identify V^{**} with V.

Proof. Linearity is standard. To see it is injective, let $e \in V \setminus \{0\}$ and extend $\{e\}$ to a basis $\{e, e_2, \ldots, e_n\}$ of V. So the dual basis $(\epsilon, \epsilon_2, \ldots, \epsilon_n)$ would have $\hat{e}(\epsilon) = \epsilon(e) = 1$. Therefore $\hat{}$ has trivial kernel, hence injective. It then follows that it is an isomorphism as $\dim V = \dim V^* = \dim V^{**}$. □

Remark. In further linear analysis and functional analysis, we will see that $\hat{}$ remains injective for a huge class of infinite dimensional vector spaces (those of interests are often space of functions). And there are many of them (called reflexive spaces) where $\hat{}$ is actually an isomorphism. The theories emerged from here have numerous applications in analysis.

Lemma 9.3. *Let V be a finite dimensional vector space over F and $U \leq V$. Define $\hat{U} = \{\hat{u} : u \in U\} \leq V^{**}$. Then $\hat{U} = U^{\circ\circ} = (U^\circ)^\circ$.*

Thus we can identify $U^{\circ\circ}$ with U too.

Proof. Trivial. □

Lemma 9.4. *Let V be finite dimensional vector space over F and $U_1, U_2 \leq V$, then:*
1. $(U_1 + U_2)^\circ = U_1^\circ \cap U_2^\circ$.
2. $(U_1 \cap U_2)^\circ = U_1^\circ + U_2^\circ$.

Proof. Just write it out. □

10 Bilinear Forms

Definition 10.1. Let U, V be vector spaces over F, then $\phi : U \times V \to F$ is a bilinear form if $\phi(u, \cdot) \in V^*$ and $\phi(\cdot, v) \in U^*$ for any $u \in U, v \in V$.

We write $\phi_L \in L(U, V^*)$ to be the map $u \mapsto \phi(u, \cdot)$ and $\phi_R \in L(V, U^*)$ to be $v \mapsto \phi(\cdot, v)$ respectively, so we have $\phi_L(u)(v) = \phi(u, v) = \phi_R(v)(u)$.

Example 10.1. 1. The map $V \times V^* \to F$ via $(v, \theta) \mapsto \theta(v)$ is a bilinear form.
2. The scalar product on F^n, that is

$$\left(\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}\right) \mapsto \sum_{i=1}^n x_i y_i$$

is a bilinear form.
3. Take $U = V = C([0,1], \mathbb{R})$, then

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

is a bilinear form.

Definition 10.2. Take a basis $B = \{e_1, \ldots, e_m\}$ of U and $C = \{f_1, \ldots, f_n\}$ basis of V and $\phi : U \times V \to F$ a bilinear form, then the matrix of ϕ with respect to B, C is

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

Lemma 10.1. *We have* $\phi(u,v) = [u]_B^\top[\phi]_{B,C}[v]_C$ *for any* $u \in U, v \in V$.

Proof. If $u = \sum_i \lambda_i e_i$ and $v = \sum_j \mu_j f_j$, then by linearity,

$$\phi(u,v) = \sum_{i=1}^n \sum_{j=1}^n \lambda_i \mu_j \phi(e_i, f_j) = [u]_B^\top[\phi]_{B,C}[v]_C$$

by simple expansion. □

Remark. The matrix $[\phi]_{B,C}$ is the unique matrix such that the previous lemma holds.

Lemma 10.2. *Take a basis* $B = \{e_1, \dots, e_m\}$ *of* U *and the dual basis* $B^* = \{\epsilon_1, \dots, \epsilon_m\}$ *of* U^*. *Similarly take a basis* $C = \{f_1, \dots, f_n\}$ *of* V *and the dual basis* $\{\eta_1, \dots, \eta_n\}$ *of* V^*. *If* $A = [\phi]_{B,C}$ *where* $\phi : U \times V \to F$ *is a bilinear form, then* $[\phi_R]_{C,B^*} = A$ *and* $[\phi_L]_{B,C^*} = A^\top$.

Proof. Pretty much from definition. □

Definition 10.3. $\ker \phi_L$ is called the left kernel of ϕ and $\ker \phi_R$ is called the right kernel of ϕ.
ϕ is nondegenerate if both kernels are zero. Otherwise, we say ϕ is degenerate.

Lemma 10.3. *Let* B, C *be bases of* U, V *respectively and* $\phi : U \times V \to F$ *be bilinear. Let* $A = [\phi]_{B,C}$, *then* ϕ *is nondegenerate iff* A *is invertible.*

Proof. Immediate from the preceding lemma. □

Corollary 10.4. *If* ϕ *is nondegenerate, then* $\dim U = \dim V$.

Proof. All invertible matrices are square. □

Example 10.2. So the dot product on $\mathbb{R}^n$ is nondegenerate.

Corollary 10.5. *If* U, V *are finite dimensional vector spaces over* F, *then choosing a nondegenerate bilinear form* $U \times V \to F$ *is just choosing an isomorphism* $\phi_L : U \to V^*$.

Proof. Obvious. □

Definition 10.4. For $T \subset U$, we define $T^\perp = \{v \in V : \forall t \in T, \phi(t,v) = 0\}$ and for $S \subset V$ we define ${}^\perp S = \{u \in U : \forall s \in S, \phi(u,s) = 0\}$

Of course we want to change the basis.

Proposition 10.6. *Let* B, B' *be bases of* U *and* $P = [\mathrm{id}]_{B',B}$ *and* C, C' *be basis of* V *and* $Q = [\mathrm{id}]_{C',C}$ *and let* $\phi : U \times V \to F$ *be a bilinear form. Then* $[\phi]_{B',C'} = P^\top[\phi]_{B,C}Q$.

Proof. We have

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

So necessarily $[\phi]_{B',C'} = P^\top[\phi]_{B,C}Q$. □

Lemma 10.7. *The rank of the matrix of* ϕ *is invariant under change of basis.*

Proof. Immediate. □

Definition 10.5. The rank $r(\phi)$ of ϕ is the rank of its matrix in any basis.

Remark. We have $r(\phi) = r(\phi_R) = r(\phi_L)$.

11 Trace and Determinant

Definition 11.1. Let $A \in M_n(F) = M_{n,n}(F)$ be a square $n \times n$ matrix. The trace of A is defined to be

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

Remark. The map sending a matrix to its trace is a linear form.

Lemma 11.1. $\operatorname{tr}(AB) = \operatorname{tr}(BA)$.

Proof. Write stuff out. □

Corollary 11.2. *Similar matrices have the same trace.*

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

Definition 11.2. If $\alpha : V \to V$ is linear, then $\operatorname{tr} \alpha = \operatorname{tr}[\alpha]_B$ for any choice of basis B of V. It is well-defined by the preceding corollary.

Lemma 11.3. *Let $\alpha : V \to V$ be linear and $\alpha^* : V^* \to V^*$ be the dual map, then $\operatorname{tr} \alpha = \operatorname{tr} \alpha^*$.*

Proof. Choose any basis B of V, then

$$\operatorname{tr} \alpha = \operatorname{tr}[\alpha]_B = \operatorname{tr}[\alpha]_B^\top = \operatorname{tr}[\alpha^*]_{B^*} = \operatorname{tr} \alpha^*$$

as desired. □

Recall that we can decompose any permutation $\sigma \in S_n$ into a product of transpositions.

Definition 11.3. The signature of a permutation is the (necessarily unique) homomorphism $\epsilon : S_n \to \{1, -1\}$ that sends any transposition to -1.

This map ϵ is well-defined as we know that the parity of the number of transpositions that builds up a permutation is fixed.

Definition 11.4. Let $A = (a_{ij}) \in M_n(F)$. We define the determinant of A as

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

Example 11.1. For $n = 2$, we have

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

Lemma 11.4. *If $A = (a_{ij})$ is an upper (resp. lower) triangular matrix, i.e. $a_{ij} = 0$ for $i > j$ (resp. $i < j$), then $\det A = 0$.*

Proof. The only permutation σ such that $\sigma(j) \leq j$ (resp. $\sigma(j) \geq j$) for all j is the identity. □

Lemma 11.5. $\det A = \det A^\top$.

Proof. For any $\sigma \in S_n$ we know that $\epsilon(\sigma) = \epsilon(\sigma^{-1})$. □

Definition 11.5. A volumn form d in F^n is a function $(F^n)^n \to F$ such that:
1. It is multilinear: For any $i \in \{1, \ldots, n\}$ and $v_1, \ldots, v_{i-1}, v_{i+1}, \ldots, v_n \in F^n$, the map

$$v \mapsto d(v_1, \ldots, v_{i-1}, v, v_{i+1}, \ldots, v_n \in F^n)$$

is linear.
2. It is an alternating form: If $v_i = v_j$ for some $i \neq j$, then $d(v_1, \ldots, v_n) = 0$.

What we want to prove that there is only one volumn form (up to multiplicative constant). If this is true, then it necessarily equals det in the following way:

Lemma 11.6. det *is a volumn form via the obvious identification* $M_n(F) = (F^n)^n$ *by grouping the* n *column vectors as a tuple.*

Proof. det is linear as it is linear in any entry. It is an alternating form as ϵ sends any transposition to -1. □

Lemma 11.7. *Let* d *be a volumn form, then swapping two entries changes the sign.*

Proof. For any $i \neq j$, $d(v_1, \ldots, v_i, \ldots, v_j, \ldots, v_n) + d(v_1, \ldots, v_j, \ldots, v_i, \ldots, v_n) = d(v_1, \ldots, v_i + v_j, \ldots, v_i + v_j, \ldots, v_n) = 0$. □

Corollary 11.8. *For any* $\sigma \in S_n$ *and volume form* d*,*

$$d(v_{\sigma_1}, \ldots, v_{\sigma_n}) = \epsilon(\sigma) d(v_1, \ldots, v_n)$$

Proof. Just decompose σ into transpositions. □

Theorem 11.9. *Let* $A \in M_n(F)$ *and let* $A^{(i)}$ *be the* i^{th} *column of* A*. For any volumn form* d*, we have*

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

where $(e_i)_j = \delta_{ij}$.

This is what we wanted.

Proof. Just expand using linearity and the preceding corollary. □

Corollary 11.10. det *is the unique volumn form that maps* $(e_1, \ldots, e_n)$ *to* 1.

12 Some Properties of Determinant

Lemma 12.1. *Let* A, B *be sqaure matrices, then* $\det(AB) = \det(A)\det(B)$.

Proof. Direct expansion does the trick, but alternatively we can define d_A via $d_A(B) = \det(AB)$ which is obviously a volume form. Therefore $d_A(B) = d_A(I)\det(B) = \det(A)\det(B)$. □

Definition 12.1. Let $A \in M_n(F)$. We say A is singular if $\det A = 0$, otherwise we say A is nonsingular.

Lemma 12.2. *If* A *is invertible then it is non-singular.*

Proof. Suppose B is an inverse of A, then $\det(A)\det(B) = \det(AB) = \det(I) = 1 \neq 0$ therefore $\det A \neq 0$. $\square$

Remark. In particular $\det(A^{-1}) = (\det A)^{-1}$ if A is invertible.

Theorem 12.3. *Let $A \in M_n(F)$, the followings are equivalent:*
1. A is invertible.
2. A is non-singular.
3. $r(A) = n$.

Proof. We just need to show (ii) implies (iii) since we have already done all others. Suppose $r(A) < n$, then $\dim \operatorname{span}(A^{(1)}, \ldots, A^{(n)}) < n$, so there is some $\lambda_1, \ldots, \lambda_n$ not all zero such that $\sum_i \lambda_i A^{(i)} = 0$. In particular, there is some j such that $\lambda_j \neq 0$ and hence $c_j = -\sum_{i\neq j}(\lambda_i/\lambda_j)A^{(i)}$. Expand $\det A$ using multilinearity gives a linear combination of determinants with repeated entries, which is zero as det is an alternating form. $\square$

Remark. By the theorem, it follows easily that the equation $Ax = y$ where $A \in M_n(F), x, y \in F^n$ has a unique solution iff $\det A \neq 0$.

Lemma 12.4. *Similar matrices have the same determinant.*

Proof. $\det(PAP^{-1}) = \det(P)\det(A)\det(P)^{-1} = \det(A)$. $\square$

Therefore the following definition makes sense.

Definition 12.2. If $\alpha : V \to V$ be an endomorphism, then $\det \alpha = \det[\alpha_{B,B}]$ for any B basis of V.

Theorem 12.5. $\det : L(V, V) \to F$ *satisfies:*
1. $\det(\mathrm{id}_V) = 1$.
2. $\det(\alpha \circ \beta) = \det(\beta)\det(\alpha)$.
3. $\det \alpha \neq 0$ *iff α is invertible. If indeed* $\det \alpha \neq 0$*, then* $\det(\alpha^{-1}) = (\det \alpha)^{-1}$.

Proof. Choose any basis and the rest follows from previous discussions. $\square$

Lemma 12.6. *Let $A \in M_k(F), B \in M_l(F), C \in M_{k,l}(F)$. Consider*

$$M_{k+l}(F) \ni M = \left(\begin{array}{c|c} A & C \\ \hline 0 & B \end{array}\right)$$

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

Proof. Write $n = k + l$ and $M = (m_{ij})$. Observe that $m_{\sigma(i)i} = 0$ if $i \leq k$ and $\sigma(i) > k$. So for $m_{\sigma(1)1} \cdots m_{\sigma(n)n} \neq 0$, we must have the decomposition $\sigma_1 \circ \sigma_2$ such that σ_1 fixes anything but $1, \ldots, k$ and σ_2 fixes anything but $k+1, \ldots, n$. But then for such σ, we have $m_{\sigma(j)j} = a_{\sigma_1(j)j}$ for any $j \in \{1, \ldots, k\}$ and $m_{\sigma(j)j} = b_{\sigma_2(s)s}$ where $s = j - k$ for any $j \in \{k+1, \ldots, n\}$. Observe also that $\epsilon(\sigma) = \epsilon(\sigma_1)\epsilon(\sigma_2)$, therefore

$$\begin{aligned}
\det M &= \sum_{\sigma \in S_n} \epsilon(\sigma) m_{\sigma(1)1} \cdots m_{\sigma(n)n} \\
&= \sum_{\sigma_1 \in S_k, \sigma_2 \in S_l} \epsilon(\sigma_1)\epsilon(\sigma_2) a_{\sigma_1(1)1} \cdots a_{\sigma_1(k)k} b_{\sigma_2(1)1} \cdots b_{\sigma_2(l)l} \\
&= \sum_{\sigma_1 \in S_k} \epsilon(\sigma_1) a_{\sigma_1(1)1} \cdots a_{\sigma_1(k)k} \sum_{\sigma_2 \in S_l} \epsilon(\sigma_2) b_{\sigma_2(1)1} \cdots b_{\sigma_2(l)l} \\
&= (\det A)(\det B)
\end{aligned}$$

as desired. ☐

Corollary 12.7. *If $A_1, \dots, A_k$ are square matrices, then*

$$\det \begin{pmatrix} A_1 & & & * \\ & A_2 & & \\ & & \ddots & \\ 0 & & & A_k \end{pmatrix} = (\det A_1)(\det A_2)\cdots(\det A_k)$$

Proof. Induction on k. ☐

In particular,

$$\det \begin{pmatrix} \lambda_1 & & & * \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_k \end{pmatrix} = \lambda_1 \cdots \lambda_k$$

Remark. Why is it called a volume from? Consider the map $(\mathbb{R}^3)^3 \to F$ via $(a, b, c) \mapsto a \cdot (b \times c)$ which is an example of a volume form. We know that geometrically this gives the (signed) volume of the parallelopiped formed by a, b, c.

13 The Adjugate Matrix

Given a square matrix $A \in M_n(F)$ with columns $A^{(i)}$. If we swap two neighbouring columns (or rows), then $\det A$ changes sign.

Remark. We can prove properties of determinant using the decomposition of A into elementary matrices.

A column expansion is a strategy to compute the determinant of a matrix by using its linkage to some of its submatrices.

Definition 13.1. Let $A \in M_n(F)$ and pick $i, j \in \{1, \dots, n\}$. We define $A_{\widehat{ij}} \in A_{n-1}(F)$ to be the matrix obtained by deleting the i^{th} row and j^{th} column from A.

Example 13.1. Take

$$A = \begin{pmatrix} 1 & 2 & -7 \\ 2 & 1 & 0 \\ -3 & 6 & 1 \end{pmatrix}$$

Then

$$A_{\widehat{32}} = \begin{pmatrix} 1 & -7 \\ 2 & 0 \end{pmatrix}$$

Lemma 13.1 (Expansion of Determinant)**.** *Let $A \in M_n(F)$.*
1. We have the expansion along the j^{th} column

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

2. We have the expansion along the i^{th} row

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

Example 13.2. Take

$$A = \begin{pmatrix} 1 & 2 & -1 \\ 3 & -1 & 1 \\ 4 & 2 & -7 \end{pmatrix}$$

So expanding along the second column gives

$$\det A = -(2) \det \begin{pmatrix} 3 & 1 \\ 4 & -7 \end{pmatrix} + (-1) \det \begin{pmatrix} 1 & -1 \\ 4 & -7 \end{pmatrix} - 2 \det \begin{pmatrix} 1 & -1 \\ 3 & 1 \end{pmatrix}$$

Proof. Suffices to show the first part. Pick $1 \leq j \leq n$ and write $A = (a_{ij})$, $A^{(j)} = \sum_i a_{ij} e_i$. Then

$$\begin{aligned}
\det A &= \det \left(A^{(1)}, \dots, \sum_{i=1}^{n} a_{ij} e_i, \dots, A^{(n)} \right) \\
&= \sum_{i=1}^{n} a_{ij} \det(A^{(1)}, \dots, e_i, \dots, A^{(n)}) \\
&= \sum_{i=1}^{n} a_{ij} (-1)^{j-1} \det(e_i, A^{(1)}, \dots, A^{(n)}) \\
&= \sum_{i=1}^{n} a_{ij} (-1)^{j-1} (-1)^{i-1} \det \begin{pmatrix} 1 & * & \cdots & * \\ 0 & & & \\ \vdots & & A_{\widehat{ij}} & \\ 0 & & & \end{pmatrix} \\
&= \sum_{i=1}^{n} (-1)^{i+j} a_{ij} \det A_{\widehat{ij}}
\end{aligned}$$

as desired. □

Definition 13.2. Let $A \in M_n(F)$, then the adjugate matrix $\operatorname{adj} A$ of A is the $n \times n$ matrix with entries

$$(\operatorname{adj} A)_{ji} = (-1)^{i+j} \det A_{\widehat{ij}} = \det(A^{(1)}, \dots, A^{(j-1)}, e_i, A^{(j+1)}, \dots, A^{(n)})$$

Theorem 13.2. *Let $A \in M_n(F)$, then $\operatorname{adj}(A)A = (\det A)I$ where I is the identity matrix.*

In particular $A^{-1} = (\det A)^{-1} \operatorname{adj} A$ whenever A is invertible.

Proof. For any j, the preceding lemma translates to

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

Now for $j < k$, let A' be the matrix obtained by putting $A^{(k)}$ in the place of $A^{(j)}$, then

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

which implies the result. □

Proposition 13.3. *Let $A \in M_n(F)$ be invertible, and let $b \in F^n$, then the unique solution to $Ax = b$ is given by*

$$x_i = \frac{1}{\det A} \det A_{\hat{i}b}$$

where $A_{\hat{i}b}$ is the matrix obtained by replacing the i^{th} column of A by b.

Proof. As A is invertible, such an x exists and is unique. Let x be a solution, then note that

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

just as we wanted. □

14 Eigenvectors, Eigenvalues and Diagonal Matrices

This is the first step into the wonderful land of diagonalisation of endomorphisms. Consider a vector space V over F with $\dim V = n < \infty$ and let $\alpha : V \to V$ be an endomorphism. The general problem is whether we can find a basis B of V such that $[\alpha]_B$ is in a nice enough form. In other words, by our change-of-basis formula, we want to know when can a matrix be conjugate to another matrix in a nice form.

Definition 14.1. 1. $\alpha \in L(V) = L(V, V)$ is diagonalisable if there exists a basis B of V such that $[\alpha]_B$ is diagonal, i.e. $([\alpha]_B)_{ij} = 0$ for $i \neq j$.
2. $\alpha \in L(V)$ is triangulable if there exists a basis B of V such that $[\alpha]_B$ is (upper) triangular.

Remark. A matrix is diagonalisable (resp. triangulable) iff it is conjugate to a diagonal (resp. triangular) matrix.

Definition 14.2. 1. $\lambda \in F$ is an eigenvalue of α if $\alpha(v) = \lambda v$ for some $v \neq 0$.
2. $v \in V$ is an eigenvector of α if $v \neq 0$ and there exists some $\lambda \in F$ such that $\alpha(v) = \lambda v$.
3. $V_\lambda = \{v \in V : \alpha(v) = \lambda v\} \leq V$ is called the eigenspace of α associated to λ.

Lemma 14.1. *If $\alpha \in L(V)$ and $\lambda \in F$, then λ is an eigenvalue iff* $\det(\alpha - \lambda \operatorname{id}_V) = 0$.

Proof. Follows from the fact that matrices with nonzero determinant have zero kernel. ☐

Remark. If $\alpha(v_j) = \lambda v_j$ for $v_j \neq 0$, then completing it into a basis $B = \{v_1, \ldots, v_j, \ldots, v_n\}$ of V gives $([\alpha]_B)_{ij} = \lambda \delta_{ij}$.

Recall that for a field F, a polynomail in F is $f(t) = a_n t^n + \cdots + a_0 \in F[t]$ with $a_i \in F$. Let $\deg f$ be the largest m such that $a_m \neq 0$, then we know that $\deg(f + g) \leq \max\{\deg f, \deg g\}$ and $\deg(fg) = \deg(f) + \deg(g)$. We say λ is a root of f iff $f(\lambda) = 0$, and $g(t)$ divides $f(t)$ if there is some $q(t) \in F[t]$ such that $f(t) = g(t)q(t)$.

Lemma 14.2. *If λ is a root of f, then $x - \lambda$ divides f.*

Proof. Write $f(t) = f(t) - f(\lambda)$ and factrorise. ☐

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

Example 14.1. $f(t) = (t - 1)^2 (t - 2)^3$ has roots 1 with multiplicity 2 and 2 with multiplicity 3.

Corollary 14.3. *A polynomial of degree n has at most n roots, counted with multiplicity.*

Proof. Induction. ☐

Corollary 14.4. *For polynomials f_1, f_2 of degree less than n with $f_1(t_i) = f_2(t_i)$ for distinct $t_1, \ldots, t_n$, then $f_1 = f_2$.*

Proof. $\deg(f_1 - f_2) < n$. ☐

Theorem 14.5 (Fundamental Theorem of Algebra)**.** *Any polynomial $f \in \mathbb{C}[t]$ of positive degree has a root.*

Proof. Omitted. ☐

Consequently, f has exactly $\deg f$ many roots counted with multplicity. This means that any $f \in \mathbb{C}[t]$ can be written as

$$f(t) = c\prod_{i=1}^{n}(t-\lambda_i)^{\alpha_i}, c, \lambda_i \in \mathbb{C}, \alpha_i \in \mathbb{N}, \sum_{i=1}^{n} \alpha_i = \deg f$$

Definition 14.3. For $\alpha \in L(V)$, the characteristic polynomial of α is $\chi_\alpha(t) = \det(\alpha - t\,\mathrm{id}_V) \in F[t]$.

Remark. Conjugate matrices then have the same characteristic polynomial.

Theorem 14.6. *$\alpha \in L(V)$ is triangulable iff*

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

for some $c, \lambda_i \in F$.

Consequently, any matrix in $\mathbb{C}$ is triangulable.

Proof. The "only if" part is trivial. For the "if" direction, we do induction on $n = \dim V$. The $n = 1$ case is trivial. For $n > 1$, there is λ such that $\chi_\alpha(\lambda) = 0$ by assumption. Let $\{v_1, \dots, v_k\}$ be a basis of $U = V_\lambda$ and extend it to a basis $B = \{v_1, \dots, v_n\}$ of V. We then have

$$[\alpha]_B = \left(\begin{array}{c|c} \lambda I_k & * \\ \hline 0 & C \end{array}\right)$$

So the induced endomorphism $\bar{\alpha} : V/U \to V/U$ has matrix C under the basis $\{v_{k+1}+U, \dots, v_n+U\}$. Then by the induction hypothesis, we can choose another set of basis $\{\tilde{v}_{k+1}+U, \dots, \tilde{v}_n+U\}$ so that C is triangular. Hence α is triangular under the basis $\{v_1, \dots, v_k, \tilde{v}_{k+1}, \dots, \tilde{v}_n\}$. This completes the proof. □

Lemma 14.7. *Suppose V is a vector space over $F = \mathbb{R}$ or $\mathbb{C}$ such that $\dim V = n < \infty$ and suppose $\alpha \in L(V)$ is an endomorphism with matrix A. Say $\chi_\alpha(t) = (-1)^n t^n + c_{n-1}t^{n-1} + \cdots + c_0$, then $c_0 = \det A$, $c_{n-1} = (-1)^{n-1}\operatorname{tr} A$.*

Proof. $\det A = \chi_A(0) = c_0$. For c_{n-1}, note that the statement is true for triangular A, so we are done by the preceding theorem. □

15 Diagonalisation Criterion and Minimal Polynomial

For a polynomial $p(t) \in F[t]$ in the form $p(t) = a_n t^n + \cdots + a_0$ for $a_i \in F$ and $A \in M_n(F)$, we define $p(A) = a_n A^n + \cdots + a_0 I$. Similarly for $\alpha \in L(V)$, we write $p(\alpha) = a_n\alpha^n + \cdots + a_0\,\mathrm{id}_V$ where α^i is α composed with itself for i times.

Theorem 15.1. *Let V be a finite-dimensional vector space over F and $\alpha \in L(V)$. Then α is diagonalisable iff there exists a polynomial $p \in F[t]$ which is the product of distinct linear factors in $F[t]$ such that $p(\alpha) = 0$.*

Proof. Suppose α is diagonalisable with distinct eigenvalues $\lambda_1, \dots, \lambda_k$ which might or might not have unit multiplicity. Then take $p(t) = (t-\lambda_1)\cdots(t-\lambda_k)$. Let B be the basis in which α is diagonal, then for any $v \in B$ we have $\alpha(v) = \lambda_i v$ for some i, which means $(\alpha - \lambda_i \operatorname{id}_V)v = 0$. Then note that the factors $\alpha - \lambda_i \operatorname{id}_V$ always commute with each other, which means $p(\alpha)v = 0$ for all $v \in B$, hence $p(\alpha) = 0$.
Conversely, suppose $p(\alpha) = 0$ for some $p(t) = (t-\lambda_1)\cdots(t-\lambda_k)$ with $\lambda_i \neq \lambda_j$ whenever $i \neq j$. Let $V_{\lambda_i} = \ker(\alpha - \lambda_i \operatorname{id}_V)$. We claim that $V = \bigoplus_i V_{\lambda_i}$. Indeed, take

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

Then $q_j(\lambda_i) = \delta_{ij}$. Consider $q(t) = q_1(t) + \cdots + q_k(t)$, then q has degree at most $k-1$ and $q(\lambda_i) = 1$ for all i, which then means $q(t) = 1$ for all t. Let $\pi_j = q_j(\alpha) \in L(V)$, then by construction $\pi_1 + \cdots + \pi_k = q(\alpha) = \operatorname{id}_V$. Now,

$$(\alpha - \lambda_j \operatorname{id}_V) q_j(\alpha)(v) = \frac{1}{\prod_{i \neq j}(\lambda_j - \lambda_i)} p(\alpha)(v) = 0$$

This means that for any j, $\pi_j(v) \in V_{\lambda_j}$ for any v, so V is indeed the sum of all V_{λ_i}.
It remains to prove that the sum is direct. Take $v \in V_{\lambda_j} \cap \sum_{i \neq j} V_{\lambda_i}$, then since $v \in V_{\lambda_j}$,

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

But also $v \in \sum_{i \neq j} V_{\lambda_i}$, so $\pi_j(v) = 0$, which implies $v = 0$, so the sum is direct, and hence α is diagonalisable. □

Remark. We have shown in part of our proof above that if $\lambda_1, \dots, \lambda_k$ are k distinct eigenvalues of α, then the sum $\sum_i V_{\lambda_i}$ is always direct. So the only way diagonalisation fails is when the sum of eigenspaces is properly contained in V.

Corollary 15.2. *If $A \in M_n(\mathbb{C})$ has finite order, then A is diagonalisable.*

Proof. $t^m - 1$ is the product of $t - \zeta_m^j$ where ζ_m^j are the m^{th} roots of unity. □

Theorem 15.3 (Simultaneous Diagonalisation)**.** *Let $\alpha, \beta \in L(V)$ be diagonalisable, then α, β are simultaneously diagonalisable, that is there exists a basis in which both matrices are diagonal, iff α, β commute.*

Proof. If there is a basis B in which $[\alpha]_B = D_1, [\beta]_B = D_2$ are diagonal matrices, then obviously D_1, D_2 commutes, thus so does α and β.
Conversely, if α, β are diagonalisable and commute, then $V = \bigoplus_i V_{\lambda_i}$ where λ_i are distinct eigenvalues of α. Now $\beta(V_{\lambda_j}) \leq V_{\lambda_j}$ since for any $v \in V_{\lambda_j}$ we have $\alpha \circ \beta(v) = \beta \circ \alpha(v) = \beta(\lambda_j v) = \lambda_j \beta(v)$. Take a polynomial p that is the product of distinct linear factors such that $p(\beta) = 0$. Consequently $p(\beta|_{V_{\lambda_i}}) = 0$, so $\beta|_{V_{\lambda_i}}$ is diagonalisable. Then the union of bases in V_{λ_i} so that $\beta|_{V_{\lambda_i}}$ is diagonal is a basis of V that makes α, β both diagonal. □

Recall that we can do division algorithm on polynomials, so if we have $a(t), b(t) \in F[t]$ for nonconstant b, we have $q(t), r(t) \in F[t]$ such that $\deg r < \deg b$ and $a = qb + r$.

Definition 15.1. Let V be a vector space over F and $\alpha \in L(V)$. A polynomial $m_\alpha(t) \in F[t]$ is a minimal polynomial of α if $m_\alpha(\alpha) = 0$ and $\deg m_\alpha$ is minimal.

Remark. This is well-defined as there must exists a polynomial $m(t) \in F[t]$ such that $m(\alpha) = 0$ by considering the linearly dependent set $\{\mathrm{id}, \alpha, \dots, \alpha^{n^2}\}$ in $L(V)$.

Note also that we have defined it as "the" minimal polynomial instead of "a" minimal polynomial. To justify it, we have

Lemma 15.4. *For $\alpha \in L(V)$, let m_α be a minimal polynomial of α and $p(t) \in F[t]$, then $p(\alpha) = 0$ iff $m_\alpha | p$.*

In particular, minimal polynomial is unique up to a nonzero constant.

Proof. The "if" direction is trivial. For the "only if" direction, we can write $p = m_\alpha q + r$ for some polynomials q, r with $\deg r < \deg m_\alpha$. $r(\alpha) = 0$ by assumption, so by minimality of $\deg m_\alpha$ we must have $r = 0$, i.e. $m_\alpha | p$. □

Example 15.1. Consider $V = F^2$ and

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

They $m_A = t - 1$ and $m_B = (t-1)^2$. Consequenly A is diagonalisable but B is not.

16 Cayley-Hamilton Theorem and Multiplicity of Eigenvalues

Theorem 16.1 (Cayley-Hamilton Theorem)**.** *Let V be a finite dimensional vector space over F and $\alpha \in L(V)$ with characteristic polynomial $\chi_\alpha(t) = \det(\alpha - t\,\mathrm{id})$, then $\chi_\alpha(\alpha) = 0$.*

Consequently, $m_\alpha | \chi_\alpha$.

Proof for $F = \mathbb{C}$. We know that α is triangulable, so there is a basis B such that it has matrix

$$[\alpha]_B = \begin{pmatrix} a_1 & & * \\ & \ddots & \\ 0 & & a_n \end{pmatrix}$$

Therefore $\chi_\alpha(t) = (t - a_1) \cdots (t - a_n)$ (up to sign). But then easily $\chi_\alpha(\alpha) = \chi_\alpha([\alpha]_B) = 0$. □

Proof for the General Case. For $A \in M_n(F)$, we write

$$(-1)^n \chi_A(t) = \det(t\,\mathrm{id} - A) = t^n + a_{n-1} t^{n-1} \cdots + a_0$$

for some $a_i \in F$. Now if $\mathrm{adj}(t\,\mathrm{id} - A) = B_{n-1} t^{n-1} + \cdots B_0$ for matrices B_i, then

$$(t\,\mathrm{id} - A)(B_{n-1} t^{n-1} + \cdots B_0) = (t^n + a_{n-1} t^{n-1} \cdots + a_0)\,\mathrm{id}$$

Equating the coefficients gives

$$\mathrm{id} = B_{n-1}, a_{n-1}\,\mathrm{id} = B_{n-2} - AB_{n-1}, \dots, a_0\,\mathrm{id} = -AB_n$$

Therefore

$$\begin{aligned}(-1)^n \chi_A(A) &= A^n + a_{n-1}A^{n-1} \cdots + a_0\,\mathrm{id} \\ &= A^n B_{n-1} + A^{n-1}(B_{n-2} - AB_{n-1}) + \cdots + A^0(-AB_0) \\ &= 0\end{aligned}$$

by telescoping. □

Definition 16.1. Let $\alpha \in L(V)$ and λ an eigenvalue of α, then the algebraic multiplicity a_λ of λ is the multiplicity of λ as a root of $\chi_\alpha(t)$.
The geometric multiplicity g_λ of λ is $\dim\ker(\alpha - \lambda\,\mathrm{id})$.

Remark. Obviously $a_\lambda, g_\lambda \geq 1$.

Lemma 16.2. $g_\lambda \leq a_\lambda$.

Proof. Let $\{v_1, \dots, v_{g_\lambda}\}$ be a basis of $V_\lambda = \ker(\alpha - \lambda\,\mathrm{id})$ and extend it to a basis $B = \{v_i\}$ of V. Then

$$[\alpha]_B = \begin{pmatrix} \lambda\,\mathrm{id}_{g_\lambda} & * \\ 0 & A_1 \end{pmatrix}$$

for some A_1. Then

$$\det(\alpha - \lambda\,\mathrm{id}) = \det \begin{pmatrix} (\lambda - t)\,\mathrm{id}_{g_\lambda} & * \\ 0 & A_1 - t\,\mathrm{id} \end{pmatrix} = (\lambda - t)^{g_\lambda} \chi_{A_1}(t)$$

which implies the claim. □

Lemma 16.3. *Let λ be an eigenvalue of α and write c_λ as the multiplicity of λ as a root of the minimal polynomial m_α. Then $1 \leq c_\lambda \leq a_\lambda$.*

Proof. $c_\lambda \leq a_\lambda$ is obvious as $m_\alpha | \chi_\alpha$. To see $c_\lambda \geq 1$, as λ is an eigenvalue, we can find $v \neq 0$ such that $\alpha(v) = \lambda v$, so $\alpha^p(v) = \lambda^p v$. Hence $0 = m_\alpha(\alpha)v = (m_\alpha(\lambda))v$ which means $m_\alpha(\lambda) = 0$, hence $c_\lambda \geq 1$. □

Example 16.1. 1. Consider

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

then $\chi_A(t) = (t-1)^2(t-2)$. So $m_A(t)$ is either $(t-1)^2(t-2)$ or $(t-1)(t-2)$ by the preceding lemma. Indeed the latter works and has a smaller degree, hence $m_A(t) = (t-1)(t-2)$.
2. Let A be the Jordan block

$$A = \begin{pmatrix} \lambda & 1 & & \\ & \ddots & \ddots & \\ & & \lambda & 1 \\ & & & \lambda \end{pmatrix}$$

Then as one can check, $g_\lambda = 1, a_\lambda = c_\lambda = n$.
3. Take $A = \lambda\,\mathrm{id}$, then $g_\lambda = a_\lambda = n$ and $c_\lambda = 1$.

Lemma 16.4. *Take $F = \mathbb{C}$, V a finite dimensional vector space over F and $\alpha \in L(V)$, then the followings are equivalent:*
(i) α is diagonalisable.
(ii) For any eigenvalue λ of α we have $a_\lambda = g_\lambda$.
(iii) For any eigenvalue λ of α we have $c_\lambda = 1$.

Proof. We already know that (i) is equivalent to (iii) by Theorem 15.1. To see (i) is equivalent to (ii), let $\lambda_1, \dots, \lambda_k$ be distinct eigenvalues of α. We have already seen that α is diagonalisable iff $V = \bigoplus_i V_{\lambda_i}$. But $\dim V = n = \deg \chi_\alpha = \sum_i a_{\lambda_i}$ by FTA and $\dim \bigoplus_i V_{\lambda_i} = \sum_i g_\lambda$. So α is diagonalisable iff $\sum_i g_{\lambda_i} = \sum_i a_{\lambda_i}$ iff $a_{\lambda_i} = g_{\lambda_i}$ for all i by Lemma 16.2. □

17 The Jordan Normal Form

We are interested in how nice a matrix can an arbitrary $\alpha \in L(\mathbb{C}^n)$ possibly have.

Definition 17.1. Let $A \in M_n(\mathbb{C})$. We say A is in Jordan Normal Form (JNF) if it is a block diagonal matrix

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

for $\{\lambda_i\} \in \mathbb{C}$ not necessarily distinct, $\sum_i n_i = n$ and $J_r(\lambda) \in M_r(\mathbb{C})$ are Jordan blocks of the form

$$J_r(\lambda) = \begin{pmatrix} \lambda & 1 & & \\ & \ddots & \ddots & \\ & & \lambda & 1 \\ & & & \lambda \end{pmatrix}$$

Theorem 17.1. *Every matrix $A \in M_n(\mathbb{C})$ is similar to a matrix in JNF which is unique up to reordering the Jordan blocks.*

Proof. Omitted. □

Example 17.1. For $n = 2$, the possible JNFs are simply $(\lambda, \lambda_1, \lambda_2 \in F, \lambda_1 \neq \lambda_2)$

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

with minimal polynomials $(t - \lambda_1)(t - \lambda_2), t - \lambda, (t - \lambda)^2$ respectively.
For $n = 3$, the JNFs (and their respective minimal polynomials) are, up to reordering of the blocks, $(\lambda, \lambda_1, \lambda_2, \lambda_3 \in F,\ \lambda_1, \lambda_2, \lambda_3$ all distinct)

$$\begin{pmatrix} \lambda_1 & & \\ & \lambda_2 & \\ & & \lambda_3 \end{pmatrix} : (t - \lambda_1)(t - \lambda_2)(t - \lambda_3); \begin{pmatrix} \lambda_1 & & \\ & \lambda_2 & \\ & & \lambda_2 \end{pmatrix} : (t - \lambda_1)(t - \lambda_2)$$

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

$$\begin{pmatrix}\lambda & & \\ & \lambda & 1 \\ & & \lambda\end{pmatrix} : (t-\lambda)^2; \begin{pmatrix}\lambda & 1 & \\ & \lambda & 1 \\ & & \lambda\end{pmatrix} : (t-\lambda)^3$$

Theorem 17.2 (Generalised Eigenspace Decomposition)**.** *Let V be a finite dimensional vector space over $\mathbb{C}$ and $\alpha \in L(V)$. Let $\lambda_1, \dots, \lambda_k$ be distinct eigenvalues of α such that $m_\alpha(t) = (t-\lambda_1)^{c_1} \cdots (t-\lambda_k)^{c_k}$, then*

$$V = \bigoplus_{i=1}^k V_j, V_j = \ker((\alpha - \lambda_j \operatorname{id})^{c_j})$$

Here V_j is called the generalised eigenspace.

Remark. When α is diagonalisable, then $c_j = 1$ for all j, consequently $V = \bigoplus_j \ker(\alpha - \lambda_j \operatorname{id})$ as we already know.

Proof. Define $p_j(t) = \prod_{i \neq j}(t - \lambda_i)^{c_i}$, then $\{p_j\}$ has no common factor, so we can find $q_1, \dots, q_k$ such that $q_1 p_1 + \cdots + q_k p_k = 1$. Define $\pi_j = q_j p_j(\alpha)$, then it follows that $\sum_j \pi_j = \operatorname{id}$. Also $(\alpha - \lambda_j \operatorname{id})^{c_j} \pi_j = 0$, so $\operatorname{Im} \pi_j \subset V_j$, hence V is the sum of all V_j. To see this sum is direct, simply observe that $\pi_i \pi_j = \delta_{ij} \pi_i$ for all i, j, so $\pi_i|_{V_j} = \delta_{ij} \operatorname{id}$. This completes the proof. □

Remark. 1. This decomposition allows us to reduce the proof of JNF to just one eigenvalue, which can be done via the study of nilpotent matrices. [1] The relation is found from the observation

$$(J_m(\lambda) - \lambda \operatorname{id})^k = \begin{cases} \begin{pmatrix} 0 & I_{m-k} \\ 0 & 0 \end{pmatrix}, \text{ if } k < m \\ 0, \text{ otherwise} \end{cases}$$

2. We can very easily compute a_λ, g_λ and c_λ if we know the JNF. Indeed, a_λ is the sum of sizes if the blocks with eigenvalue λ, g_λ is the number of Jordan blocks with eigenvalue λ and c_λ is the size of the largest Jordan block with eigenvalue λ. This can (sometimes) be used to compute the JNF as well.

Example 17.2. Take

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

We want to find a basis in which A is in JNF. Now $\chi_A(t) = m_A(t) = (t-1)^2$, so the JNF is in the form

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

We want the basis, which naturally consists of eigenvectors. Indeed, $\ker(A - \operatorname{id})$ is spanned by $(1, -1)^\top$. Choose v_2 such that $(A - \operatorname{id})v_2 = v_1$, which is nonunique but we can take $v_2 = (-1, 0)^\top$. So take the basis $\{v_1, v_2\}$ works. To put it concretely,

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

[1] That is if you want to do it the linear algebra way – I like the $\mathbb{C}[X]$-module approach more.

18 More on Bilinear Forms

Guess what? We are back to bilinear forms again! But this time, we are interested in bilinear forms $\phi : V \times V \to F$ for a finite dimensional vector space V over F. For a basis B of V, we write $[\phi]_B = [\phi]_{B,B}$.

Lemma 18.1. *Let B, B' be bases of V and $P = [\mathrm{id}]_{B',B}$, then $[\phi]_{B'} = P^\top [\phi]_B P$.*

Proof. Proposition 10.6. □

Definition 18.1. Matrices $A, B \in M_n(F)$ are congruent if there is invertible P such that $A = P^\top BP$.

Remark. Easily congruence is an equivalence relation.

Definition 18.2. A bilinear form ϕ in V is symmetric if $\phi(u, v) = \phi(v, u)$ for any $u, v \in V$.

Remark. ϕ is symmetric iff $[\phi]_B$ is symmetric in some basis B.
If we want to diagonalise ϕ, then it has to be symmetric by the preceding lemma.

Definition 18.3. A map $Q : V \to F$ is a quadratic form iff there exists a bilinear form $\phi : V \times V \to F$ such that $Q(u) = \phi(u, u)$.

Remark. If $B = \{e_i\}_i$ and $A = [\phi]_B = (\phi(e_i, e_j))_{i,j}$, then for $u = \sum_i x_i e_i$ we have

$$Q(u) = \phi\left(\sum_{i=1}^n x_i e_i, \sum_{i=1}^n x_i e_i\right) = \sum_{i=1}^n \sum_{j=1}^n x_i x_j \phi(e_i, e_j) = x^\top Ax$$

where $x = (x_1, \dots, x_n)^\top$. Also observe that $x^\top Ax = x^\top Sx$ where $S = (A + A^\top)/2$ is symmetric.

Proposition 18.2. *If $Q : V \times V \to F$ is a quadratic form, then there exists a unique symmetric bilinear form $\phi : V \times V \to F$ such that $Q(u) = \phi(u, u)$ for all $u \in V$.*

Proof. We know that $Q(u) = \psi(u, u)$ for some bilinear ψ. Take $\phi = (\psi + \psi^\top)/2$ where $\psi^\top(u, v) = \psi(v, u)$ works. To see it is unique, just observe that for a symmetric ϕ,

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

which implies that necessarily $\phi(u, v) = (Q(u + v) - Q(u) - Q(v))/2$ (known as the polarisation identity), so in particular ϕ is uniquely determined. □

Theorem 18.3 (Diagonalisation of Bilinear Forms)**.** *Let $\phi : V \times V \to F$ be a symmetric bilinear form and* $\dim V < \infty$*, then there exists a basis of V such that $[\phi]_B$ is diagonal.*

Proof. We proceed by induction on $n = \dim V$. If $\phi(u, u) = 0$ for any $u \in V$, then ϕ is identically zero by the polarisation identity. Otherwise, we can always find $u \in V \setminus \{0\}$ such that $\phi(u, u) \neq 0$. Write $u = e_1$ and define

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

So $\dim U = n - 1$ and $U + e_1 = U \oplus e_1$. Pick a basis $\{e_2, \dots, e_n\}$ of U such that $\phi|_U$ is diagonal in this basis, which is possible by induction hypothesis. Then ϕ is diagonal in $\{e_1, \dots, e_n\}$. □

Example 18.1. Take $V = \mathbb{R}^3$ and

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

By inspection, if we take

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

then $Q(x) = x^\top Ax$. Of course, we can follow the algorithm illustrated in the proof. We can alternatively complete the square to get

$$\begin{aligned} Q(x) &= x_1^2 + x_2^2 + 2x_3^3 + 2x_1x_2 + 2x_1x_3 - 2x_2x_3 \\ &= (x_1 + x_2 + x_3)^2 + (x_3 - 2x_2)^2 - (2x_2)^2 \end{aligned}$$

then under the new basis induced by the change of variables $x_1' = x_1 + x_2 + x_3, x_2' = x_3 - 2x_2, x_3' = 2x_2$, A has the matrix

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

More concretely, we have $A' = P^\top AP$ where

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

19 Sylvester's Law, Sesquilinear Forms

We start by looking at some immediate corollaries of Theorem 18.3.

Corollary 19.1. *For a finite dimensional complex vector space V and a symmetric bilinear form ϕ on V, there is a basis B of V such that*

$$[\phi]_B = \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}$$

where $r = r(\phi)$.

Proof. Square roots always exist as $F = \mathbb{C}$. □

Corollary 19.2. *Every symmetric matrix in $\mathbb{C}$ is congruent to a unique matrix of the form*

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

Proof. Immediate. □

Corollary 19.3. *If $F = \mathbb{R}$, $\dim V = n < \infty$ and ϕ a symmetric bilinear form of V, then there exists a basis $\{v_1, \dots, v_n\}$ of V such that*

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

Proof. Every positive number in $\mathbb{R}$ has a square root. $\square$

Definition 19.1. $s(\phi) = p - q$ is called the signature of the real symmetric bilinear form ϕ.

To see it is well-defined,

Theorem 19.4 (Sylvester's Law of Inertia)**.** *If a real symmetric bilinear form ϕ has*

$$[\phi]_B = \begin{pmatrix} I_p & & \\ & -I_q & \\ & & 0 \end{pmatrix}, [\phi]_{B'} = \begin{pmatrix} I_{p'} & & \\ & -I_{q'} & \\ & & 0 \end{pmatrix}$$

then $p = p', q = q'$.

Definition 19.2. Let ϕ be a real symmetric bilinear form. We say that ϕ is positive semidefinite if $\phi(u, u) \geq 0$ for any $u \in V$, and is positive definite if $\phi(u, u) > 0$ for any $u \in V \setminus \{0\}$. Similarly, ϕ is positive semidefinite if $\forall u \in V, \phi(u, u) \leq 0$ for any $u \in V$, and negative definite if $\forall u \in V \setminus \{0\}, \phi(u, u) < 0$.

Example 19.1. The matrix

$$\begin{pmatrix} I_p & 0 \\ 0 & 0 \end{pmatrix} \in M_n(\mathbb{R})$$

is always positive semidefinite, and is positive definite iff $p = n$.

Proof. Indeed p is the largest dimension of subspace of V in which ϕ is positive definite. Similarly q is the largest dimension of a subspace in which ϕ is negative definite. These descriptions are independent of the choice of basis, so we are done. $\square$

Definition 19.3. The kernel of the bilinear form $\phi : V \times V \to F$ is the set $K(\phi) = \{v \in V : \forall u \in V, \phi(u, v) = 0\}$.

Remark. 1. $\dim K(\phi) + r(\phi) = 0$.
2. For $F = \mathbb{R}$, we now know from the preceding theorem that there is a subspace T of dimension $n - (p + q) + \min\{p, q\}$ such that $\phi|_T = 0$. More over, this can easily be shown to be the largest dimension such that such a subspace T exists.

Recall that the standard inner product on $\mathbb{C}^n$, that is

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

is not a bilinear form.

Definition 19.4. Let V, W be vector spaces over $\mathbb{C}$. A map $\phi : V \times W \to \mathbb{C}$ is a sesquilinear form if for any $w \in W$, $\phi(\cdot, w)$ is linear and for any $v \in V, \lambda_1, \lambda_2 \in \mathbb{C}, w_1, w_2 \in W$,

$$\phi(v, \lambda_1 w_1 + \lambda_2 w_2) = \bar{\lambda}_1 \phi(v, w_1) + \bar{\lambda}_2 \phi(v, w_2)$$

Definition 19.5. With notation as above, for bases $B = \{v_1, \ldots, v_m\}$ of V and $C = \{w_1, \ldots, w_n\}$ of W, the matrix of ϕ is $([\phi]_{B,C})_{ij} = (\phi(v_i, w_j))$.

Lemma 19.5. $\phi(v,w) = [v]_B^\top [\phi]_{B,C} \overline{[w]_C}$.

Proof. Expand. □

Lemma 19.6. *If B, B' are bases of V and C, C' of W and $P = [\mathrm{id}_V]_{B',B}, Q = [\mathrm{id}_W]_{C',C}$, then $[\phi]_{B',C'} = P^\top [\phi]_{B,C} \bar{Q}$*

Proof. Analogous to the bilinear case. □

20 Hermitian Forms and Real Skew-Symmetric Forms

Definition 20.1. A sesquilinear form $\phi : V \times V \to \mathbb{C}$ is called Hermitian if $\phi(u,v) = \overline{\phi(v,u)}$.

Remark. In particular, $\phi(u,u) = \overline{\phi(u,u)}$, so $\phi(u,u)$ is real. Moreover, for any $\lambda \in \mathbb{C}$ we have $\phi(\lambda u, \lambda u) = |\lambda|^2 \phi(u,u)$. Therefore it makes sense to talk about positive/negative (semi)definite Hermitian forms.

Lemma 20.1. *A sesquilinear form $\phi : V \times V \to \mathbb{C}$ is Hermitian iff for any basis B of V, $[\phi]_B = \overline{[\phi]}_B^\top$.*

Proof. If ϕ is Hermitian, then write $A = [\phi]_B = (a_{ij})_{i,j} = (\phi(e_i, e_j))_{i,j}$ where we have $a_{ji} = \phi(e_j, e_i) = \overline{\phi(e_i, e_j)} = \bar{a}_{ij}$. Conversely if $[\phi]_B = A$ with $A = (a_{ij})_{ij} = \bar{A}^\top$ and $u = \sum_i \lambda_i e_i, v = \sum_i \mu_i e_i$, then

$$\begin{aligned}\phi(u,v) &= \phi\left(\sum_{i=1}^n \lambda_i e_i, \sum_{j=1}^n \mu_j e_j\right) = \sum_{i=1}^n \sum_{j=1}^n \lambda_i \bar{\mu}_j a_{ij} \\ &= \overline{\sum_{i=1}^n \sum_{j=1}^n \bar{\lambda}_i \mu_j a_{ji}} \\ &= \overline{\phi\left(\sum_{j=1}^n \mu_j e_j, \sum_{i=1}^n \lambda_i e_i\right)} \\ &= \overline{\phi(v,u)}\end{aligned}$$

So ϕ is Hermitian. □

The polarisation identity becomes

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

for a Hermitian ϕ and $Q(w) = \phi(w,w)$.

Theorem 20.2 (Hermitian Formulation of Sylvester's Law)**.** *Let V be an n-dimensional vector space over $\mathbb{C}$ and $\phi : V \times V \to \mathbb{C}$ a Hermitian form on V, then V has a basis $\{v_1, \dots, v_n\}$ such that*

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

where p, q depends only on ϕ.

The proof is nearly identical to what we did for the real symmetric case.

Proof. If $\phi = 0$ then we are done. Otherwise, by the polarisation identity, there exists $e_1 \neq 0$ such that $\phi(e_1, e_1) \neq 0$. Set $v_1 = e_1/\sqrt{|\phi(e_1, e_1)|}$, then we get $\phi(v_1, v_1) = \pm 1$. Consider $W = \{w \in V : \phi(v, w) = 0\}$, then easily $V = \langle\{v_1\}\rangle \oplus W$. Then we can do induction on the dimension to show that ϕ is diagonal in some basis, which implies the existence of p, q. The uniqueness follows from the observation that p (resp. q) is the minimal dimension of a subspace on which ϕ is positive (resp. negative) definite. $\square$

Definition 20.2. Let V be a vector space over $\mathbb{R}$. A bilinear form on a real vector space is skew-symmetric if $\phi(u, v) = -\phi(v, u)$ for all $u, v \in V$.

Remark. 1. For any $u \in V$, $\phi(u, u) = -\phi(u, u)$ therefore $\phi(u, u) = 0$.
2. The definition is equivalent to say that for any basis B of V we have $[\phi]_B = -[\phi]_B^\top$.
3. For any $A \in M_n(\mathbb{R})$, we can decompose it

$$A = \frac{A + A^\top}{2} + \frac{A - A^\top}{2}$$

into symmetric and skew-symmetric parts.

Theorem 20.3 (Sylvester Form)**.** *Let ϕ be a skew-symmetric bilinear form over a real vector space V with $\dim V = n < \infty$, then there is a basis $B = \{v_1, w_1, \dots, v_m, w_m, v_{2m+1}, \dots, v_n\}$ of V such that*

$$[\phi]_B = \begin{pmatrix} A & & & \\ & \ddots & & \\ & & A & \\ & & & 0 \end{pmatrix}, A = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$$

where there are m copies of A.

Proof. Induction on n. If $\phi = 0$, then we are done. Otherwise $\phi \neq 0$, so there is some v_1, w_1 such that $\phi(v_1, w_1) \neq 0$. After rescaling we might as well assume that $\phi(v_1, w_1) = 1$, so correspondingly $\phi(w_1, v_1) = -1$. We know that v_1, w_1 has to be linearly independent as ϕ is skew-symmetric. Let $U = \langle\{v_1, v_2\}\rangle$ and $W = \{v \in V : \phi(v_1, v) = \phi(w_1, v) = 0\}$, then $V = U \oplus W$. We are then done since we can use induction hypothesis on W and $[\phi|_U] = A$. $\square$

Corollary 20.4. *Skew-symmetric bilinear forms have even rank.*

Proof. Immediate. $\square$

Definition 20.3. Let V be a vector space over $\mathbb{R}$ (resp. $\mathbb{C}$). An inner product on V is a positive definite symmetric (resp. Hermitian) form ϕ on V. The pair (V, ϕ) is then called a real (resp. complex) inner product space.

Sometimes we write $\langle u, v\rangle = \phi(u, v)$ if it is understood.

Example 20.1. In $\mathbb{R}^n$, the usual real scalar product is an inner product.
In $\mathbb{C}^n$, the usual complex scalar product is an inner product.
In $C([0,1],\mathbb{C})$ (over $\mathbb{C}$), the form

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

is an inner product for any $w \in C([0,1],\mathbb{R}_+)$.

Definition 20.4. Let $\langle\cdot,\cdot\rangle$ be an inner product, its induced norm is $\|v\| = \sqrt{\langle v, v\rangle}$.

Remark. $\|v\| \geq 0$ and the equality holds iff $v = 0$.

21 Gram-Schmidt and Orthogonal Complement

Here we only interested in inner product spaces over $F = \mathbb{R}$ or $\mathbb{C}$.

Lemma 21.1 (Cauchy-Schwartz Inequality)**.** $|\langle u, v\rangle| \leq \|u\|\|v\|$.

In particular, equality hold iff u, v are linearly dependent.

Proof. For $t \in F$, expanding $\langle tu - v, tu - v\rangle \geq 0$ gives

$$0 \leq |t|^2\|u\|^2 - 2\operatorname{Re}(t\langle u, v\rangle) + \|v\|^2$$

Picking $t = \overline{\langle u, v\rangle}/\|u\|^2$ ends the proof. □

Corollary 21.2 (Triangle Inequality)**.** $\|u + v\| \leq \|u\| + \|v\|$.

Consequently $\|\cdot\|$ is a indeed a norm.

Proof. Square both sides and use Cauchy-Schwartz. □

Definition 21.1. Fix an inner product $\langle\cdot,\cdot\rangle$. A set $\{e_1, \ldots, e_k\}$ of vectors in V is orthogonal if $\langle e_i, e_j\rangle = 0$ for $i \neq j$ and orthonormal if in addition they all have norm 1, that is $\langle e_i, e_j\rangle = \delta_{ij}$.

Note that both notion depends on our choice of inner product.

Lemma 21.3. *A set of orthogonal vectors $e_1, \ldots, e_k$ has to be linearly independent. In fact, if $v = \sum_i \lambda_i e_i$ then $\lambda_i = \langle v, e_i\rangle/\|e_i\|$.*

Proof. Immediate from bilinearity. □

Lemma 21.4 (Parseval's Identity)**.** *If V is a finite dimensional inner product space and $e_1, \ldots, e_n$ is an orthonormal basis, then*

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

Proof. Obvious from the preceding lemma. □

In particular, in an orthogonal basis, $\|v\|^2 = \sum_i |\langle v, e_i\rangle|^2$. Does an orthogonal basis always exist?

Theorem 21.5 (Gram-Schmidt Orthogonalisation)**.** *If we have an inner product space V and a sequence of linearly independent vectors $(v_i)_{i\in I} \in V$ where $I = \{1, 2, \dots\}$ (which may or may not terminate), then there exists a sequence $(e_i)_{i\in I}$ of orthonormal vectors such that $\langle v_1, \dots, v_k\rangle = \langle e_1, \dots, e_k\rangle$ for any $k \in I$.*

Proof. We shall define (e_i) inductively on k. For $k = 1$, just take $e_1 = v_1/\|v_1\|$. Say we have found $e_1, \dots, e_k$, then define

$$e'_{k+1} = v_{k+1} - \sum_{i=1}^{k} \langle v_{k+1}, e_i\rangle e_i, e_{k+1} = \frac{1}{|e'_{k+1}|} e'_{k+1}$$

This is well-defined as (v_i) is linearly independent (so $e'_{k+1} \neq 0$) and it is easy to verify that $\langle v_1, \dots, v_{k+1}\rangle = \langle e_1, \dots, e_{k+1}\rangle$. This completes the proof. □

So not only does there exists such a set of orthonormal vectors, we also get an algorithm to compute it.

Corollary 21.6. *Let V be a finite dimensional inner product space, then any orthonormal set of vectors can be extend to an orthonormal basis of V.*

Proof. Extend it to a basis, then apply the Gram-Schmidt algorithm (which fixes the original set). □

Note. A matrix $A \in M_{m,n}(F)$ has orthogonal columns if $A^\top \bar{A} = I$.

Definition 21.2. $A \in M_n(\mathbb{R})$ is orthogonal if $A^\top A = I$. $A \in M_n(\mathbb{C})$ is unitary if $A^\top \bar{A} = I$.

Proposition 21.7. *Any nonsingular $A \in M_n(\mathbb{R})$ (resp. $M_n(\mathbb{C})$) can be written as $A = RT$ where T is upper-triangular and R is orthogonal (resp. unitary).*

Proof. Do Gram-Schmidt on columns of A. □

Definition 21.3. Let V be an inner product space and $V_1, V_2 \leq V$. We say V is the orthogonal sum of V_1, V_2 (written as $V = V_1 \oplus^\perp V_2$) if $V = V_1 \oplus V_2$ and $\forall v_1, \in V_1, v_2 \in V_2$, we have $\langle v_1, v_2\rangle = 0$.

Definition 21.4. Let V be an inner product space and $W \leq V$. We define $W^\perp = \{v \in V : \forall w \in W, \langle v, w\rangle = 0\}$.

Lemma 21.8. *$W \oplus^\perp W^\perp = V$ if V is finite dimensional.*

Proof. Clearly $W^\perp \leq V$ and by definition the sum $W + W^\perp$ is direct and orthogonal. So it suffices to show that $V = W + W^\perp$, which is obvious since we can obtain a basis of $W^\perp$ of the right size by extending an orthonormal basis on W orthonormally to V. □

22 Orthogonal Complement and Adjoint Map

Definition 22.1. Suppose $V = U \oplus W$. The projection operator $\pi = \pi_W : V \to W$ into W is defined via $u + w \mapsto w$ for any $u \in U, w \in W$.

Easy to see that π is linear and $\pi^2 = \pi$

Remark. We have $\pi_U = \mathrm{id} - \pi_W$.

Of course, in the case where $U = W^\perp$, we can have something better.

Lemma 22.1. *Let V be an inner product space and $W \leq V$ finite dimensional subspace of V, then:*
(a) If $\{e_i\}$ is an orthonormal basis of W, then $\forall v \in V, \pi(v) = \sum_i \langle v, e_i \rangle e_i$.
(b) $\forall v \in V, w \in W, \|v - \pi(v)\| \leq \|v - w\|$ with equality iff $w = \pi(v)$.

Proof. Just observe that $v - \pi(v) \in W^\perp$ which is known to be a complementary subspace of W. This gives (a) immediately, and for (b) we have $\|v - w\|^2 = \|v - \pi(v) + \pi(v) - w\|^2 = \|v - \pi(v)\|^2 + \|\pi(v) - w\|^2 \geq \|v - \pi(v)\|^2$. □

Proposition 22.2. *Let V, W be finite dimensional inner product spaces and $\alpha \in L(V, W)$. Then there is a unique linear map $\alpha^* : W \to V$ such that $\forall v \in V, w \in W, \langle \alpha(v), w \rangle = \langle v, \alpha^*(w) \rangle$. Moreover, if B, C are orthonormal bases of V, W, then $[\alpha^*]_{C,B} = (\overline{[\alpha]_{B,C}})^\top$.*

Proof. Brute-force computation. □

Definition 22.2. This map α^* is called the adjoint of α.

Remark. One might notice that we used the same notation for adjoint and dual of a map. This (intentional) abuse of notation can be justified by considering the linear isomorphisms $\psi_{R,V} : V \to V^*, \psi_{R,W} : W \to W^*$ via $\psi_{R,V}(v) = \langle \cdot, v \rangle, \psi_{R,W}(w) = \langle \cdot, w \rangle$ which immediately satisfies $\alpha^*_{\text{adjoint}} = \psi_{R,V}^{-1} \circ \alpha^*_{\text{dual}} \circ \psi_{R,W}$.

$$\begin{array}{ccc} W^* & \xrightarrow{\alpha^*_{\text{dual}}} & V^* \\ {\scriptstyle \psi_{R,W}}\uparrow & & \uparrow{\scriptstyle \psi_{R,V}} \\ W & \xrightarrow[\alpha^*_{\text{adjoint}}]{} & V \end{array}$$

Definition 22.3. Let V be an inner product space. A map $\alpha \in L(V)$ is self-adjoint if $\alpha = \alpha^*$, i.e. $\forall v, w \in V, \langle \alpha(v), w \rangle = \langle v, \alpha(w) \rangle$.
It is called an isometry if $\alpha^* \circ \alpha = \mathrm{id}$, or $\langle \alpha(v), \alpha(w) \rangle = \langle v, w \rangle$ for any $v, w \in V$.

Remark. By the polarisation identity, α is an isometry iff $\|\alpha(v)\| = \|v\|$ for any $v \in V$.

Lemma 22.3. *Let V be a finite dimensional inner product space over $\mathbb{R}$ (resp. $\mathbb{C}$). Then $\alpha \in L(V)$ is self-adjoint iff for any orthonormal basis B of V, $[\alpha]_B$ is symmetric (resp. Hemitian). It is an isometry iff for any orthonormal basis B of V, $[\alpha]_B$ is orthonormal (resp. unitary).*

Proof. Immediate. □

The collection of isometries are naturally subgroups of $L(V)$.

Definition 22.4. Let V be a finite dimensional inner product space over a field $F = \mathbb{R}$ or $\mathbb{C}$. The subgroup of isometries $\{\alpha \in L(V) : \alpha^* \circ \alpha = \mathrm{id}\} \leq L(V)$ is called the orthogonal group $O(V)$ of V when $F = \mathbb{R}$ and the unitary group $U(V)$ of V when $F = \mathbb{C}$.

Remark. Fix an orthonormal basis $\{e_i\}$ of V. Then there is a one-to-one correspondence between the isometries in V and the orthonormal bases of V via $\alpha \leftrightarrow \{\alpha(e_i)\}$.

23 Spectral Theory

Spectral Theory is the study of spectrum (eigen-stuff) of operators, which is very important in boths maths and physics. Fix an inner product space V. Recall that an opeator $\alpha \in L(V)$ is self-adjoint if $\alpha = \alpha^*$.

Lemma 23.1. *A self-adjoint opeator $\alpha \in L(V)$ has real eigenvalues and eigenvectors with different eigenvalues are orthogonal.*

Proof. If $v \neq 0$ and $\alpha(v) = \lambda v$, then $\lambda \|v\|^2 = \langle \lambda v, v\rangle = \langle \alpha(v), v\rangle = \langle v, \alpha(v)\rangle = \langle v, \lambda(v)\rangle = \bar{\lambda}\|v\|^2$, so $\lambda = \bar{\lambda} \implies \lambda \in \mathbb{R}$.
Now if $v, w \neq 0, \lambda \neq \mu$ have $\alpha(v) = \lambda v, \alpha(w) = \mu w$, then $\lambda, \mu \in \mathbb{R}$ and hence $\lambda\langle v, w\rangle = \langle \alpha(v), w\rangle = \langle v, \alpha(w)\rangle = \langle v, \mu w\rangle = \bar{\mu}\langle v, w\rangle = \mu\langle v, w\rangle \implies \langle v, w\rangle = 0$ since $\lambda \neq \mu$. ☐

Theorem 23.2 (Spectral Theorem for Self-Adjoint Operators in Finite Dimensions)**.** *Let V be a finite dimensional inner product space over $F = \mathbb{R}$ or $\mathbb{C}$ and $\alpha \in L(V)$ is self-adjoint. Then V has an orthonormal basis of eigenvectors of α.*

Consequently, α can be diagonalised in an orthonormal basis of V.

Proof. We proceed by induction on $n = \dim V$. $n = 1$ is trivial. Now assume it is true for $n-1$. Let λ be a root of χ_A (exists by FTA). $\lambda \in \mathbb{R}$ by the preceding lemma, so $\alpha - \lambda\,\mathrm{id}_V$ is indeed a (singular) linear map $V \to V$, therefore we can choose $v \in V \backslash \{0\}$ such that $\alpha(v) = \lambda v$. Normalise v. Let $U = \langle\{v\}\rangle^\perp \leq V$, then $\alpha(U) \leq U$ since $\langle \alpha(u), v\rangle = \langle u, \alpha(v)\rangle = \langle u, \lambda v\rangle = \bar{\lambda}\langle u, v\rangle = 0$ for any $u \in U$. Also $\dim U = n-1$. Adding v to a basis of U as in the induction hypothesis then gives an orthonormal basis of V consists of eigenvectors of α. ☐

Corollary 23.3. *If V is a finite dimensional inner product space and $\alpha \in L(V)$ is self-adjoint, then V is the orthogonal direct sum of all the eigenspaces of α.*

Proof. Follows directly. ☐

Recall that $\alpha \in L(V)$ is an isometry iff $\alpha^* \circ \alpha = \mathrm{id}$. It is called unitary when V is a vector space over $\mathbb{C}$.

Lemma 23.4. *Let V be a complex inner product space and $\alpha \in L(V)$ be unitary. Then:*
(i) All eigenvalues of α are in the unit circle S^1.
(ii) Eigenvectors with distinct eigenvalues are orthogonal.

Proof. If $v \neq 0, \alpha(v) = \lambda v$ we have $\lambda \neq 0$ as $\alpha \neq 0$ and that $\lambda\|v\|^2 = \langle \alpha(v), v\rangle = \langle v, \alpha^{-1}(v)\rangle = \langle v, \lambda^{-1}(v)\rangle = \bar{\lambda}^{-1}\|v\|^2$, so $\lambda\bar{\lambda} = 1 \implies \lambda \in S^1$.
Now if $v, w \neq 0$ and $\alpha(v) = \lambda v, \alpha(w) = \mu w$ for $\lambda \neq \mu$, then $\lambda\langle v, w\rangle = \langle \alpha(v), w\rangle = \langle v, \alpha^{-1}(w)\rangle = \langle v, \mu^{-1} w\rangle = \bar{\mu}^{-1}\langle v, w\rangle = \mu\langle v, w\rangle \implies \langle v, w\rangle = 0$. ☐

Theorem 23.5 (Spectral Theorem for Unitary Operators in Finite Dimensions)**.** *Let V be a finite dimensional inner product space over $F = \mathbb{C}$ and $\alpha \in L(V)$ is unitary. Then V has an orthogonal basis of eigenvectors of α.*

Equivalently, α can be diagonalised in an orthonormal basis.

Proof. Same idea as in the case for self-adjoint operators. ☐

Sadly, we cannot tell the same tale for real orthogonal matrices since it can have complex eigenvalues.

24 Application to Bilinear Form

We want to analyse bilinear forms by our study in spectral theory.

Corollary 24.1. *Let $A \in M_n(F)$ for $F = \mathbb{R}$ (resp. $\mathbb{C}$) be a symmetric (resp. Hermitian) matrix, then there is an orthogonal (resp. unitary) matrix P such that $P^\top AP$ is a real diagonal matrix.*

Proof. Just take P to be the basis of orthonormal eigenvectors that spans V which exists by spectral theorem. □

Corollary 24.2. *Let V be a finite dimensional inner product space over $F = \mathbb{R}$ (resp. $\mathbb{C}$) and $\phi : V \times V \to F$ be a symmetric bilinear (resp. Hermitian) form. Then there is an orthogonal basis of V in which ϕ is diagonal.*

Proof. Follows directly from the preceding corollary and the change-of-basis formula for bilinear/sesquilinear forms. □

Remark. The diagonal entries in above corollaries are, of course, the eigenvalues.

Corollary 24.3 (Simultaneous Diagonalisation)**.** *Let V be a finite dimensional inner product space over $F = \mathbb{R}$ (resp. $\mathbb{C}$) and $\phi, \psi : V \times V \to F$ be symmetric bilinear (resp. Hermitian) forms. Assume ϕ is positive definite, then there exists a basis of V in which both are diagonalised.*

Proof. Define a new scalar product by $\langle v, w\rangle = \phi(v, w)$ which works as ϕ is positive definite. Then just take the basis to be the basis in which ψ is diagonal under this new inner product. □

Note that in this new basis that we described, ϕ is actually represented by the identity matrix.

Corollary 24.4. *Let $A, B \in M_n(F)$ where $F = \mathbb{R}$ (resp. $\mathbb{C}$). Suppose they are both symmetric (resp. Hermitian) and assume that for any $x \neq 0, x^\top Ax > 0$, then there exists $Q \in M_n(F)$ such that both $Q^\top AQ$ and $Q^\top BQ$ are diagonal.*

Proof. Just a restatement of the preceding corollary. □