Abstract and Linear Algebra

Author: Rex Tse

Last Updated: Dec 2023

Contents

1 Introduction

This set of study notes is written in the summer of 2023 in preparation of MATH2131 which the author would take in the following semester. This is not meant to be a substitute of the course content, instead, it is just a concise summary of what the author has read in order to enhance understanding. If you find any error in this document, you are most welcome to contact the author at yltseaf@connect.ust.hk.

Although no previous knowledge in abstract and linear algebra is required, readers are expected to a strong grasp of fundamental mathematical concepts, including functions, equivalence classes, polynomials, real and complex numbers, etc.

1.1 Acknowledgement

The set of notes takes heavy inspiration from the following sources:

- *Linear Algebra: An Introduction to Abstract Mathematics* by *Robert J. Valenza*
- *Linear Algebra* by *Kenneth Hoffman & Ray Kunze*
- *Prof. Ivan Ip*'s *MATH2131 Lecture Notes*

The author would like to express his sincere gratitude for the help of their work in the creation of this set of notes.

2 Groups, Rings and Fields

2.1 Groups

Definition 2.1.1: Magma

A magma $(M, *)$ is a set M with an operation $* : M \times M \to M$.
In other words, M is closed under $*$.

For example, $(\mathbb{R}^+, +)$ is a magma, but $(\mathbb{R}^+, -)$ is not because $1 - 2 = -1 \notin R^+$.

Definition 2.1.2: Semigroup

A semigroup $(M, *)$ is a magma with the additional property of $*$ being associative,
i.e. $(x * y) * z = x * (y * z)$.

For example, $(\mathbb{R}^+, +)$ from above is also a semigroup because of the associativity of addition.
However, $(\mathbb{N}, \wedge)$, where $x \wedge y = x^y$, is a magma but not a semigroup because $(a^b)^c = a^{bc} \neq a^{(b^c)}$ in general.

Definition 2.1.3: Monoid

A monoid $(M, *)$ is a semigroup with an identity element $e \in M$ such that $e * x = x * e = x$.

For example, $(\mathbb{R}^+, +)$ is not a monoid because $0 \notin \mathbb{R}^+$, but $(\mathbb{R}^+, \times)$ is because $1 \in \mathbb{R}^+$.

Definition 2.1.4: Group

A group $(M, *)$ is a monoid in which every element x has an inverse $x^{-1} \in M$ such that $x * x^{-1} = x^{-1} * x = e$.

Definition 2.1.5: Abelian group

A group $(M, *)$ is Abelian if $*$ is also commutative, i.e. $x * y = y * x$.

As groups are the center of this section, we will look at more examples of groups, some of which we will reuse to illustrate the properties of groups.

Example 1: $(\mathbb{R}^+, \times)$
As $\forall x \in \mathbb{R}^+ : \frac{1}{x} > 0 \implies \frac{1}{x} \in \mathbb{R}^+$, so $(\mathbb{R}^+, \times)$ is also a group.

Example 2: Set of integers modulo n $\mathbb{Z}_n = \{0, 1, ..., n-1\}$
This set contains all the possible moduloes when an integer is divided by n.
Define the addition operator $\oplus$ on this set such that $\forall x, y \in \mathbb{Z}_n : x \oplus y = (x + y) \bmod n$. We will check if $(\mathbb{Z}_n, \oplus)$ is a group:
Closure True by definition.
Associativity True $\because (((x+y) \bmod n) + z) \bmod n = (x + y + z) \bmod n = ((x \bmod n) + y + z) \bmod n$. *Identity* 0 is the identity $\because (x + 0) \bmod n = (0 + x) \bmod n = x \bmod n$.
Inverse $x^{-1} = n - x \because (x + n - x) \bmod n = (n - x + x) \bmod n = 0$.
Therefore, $(\mathbb{Z}_n, \oplus)$ is a group. Moreover, it is an Abelian group as $(x + y) \bmod n = (y + x) \bmod n$.

Example 3: The set of all real polynomials $\mathbb{R}[x]$ **under addition**
The proof for this one is left as an exercise for the readers.

From this page onwards, I will omit the operator (so $x * y$ becomes xy) for simplicity.

2.2 Properties of Groups

Proposition 2.2.1: Cancellation

$xy = xz \implies y = z.$

Proof $$\begin{aligned} xy = xz &\implies x^{-1}(xy) = x^{-1}(xz) \text{ (Inverse)} \\ &\implies (x^{-1}x)y = (x^{-1}x)z \text{ (Associativity)} \\ &\implies y = z \end{aligned}$$

Similarly, $yx = zx \implies y = z$.

Proposition 2.2.2: Uniqueness of identity

The identity element of a group is unique.

Proof Suppose $\exists e, e' \in M$ such that $\forall x \in M : ex = e'x = x$.
By Proposition 2.2.1, this implies $e = e'$, hence the identity is unique.

Proposition 2.2.3: Uniqueness of inverse

The inverse of any element of a group is unique.

Proof $\forall x \in M$: Suppose $\exists p, q \in M$ such that $px = qx = e$.
By Proposition 2.2.1, this implies $p = q$, hence the inverse is unique.

Proposition 2.2.4

$xx = x \iff x = e$

Proof ($\rightarrow$) $xx = x \implies xx = xe \implies x = e$ (Proposition 2.2.1)
The proof of the converse is trivial.

Proposition 2.2.5

$(x^{-1})^{-1} = x$

Proof $xx^{-1} = e$ and $x^{-1}x = e$

Proposition 2.2.6

$(xy)^{-1} = y^{-1}x^{-1}$

Proof $$\begin{aligned} (xy)(y^{-1}x^{-1}) &= x(yy^{-1})x^{-1} \text{ (Associativity)} \\ &= xex^{-1} \\ &= xx^{-1} \\ &= e \end{aligned}$$

2.3 Subgroups

Definition 2.3.1: Subgroup

$H \subset M$ is a subgroup of a group M ($H \leq M$) if:

1. $\forall x, y \in H : xy \in H$ (Closure)
2. $e \in H$ (Identity)
3. $\forall x \in H : x^{-1} \in H$ (Inverse)

As H preserves the associativity property of M, it is obvious that H is also a group.

Example 1: $n\mathbb{Z} = \{nk \mid k \in \mathbb{Z}\} \leq Z$ **under** $+$
Closure $nk_1 + nk_2 = n(k_1 + k_2) \in n\mathbb{Z}$
Identity $0 \in n\mathbb{Z}$
Inverse $\forall k \in \mathbb{Z} : -nk = n(-k) \in n\mathbb{Z}$ and $nk + (-nk) = (-nk) + nk = 0$

Example 2: $\{e\} \leq M$ **for any group** M
The proof for this one is trivial. (In fact, $\{e\}$ is also known as the trivial subgroup.)

Theorem 2.3.1: Subgroup test

Given a group M and $H \subseteq M$, $M \leq H \iff \forall x, y \in H : xy^{-1} \in H$

Proof ($\leftarrow$) *Identity* Find any $x \in H$. Then $xx^{-1} = e \in H$.
Inverse $\forall x \in H : ex^{-1} = x^{-1} \in H$
Closure $\forall x, y \in H : x, y^{-1} \in H \implies x(y^{-1})^{-1} = xy \in H$ (Proposition 2.2.5)
The proof of the converse is trivial.

Let's try to use the subgroup test on **Example 1** above.
$\forall x, y \in n\mathbb{Z} : xy^{-1} = nk_1 + (-nk_2) = n(k_1 - k_2) \in n\mathbb{Z} \implies n\mathbb{Z} \leq \mathbb{Z}$

2.4 Group Homomorphisms

Definition 2.4.1: Group homomorphism

Given groups $(G, *)$ and $(H, \cdot)$, if there is a function $\phi : G \to H$ such that $\forall x_1, x_2 \in G : \phi(x_1 * x_2) = \phi(x_1) \cdot \phi(x_2)$, ϕ is a group homomorphism.

Example 1: $f : (\mathbb{Z}, +) \to (\mathbb{Q}, \times), f(x) = 2^x$
$f(x_1 + x_2) = 2^{x_1 + x_2} = 2^{x_1} \times 2^{x_2} = f(x_1)f(x_2) \implies f$ is a group homomorphism

Example 2: $g : (\mathbb{Z}, +) \to (\mathbb{Z}_n, \oplus), g(x) = x \bmod n$
$g(x_1 + x_2) = (x_1 + x_2) \bmod n = ((x_1 \bmod n) + (x_2 \bmod n)) \bmod n = g(x_1) \oplus g(x_2)$
Therefore, g is a group homomorphism.

Example 3: $h : (\mathbb{R}[t], +) \to (\mathbb{R}[t], +), h(x) = \int x dt$
$h(x_1(t) + x_2(t)) = \int (x_1(t) + x_2(t))dt = \int x_1(t)dt + \int x_2(t)dt = h(x_1(t)) + h(x_2(t))$
Therefore, h is a group homomorphism.

From now on, we will assume $\phi : (G, *) \to (H, \cdot)$ and $\theta : (H, \cdot) \to (K, \otimes)$ are group homomorphisms.

Proposition 2.4.1: Composition of group homomorphisms

$\theta \circ \phi : (G, *) \to (K, \otimes)$ is a group homomorphism.

Proof $\theta(\phi(x_1 * x_2)) = \theta(\phi(x_1) \cdot \phi(x_2)) = \theta(\phi(x_1)) \otimes \theta(\phi(x_2))$

Proposition 2.4.2

Suppose e_G and e_H are the identity elements of G and H respectively. Then $\phi(e_G) = e_H$.

Proof Choose any $x \in G$. Then $\phi(x) \cdot e_H = \phi(x) = \phi(x * e_G) = \phi(x)\phi(e_G)$.
By Proposition 2.2.1, $\phi(e_G) = e_H$.

Proposition 2.4.3

$\phi(x^{-1}) = \phi(x)^{-1}$

Proof $\phi(x^{-1}) \cdot \phi(x) = \phi(x^{-1} * x) = \phi(e_G) = e_H$ (Proposition 2.4.2)
Similarly, $\phi(x) \cdot \phi(x^{-1}) = e_H$

The above propositions show that a group homomorphism is a group structure preserving mapping.

Definition 2.4.2: Kernel

$\ker \phi = \{x \in G \mid \phi(x) = e_H\} = \phi^{-1}[\{e_H\}]$

Definition 2.4.3: Image

$\operatorname{im} \phi = \{y \in H \mid \exists x \in H : \phi(x) = y\} = \phi(G)$

Proposition 2.4.4

$\ker \phi \leq G$

Proof $\forall x_1, x_2 \in \ker \phi :$

$$\begin{aligned} \phi(x_1 * {x_2}^{-1}) = \phi(x_1) \cdot \phi({x_2}^{-1}) &= \phi(x_1) \cdot \phi(x_2)^{-1} \text{ (Proposition 2.4.3)} \\ &= e_H \cdot {e_H}^{-1} \text{ (By definition)} \\ &= e_H \end{aligned}$$

$\therefore x_1 * {x_2}^{-1} \in \ker \phi \implies$ By the subgroup test, $\ker \phi \leq G$

Proposition 2.4.5

$\operatorname{im} \phi \leq H$

Proof $\forall y_1, y_2 \in \operatorname{im} \phi :$

$$\begin{aligned} y_1 \cdot {y_2}^{-1} = \phi(x_1) \cdot \phi(x_2)^{-1} &= \phi(x_1)\phi({x_2}^{-1}) \text{ (Proposition 2.4.3)} \\ &= \phi(x_1 * {x_2}^{-1}) \in \operatorname{im} \phi \text{ (By definition)} \end{aligned}$$

By the subgroup test, $\operatorname{im} \phi \leq H$

Theorem 2.4.1: Group isomorphism

ϕ is a group isomorphism (bijective homomorphism) $\iff$ $\ker\phi = \{e_G\}$ and $\operatorname{im}\phi = H$

Proof (⟵) Suppose $\forall y \in H : \exists x_1, x_2 \in G : \phi(x_1) = \phi(x_2) = y$.

$$\begin{aligned} \text{Then } \phi(x_1)\cdot\phi(x_2)^{-1} = y \cdot y^{-1} &\implies \phi(x_1)\cdot\phi({x_2}^{-1}) = e_H \text{ (Proposition 2.4.3)} \\ &\implies \phi(x_1 * {x_2}^{-1}) = e_H \\ &\implies x_1 * {x_2}^{-1} = e_G \ (\because \ker\phi = \{e_G\}) \\ &\implies x_1 = x_2 \implies \phi \text{ is injective} \end{aligned}$$

$\operatorname{im}\phi = H \implies H$ is surjective

The proof of the converse is trivial.

If there exists an isomorphism between two groups, we say that the two groups are isomorphic. We can also represent this using the notation: $(G, *) \cong (H, \cdot)$ or $G \cong H$ if the group operations are obvious.

2.5 Rings

Definition 2.5.1: Ring

A ring $(R, +, *)$ (with unity) is a set which:

1. $(R, +)$ is an Abelian group with the identity element $0 \in R$
2. $(R, *)$ is a monoid with the identity element $1 \in R$
3. $\forall x, y, z \in R : x * (y + z) = x * y + x * z$ and $(x + y) * z = x * z + y * z$ (Distributivity)

Additionally, if $*$ is commutative, R is also known as a commutative ring.

By convention, we express the additive inverse of $x \in R$ as $-x$ and the multiplicative inverse of $x \in R$ as x^{-1}.

Example 1: $\mathbb{Z}, \mathbb{Q}, \mathbb{R}, \mathbb{C}$

Under regular addition and multiplication, they form commutative rings. The proof of this is left as an exercise to the readers.

Example 2: $R[x]$ **for any ring** R

Under polynomial addition and multiplication, $R[x]$ forms a polynomial ring. The proof of this is also left as an exercise to the readers.

Example 3: $(\mathbb{Z}_n, \oplus, \otimes)$

We will define $\otimes$ in a similar way to $\oplus$, i.e. $\forall x, y \in \mathbb{Z}_n : x \otimes y = xy \bmod n$.

We will check if $(\mathbb{Z}_n, \oplus, \otimes)$ forms a ring:

Additive Abelian Group Proved in Section 2.1.

Multiplicative Monoid

- $\mathbb{Z}_n$ is closed under $\otimes$.
- The identity element $1 \in \mathbb{Z}_n$.
- $(x \otimes y) \otimes z = ((xy \bmod n)z) \bmod n = xyz \bmod n = (x(yz \bmod n)) \bmod n = x \otimes (y \otimes z)$

Distributivity

$$\begin{aligned} x \otimes (y \oplus z) &= (x((y + z) \bmod n)) \bmod n \\ &= (x(y + z)) \bmod n \\ &= (xy + xz) \bmod n \\ &= ((xy \bmod n) + (xz \bmod n)) \bmod n = (x \otimes y) \oplus (x \otimes z) \end{aligned}$$

Similarly, we can show that $(x \oplus y) \otimes z = (x \otimes z) \oplus (y \otimes z)$.
Therefore, we have shown that $(\mathbb{Z}_n, \oplus, \otimes)$ is a ring. Moreover, as $\otimes$ is commutative, it is a commutative ring too.

2.6 Properties of Rings

Rings share most of the properties of signed numbers in the real number system we are familiar with.

Proposition 2.6.1

$0 * x = x * 0 = 0$

Proof By distributivity, $(0+1) * x = 0 * x + 1 * x$.
$(0+1) * x = 1 * x = 0 + 1 * x$
By Proposition 2.2.1, $0 * x + 1 * x = 0 + 1 * x \implies 0 * x = 0$.
$x * 0 = 0$ can be proved in a similar way.

Proposition 2.6.2

If $0 = 1$, then the ring is the zero ring $\{0\}$.

Proof By Proposition 3.2.1, $\forall x \in R : x = x * 0 = 0$.

Proposition 2.6.3

$x(-y) = (-x)y = -(xy)$

Proof By Proposition 3.2.1, $x(y - y) = x * 0 = 0$.
By distributivity and additive commutativity, $x(y-y) = xy + x(-y) = x(-y) + xy$.
Therefore, $xy + x(-y) = x(-y) + xy = 0 \implies x(-y) = -(xy)$.
$(-x)y = -(xy)$ can be proved in a similar way.

Proposition 2.6.4

$(-x)(-y) = xy$

Proof By Proposition 3.2.1, $(-x)(y-y) = (-x) * 0 = 0$.
By distributivity and Proposition 3.2.3, $(-x)(y-y) = (-x)y + (-x)(-y) = (-xy) + (-x)(-y)$.
By the above and additive commutativity, $(-xy) + (-x)(-y) = (-x)(-y) + (-xy) = 0$.
By Proposition 2.2.5, $(-x)(-y) = -(-xy) = xy$.

Proposition 2.6.5

$(-1)x = -x$

Proof By Proposition 3.2.1, $(1-1)x = 0 * x = 0$.
By distributivity and additive commutativity, $(1-1)x = 1 * x + (-1)x = x + (-1)x = (-1)x + x$.
Therefore, $x + (-1)x = (-1)x + x = 0 \implies (-1)x = -x$.

2.7 Fields

Definition 2.7.1: Field

A ring $(R, +, *)$ is also a field if $(R \setminus \{0\}, *)$ is an Abelian group.

Example 1: $\mathbb{Q}, \mathbb{R}, \mathbb{C}$
Under regular addition and mulitplication, they also form fields. The proof is left as an exercise to the readers.

Example 2: $\mathbb{Z}_n$
We will show that $(\mathbb{Z}_n, \oplus, \otimes)$ is a field if and only if $n \in \mathbb{P}$ (the set of all prime numbers). However, before that, we will first have to prove Bézout's identity:

Theorem 2.7.1: Bézout's Identity

Suppose d is the highest common factor (HCF) of two integers n and m. Then $\exists x, y \in \mathbb{Z}$ such that $nx + my = d$.

Proof Let $S = \{nx + my : x, y \in \mathbb{Z} \text{ and } nx + my > 0\}$.
As $S \subset N$ and is non-empty (consider $x = \pm 1, y = 0$), we can let $d = \min S = ns + mt$.
Let $n = dq + r$, where $0 \leq r < d$. Then $r = n - dq = n - q(ns + mt) = n(1 - qs) - mqt \in S \cup \{0\}$.
Yet, if $r \in S$, $r < q = \min S$. Therefore, $r = 0$ and hence d is a factor of n. We can show that d is a factor of m in a similar way.
We have thus established that d is a common factor of n and m.
For any c which is a common factor of n and m, $d = ns + mt = cus + cvt = c(us + vt)$, which means c is a factor of d.
As $d > 0$, it is a must that $c < d$, hence d is the HCF n and m.

$(\to)$ As n is a prime number, the HCF of n and any $m \in \mathbb{Z}_n \setminus \{0\}$ is 1.
Then by Bézout's identity, $\exists x, y \in \mathbb{Z}$ such that $nx + my = 1$.

$$\begin{aligned} nx + my = 1 \implies (nx + my) \bmod n = 1 \implies nx \bmod n + my \bmod n = 1 &\implies m(y \bmod n) \bmod n = 1 \\ \text{(a)} &\implies m \otimes (y \bmod n) = 1 \\ \text{(Commutative ring)(b)} &\implies (y \bmod n) \otimes m = 1 \\ \text{(a,b)} &\implies m^{-1} = y \bmod n \end{aligned}$$

As we have proved $(\mathbb{Z}_n, \oplus, \otimes)$ is a commutative ring in Section 3.1 already, with the fact that every element of $\mathbb{Z}_n \setminus \{0\}$ has an inverse, we have shown that it is also a field.
$(\leftarrow)$ Suppose $\exists n \notin \mathbb{P}$ such that $(\mathbb{Z}_n, \oplus, \otimes)$ is a ring. By definition, every $m \in \mathbb{Z}_n$ should have an inverse.
By the definition of prime numbers, $\exists 1 < m < n$ such that m is a factor of n. Suppose $n = cm$.
Then $\forall x \in \mathbb{Z}_n$:
If $xm \leq n$: $xm \bmod n = xm \neq 1$
If $xm > n$: Suppose $xm = qn + d$. By definition, $d = xm \bmod n$

$$\text{Then } xm = qcm + d \implies x = qc + \frac{d}{m}.$$

$$x, qc \in \mathbb{N} \implies \frac{d}{m} \in \mathbb{N} \implies d \text{ is a multiple of } m \implies d > 1$$

Therefore, m does not have a multiplicative inverse, which leads to a contradiction.

To illustrate the above, we can consider the multiplication table of $\mathbb{Z}_5$ and $\mathbb{Z}_6$:

⊗	0	1	2	3	4
0	0	0	0	0	0
1	0	1	2	3	4
2	0	2	4	1	3
3	0	3	1	4	2
4	0	4	3	2	1

⊗	0	1	2	3	4	5
0	0	0	0	0	0	0
1	0	1	2	3	4	5
2	0	2	4	0	2	4
3	0	3	0	3	0	3
4	0	4	2	0	4	2
5	0	5	4	3	2	1

This concludes the abstract algebra part of this set of notes. In the following sections, we will apply the above to prove results in linear algebra.

3 Vector Spaces

3.1 Vector Spaces

Definition 3.1.1: Vector space

Given a field $(k, +, *)$ and a set V with the following operations:

1. $+ : V \times V \to V$ (Vector addition)
2. $\ : k \times V \to V$ (Scalar multiplication),

then V is a vector space over k if $\forall r, s \in k, \vec{u}, \vec{v} \in V$:

1. $(V, +)$ is an Abelian group.
2. $r(s\vec{v}) = (rs)\vec{v}$ (Scalar associativity)
3. $(r + s)\vec{v} = r\vec{v} + s\vec{v}$ (Vector distributivity) and $r(\vec{u} + \vec{v}) = r\vec{u} + r\vec{v}$ (Scalar distributivity)
4. $1\vec{v} = \vec{v}$ (Identity scalar)

By convention, we express the identity element of the vector additive Abelian group as $\vec{0}$ and the additive inverse of any $\vec{v} \in V$ as $-\vec{v}$.

Example 1: The trivial vector space $\{\vec{0}\}$
For any field k, We define $\vec{0}$ to have the following properties:

1. $\vec{0} + \vec{0} = \vec{0}$
2. $r\vec{0} = \vec{0}$

Then $\{\vec{0}\}$ is a vector space over k. The proof for this one is trivial.

Example 2: A field k over itself
If we define vector addition and scalar multiplication identically to addition and multiplication in k respectively, then k is a vector space over itself. The proof for this one is also trivial.

Example 3: k^n over k for any field $(k, +, *)$
k_n is the Cartesian product of n number of ks, i.e. $k^n = \{\langle v_1, v_2, ..., v_n\rangle \mid v_i \in k\}$.
If we define vector addition and scalar multiplication as the following:

1. $\vec{u} + \vec{v} = \langle u_1, u_2, ..., u_n\rangle + \langle v_1, v_2, ..., v_n\rangle = \langle u_1 + v_1, u_2 + v_2, ..., u_n + v_n\rangle \in k^n$
2. $r\vec{v} = r\langle v_1, v_2, ..., v_n\rangle = \langle rv_1, rv_2, ..., rv_n\rangle \in k^n$,

then k^n is a vector space over k:
Vector Additive Abelian Group

- *Closure* True by definition
- *Associativity* True $\because (k, +)$ is an Abelian group
- *Identity* $\vec{0} = \langle 0, 0, ..., 0\rangle$
- *Inverse* $-\vec{v} = -\langle v_1, v_2, ..., v_n\rangle = \langle -v_1, -v_2, ..., -v_n\rangle$
- *Commutativity* True $\because (k, +)$ is an Abelian group

Scalar Associativity $r(s\vec{v}) = r(s\langle v_1, v_2, ..., v_n\rangle) = r\langle sv_1, sv_2, ..., sv_n\rangle = \langle rsv_1, rsv_2, ..., rsv_n\rangle = rs\vec{v}$
Vector Distributivity $(r+s)\vec{v} = (r+s)\langle v_1, v_2, ..., v_n\rangle = \langle (r + s)v_1, (r + s)v_2, ..., (r + s)v_n\rangle = \langle rv_1+sv_1, rv_2+sv_2, ..., rv_n + sv_n\rangle = \langle rv_1, rv_2, ..., rv_n\rangle + \langle sv_1, sv_2, ..., sv_n\rangle = r\vec{v} + s\vec{v}$
Scalar Distributivity $r(\vec{u} + \vec{v}) = r(\langle u_1, u_2, ..., u_n\rangle + \langle v_1, v_2, ..., v_n\rangle) = r\langle u_1 + v_1, u_2 + v_2, ..., u_n + v_n\rangle = \langle ru_1 + rv_1, ru_2 + rv_2, ..., ru_n + rv_n\rangle = r\langle u_1, u_2, ..., u_n\rangle + r\langle v_1, v_2, ..., v_n\rangle = r\vec{u} + r\vec{v}$
Identity Scalar $1\vec{v} = 1\langle v_1, v_2, ..., v_n\rangle = \langle 1 * v_1, 1 * v_2, ..., 1 * v_n\rangle = \vec{v}$

An important application of this result is the $\mathbb{R}^n$ vector space over $\mathbb{R}$, also known as Euclidean vectors. They are the “vectors” we commonly use in other disciplines, such as physics and computer science. Geometrically, they represent the movement of point objects in an Euclidean space. To visualise this, we will focus on the $\mathbb{R}^2$ space below.

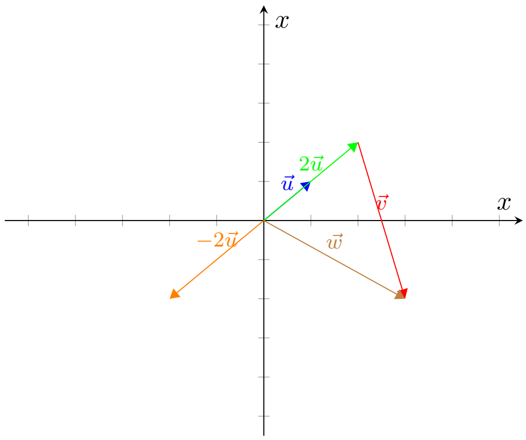

Figure 3.1.1: The $\mathbb{R}^2$ vector space

Suppose a person is standing at the origin $(0, 0)$ in the beginning. A vector $\vec{u} = \langle 1, 1\rangle$ would translate to the person moving one unit to the right and one unit upwards. We can scale a vector by a factor $r \in \mathbb{R}$, which would change the magnitude of the vector but not its direction. This can be done via scalar multiplication, for example, if we scale $\vec{u}$ by two times, we get $2\vec{u} = 2\langle 1, 1\rangle = \langle 2, 2\rangle$.
We can also combine two vectors. If a person moves two unit to the right and upwards (as in $2\vec{u}$), then one units to the right and three units downwards (as in $\vec{v} = \langle 1, -3\rangle$), then we can say the person has moved three units to the right and two units downwards in total (as in $\vec{w}$). This can be represented by vector addition: $\vec{w} = 2\vec{u} + \vec{v}$.

Another application would be the $\mathbb{Z}_p^n$ vector space over $\mathbb{Z}_p$, where $p \in \mathbb{P}$. In Section 3.3, we have shown that $(\mathbb{Z}_p, \oplus, \otimes)$ is a field. Thus by the above, we can establish that $\mathbb{Z}_p^n$ is a vector space over $\mathbb{Z}_p$. A way to get an intuition of this vector space is to imagine a finite grid which repeats itself indefinitely. Vector addition and scalar multiplication work similarly to that in the $\mathbb{R}^n$ vector space, except if the vector gets "out-of-bounds", it lands on another copy of the grid.

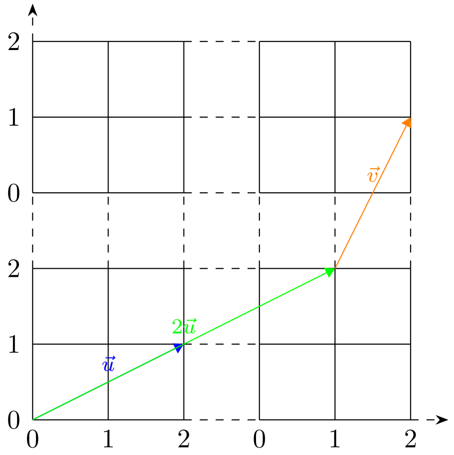

Figure 3.1.2: The $\mathbb{Z}_3$ vector space. As shown in the figure, if $\vec{u} = \langle 0, 2\rangle$ and $\vec{v} = \langle 1, 2\rangle$, we have $2\vec{u} = \langle 1, 2\rangle$ and $2\vec{u} \oplus \vec{v} = \langle 2, 1\rangle$.

Example 4: $k[x]$ over k for any field $(k, +, *)$
$k[x]$ forms a vector space over k if we define vector addition and scalar multiplication as polynomial addition and constant multiplication respectively. The proof for this is trivial.

Example 5: $\mathcal{C}^n$ over $\mathbb{R}$
The set of n-times differentiable functions $\mathcal{C}^n$ forms a vector space over $\mathbb{R}$ under regular addition and multiplication due to properties of differentiability. The detailed proof is left as an exercise to the readers.

3.2 Properties of Vector Spaces

Proposition 3.2.1

$r\vec{0} = \vec{0}$

Proof By scalar distributivity, $r(\vec{0}+\vec{0}) = r\vec{0} + r\vec{0}$.
As $\vec{0}$ is the additive identity, $r(\vec{0}+\vec{0}) = r\vec{0} = \vec{0} + r\vec{0}$.
By Proposition 2.2.1, $r\vec{0} = \vec{0}$.

Proposition 3.2.2

$0\vec{v} = 0$

Proof By scalar distributivity, $(0+1)\vec{v} = 0\vec{v} + 1\vec{v}$.
As $\vec{0}$ is the additive identity, $(0+1)\vec{v} = 1\vec{v} = \vec{0} + 1\vec{v}$.
By Proposition 2.2.1, $0\vec{v} = 0$.

Proposition 3.2.3

$(-r)\vec{v} = -(r\vec{v})$

Proof By scalar distributivity and commutativity, $(r-r)\vec{v} = r\vec{v} + (-r)\vec{v} = (-r)\vec{v} + r\vec{v}$.
By Proposition 3.2.2,$(r-r)\vec{v} = 0\vec{v} = 0$.
Therefore, $r\vec{v} + (-r)\vec{v} = (-r)\vec{v} + r\vec{v} = 0$, hence by definition $(-r)\vec{v} = -(r\vec{v})$.

A corollary of this proposition is that $(-1)\vec{v} = -\vec{v}$.

Proposition 3.2.4

$r\vec{v} = \vec{0} \iff r = 0 \text{ or } \vec{v} = \vec{0}$

Proof (→) We already know by Proposition 3.2.2 that $0\vec{v} = \vec{0}$, so we consider the case when $r \neq 0$.
As $r \in k \setminus \{0\}$, r has a multiplicative inverse r^{-1}, thus we have:

$$\begin{aligned} r^{-1}(r\vec{v}) = r^{-1}\vec{0} &\implies (r^{-1}r)\vec{v} = \vec{0} \text{ (Associativity, Proposition 3.2.1)} \\ &\implies 1\vec{v} = 0 \\ &\implies \vec{v} = 0 \text{ (Identity scalar)} \end{aligned}$$

(←) Proved in Proposition 3.2.1 and Proposition 3.2.2.

3.3 Vector Subspaces

Definition 3.3.1: Vector subspace

Suppose V is a vector space over k and $W \subset V$. W is a subspace of V if:

1. $(W,+)$ is a subgroup of $(V,+)$.
2. $\forall \vec{w} \in W, r \in k : r\vec{w} \in W$

As a subspace retains all properties of the parent vector space, it is also a vector space (over the same field).

Theorem 3.3.1: Subspace test

$W \subseteq V$ is a subspace $\iff \forall \vec{v}, \vec{w} \in W, r \in k : r\vec{v}, \vec{v} + \vec{w} \in W$

Proof **(←)** The closure of vector addition and scalar multiplication is given.
Identity By Proposition 3.2.2, $\vec{0} = 0\vec{v} \in W$.
Inverse By Proposition 3.2.3, $-\vec{v} = (-1)\vec{v} \in W$.
The proof of the converse is trivial.

Example 1: $W_1 = \{\langle r_0v_0, r_1v_1, ..., r_nv_n\rangle \mid v_i \in k\} \subseteq k^n$ **over any field** k**, where** $r_i \in k$ **are constants**
Notice that $W_1 \subseteq W$ because scalar multiplication is closed.
$\forall \vec{v}, \vec{w} \in W$:

1. $$\begin{aligned} s\vec{v} &= \langle s(r_0v_0), s(r_1v_1), ..., s(r_nv_n)\rangle \\ &= \langle (sr_0)v_0, (sr_1)v_1, ..., (sr_n)v_n\rangle \text{ (Associativity)} \\ &= \langle (r_0s)v_0, (r_1s)v_1, ..., (r_ns)v_n\rangle \text{ (Commutativity, including 0)} \\ &= \langle r_0(sv_0), r_1(sv_1), ..., r_n(sv_n)\rangle \text{ (Associativity)} \\ &\in W_1 \text{ (Closure)} \end{aligned}$$
2. $$\begin{aligned} \vec{v} + \vec{w} &= \langle r_0v_0, r_1v_1, ..., r_nv_n\rangle + \langle r_0w_0, r_1w_1, ..., r_nw_n\rangle \\ &= \langle r_0(v_0 + w_0), r_1(v_1 + w_1), ..., r_n(v_n + w_n)\rangle \text{ (Distributivity)} \\ &\in W_1 \text{ (Closure)} \end{aligned}$$

Therefore, by the subspace test, we have proved that W_1 is a subspace of k^n.

The geometric meaning of this result is that we can "stretch" a vector space, and it would remain a vector space. Let's consider the $\mathbb{R}^2$ vector space. If we let $W_1 = \{\langle 2x, 3y\rangle \mid x, y \in \mathbb{R}\} \subseteq \mathbb{R}^2$, we are scaling every vector in the $\mathbb{R}^2$ vector space by two times in the x direction and three times in the y direction.

A more interesting example would be when one (or more) of the rs is 0, in which case one of the axis would collapse. For example, if $W_1 = \{\langle x, 0\rangle \mid x \in \mathbb{R}\} \subseteq \mathbb{R}^2$, then every vector would "fall" on the x-axis.

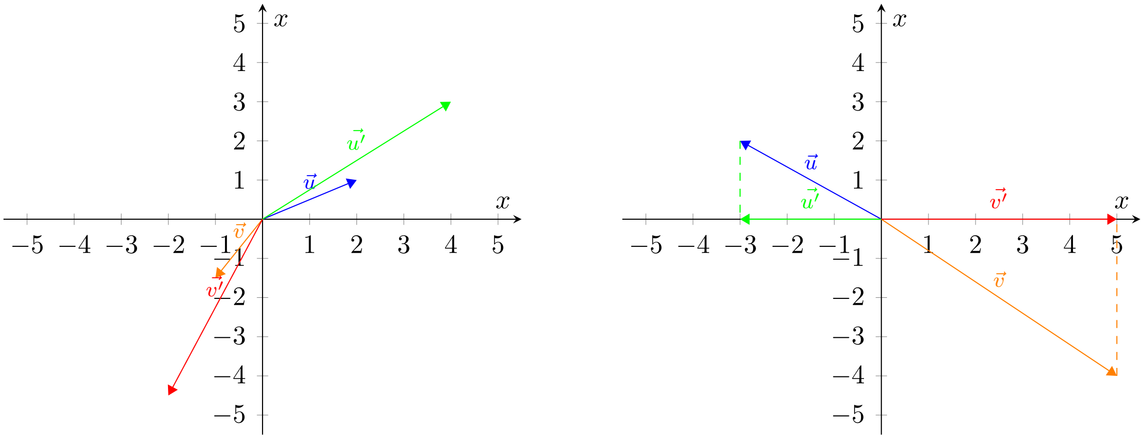

Figure 3.3.1: Examples of W_1 vector subspaces.
(Left: $W_1 = \{\langle 2x, 3y\rangle \mid x, y \in \mathbb{R}\}$, Right: $W_1 = \{\langle x, 0\rangle \mid x \in \mathbb{R}\}$)

Example 2: $W_2 = \{\langle r_1v_1, r_2v_2, ..., r_nv_n\rangle \mid v_i \in k, r_1v_1 + r_2v_2 + ... + r_nv_n = 0\} \subseteq k^n$ **over any field** k**, where** $r_i \in k$ **are constants**
$\forall \vec{v}, \vec{w} \in W$:

1. $$\begin{aligned} r_1v_1 + r_2v_2 + ... + r_nv_n = 0 &\implies sr_1v_1 + sr_2v_2 + ... + sr_nv_n = s * 0 = 0 \text{ (Distributivity, Proposition 3.2.1)} \\ &\implies r_1(sv_1) + r_2(sv_2) + ... + r_n(sv_n) = 0 \text{ (}\textbf{Example 1}\text{: } r_i(sv_i) = s(r_iv_i)\text{)} \\ &\implies s\vec{v} \in W_2 \end{aligned}$$

2. $r_1(v_1+w_1)+r_2(v_2+w_2)+...+r_n(v_n+w_n)$
   $=r_1v_1+r_1w_1+r_2v_2+r_2w_2+...+r_nv_n+r_nw_n$ (Distributivity)
   $=(r_1v_1+r_2v_2+...+r_nv_n)+(r_1w_1+r_2w_2+...+r_nw_n)$ (Associativity, Commutativity)
   $=0+0=0$
   $\implies \vec{v}+\vec{w}\in W_2$

Therefore, by the subspace test, W_2 is a subspace of k^n.

In the $\mathbb{R}^2$ vector space, this result means that vectors that lie on a line passing through the origin form a subspace. (In the $\mathbb{R}^3$ vector space, it would be vectors that lie on a plane passing through the origin, etc.)

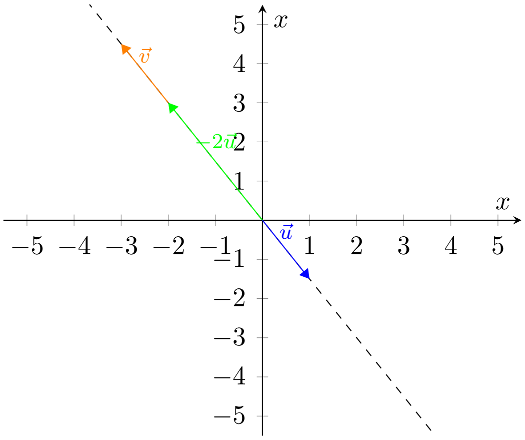

Figure 3.3.2: Example of a W_2 vector subspace, where $W_2=\{\langle x,y\rangle \mid x,y\in\mathbb{R}, 3x+2y=0\}$

Example 3: $k_n[x]\subset k[x]$ over k
We have shown in the last section that $k[x]$ forms a vector space over k.
If we define $k_n[x]$ to be polynomials in x with coefficients in k and degree less than or equal to n, as the sum and constant multiple of such polynomials have degree less than or equal to k, by the subspace test, $k_n[x]$ is a vector subspace of $k[x]$.

Example 4: $\mathcal{C}^k\subseteq\mathcal{C}^n$ over $\mathbb{R}$, where $k\geq n$
We have shown in the last section that $\mathcal{C}^n$ forms a vector space over $\mathbb{R}$. As the sum and constant multiple of n-times differentiable functions are still n-times differentiable, by the subspace test, $\mathcal{C}^k$ is a vector subspace of $\mathcal{C}^n$.

Example 5: $\mathcal{W}_1=\{f\mid f(k)=0\}\subset\mathcal{C}^0$ over $\mathbb{R}$, where $k\in\mathbb{R}$ is a constant
For any $f,g\in\mathcal{W}_1$, we have $rf(k)=0$ and $f(k)+g(k)=0+0=0$. Hence, by the subspace test, $\mathcal{W}_1$ is a vector subspace of $\mathcal{C}^0$.

Example 6: $\mathcal{W}_2=\{f\mid f'(x)=f(x)\}\subset\mathcal{C}^1$ over $\mathbb{R}$
For any $f,g\in\mathcal{W}_1$, by derivative rules, we have $(rf)'=rf'=rf$ and $(f+g)'=f'+g'=f+g$. Hence, by the subspace test, $\mathcal{W}_2$ is a vector subspace of $\mathcal{C}^1$.

Definition 3.3.2: Linear combination

Given a vector space V over k, the linear combination of $\vec{v_i}\in V$ with weights $r_i\in k$ is $\sum r_i\vec{v_i}$. Additionally, the set of all linear combinations of a set of vectors is called the span, i.e. $\text{span}\{v_i\}=\{\sum r_i\vec{v_i}\mid r_i\in k\}$.

Proposition 3.3.1

For any finite subset v of a vector space V, $\operatorname{span} v$ is a vector subspace of V.

Proof $\forall \vec{u}, \vec{w} \in V$:
Let $\vec{u} = r_1\vec{v_1} + r_2\vec{v_2} + ... + r_n\vec{v_n}$ and $\vec{w} = s_1\vec{v_1} + s_2\vec{v_2} + ... + s_n\vec{v_n}$, where $r_i, s_i \in k$.
$\forall t \in k$: By scalar distributivity, $t\vec{u} = t(r_1\vec{v_1} + r_2\vec{v_2} + ... + r_n\vec{v_n}) = tr_1\vec{v_1} + tr_2\vec{v_2} + ... + tr_n\vec{v_n} \in \operatorname{span} v$

$$\begin{aligned}\vec{u} + \vec{w} &= (r_1\vec{v_1} + r_2\vec{v_2} + ... + r_n\vec{v_n}) + (s_1\vec{v_1} + s_2\vec{v_2} + ... + s_n\vec{v_n}) \\ &= (r_1\vec{v_1} + s_1\vec{v_1}) + (r_2\vec{v_2} + s_2\vec{v_2}) + ... + (r_n\vec{v_n} + s_n\vec{v_n}) \text{ (Associativity, Distributivity)} \\ &= (r_1 + s_1)\vec{v_1} + (r_2 + s_2)\vec{v_2} + ... + (r_n + s_n)\vec{v_n} \text{ (Vector distributivity)} \\ &\in \operatorname{span} v\end{aligned}$$

Therefore, by the subspace test, $\operatorname{span} v$ is a vector subspace of V.

3.4 Linear Transformations

Definition 3.4.1: Linear transformation

Suppose $(V, +,)$ and $(W, \oplus, \circ)$ are vector spaces over the same field k. Then a function $\phi : V \to W$ is a linear transformation if:

1. $\forall \vec{v}, \vec{w} \in V : \phi(\vec{v} + \vec{w}) = \phi(\vec{v}) \oplus \phi(\vec{w})$ (Homomorphism)
2. $\forall \vec{v} \in V, r \in k : \phi(r\vec{v}) = r \circ \phi(\vec{v})$ (Homogeneity)

Additionally, if ϕ is bijective, it is a vector space isomorphism. If an isomorphism exists between V and W, $V \cong W$.

In this section, we will represent vectors in V using overhead arrows (e.g. $\vec{v}$) and vectors in W using bold text (e.g. $\boldsymbol{v}$).

Definition 3.4.2: Kernel

$\ker \phi = \{\vec{v} \in V \mid \phi(\vec{v}) = \mathbf{0}\}$

Definition 3.4.3: Image

$\operatorname{im} \phi = \{\boldsymbol{v} \in W \mid \exists \vec{v} \in V : \boldsymbol{v} = \phi(\vec{v})\}$

As linear transformation is homomorphic and vectors in a vector space form an additive group, linear transformations are group homomorphisms too. As a result, some of the results from group homomorphisms directly apply:

- By Proposition 2.4.2, $\phi(\vec{0}) = \mathbf{0}$
- By Proposition 2.4.3, $\phi(-\vec{v}) = -\boldsymbol{v}$
- By Theorem 2.4.1, if $\ker \phi = \{\vec{0}\}$ and $\operatorname{im} \phi = W$, ϕ is bijective and hence a vector space isomorphism

Most of the other properties also apply to linear transformations:

Proposition 3.4.1: Composition of linear transformations

Suppose $\theta : (W, \oplus, \circ) \to (X, \boxplus, \cdot)$ is a linear transformation. Then $\theta \circ \phi$ is also a linear transformation.

Proof *Homomorphism* $\theta(\phi(\vec{v} + \vec{w})) = \theta(\phi(\vec{v}) \oplus \phi(\vec{w})) = \theta(\phi(\vec{v})) \boxplus \theta(\phi(\vec{w}))$
Homogeneity $\theta(\phi(r\vec{v})) = \theta(r \circ \phi(\vec{v})) = r \cdot \theta(\phi(\vec{v}))$

Proposition 3.4.2

$\ker\phi$ is a vector subspace of V.

Proof $\forall\vec{v},\vec{w}\in\ker\phi, r\in k$:
$\phi(r\vec{v})=r\phi(\vec{v})=r\circ\mathbf{0}=\mathbf{0}$ (Homogeneity, Proposition 3.2.1) $\implies r\vec{v}\in\ker V$
$\phi(\vec{v}+\vec{w})=\phi(\vec{v})+\phi(\vec{w})=\mathbf{0}\oplus\mathbf{0}=\mathbf{0}$ (Homomorphism) $\implies \vec{v}+\vec{w}\in\ker V$
By the subspace test, $\ker\phi$ is a vector subspace of V.

Proposition 3.4.3

$\operatorname{im}\phi$ is a vector subspace of W.

Proof $\forall\boldsymbol{v},\boldsymbol{w}\in W, r\in k$: Suppose $\phi(\vec{v})=\boldsymbol{v}$ and $\phi(\vec{w})=\boldsymbol{w}$.
Then $\phi(r\vec{v}\in V)=r\phi(\vec{v})=r\circ\boldsymbol{v}$ (Homogeneity) $\implies r\circ\boldsymbol{v}\in\operatorname{im}\phi$,
$\phi(\vec{v}+\vec{w}\in V)=\phi(\vec{v})+\phi(\vec{w})=\boldsymbol{v}+\boldsymbol{w}$ (Homomorphism) $\implies \boldsymbol{v}+\boldsymbol{w}\in W$.
Therefore, by the subspace test, $\operatorname{im}\phi$ is a vector subspace of W.

We will now look at some examples of linear transformations/ vector space isomorphisms:
Example 1: The zero map $O:V\to W, O(\vec{v})=\mathbf{0}$ for any vector spaces V and W
Homomorphism $O(\vec{v}+\vec{w})=\mathbf{0}=\mathbf{0}\oplus\mathbf{0}=O(\vec{v})+O(\vec{w})$
Homogeneity $O(r\vec{v})=\mathbf{0}=r\circ\mathbf{0}=r\circ O(\vec{v})$ (Proposition 3.2.1)
Therefore, O is a linear transformation. Moreover, Z is injective if and only if V is the trivial vector space, and is surjective if and only if W is the trivial vector space.

Example 2: The identity map $\mathrm{id}_V:V\to V, \mathrm{id}_V(\vec{v})=\vec{v}$ for any vector space V
Obviously, the function is both homomorphic and homogeneous, thus it is a linear transformation.
It is also easy to see that it is a vector space isomorphism.

Example 3: $f:k^n\to k, f(\langle v_1,v_2,...,v_n\rangle)=r_1v_1+r_2v_2+...+r_nv_n$ for any field k and constants $r_i\in k$
From here onwards, if I treat a field k as a vector space, the vector space is assumed to be k over itself.
Homomorphism

$$\begin{aligned}f(\vec{v}+\vec{w})&=f(\langle v_1,v_2,...,v_n\rangle+\langle w_1,w_2,...,w_n\rangle)\\&=f(\langle v_1+w_1,v_2+w_2,...,v_n+w_n\rangle)\\&=r_1(v_1+w_1)+r_2(v_2+w_2)+...+r_n(v_n+w_n)\\&=(r_1v_1+r_1w_1)+(r_2v_2+r_2w_2)+...+(r_nv_n+r_nw_n)\text{ (Scalar distributivity)}\\&=(r_1v_1+r_1v_2+...+r_nv_n)+(r_1w_1+r_2w_2+...+r_nw_n)\text{ (Associativity, Distributivity)}\\&=f(\vec{v})+f(\vec{w})\end{aligned}$$

Homogeneity

$$\begin{aligned}f(s\vec{v})=f(s\langle v_1,v_2,...,v_n\rangle)&=f(\langle sv_1,sv_2,...,sv_n\rangle)\\&=r_1(sv_1)+r_2(sv_2)+...+r_n(sv_n)\\&=s(r_1v_1)+s(r_2v_2)+...+(sr_nv_n)\text{ (Associativity, Commutativity)}\\&=s(r_1v_1+r_2v_2+...+r_nv_n)\text{ (Scalar distributivity)}\\&=sf(\vec{v})\end{aligned}$$

Hence, f is a linear transformation. However, it is injective if and only if $n=1$ and $r_1\neq 0$:
($\rightarrow$) Case 1: $n>1, r_i\neq 0$
Suppose f is injective and $\exists j\neq k$ such that $r_j,r_k\neq 0$.
We construct a vector $\vec{v}\in k^n$ such that $v_j=1, v_k={r_k}^{-1}(-r_j), v_{i\neq j,k}=0$.

Then by Proposition 3.2.1 and associativity, $f(\vec{v}) = r_j + r_k({r_k}^{-1}(-r_j)) = r_j + (r_k {r_k}^{-1})(-r_j) = r_j + (-r_j) = 0$. This means that $\ker f \neq \{\vec{0}\}$, which means that f is not injective.
Case 2: $n \geq 1, \exists j$ such that $r_j = 0$
Construct a vector $\vec{v} \in k$ such that $v_i = 1, v_{i \neq j} = 0$.
Then by Proposition 3.2.1, $f(\vec{v}) = 1 * 0 = 0 \implies \ker f \neq \{\vec{0}\} \implies f$ is not injective.
($\leftarrow$) $\forall x \in k : r_1 x = 0 \implies ({r_1}^{-1} r_1)x = {r_1}^{-1} * 0 \implies x = 0$ (Associativity, Proposition 3.2.1)
Hence, $\ker f = \{0\}$ and by Proposition 3.2.1, f is injective.

Example 4: $g : k^n \to k^m, g(\langle v_1, v_2, ..., v_n \rangle) = \langle v_1, v_2, ..., v_n, 0, ..., 0 \rangle$ for any field k and $m > n$
g is a vector space isomorphism. The proof of this is left as an exercise to the readers.

Example 5: The formal derivative $D : k[x] \to k[x]$ for any polynomial vector space $k[x]$ over k
As derivatives only apply to real functions, to extend it to any polynomial ring $R[x]$, we will have to define a new function called the formal derivative:
$\forall a_i \in R : D(a_n x^n + a_{n-1}x^{n-1} + ... + a_0) = (na_n x^{n-1} + (n-1)a_{n-1}x^{n-2} + ... + 0)$ and $na = \underbrace{a + ... + a}_{n \text{ times}}$
It is easy to see that almost all of the properties of regular derivatives apply to formal derivatives, including the sum rule and the constant multiple rule. Hence, the formal derivative is a vector space homomorphism on polynomial vector spaces.
Yet, it is not injective, for example: $D(x+1) = D(x) = 1$.

Example 6: $I_{[a,b]} : F \to \mathbb{R}, I_{[a,b]}(f) = \int_a^b f dx$, where F is the set of all integrable functions on $[a, b]$
$I_{[a,b]}$ is a linear transformation due to the properties of integration, but is not injective. The detailed proof is left as an exercise to the readers.

3.5 Direct Product and Sum of Vector Spaces

Definition 3.5.1: Direct product of vector spaces

Given two vector spaces V and W over the same field k, the direct product of V and W is $V \times W = \{(\vec{v}, \vec{w}) \mid \vec{v} \in V, \vec{w} \in W\}$, which has the following operations:

1. $(\vec{v_1}, \vec{w_1}) + (\vec{v_2}, \vec{w_2}) = (\vec{v_1} + \vec{v_2}, \vec{w_1} + \vec{w_2})$ (Vector addition)
2. $r(\vec{v}, \vec{w}) = (r\vec{v}, r\vec{w})$ (Scalar multiplication)

Proposition 3.5.1

$V \times W$ is a vector space over k.

Proof *Additive Abelian Group* True because $(V, +)$ and $(W, +)$ are both Abelian groups.
Scalar Associativity $r(s(\vec{v}, \vec{w})) = r(s\vec{v}, s\vec{w}) = (r(s\vec{v}), r(s\vec{w})) = ((rs)\vec{v}, (rs)\vec{w}) = (rs)(\vec{v}, \vec{w})$
Vector Distributivity $(r+s)(\vec{v}, \vec{w}) = ((r+s)\vec{v}, (r+s)\vec{w}) = (r\vec{v} + s\vec{v}, r\vec{w} + s\vec{w}) = r(\vec{v}, \vec{w}) + s(\vec{v}, \vec{w})$
Scalar Distributivity $r((\vec{v_1}, \vec{w_1}) + (\vec{v_2}, \vec{w_2})) = r(\vec{v_1} + \vec{v_2}, \vec{w_1} + \vec{w_2}) = (r\vec{v_1} + r\vec{v_2}, r\vec{w_1} + r\vec{w_2})$
$= r(\vec{v_1}, \vec{w_1}) + r(\vec{v_2}, \vec{w_2})$
Identity Scalar $1(\vec{v}, \vec{w}) = (1\vec{v}, 1\vec{w}) = (\vec{v}, \vec{w})$

Example 1: $k^n \times k^m$ for any field k
In fact, $k^n \times k^m \cong k^{n+m}$ if we consider the function $\phi(\langle v_1, v_2, ..., v_n \rangle, \langle w_1, w_2, ..., w_n \rangle) = \langle v_1, ..., v_n, w_1, ..., w_n \rangle$. The proof of ϕ being an isomorphism is trivial.

Theorem 3.5.1: Universal property for direct product

Define $\pi_V : V \times W \to V, \pi_V((\vec{v}, \vec{w})) = \vec{v}$ and $\pi_W : V \times W \to W, \pi_W((\vec{v}, \vec{w})) = \vec{w}$.
Suppose U,V and W are vector spaces over the same field k, and $\phi_V : U \to V$ and $\phi_W : U \to W$ are linear transformations. Then there exists an unique linear transformation $\theta : U \to (V \times W)$ such that $\pi_V \circ \theta = \phi_V$ and $\pi_W \circ \theta = \phi_W$.

Proof It is easy to see that $\theta(\vec{u}) = (\phi_V(\vec{u}), \phi_W(\vec{u}))$. We now have to verify it:

Homomorphism $\theta(\vec{u_1} + \vec{u_2}) = (\phi_V(\vec{u_1} + \vec{u_2}), \phi_W(\vec{u_1} + \vec{u_2})) = (\phi_V(\vec{u_1}) + \phi_V(\vec{u_2}), \phi_W(\vec{w_1}) + \phi_W(\vec{w_2}))$

$$= (\phi_V(\vec{u_1}), \phi_W(\vec{w_1})) + (\phi_V(\vec{u_2}), \phi_W(\vec{w_2}))$$

$$= \theta(\vec{u_1}) + \theta(\vec{u_2})$$

Homogeneity $\theta(r\vec{u}) = (\phi_V(r\vec{u}), \phi_W(r\vec{w})) = (r\phi_V(\vec{u}), r\phi_W(\vec{w})) = r(\phi_V(\vec{u}), \phi_W(\vec{u})) = r\theta(\vec{u})$

The proof of θ being unique is trivial.

We can represent the above result using a diagram:

$$\begin{tikzcd}
U \arrow[drr, bend left, "\phi_V"] \arrow[dr, dashed, "\theta"] \arrow[ddr, bend right, "\phi_W"'] & & \\
& V \times W \arrow[r, "\pi_V"'] \arrow[d, "\pi_W"] & V \\
& W &
\end{tikzcd}$$

Figure 3.5.1: An illustration of Theorem 3.5.1: The existence of $\rightarrow$ implies $\dashrightarrow$.

Definition 3.5.2: Sum of vector spaces

Suppose U is a vector space and $V, W \subseteq U$ are vector subspaces of U. The sum of V and W is $V + W = \{\vec{v} + \vec{w} \mid \vec{v} \in V, \vec{w} \in W\}$. Moreover, if the combination of $\vec{v}$ and $\vec{w}$ is unique for every vector in $V + W$, $V + W = V \oplus W$ is the direct sum of V and W.

We can also talk of the sum and direct sum of multiple subspaces. In fact, the sum and direct sum operations are associative and commutative, i.e. $V + W + X = (V + W) + X = V + (W + X)$ and $V + W = W + V$ for $V, W, X \subseteq U$. The proof of these two properties is left as an exercise to the readers.

Proposition 3.5.2

$V + W \subseteq U$ is a vector subspace of U.

Proof $(\vec{v_1} + \vec{w_1}) + (\vec{v_2} + \vec{w_2}) = (\vec{v_1} + \vec{v_2}) + (\vec{w_1} + \vec{w_2}) \in V + W$ (Associativity, Commutativity)
$r(\vec{v} + \vec{w}) = r\vec{v} + r\vec{w} \in V + W$ (Scalar distributivity)
Hence, by the subspace test, $V + W \subseteq U$.

Proposition 3.5.3

Given subspaces V and W of a vector space U, $V + W = V \oplus W \iff (\exists \vec{v} \in V, \vec{w} \in W : \vec{v} + \vec{w} = \vec{0} \implies \vec{v} = \vec{0}, \vec{w} = \vec{0})$.

Proof ($\leftarrow$) For any $\vec{u} \in V + W$, suppose there exist $\vec{v_1}, \vec{v_2} \in V$ and $\vec{w}_1, \vec{w}_2 \in W$ such that $\vec{u} = \vec{v_1} + \vec{w_1} = \vec{v_2} + \vec{w_2}$. Then by associativity and commutativity, $(\vec{v_1} - \vec{v_2}) + (\vec{w_1} - \vec{w_2}) = \vec{0}$. Hence, $\vec{v_1} = \vec{v_2} = \vec{0}, \vec{w_1} - \vec{w_2} = \vec{0} \implies \vec{v_1} = \vec{v_2}, \vec{w_1} = \vec{w_2}$, which means the combination is unique.
The converse is trivial.

Note that the proposition can be extended to sums of any (finite) number of subspaces, i.e. $V_1 + V_2 + ... + V_n = V_1 \oplus V_2 \oplus ... \oplus V_n$ if and only if for $\vec{v_i} \in V_i$, $\sum \vec{v_i} = 0 \implies \vec{v_i} = \vec{0}$.

Proposition 3.5.4

Given subspaces V and W of a vector space U, $V + W = V \oplus W \iff V \cap W = \{\vec{0}\}$.

Proof ($\rightarrow$) Suppose there exists $\vec{u} \in V \cap W$ and $\vec{u} \neq \vec{0}$. Then $\vec{u} = \vec{0} + \vec{u} = \vec{u} + \vec{0}$, meaning its representation is not unique. This is impossible if $V + W = V \oplus W$.
($\leftarrow$) For any non-zero $\vec{v} \in V$, $-\vec{v} \notin W$. As the additive inverse is unique (Proposition 2.2.3), for $\vec{v} \in V$ and $\vec{w} \in W$, $\vec{v} + \vec{w} = \vec{0} \implies \vec{v} = \vec{0}, \vec{w} = \vec{0}$. By Proposition 3.5.3, $V + W = V \oplus W$.

Note that the leftward direction does not apply to sums of more than two subspaces, i.e. $V_1 \cap V_2 \cap V_3 \cap ... \cap V_n = \{\vec{0}\}$ does not imply a direct sum of the subspaces exists.

Example 1: $I = \{\langle x, 0\rangle \mid x \in k\}, J = \{\langle 0, y\rangle \mid y \in k\} \subset k^2$ for any field k
Obviously, $I \cap J = \{\vec{0}\} = \{\langle 0, 0\rangle\}$ and $I \oplus J = k^2$. In the $\mathbb{R}^2$ vector space, I and J would represent the set of all x-axis and y-axis direction vectors respectively.

Example 2: All even polynomials $E = k[x^2]$ and all odd polynomials $O = k[x^2]x$ for any field k
It is obvious that $E \cap O = 0$ (which is the zero vector in $k[x]$) and $E \oplus O = k[x]$.

Theorem 3.5.2: Universal property for direct sum

Define $\iota_V : V \to V \oplus W, \iota_V(\vec{v}) = \vec{v}$ and $\iota_W : W \to V \oplus W, \iota_W(\vec{w}) = \vec{w}$.
Suppose Z is a vector space with with two linear transformations $\phi_V : V \to Z$ and $\phi_W : W \to Z$. Then there exists an unique linear transformation $\psi : V \oplus W \to Z$ such that $\psi \circ \iota_V = \phi_V$ and $\psi \circ \iota_W = \phi_W$.

Proof It is easy to see that $\psi(\vec{v} + \vec{w}) = \phi_V(\vec{v}) + \phi_W(\vec{w})$. This function is well-defined due to the definition of the direct sum. We now have to prove ψ is a linear transformation:
Homomorphism $\psi(\vec{v} + \vec{w}) = \phi_V(\vec{v}) + \phi_W(\vec{w}) = \phi_V(\vec{v}) + \vec{0} + \phi_W(\vec{w}) + \vec{0} = \phi_V(\vec{v}) + \phi_V(\vec{0}) + \phi_W(\vec{w}) + \phi_W(\vec{0}) = \psi(\vec{v} + \vec{0}) + \psi(\vec{w} + \vec{0}) = \psi(\vec{v}) + \psi(\vec{w})$. (Proposition 2.4.2)
Homogeneity $\psi(r(\vec{v}+\vec{w})) = \psi(r\vec{v}+r\vec{w}) = \phi_V(r\vec{v})+\phi_W(r\vec{w}) = r\phi_V(\vec{v})+r\phi_W(\vec{w}) = r(\phi_V(\vec{v})+\phi_W(\vec{w})) = r\psi(\vec{v} + \vec{w})$ (Scalar distributivity)
Finally, to see that ψ is unique, for any other such linear transformation ψ', $\psi'(\vec{v} + \vec{w}) = \psi'(\vec{v}) + \psi'(\vec{w}) = \phi_V(\vec{v}) + \phi_W(\vec{w}) = \psi(\vec{v} + \vec{w})$. (Homomorphism)

Again, we can represent this result using a diagram:

$$\begin{tikzcd}
V \arrow[dr, "\iota_V"] \arrow[drr, bend left, "\phi_V"] & & \\
& V \oplus W \arrow[r, dashed, "\psi"] & Z \\
W \arrow[ur, "\iota_W"] \arrow[urr, bend right, "\phi_W"'] & &
\end{tikzcd}$$

Figure 3.5.2: An illustration of Theorem 3.5.2: The existence of → implies ⇢.

3.6 Linear Independence

Definition 3.6.1: Linear independence

Suppose V is a vector space over k and $S = \{\vec{v_i}\} \subsetneq V$:

- S is linearly independent if $r_i \in k, \sum r_i\vec{v_i} = \vec{0} \implies r_i = \vec{0}$.
- S is linearly dependent if $\exists r_i \in k$ and $\{r_i\} \neq \{0\}$ such that $\sum r_i\vec{v_i} = \vec{0}$.

Note that S is either linearly independent or linearly dependent.

Example 1: $\{\langle 1,2\rangle, \langle 3,-4\rangle\} \subset \mathbb{R}^2$ **over** $\mathbb{R}$
By solving $r\langle 1,2\rangle + s\langle 3,-4\rangle = \langle 0,0\rangle \implies r + 3s = 0, 2r - 4s = 0$, we get $r = 0, s = 0$, hence it is linearly independent.

Example 2: $\{x^2+2, 3x+5\} \subset \mathbb{R}[x]$ **over** $\mathbb{R}$
$r(x^2+2) + s(3x+5) = 0 \implies rx^2 + 3sx + (2r+5s) = 0$, by comparing coefficients, we get $r = 0, s = 0$. Hence, it is linearly independent.

Proposition 3.6.1

If S is finite, S is linearly dependent $\iff \exists \vec{v_j} \in S$ such that $\vec{v_j}$ can be expressed as a linear combination of $\vec{v_{i\neq j}}$.

Proof (→) Find the weight $r_j \neq 0$ in $\{r_i\}$.

$$\begin{aligned}
r_j\vec{v_j} + \sum_{i\neq j} r_i\vec{v_i} = 0 &\implies {r_j}^{-1}(r_j v_j + \sum_{i\neq j} r_i\vec{v_i}) = {r_j}^{-1} * 0 = 0 \text{ (Proposition 3.2.1)} \\
&\implies v_j + \sum_{i\neq j} {r_j}^{-1} r_i\vec{v_i} = 0 \text{ (Scalar distributivity, Associativity)} \\
&\implies v_j = -\sum_{i\neq j} {r_j}^{-1} r_i\vec{v_i} = (-1)\sum_{i\neq j} {r_j}^{-1} r_i\vec{v_i} \text{ (Proposition 3.2.3)} \\
&\implies v_j = \sum_{i\neq j} -{r_j}^{-1} r_i\vec{v_i} \text{ (Scalar distributivity, Associativity, Proposition 3.2.3)}
\end{aligned}$$

(←) Suppose $v_j = \sum_{i\neq j} r_i\vec{v_i}$. Then by associativity and commutativity, $1 * \vec{v_j} + \sum_{i\neq j} -r_i\vec{v_i} = 0$. (To see that $1 \neq 0$, consider Proposition 3.2.2 and the definition of fields.)

Notice that the above proof only works when S is finite because properties of vector spaces are only guaranteed to apply to finitely many vectors at a time. This can be shown by induction.

3.7 Bases

Definition 3.7.1: Basis

An ordered set of vectors $B \subseteq V$ is a basis for a vector space V if:

1. B is linearly independent (Linear independence)
2. $\operatorname{span} B = V$ (Spanning property)

There are some standard bases (or canonical bases) for common vector spaces:

- $\{\langle 1,0,...,0\rangle, \langle 0,1,0,...,0\rangle, \langle 0,...,0,1\rangle\} = \{\vec{e_1}, \vec{e_2}, ..., \vec{e_n}\} = \{\hat{\imath}, \hat{\jmath}, \hat{k}, ...\}$ for k^n, where k is any field
- $\{1, x, x^2, ...\}$ for $k[x]$

The proofs of the above being bases are easy and left as an exercise to the readers.

Theorem 3.7.1: Unique representation theorem

B is a finite basis $\iff$ every vector $\vec{v} \in V$ can be expressed as an unique linear combination of B

Proof **(→)** By the spanning property, it is guaranteed that v can be expressed as a linear combination of B. We only have to show it is unique:
Suppose $\exists \{r_i\}, \{s_i\} \subseteq k$ such that $r_1\vec{v_1} + r_2\vec{v_2} + ... + r_n\vec{v_n} = \vec{v}$ and $s_1\vec{v_1} + s_2\vec{v_2} + ... + s_n\vec{v_n} = \vec{v}$, where $\vec{v_i} \in B$.
$r_1\vec{v_1} + r_2\vec{v_2} + ... + r_n\vec{v_n} - (s_1\vec{v_1} + s_2\vec{v_2} + ... + s_n\vec{v_n} = \vec{v}) = \vec{v} - \vec{v} = \vec{0}$
By what we did in Proposition 3.6.1 and vector distributivity, $(r_1 - s_1)\vec{v_1} + ... + (r_n - s_n)\vec{v_n} = \vec{0}$.
As B is linearly independent, $r_i - s_i = 0 \implies -s_i + r_i = 0 \implies r_i = s_i$. (Commutativity)
(←) The spanning property is fulfilled by definition.
For linear independence, by Proposition 3.2.2, $0 * \vec{v_1} + 0 * \vec{v_2} + ... + 0 * \vec{v_n} = \vec{0}$.
As this is the unique linear combination, we know that B must be linearly independent.

Definition 3.7.2: Coordinate map

Given a vector space V over k with basis $B = \{\vec{v_1}, \vec{v_2}, ..., \vec{v_n}\}$, the coordinate map $\gamma_B : V \to k^n$ is defined such that $\gamma_B(\vec{v}) = \langle r_1, r_2, ..., r_n\rangle$, where $r_1\vec{v_1} + r_2\vec{v_2} + ... + r_n\vec{v_n} = \vec{v}$,

In other words, the coordinate map finds the weights of the linear combination that makes up $\vec{v}$.

Proposition 3.7.1

The coordinate map (with respect to a finite basis) is a linear transformation.

Proof By the properties of vector spaces, for $\vec{v} = \sum r_i\vec{v_i}$, $\vec{w} = \sum s_i\vec{v_i}$ and $t \in k$:
Homomorphism $\sum r_i\vec{v_i} + \sum s_i\vec{v_i} = \sum(r_i + s_i)\vec{v_i} \implies \gamma_B(\vec{v} + \vec{w}) = \gamma_B(\vec{v}) + \gamma_B(\vec{w})$
Homogeneity $\sum tr_i\vec{v_i} = t\sum r_i\vec{v_i} \implies \gamma_B(t\vec{v}) = t\gamma_B(\vec{v})$

Let's look at an example to see how the above applies: $B = \{\langle 1,3\rangle, \langle -2,-4\rangle\} \subset \mathbb{R}^2$.
First we have to check if it is a basis:
Linear independence
As neither vector is a scalar multiple of the other, by Proposition 3.6.1, B is linearly independent.
Spanning property
For every $\langle a,b\rangle \in \mathbb{R}^2$, we have the equation $x\langle 1,3\rangle + y\langle -2,-4\rangle = \langle a,b\rangle$, which is the same as:

$$\begin{cases} x - 2y = a & (1) \\ 3x - 4y = b & (2) \end{cases}$$

(1): $x - 2y = a \implies x = 2y + a$
→(3): $3(2y + a) - 4y = b \implies 2y = b - 3a \implies y = \frac{b-3a}{2}$
→(1): $x - 2(\frac{b-3a}{2}) = a \implies x - 2b + 3a = a \implies x = 2b - 2a$
Therefore, every vector can be written as a linear combination of the two vectors.
Hence, B is a basis for $\mathbb{R}^2$. Moreover, with the solution above, we know that $\gamma_B(\langle a, b\rangle) = \langle \frac{b-3a}{2}, 2b - 2a\rangle$.

3.8 Dimension

Suppose V is a vector space over k, then we can define the following two sets:

Definition 3.8.1: Maximal linearly independent set

B is a maximal linearly independent set if B is linearly independent and every S such that $B \subset S \subseteq V$ is linearly dependent.

Definition 3.8.2: Minimal spanning set

B is a minimal spanning set if $\operatorname{span} B = V$ and for every $S \subset B$, $\operatorname{span} S \subsetneq V$.

Theorem 3.8.1

If B is finite, the following are equivalent:

1. B is a maximal linearly independent set
2. B is a minimal spanning set
3. B is a basis for V

Proof (1)→(3) Suppose B is not spanning, i.e. $\exists \vec{v} \in V$ such that $\forall r_i \in k : \vec{v_i} \in V, \sum r_i\vec{v_i} \neq \vec{v}$.
Notice this means that $\sum r_i\vec{v_1} \neq r\vec{v}$ for all $r \in k \setminus \{0\}$, because otherwise $r^{-1}\sum r_i\vec{v_i} = r^{-1}r\vec{v} \implies \sum r^{-1}r_i\vec{v_i} = \vec{v}$. (Scalar distributivity, Associativity)
Hence, $(\sum r_i\vec{v_i}) + r\vec{v} \neq 0$ for $r \neq 0$, meaning $B \cup \vec{v}$ is a linearly independent set, which is impossible.
Therefore, B is both linearly independent and spanning, thus it is a basis for V.
(3)→(1) For any vector $\vec{v} \in V$, there is a linear combination of B that can represent $\vec{v}$.
By Proposition 3.6.1, this means that $B \cup \vec{v}$ and hence any superset of B is linearly dependent.
(2)→(3) Suppose B is not linearly independent.
Then by Proposition 3.6.1, $\exists \vec{v_j} \in B$ such that $B = \sum_{i \neq j} s_i\vec{v_i}$, where $s_i \in k$.
As B is spanning, for any $v \in V$, $\exists r_i \in k$ such that $\sum r_i\vec{v_i} = \vec{v} \implies (\sum_{i\neq j} r_i\vec{v_i}) + r_j\vec{v_j} = \vec{v} \implies \sum_{i\neq j} r_i\vec{v_i} + r_j\sum_{i\neq j} s_i\vec{v_i} = \vec{v} \implies \sum_{i\neq j}(r_i + r_js_i)\vec{v_i} = \vec{v}$. (Properties of vector spaces)
In other words, $\{\vec{v}_{i\neq j}\} \subset B$ is already a spanning set, which is a contradiction.
As B is both linearly independent and spanning, B is a basis for V.
(3)→(2) Suppose $\exists B' \subset B$ such that B' is a spanning set.
Then for any $\vec{v} \in B \setminus B'$, there is a linear combination of B' that can represent $\vec{v}$.
By Proposition 3.6.1, this means that B is not linearly independent, which is a contradiction.

Theorem 3.8.2: Basis extension theorem

If $S = \{\vec{v_1}, \vec{v_2}, ..., \vec{v_n}\} \subseteq V$ is linearly independent and $T = \{\vec{w_1}, \vec{w_2}, ..., \vec{w_n}\} \subseteq V$ is spanning, then $\exists T' \subseteq T$ such that $S \cup T'$ is a basis for V.

Proof Let $T' = \emptyset$ at the start. Construct the set by adding $\vec{w_i} \notin \operatorname{span} S \cup T'$ to T' until this is impossible. By Proposition 3.6.1, we know that $S \cup T'$ is linearly independent.

For any $\vec{v} \in V$, T is spanning $\implies \exists r_i \in k$ such that $\sum_{i\neq j} r_i\vec{w_i} + \sum r_j\vec{w_j} = \vec{v}$. (Commutativity)
By definition, $\vec{w}_{i\neq j} \in \operatorname{span} S \cup T'$, thus by the properties of vector spaces, $\vec{v}$ can be expressed as a linear combination of $S \cup T'$, which means $S \cup T'$ is spanning.

Theorem 3.8.3: Spanning set theorem

If S is a finite spanning set, a basis can be formed by removing elements from S that can be expressed as a linear combination of the others one at a time until this is impossible.

Proof By Proposition 3.6.1, we know that the end result is linearly independent.
For any $\vec{v} \in V$, $\exists r_i \in k$ such that $\sum r_i\vec{v_i} = \vec{v}$, where $\vec{v_i} \in S$. If we remove a linearly dependent element $\vec{v_j}$ from S, then there exists $\{s_i\} \subseteq k$ such that:

$$\vec{v} = \sum_{i\neq j} s_i\vec{r_i} + r_j\vec{v_j} = \sum_{i\neq j} r_i\vec{v_i} + r_j(\sum_{i\neq j} s_i\vec{v_i}) = \sum_{i\neq j}(r_i + r_js_i)\vec{v_i}$$

This means that the new set is still spanning.
Hence, by induction, we know that the end result will be spanning too.

Theorem 3.8.4: Replacement theorem

If $S = \{\vec{v_1}, \vec{v_2}, ..., \vec{v_n}\}$ is a spanning set and $\{\vec{w_1}, \vec{w_2}, ..., \vec{w_m}\}$ is linearly independent, then $|S| = n \geq m$.

Proof Suppose $n < m$. For any $i \in \{1, ..., n\}$, we can express $\vec{w_i}$ as $r_1\vec{v_1} + r_2\vec{v_2}... + r_n\vec{v_n}$ for some $r_j \in k$. Note that for any $\vec{u} \in V$, there exists $\{s_j\} \subseteq k$ such that:

$$\begin{aligned}\vec{w} &= s_1\vec{v_1} + ... + s_i\vec{v_i} + ...s_n\vec{v_n}\\ &= (s_1 - r_1)\vec{v_1} + ... + \vec{w_i} + ... + (s_n - r_n)\vec{v_n}\end{aligned}$$

In other words, we can ”replace” $\vec{v_i}$ by $\vec{w_i}$ and still retain the span of S. By assumption, we can replace the entirety of S with $\vec{w_1}, ..., \vec{w_n}$. However, this would mean that $\vec{w_m}$ can be represented as a linear combination of the vectors, which per Proposition 3.6.1, contradicts linear independence.

Definition 3.8.3: Dimension

The dimension of a vector space V is $\dim V = |B|$, where B is a basis for V.

Proposition 3.8.1

$\dim V$ is well-defined.

Proof In other words, we have to prove that the size of every basis for a vector space is the same.
Suppose B and B' are two bases for V.
As B is spanning and B' is linearly independent, by the replacement theorem, we have $|B| \geq |B'|$.
Meanwhile, as B is linearly independent and B' is spanning, we also have $|B| \leq |B'|$.
Therefore, $|B| = |B'|$.

Note that the dimension of a vector space (e.g. $k[x]$) can be infinite. The case when the dimension is infinite should be dealt with separately. We will assume the dimension henceforth unless otherwise specified.

Example 1: The trivial vector space $\{\vec{0}\}$
The only basis for the trivial vector space is the empty set, because we define the sum of an empty set as $\vec{0}$ by convention. Hence, the dimension of the trivial vector space is 0. The converse is also true: the only vector space with dimension 0 is the trivial vector space.

Example 2: $V = \{\langle v_1, v_2, ..., v_n\rangle \mid v_i \in k, v_1 + v_2 + ... + v_n = 0\} \subset k^n$ **for any field** k
In Section 3.3, we have proved that V is a subspace of $\mathbb{R}^n$. We now have to find its basis and dimension. Let $B_V = \{\langle 1, 0, ..., 0, -1\rangle, \langle 0, 1, 0, ..., 0, -1\rangle, \langle 0, ..., 0, 1, -1\rangle\}$. It is easy to see that B_V is both linearly independent and spanning.
Therefore, B_V is a basis for V and $\dim V = n - 1$.

Example 3: $S = \{f \mid f(c) = 0\} \subset \mathbb{R}_n[x]$**, where** $c \in \mathbb{R}$
$f \in S$ can be written as $(x - c)(r_0 + r_1x + ... + r_{n-1}x^{n-1}) = r_0(x - c) + r_1x(x - c) + ... + r_{n-1}x^{n-1}(x - c)$. Let $B_S = \{x - c, x(x - c), ..., x^{n-1}(x - c)\}$, which is obviously spanning by the above. It is also easy to see that B_S is linearly independent because every element has different degrees.
Therefore, B_S is a basis for S and $\dim S = n - 1$.

Proposition 3.8.2

W is a subspace of $V \implies \dim W \leq \dim V$

Proof Suppose B is a basis for V. Be definition, this means that B is spanning for V and hence W. Hence, we can find a $B' \subseteq B$ that is a minimal spanning set for V.
By Theorem 3.8.1, B' is a basis, thus $\dim B' = |B'| \leq |B| = \dim B$.

Proposition 3.8.3

Suppose $\dim V = n$ and $|S \subseteq V| = m$. Then

1. $m > n \implies S$ is linearly dependent
2. $m < n \implies S$ is not spanning

Proof We will prove them by their converses.
(1) Suppose S is linearly independent. Then $\exists S' \supseteq S$ such that S' is a maximal linearly independent set. By Theorem 3.8.1, S' is a basis. Therefore, $\dim V = |S'| = n \geq m = |S|$.
(2) Suppose S is spanning. Then $\exists S' \subseteq S$ such that S' is a minimal spanning set. By Theorem 3.8.1, S' is a basis. Therefore, $\dim V = |S'| = n \leq m = |S|$

Theorem 3.8.5: Basis theorem

Suppose $\dim V = n$ and $B = \{\vec{v_1}, \vec{v_2}, ..., \vec{v_n}\}$. Then the following are equivalent:

1. B is linearly independent
2. B is spanning
3. B is a basis for V

Proof (1)→(3) By Proposition 3.8.3, we know that S is linearly independent implies $m \leq n$. Hence, as $|B| = n$, B is a maximal linearly independent set, which by Theorem 3.8.1 means B is a basis for V.
(2)→(3) By Proposition 3.8.3, we know that S is spanning implies $m \geq n$. Hence, as $|B| = n$, B is a minimal spanning, which by Theorem 3.8.1 means B is a basis for V.
(3)→(1,2) True by definition.

Theorem 3.8.6: Dimension formula

If V and W are subspaces of U and $\dim U < \infty$, then $\dim V + \dim W = \dim(V + W) + \dim(V \cap W)$.

Proof It is trivial to show that $X = V \cap W$ is a vector space. We first find a basis $B_X = \{\vec{x_1}, \vec{x_2}, ..., \vec{x_n}\}$ for X.
Then we find a finite spanning set for V. By the basis extension theorem, $\exists B_V = \{\vec{v_1}, \vec{v_2}, ..., \vec{v_m}\} \subseteq V$ such that $B_X \cup B_V$ is a basis for V. Notice that $B_V \subseteq V \setminus X$ as otherwise $\vec{v_i}$ can be expressed as a linear combination of $\vec{x_i}$, which by Proposition 3.6.1 is impossible.
We can do the same for W and obtain $B_W = \{\vec{w_1}, \vec{w_2}, ..., \vec{w_k}\} \subseteq W$ where $B_W \subseteq W \setminus X$.
We now check if $B_X \cup B_V \cup B_W$ is a basis for $V + W$. As $B_X \cap B_V$ is a basis for V and $B_X \cap B_W$ is a basis for W, it is easy to see that it is spanning for $V + W$.
To see if it is linearly independent, we can decompose the set into three components $\vec{x} = \sum \vec{x_i}$, $\vec{v} = \sum \vec{v_i}$ and $\vec{w} = \sum \vec{w_i}$.
Hence, $\vec{x} + \vec{v} + \vec{w} = \vec{0} \implies \vec{v} = -(\vec{x} + \vec{w}) \in W$ and $\vec{w} = -(\vec{x} + \vec{v}) \in V \implies \vec{x} = \vec{v} = \vec{w} = \vec{0}$ (Properties of vector spaces). As B_X,B_V and B_W are all linearly independent, this means the only solution for $\sum B_X \cup B_V \cup B_W = 0$ is the trivial solution, which means it is linearly independent.
Therefore, $\dim V + \dim W = (n + m) + (n + k) = (n + m + k) + n = \dim(V + W) + \dim(V \cap W)$.

There are two corollaries of the dimension formula:

1. As the dimension of the trivial vector space is 0, $\dim(V \oplus W) = \dim V + \dim W$.
2. If we let $V' = \{\langle \vec{v}, 0 \rangle \mid \vec{v} \in V\}$ and $W' = \{\langle 0, \vec{w} \rangle \mid \vec{w} \in W\}$, by the above, we get $\dim(V \times W) = \dim(V' \oplus W') = \dim V + \dim W$.

3.9 Rank and Nullity

For any linear transformation $\phi : (V, +,) \to (W, +,)$, by Proposition 3.4.2 and 3.4.3, we know that the kernel and image of ϕ are vector subspaces of V and W respectively. Hence, we can define the following:

Definition 3.9.1: Nullity

The nullity of ϕ is the dimension of $\ker \phi$.

Definition 3.9.2: Rank

The rank of ϕ is the dimension of $\operatorname{im} \phi$.

Theorem 3.9.1: Rank-nullity theorem

If $\dim V < \infty$, then $\dim \ker \phi + \dim \operatorname{im} \phi = \dim V$.

Proof We first find a basis for $\ker \phi$. Suppose it is $B = \{\vec{v_1}, \vec{v_2}, ..., \vec{v_n}\}$. Then $\dim \ker \phi = n$.
Obviously, B is still linearly independent in V. Hence, by Theorem 3.8.1, there is a maximal linearly independent $B' \supseteq B$, which is also the basis for V.
Suppose $B' \cap B = \{\vec{v'}_1, \vec{v'}_2, ..., \vec{v'}_k\}$. Then $\dim V = n + k$. Hence, we now have to prove that $\phi(B' \cap B)$ is a basis for $\operatorname{im} \phi$, which would mean $\dim \operatorname{im} \phi = k$.
By definition, $\phi(B')$ is a finite spanning set for $\operatorname{im} \phi$. By the spanning set theorem, we can construct a basis from $\phi(B)$ by removing linearly dependent elements.
Considering that as $B \subseteq \ker \phi$, $\phi(B) = \{0\}$, we can remove all elements of B from the set.
For the remaining elements (i.e. $B' \cap B$), suppose $\phi(B' \cap B)$ is not linearly independent. Then

$\exists\{r_i\} \neq \{0\}$ such that $\sum r_i\phi(\vec{v'}_i) = \sum \phi(r_i\vec{v'}_i) = \phi(\sum r_i\vec{v'}_i) = 0$. (Homogeneity, Homomorphism)
Therefore, $\sum r_i\vec{v'}_i \in \ker\phi \implies \exists\{s_i\} \neq \{0\} : \sum r_i\vec{v'}_i = \sum s_i\vec{v_i} \implies \sum r_i\vec{v'}_i - \sum s_i\vec{v_i} = \vec{0}$.
By the properties of vector spaces, this means that B is linearly dependent, which is a contradiction.
Therefore, we have shown that $B \cap B'$ is already linearly independent and hence a basis.

Four corollaries directly follow this theorem:

1. If ϕ is injective, by Theorem 2.4.1, $\ker\phi = \{0\}$ and hence $\dim\operatorname{im}\phi = \dim V$. Furthermore, as $\operatorname{im}\phi \subseteq W$, by Proposition 3.8.2, $\dim V \leq \dim W$.
2. If ϕ is surjective, $\operatorname{im}\phi = W$, thus $\dim\ker\phi = \dim V - \dim W$. Furthermore, $\dim V \geq \dim W$.
3. If ϕ is bijective, in other words, if $V \cong W$, by (1) and (2), $\dim V = \dim W$.
4. Given that $\dim V = \dim W < \infty$, if ϕ is injective, by (1) and (3), $\dim W = \dim\operatorname{im}\phi = \dim V$. Suppose B is a basis for $\operatorname{im}\phi \subseteq W$. By definition, B is linearly independent. Then by the basis theorem, B is also a basis for W, implying that $W = \operatorname{im}\phi$ or ϕ is surjective. On the other hand, if ϕ is surjective, by (2) and (3), $\dim\ker\phi = \dim V - \dim W = 0$, which means $\ker\phi$ is the trivial vector space. By Theorem 2.4.1, ϕ is injective. In short, if $\dim V = \dim W < \infty$, the injectivity and surjectivity of ϕ are equivalent.

4 Matrices

4.1 Matrices

Definition 4.1.1: Matrix

Given a field k, a $n \times m$ matrix A is a rectangular array of elements of k with m rows and n columns:

$$A = \begin{pmatrix} a_{1,1} & a_{1,2} & \ldots & a_{1,m} \\ a_{2,1} & a_{2,2} & \ldots & a_{2,m} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n,1} & a_{n,2} & \ldots & a_{n,m} \end{pmatrix} = (a_{i,j})$$

If the field under a matrix is not specified, it is assumed to be $\mathbb{R}$.
There are some usual notations we use on matrices for simplicity:

- $\operatorname{Mat}_{n\times m} k$ is the set of all $n \times m$ matrices over k
- $A_i = \langle a_{i,1}, a_{i,2}, \ldots, a_{i,m} \rangle \in k^m$ is the i-th row of A
- $A^i = \langle a_{1,i}, a_{2,i}, \ldots, a_{n,i} \rangle \in k^n$ is the i-th column of A

Definition 4.1.2: Matrix addition

If $A = (a_{i,j}), B = (b_{i,j}) \in \operatorname{Mat}_{n\times m} k$, then $A + B = (a_{i,j} + b_{i,j}) \in \operatorname{Mat}_{n\times m} k$.

In simpler terms, matrix addition is adding entries in the same position together. For example:

1. $\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} + \begin{pmatrix} 5 & 6 \\ 7 & 8 \end{pmatrix} = \begin{pmatrix} 6 & 8 \\ 10 & 12 \end{pmatrix} \in \operatorname{Mat}_{2\times 2} \mathbb{R}$
2. $\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \oplus \begin{pmatrix} 2 & 3 \\ 4 & 5 \end{pmatrix} = \begin{pmatrix} 3 & 0 \\ 2 & 4 \end{pmatrix} \in \operatorname{Mat}_{2\times 2} \mathbb{Z}_5$

Definition 4.1.3: Matrix scalar multiplication

Given $r \in k$ and $A = (a_{i,j}) \in \operatorname{Mat}_{n\times m} k$, $rA = (ra_{i,j}) \in \operatorname{Mat}_{n\times m} k$.

Some examples:

1. $2\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} = \begin{pmatrix} 2 & 4 & 6 \\ 8 & 10 & 12 \end{pmatrix} \in \operatorname{Mat}_{2\times 3} \mathbb{R}$
2. $2 \otimes \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} = \begin{pmatrix} 2 & 4 & 6 \\ 1 & 3 & 5 \end{pmatrix} \in \operatorname{Mat}_{2\times 3} \mathbb{Z}_7$

Theorem 4.1.1

For any field k, $(\operatorname{Mat}_{n\times m} k, +,)$ forms a nm-dimensional vector space over k.

Proof *Additive Abelian Group* True as k is a field, with the zero matrix being the identity.
Scalar associativity For $r, s \in k$, $r(sA) = r(s(a_{i,j})) = r(sa_{i,j}) = (rsa_{i,j}) = rs(a_{i,j})$
Vector distributivity $(r + s)A = (r + s)(a_{i,j}) = ((r + s)a_{i,j}) = (ra_{i,j} + sa_{i,j}) = rA + sA$
Scalar distributivity $r(A + B) = r((a_{i,j}) + (b_{i,j})) = (r(a_{i,j} + b_{i,j})) = (ra_{i,j} + rb_{i,j}) = rA + rB$
Identity Scalar $1A = 1(a_{i,j}) = (1a_{i,j}) = (a_{i,j}) = A$
Therefore, it is a vector space. To determine its dimension, consider that a $n \times m$ matrix is composed of n m-dimensional vectors, i.e. $\operatorname{Mat}_{n\times m} k \cong \underbrace{k^m \times \ldots \times k^m}_{n \text{ times}}$.
Hence, by the dimension formula, $\dim \operatorname{Mat}_{n\times m} k = nm$.

Definition 4.1.4: Dot Product

Given two column matrices $A = \langle v_1, v_2, ..., v_n\rangle$ and $B = \langle w_1, w_2, ..., w_n\rangle$, where v_i and w_i belong to the same field k, the dot product $A \cdot B = v_1w_1 + v_2w_2 + ... + v_nw_n$.

Some Examples:

1. $\langle 1, 2, 3\rangle \cdot \langle 4, 5, 6\rangle = 4 + 10 + 18 = 32 \in \mathbb{R}$
2. $\langle 1, 2, 3\rangle \cdot \langle 4, 5, 6\rangle = 4 \oplus 3 \oplus 4 = 4 \in \mathbb{Z}_7$

By the properties of fields, we can show the following properties of the dot product:

- $A \cdot B = v_1w_1 + v_2w_2 + ... + v_nw_n = w_1v_1 + w_2v_2 + ... + w_nv_n = B \cdot A$ (Commutativity)
- $rA \cdot sB = rv_1sw_1 + rv_2sw_2 + ... + rv_nsw_n = rsv_1w_1 + rsv_2w_2 + ... + rsv_nw_n = rs(A \cdot B)$ (Homogeneity)
- $(A+B)\cdot C = (v_1+w_1)u_1+(v_2+w_2)u_2+...+(v_n+w_n)u_n = v_1u_1+w_1u_1+v_2u_2+w_2u_2+...+v_nu_n+w_nu_n = A \cdot C + B \cdot C$
- By commutativity, we know that $A \cdot (B + C) = A \cdot B + A \cdot C$. In other words, the dot product is homomorphic in both arguments.

We will talk more about the dot product and its generalisation in later sections.

Definition 4.1.5: Matrix multiplication

Given $A = (a_{i,j}) \in \operatorname{Mat}_{n\times m} k$ and $B = (b_{i,j}) \in \operatorname{Mat}_{m\times k} k$, $AB = (A_i \cdot B^j) = (\sum_k a_{i,k}b_{k,j}) \in \operatorname{Mat}_{n\times k} k$.

This definition is very hard to understand without examples:

1. $\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} \in \operatorname{Mat}_{2\times 3}\mathbb{R} \begin{pmatrix} 7 & 8 \\ 9 & 10 \\ 11 & 12 \end{pmatrix} \in \operatorname{Mat}_{3\times 2}\mathbb{R} = \begin{pmatrix} 58 & 64 \\ 139 & 154 \end{pmatrix} \in \operatorname{Mat}_{2\times 2}\mathbb{R}$
2. $\begin{pmatrix} 1 & 2 \\ 4 & 5 \end{pmatrix} \in \operatorname{Mat}_{2\times 2}\mathbb{R} \begin{pmatrix} 6 & 7 \\ 8 & 9 \end{pmatrix} \in \operatorname{Mat}_{2\times 2}\mathbb{R} = \begin{pmatrix} 19 & 22 \\ 43 & 50 \end{pmatrix} \in \operatorname{Mat}_{2\times 2}\mathbb{R}$

Note that matrix multiplication is not necessarily commutative, even if the shapes are compatible.

Definition 4.1.6: Square matrix

A is a square matrix if $A \in M_n\ k = \operatorname{Mat}_{n\times n} k$.

For example, $\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$ is a square matrix but $\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}$ is not.

Definition 4.1.7: Diagonal matrix

Given a square matrix $A = (a_{i,j})$, if its non-zero entries are only located on the diagonal (i.e. $a_{i\neq j} = 0$), then A is a diagonal matrix.

For example, $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 0 \end{pmatrix}$ is a diagonal matrix but $\begin{pmatrix} 1 & 0 \\ 2 & 3 \end{pmatrix}$ is not.

Definition 4.1.8: Identity matrix

An $n \times n$ identity matrix I_n is a diagonal matrix of which the entries on the diagonal are all 1.

For example, $I_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ and $I_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$.

Proposition 4.1.1

If $A \in \mathrm{Mat}_{n\times m}$, then $I_nA = AI_m = A$.

Proof Let $A = (a_{i,j})$. Then $I_nA = (I_i \cdot A^j) = (\langle \underbrace{0, ..., 0}_{i-1 \text{ items}}, 1, 0, ..., 0\rangle \cdot \langle a_{1,j}, a_{2,j}, ..., a_{m,j}\rangle) = (a_{i,j}) = A$.

Similarly, $AI_m = (A_i \cdot I_j) = (a_{i,j}) = A$.

Proposition 4.1.2

If the shapes of the matrices are compatible, for any $r \in k$, $r(AB) = (rA)B = A(rB)$.

Proof $r(AB) = (r(A_i \cdot B^j)) = ((rA_i) \cdot B^j) = (rA)B$
$= (A_i \cdot (rB_j)) = A(rB)$ (Homomorphism)

Proposition 4.1.3: Matrix distributivity

If the shapes of the matrices are compatible, $A(B + C) = AB + AC$ and $(A + B)C = AC + BC$.

Proof $A(B + C) = (A_i \cdot (B + C)^j) = (A_i \cdot (B^j + C^j)) = (A_i \cdot B^j + A_i \cdot C^j) = AB + AC$.
The proof for the second case is similar (by using the homomorphism property of the dot product).

Proposition 4.1.2 and Proposition 4.1.3 prove that matrix multiplication is both homomorphic and homogeneous. Together with Theorem 4.1.1, matrix multiplication can be thought as a linear transformation, i.e. for $A \in \mathrm{Mat}_{n\times m}\, k$ and $B \in \mathrm{Mat}_{m\times k}\, k$, $AB = \phi_A(B) = \phi_B(A)$, where $\phi_A : \mathrm{Mat}_{m\times k}\, k \to \mathrm{Mat}_{n\times k}\, k$ and $\phi_B : \mathrm{Mat}_{n\times m}\, k \to \mathrm{Mat}_{n\times k}\, k$ are linear transformations. In fact, as both matrices can be interpreted as a linear transformation for the other, matrix multiplication can also be considered as a bilinear map.

Proposition 4.1.4: Associativity of matrix mulitplication

If the shapes of the matrices are compatible, $(AB)C = A(BC)$.

Proof Suppose $A = (a_{i,j})$, $B = (b_{i,j}$ and $C = (c_{i,j})$.
$(AB)C = (\sum_k a_{i,k}b_{k,j})C = (\sum_t \sum_k a_{i,k}b_{k,t}c_{t,j}) = (\sum_k \sum_t a_{i,k}b_{k,t}c_{t,j}) = A(\sum_t b_{i,t}c_{t,j}) = A(BC)$
The summation signs can be swapped as the products are just sumed in a different order.

As you can see, this proof is very complicated and difficult to understand. A more elegant proof will be presented in a later section after we introduce the space of linear transformations.

Theorem 4.1.2: Ring of square matrices

$M_n\, k$ forms a ring under matrix addition and multiplication.

Proof *Additive Abelian Group* True as k is a field.
Multiplicative Monoid

- *Closure* True by the definition of matrix multiplication.
- *Associativity* Proved in Proposition 4.1.4.
- *Identity* $1 = I_n$

Distributivity Proved in Proposition 4.1.3.

Definition 4.1.9: Transpose

For $A = (a_{i,j}) \in \operatorname{Mat}_{n\times m} k$, the transpose of A is $A^T = (a_{j,i}) \in \operatorname{Mat}_{m\times n}$.

For example, the transpose of $\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}$ is $\begin{pmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{pmatrix}$.

If a matrix is identical to its transpose, it is symmetric, e.g. $\begin{pmatrix} 1 & 2 \\ 2 & 3 \end{pmatrix}$.

Proposition 4.1.5

For any $A, B \in \operatorname{Mat}_{n\times m} k$, $(A+B)^T = A^T + B^T$.

Proof Let $A = (a_{i,j})$ and $B = (b_{i,j})$. Then $(A+B)^T = (a_{i,j} + b_{i,j})^T = (a_{j,i} + b_{j,i}) = A^T + B^T$.

Proposition 4.1.6

For $r \in k$ and $A \in \operatorname{Mat}_{n\times m} k$, $(rA)^T = rA^T$.

Proof Let $A = (a_{i,j})$. Then $(rA)^T = (ra_{i,j})^T = (ra_{j,i}) = rA^T$.

Similar to matrix multiplication, Proposition 4.1.5 and Proposition 4.1.6 show that the transpose operation is both homomorphic and homogeneous. Hence, the transpose operation is a linear transformation.

Proposition 4.1.7

If the shapes of the matrices are compatible, $(AB)^T = B^T A^T$.

Proof $(AB)^T = (A_i \cdot B^j)^T = (A_j \cdot B^i) = (B^i \cdot A_j) = B^T A^T$ (Commutativity)

Again, this can be proved more elegantly when we introduce the space of linear transformations.

4.2 Invertibility

Definition 4.2.1: Invertible matrix

A square matrix $A \in M_n\ k$ is invertible (or non-singular) if there exists $A^{-1} \in M_n\ k$ such that $AA^{-1} = A^{-1}A = I_n$.

Definition 4.2.2: General linear group

The general linear group of rank-n matrices over k $GL_n\ k$ is the set of all invertible matrices in $M_n\ k$.

Proposition 4.2.1

$GL_n\ k$ is a group under multiplication for all n and is non-commutative for all $n > 1$.

Proof As we have already proved that $M_n\ k$ is a multiplicative monoid in Theorem 4.1.2, and all invertible matrices are invertible, the only thing left to prove is that it is closed:

$$\forall A, B \in GL_n\, k : (AB)(B^{-1}A^{-1}) = A(B^{-1}B)A^{-1} = AI_nA^{-1} = AA^{-1} = I_n$$
$$(B^{-1}A^{-1})(AB) = B^{-1}(A^{-1}A)B = B^{-1}I_nB = B^{-1}B = I_n \text{ (Associativity)}$$

To see that it is non-commutative, consider the following example in $GL_2\, k$ (for any k):

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} \text{ but } \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} \text{ (Proposition 2.6.1)}$$

For matrices of higher ranks, we can "pad" the above matrices with 0 (e.g. from $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \in GL_2\, k$ to $\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \in GL_3\, k$). The products will be the padded versions of the above too, the proof of which is trivial.

Before discussing more about invertibility, we need to see the linkage between matrices and systems of linear equations. In fact, every linear system of n equations with m unknowns (in a field k) can be written as a $\mathrm{Mat}_{n\times m}\, k \times k^m \to k^n$ matrix multiplication, and vice versa:

$$\begin{cases} a_{1,1}x_1 + a_{1,2}x_2 + \ldots + a_{1,m}x_m & = y_1 \\ a_{2,1}x_1 + a_{2,2}x_2 + \ldots + a_{2,m}x_m & = y_2 \\ & \vdots \\ a_{n,1}x_1 + a_{n,2}x_2 + \ldots + a_{n,m}x_m & = y_n \end{cases} \iff \begin{pmatrix} a_{1,1} & a_{1,2} & \ldots & a_{1,m} \\ a_{2,1} & a_{2,2} & \ldots & a_{2,m} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n,1} & a_{n,2} & \ldots & a_{n,m} \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \iff A\vec{x} = \vec{y}$$

Moreover, it can be expressed as a linear combination: $x_1\begin{pmatrix} a_{1,1} \\ a_{2,1} \\ \vdots \\ a_{n,1} \end{pmatrix} + x_2\begin{pmatrix} a_{1,2} \\ a_{2,2} \\ \vdots \\ a_{n,2} \end{pmatrix} + \ldots + x_n\begin{pmatrix} a_{1,m} \\ a_{2,m} \\ \vdots \\ a_{n,m} \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$.

A system of linear equations can have zero, one unique, or infinite solutions. To illustrate this, consider three equations: $x + 1 = x$, $x = 0$ and $x_1 + x_2 = 0$.

Proposition 4.2.2

1. The set of all $\vec{x}$ such that $A\vec{x} = \vec{0}$ (i.e. the solution set) is a subspace of k^n
2. The set of all $\vec{y}$ such that $A\vec{x} = \vec{y}$ is solvable is a subspace of k^m

Proof We have proved in the previous section that matrix multiplication is a linear transformation. Hence, set (1) and (2) are the kernel and the image respectively, which according to Proposition 3.4.2 and 3.4.3, are subspaces of k^n and k^m.

Proposition 4.2.3

If $m > n$ (i.e. more variables than equations), then $A\vec{x} = \vec{0}$ has non-trivial (non-zero) solutions. Furthermore, the dimension of the solution set (i.e. the nullity of ϕ_A) is at least $m - n$.

Proof The solution set of $A\vec{x} = \vec{0}$ is $\ker\phi_A$. By the rank-nullity theorem, as $m > n$, ϕ_A cannot be injective. Hence, by Theorem 2.4.1, $\ker\phi_A \supset \{0\}$.
By Proposition 3.8.2, $\operatorname{im}\phi_A \subseteq k^m \implies \dim\operatorname{im}\phi_A \leq m$. Therefore, by the rank-nullity theorem, $m = \dim\ker\phi_A + \dim\operatorname{im}\phi_A \implies \dim\ker\phi_A \geq m - n$.

Theorem 4.2.1: Invertible matrix theorem (1)

For $A \in M_n\, k$ and $\vec{x}, \vec{y} \in k^n$, the following are equivalent:

1. $A\vec{x} = \vec{y}$ is solvable for all $\vec{y}$.
2. $A\vec{x} = \vec{y}$ has a unique solution for all $\vec{y}$.
3. The only solution to $A\vec{x} = \vec{0}$ is the trivial solution (i.e. $\vec{0}$).
4. Columns of A (i.e. $\{A^i\}$, which is a set of vectors) are linearly independent.
5. Columns of A span k^n.
6. Columns of A is a basis for k^n.

Moreover, A being invertible is a sufficient condition to any of the above.

Proof (1) implies surjectivity and (3) implies injectivity of ϕ_A (Theorem 2.4.1). By the rank-nullity theorem, as $\phi : k^n \to k^n$, each of them are equivalent to (2), which is bijectivity.
By the basis theorem, as every column has n entries, (4), (5) and (6) are equivalent.
(2) and (4) are equivalent by the unique representation theorem. This shows the equivalence of (1) to (6).
Finally, to see that A is invertible implies (1) to (6), consider that $A\vec{x} = \vec{y} \iff A^{-1}A\vec{x} = A^{-1}\vec{y} \iff I_n\vec{x} = \vec{x} = A^{-1}\vec{y}$, which implies (1).

In a later section, we will prove that A being invertible is not only sufficient, but also necessary.

4.3 Gauss-Jordan Elimination

In the last section, we discussed how systems of linear equations can be written in terms of matrix multiplication. We can further simplify the notation by representing it in the augmented matrix form:

$$(A \mid \vec{y}) = \left(\begin{array}{cccc|c} a_{1,1} & a_{1,2} & \dots & a_{1,m} & y_1 \\ a_{2,1} & a_{2,2} & \dots & a_{2,m} & y_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{n,1} & a_{n,2} & \dots & a_{n,m} & y_n \end{array}\right)$$

Definition 4.3.1: Elementary row operations

There are three elementary row operations, which produce a new system of linear equation $(B \mid \vec{y})$:

1. $rR_i + R_j \to R_j$: $B_j = rA_i + A_j$ (Adding scalar multiple of one row to another)
2. $rR_i \to R_i$, where $r \neq 0$: $B_i = rA_i$ (Scalar multiplication of a row)
3. $R_i \leftrightarrow R_j$: $B_i = A_j$, $B_j = A_i$ (Swapping)

We can illustrate these operations using an example:

$$\left(\begin{array}{cc|c} 1 & 2 & 5 \\ 3 & 4 & 11 \end{array}\right) \xrightarrow[R_1 \leftrightarrow R_2]{} \left(\begin{array}{cc|c} 3 & 4 & 11 \\ 1 & 2 & 5 \end{array}\right) \xrightarrow[3R_2 \to R_2]{} \left(\begin{array}{cc|c} 3 & 4 & 11 \\ 3 & 6 & 15 \end{array}\right) \xrightarrow[-R_1 + R_2 \to R_2]{} \left(\begin{array}{cc|c} 3 & 4 & 11 \\ 0 & 2 & 4 \end{array}\right)$$

It is obvious that elementary row operations do not affect the solution set. There are also elementary column operations that work similarly on columns, which will not be elaborated on extensively.

Definition 4.3.2: Row-echelon form

Matrices (not necessarily square matrices) are said to be in row-echelon form if:

1. All (if any) zero rows lie at the bottom.
2. The number of leading zeros strictly increases with row number.
3. The first non-zero entry of each row is 1.

Also, if all entries above and below the leading 1s are 0, the matrix is in reduced row-echelon form.

For example, $\left(\begin{array}{cc|c} 3 & 4 & 11 \\ 0 & 2 & 4 \end{array}\right)$ from above is in row-echelon form but $\left(\begin{array}{cc|c} 0 & 0 & 0 \\ 3 & 6 & 15 \end{array}\right)$ and $\left(\begin{array}{cc|c} 0 & 4 & 8 \\ 0 & 6 & 12 \end{array}\right)$ are not.

One example of a matrix in reduced row-echelon form would be $\left(\begin{array}{ccc|c} 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 2 \\ 0 & 0 & 0 & 0 \end{array}\right)$.

We are especially interested in augmented matrices in (reduced) row-echlon form because they directly lead to the solution. The Gauss-Jordan elimination method that will be introduced below aims to translate any augmented matrix to its row-echelon form using elementary row operations.

Definition 4.3.3: Gauss-Jordan elimination

Given any system of linear equation $A\vec{x} = \vec{y}$, where A has n rows:

1. If $n = 1$, the system can be solved directly.
2. Otherwise, if $n > 1$, swap the rows below the first row such that the number of leading zeros increases with row number.
3. Multiply each row by a scalar such that the leading non-zero entry becomes 1.
4. Eliminate the first entry of each row below the first row by adding scalar multiples of the first row to rows below the first row.
5. Repeat (2) to (4), eliminating the second entry of each row below the second row, and so on until we reach row-echelon form.
6. Start from the last row. Eliminate all entries above the leading 1 by adding scalar multiples of the last row to rows above the first row.
7. Repeat (4) to rows above the last row until we reach reduced row-echelon form.

Here are some examples to help explain the algorithm:

1. $$\left(\begin{array}{ccc|c} 1 & 2 & 3 & 14 \\ 2 & 3 & 4 & 20 \\ 3 & 4 & 6 & 29 \end{array}\right) \xrightarrow[\substack{-2R_1+R_2\to R_2 \\ -3R_1+R_3\to R_3}]{} \left(\begin{array}{ccc|c} 1 & 2 & 3 & 14 \\ 0 & -1 & -2 & -8 \\ 0 & -2 & -3 & -13 \end{array}\right) \xrightarrow[\substack{-R_2\to R_2 \\ -\frac{1}{2}R_3\to R_3}]{} \left(\begin{array}{ccc|c} 1 & 2 & 3 & 14 \\ 0 & 1 & 2 & 8 \\ 0 & 1 & \frac{3}{2} & \frac{13}{2} \end{array}\right) \xrightarrow[-R_2+R_3\to R_3]{}$$
$$\left(\begin{array}{ccc|c} 1 & 2 & 3 & 14 \\ 0 & 1 & 2 & 8 \\ 0 & 0 & -\frac{1}{2} & -\frac{3}{2} \end{array}\right) \xrightarrow[-2R_3\to R_3]{} \left(\begin{array}{ccc|c} 1 & 2 & 3 & 14 \\ 0 & 1 & 2 & 8 \\ 0 & 0 & 1 & 3 \end{array}\right) \xrightarrow[\substack{-3R_3+R_1\to R_1 \\ -2R_2+R_2\to R_2}]{} \left(\begin{array}{ccc|c} 1 & 2 & 0 & 5 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 3 \end{array}\right) \xrightarrow[-2R_2+R_1\to R_1]{} \left(\begin{array}{ccc|c} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 3 \end{array}\right)$$
2. $$\left(\begin{array}{ccc|c} 1 & 2 & 3 & 14 \\ 2 & 3 & 4 & 20 \\ 3 & 4 & 5 & 26 \end{array}\right) \xrightarrow[\substack{-2R_1+R_2\to R_2 \\ -3R_1+R_3\to R_3}]{} \left(\begin{array}{ccc|c} 1 & 2 & 3 & 14 \\ 0 & -1 & -2 & -8 \\ 0 & -2 & -4 & -16 \end{array}\right) \xrightarrow[\substack{-R_2\to R_2 \\ -\frac{1}{2}R_3\to R_3}]{} \left(\begin{array}{ccc|c} 1 & 2 & 3 & 14 \\ 0 & 1 & 2 & 8 \\ 0 & 1 & 2 & 8 \end{array}\right) \xrightarrow[-R_2+R_3\to R_3]{}$$
$$\left(\begin{array}{ccc|c} 1 & 2 & 3 & 14 \\ 0 & 1 & 2 & 8 \\ 0 & 0 & 0 & 0 \end{array}\right) \xrightarrow[-2R_2+R_1\to R_1]{} \left(\begin{array}{ccc|c} 1 & 0 & -1 & -2 \\ 0 & 1 & 2 & 8 \\ 0 & 0 & 0 & 0 \end{array}\right) \iff \begin{cases} x_1 = t - 2 \\ x_2 = 8 - 2t \\ x_3 = t \end{cases}$$
3. $$\left(\begin{array}{ccc|c} 1 & 2 & 3 & 14 \\ 2 & 3 & 4 & 20 \\ 3 & 4 & 5 & 28 \end{array}\right) \xrightarrow[\substack{-2R_1+R_2\to R_2 \\ -3R_1+R_3\to R_3}]{} \left(\begin{array}{ccc|c} 1 & 2 & 3 & 14 \\ 0 & -1 & -2 & -8 \\ 0 & -2 & -4 & -14 \end{array}\right) \xrightarrow[\substack{-R_2\to R_2 \\ -\frac{1}{2}R_3\to R_3}]{} \left(\begin{array}{ccc|c} 1 & 2 & 3 & 14 \\ 0 & 1 & 2 & 8 \\ 0 & 1 & 2 & 7 \end{array}\right) \xrightarrow[-R_2+R_3\to R_3]{}$$
$$\left(\begin{array}{ccc|c} 1 & 2 & 3 & 14 \\ 0 & 1 & 2 & 8 \\ 0 & 0 & 0 & -1 \end{array}\right) \implies \text{No solution}$$

Example (1) is a system with a unique solution, while (2) is one with infinite solutions (also known as an indeterminate system). In constrast, example (3) is an inconsistent system, meaning it is unsolvable.

Theorem 4.2.1 shows that every invertible matrix corresponds to a linear system that has a unique solution for all $\vec{y}$. Suppose $A \in M_n\ k$ is invertible. Then we know that $A\vec{x} = \vec{e_1} = \langle 1, 0, ..., 0\rangle$ has a unique solution. Yet, also consider that $A\vec{x_1} = \vec{e_1} \implies A^{-1}A\vec{x_1} = A^{-1}\vec{e_1} \implies A^{-1}\vec{e_1} = I_n\vec{x_1} = \vec{x_1} \implies (A^{-1})^1 = \vec{x_1}$. Following the same logic, we can find the i-th column of A^{-1} by finding the unique solution $\vec{x_i}$ to the linear system $A\vec{x_i} = \vec{e_i}$. This means that using the Gauss-Jordan elimination method introduced above, we are now able to find the inverse of any matrix that we know is invertible. Specifically, this can be done by solving the linear system:

$$(A \mid \begin{pmatrix}\vec{e_1} & \vec{e_2} & ... & \vec{e_n}\end{pmatrix} = I_n) = \left(\begin{array}{cccc|cccc} a_{1,1} & a_{1,2} & ... & a_{1,m} & 1 & 0 & ... & 0 \\ a_{2,1} & a_{2,2} & ... & a_{2,m} & 0 & 1 & ... & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{n,1} & a_{n,2} & ... & a_{n,m} & 0 & 0 & ... & 1 \end{array}\right)$$

Notice we combined the $\vec{e_i}$s together so that we can find the inverse in one go.

We will demonstrate this by trying to find the general formula for the inverse of matrices in $M_2\ k$:

$$\left(\begin{array}{cc|cc} a & b & 1 & 0 \\ c & d & 0 & 1 \end{array}\right) \xrightarrow[-\frac{c}{a}R_1 + R_2 \to R_2]{} \left(\begin{array}{cc|cc} a & b & 1 & 0 \\ 0 & -\frac{bc}{a} + d & -\frac{c}{a} & 1 \end{array}\right) \xrightarrow[\substack{a^{-1}R_1 \to R_1 \\ -\frac{a}{ad-bc}R_2 \to R_2}]{} \left(\begin{array}{cc|cc} 1 & \frac{b}{a} & a^{-1} & 0 \\ 0 & 1 & \frac{c}{ad-bc} & -\frac{a}{ad-bc} \end{array}\right) \xrightarrow[-\frac{b}{a}R_2 + R_1 \to R_1]{}$$

$$\left(\begin{array}{cc|cc} 1 & 0 & -\frac{d}{ad-bc} & \frac{b}{ad-bc} \\ 0 & 1 & \frac{c}{ad-bc} & -\frac{a}{ad-bc} \end{array}\right) \implies \begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = (ad-bc)^{-1}\begin{pmatrix} -d & b \\ c & -a \end{pmatrix}$$

$(\frac{x}{y} = xy^{-1})$
This solution raises a few questions:

- What happens if $a = 0$? The first step would not hold in this case.
  Per the algorithm, we would have to swap the first and the second row. The solution would remain the same as it is an elementary row operation.
- What if $ad - bc = 0$? The derivation would not hold since the second step.
  $ad - bc = 0 \implies ad = bc \implies \frac{a}{b} = \frac{c}{d}$, which means $\langle a, c\rangle$ and $\langle b, d\rangle$ are not linearly independent. By the invertible matrix theorem, this suggests that A is not invertible, which violates our assumption.
- Does the existence of the result imply the existence of an inverse?
  The result exists if and only if $ad - bc \neq 0$, which is equivalent to the linear independence of $\langle a, c\rangle$ and $\langle b, d\rangle$. (The proof is trivial.) If the converse of the invertible matrix theorem holds (which in fact does), then the existence of an inverse is guaranteed. However, the proof will have to wait until a later section.

5 Matrix and Linear Transformation

5.1 Representation of $\hom(k^m, k^n)$

Definition 5.1.1: Set of Linear Transformations

Suppose V and W are vector spaces over k. Then $\hom(V, W)$ is the set of all linear transformations $\phi : V \to W$.

Theorem 5.1.1

Define the following operations on $\hom(V, W)$ for all $\phi, \theta, \psi \in \hom(V, W)$, $\vec{v} \in V$ and $r, s \in k$:

1. $+ : (\phi + \theta)(\vec{v}) = \phi(\vec{v}) + \theta(\vec{v})$ (Addition)
2. $\quad : (r\phi)(\vec{v}) = r\phi(\vec{v})$ (Scalar multiplication)

Then $\hom(V, W)$ is a vector space over k.

Proof We first have to show that both operations are closed. For $+$:
Homomorphism $(\phi + \theta)(\vec{v_1} + \vec{v_2}) = \phi(\vec{v_1} + \vec{v_2}) + \theta(\vec{v_1} + \vec{v_2}) = \phi(\vec{v_1}) + \phi(\vec{v_2}) + \theta(\vec{v_1}) + \theta(\vec{v_2}) = \phi(\vec{v_1} + \vec{v_2}) + \theta(\vec{v_1} + \vec{v_2}) = (\phi + \theta)(\vec{v_1} + \vec{v_2})$
Homogeneity $(\phi + \theta)(r\vec{v}) = \phi(r\vec{v}) + \theta(r\vec{v}) = r\phi(\vec{v}) + r\theta(\vec{v}) = r(\phi + \theta)(\vec{v})$
Similarly, for scalar multiplication:
Homomorphism $(r\phi)(\vec{v_1} + \vec{v_2}) = r\phi(\vec{v_1} + \vec{v_2}) = r\phi(\vec{v_1}) + r\phi(\vec{v_2}) = (r\phi)(\vec{v_1}) + (r\phi)(\vec{v_2})$
Homogeneity $(r\phi)(s\vec{v}) = r\phi(s\vec{v}) = rs\phi(\vec{v}) = s(r\phi(\vec{v})) = s(r\phi)(\vec{v})$
Next, we have to prove that $\hom(V, W)$ is a vector space:
Additive Abelian Group

- *Closure* Proved above.
- *Associativity* $((\phi + \theta) + \psi)(\vec{v}) = (\phi(\vec{v}) + \theta(\vec{v})) + \psi(\vec{v}) = \phi(\vec{v}) + (\theta(\vec{v}) + \psi(\vec{v})) = (\phi + (\theta + \psi))(\vec{v})$
- *Identity* The zero map O, where $O(\vec{v}) = \vec{0}$.
- *Commutativity* True as k is a field.
- *Inverse* The inverse of ϕ is $-\phi = (-1)\phi$. To show this, consider $(\phi - \phi)(\vec{v}) = \phi(\vec{v}) + (-1)\phi(\vec{v}) = \phi(\vec{v}) - \phi(\vec{v}) = \vec{0}$, hence $\phi - \phi = O$. Also, $-\phi + \phi = O$ due to commutativity.

Scalar Associativity $(r(s\phi))(\vec{v}) = r((s\phi)(\vec{v})) = r(s\phi(\vec{v})) = (rs)\phi(\vec{v}) = ((rs)\phi)(\vec{v})$
Vector Distributivity $((r + s)\phi)(\vec{v}) = (r + s)\phi(\vec{v}) = r\phi(\vec{v}) + s\phi(\vec{v}) = (r\phi)(\vec{v}) + (s\phi)(\vec{v})$
Scalar Distributivity $(r(\phi + \theta))(\vec{v}) = r(\phi(\vec{v}) + \theta(\vec{v})) = r\phi(\vec{v}) + r\theta(\vec{v}) = (r\phi)(\vec{v}) + (r\theta)(\vec{v})$
Identity Scalar $(1\phi)(\vec{v}) = 1\phi(\vec{v}) = \phi(\vec{v})$

Definition 5.1.2: k-algebra

A vector space V over a field k, equipped with another operation $\cdot : V \times V \to V$, is a k-algebra if:

1. V is a ring with unity under vector addition and $\cdot$.
2. $\forall r, s \in k, \vec{v}, \vec{w} \in V : (r\vec{v}) \cdot (s\vec{w}) = (rs)(\vec{v} \cdot \vec{w})$ (Compatibility with scalar)

Theorem 5.1.2

Given any vector spaces V, $\hom(V, V)$ is a k-algebra under composition.

Proof *Vector Ring*

- *Additive Abelian Group* Proved in Theorem 5.1.1.
- *Multiplicative Monoid* Composition of compatible linear transformations is closed (Proposition 3.4.1) and associative. The identity element is the identity map id_V.

- *Distributivity* $\forall \phi, \theta, \psi \in \hom(V,V) : \phi \circ (\theta + \psi) = \psi \circ \theta + \phi \circ \psi$ and $(\phi + \theta) \circ \psi = \phi \circ \psi + \theta \circ \psi$ *Compatibility with Scalar* $(r\phi) \circ (s\theta) = s(r(\phi \circ \theta)) = (rs)(\phi \circ \theta)$

Proposition 5.1.1

Suppose $\phi, \theta \in \hom(V,W)$, where V and W are finite dimensional vector spaces. If $\phi(\vec{v_i}) = \theta(\vec{v_i})$ for $\vec{v_i} \in S$, where S is a spanning set for V, then $\phi = \theta$.

Proof For any $\vec{v} \in V$, there exists $r_i \in k$ such that $\sum r_i \vec{v_i} = \vec{v}$.
$\phi(\vec{v}) = \phi(\sum r_i \vec{v_i}) = \sum r_i \phi(\vec{v_i}) = \sum r_i \theta(\vec{v_i}) = \theta(\sum r_i \vec{v_i}) = \theta(\vec{v})$

Proposition 5.1.2

Given $A \in \operatorname{Mat}_{n\times m} k$ and $B \in \operatorname{Mat}_{m\times k} k$, $\phi_{AB} = \phi_A \circ \phi_B$.

Proof $\phi_{AB}(\vec{e_i} \in k^k) = (AB)(\vec{e_i}) = (AB)^i$ while $\phi_A(\phi_B(\vec{e_i})) = A(B\vec{e_i}) = A(B^i) = (AB)^i$.
As $\{\vec{e_i}\}$ is a basis for k^k, by Proposition 5.1.1, $\phi_{AB} = \phi_A \circ \phi_B$.

Proposition 5.1.3: Matrix of linear transformation

Given any field k, suppose $V = k^m$, $W = k^n$, and $\phi \in \hom(V,W)$. Then there exists a $M(\phi) \in \operatorname{Mat}_{n\times m} k$ such that $\forall \vec{v} \in V : M(\phi)\vec{v} = \phi(\vec{v})$.

Proof For any $\vec{v} = \langle v_1, v_2, ..., v_n \rangle \in V$, suppose $M = (\phi(\vec{e_1}), \phi(\vec{e_2}), ..., \phi(\vec{e_m}))$ (joined horizontally).

$$\begin{aligned} M\vec{v} = M(v_1\vec{e_1} + v_2\vec{e_2} + ... + v_n\vec{e_m}) &= M(v_1\vec{e_1}) + M(v_2\vec{e_2}) + ... + M(v_m\vec{e_m}) \text{ (Proposition 4.1.3)} \\ &= v_1(M\vec{e_1}) + v_2(M\vec{e_2}) + ... + v_m(M\vec{e_m}) \text{ (Distributivity)} \\ &= v_1\phi(\vec{e_1}) + v_2\phi(\vec{e_2}) + ... + v_n\phi(\vec{e_n}) \; (\because M\vec{e_i} = M^i) \\ &= \phi(v_1\vec{e_1} + v_2\vec{e_2} + ... + v_n\vec{e_n}) = \phi(\vec{v}) \end{aligned}$$

Theorem 5.1.3

$\hom(k^m, k^n) \cong \operatorname{Mat}_{n\times m} k$

Proof We first have to prove ϕ is a linear transformations. For $A, B \in \operatorname{Mat}_{n\times m} k$, $\vec{v} \in k^m$ and $r \in k$:
Homomorphism By distributivity, $(A+B)\vec{v} = A\vec{v} + B\vec{v}$. Hence, $\phi_{A+B}(\vec{v}) = \phi_A(\vec{v}) + \phi_B(\vec{v})$.
Homogeneity By Proposition 4.1.4, $(rA)\vec{v} = r(A\vec{v})$. Hence, $\phi_{rA}(\vec{v}) = r\phi_A(\vec{v})$.
We then have to show ϕ is an isomorphism. Its surjectivity is already guaranteed by Proposition 5.1.3. Now suppose the kernel of ϕ contains M which is not the zero matrix. This means we can find an entry $M_{i,j} \neq 0$. Then by constructing a vector $\vec{v} = \langle v_i \rangle \in k^m$ such that $v_j = 1$, $M\vec{v} \neq \vec{0}$, which means $\phi_M \neq Z$, leading to a contradiction. Hence, $\ker \phi = \{(0)\}$, which by Theorem 2.4.1 implies ϕ is injective. As a result, we have shown that $\operatorname{Mat}_{n\times m} k \cong \hom(k^m, k^n)$.
With this and the fact that $\phi_{M(\phi)} = \phi$, M is the inverse function of ϕ. (Notice we have not proved the uniqueness of $M(\phi)$ until now.) Thus, M is also a vector space isomorphism.

There are two important corollaries to this theorem:

1. By the rank-nullity theorem and Theorem 4.1.1, $\dim \hom(k^m, k^n) = \dim \operatorname{Mat}_{n\times m} k = nm$.
2. Given any three matrices A, B and C of compatible shapes, by Proposition 5.1.2, we have $\phi_{ABC} = \phi_A \circ \phi_B \circ \phi_C$. Since matrices can be uniquely represented by corresponding linear transformations, $(\phi_A \circ \phi_B) \circ \phi_C = \phi_A \circ (\phi_B \circ \phi_C) \implies (AB)C = A(BC)$, which is the associativity of matrix

multiplication. We have hence produced a more concise proof than that in Proposition 4.1.4.

Proposition 5.1.4

Given two linear transformations $\phi : k^m \to k^k$ and $\theta : k^n \to k^m$, $M(\phi \circ \theta) = M(\phi)M(\theta)$.

Proof By Proposition 5.1.2 and Theorem 5.1.3, $M(\phi \circ \theta) = M(\phi_{M(\phi)} \circ \phi_{M(\theta)}) = M(\phi_{M(\phi)M(\theta)}) = M(\phi)M(\theta)$.

Definition 5.1.3: Algebra isomorphism

Given two k-algebras V and W, if ϕ is a vector space isomorphism between V and W, and $\forall \vec{v}, \vec{w} \in V : \phi(\vec{v} \cdot \vec{w}) = \phi(\vec{v}) \cdot \phi(\vec{w})$ (homomorphism), then ϕ is also an algebra isomorphism.

Theorem 5.1.4

$\phi : M_n\ k \to \hom(k^n, k^n)$ and $M : \hom(k^n, k^n) \to M_n\ k$ are algebra isomorphisms.

Proof First of all, we have to verify if $M_n\ k$ is a k-algebra, which is easy by Theorem 4.1.1, Theorem 4.1.2 and Proposition 4.1.2.
We know from Proposition 5.1.3 that ϕ and M are vector space isomorphisms, while Proposition 5.1.2 and Proposition 5.1.4 show they are also homomorphic.
Hence, we can now establish that ϕ and M are algebra isomorphisms.

Before concluding this section, let's look at an example to see how the above applies.

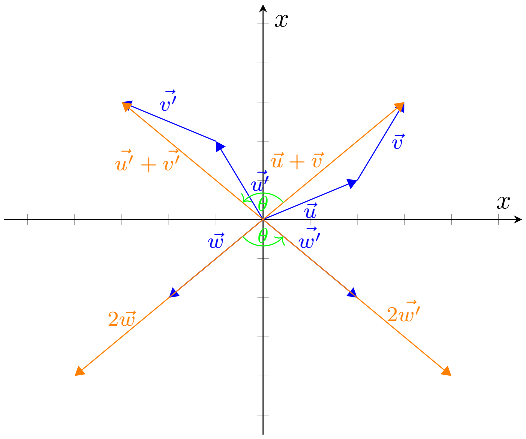

Figure 5.1.1: Visual proof of linearity of R_θ: $\vec{u}$ and $\vec{v}$ show homomorphism while $\vec{w}$ shows homogeneity.

Suppose $R_\theta : \mathbb{R}^2 \to \mathbb{R}^2$ is a mapping that rotates a vector $\vec{v} \in \mathbb{R}^2$ by θ anticlockwise about the origin. We can easily show that this is linear geometrically. Thus, by Proposition 5.1.3, we can find a matrix M_θ that represents this tranformation: By elementary trigonometry, we obtain $R_\theta(\langle 1, 0 \rangle) = \langle \cos\theta, \sin\theta \rangle$ and $R_\theta(\langle 0, 1 \rangle) = \langle -\sin\theta, \cos\theta \rangle$. Hence $M_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$, which is also known as a rotation matrix.

Now let's study what happens when a vector is rotated twice, i.e. the composition of rotation transformations. It is obvious that $R_\phi \circ R_\theta = R_{\theta+\phi}$. Meanwhile, Proposition 5.1.4 suggests that $M(R_\phi \circ R_\theta) = M_\phi M_\theta$. Since the matrix of a linear transformation is unique (by isomorphism), $M(R_\phi \circ R_\theta) = M(R_{\theta+\phi}) = M_{\theta+\phi} = M_\phi M_\theta$. Hence, $\begin{pmatrix} \cos(\theta+\phi) & -\sin(\theta+\phi) \\ \sin(\theta+\phi) & \cos(\theta+\phi) \end{pmatrix} = \begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix} \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} = \begin{pmatrix} \cos\theta\cos\phi - \sin\theta\sin\phi & -\sin\theta\cos\phi - \cos\theta\sin\phi \\ \sin\theta\cos\phi + \cos\theta\sin\phi & \cos\theta\cos\phi - \sin\theta\sin\phi \end{pmatrix}$, which are the compound angle formulae.

5.2 Representation of $\hom(V, W)$

Definition 5.2.1: Matrix of linear transformation with respect to bases

Suppose V and W are two vector spaces of dimensions n and m over the same field k, and B_V and B_W are bases for the two spaces respectively. Then there are two isomorphisms $\gamma_V : V \to k^n$ and $\gamma_W : W \to k^m$ which are the coordinate maps of V and W with respect to B_V and B_W. For any linear transformation $\phi : V \to W$, define the matrix of ϕ with respect to B_V and B_W as $M_{B_V,B_W}(\phi) = M(\gamma_W \circ \phi \circ \gamma_V^{-1}), M_{B_V,B_W}(\phi) \in \operatorname{Mat}_{m\times n} k$.

We know $\gamma_W \circ \phi \circ \gamma_V^{-1}$ is a linear transformation by Proposition 3.4.1 and Proposition 3.7.1.

This definition may seem incomprehensible at first, but the meaning of the matrix will become immediately apparent with the following diagram:

$$\begin{tikzcd}
V \arrow[r, "\phi"] \arrow[d, shift right, "\gamma_V"'] & W \arrow[d, shift right, "\gamma_W"'] \\
k^n \arrow[u, shift right, "B_V \cdot \vec{r}"'] \arrow[r, "M_{B_V,B_W}(\phi)"'] & k^m \arrow[u, shift right, "B_W \cdot \vec{s}"']
\end{tikzcd}$$

Figure 5.2.1: Visual representation of $M_{B_V,B_W}(\phi)$

Theorem 5.2.1

$M_{B_V,B_W} : \hom(V, W) \to \operatorname{Mat}_{m\times n} k$ is a vector space isomorphism and thus $\hom(V, W) \cong \operatorname{Mat}_{m\times n} k$.

Proof *Injectivity* Suppose there exists $\phi \neq Z$ such that $M_{B_V,B_W}(\phi)$ is the zero matrix.
Then there exists $\vec{v} \in V$ such that $\phi(\vec{v}) = \vec{w} \neq \vec{0}$.
$M_{B_V,B_W}(\phi)\gamma_V(\vec{v}) = \gamma_W(\phi(\gamma_V^{-1}(\gamma_V(\vec{v})))) = \gamma_W(\vec{w}) \neq \vec{0}$ (By the definition of a basis)
This is a contradiction. Hence, $\ker M_{B_V,B_W} = \{Z\}$ and by Theorem 2.4.1, it is injective.
Surjectivity For any $M \in \operatorname{Mat}_{m\times n} k$, let $\phi = \gamma_W^{-1} \circ \phi_M \circ \gamma_V$.
By Theorem 5.1.3, $M_{B_V,B_W}(\phi) = M(\phi_M) = M$.

Together with the rank-nullity theorem and Theorem 4.1.1, we obtain $\dim \hom(V, W) = \dim \operatorname{Mat}_{m\times n} k = nm$.

Proposition 5.2.1

Suppose U, V and W are vector spaces over the same field k, with bases B_U, B_V and B_W respectively. Let $\phi : U \to V$ and $\theta : V \to W$ be linear transformations. Then $M_{B_U,B_W}(\theta \circ \phi) = M_{B_V,B_W}(\theta)M_{B_U,B_V}(\phi)$.

Proof
$$\begin{aligned}
M_{B_U,B_W}(\theta \circ \phi) &= M(\gamma_W \circ \theta \circ \phi \circ \gamma_U^{-1}) \\
&= M(\gamma_W \circ \theta \circ \gamma_V^{-1} \circ \gamma_V \circ \phi \circ \gamma_U^{-1}) \\
&= M(\gamma_W \circ \theta \circ \gamma_V^{-1})M(\gamma_V \circ \phi \circ \gamma_U^{-1}) \text{ (Proposition 5.1.4)} \\
&= M_{B_V,B_W}(\theta)M_{B_U,B_V}(\phi)
\end{aligned}$$

We can visualise the above proof using a diagram:

$$\begin{tikzcd}
U \arrow[r, "\phi"] \arrow[d, shift right, "\gamma_U"'] & V \arrow[r, "\theta"] \arrow[d, shift right, "\gamma_V"'] & W \arrow[d, shift right, "\gamma_W"'] \\
k^n \arrow[u, shift right, "B_U \cdot \vec{r}"'] \arrow[r, "M_{B_U,B_V}(\phi)"'] & k^m \arrow[u, shift right, "B_V \cdot \vec{s}"'] \arrow[r, "M_{B_V,B_W}(\theta)"'] & k^k \arrow[u, shift right, "B_W \cdot \vec{t}"']
\end{tikzcd}$$

Figure 5.2.2: Visual representation of $M_{B_U,B_W}(\theta \circ \phi)$

Definition 5.2.2: Matrix of endomorphism with respect to basis

Suppose V is a vector space over k with a basis B, and $\phi : V \to V$ is a linear transformation. Then ϕ is an endomorphism of V and the matrix of ϕ with respect to B is $M_B(\phi) = M_{B,B}(\phi)$.

Theorem 5.2.2

Given any vector space V of dimension n over k with basis B, $M_B : \hom(V, V) \to M_n\ k$ is an algebra isomorphism.

Proof We have already shown M_B is a vector space isomorphism in Theorem 5.2.1, while homomorphism is proved in Proposition 5.2.1.

Similar to the last section, we will end this section with an example utilising the above results. Suppose F is the function space over $\mathbb{R}$ formed by the basis $B = \{\sin(x), \cos(x)\}$. Then $D : F \to F$, which outputs the derivative of the input function, is an endomorphism, with its linearity guaranteed by the sum rule and the constant multiple rule. To obtain the matrix of D with respect to B, we again use the method described in Proposition 5.1.3: As $\gamma_B \circ D \circ \gamma_B^{-1}(\langle 1, 0\rangle) = \gamma_B \circ D(\sin x) = \gamma_B(\cos x) = \langle 0, 1\rangle$ and similarly $\gamma_B \circ D \circ \gamma_B^{-1}(\langle 0, 1\rangle) = \langle -1, 0\rangle$, $M = M_B(D) = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$.

Now we study what happens when happens when M multiplies itself for n times, which can be expressed as $(M)^n$. From Proposition 5.2.1, we know that $(M)^n = M_B(\underbrace{D \circ D \circ \ldots \circ D}_{n \text{ times}})$. At the same time, $\dfrac{\mathrm{d}^n}{\mathrm{d}x^n} \sin x = \begin{cases} n \text{ is odd: } (-1)^{\frac{n-1}{2}} \cos x \\ n \text{ is even: } (-1)^{\frac{n}{2}} \sin x \end{cases}$ and $\dfrac{\mathrm{d}^n}{\mathrm{d}x^n} \cos x = \begin{cases} n \text{ is odd: } (-1)^{\frac{n+1}{2}} \sin x \\ n \text{ is even: } (-1)^{\frac{n}{2}} \cos x \end{cases}$. Combining the two statements, we get $(M)^n = \begin{cases} n \text{ is odd: } \begin{pmatrix} 0 & (-1)^{\frac{n+1}{2}} \\ (-1)^{\frac{n-1}{2}} & 0 \end{pmatrix} \\ n \text{ is even: } \begin{pmatrix} (-1)^{\frac{n}{2}} & 0 \\ 0 & (-1)^{\frac{n}{2}} \end{pmatrix} \end{cases}$.

5.3 Dual Spaces

Theorem 5.3.1: Extension by linearity

Let V and W be vector spaces over the same field k, and B be a finite basis for V. Then for any function (not necessarily linear) $f : B \to W$, we can extend it to a unique linear transformation $\bar{f} : V \to W$ such that for $\vec{u} \in B$, $\bar{f}(\vec{u}) = f(\vec{u})$.

Proof By the unique representation theorem, any $\vec{v} \in V$ can be uniquely represented by $\sum r_i \vec{u_i}$, where $r_i \in k$ and $\vec{u_i} \in B$. We define $\bar{f}(\vec{v}) = \sum r_i f(\vec{u_i})$. We can easily see it meets the requirements.

As for unqiueness, suppose there exists $\bar{f}' : V \to W$ with the same properties. Then $\bar{f}'(\vec{v}) = \bar{f}'(\sum r_i \vec{u_i}) = \sum r_i \bar{f}'(\vec{u_i}) = \sum r_i f(\vec{u_i}) = \bar{f}(\vec{v})$.

Definition 5.3.1: Dual space

For any vector space V over k, the dual space of V is defined as $V^* = \hom(V, k)$.

Definition 5.3.2: Transpose map

Suppose V and W are vector spaces over the same field k, and $\phi : V \to W$ is a linear transformation. Then the transpose map of ϕ is a function $\phi^* : W^* \to V^*$, where $\phi^*(f \in W^*) = f \circ \phi$.

$$\begin{tikzcd}
V \arrow[rr, "\phi \in \hom(V, W)"] \arrow[dr, "\phi^*(f) \in V^*"'] & & W \arrow[dl, "f \in W^*"] \\
& k &
\end{tikzcd}$$

Figure 5.3.1: Visual representation of the transpose map

In fact, ϕ^* is also a linear transformation, the proof of which is left as an exercise to the readers.

Proposition 5.3.1

$(\mathrm{id}_V)^* = \mathrm{id}_{V^*}$

Proof For any $f \in V^*$, $(\mathrm{id}_V)^*(f) = f \circ \mathrm{id}_V = f = \mathrm{id}_{V^*}(f)$

Proposition 5.3.2: Contravariance of transposition

Let $\phi : U \to V$ and $\theta : V \to W$ be linear transformations. Then $(\theta \circ \phi)^* = \phi^* \circ \theta^*$.

Proof For any $f \in W^*$, $(\theta \circ \phi)^*(f) = f \circ \theta \circ \phi = \theta^*(f) \circ \phi = (\phi^* \circ \theta^*)(f)$.

Proposition 5.3.3

Suppose V and W are (possibly infinitely dimensional) vector spaces over k, and $\phi : V \to W$ is a linear transformation, then we have:

1. If ϕ is injective, then ϕ^* is surjective.
2. If ϕ is surjective, then ϕ^* is injective.

Proof (1) Given a $g \in V^*$, we try to construct a $f \in W^*$ such that $\phi^*(f) = g$. It can be done by defining f such that $f(\phi(\vec{v})) = g(\vec{v})$. When ϕ is injective, such a map is well-defined, meaning that ϕ^* is surjective. (The linearity of f is trivial.)
(2) To determine injectivity, we only have to check the kernel of ϕ^*. Recall from Theorem 5.1.1 that the zero element of V^* is the zero map O. Let $f \in W^*$ be a map such that $\phi^*(f) = f \circ \phi = O$. As ϕ is surjective, f must map all vectors in W to 0. In other words, f is the zero map. Hence, by Theorem 2.4.1, ϕ^* is injective.

Theorem 5.3.2

Suppose V and W are finite dimensional vector spaces over k and $\phi : V \to W$ is a linear transformation. Then the rank of ϕ is equal to the rank of ϕ^*.

Proof We first decompose ϕ into two functions: $\phi' : V \to \operatorname{im}\phi$ and the inclusion map $i : \operatorname{im}\phi \to W$. ϕ' is identical to ϕ output-wise (i.e. $\phi'(\vec{v}) = \phi(\vec{v})$) and only differ from ϕ in its codomain, while $i(\vec{w} \in \operatorname{im}\phi) = \vec{w} \in W$. (It functions similarly to an identity map, except its domain and codomain are different.) Obviously, $\phi(\vec{v}) = i \circ \phi'(\vec{v})$.
By countravariance, $\phi^* = (i \circ \phi')^* = \phi'^* \circ i^*$. We also know from Proposition 5.3.3 as ϕ' is surjective and i is injective (by definition), ϕ'^* and i^* are injective and surjective respectively. Hence, by the rank-nullity theorem, we have $\dim\operatorname{im}\phi^* = \dim\operatorname{im}\phi'^* = \dim\hom(\operatorname{im}\phi, k) = \dim\operatorname{im}\phi$. (The last step follows Theorem 5.2.1.)

5.4 Dual Bases

Proposition 5.4.1: Dual basis

Given any finite dimensional vector space V over k with a basis $B = \langle v_1, v_2, ..., v_n\rangle$. Then the dual basis to B, $B^* = \{\vec{v_i}^* : V \to k, \vec{v_i}^*(\vec{v}) = \gamma_B(\vec{v}) \cdot \vec{e_i}\} \subseteq V^*$, is a basis for V^*.

Proof Theorem 5.2.1 shows that $\dim V^* = \dim V = n$. Thus, by the basis theorem, we would only have to show that B^* is linearly independent in order to establish it is a basis:
For any $\{r_1, r_2, ..., r_n\} \subseteq k$, suppose there exists $r_j \neq 0$. As $\gamma_B(\vec{v_j}) = \vec{e_j}$, $(\sum r_i\vec{v_i}^*)(\vec{v_i}) = r_j \neq 0$, implying that $\sum r_i\vec{v_i}^*$ is not the zero map. In other words, $\sum r_i\vec{v_i}^* = O \implies r_i = \{0\}$, confirming linear independence.

Proposition 5.4.2

Extend the operation * described above to the entirety of V such that $(\vec{v} \in V)^* = (\sum r_i\vec{v_i})^* = \sum r_i\vec{v_i}^*$. Then $^* : V \to V^*$ is a vector space isomorphism and hence $V \cong V^*$.

Proof * is obviously a linear transformation by definition.
For any $f \in V^*$, $f = \gamma_{B^*}(f) \cdot B^* = (\gamma_{B^*} \cdot B)^*$, showing the surjectivity of *. By the rank-nullity theorem, as $|B| = |B^*|$, this is sufficient to ensure the bijectivity of *.

Proposition 5.4.3

Suppose V is a vector space with a finite basis B over k. Then for any $f \in V^*$, $\gamma_{B^*}(f) = \langle f(\vec{v_1}), f(\vec{v_2}), ..., f(\vec{v_n})\rangle$.

Proof As B^* is a basis for V^*, we know that there exists a unique $\{r_i \in k\}$ such that $f = \sum r_i\vec{v_i}^*$. Then $f(\vec{v_i}) = \sum r_i\vec{v_i}^*(\vec{v_i}) = r_i$.

Theorem 5.4.1

Suppose V and W are vector spaces with bases $B_V = \langle\vec{v_1}, \vec{v_2}, ..., \vec{v_n}\rangle$ and $B_W = \langle\vec{w_1}, \vec{w_2}, ..., \vec{w_m}\rangle$. Then for any linear transformation $\phi : V \to W$, $M_{B_W^*, B_V^*}(\phi^*) = M_{B_V, B_W}(\phi)^T$.

Proof Suppose $M = M_{B_V,B_W}(\phi) = (m_{i,j})$. Then

$$\begin{aligned}\phi^*(\vec{w_j}^*) &= \vec{w_j}^* \circ \phi \\ &= \sum_i \vec{w_j}^*(\phi(\vec{v_i}))\vec{v_i}^* \text{ (Proposition 5.4.3)} \\ &= \sum_i \vec{w_j}^*(\gamma_W^{-1} \circ (\gamma_W \circ \phi \circ \gamma_V^{-1}) \circ \gamma_V(\vec{v_i}))\vec{v_i}^* \\ &= \sum_i \vec{w_j}^*(B_W \cdot M\vec{e_i})\vec{v_i}^* \text{ (Proposition 5.1.4)} \\ &= \sum_i \vec{w_j}^*(B_W \cdot M^i)\vec{v_i}^* \\ &= \sum_i m_{j,i}\vec{v_i}^*\end{aligned}$$

By Definition 5.2.1 and Proposition 5.1.3, this means $M_{B_W^*,B_V^*}(\phi^*) = (m_{j,i}) = M^T$.

$$\begin{tikzcd}
V^* \arrow[d, shift right] \arrow[dd, bend right=60, shift right, "\gamma_{V^*}"'] & W^* \arrow[l, "\phi^*"'] \arrow[d, shift right] \arrow[dd, bend left=60, shift left, "\gamma_{V^*}"] \\
V \arrow[u, shift right, "*"'] \arrow[r, "\phi"] \arrow[d, shift right, "\gamma_V"'] & W \arrow[u, shift right, "*"'] \arrow[d, shift right, "\gamma_W"'] \\
k^n \arrow[u, shift right] \arrow[uu, bend left=60, shift left] \arrow[r, "\begin{array}{c}M_{B_V,B_W}(\phi)\\ M_{B_W^*,B_V^*}(\phi^*)\end{array}"'] & k^m \arrow[u, shift right] \arrow[uu, bend right=60, shift right]
\end{tikzcd}$$

Figure 5.4.1: Visual representation of the relationship between vector spaces, their bases and dual spaces

This theorem immediately leads to an alternative proof to Proposition 4.1.7:

$$\begin{aligned}(AB)^T &= M(\phi_A \circ \phi_B)^T \text{ (Proposition 5.1.2 and Theorem 5.1.3)} \\ &= M((\phi_A \circ \phi_B)^*) \\ &= M(\phi_B^* \circ \phi_A^*) \text{ (Proposition 5.3.2)} \\ &= M(\phi_B^*)M(\phi_A^*) \text{ (Proposition 5.1.4)} \\ &= B^T A^T\end{aligned}$$

Note that the bases in the matrix notation (M) are not explicitly shown for simplicity.

Proposition 5.4.4

The rank of a matrix $M \in \mathrm{Mat}_{n\times m}\, k$ (i.e. $\mathrm{rk}\, M = \dim \mathrm{im}\, \phi_M$) is equal to the rank of M^T.

Proof Let $V = k^m$ and $W = k^n$, where k is a field, and e_V and e_W be their standard bases. Note that for any $\vec{v} \in V$, $\gamma_{e_V}(\vec{v}) = \vec{v}$, and the same thing goes for γ_{e_W}.
Consider that $M_{e_V,e_W}(\phi_M) = M(\gamma_{e_W} \circ \phi_M \circ \gamma_{e_V}^{-1}) = M(\phi_M) = M$. (Inverse function)
Then by Theorem 5.4.1, $M^T = M_{e_V,e_W}(\phi_M)^T = M_{e_W^*,e_V^*}(\phi^*) = M(\gamma_{e_W^*} \circ \phi^* \circ \gamma_{e_V^*}^{-1})$.
As $\gamma_{e_V^*}^{-1}$ is injective, $\mathrm{im}\, \phi^* \circ \gamma_{e_V^*}^{-1} = \mathrm{im}\, \phi^*$. Moreover, due to $\gamma_{e_W^*}$ being isomorphic, by the rank-nullity theorem, $\dim \mathrm{im}\, \gamma_{e_W^*} \circ \phi^* \circ \gamma_{e_V^*}^{-1} = \dim \mathrm{im}\, \phi^* \circ \gamma_{e_V^*}^{-1} = \dim \mathrm{im}\, \phi^*$.
Finally, by Theorem 5.3.2, $\mathrm{rk}\, M = \dim \mathrm{im}\, \phi = \dim \mathrm{im}\, \phi^* = \mathrm{rk}\, M^T$.

Theorem 5.4.2: Invertible matrix theorem (2)

The following are equivalent to (1) to (6) in Theorem 4.2.1:

7. The rows of A are linearly independent.
8. The rows of A span k^n.
9. The rows of A is a basis for k^n.
10. A is invertible.

Proof By Proposition 5.4.4, (5) implies that $\operatorname{rk} A^T = \operatorname{rk} A = n$. Hence, by the rank-nullity theorem, $\dim \ker \phi_{A^T} = \dim k^n - \operatorname{rk} A = 0$. As the only 0-dimensional vector space is the trivial vector space, according to Theorem 2.4.1, this means ϕ_{A^T} is injective. Notice that for any $\vec{w} \in \operatorname{im} \phi_{A^T}$, there exists an unique $\vec{v} = \langle v_1, v_2, ..., v_n \rangle \in k^n$ such that $\phi_{A^T} = A^T\vec{v} = \sum v_i A_i = \vec{w}$. By the unique representation theorem, the rows of A is a basis for $\operatorname{im} \phi_{A^T} \subseteq k^n$, implying (7). By the basis theorem, because $|A| = n$, (7) is equivalent to (8) and (9).
Conversely, considering that $(A^T)^T = A$, using the same logic, we can prove that (7) to (9) implies (5). With this, we have established that (1) to (9) are equivalent.
We now have to prove that (1) to (9) lead to (10). Using the same methodology to the one we introduced in Section 4.3, we can find a right inverse B to A by solving $A\vec{v_i} = \vec{e_i}$ (the soluability of which is guaranteed by (2)), then joining the column vectors together. Similarly, by finding the solution to $\vec{v_i}^T A = \sum v_i A_i = \vec{e_i}^T$ (which must exist by (8)), and joining the row vectors together, we are able to find the left inverse C. By associativity, $B = I_n B = (CA)B = C(AB) = CI_n = C$, hence $A^{-1} = B = C$.

5.5 Evaluation

Definition 5.5.1: Evaluation

Suppose V is a vector space over k. Then given $\vec{v} \in V$ and $f \in V^*$, the evaluation at $\vec{v}$, $e_{\vec{v}}$, is defined such that $e_{\vec{v}}(f) = f(\vec{v}), e_{\vec{v}} \in V^{**}$.

Theorem 5.5.1

The map $e : V \to V^{**}$ is a vector space isomorphism for any finite dimensional vector space V.

Proof It is obvious that e is a linear transformation.
To show the map is injective, suppose there exists $\vec{v} \neq \vec{0}$ in V such that $e_{\vec{v}} = O$. Then for any arbitrary basis $B = \{\vec{u_i}\}$ of V, $\gamma_B(\vec{v}) = \langle r_1, r_2, ..., r_n \rangle \neq \vec{0}$. If $r_j \neq 0$, then $e_{\vec{v}}(\vec{u_j}^*) = r_j \neq 0$, meaning that $e_{\vec{v}}$ is not the zero map, which is a contradiction.
By Proposition 5.4.2, we have $V \cong V^* \cong V^{**}$. By the rank-nullity theorem, this means that $\dim V = \dim V^{**}$. Hence, injectivity of e already implies its surjectivity.

5.6 Annihilators

Definition 5.6.1: Annihilators

Let V be a vector space. The annihilator of a subset $S \subseteq V$, denoted as S^0, is the set $\{f \mid \forall \vec{v} \in S : f(\vec{v}) = 0\} \subseteq V^*$.

Proposition 5.6.1

Annihilators are vector subspaces.

Proof By the subspace test, for all $f, g \in S^0, r \in k$:

$$(f+g)(\vec{v}) = f(\vec{v}) + g(\vec{v}) = 0 + 0 = 0$$

$$(rf)(\vec{v}) = rf(\vec{v}) = 0r = 0$$

Proposition 5.6.2

Suppose V is a vector space with a finite basis B and a subspace W. Then $W \oplus W^0 = V$ under the dual basis B^*, that is, $W + \gamma_B^{-1} \circ \gamma_{B^*}(W^0) = V$.

Proof We first find a basis B_W for W. By the basis extension theorem, we can extend B_W to B' which is a basis for V. Then $\operatorname{span}(B' \setminus B_W)^*$ is obviously the annihilator of W and $W \oplus W^0 = V$ under B'.
Next, to see that this argument works for any basis B, note that the map $\gamma_B^{-1} \circ \gamma_{B^*}$ is natural, i.e. does not depend on the choice of B.

Definition 5.6.2: Hyperplane

Let V be a non-zero n-dimensional vector space. Then any subspace of V with dimension $n-1$ is a hyperplane of V.

Hyperplane is a generalisation of the notion of a line in $\mathbb{R}^2$ and a plane in $\mathbb{R}^3$.

Proposition 5.6.3

Any m-dimension subspace W of a vector space V is the intersection of $n-m$ hyperplanes.

Proof By Proposition 5.6.2 and the dimension formula, the annihilator of (the span of) a single vector is a hyperplane. Also by , we know that W^0 is $(n-m)$-dimensional. Thus, if W^0 has a basis $\{\vec{u_1}, \vec{u_2}, ..., \vec{u}_{n-m}\}$, then $W = (\vec{u_1})^0 \cap ... \cap (\vec{u}_{n-m})^0$.

Proposition 5.6.4

For any subspace W of a finitely dimensional vector space V, $W = (W^0)^0 = \ker W^0$ under the natural evaluation, where $\ker W^0$ is the intersection of $\{\ker f \mid f \in W^0\}$.

Proof It is obvious that $W \subseteq \ker W^0$. By Proposition 5.6.2 and the dimension formula:

$$\dim W + \dim W^0 = \dim V$$

$$\dim W^0 + \dim (W^0)^0 = \dim V$$

Hence, $\dim W = \dim (W^0)^0$. By the basis theorem, the first equality is true.
The second equality holds because any $F \in V^{**}$ can be translated to and from $e_{\vec{v}}$ for some $\vec{v} \in V$ by Theorem 5.5.1. Then for any $f \in W^0$:

$$F(f) = e_{\vec{v}}(f) = f(\vec{v}) = 0$$

The proposition can be easily generalised to the following version: For any subset S of a finite-dimensional vector space V, $S \subseteq (S^0)^0$. By considering the annihilator of the span of S, this should be apparent.
In fact, this proposition holds for all vector spaces, including infinitely dimensional ones. However, such a

proof would require the axiom of choice, which is out of the scope of this set of notes.

Finally, we will give another interpretation to the transpose map using annihilators.

Theorem 5.6.1

Let V and W be two finitely dimensional vector spaces. Then for any $\phi \in \hom(V, W)$:

1. $\ker \phi^* = (\operatorname{im} \phi)^0$
2. $\dim \operatorname{im} \phi^* = \dim \operatorname{im} \phi$
3. $\operatorname{im} \phi^* = (\ker \phi)^0$

Proof (1) By definition.
(2) By Proposition 5.6.2 and the dimension formula:

$$\dim (\operatorname{im} \phi)^0 = \dim W - \dim \operatorname{im} \phi$$

Then, by the rank-nullity theorem and (1):

$$\dim \operatorname{im} \phi^* = \dim W^* - \dim \ker \phi^* = \dim W - (\dim W - \dim \operatorname{im} \phi) = \dim \operatorname{im} \phi$$

(3) It is to see that $(\ker \phi)^0 \subseteq \operatorname{im} \phi^*$. (2), Proposition 5.6.2 and the dimension formula show that $\dim \operatorname{im} \phi^* = \dim (\ker \phi)^0$. By the basis theorem, $\operatorname{im} \phi^* = (\ker \phi)^0$.

5.7 Change of Basis

Definition 5.7.1: Transition matrix

Let V be a vector space over k with bases B and B'. Then the transition matrix from B to B' is $P_{B,B'} = M(\gamma_{B'} \circ \gamma_B^{-1})$.

This definition is more intuitive with visualisation:

$$\begin{tikzcd}
 & V \arrow[dl, shift right, "\gamma_B"'] \arrow[dr, shift left, "\gamma_{B'}"] & \\
k^n \arrow[ur, shift right, "B \cdot \vec{r}"'] \arrow[rr, shift left, "P_{B,B'}"] & & k^n \arrow[ul, shift left, "B \cdot \vec{r'}"] \arrow[ll, shift left, "P_{B',B}"]
\end{tikzcd}$$

Figure 5.7.1: Visual representation of a transition matrix

In essence, the transition matrix transforms the expression of a vector based on B to another based on B'.

Proposition 5.7.1

$P_{B,B'} = (P_{B',B})^{-1}$

Proof $P_{B,B'}P_{B',B} = M(\gamma_{B'} \circ \gamma_B^{-1})M(\gamma_B \circ \gamma_{B'}^{-1}) = M(\gamma_{B'} \circ \gamma_B^{-1} \circ \gamma_B \circ \gamma_{B'}^{-1})$ (Proposition 5.1.4)

$$= M(\mathrm{id}_{k^n}) = I_n$$

Similarly, $P_{B',B}P_{B,B'} = I_n$. Hence, $P_{B,B'} = (P_{B',B})^{-1}$.

Theorem 5.7.1: Change of basis formula

Suppose V is a vector space over k with finite bases B and B', and $\phi : V \to V$ is an endomorphism. Then $M_{B'}(\phi) = P_{B,B'}M_B(\phi)P_{B',B}$.

Proof $P_{B,B'}M_B(\phi)P_{B',B} = M((\gamma_{B'} \circ \gamma_B^{-1}) \circ (\gamma_B \circ \phi \circ \gamma_B^{-1}) \circ (\gamma_B \circ \gamma_{B'}^{-1}))$ (Proposition 5.1.4)
$= M(\gamma_{B'} \circ \phi \circ \gamma_{B'}^{-1}) = M_{B'}(\phi)$

Definition 5.7.2: Similarity of square matrices

Given $A, B \in M_n\ k$, A is similar to B ($A \sim B$) if there exists $P \in GL_n\ k$ such that $B = P^{-1}AP$.

Proposition 5.7.2

$\sim$ defined above is an equivalence relation on $M_n\ k$.

Proof *Reflexivity* $A = I_n^{-1}AI_n \implies A \sim A$
Symmetry $A \sim B \implies B = P^{-1}AP \implies PBP^{-1} = PP^{-1}APP^{-1}$
$\implies (P^{-1})^{-1}BP^{-1} = A$ (Proposition 2.2.5)
$\implies B \sim A$
Transitivity $A \sim B, B \sim C \implies B = P_1^{-1}AP_1, C = P_2^{-1}BP_2 \implies C = P_1^{-1}P_2^{-1}AP_2P_1$
$\implies A \sim C\ (\because (P_2P_1)^{-1} = P_1^{-1}P_2^{-1})$

Proposition 5.7.3

Suppose V is an n-dimensional vector space over k and $A, A' \in M_n\ k$. Then $A \sim A' \iff$ there exists bases B and B' such that $A' = P_{B,B'}AP_{B',B}$.

Proof **($\rightarrow$)** Suppose $A' = P^{-1}AP$. Our task is to find B and B' such that $P = P_{B',B}$.
Let $B = \{(P^{-1})^j\}$ (which must be a basis by the invertible matrix theorem) and B' be the standard basis.
Then $M_BP = I_n \implies \gamma_B^{-1}(P^j) = \vec{e_j} \implies P^j = \gamma_B(\vec{e_j}) = \gamma_B \circ \gamma_{B'}^{-1}(\vec{e_j}) \implies P = M(\gamma_B \circ \gamma_{B'}^{-1}) = P_{B',B}$ (Proposition 5.1.3), where M_B is the matrix constructed by joining basis vectors in B together as column vectors.
The converse is obvious with the change of basis formula and Proposition 5.6.1.

6 Inner Product Spaces

6.1 Real Inner Product Spaces

Definition 6.1.1: Real inner product space

A real inner product space V is a vector space over $\mathbb{R}$ with an additional map $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{R}$, which is known as the inner product and has the following properties for all $\vec{v}, \vec{w} \in V$ and $r, s \in \mathbb{R}$:

1. $\langle \vec{v}, \vec{v} \rangle > 0$ for $\vec{v} \neq \vec{0}$ and $\langle \vec{0}, \vec{0} \rangle = 0$ (Positive definiteness)
2. $\langle \vec{v}, \vec{w} \rangle = \langle \vec{w}, \vec{v} \rangle$ (Symmetry)
3. $\langle r\vec{v_1} + s\vec{v_2}, \vec{w} \rangle = r\langle \vec{v_1}, \vec{w} \rangle + s\langle \vec{v_2}, \vec{w} \rangle$ (Linearity in first argument)

Note that by (2) and (3), the inner product must also be linear in the second argument. Hence, the real inner product is bilinear.

The inner product is a generalisation of the dot product in $\mathbb{R}^n$. As a reminder, for any $\vec{v} = \langle v_1, v_2, ..., v_n \rangle$ and $\vec{w} = \langle w_1, w_2, ..., w_n \rangle$ in $\mathbb{R}^n$, $\vec{v} \cdot \vec{w} = v_1 w_1 + v_2 w_2 + ... + v_n w_n$. Apparently, $\mathbb{R}^n$ with the dot product is a real inner product space. (In fact, we have already proved its symmetry and bilinearity.)

There is also an alternative geometric definition of the dot product. Define the magnitude of a vector in $\mathbb{R}^n$ to be the Euclidean distance between its origin and destination. Denote the magnitude of a vector $\vec{v}$ to be $||\vec{v}||$. Then the dot product can also be defined as the product of the magnitude of $\vec{v}$ and the magnitude of the projection of $\vec{w}$ on the span of $\vec{v}$ ($\vec{v}$ and $\vec{w}$ can be swapped as the dot product is symmetric). Or more concisely, $\vec{v} \cdot \vec{w} = ||\vec{v}||\ ||\vec{w}|| \cos\theta$, where $\theta \in [0, \pi)$ is the angle between the two vectors.

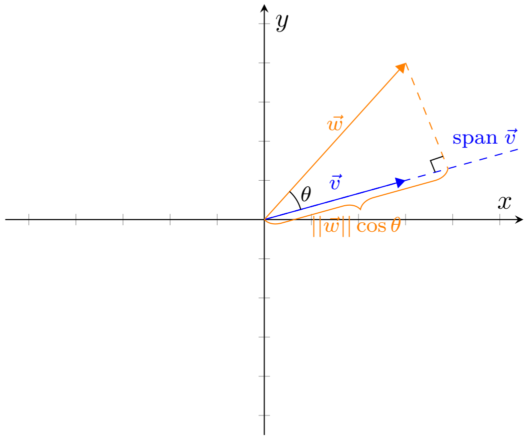

Figure 6.1.1: Geometric definition of the dot product in $\mathbb{R}^2$

To see that the two definitions are equivalent, we have to use the cosine law:

$$\begin{aligned} ||\vec{v} - \vec{w}||^2 = ||\vec{v}||^2 + ||\vec{w}||^2 - 2||\vec{v}||\ ||\vec{w}|| \cos\theta &\implies \sum (v_i - w_i)^2 = (\sum v_i^2) + (\sum w_i^2) - 2||\vec{v}||\ ||\vec{w}|| \cos\theta \\ &\implies -2\sum v_i w_i = -2||\vec{v}||\ ||\vec{w}|| \cos\theta \\ &\implies \sum v_i w_i = ||\vec{v}||\ ||\vec{w}|| \cos\theta \end{aligned}$$

This alternative definition is important because we can then extend geometric concepts, such as magnitude and angle, to abstract inner product spaces.

Proposition 6.1.1

$\langle \vec{0}, \vec{v} \rangle = \langle \vec{v}, \vec{0} \rangle = 0$

Proof By Proposition 3.2.2 and bilinearity, $\langle \vec{0}, \vec{v} \rangle = \langle 0\vec{v}, \vec{v} \rangle = 0\langle \vec{v}, \vec{v} \rangle = 0$.
The second equality follows a similar proof.

Definition 6.1.2: Magnitude

For any real inner product space V, the magnitude of a vector $\vec{v} \in V$ is denoted as $||\vec{v}||$ and $||\vec{v}|| = \sqrt{\langle \vec{v}, \vec{v} \rangle}$.

This definition comes directly from the Euclidean distance formula in $\mathbb{R}^n$.

Proposition 6.1.2

$||r\vec{v}|| = |r|\ ||\vec{v}||$

Proof By bilinearity, $\sqrt{\langle r\vec{v}, r\vec{v} \rangle} = \sqrt{r^2\langle \vec{v}, \vec{v} \rangle} = |r|\ ||\vec{v}||$.

Definition 6.1.3: Unit vector

A vector $\vec{v}$ in a real inner product space is a unit vector if $||\vec{v}|| = 1$.

Theorem 6.1.1: Cauchy-Schwarz inequality

For any $\vec{v}$ and $\vec{w}$ in a real inner product space V, $|\langle \vec{v}, \vec{w} \rangle| \leq ||\vec{v}||\ ||\vec{w}||$.

Proof For any $\vec{x} \in \mathbb{R}$, $\langle \vec{v} + x\vec{w}, \vec{v} + x\vec{w} \rangle \geq 0$ (Positive definiteness)

$$\begin{aligned}
&\implies \langle \vec{v}, \vec{v} + x\vec{w} \rangle + x\langle \vec{w}, \vec{v} + x\vec{w} \rangle \geq 0 \text{ (Bilinearity)} \\
&\implies \langle \vec{v}, \vec{v} \rangle + x\langle \vec{v}, \vec{w} \rangle + x\langle \vec{w}, \vec{v} \rangle + x^2\langle \vec{w}, \vec{w} \rangle \geq 0 \text{ (Bilinearity)} \\
&\implies ||\vec{w}||^2x^2 + 2\langle \vec{v}, \vec{w} \rangle x + ||\vec{v}||^2 \geq 0 \text{ (Symmetry)}
\end{aligned}$$

This implies the discriminant of the equation $||\vec{w}||^2x^2 + 2\langle \vec{v}, \vec{w} \rangle x + ||\vec{v}||^2 = 0$ is less than or equal to 0. Hence, $4\langle \vec{v}, \vec{w} \rangle^2 - 4||\vec{v}||^2||\vec{w}||^2 \leq 0 \implies \langle \vec{v}, \vec{w} \rangle^2 \leq ||\vec{v}||^2||\vec{w}||^2 \implies |\langle \vec{v}, \vec{w} \rangle| \leq ||\vec{v}||\ ||\vec{w}||$.

Theorem 6.1.2: Triangle inequality

For any $\vec{v}$ and $\vec{w}$ in a real inner product space V, $||\vec{v} + \vec{w}|| \leq ||\vec{v}|| + ||\vec{w}||$.

Proof By the Cauchy-Schwarz inequality,

$$\begin{aligned}
|\langle \vec{v}, \vec{w} \rangle| \leq ||\vec{v}||\ ||\vec{w}|| &\implies 2\langle \vec{v}, \vec{w} \rangle \leq 2||\vec{v}||\ ||\vec{w}|| \\
&\implies \langle \vec{v}, \vec{v} \rangle + \langle \vec{v}, \vec{w} \rangle + \langle \vec{w}, \vec{v} \rangle + \langle \vec{w}, \vec{w} \rangle \leq ||\vec{v}||^2 + 2||\vec{v}||\ ||\vec{w}|| + ||\vec{w}||^2 \text{ (Symmetry)} \\
&\implies \langle \vec{v} + \vec{w}, \vec{v} + \vec{w} \rangle \leq (||\vec{v}|| + ||\vec{w}||)^2 \\
&\implies ||\vec{v} + \vec{w}|| \leq ||\vec{v}|| + ||\vec{w}||
\end{aligned}$$

Definition 6.1.4: Directional cosine

The angle θ between two vectors $\vec{v}$ and $\vec{w}$ in a real inner product space is defined such that $\cos\theta = \frac{\langle \vec{v}, \vec{w} \rangle}{||\vec{v}||\ ||\vec{w}||}$ and $\theta \in [0, \pi)$.

This formula is derived from the aforementioned geometric definition of the dot product. Its validity (concerning whether $0 \leq \cos\theta \leq 1$) is guaranteed by the Cauchy-Schwarz inequality.

Definition 6.1.5: Orthogonality

Two vectors $\vec{v}$ and $\vec{w}$ in a real inner product space are orthogonal (written as $\vec{v} \perp \vec{w}$) if $\theta = \frac{\pi}{2}$, or equivalently, $\langle \vec{v}, \vec{w} \rangle = 0$.

Orthogonality is an extension of the geometric notion of perpendicularity.

Definition 6.1.6: Orthogonal family

A finite subset $\{\vec{v_i} \neq \vec{0}\}$ of a real inner product space is an orthogonal family if $\vec{v_i} \perp \vec{v_j}$ for $i \neq j$.

An orthogonal family is also an orthonormal family if it only consists of unit vectors.

Proposition 6.1.3

An orthogonal family is linearly independent.

Proof Suppose $A = \{\vec{v_1}, \vec{v_2}, ..., \vec{v_n}\}$ is a linearly dependent orthogonal family. Then there exists at least two non-zero weights, including r_j, such that $r_1\vec{v_1} + r_2\vec{v_2} + ... + r_n\vec{v_n} = 0$. (As $\vec{v_i} \neq \vec{0}$ for all i.)

By positive definiteness, $\langle \vec{v_j}, \vec{v_j} \rangle > 0 \implies \langle \vec{v_j}, -\sum_{i \neq j} r_i \vec{v_i} \rangle > 0$

$$\begin{aligned} &\implies -\sum_{i \neq j} r_i \langle \vec{v_j}, \vec{v_i} \rangle > 0 \text{ (Bilinearity)} \\ &\implies -r_k \langle \vec{v_j}, \vec{v_k} \rangle > 0 \text{ for some } k \neq j \text{ and } r_k \neq 0 \\ &\implies \langle \vec{v_j}, \vec{v_k} \rangle \neq 0 \text{, which is a contradition.} \end{aligned}$$

Theorem 6.1.3: Pythagorean theorem

Suppose $\vec{v_1}, \vec{v_2}, ..., \vec{v_n}$ constitute an orthogonal family in a real inner product space. Then $||\sum \vec{v_i}||^2 = \sum ||\vec{v_i}||^2$.

Proof Consider that for any i, $\langle \vec{v_i}, \vec{v_1} + \vec{v_2} + ... + \vec{v_n} \rangle = \langle \vec{v_i}, \vec{v_1} \rangle + ... + \langle \vec{v_i}, \vec{v_n} \rangle$ (Bilinearity)

$$= \langle \vec{v_i}, \vec{v_i} \rangle = ||\vec{v_i}||^2 \text{ (Orthogonality)}$$

Hence, by bilinearity, $||\sum \vec{v_i}||^2 = \langle \sum \vec{v_i}, \sum \vec{v_i} \rangle = \sum \langle v_i, \sum \vec{v_i} \rangle = \sum ||\vec{v_i}||^2$.

6.2 Orthogonal Bases and Orthogonal Projection

Definition 6.2.1: Orthonormal basis

A (finite) basis of a real inner product space is orthonormal if it is also an orthonormal family.

Now suppose $\{\vec{u_1}, \vec{u_2}, ..., \vec{u_n}\}$ is an orthonormal basis for a real inner product space V. We then have the following two propositions:

Proposition 6.2.1

For all $\vec{v} \in V$, $\vec{v} = \sum \langle \vec{v}, \vec{u_i} \rangle \vec{u_i}$.

Proof Suppose $\vec{v} = \sum r_i\vec{u_i}$, where $r_i \in \mathbb{R}$. Then

$$\begin{aligned}\sum\langle\vec{v},\vec{u_i}\rangle\vec{u_i} &= \sum_i\langle\sum_j r_j\vec{u_j},\vec{u_i}\rangle\vec{u_i}\\ &= \sum_{i,j} r_j\langle\vec{u_j},\vec{u_i}\rangle\vec{u_i} \text{ (Bilinearity)}\\ &= \sum_i r_i\langle\vec{u_i},\vec{u_i}\rangle\vec{u_i} \text{ (Orthogonality)}\\ &= \sum_i r_i\vec{u_i} = \vec{v} \text{ (Unit vectors)}\end{aligned}$$

Proposition 6.2.2

For all $\vec{v} \in V$, $||\vec{v}||^2 = \sum\langle\vec{v},\vec{u_i}\rangle^2$.

Proof By Proposition 6.2.1,

$$\begin{aligned}||\vec{v}||^2 &= \sum||\langle\vec{v},\vec{u_i}\rangle\vec{u_i}||^2 \text{ (Pythagorean Theorem)}\\ &= \sum\langle\vec{v},\vec{u_i}\rangle^2||\vec{u_i}||^2 \text{ (Proposition 6.1.2)}\\ &= \sum\langle\vec{v},\vec{u_i}\rangle^2 \text{ (Unit vectors)}\end{aligned}$$

Definition 6.2.2: Orthogonal projection

Suppose V is a real inner product space and W is a vector subspace of V with an orthonormal basis $\{\vec{u_1},\vec{u_2},...,\vec{u_m}\}$. Then the orthogonal projection of a vector $\vec{v} \in V$ onto W is $\text{pr}_W(\vec{v}) = \sum\langle\vec{v},\vec{u_i}\rangle\vec{u_i}$.

We can also project a vector on another vector similar to that in the geometric definition of the dot product: As the span of a vector is a subspace of the larger vector space (by Proposition 3.3.1), we can define this as the projection of a vector onto the span of the other vector.

Orthogonal projection generalises the notion of "projection" with which we are familiar in Euclidean spaces. To make sense of why orthogonal projection is defined in such a way, consider a finite real inner product space V which has a subspace W with an orthonormal basis $B_W = \{\vec{u_1},\vec{u_2},...,\vec{u_m}\}$. Extend B_W to $B_V = \{\vec{u_1},...,\vec{u_m},...,\vec{u_n}\}$ such that it is an orthonormal basis for V. (The proof of its existence will be presented later.)

For every vector in V, we can express it as a linear combination of B_N. If we extract the weights with respect to basis vectors in B_N in the expression, we get a projection vector lying in W. We can represent the weights using inner product as in Proposition 6.2.1. This is very much akin to how we obtain projections in an Euclidean space:

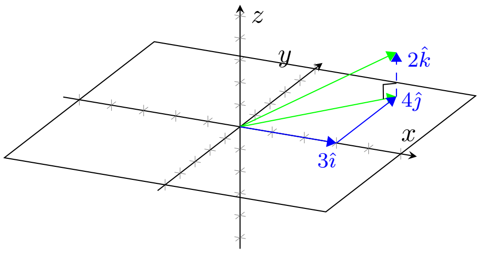

Figure 6.2.1: Example of a 3-dimensional to 2-dimensional projection in $\mathbb{R}^3$

Proposition 6.2.3

Orthogonal projection is a linear transformation.

Proof Given by the bilinearity of inner products.

Proposition 6.2.4

Given a finite dimensional real inner product space V with a subspace W, for any $\vec{v} \in V$ and $\vec{w} \in W$, $\vec{v} - \mathrm{pr}_W(\vec{v}) \perp \vec{w}$.

Proof Suppose W has an orthonormal basis $\{\vec{u_1}, \vec{u_2}, ..., \vec{u_m}\}$ and $\vec{w} = r_1\vec{u_1} + r_2\vec{u_2} + ... + r_m\vec{u_m}$.
Then
$$\begin{aligned}\langle \vec{v} - \mathrm{pr}_W(\vec{v}), \vec{w}\rangle = \langle \vec{v} - \sum \langle \vec{v}, \vec{u_i}\rangle \vec{u_i}, \vec{w}\rangle &= \langle \vec{v}, \vec{w}\rangle - \langle \sum \langle \vec{v}, \vec{u_i}\rangle \vec{u_i}, \sum r_i \vec{u_i}\rangle \\ &= \langle \vec{v}, \vec{w}\rangle - \sum_{i,j} \langle \langle \vec{v}, \vec{u_i}\rangle \vec{u_i}, r_j \vec{u_j}\rangle \text{ (Bilinearity)} \\ &= \langle \vec{v}, \vec{w}\rangle - \sum_{i,j} r_j \langle \vec{v}, \vec{u_i}\rangle \langle \vec{u_i}, \vec{u_j}\rangle \text{ (Bilinearity)} \\ &= \langle \vec{v}, \vec{w}\rangle - \sum_{i} r_i \langle \vec{v}, \vec{u_i}\rangle \text{ (Orthonormality)} \\ &= \langle \vec{v}, \vec{w}\rangle - \langle \vec{v}, \vec{w}\rangle = 0 \text{ (Bilinearity)}\end{aligned}$$

The above proof actually implicitly requires the existence of an orthonormal basis for W, which will be proven in the theorem below. Note that although the following proof relies on the above proposition, it does not depend on this fact. In other words, there is no circular reasoning.

Theorem 6.2.1: Gram-Schmidt orthonormalisation process

Every finite dimensional real inner product space has a (finite) orthonormal basis.

Proof Suppose the real inner product space has a basis $\{\vec{v_1}, \vec{v_2}, ..., \vec{v_n}\}$. We can obtain an orthonormal basis $\{\vec{u_1} = \frac{\vec{v_1}}{||\vec{v_1}||}, \vec{u_i} = \frac{\vec{v_i} - \mathrm{pr}_{W_{i-1}}(\vec{v_i})}{||\vec{v_i} - \mathrm{pr}_{W_{i-1}}(\vec{v_i})||}\}$, where $W_i = \mathrm{span}\{\vec{u}_{j \leq i}\}$. This is called the Gram-Schmidt orthonormalisation process.
It is obvious that the constructed set only contains unit vectors. Proposition 6.2.4 also guarantees that $\vec{u_i} \perp \vec{u_j}$ for $j < i$. By the symmetry of inner products, orthogonality is reflexive and hence the set is orthonormal. Furthermore, by Proposition 6.1.3 and the basis theorem, it is an orthonormal basis for the real inner product space.

An interesting application of the Gram-Schmidt process is to generate Legendre polynomials. Consider $\mathbb{R}_2[x]$, the vector space of real polynomials of degree less than or equal to 2. Define an inner product on it such that $\langle f, g\rangle = \int_{-1}^{1} fg\ dx$, where $f, g \in \mathbb{R}_2[x]$. (The proof of the validity of the definition is left as an exercise to the readers.) We can then apply the Gram-Schmidt process on its standard basis, $\{1, x, x^2\}$, to acquire an orthonormal basis for the space:

1. As $\langle 1, 1\rangle = \displaystyle\int_{-1}^{1} 1\ dx = 2$, $\vec{u_1} = \dfrac{1}{\sqrt{2}} = \dfrac{\sqrt{2}}{2}$.
2. $\mathrm{pr}_{W_1}(x) = \langle x, \dfrac{\sqrt{2}}{2}\rangle 1 = \displaystyle\int_{-1}^{1} \frac{\sqrt{2}}{2} x\ dx = 0$, and $||x|| = \sqrt{\displaystyle\int_{-1}^{1} x^2\ dx} = \sqrt{[\frac{x^3}{3}]_{-1}^{1}} = \dfrac{\sqrt{6}}{3}$.

   Hence, $\vec{u_2} = \dfrac{x - 0}{||x - 0||} = \dfrac{x}{\frac{\sqrt{6}}{3}} = \dfrac{\sqrt{6}}{2} x$.

3. $$\begin{aligned}\text{pr}_{W_2}(x) &= \langle x^2, \frac{\sqrt{2}}{2}\rangle \frac{\sqrt{2}}{2} + \langle x^2, \frac{\sqrt{6}}{2}x\rangle \frac{\sqrt{6}}{2}x \\ &= (\int_{-1}^{1} \frac{\sqrt{2}x^2}{2}\,dx)\frac{\sqrt{2}}{2} + (\int_{-1}^{1} \frac{\sqrt{6}x^3}{2}\,dx)\frac{\sqrt{6}}{2}x \\ &= [\frac{\sqrt{2}x^3}{6}]_{-1}^{1}\frac{\sqrt{2}}{2} + 0 \\ &= \frac{1}{3}\end{aligned}$$

$$||x^2 - \frac{1}{3}|| = \sqrt{\int_{-1}^{1} (x^2 - \frac{1}{3})^2\,dx} = \sqrt{\int_{-1}^{1} x^4 - \frac{2}{3}x^2 + \frac{1}{9}\,dx} = \sqrt{[\frac{x^5}{5} - \frac{2}{3}\cdot\frac{x^3}{3} + \frac{1}{9}x]_{-1}^{1}} = \frac{2\sqrt{10}}{15}$$

Hence, $\vec{u_2} = \frac{x^2 - \frac{1}{3}}{||x^2 - \frac{1}{3}||} = \frac{x^2 - \frac{1}{3}}{\frac{2\sqrt{10}}{15}} = \frac{\sqrt{10}}{4}(3x^2 - 1)$.

Definition 6.2.3: Orthogonal complementation

Given a real inner product space V with a subspace W, the orthogonal complementation of W, $W^{\perp} = \{\vec{v} \in V \mid \vec{v} \perp \vec{w}, \vec{w} \in W\}$.

Proposition 6.2.5

V is the direct sum of W and $W^{\perp}$, i.e. $V = W \oplus W^{\perp}$, for any subspace W.

Proof We first have to make sure that $W^{\perp}$ is a vector subspace of V, which can be achieved by the subspace test:

1. Suppose $\vec{w_1'}, \vec{w_2'} \in W^{\perp}$, and $\vec{w_1}$ and $\vec{w_2}$ are their corresponding orthogonal vectors in W respectively. Note that $\text{pr}_W(\vec{w_i} + \vec{w_i'}) = \vec{w_i}$. (This can be verified by simple calculations.) Thus by linearity, $\text{pr}_W(\vec{w_1} + \vec{w_1'} + \vec{w_2} + \vec{w_2'}) = \vec{w_1} + \vec{w_2}$. By Proposition 6.2.4, $\vec{w_1'} + \vec{w_2'} \perp \vec{w_1} + \vec{w_2} \in W$. In other words, $\vec{w_1'} + \vec{w_2'} \in W^{\perp}$.
2. For any $r \in \mathbb{R}$, by bilinearity, $\langle r\vec{w'}, \vec{w}\rangle = r\langle \vec{w'}, \vec{w}\rangle = 0$. Hence, $r\vec{w'} \in W^{\perp}$.

Next, Proposition 6.2.4 implies every vector in V can expressed as the sum of a vector from W and another from $W^{\perp}$.

Finally, the positive definiteness of inner products ensures that only the zero vector is orthogonal to itself, meaning that $W \cap W^{\perp} = \{\vec{0}\}$. By Proposition 3.5.4, $V = W + W^{\perp} = W \oplus W^{\perp}$.

Proposition 6.2.6

Orthogonal projection is well-defined, i.e. it does not depend on the choice of basis for the subspace.

Proof As $V = W \oplus W^{\perp}$ (Proposition 6.2.5), a vector $\vec{v}$ in V can be uniquely represented in the form of $\vec{w} + \vec{w'}$, where $\vec{w} \in W$, $\vec{w'} \in W^{\perp}$.

Suppose $\vec{v}$ has two distinct projections onto W. Let them be $\vec{p_1}$ and $\vec{p_2}$. Then either $\vec{v} - \vec{p_1} \notin W^{\perp}$ or $\vec{v} - \vec{p_2} \notin W^{\perp}$. This violates Proposition 6.2.4. We have hence proved the statement by contradiction.

Proposition 6.2.7

Suppose V is a real inner product space and W is a subspace of it. Then for any $\vec{v} \in V$ and $\vec{w} \in W$, $||\vec{v} - \text{pr}_W(\vec{v})|| < ||\vec{v} - \vec{w}||$ if $\vec{w} \neq \text{pr}_W(\vec{v})$.

Proof By Proposition 6.2.4, $\vec{v} - \mathrm{pr}_W(\vec{v}) \perp \mathrm{pr}_W(\vec{v}) - \vec{w}$. Hence, by the Pythagorean theorem:

$$||\vec{v} - \vec{w}||^2 = ||\vec{v} - \mathrm{pr}_W(\vec{v}) + \mathrm{pr}_W(\vec{v}) - \vec{w}||^2 = ||\vec{v} - \mathrm{pr}_W(\vec{v})||^2 + ||\,\mathrm{pr}_W(\vec{v}) - \vec{w}||^2$$

This implies $||\vec{v} - \vec{w}||$ is smallest when $||\,\mathrm{pr}_W(\vec{v}) - \vec{w}|| = 0$, which occurs only when $\vec{w}$ is the projection vector. (Proposition 6.2.6)

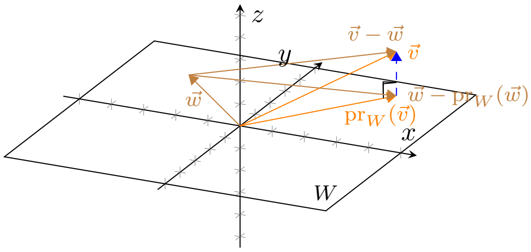

Figure 6.2.2: Visual illustration of the proof of Proposition 6.2.7

6.3 Complex inner product spaces

This section requires basic knowledge of complex numbers.

Complex inner product spaces are a generalisation of real inner product spaces, the definition of which mimic that of real inner product spaces. Therefore, most results from the previous two sections will still hold, albeit requiring new proofs.

Definition 6.3.1: Complex inner product space

A complex inner product space V is a vector space over $\mathbb{C}$ with an additional map $\langle\cdot,\cdot\rangle : V \times V \to \mathbb{C}$, which is known as the inner product and has the following properties for all $\vec{v}, \vec{w} \in V$ and $r, s \in \mathbb{C}$:

1. $\langle \vec{v}, \vec{v}\rangle \in \mathbb{R}$, $\langle \vec{v}, \vec{v}\rangle > 0$ for $\vec{v} \neq \vec{0}$ and $\langle \vec{0}, \vec{0}\rangle = 0$ (Positive definiteness)
2. $\langle \vec{v}, \vec{w}\rangle = \overline{\langle \vec{w}, \vec{v}\rangle}$ (Conjugate symmetry)
3. $\langle \vec{v_1} + \vec{v_2}, \vec{w}\rangle = \langle \vec{v_1}, \vec{w}\rangle + \langle \vec{v_2}, \vec{w}\rangle$ and $\langle r\vec{v}, \vec{w}\rangle = r\langle \vec{v}, \vec{w}\rangle$ (Linearity in first argument)

By (2) and (3), the inner product is antilinear in the second argument, i.e. $\langle \vec{v}, \vec{w_1} + \vec{w_2}\rangle = \langle \vec{v_1}, \vec{w_1}\rangle + \langle \vec{v}, \vec{w_2}\rangle$ and $\langle r\vec{v}, \vec{w}\rangle = \overline{r}\langle \vec{v}, \vec{w}\rangle$.

Similar to real vector spaces, we also have a dot product (sometimes called the canonical inner product) for complex vector spaces: For any $\vec{v} = \langle v_1, v_2, ..., v_n\rangle$ and $\vec{w} = \langle w_1, w_2, ..., w_n\rangle$ in $\mathbb{C}^n$, $\vec{v} \cdot \vec{w} = \sum v_i \overline{w_i}$.
We now verify it meets the requirements in Definition 6.3.1:
Positive Definiteness $\vec{v} \cdot \vec{v} = \sum v_i \overline{v_i} = \sum \Re(v_i)^2 > 0$ for $\vec{v} \neq \vec{0}$ and $\vec{0} \cdot \vec{0} = \sum 0 \cdot 0 = 0$
Conjugate Symmetry $\vec{v} \cdot \vec{w} = \sum v_i \overline{w_i} = \overline{\sum \overline{v_i \overline{w_i}}} = \overline{\sum \overline{v_i} w_i} = \overline{\vec{w} \cdot \vec{v}}$
Linearity in First Argument $(\vec{v} + \vec{v'}) \cdot \vec{w} = \sum (v_i + v_i')w_i = \sum v_i w + v_i' w = \vec{v} \cdot \vec{w} + \vec{v'} \cdot \vec{w}$

$$(r\vec{v}) \cdot \vec{w} = \sum r v_i w_i = r \sum v_i w_i = r(\vec{v} \cdot \vec{w})$$

The magnitude and orthogonality of complex vectors are defined in the same way as that of real vectors. Yet, although the Cauchy-Schwarz inequality and the triangle inequality are still valid, they require different proofs as the complex inner product is not symmetric. (It is conjugate symmetric instead.)

Proof of Cauchy-Schwarz inequality
For any vectors $\vec{v}$ and $\vec{w}$ in a complex inner product space, let $\lambda = \frac{\langle \vec{v}, \vec{w}\rangle}{||\vec{w}||^2}$. Observe that $\lambda\vec{w}$ is exactly the projection of $\vec{v}$ onto $\vec{w}$. Thus, by Proposition 6.2.4, $\vec{v} - \lambda\vec{w} \perp \vec{w}$. (Readers should verify for themselves

that the proof of the proposition still applies (with a little modification).) Hence,

$$\begin{aligned}0 \leq ||\vec{v}-\lambda\vec{w}||^2 = \langle \vec{v}-\lambda\vec{w}, \vec{v}-\lambda\vec{w}\rangle &= \langle \vec{v}-\lambda\vec{w}, \vec{v}\rangle - \overline{\lambda}\langle \vec{v}-\lambda\vec{w}, \vec{w}\rangle \text{ (Antilinearity)}\\ &= \langle \vec{v}-\lambda\vec{w}, \vec{v}\rangle \text{ (Orthogonality)}\\ &= \langle \vec{v}, \vec{v}\rangle - \lambda\langle \vec{w}, \vec{v}\rangle \text{ (Linearity)}\\ &= ||\vec{v}||^2 - \frac{\langle \vec{v}, \vec{w}\rangle\langle \vec{w}, \vec{v}\rangle}{||\vec{w}||^2}\\ &= ||\vec{v}||^2 - \frac{|\langle \vec{v}, \vec{w}\rangle|^2}{||\vec{w}||^2} \text{ (Conjugate symmetry)}\end{aligned}$$

By simple rearrangement, we get $|\langle \vec{v}, \vec{w}\rangle|^2 \leq ||\vec{v}||^2||\vec{w}||^2$ and thus $|\langle \vec{v}, \vec{w}\rangle| \leq ||\vec{v}||\ ||\vec{w}||$.

Proof of triangle inequality

By the Cauchy-Schwarz inequality,

$$\begin{aligned}&|\langle \vec{v}, \vec{w}\rangle| \leq ||\vec{v}||\ ||\vec{w}||\\ \implies &2\Re(\langle \vec{v}, \vec{w}\rangle) \leq 2||\vec{v}||\ ||\vec{w}||\\ \implies &\langle \vec{v}, \vec{v}\rangle + \langle \vec{v}, \vec{w}\rangle + \langle \vec{w}, \vec{v}\rangle + \langle \vec{w}, \vec{w}\rangle \leq ||\vec{v}||^2 + 2||\vec{v}||\ ||\vec{w}|| + ||\vec{w}||^2 \text{ (Conjugate symmetry)}\\ \implies &\langle \vec{v}+\vec{w}, \vec{v}+\vec{w}\rangle \leq (||\vec{v}|| + ||\vec{w}||)^2\\ \implies &||\vec{v}+\vec{w}|| \leq ||\vec{v}|| + ||\vec{w}||\end{aligned}$$

Next, the definition for directional cosine has to be rewritten as complex cosine values do not make sense if we want the angle to be analogous to its geometric counterpart.

Definition 6.3.2: Directional cosine

The angle θ between two vectors $\vec{v}$ and $\vec{w}$ in a complex inner product space is defined such that $\cos\theta = \frac{\Re(\langle \vec{v}, \vec{w}\rangle)}{||\vec{v}||\ ||\vec{w}||}$ and $\theta \in [0, \pi)$.

Finally, notice that all new definitions introducted in this section are compatible with their real counterparts. Hence, all real inner product spaces are complex inner product spaces.

6.4 Representations of Inner Products

Inner product spaces are complex by default in this section. All results naturally apply to real inner product spaces as well.

Definition 6.4.1: Positive definite matrix

A matrix $C \in M_n\ \mathbb{C}$ is positive definite if for all $\vec{v} \in \mathbb{C}^n \setminus \{\vec{0}\}$, $\overline{\vec{v}^T} C\vec{v} > 0$.

Definition 6.4.2: Positive semidefinite matrix

A matrix $C \in M_n\ \mathbb{C}$ is positive semidefinite if for all $\vec{v} \in \mathbb{C}^n \setminus \{\vec{0}\}$, $\overline{\vec{v}^T} C\vec{v} \geq 0$.

Proposition 6.4.1

A positive definite matrix C is invertible.

Proof By definition, the kernel of C is $\vec{0}$. By the invertible matrix theorem, C is invertible.

Theorem 6.4.1

Suppose V is a vector space over $\mathbb{C}$ with a basis $B = \{\vec{u_1}, \vec{u_2}, ..., \vec{u_n}\}$. A inner product space can be constructed by defining the inner product as $\langle \vec{v}, \vec{w} \rangle = \overline{\gamma_B(\vec{v})^T} C \gamma_B(\vec{w})$, where C is a symmetric positive definite matrix. Conversely, the definition of every inner product defined on complex vector spaces can be expressed in the above form given an orthonormal basis B.

Proof **(→)** *Positive Definiteness* By definition.
Conjugate Symmetry $\overline{\gamma_B(\vec{v})^T} C \gamma_B(\vec{w}) = (\overline{\gamma_B(\vec{v})^T} C \gamma_B(\vec{w}))^T$ (Scalars are symmetric)

$$= \gamma_B(\vec{w})^T C^T \overline{\gamma_B(\vec{v})} \text{ (Contravariance)}$$

$$= \overline{\overline{\gamma_B(\vec{w})^T} C \gamma_B(\vec{v})} \text{ (Symmetry of } C\text{)}$$

Linearity Given by Proposition 4.1.2 4.1.2 and Proposition 4.1.3.
(←) Let $C = (\langle \vec{u_j}, \vec{u_i} \rangle)$. Then $\overline{\gamma_B(\vec{v})^T} C \gamma_B(\vec{w}) = \overline{\gamma_B(\vec{v})^T} \left\langle \sum (\gamma_B(\vec{w}))_i \langle \vec{u_i}, \vec{u_1} \rangle, ..., \sum (\gamma_B(\vec{w}))_i \langle \vec{u_i}, \vec{u_n} \rangle \right\rangle$

$$= \overline{\gamma_B(\vec{v})^T} \langle \langle \vec{w}, \vec{u_1} \rangle, ..., \langle \vec{w}, \vec{u_n} \rangle \rangle \text{ (Linearity)}$$

$$= \langle \vec{v}, \vec{w} \rangle \text{ (Conjugate linearity)}$$

Definition 6.4.3: Isometry

A map is isometric if it preserves distance.

Theorem 6.4.2

Every inner product space is isometrically isomorphic with the dot product space.

Proof Find any orthonormal basis $B = \{\vec{u_1}, \vec{u_2}, ..., \vec{u_n}\}$ for the inner product space V. Then γ_B is already an isometric isomorphism because by Pythagorean theorem, for $r_i \in \mathbb{C}$:

$$||r_1\vec{u_1} + ... + r_n\vec{u_n}||_V^2 = |r_1|^2 + ... + |r_n|^2 = ||r_1\vec{e_1} + ... + r_n\vec{e_n}||_{\mathbb{C}^n}^2$$

Proposition 6.4.2

Suppose V and W are inner product spaces. If $\phi : V \to W$ is a linear isometry, ϕ is injective and preserves the inner product.

Proof For any $\vec{v} \in \ker\phi$, $||\phi(\vec{v})||_W = ||\vec{v}||_V = 0$. By positive definiteness, $\vec{v} = \vec{0}$. In other words, $\ker\phi = \{\vec{0}\}$, which means ϕ must be injective.
For any $\vec{v}, \vec{w} \in V$:

$$\langle \vec{v} + \vec{w}, \vec{v} + \vec{w} \rangle_V = \langle \phi(\vec{v} + \vec{w}), \phi(\vec{v} + \vec{w}) \rangle_W$$
$$\implies ||\vec{v}||_V^2 + ||\vec{w}||_V^2 + \Re(\langle \vec{v}, \vec{w} \rangle_V) = ||\phi(\vec{v})||_W^2 + ||\phi(\vec{w})||_W^2 + \Re(\langle \phi(\vec{v}), \phi(\vec{w}) \rangle_W)$$
$$\implies \Re(\langle \vec{v}, \vec{w} \rangle_V) = \Re(\langle \phi(\vec{v}), \phi(\vec{w}) \rangle_W)$$

$$\langle i\vec{v} + \vec{w}, i\vec{v} + \vec{w} \rangle_V = \langle \phi(i\vec{v} + \vec{w}), \phi(i\vec{v} + \vec{w}) \rangle_W$$
$$\implies -||\vec{v}||_V^2 + ||\vec{w}||_V^2 + \Re(i\langle \vec{v}, \vec{w} \rangle_V) = -||\phi(\vec{v})||_W^2 + ||\phi(\vec{w})||_W^2 + \Re(i\langle \phi(\vec{v}), \phi(\vec{w}) \rangle_W)$$
$$\implies \Im(\langle \vec{v}, \vec{w} \rangle_V) = \Im(\langle \phi(\vec{v}), \phi(\vec{w}) \rangle_W)$$

Definition 6.4.4: Inner product functional

Suppose V is a vector space over $\mathbb{C}$ with an inner product $\langle .,. \rangle$. For every $\vec{u} \in V$, the inner product functional on $\vec{u}$, $l_{\vec{u}} \in V^*$, is defined by $l_{\vec{u}}(\vec{v}) = \langle \vec{v}, \vec{u} \rangle$.

Theorem 6.4.3: Riesz representation theorem

For any finitely dimensional inner product space V, l is a bijection between V and V^*.

Proof *Injectivity* Suppose there exist $\vec{v}, \vec{w} \in V$ such that $l_{\vec{v}} = l_{\vec{w}}$. Then:

$$\begin{aligned} & l_{\vec{v}}(\vec{v} - \vec{w}) = l_{\vec{w}}(\vec{v} - \vec{w}) \\ \implies & \langle \vec{v} - \vec{w}, \vec{v} \rangle = \langle \vec{v} - \vec{w}, \vec{w} \rangle \\ \implies & \langle \vec{v} - \vec{w}, \vec{v} - \vec{w} \rangle = 0 \text{ (Conjugate linearity)} \\ \implies & \vec{v} - \vec{w} = \vec{0} \text{ (Positive definiteness)} \\ \implies & \vec{v} = \vec{w} \end{aligned}$$

Surjectivity Find an orthonormal basis $B = \{\vec{u_1}, \vec{u_2}, ..., \vec{u_n}\}$ for V. By Proposition 5.4.1, $B^* = \{\vec{u_1}^*, \vec{u_2}^*, ..., \vec{u_n}^*\}$ is a basis for V^*. For any $f \in V^*$, if $f = r_1\vec{u_1}^* + ... + r_n\vec{u_n}^*$, for any $\vec{v} = s_1\vec{u_1} + ... + s_n\vec{u_n} \in V$:

$$\begin{aligned} l_{\overline{r_1}\vec{u_1}+...+\overline{r_n}\vec{u_n}}(\vec{v}) &= \langle \vec{v}, \overline{r_1}\vec{u_1} + ... + \overline{r_n}\vec{u_n} \rangle \\ &= r_1\langle \vec{v}, \vec{u_1} \rangle + ... + r_n\langle \vec{v}, \vec{u_n} \rangle \text{ (Conjugate linearity)} \\ &= r_1 s_1 + ... + r_n s_n \text{ (Proposition 6.2.1)} \\ &= f(\vec{v}) \text{ (By definition)} \end{aligned}$$

Proposition 6.4.3

For any subspace W of a finitely dimensional inner product space, $W^\perp = W^0$ under l.

Proof By definition.

Proposition 6.4.4

For any subspace W of a finitely dimensional inner product space, $(W^\perp)^\perp = W$.

Proof By Proposition 6.4.3 and Proposition 5.6.4.

Note that this result is not necessarily true for infinitely dimensional spaces. In such a case, only $W \subseteq (W^\perp)^\perp$ is guaranteed. This is because l is not a bijection when the dimension is infinite.

Proposition 6.4.5

For any $A \in \mathrm{Mat}_{m \times n}\,\mathbb{C}$:

1. $\ker \overline{A^T} = (\operatorname{im} A)^\perp$
2. $\operatorname{im} \overline{A^T} = (\ker A)^\perp$

with respect to the dot product.

Proof First turn A into a linear transformation $\phi_A : \mathbb{C}^n \to \mathbb{C}^m$. By Theorem 5.6.1, we have:

$$\ker \phi_A^* = (\operatorname{im} \phi_A)^0 \tag{1}$$

$$\operatorname{im} \phi_A^* = (\ker \phi_A)^0 \tag{2}$$

Observe that for any $f = r_1\vec{e_1}^* + ... + r_n\vec{e_n}^*$ and $\vec{v} = s_1\vec{e_1} + ... + s_n\vec{v_n}$, where $r_i, s_i \in \mathbb{C}$:

$$\begin{aligned} \langle \vec{v}, \overline{r_1}\vec{e_1} + ... + \overline{r_n}\vec{e_n} \rangle &= r_1\langle \vec{v}, \vec{e_1} \rangle + ... + r_n\langle \vec{v}, \vec{e_n} \rangle \text{ (Conjugate symmetry)} \\ &= r_1 s_1 + ... + r_n s_n \text{ (Proposition 6.2.1)} \\ &= f(\vec{v}) \end{aligned}$$

$$\implies l^{-1}(f) = \langle \overline{r_1}, ..., \overline{r_n} \rangle$$

For $r_i, s_i \in \mathbb{C}$:

$$\overline{A^T}\langle r_1, ..., r_m \rangle = \vec{0} \iff A^T\langle \overline{r_1}, ..., \overline{r_m} \rangle = \vec{0}$$

$$\overline{A^T}\langle r_1, ..., r_m \rangle = \langle s_1, ..., s_n \rangle \iff A^T\langle \overline{r_1}, ..., \overline{r_m} \rangle = \langle \overline{s_1}, ..., \overline{s_n} \rangle$$

, which respectively implies:

$$\langle r_1, ..., r_m \rangle \in \ker \overline{A^T} \iff \langle \overline{r_1}, ..., \overline{r_m} \rangle \in \ker A^T$$

$$\langle s_1, ..., s_n \rangle \in \operatorname{im} \overline{A^T} \iff \langle \overline{s_1}, ..., \overline{s_n} \rangle \in \operatorname{im} A^T$$

Finally, recall Theorem 5.4.1, which states that $M_{(B')^*,B^*}(\phi_A^*) = A^T$, where B and B' are the standard bases for $\mathbb{C}^n$ and $\mathbb{C}^m$ respectively. Therefore, if we apply l^{-1} to both sides of (1) and (2), by Proposition 6.4.3, we obtain the proposition.

Of course, one could also derive the proposition by considering the mechanisms of matrix multiplication and the complex dot product.

7 Determinants

7.1 Determinants

For simplicity, we will introduce the ∂ notation to signify row and column deletion:
For any $A = (a_{i,j}) \in M_n\ k$, where $n > 1$, $\partial_{k,p} A = (a_{i \neq k, j \neq p}) \in M_{n-1}\ k$. $\partial_{k,p}$ is also known as a <u>minor</u> of A.
For example, $\partial_{1,1} \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} = \begin{pmatrix} \cancel{1} & \cancel{2} & \cancel{3} \\ \cancel{4} & 5 & 6 \\ \cancel{7} & 8 & 9 \end{pmatrix} = \begin{pmatrix} 5 & 6 \\ 8 & 9 \end{pmatrix}$

Theorem 7.1.1: Fundamental theorem of determinants

There exists a <u>unique</u> function $\det : M_n\ k \to k$ such that:

1. If $A^k = r_1 C_1 + r_2 C_2$, where $r_i \in k$ and $C_i \in k^n$, then $\det A = r_1 \det(A^1 \ \dots \ \underset{\substack{\uparrow \\ k^{\text{th}} \text{ column}}}{C_1} \ \dots \ A^n) + r_2 \det(A^1 \ \dots \ \underset{\substack{\uparrow \\ k^{\text{th}} \text{ column}}}{C_2} \ \dots \ A^n)$. (Multilinearity)
2. If two adjacent columns of A is identical, then $\det A = 0$. (Alternation of sign)
3. $\det I_n = 1$ (Normalisation)

Proof We will only prove the existence of such a function here. The proof of its uniqueness has to wait until later.
Define $\det \begin{pmatrix} a \end{pmatrix} = a$. Then for $n > 1$, suppose $A = (a_{i,j})$, the determinant is defined such that:

$$\det A = \sum_{j=1}^{n} (-1)^{j+1} a_{1,j} \det \partial_{1,j} A$$

If the determinant of $\partial_{1,j} A$ has not been defined yet, reuse the above definition until we get to $n = 1$.
We now have to show that the above definition fits with the requirements, which will be done through induction as the function is defined recursively:
The requirements are obviously met in the base case $(n = 1)$.
Suppose (1) to (3) is true in $M_n\ k$. Then for $A \in M_{n+1}\ k$:
Multilinearity
Let $A' = \begin{pmatrix} A^1 & \dots & C_1 & \dots & A^{n+1} \end{pmatrix} = (a'_{i,j})$ and $A'' = \begin{pmatrix} A^1 & \dots & C_2 & \dots & A^{n+1} \end{pmatrix} = (a''_{i,j})$. Then:

$$\begin{aligned} \det A &= (\sum_{j \neq k} (-1)^{j+1} a_{1,j} \partial \det \partial_{1,j} A) + (-1)^{k+1} a_{1,k} \det \partial_{1,k} A \\ &= (\sum_{j \neq k} (-1)^{j+1} a_{1,j} (r_1 \det \partial_j A' + r_2 \det \partial_j A'')) + (-1)^{k+1} (r_1 {a'}_{1,k} + r_2 {a''}_{1,k}) \det \partial_{1,k} A \\ &\because \text{For } j \neq k,\ a_{1,j} = {a'}_{1,j} = {a''}_{1,j};\ \partial_{1,k} A = \partial_{1,k} A' = \partial_{1,k} A'' \\ &= r_1 (\sum_{j=1}^{n+1} (-1)^{j+1} {a'}_{1,j} \det \partial_j A') + r_2 (\sum_{j=1}^{n+1} (-1)^{j+1} {a''}_{1,j} \det \partial_j A'') = r_1 \det A' + r_2 \det A'' \end{aligned}$$

Alternation of Sign
Suppose $A^k = A^{k+1}$. Then $\det A = (-1)^{k+1} a_{1,k} \det \partial_{1,k} A + (-1)^{k+2} a_{1,k+1} \det \partial_{1,k=1} A$.
(Determinants of other columns are all eliminated as both A^k and A^{k+1} exist in the reduced matrix.)
As $a_{1,k} = a_{1,k+1}$ and $\partial_{1,k} A = \partial_{1,k+1} A$ (by definition), the above equates to 0.
Normalisation

$$\det I_{n+1} = \det \left(\begin{array}{c:c} 1 & \mathbf{0} \\ \hdashline \mathbf{0} & I_n \end{array} \right) = \det I_n = 1$$

We will demonstrate our definition above by evaluating the determinant of a 3x3 square matrix:

$$\det\begin{pmatrix}1 & 2 & 3\\ 4 & 5 & 6\\ 7 & 8 & 9\end{pmatrix} = 1\det\begin{pmatrix}1 & 2 & 3\\ 4 & 5 & 6\\ 7 & 8 & 9\end{pmatrix} - 2\det\begin{pmatrix}1 & 2 & 3\\ 4 & 5 & 6\\ 7 & 8 & 9\end{pmatrix} + 3\det\begin{pmatrix}1 & 2 & 3\\ 4 & 5 & 6\\ 7 & 8 & 9\end{pmatrix}$$

$$= 1 * (5 * 6 - 6 * 8) - 2 * (4 * 9 - 6 * 7) + 3 * (4 * 8 - 5 * 7) = 0$$

Proposition 7.1.1

Suppose we swap any two columns in A and obtain A'. Then $\det A' = -\det A$.

Proof We first prove that the proposition is true when swapping two adjacent columns.

$$\begin{aligned}
&\det\begin{pmatrix}\dots & A^j & A^{j+1} & \dots\end{pmatrix} + \det\begin{pmatrix}\dots & A^{j+1} & A^j & \dots\end{pmatrix}\\
=&\det\begin{pmatrix}\dots & A^j & A^j & \dots\end{pmatrix} + \det\begin{pmatrix}\dots & A^j & A^{j+1} & \dots\end{pmatrix} +\\
&\det\begin{pmatrix}\dots & A^{j+1} & A^j & \dots\end{pmatrix} + \det\begin{pmatrix}\dots & A^{j+1} & A^{j+1} & \dots\end{pmatrix} \text{ (Alternation of sign)}\\
=&\det\begin{pmatrix}\dots & A^j & A^j + A^{j+1} & \dots\end{pmatrix} + \begin{pmatrix}\dots & A^{j+1} & A^j + A^{j+1} & \dots\end{pmatrix} \text{ (Multilinearity)}\\
=&\det\begin{pmatrix}\dots & A^j + A^{j+1} & A^j + A^{j+1} & \dots\end{pmatrix} \text{ (Multilinearity)}\\
=&\,0 \text{ (Alternation of sign)}\\
\implies& \det\begin{pmatrix}\dots & A^j & A^{j+1} & \dots\end{pmatrix} = -\det\begin{pmatrix}\dots & A^{j+1} & A^j & \dots\end{pmatrix}
\end{aligned}$$

To swap columns that are not adjacent to each other, we can perform multiple adjacent swappings:

$$\begin{pmatrix}\dots & A^j & \dots & A^{j+m} & \dots\end{pmatrix} \xrightarrow[m-1 \text{ steps}]{} \begin{pmatrix}\dots & A^j & A^{j+m} & \dots & \dots\end{pmatrix} \xrightarrow[m+2 \text{ steps}]{} \begin{pmatrix}\dots & A^{j+m} & \dots & A^j & \dots\end{pmatrix}$$

Hence, $\det A' = (-1)^{2m+1} \det A = -\det A$.

Proposition 7.1.2

If any two columns of a square matrix A are identical, $\det A = 0$.

Proof Construct A' by swapping two columns of A such that the columns in question are adjacent to each other. Then by Proposition 7.1.1 and alternation of sign, $\det A = -\det A' = 0$.

Proposition 7.1.3

If columns of a finite square matrix A are linearly dependent, then $\det A = 0$.

Proof By Proposition 3.6.1, there exists $A^k = \sum_{j \neq k} r_j A^j$, where $r_i \in k$. Then by multilinearity and Proposition 7.1.2, $\det A = \sum r_i \det(\dots \underset{k^{\text{th}} \text{ column}}{\underset{\uparrow}{A^j}} \dots) = 0$.

Proposition 7.1.4

If $A \in M_n\, k$ contains at least one zero row, then $\det A = 0$.

Proof Suppose there are m zero rows. Then columns of A belong to a $(n-m)$-dimensional subspace of k^n. (The proof of which is trivial.) By Proposition 3.8.3, as $n > n - m$, the columns of A are linearly dependent. Hence, per Proposition 7.1.3, $\det A = 0$.

Note that the above properties do not depend on the formula given in Theorem 7.1.1, but only the definition of a determinant. Such a distinction is important when we prove the uniqueness of determinants in the following theorem:

Theorem 7.1.2: Leibniz expansion

For all $A = (a_{i,j}) \in M_n\ k$, $\det A = \sum_{\pi \in S_n} \sigma(\pi) a_{\pi(1),1} a_{\pi(2),2} ... a_{\pi(n),n}$, where:

- The symmetric group on n letters, S_n, is the set of all bijective functions $\pi : \{1, ..., n\} \to \{1, ..., n\}$.
- The sign homomorphism, $\sigma : S_n \to \{\pm 1\} \subset k$, is defined such that $\sigma(\pi) = \begin{cases} \text{No. of reversals is even: } 1 \\ \text{No. of reversals is odd: } -1 \end{cases}$, where a reversal is defined as when $i < j < n$ but $\pi(i) > \pi(j)$.

Proof $\det A = \det \begin{pmatrix} \sum a_{i,1}\vec{e_i} & \sum a_{i,2}\vec{e_i} & ... & \sum a_{i,n}\vec{e_i} \end{pmatrix}$

$$= \sum_{i=1}^{n} a_{i,1} \det \begin{pmatrix} \vec{e_i} & \sum a_{i,2}\vec{e_i} & ... & \sum a_{i,n}\vec{e_i} \end{pmatrix} \text{ (Multilinearity)}$$

If φ is any function from $\{1, ..., n\}$ to $\{1, ..., n\}$, then by the same logic:

$$= \sum_{\varphi} a_{\varphi(1),1} a_{\varphi(2),2} ... a_{\varphi(n),n} \det \begin{pmatrix} \vec{e}_{\varphi(1)} & \vec{e}_{\varphi(2)} & ... & \vec{e}_{\varphi(n)} \end{pmatrix}$$

If φ is not injective, there are identical columns, so $\det = 0$. (Proposition 7.1.2)

If φ is not surjective, there is at least one zero row, so $\det = 0$. (Proposition 7.1.4)

$$= \sum_{\pi \in S_n} a_{\pi(1),1} a_{\pi(2),2} ... a_{\pi(n),n} \det \begin{pmatrix} \vec{e}_{\pi(1)} & \vec{e}_{\pi(2)} & ... & \vec{e}_{\pi(n)} \end{pmatrix}$$

We can swap the columns of the matrix to get the identity matrix so that we can get rid of the determinant term. However, we would have to figure out the number of necessary swaps because by Proposition 7.1.1, they change the sign of the determinant.

To do this, we can use the insertion sort algorithm:

1. Start with the first column ($\vec{e}_{\pi(1),1}$) and compare it with the second. If $\pi(1) > \pi(2)$ (a reversal), swap the first and second column.
2. Repeat the above step until $\pi(1)$ has been compared with $\pi(i)$ for all $1 < i \leq n$.
3. $\vec{e}_{\pi(1),1}$ is now in the correct position (as in the identity matrix).
4. Repeat (1) and (2) to all $\pi(j)$ for $1 < j \leq n$, comparing each to all $\pi(i)$ for $j < i \leq n$.

In the above process, the number of swaps performed is equal to the number of reversals. Hence, if $\det A$ exists, $\det A = \sum_{\pi \in S_n} \sigma(\pi) a_{\pi(1),1} a_{\pi(2),2} ... a_{\pi(n),n}$, which is a unique value for each A. This completes the proof of Theorem 7.1.1.

The naming of S_n and σ is not arbitrary: S_n is a group under composition and σ is a group homomorphism from $(S_n, \circ)$ to $(\{\pm 1\}, \times)$. We will only prove the latter as the former is trivial:

Suppose $\pi_1 \in S_n$, $\pi_2 \in S_m$ and C is a finite subset of $C_{\mathbb{N}} = \{\{i, j\} \mid i, j \in \mathbb{N}, i < j\}$. Then for $c \in C$, we can observe the following:

$\boldsymbol{\pi_2}(\pi_1(c))$	$\boldsymbol{\pi_1}(c)$	$\boldsymbol{\pi_1 \circ \pi_2}(c)$
r	r	n
r	n	r
n	r	r
n	n	n

("r" stands for reversal, "n" stands for no reversal; $\pi(c) \subset C_{\mathbb{N}}$ because π_i is injective by definition.)

Suppose $\sigma_C(\pi) =$ The number of reversals π makes on C, where $\pi \in S_n$ and $C \subseteq C_{\{1,...,n\}}$. We will prove $\sigma_C(\pi_2 \circ \pi_1) = \sigma_{\pi_1(C)}(\pi_2)\sigma_C(\pi_1)$ by induction:

When $|C| = 1$, if π reverses the unique pair $c \in C$, $\sigma(\pi) = -1$; otherwise, if π does not reverse c, $\sigma(\pi) = 1$.

As the multiplication table of $\{\pm 1\}$ is analogous to the table above, the statement is true in this case. Suppose the statement is true for $|C| = n$. Then for $|C| = n+1$, construct C' by selecting n pairs from C. Then by assumption, $\sigma_{C'}(\pi_2 \circ \pi_1) = \sigma_{\pi_1(C')}(\pi_2)\sigma_{C'}(\pi_1)$. Let the pair left out by C' be c. When we add c to the C', four cases arise:

$\boldsymbol{\pi_2}(\pi_1(c))$	$\sigma_{C'}(\pi_1)\times$	$\boldsymbol{\pi_1}(c)$	$\sigma_{\pi_1(C')}(\pi_2)\times$	$\boldsymbol{\pi_1 \circ \pi_2}(c)$	$\sigma_{C'}(\pi_1 \circ \pi_2)\times$
r	-1	r	-1	n	1
r	-1	n	1	r	-1
n	1	r	-1	r	-1
n	1	n	1	n	1
	$= \sigma_C(\pi_1)$		$= \sigma_{\pi_1(C)}(\pi_2)$		$= \sigma_C(\pi_2 \circ \pi_1)$

($\pi_1(c) \notin \pi_1(C')$ because π_1 is injective by definition.)

As all of the above cases agree with the statement, it is also true in this case.
Finally, let $C_n = C_{\{1,\ldots,n\}}$. Consider that for any n and $\pi \in S_n$, $\pi(C_n) = C_n$. Hence, by the above, $\sigma_{C_n}(\pi_2 \circ \pi_1) = \sigma_{C_n}(\pi_2)\sigma_{C_n}(\pi_1)$. In other words, $\sigma(\pi_2 \circ \pi_1) = \sigma(\pi_2)\sigma(\pi_1)$, which means σ is a group homomorphism.

7.2 Properties of Determinants

Having proved the uniqueness of a determinant, we can derive more properties of determinants.

Proposition 7.2.1

For any $A = (a_{i,j}) \in M_n\, k$, $\det A^T = \det A$.

Proof $\det A = \sum_{\pi \in S_n} \sigma(\pi) a_{\pi(1),1} a_{\pi(2),2} \ldots a_{\pi(n),n}$ (Theorem 7.1.2)

$$= \sum_{\pi \in S_n} \sigma(\pi^{-1})^{-1} a_{1,\pi^{-1}(1)} a_{2,\pi^{-1}(2)} \ldots a_{n,\pi^{-1}(n)} \text{ (Proposition 2.4.3, Proposition 2.2.5, *)}$$

$$= \sum_{\pi \in S_n} \sigma(\pi^{-1}) a_{1,\pi^{-1}(1)} a_{2,\pi^{-1}(2)} \ldots a_{n,\pi^{-1}(n)} \; (\because 1^{-1} = 1,\ (-1)^{-1} = -1)$$

$$= \det A^T \; (\because \{\pi^{-1}\} = S_n)$$

$$A = \begin{array}{|c||c|c|c|}\hline & 1 & 2 & 3 \\ \hline\hline 1 & & & a_{\pi(3),3} \\ \hline 2 & a_{\pi(1),1} & & \\ \hline 3 & & a_{\pi(2),2} & \\ \hline \end{array} \longrightarrow A^T = \begin{array}{|c||c|c|c|}\hline & 1 & 2 & 3 \\ \hline\hline 1 & & a_{1,\pi^{-1}(1)} & \\ \hline 2 & & & a_{2,\pi^{-1}(2)} \\ \hline 3 & a_{3,\pi^{-1}(3)} & & \\ \hline \end{array}$$

Figure 7.2.1: Explanation of (*)

Theorem 7.2.1: Laplace Expansion

Define the formula given in Theorem 7.1.1 to be an expansion by the first row. We can similarly obtain the determinant of a square matrix A by expanding by any row or column in A:

- Expansion by the i^{th} row: $\det A = \sum_{j=1}^{n} (-1)^{i+j} a_{i,j} \det \partial_{i,j} A$
- Expansion by the j^{th} column: $\det A = \sum_{j=1}^{n} (-1)^{i+j} a_{i,j} \det \partial_{i,j} A$

Proof Using a similar proof to that in Theorem 7.1.1, we can show that expanding by any row still satisfies multilinearity, alternation of sign and normalisation. Hence, expanding by any row returns the determinant.
Expansion by the i^{th} column of A is equivalent to expansion by the i^{th} row of A^T. By Proposition 7.2.1, they would result in the same value.

Proposition 7.2.2

The determinant of a triangular matrix A (a square matrix of which entries are entirely zero on one side of the diagonal) is the product of its diagonal entries.

Proof If A is an upper triangular matrix (meaning that entries below the diagonal of A are all zeroes), by expanding by the first row, we get:

$$\det A = \sum_{j=1}^{n} (-1)^{j+1} a_{1,j} \det \partial_{1,j} A = a_{1,1} \det \partial_{1,1} A \ (\because (\partial_{1,j\neq 1} A)^1 = \vec{0})$$

$$= a_{1,1} \sum_{j=1}^{n} (-1)^{j+1} a_{1,j} \det \partial_{1,j} (\partial_{1,1} A)$$

$$= a_{1,1} a_{2,2} \ldots a_{n,n} \text{ (By a similar logic)}$$

Similarly, we can get the same result if we expand by the first column of a lower triangular matrix.

For example, $\begin{pmatrix} 1 & 2 & 3 \\ 0 & 4 & 5 \\ 0 & 0 & 6 \end{pmatrix}$ is an upper triangular matrix and $\begin{pmatrix} 1 & 0 & 0 \\ 2 & 3 & 0 \\ 4 & 5 & 6 \end{pmatrix}$ is a lower triangular matrix.

Proposition 7.2.3

$\det \left(\begin{array}{c:c} M & P \\ \hdashline \mathbf{0} & N \end{array}\right) = \det \left(\begin{array}{c:c} M & \mathbf{0}^T \\ \hdashline P^T & N \end{array}\right) = \det M \det N$, where:

1. M is a $m \times m$ square matrix,
2. N is a $n \times n$ square matrix,
3. P is a $m \times n$ matrix,
4. $\mathbf{0}$ is a $n \times m$ zero matrix.

Proof Let A be the matrix. We will prove the above by induction. For any n:
When $m = 1$, let $M = \begin{pmatrix} m \end{pmatrix}$. By expanding by the first column, $\det A = m \det N = \det M \det N$.
Suppose the statement is true for when $m = k$. Then for $m = k + 1$:

$$\det A = \sum_{i=1}^{m} (-1)^{i+1} a_{i,1} \det \left(\begin{array}{c:c} \partial_{i,1} M & \partial_i P \\ \hdashline \partial_{,1} \mathbf{0} & N \end{array}\right) \text{ (Expansion by the first column)}$$

$$= \sum_{i=1}^{m} (-1)^{i+1} a_{i,1} \det \partial_{i,1} M \det N \text{ (By assumption)}$$

$$= \det N \sum_{i=1}^{m} (-1)^{i+1} a_{i,1} = \det N \det M$$

The second equality directly follows the above by Proposition 7.2.1.

Proposition 7.2.4

Suppose $A, B \in M_n\ k$. Then $\det AB = \det A \det B$.

Proof Suppose $A = (a_{i,j})$ and $B = (b_{i,j})$. Notice that $(AB)^j = b_{1,j}A^1 + b_{2,j}A^2 + ... + b_{n,j}A^n$. If we let $\varphi : \{1, ..., n\} \to \{1, ..., n\}$ be any function, we have:

$$\det AB = \sum_{\varphi} b_{\varphi(1),1}b_{\varphi(2),2}...b_{\varphi(n),n} \det \begin{pmatrix} A^{\varphi(1)} & A^{\varphi(2)} & ... & A^{\varphi(n)} \end{pmatrix} \text{ (Multilinearity)}$$

If φ is not bijective, there are identical columns. By Proposition 7.1.2, $\det = 0$.

$$= \sum_{\pi \in S_n} b_{\pi(1),1}b_{\pi(2),2}...b_{\pi(n),n} \det \begin{pmatrix} A^{\pi(1)} & A^{\pi(2)} & ... & A^{\pi(n)} \end{pmatrix}$$

$$= \sum_{\pi \in S_n} \sigma(\pi) b_{\pi(1),1}b_{\pi(2),2}...b_{\pi(n),n} \det \begin{pmatrix} A^1 & A^2 & ... & A^n \end{pmatrix} \text{ (Proved in Theorem 7.1.2)}$$

$$= \det A \det B$$

Proposition 7.2.5

If A is an invertible square matrix, $\det A = (\det A^{-1})^{-1}$.

Proof By Proposition 7.2.4, $AA^{-1} = I_n \implies \det AA^{-1} = \det I_n \implies \det A \det A^{-1} = 1 \implies \det A = (\det A^{-1})^{-1}$.

7.3 Determinant and Invertibility

Theorem 7.3.1: Invertible matrix theorem (3)

For any square matrix A, $\det A \neq 0 \iff A$ is invertible.

Proof **(→)** We will prove this by its negation. Suppose A is not invertible. By the previous invertible matrix theorem, this is equivalent to the linear dependence of the columns of A. By Proposition 7.1.3, this implies $\det A = 0$.
(←) Proposition 7.2.5 would not hold otherwise.

Theorem 7.3.2

$\det : (GL_n\ ,\ ) \to (k \setminus \{0\}, *)$ is group homomorphism.

Proof The validity of the codomain of the determinant is given by the invertible matrix theorem, while its homomorphism is given by Proposition 7.2.4.

The kernel of det, or in other words, the set of all $n \times n$ square matrices of determinant 1, is also known as the special linear group and is denoted as $\mathrm{SL}\, nk$.

Theorem 7.3.3: Cramer's rule

If $A \in M_n\ k$ and $\det A \neq 0$, then the unique solution to $A\vec{x} = \vec{y}$, where $\vec{x} = \langle x_1, x_2, ..., x_n \rangle, \vec{y} \in k^n$, is $x_i = (\det(A^1 \ ... \ \underset{\underset{i^{\text{th}}\text{ column}}{\uparrow}}{\vec{y}} \ ... \ A^n))(\det A)^{-1}$.

Proof Recall that $\vec{y} = x_1A^1 + x_2A^2 + ... + x_nA^n$. Hence, by multilinearity and alternation of sign, $\det\left(A^1 \;...\; \underset{\underset{i^{\text{th}}\text{ column}}{\uparrow}}{\vec{y}} \;...\; A^n\right) = \sum_j x_j \det\left(A^1 \;...\; \underset{\underset{i^{\text{th}}\text{ column}}{\uparrow}}{A^j} \;...\; A^n\right) = x_i \det A$.

The existence and uniqueness of this solution is given by the the invertible matrix theorem.

Definition 7.3.1: Adjoint matrix

The adjoint (or adjugate) matrix of a square matrix A, denoted as $\operatorname{adj} A$, is defined as the transpose of the cofactor matrix C, where $C = ((-1)^{i+j} \det \partial_{i,j}A)$.

In other words, $\operatorname{adj} A = C^T = ((-1)^{i+j} \det \partial_{j,i}A)$.

Proposition 7.3.1

If A is a square matrix and $\det A \neq 0$, $A^{-1} = (\det A)^{-1} \operatorname{adj} A$.

Proof The existence of the inverse is given by the invertible matrix theorem.

Recall that $(A^{-1})^j$ is the solution to the equation $A\vec{x} = \vec{e_j}$. Hence, by Cramer's rule, $A^{-1} = ((\det\left(A^1 \;...\; \underset{\underset{i^{\text{th}}\text{ column}}{\uparrow}}{\vec{e_j}} \;...\; A^n\right))(\det A)^{-1})$.

By expanding by the i^{th} column, we get $A^{-1} = (\det A)^{-1}((-1)^{i+j}\partial_{j,i}A) = (\det A)^{-1} \operatorname{adj} A$.

Cramer's rule may seem a more elegant way of solving systems of linear equations and finding inverses. Yet, in practice, unless only a partial solution is needed, it is a very inefficient algorithm because for a $n \times n$ matrix, n determinants of matrices of the same size are required.

7.4 Determinant and Volume

In this section, "volume" refers to the generalised notion of "length", "area", "volume", etc.

A parallelepiped is a linear object defined by a collection of vectors in a vector space. In Euclidean spaces, its meaning is easily comprehensible: it is the region bounded by the specified vectors. More specifically, if a parallelepiped is bounded by vectors $\vec{v_1}, \vec{v_2}, ..., \vec{v_n} \in \mathbb{R}^n$, it is defined as $\{\sum r\vec{v_i} \mid 0 \leq r \leq 1\}$.

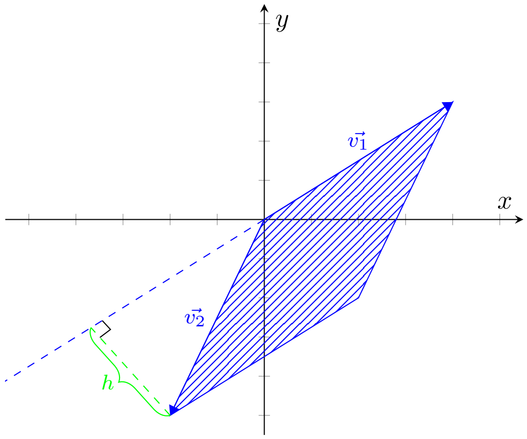

Figure 7.4.1: A parallelepiped bounded by $\vec{v_1}$ and $\vec{v_2}$ in $\mathbb{R}^2$

The volume of a parallelepiped in an Euclidean space should not be unfamiliar to us. The n-dimensional volume of a parallelepiped can be found using the iterative formula: Volume = Base Volume × Height, The base volume is the $(n-1)$-dimensional volume of the base of the parallelepiped, while the height is the magnitude of the vector not belonging to the base minus its orthogonal projection on the base. For instance, in Figure 7.4.1, if $\vec{v_1}$ is selected as the base, the corresponding height is represented by length h.

One more concept has to be introduced before we start our main exposition. Orientation of a parallelepiped describes whether it can be attained by (continuously) rotating and scaling basis vectors without entering the span of other basis vectors, while preserving the order of its defining vectors. If this is possible, the parallelepiped is in consistent orientation with the basis; otherwise, it is in inconsistent orientation. More intuitively, in the latter case, one would have to "flip" the space in order to get the parallelepiped. In the example of Figure 7.4.1, the parallelepiped is in inconsistent orientation with the standard basis. This is because the anticlockwise turn from $\hat{\imath}$ to $\hat{\jmath}$ cannot exceed π no matter how they are transformed to maintain orientation.

In higher dimensions, this definition becomes inconvenient as we have to consider all combinations of basis vectors to determine if a non-orientation-changing transformation is possible. Before an alternative method is discussed, observe that no defining vector of a parallelepiped is redundant information to its orientation. In other words, we cannot judge the orientation of a parallelepiped by any sub-region that has a lower dimension than its parent. The reason behind this is that with an extra vector, we can "flip" the base without changing its orientation. To see this in action, let's put the parallelepiped in Figure 7.4.1 in the $\mathbb{R}^3$ space. As the Figure 7.4.2 would illustrate, by rotating $\hat{\jmath}$ and $\hat{k}$ together by π along the yz-plane, $\hat{\imath}$ and $\hat{\jmath}$ can align with $\vec{v_1}$ and $\vec{v_2}$ respectively without crossing the span of other standard basis vectors.

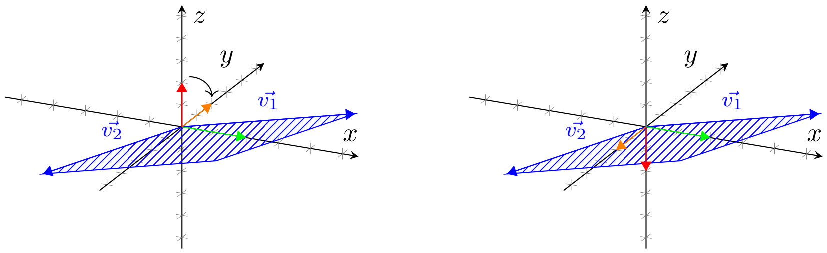

Figure 7.4.2: Parallelepiped in Figure 7.4.1 in the $\mathbb{R}^3$ space
(Green vector represents $2\hat{\imath}$, orange vector represents $2\hat{\jmath}$, red vector represents $2\hat{k}$)

Also observe that if we "flip" one of a consistently oriented parallelepiped's defining vectors (substituting it with its inverse), or in other words, replace the parallelepiped with its reflection, the orientation of the parallelepiped must change.

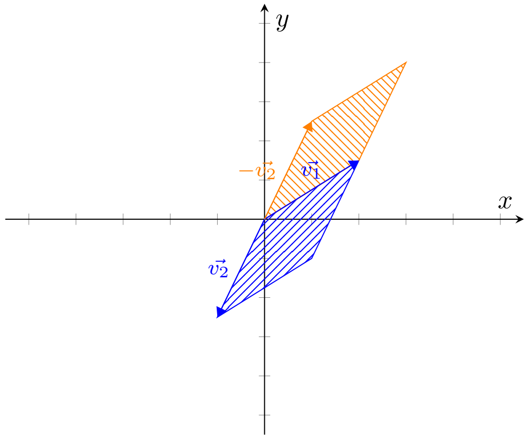

Figure 7.4.3: Parallelepiped in Figure 7.4.1 with its reflection
(Resized to preserve space)

Using these two facts, we can devise a non-geometric avenue of evaluating orientation. The basic idea is given a reference basis $\{\vec{u_1}, ..., \vec{u}_{n-1}, \vec{u_n}\}$ and the base of a parallelepiped, we should be able to produce a vector $\vec{w}$ which is orthogonal to the base and has consistent orientation with the basis. Then we can project the remaining defining vector of the parallelepiped $\vec{v}$ onto $\vec{w}$, which we will call $\vec{v'}$, and compare it with $\vec{w}$ to determine the orientation of the parallelepiped. Suppose $\vec{v'} = r\vec{w}$. If $r > 0$, the parallelepiped is of consistent orientation with the reference basis; in the contrary, if $r < 0$, the parallelepiped is ”fliped” and hence of inconsistent orientation. (In the $\mathbb{R}^3$ space, if the standard basis is taken as reference, $\vec{w}$ is the cross product of the base vectors.)

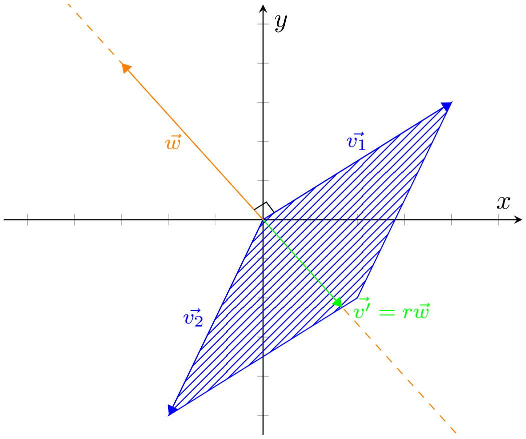

Figure 7.4.4: Graphical demonstration of the process described above
(As $r < 0$, the orientation of the parallelepiped in Figure 7.4.1 is inconsistent with the standard basis.)

There are three important properties of this process:

1. $\vec{w}$ depends linearly on the base vectors (illustrated below). As projection is a linear operation as well (Proposition 6.2.3), by Proposition 3.4.1, r is also the output of a linear map of the base vectors.
2. $r = 0$ when the base is defined by duplicated vectors. This is because the dimension of the span of any combination of vectors containing identical vectors or the zero vector must be smaller than $n - 1$, meaning that no vector added to the combination can make its orientation determinate.
3. $\vec{w} = \vec{u_n}$ when the base is defined by $\vec{u_1}, ..., \vec{u}_{n-1}$, and thus meanwhile $r = 0$.

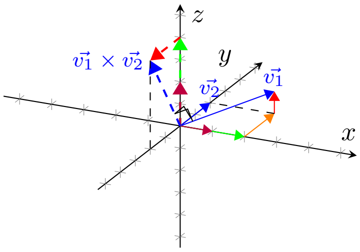

Figure 7.4.5: Geometric explanation of the linearity of $\vec{w}$
(Each dashed vector represents the cross product of its corresponding solid vector and $\vec{v_2}$.)

Observant readers may have noticed that these three properties exactly correspond to the three axioms of determinant in Theorem 7.1.1. As the determinant is unique, we can define orientation of a parallelepiped as the sign of the determinant of its matrix form in terms of a reference basis. (The formal definition will be given below.)

With the above, we can finally rigourously define ”volume” in an abstract real inner product space V of dimension $n < \infty$.

Definition 7.4.1: Parallelepiped

A parallelepiped is a linear object defined by n vectors in V. If a parallelepiped is defined by $\vec{v_1}, \vec{v_2}, ..., \vec{v_n}$, it is denoted as $P(\vec{v_1}, \vec{v_2}, ..., \vec{v_n}) = \{r_1\vec{v_1} + r_2\vec{v_2} + ... + r_n\vec{v_n} \mid r_i \leq 1, r_i \in \mathbb{R}\}$.

Definition 7.4.2: Proper parallelepiped

A parallelepiped is proper if its defining vectors are linearly independent.

Definition 7.4.3: Volume

Given a parallelepiped $P = P(\vec{v_1}, ..., \vec{v_n})$, if $n = 1$, its volume $|\alpha(P)| = ||\vec{v_1}||$. Otherwise, its volume is defined by the iterative formula $|\alpha(P)| = |\alpha(P(\vec{v_1}, ..., \vec{v}_{n-1}))| \; ||\vec{v_n} - \mathrm{pr}_{U_n}(\vec{v_n})||$, where $U_n = \mathrm{span}\{\vec{v_1}, ..., \vec{v}_{n-1}\}$.

In this case, $\vec{v_n} - \mathrm{pr}_B(\vec{v_n})$ is the height of P with respect to base U.

Proposition 7.4.1

Improper parallelepipeds have zero volume.

Proof Suppose an improper parallelepiped is defined by $\vec{v_1}, \vec{v_2}, ..., \vec{v_n} \in V$. As the defining vectors are linearly dependent, $\vec{v_n}$ can be expressed as a linear combination of $\vec{v_1}, ..., \vec{v}_{n-1}$. In other words, $\vec{v_n} \in U_n$ and thus height and consequently volume equal zero.

Proposition 7.4.2

Volume does not depend on the choice of base, that is, in the context of Definition 7.4.3, when $n \geq 2$, $|\alpha(P)| = |\alpha(P(..., \vec{v}_{i-1}, \vec{v}_{i+1}, ...))| \; ||\vec{v_i} - \mathrm{pr}_{U_i}(\vec{v_i})||$ is identical for all i.

Proof The proof is divided into two parts with respect to n, the dimension of the parallelepiped.
<u>When $n = 2$</u>

$$\begin{aligned}
|\alpha(P(\vec{v_1}))| \; ||\vec{v_2} - \mathrm{pr}_{\vec{v_1}}(\vec{v_2})|| &= ||\vec{v_1}|| \sqrt{||\vec{v_2}||^2 - ||\,\mathrm{pr}_{\vec{v_1}}(\vec{v_2})||^2} \text{ (Proposition 6.2.4, Pythagorean theorem)} \\
&= ||\vec{v_1}|| \sqrt{||\vec{v_2}||^2 - \left|\left|\langle \vec{v_2}, ||\vec{v_1}||^{-1}\vec{v_1}\rangle ||\vec{v_1}||^{-1}\vec{v_1}\right|\right|^2} \\
&= ||\vec{v_1}|| \sqrt{||\vec{v_2}||^2 - \langle \vec{v_2}, \vec{v_1}\rangle^2 ||\vec{v_1}||^{-4} ||\vec{v_1}||^2} \text{ (Linearity, Proposition 6.1.2)} \\
&= \sqrt{||\vec{v_1}||^2 ||\vec{v_2}||^2 - \langle \vec{v_2}, \vec{v_1}\rangle^2} \\
&= \sqrt{||\vec{v_1}||^2 ||\vec{v_2}||^2 - \langle \vec{v_1}, \vec{v_2}\rangle^2} \text{ (Symmetry)} \\
&= |\alpha(P(\vec{v_2}))| \; ||\vec{v_1} - \mathrm{pr}_{\vec{v_2}}(\vec{v_1})|| \text{ (Similar derivation to the above)}
\end{aligned}$$

<u>When $n \geq 3$</u>
For any i and j, by the Gram-Schimdt process, we can obtain an orthonormal basis $B = \{\vec{u_1}, ..., \vec{u}_{i-1}, \vec{u}_{i+1}, ..., \vec{u}_{j-1}\}$ for $U_{i,j}$. Then there exist non-zero $r_i, s_i, r'_i, s'_i \in \mathbb{R}$ and $\vec{u_i}, \vec{u_j}, \vec{u'_i}, \vec{u'_j} \in V$ such that $\{\vec{u_1}, ..., \vec{u_i}, ..., \vec{u_j}, ..., \vec{u_n}\}$ and $\{\vec{u_1}, ..., \vec{u'_i}, ..., \vec{u'_j}, ..., \vec{u_n}\}$ are orthonormal families and:

$$\vec{v_i} = r_1\vec{u_i} + ... + r_i\vec{u_i} + ... + r_{j-1}\vec{u}_{j-1} + r_{j+1}\vec{u}_{j+1} + ...$$

$$\vec{v_j} = s_1\vec{u_i} + ... + s_i\vec{u_i} + ... + s_j\vec{u_j} + ...$$

(The above is obtained by first projecting $\vec{v_i}$ on $U_{i,j}$, then $\vec{v_j}$ on U_j.)

$$\vec{v_j} = r'_1\vec{u_i} + ... + r'_{i-1}\vec{u}_{i-1} + r'_{i+1}\vec{u}_{i+1} + ... + r'_{j-1}\vec{u}_{j-1} + r'_j\vec{u'_j} + r'_{j+1}\vec{u}_{j+1} + ...$$

$$\vec{v_i} = s'_1\vec{u_i} + ... + s'_{i-1}\vec{u}_{i-1} + s'_i\vec{u'_i} + s'_{i+1}\vec{u}_{i+1} + ... + s'_{j-1}\vec{u}_{j-1} + s'_j\vec{u'_j} + s'_{j+1}\vec{u}_{j+1} + ...$$

(The above is obtained by first projecting $\vec{v_j}$ on $U_{i,j}$, then $\vec{v_i}$ on U_i.)
Recall that $V = U_{i,j} \oplus U_{i,j}^{\perp}$ (Proposition 6.2.5). Hence, every vector in V is represented by a unique linear combination of a vector from $U_{i,j}$ and another from $U_{i,j}^{\perp}$. As $\vec{u_i}, \vec{u_j}, \vec{u'_i}, \vec{u'_j} \in U_{i,j}^{\perp}$, by comparing

the two representations of $\vec{v_i}$ and $\vec{v_j}$ above respectively, the following can be concluded:

$$s_i\vec{u_i} + s_j\vec{u_j} = r'_j\vec{u'_j} \tag{1}$$

$$r_i\vec{u_i} = s'_i\vec{u'_j} + s'_j\vec{u'_j} \tag{2}$$

Furthermore, by the Pythagorean theorem and Proposition 6.2.4:

$$(1) \implies s_i^2||\vec{u_i}||^2 + s_j^2||\vec{u_j}||^2 = (r'_j)^2||\vec{u'_j}||^2 \implies s_i^2 + s_j^2 = (r'_j)^2 \tag{3}$$

$$(2) \implies r_i^2||\vec{u_i}||^2 = (s'_i)^2||\vec{u'_i}||^2 + (s'_j)^2||\vec{u'_j}||^2 \implies r_i^2 = (s'_i)^2 + (s'_j)^2 \tag{4}$$

$$(1) + (2) \implies (s_i + r_i)\vec{u_i} + s_j\vec{u_j} = s'_i\vec{u'_i} + (r'_j + s'_j)\vec{u'_j}$$

By the Pythagorean theorem and Proposition 6.2.4,

$$\begin{aligned}
&\implies (s_i + r_i)^2||\vec{u_i}||^2 + s_j^2||\vec{u_j}||^2 = (s'_i)^2||\vec{u'_i}||^2 + (r'_j + s'_j)^2||\vec{u'_j}||^2 \\
&\implies (s_i + r_i)^2 + s_j^2 = (s'_i)^2 + (r'_j + s'_j)^2 \\
&\implies s_i^2 + 2r_is_i + r_i^2 + s_j^2 = (s'_i)^2 + (r'_j)^2 + 2r'_js'_j + (s'_j)^2 \\
&\implies 2r_is_i = 2r'_js'_j \text{ (By (3) and (4))} \\
&\implies r_is_i = r'_js'_j && (5)
\end{aligned}$$

$$\begin{aligned}
(3) \times (4) &\implies r_i^2s_i^2 + r_i^2s_j^2 = (r'_j)^2(s'_i)^2 + (r'_j)^2(s'_j)^2 \\
&\implies r_i^2s_j^2 = (r'_j)^2(s'_i)^2 \text{ (By (5))} \\
&\implies |r_is_j| = |r'_js'_i| && (6)
\end{aligned}$$

Therefore:

$$\begin{aligned}
|\alpha(P(..., \vec{v}_{j-1}, \vec{v}_{j+1}, ...))| \, ||\vec{v_j} - \text{pr}_{U_j}(\vec{v_j})|| &= |\alpha(P(..., \vec{v}_{i-1}, \vec{v}_{i+1}, ..., \vec{v}_{j-1}, \vec{v}_{j+1}, ...))| \, ||\vec{v_i} - \text{pr}_{U_{i,j}}(\vec{v_i})|| \, |s_j| \\
&= |\alpha(P(..., \vec{v}_{i-1}, \vec{v}_{i+1}, ..., \vec{v}_{j-1}, \vec{v}_{j+1}, ...))| \, |r_i| \, |s_j|
\end{aligned}$$

$$\begin{aligned}
|\alpha(P(..., \vec{v}_{i-1}, \vec{v}_{i+1}, ...))| \, ||\vec{v_i} - \text{pr}_{U_i}(\vec{v_i})|| &= |\alpha(P(..., \vec{v}_{i-1}, \vec{v}_{i+1}, ..., \vec{v}_{j-1}, \vec{v}_{j+1}, ...))| \, ||\vec{v_j} - \text{pr}_{U_{i,j}}(\vec{v_j})|| \, |s'_i| \\
&= |\alpha(P(..., \vec{v}_{i-1}, \vec{v}_{i+1}, ..., \vec{v}_{j-1}, \vec{v}_{j+1}, ...))| \, |r'_j| \, |s'_i|
\end{aligned}$$

By (6), the two expressions are equal.

An immediate consequence of this proposition is that swapping the defining vectors of a parallelepiped does not change its volume.

Definition 7.4.4: Orientation

Two proper parallelepipeds $P = P(\vec{v_1}, \vec{v_2}, ..., \vec{v_n})$ and $Q = P(\vec{w_1}, \vec{w_2}, ..., \vec{w_n})$ in V have consistent orientation if $\det\begin{pmatrix}\gamma_P(\vec{w_1}) & \gamma_P(\vec{w_2}) & ... & \gamma_P(\vec{w_n})\end{pmatrix} = \det P_{Q,P} > 0$. This is denoted as $P \sim Q$.

Proposition 7.4.3

Orientation consistency is an equivalence relation.

Proof For any proper parallelepipeds P, Q and R in V:

Reflexivity $\det\begin{pmatrix}\gamma_P(\vec{v_1}) & \gamma_P(\vec{v_2}) & ... & \gamma_P(\vec{v_n})\end{pmatrix} = \det I_n = 1 > 0 \implies P \sim P$

Symmetry $P \sim Q \implies \det P_{Q,P} > 0 \implies P_{P,Q} > 0$ (Proposition 5.6.1,Proposition 7.3.1)

$\implies Q \sim P$

Transitivity $P \sim Q, Q \sim R \implies P_{R,P} = P(\gamma_P \circ \gamma_R^{-1}) = P(\gamma_P \circ \gamma_Q^{-1} \circ \gamma_Q \circ \gamma_R^{-1})$

$= P(\gamma_P \circ \gamma_Q^{-1})P(\gamma_Q \circ \gamma_R^{-1})$ (Proposition 5.1.4)

$= P_{R,Q}P_{Q,P} \implies P \sim R$

Proposition 7.4.4

There are only two orientations in V, i.e. there are only two equivalence classes with respect to orientation consistency.

Proof Select any basis B for V as reference. By the invertible matrix theorem, the linear independence of columns implies the determinant is not zero. Therefore, for any proper parallelepiped P in V, either $\det P_{P,B} > 0$ or $\det P_{P,B} < 0$.
By transitivity, all consistently oriented parallelepipeds (with respect to B) are consistently oriented with each other, forming one equivalence class.
For any two parallelepipeds P and Q where $\det P_{P,B}, \det P_{Q,B} < 0$, by Proposition 5.6.1 and Proposition 7.3.1, $\det P_{Q,P} = \det P_{Q,B}P_{B,P} = (\det P_{Q,B})(\det P_{P,B})^{-1} > 0$, which means $P \sim Q$. Hence, parallelepipeds not consistently oriented with B form the only other equivalence class.

Definition 7.4.5: Oriented volume

Given a reference basis B for V and a parallelepiped P, if P is proper, its oriented volume (with respect to B) $\alpha(P) = (\operatorname{sgn}\det P_{P,B})|\alpha(P)|$. Otherwise, if P is improper, its oriented volume is defined to be 0.

This explains the choice of notation for volume.

Theorem 7.4.1: Gram determinant

Given an orthonormal reference basis $B = \{\vec{u_1}, \vec{u_2}, ..., \vec{u_n}\}$ for V, the oriented volume of a parallelepiped $P = P(\vec{v_1}, \vec{v_2}, ..., \vec{v_n})$ in V equals the determinant of its Gram matrix with respect to B, i.e.:

$$\alpha(P) = \det G(P) = \det \begin{pmatrix} \langle \vec{v_1}, \vec{u_1}\rangle & \langle \vec{v_2}, \vec{u_1}\rangle & ... & \langle \vec{v_n}, \vec{u_1}\rangle \\ \langle \vec{v_1}, \vec{u_2}\rangle & \langle \vec{v_2}, \vec{u_2}\rangle & ... & \langle \vec{v_n}, \vec{u_2}\rangle \\ \vdots & \vdots & \vdots & \vdots \\ \langle \vec{v_1}, \vec{u_n}\rangle & \langle \vec{v_2}, \vec{u_n}\rangle & ... & \langle \vec{v_n}, \vec{u_n}\rangle \end{pmatrix}$$

Proof Define the function $f : M_n\,\mathbb{R} \to \mathbb{R}$ such that $f(A) = \alpha(P(A^1 \cdot B, A^2 \cdot B, ..., A^n \cdot B))$. According to the fundamental theorem of determinants, if f meets the three requirements in the theorem, it is identical to the determinant. It is trivial to show that $f(I_n) = \alpha(P(\vec{u_1}, \vec{u_2}, ..., \vec{u_n})) = 1$, while alternation of sign is true by definition. Therefore, the only item left to prove is linearity.
For any two column vectors $C_1, C_2 \in \mathbb{R}^n$, let $\vec{v} = C_1 \cdot B$ and $\vec{w} = C_2 \cdot B$. Then for any $r, s \in \mathbb{R}$, if $rC_1 + sC_2$ is the j^{th} column of $A \in M_n\,\mathbb{R}$, observe that:

$$f(A) = f(\begin{pmatrix} ... & rC_1 + sC_2 & \boldsymbol{...} \end{pmatrix})$$

$$= \alpha(P(..., r\vec{v} + s\vec{w}, \boldsymbol{...}))$$

By Proposition 7.4.2,

$$= (\operatorname{sgn}\det\begin{pmatrix} ... & rC_1 + sC_2 & \boldsymbol{...} \end{pmatrix})|\alpha(P(..., \boldsymbol{...}))|\ ||r\vec{v} + s\vec{w} - \operatorname{pr}_{U_j}(r\vec{v} + s\vec{w})||$$

By Proposition 6.2.3,

$$= \operatorname{sgn}(r\det\begin{pmatrix} ... & C_1 & \boldsymbol{...} \end{pmatrix} + s\det\begin{pmatrix} ... & C_2 & \boldsymbol{...} \end{pmatrix})|\alpha(P(..., \boldsymbol{...}))|\ ||r\vec{v} - r\operatorname{pr}_{U_j}(\vec{v}) + s\vec{w} - s\operatorname{pr}_{U_j}(\vec{w})||$$

If $\vec{w}$ is linearly dependent on other columns of A, as we have shown in Proposition 7.4.1, $s\vec{w} - s\operatorname{pr}_{U_j}(\vec{w}) = \vec{0}$. Meanwhile, $\det\begin{pmatrix} ... & C_2 & \boldsymbol{...} \end{pmatrix} = 0$ by Proposition 7.1.3. Hence, the above reduces to $\operatorname{sgn}(r\det\begin{pmatrix} ... & C_1 & \boldsymbol{...} \end{pmatrix})|\alpha(P(..., \boldsymbol{...}))|\ ||r\vec{v} - r\operatorname{pr}_{U_j}(\vec{v})|| = rf(\begin{pmatrix} ... & C_1 & \boldsymbol{...} \end{pmatrix})$, which displays linearity as $sf(\begin{pmatrix} ... & C_2 & \boldsymbol{...} \end{pmatrix}) = 0$ due to Proposition 7.1.3. The case when $\vec{w}$ is linearly dependent is the same.

Otherwise, if columns of A define a proper parallelepiped, let $\vec{h_1} = \vec{v} - \mathrm{pr}_{U_j}(\vec{v})$ and $\vec{h_2} = \vec{w} - \mathrm{pr}_{U_j}(\vec{w})$. Recall that by Proposition 6.2.5, $V = U_j \oplus U_j^{\perp}$. By the dimension formula, $\dim U_j^{\perp} = 1$. Thus, $\vec{u_j}$ is already a basis for $U_j^{\perp}$. As $\vec{h_1}, \vec{h_2} \in U_j^{\perp}$ (Proposition 6.2.4), there exist $t_1, t_2 \in \mathbb{R}$ such that $\vec{h_1} = t_1\vec{u_j}$ and $\vec{h_2} = t_2\vec{u_j}$. At the same time, $C_1 = \gamma_B(\vec{h_1}) + \gamma_B(\vec{h_1}) = \gamma_B(\vec{h_1}) + t_1\vec{e_j}$. As $\gamma_B(\vec{h_1})$ is linearly dependent on other columns, by linearity and Proposition 7.1.3, $r\det\begin{pmatrix}\dots & C_1 & \dots\end{pmatrix} = t_1\det\begin{pmatrix}\dots & \vec{u_j} & \dots\end{pmatrix}$. Similarly, $\det\begin{pmatrix}\dots & C_2 & \dots\end{pmatrix} = t_2\det\begin{pmatrix}\dots & \vec{u_j} & \dots\end{pmatrix}$. Therefore, the above can be simplified to:

$$\begin{aligned}
&\operatorname{sgn}((rt_1 + st_2)\det\begin{pmatrix}\dots & \vec{u_j} & \dots\end{pmatrix})|\alpha(P(\dots,\dots))|\ ||(rt_1 + st_2)\vec{u_j}|| \\
&= \operatorname{sgn}(rt_1 + st_2)|rt_1 + st_2|\ (\operatorname{sgn}\det\begin{pmatrix}\dots & \vec{u_j} & \dots\end{pmatrix})|\alpha(P(\dots,\dots))|\ ||\vec{u_j}|| \text{ (Proposition 6.1.2)} \\
&= (rt_1 + st_2)(\operatorname{sgn}\det\begin{pmatrix}\dots & \vec{u_j} & \dots\end{pmatrix})|\alpha(P(\dots,\dots))|\ ||\vec{u_j}|| \\
&= r(\operatorname{sgn}\det\begin{pmatrix}\dots & t_1\vec{u_j} & \dots\end{pmatrix})|\alpha(P(\dots,\dots))|\ ||t_1\vec{u_j}|| + s(\operatorname{sgn}\det\begin{pmatrix}\dots & t_2\vec{u_j} & \dots\end{pmatrix})|\alpha(P(\dots,\dots))|\ ||t_2\vec{u_j}|| \\
&= rf(\begin{pmatrix}\dots & C_1 & \dots\end{pmatrix}) + sf(\begin{pmatrix}\dots & C_2 & \dots\end{pmatrix})
\end{aligned}$$

We have thus shown that f is identical to the determinant. Finally, as $\vec{v_i} = \sum_j \langle \vec{v_i}, \vec{u_j} \rangle$ (Proposition 6.2.1), $\det G(P) = f(G(P)) = \alpha(P)$.

The following proposition provides another geometric interpretation of the determinant: scaling factor of a linear transformation.

Proposition 7.4.5

Given an orthonormal reference basis B for and an endomorphism ϕ in V, for any $\vec{v_i} \in V$, $\alpha(P(\phi(\vec{v_1}), \phi(\vec{v_2}), \dots, \phi(\vec{v_n}))) = (\det M_B(\phi))\alpha(P(\vec{v_1}, \vec{v_2}, \dots, \vec{v_n}))$.

Proof

$$\begin{aligned}
\alpha(P(\phi(\vec{v_1}), \dots, \phi(\vec{v_n}))) &= \det G(P(\phi(\vec{v_1}), \dots, \phi(\vec{v_n})) \text{ (Gram determinant)} \\
&= \det\left(\gamma_B(\phi(\vec{v_1})), \dots, \gamma_B(\phi(\vec{v_n}))\right) \text{ (Proposition 6.2.1)} \\
&\because M_B(\phi)\gamma_B(\vec{v_i}) = \gamma_B \circ \phi \circ \gamma_B^{-1}(\gamma_B(\vec{v_i})) = \gamma_B(\phi(\vec{v_i})) \\
&= \det M_B(\phi)\left(\gamma_B(\vec{v_1}), \dots, \gamma_B(\vec{v_n})\right) \\
&= \det M_B(\phi) G(P(\vec{v_1}, \dots, \vec{v_n})) \text{ (Proposition 6.2.1)} \\
&= (\det M_B(\phi))\alpha(P(\vec{v_1}, \vec{v_2}, \dots, \vec{v_n})) \text{ (Proposition 7.2.4, Gram determinant)}
\end{aligned}$$

8 Spectral Theory

8.1 Eigenvalues and Eigenvectors

Suppose V is a vector space over k and $\phi : V \to V$ is an endomorphism. Our discussion in this section will revolve around the equation $\phi(\vec{v}) = \lambda\vec{v}$, where $\vec{v} \in V$ and $\lambda \in k$.

Definition 8.1.1: Eigenvalue

If there exists λ such that the equation is solvable for $\vec{v} \neq \vec{0}$, then λ is an eigenvalue of ϕ.

Definition 8.1.2: Eigenvector

For every eigenvalue λ of ϕ, the corresponding solutions of $\vec{v}$ (including $\vec{0}$) are eigenvectors belonging to λ.

Proposition 8.1.1: Eigenspace

The set of eigenvectors belonging to an eigenvalue λ, Λ, is a vector subspace of V. This is called the eigenspace belonging to λ.

Proof Suppose $\vec{v}$, $\vec{v_1}$ and $\vec{v_2}$ are eigenvectors belonging to λ. By the subspace test,

1. $\phi(\vec{v_1} + \vec{v_2}) = \phi(\vec{v_1}) + \phi(\vec{v_2}) = \lambda\vec{v_1} + \lambda(\vec{v_2}) = \lambda(\vec{v_1} + \vec{v_2}) \implies \vec{v_1} + \vec{v_2} \in \Lambda$
2. $\phi(r\vec{v}) = r\phi(\vec{v}) = r\lambda\vec{v} = \lambda(r\vec{v}) \implies r\vec{v} \in \Lambda$

Definition 8.1.3: Eigenbasis

If there exists a basis for V that consists only of eigenvectors of ϕ, it is an eigenbasis for V with respect to ϕ.

Theorem 8.1.1: Diagonalisable endomorphism

Suppose V is a finite dimensional vector space and $\phi : V \to V$ is an endomorphism. Then $B = \{\vec{v_1}, \vec{v_2}, ..., \vec{v_n}\}$ is an eigenbasis for V with respect to ϕ, with the basis vectors belonging to eigenvalues $\lambda_1, \lambda_2, ..., \lambda_n$ respectively $\iff M_B(\phi) = \begin{pmatrix} \lambda_1 & 0 & ... & 0 \\ 0 & \lambda_2 & ... & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & ... & \lambda_n \end{pmatrix}$.

Proof **(→)** We use the method described in Proposition 5.1.3 to obtain $M_B(\phi)$. Notice that $\gamma_B \circ \phi \circ \gamma_B^{-1}(\vec{e_j}) = \gamma_B \circ \phi(\vec{v_j}) = \gamma_B(\lambda_j\vec{v_j}) = \lambda\vec{e_j}$, which is exactly the matrix above.
(←) Consider that $\phi(\vec{v_i}) = \gamma_B^{-1}(M_B\gamma_B(\vec{v_i})) = \gamma_B^{-1}(M_B\vec{e_i}) = \gamma_B^{-1}(\lambda\vec{e_i}) = \lambda\vec{v_i}$. This means λ_i is an eigenvalue of ϕ and $\vec{v_i}$ is an eigenvector belonging to it. Hence, B is an eigenbasis for V with respect to ϕ.

Such endomorphisms are called diagonalisable.

Definition 8.1.4: Diagonablisable matrix

If $A \in M_n\ k$ is similar to a diagonal matrix, A is diagonalisable.

Proposition 8.1.2

Given any $A \in M_n\ k$, A is diagonalisable $\iff$ there exists an eigenbasis for k^n with respect to ϕ_A.

Proof (→) Let the diagonal matrix be $A' \in M_n\ k$. Then by Proposition 5.6.3, there exists bases $B = \{e_i\}$ and B' for k^n such that $A' = P_{B',B}AP_{B,B'}$. By the change of basis formula, $A' = P_{B',B}M_B(\phi_A)P_{B,B'} = M_{B'}(\phi_A)$. Hence, we know from Theorem 8.1.1 that as A' is a diagonal matrix, B' is an eigenbasis for k^n.
(←) Per Theorem 8.1.1, the existence of an eigenbasis B' implies $M_{B'}(\phi_A)$ is a diagonal matrix. Let $\mathcal{E}$ be the standard basis. We can then instantly construct a transition matrix $P = P_{B',\mathcal{E}}$ such that $A = M_{\mathcal{E}}(\phi_A) = P^{-1}M_{B'}(\phi_A)P$. (Change of basis formula, Proposition 5.6.1)

Definition 8.1.5: Characteristic polynomial

Let $A \in M_n\ k$. The characteristic polynomial of A is defined to be $p_A(x) = \det(xI_n - A)$.

The characteristic polynomial comes from an instinctive way of finding eigenvalues of a square matrix:

$$A\vec{v} = \lambda\vec{v} \iff A\vec{v} = \lambda I_n\vec{v} \iff (\lambda I_n - A)\vec{v} = \vec{0} \impliedby \det(\lambda I_n - A) = 0 \text{ (Matrix distributivity)}$$

However, note that the last step is not bidirectional. This means we have to further investigate if the roots of the polynomial are exhaustive of all eigenvalues of A.

Theorem 8.1.2

For any $A \in M_n\ k$, the eigenvalues of A (or ϕ_A) are the roots of the characteristic polynomial of A.

Proof We have shown that the roots of the polynomial must be eigenvalues of A.
To prove the converse, suppose there exists $x \in k$ and $\vec{v} \in k^n$ such that $\vec{v} \neq \vec{0}$ and $A\vec{v} = x\vec{v}$. Note that $(xI_n - A)\vec{v} = xI_n\vec{v} - A\vec{v} = x\vec{v} - x\vec{v} = \vec{0}$, which means the kernel of A is non-trivial. By the invertible matrix theorem, $\det(xI_n - A) = 0$.

Proposition 8.1.3

If $A, B \in M_n\ k$ are similar, their characteristic polynomials are identical.

Proof Suppose $B = P^{-1}AP$, where $P \in GL_n\ k$.
By matrix distributivity, $xI_n - P^{-1}AP = xP^{-1}P - P^{-1}AP = P^{-1}(xP - AP) = P^{-1}(xI_n - A)P$.
Hence, by Proposition 7.2.4, $\det(xI_n - B) = \det(P^{-1}(xI_n - A)P) = \det P^{-1}\det(xI_n - A)\det P = \det(xI_n - A)$. (Proposition 7.2.5)

The characteristic polynomial of a general endomorphism $\phi : V \to V$ thus makes sense as $M_B(\phi) \sim M_{B'}(\phi)$ for any basis B and B' for V. (Think the change of basis formula.) Hence, by the above, the characteristic polynomial of ϕ does not depend on the choice of basis.

Proposition 8.1.4

Suppose $\phi : V \to V$ is an endomorphism, where V is a finite dimensional vector space. Let $\lambda_1, \lambda_2, ..., \lambda_n$ be distinct eigenvalues of ϕ, corresponding to non-zero eigenvectors $\vec{v_1}, \vec{v_2}, ..., \vec{v_n}$. Then $\vec{v_1}, \vec{v_2}, ..., \vec{v_n}$ are linearly independent.

Proof We will prove the proposition by induction:
The case when $n = 1$ is trivial.
Assume that the proposition is true when $n = m - 1$. Let λ_m be another distinct eigenvalue of ϕ and $\vec{v_m}$ be a non-zero eigenvector belonging to λ_m. Suppose $\vec{v_1}, ..., \vec{v}_{m-1}, \vec{v_m}$ are linearly dependent. Then there exists a non-trivial $\{r_i\}$ so that: (Note that r_m is impossible to be 0.)

$$\vec{v_m} = \sum_{i=1}^{m-1} r_i \vec{v_i} \implies \phi(\vec{v_m}) = \sum_{i=1}^{m-1} r_i \phi(\vec{v_i}) \implies \lambda_m \vec{v_m} = \sum_{i=1}^{m-1} r_i \lambda_i \vec{v_i} \implies \vec{v_m} = \sum_{i=1}^{m-1} r_i \lambda_m^{-1} \lambda_i \vec{v_i}$$

As $\lambda_m \neq \lambda_i$, this is a different representation of $\vec{v_m}$. However, by assumption, $S = \vec{v_1}, ..., \vec{v}_{m-1}$ is a basis for $\operatorname{span} S$, which means according to the unique representation theorem, this should be impossible. Therefore, we have proved that the proposition is true when $n = m$ as well by contradiction.

Theorem 8.1.3

If the characteristic polynomial of $A \in M_n\, k$ has n distinct roots (in k), A is diagonalisable.

Proof Let the roots of the polynomial be $\lambda_1, \lambda_2, ..., \lambda_n$. By Theorem 8.1.2, they are eigenvalues of A. Suppose they correspond to non-zero eigenvectors $\vec{v_1}, \vec{v_2}, ..., \vec{v_n}$ respectively. Then according to Proposition 8.1.4, they are linearly independent. Furthermore, by the basis theorem, they also form a basis (and thus an eigenbasis) for k^n. Hence, per Proposition 8.1.2, A is diagonalisable.

If a square matrix A is diagonalisable, we can diagonalise it, which involves finding the sqaure matrix P such that $A' = P^{-1}AP$, where A' is a diagonal matrix. Diagonalisation is extremely useful in finding powers of square matrices.

To illustrate how diagonalisation could be done with the results above, suppose $A = \begin{pmatrix} -6 & 3 \\ 4 & 5 \end{pmatrix} \in M_2\, \mathbb{R}$.

The characteristic polynomial of A is $p_A(x) = \det(xI_2 - A) = \det \begin{pmatrix} x+6 & 3 \\ 4 & x-5 \end{pmatrix} = (x+6)(x-5) - 3 \cdot 4 = x^2 + x - 42$. As its roots are -7 and 6, Theorem 8.1.3 ensures that A is diagonalisable while Theorem 8.1.2 confirms that they are the eigenvalues of A.
Next, we have to find two non-zero eigenvectors $\vec{v_1}, \vec{v_2}$ associated with the two eigenvalues respectively. This can be done by plugging in the eigenvalues back to the equation $A\vec{v} = \lambda\vec{v}$, which would give us:

$$\begin{pmatrix} -6 & 3 \\ 4 & 5 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = -7 \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \implies \begin{cases} -6v_1 + 3v_2 & = -7v_1 \\ 4v_1 + 5v_2 & = -7v_2 \end{cases} \implies \begin{cases} v_1 + 3v_2 = 0 \\ 4v_1 + 12v_2 = 0 \end{cases} \tag{1}$$

$$\begin{pmatrix} -6 & 3 \\ 4 & 5 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = 6 \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \implies \begin{cases} -6v_1 + 3v_2 & = 6v_1 \\ 4v_1 + 5v_2 & = 6v_2 \end{cases} \implies \begin{cases} -12v_1 + 3v_2 = 0 \\ 4v_1 - v_2 = 0 \end{cases} \tag{2}$$

As expected, the two systems of linear equations are underdetermined (meaning there is more than one solution), because multiples of an eigenvector are eigenvectors too. Regardless, we will choose the simplest eigenvectors, i.e. $\vec{v_1} = \langle 3, -1 \rangle$ and $\vec{v_2} = \langle 1, 4 \rangle$, for the ease of calculation.
Note that $B' = \{\vec{v_1}, \vec{v_2}\}$ is an eigenbasis for k^2. Using the method described in Proposition 8.1.2, we fix $P = P_{\mathcal{E},B'} = M(\gamma_{\mathcal{E}} \circ \gamma'_B) = \begin{pmatrix} \frac{4}{13} & -\frac{1}{13} \\ \frac{1}{13} & \frac{3}{13} \end{pmatrix}$, where $\mathcal{E}$ is the standard basis. Then by Theorem 8.1.1, $A' = P^{-1}AP = P_{B',B} M_B(\phi_A) P_{B,B'} = M_{B'}(\phi_A) = \begin{pmatrix} -7 & 0 \\ 0 & 6 \end{pmatrix}$. To see how this result is useful, observe that:

$$(A')^n = \underbrace{(P^{-1}AP)(P^{-1}AP)...(P^{-1}AP)}_{n \text{ times}} = P^{-1}(A)^n P \implies (A)^n = P \begin{pmatrix} (-7)^n & 0 \\ 0 & 6^n \end{pmatrix} P^{-1}$$

We have thus derived a general formula for the powers of A.

8.2 Hermitian Transformations

Inner product spaces in the rest of this chapter default to complex inner product spaces. Results in this section hence apply to real inner product spaces as well.

Definition 8.2.1: Adjoint endomorphism and matrix

Given an endomorphism $\phi : V \to V$, where V is an inner product space, another endomorphism ϕ^H on V is an adjoint for ϕ if for any $\vec{v}$ and $\vec{w}$ in V, $\langle \phi(\vec{v}), \vec{w} \rangle = \langle \vec{v}, \phi^H(\vec{w}) \rangle$.
Similarly, given $A, A^H \in M_n\ \mathbb{C}$, A^H is an adjoint for A if for $\vec{v}, \vec{w} \in \mathbb{C}^n$, $\langle A\vec{v}, \vec{w} \rangle = \langle \vec{v}, A^H\vec{w} \rangle$.

Proposition 8.2.1

The adjoint for an endomorphism ϕ in a finite dimensional inner product space exists and is unique.

Proof For any $\vec{w} \in V$, define the linear functional $l'_{\vec{w}} \in V^*$ by $l'_{\vec{w}}(\vec{v}) = \langle \phi(\vec{v}), \vec{w} \rangle$. By the Riesz representation theorem, there is a unique $\vec{w}^* \in V$ such that $l_{\vec{w}^*} = l'_{\vec{w}}$. Now define the map $\phi^H : V \to V$ by $\phi^H(\vec{w}) = \vec{w}^*$ in the above context.
The only task left is to prove the linearity of the map. For any $\vec{v}, \vec{w_1}, \vec{w_2} \in V$ and $r_1, r_2 \in \mathbb{C}$:

$$\langle \phi(\vec{v}), r_1\vec{w_1} + r_2\vec{w_2} \rangle = \overline{r_1}\langle \phi(\vec{v}), \vec{w_1} \rangle + \overline{r_2}\langle \vec{v}, \vec{w_2} \rangle = \langle \vec{v}, r_1\phi^H(\vec{w_1}) \rangle + \langle \vec{v}, r_2\phi^H(\vec{w_2}) \rangle$$

$$\implies \phi^H(r_1\vec{w_1} + r_2\vec{w_2}) = r_1\phi^H(\vec{w_1}) + r_2\phi^H(\vec{w_2})$$

The above process can be rephrased as $\phi^H = l^{-1} \circ \phi^* \circ l$.

Proposition 8.2.2

The adjoint for any complex square matrix $A \in M_n\ \mathbb{C}$ exists, is unique and $A^H = M(\phi_A^H)$. Moreover, if the inner product is defined as the dot product, $A^H = \overline{A^T}$.

Proof Let $\vec{v}$ and $\vec{w}$ be any two vectors in $\mathbb{C}^n$.

$$\langle A\vec{v}, \vec{w} \rangle = \langle \vec{v}, A^H\vec{w} \rangle \implies \langle \phi_A(\vec{v}), \vec{w} \rangle = \langle \vec{v}, \phi_{A^H}(\vec{w}) \rangle \implies \phi_A^H = \phi_{A^H} \text{ (Proposition 8.2.1)}$$

$$\iff A^H = M(\phi_A^H) \text{ (Theorem 5.1.3)}$$

Define the conjugate map $c : \mathbb{C}^n \to \mathbb{C}^n$ by $c(\langle r_1, ..., r_n \rangle) = \langle \overline{r_1}, ..., \overline{r_n} \rangle$, which is self-inverse. Note that in a dot product space, $l \circ c$ is just the dual operator $*$. Hence, if $\mathcal{E}$ is the standard basis of $\mathbb{C}^n$:

$$\begin{aligned} & \phi_A^H = l^{-1} \circ \phi^* \circ l \\ \implies & c \circ \phi_A^H \circ c = *^{-1} \circ \phi^* \circ * \\ \implies & M_{\mathcal{E}}(c \circ \phi_A^H \circ c) = M_{B^*,\mathcal{E}}(*^{-1})M_{B^*}(\phi_A^*)M_{\mathcal{E},B^*}(*) \\ \implies & \overline{A^H} = A^T \text{ (Theorem 5.4.1)} \end{aligned}$$

Therefore, a real square matrix in a dot product space is Hermitian if and only if it is symmetric.

Proposition 8.2.3

For any endomorphism ϕ on an inner product space V with an orthonormal basis B, if $M = M_B(\phi)$, then $\phi^H = \gamma_B^{-1} \circ \phi_{\overline{M^T}} \circ \gamma_B$.

Proof Recall that $\gamma_B : V \to \mathbb{C}^n$ and $\gamma_B^{-1} : \mathbb{C}^n \to V$ are isometries (Theorem 6.4.2), where $\mathbb{C}^n$ is equipped with the dot product. This means they preserve the inner product (Proposition 6.4.2). By Proposition 8.2.2, $M^H = \overline{M^T}$ in $\mathbb{C}^n$.

This proposition means that any property related to the adjoint of an endomorphism must also apply to its matrix (with respect to an orthonormal basis), and vice versa. For example, $\phi^H = \phi$ if and only if $M = \overline{M^T}$.

Definition 8.2.2: Hermitian transformation and matrix

An endormophism ϕ on an inner product space V is Hermitian if it is self-adjoint, i.e. $\phi^H = \phi$. Similarly, a complex square matrix A is Hermitian if it is self-adjoint, i.e. $A^H = A$.

Proposition 8.2.4

All eigenvalues of a Hermitian endomorphism are real.

Proof Let λ be an eigenvalue of the endomorphism ϕ, and $\vec{v}$ be a non-zero eigenvector belonging to λ. Then $\lambda\langle\vec{v},\vec{v}\rangle = \langle\lambda\vec{v},\vec{v}\rangle = \langle\phi(\vec{v}),\vec{v}\rangle = \langle\vec{v},\phi(\vec{v})\rangle = \langle\vec{v},\lambda\vec{v}\rangle = \overline{\lambda}\langle\vec{v},\vec{v}\rangle$. As $\langle\vec{v},\vec{v}\rangle \neq 0$, $\lambda = \overline{\lambda}$, which implies λ is real.

8.3 Unitary Transformations

Definition 8.3.1: Unitary transformation and matrix

An invertible (or bijective) endomorphism ϕ on an inner product space is unitary if $\phi^H = \phi^{-1}$. Similarly, an invertible complex matrix A is unitary if $A^H = A^{-1}$.

Observe that if ϕ is unitary, $\langle\phi(\vec{v}),\phi(\vec{w})\rangle = \langle\vec{v},\phi^{-1}(\phi(\vec{w}))\rangle = \langle\vec{v},\vec{w}\rangle$. Hence, unitary endomorphisms are length- and angle-preversing transformations. The converse is also true, in fact, a weaker condition suffices, which will be introduced shortly below.

A notable example of unitary endomorphisms is rotation transformations in $\mathbb{R}^n$ (with the inner product defined as dot product), which was introduced in Section 5.1. Clearly, it preverses length and angle, which means it must be unitary. To verify this, recall that the dot product of two vectors $\vec{v}$ and $\vec{w}$ is the product of the magnitudes of $\vec{v}$ and the projection of $\vec{w}$ on $\vec{v}$. The magnitude of the projection would be the same if we:

1. Rotate $\vec{w}$ by an angle θ in the anticlockwise direction; or
2. Rotate $\vec{v}$ by the same angle θ in the clockwise direction.

This fact directly translates to $\phi(\vec{v}) \cdot \vec{w} = \vec{v} \cdot \phi^{-1}(\vec{w})$, which is exactly the definition of a unitary endomorphism.

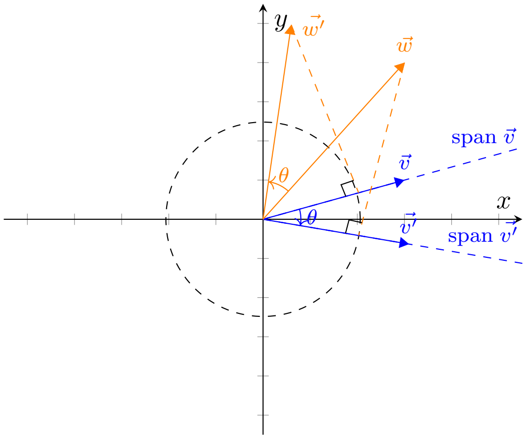

Figure 8.3.1: Geometric demonstration of the unitariness of rotation transformations

Proposition 8.3.1

Suppose ϕ is an endomorphism of an inner product space V. Then ϕ is unitary if and only if $\langle\phi(\vec{v}),\phi(\vec{v})\rangle = \langle\vec{v},\vec{v}\rangle$. In other words, ϕ is unitary if and only if ϕ is an isometry.

Proof By Proposition 6.4.2, ϕ is injective and preserves the inner product. As ϕ is an endomorphism, by the rank-nullity theorem, ϕ is also invertible. Therefore, by the above, ϕ is a unitary transformation.
The proof of the converse is trivial and has been presented above.

A geometric interpretation of this proposition revolves around congruent triangles. For any two vectors $\vec{v}$ and $\vec{w}$ in an inner product space, we construct a triangle A bounded by the vectors. Now consider the triangle A' bounded by $\phi(\vec{v})$ and $\phi(\vec{w})$, where ϕ is a length-preserving transformation. As $||\vec{v}|| = ||\phi(\vec{v})||$, $||\vec{w}|| = ||\phi(\vec{w})||$ and $||\vec{v}+\vec{w}|| = ||\phi(\vec{v}+\vec{w})|| = ||\phi(\vec{v})+\phi(\vec{w})||$, by SSS (side-side-side), $A \cong A'$ and thus the angle between the two vectors remains unchanged before and after the transformation.

Proposition 8.3.2

For any eigenvalue λ of a unitary transformation ϕ, $|\lambda| = 1$.

Proof Suppose λ is an eigenvalue of ϕ and $\vec{v}$ is a corresponding non-zero eigenvector. By Proposition 8.3.1, $\langle\vec{v},\vec{v}\rangle = \langle\phi(\vec{v}),\phi(\vec{v})\rangle = \langle\lambda\vec{v},\lambda\vec{v}\rangle = \lambda\overline{\lambda}\langle\vec{v},\vec{v}\rangle$. This implies $\lambda\overline{\lambda} = |\lambda|^2 = 1$ and hence $|\lambda| = 1$.

8.4 Normal Transformations

Definition 8.4.1: Normal transformation and matrix

An endomorphism ϕ on an inner product space is normal if it commutes with its adjoint, i.e. if $\phi \circ \phi^H = \phi^H \circ \phi$.
Similarly, a matrix $A \in M_n\,\mathbb{C}$ is normal if $AA^H = A^H A$.

Examples of normal transformations:

1. Hermitian transformations
2. Unitary transformations
3. Skew-Hermitian transformations: $\phi^H = -\phi$

Proposition 8.4.1

If ϕ is a normal transformation, then for any $\vec{v} \in V$, $||\phi(\vec{v})|| = ||\phi^H(\vec{v})||$

Proof $\langle\phi(\vec{v}),\phi(\vec{v})\rangle = \langle\vec{v},\phi^H\circ\phi(\vec{v})\rangle = \langle\vec{v},\phi\circ\phi^H(\vec{v})\rangle = \langle\phi^H(\vec{v}),\phi^H(\vec{v})\rangle$

Proposition 8.4.2

Suppose ϕ is a normal transformation. Then $\phi(\vec{v}) = \lambda\vec{v} \implies \phi^H(\vec{v}) = \overline{\lambda}\vec{v}$.

Proof $0 = ||(\phi - \lambda\,\mathrm{id}_V)(\vec{v})|| = ||(\phi - \lambda\,\mathrm{id}_V)^H(\vec{v})||$ (Proposition 8.4.1)
$= ||(\phi^H - \overline{\lambda}\,\mathrm{id}_V)(\vec{v})||$ (Proposition 8.2.3)
By positive definiteness, $(\phi^H - \overline{\lambda}\,\mathrm{id}_V)(\vec{v}) = 0 \implies \phi^H(\vec{v}) = \overline{\lambda}\vec{v}$.

Proposition 8.4.3

Suppose ϕ is a normal transformation. Then if λ_1, λ_2 are two distinct eigenvalues of ϕ, and $\vec{v_1}, \vec{v_2}$ are eigenvectors belonging to them respectively, then $\vec{v_1} \perp \vec{v_2}$.

Proof By Proposition 8.4.2, $(\lambda_1 - \lambda_2)\langle \vec{v_1}, \vec{v_2} \rangle = \langle \phi(\vec{v_1}), \vec{v_2} \rangle - \langle \vec{v_1}, \phi^H(\vec{v_2}) \rangle = 0 \implies \langle \vec{v_1}, \vec{v_2} \rangle = 0$.

Proposition 8.4.4

Suppose ϕ is a normal transformation, and $\vec{v}$ is an eigenvector of ϕ. Then if $\vec{w} \perp \vec{v}$, $\phi(\vec{w}) \perp \vec{v}$.

Proof By Proposition 8.4.2, if $\phi(\vec{v}) = \lambda\vec{v}$, $\langle \phi(\vec{w}), \vec{v} \rangle = \langle \vec{w}, \phi^H(\vec{v}) \rangle = \langle \vec{w}, \overline{\lambda}\vec{v} \rangle = 0$.

8.5 Spectral Decomposition

Definition 8.5.1: Invariant subspace

Suppose V is an inner product space and W is its vector subspace. If ϕ is an endomorphism of V and $\phi(W) \subseteq W$, then W is ϕ-invariant.

Naturally, we can extract a restricted endomorphism $\phi|_W : W \to W$ from ϕ with the following trivial properties:

1. Eigenvectors and eigenvalues of $\phi|_W$ are eigenvectors and eigenvalues of ϕ.
2. ϕ is Hermitian/unitary/normal implies $\phi|_W$ is Hermitian/unitary/normal.
3. If there exists an eigenvalue with respect to ϕ, its corresponding eigenspace is ϕ-invariant.

Theorem 8.5.1: Spectral decomposition theorem

Suppose V is a finite dimensional vector space and ϕ is an endomorphism of V. Then if

1. V is a real or complex inner product space and ϕ is Hermitian, or
2. V is a complex inner product space and ϕ is normal,

there exists an orthonormal eigenbasis for V with respect to ϕ. In other words, V can be unitarily diagonalised with respect to ϕ.

Proof Recall how we can obtain a characteristic polynomial for a linear transformation. By the fundamental theorem of algebra, an endomorphism of a finite-dimensional complex vector space (except the trivial vector space) thus must have at least one eigenvalue. (Theorem 8.1.2) This guarantees the existence of an eigenvalue for ϕ in (2). Furthermore, by Proposition 8.2.4, an eigenvalue must exist for ϕ in (1) as well.

Suppose $\dim V = n$. If $n = 1$, the statement is trivially true. Otherwise, select a unit eigenvector $\vec{v_1}$ belonging to the guaranteed eigenvalue λ_1. Let $V_1 = \operatorname{span} \vec{v_1}$. By Proposition 6.2.5, $V_1^\perp$ is also a subspace of V and $V = V_1 \oplus V_1^\perp$. Additionally, Proposition 8.4.4 implies that $V_1^\perp$ is ϕ-invariant. By the dimension formula, $\dim V_1^\perp = n - 1$. Hence, by the properties above, there exists a unit eigenvector $\vec{v_2}$ of $\phi|_{V_1^\perp}$ and hence of ϕ. By definition, $\vec{v_2} \perp \vec{v_1}$.

We can repeat the above process $n - 1$ times (until the orthogonal complement becomes the trivial vector space). Doing so gives us an orthonormal set $\{\vec{v_1}, \vec{v_2}, ..., \vec{v_n}\}$. As $V = V_1 \oplus V_2 \oplus ... \oplus V_n$, the set is an orthonormal eigenbasis for V (with respect to ϕ).

9 Triangulation and Decomposition of Endomorphisms

9.1 Cayley-Hamilton Theorem

This entire section will be dedicated to proving the Cayley-Hamilton Theorem. To this end, it is useful to revisit an algebraic structure we have touched on before: k-algebra.

For any field k, $k[x]$ is an algebra over k under polynomial addition and multiplication. (The proof is left as an exercise to the readers.) Another example of a k-algebra would be the vector space of endomorphisms under addition and composition, which have been discussed in Theorem 5.1.2.

A natural way to associate the two algebras would be through evaluation. If we define powers of an endomorphism ϕ such that $\phi^n = \underbrace{\phi \circ ... \circ \phi}_{n \text{ times}}$ (ϕ^0 is defined as the identity map), then we can evaluate a polynomial at an endomorphism. For example, the polynomial $x^2 + x + 1$ evaluated at an endomorphism $\phi : V \to V$ would give another endomorphism $\phi \circ \phi + \phi + \mathrm{id}_V$.

Such a map from $k[x]$ to $\hom(V, V)$ is also an homomorphism: Given two polynomials $p = \sum r_i x^i$ and $q = \sum s_i x^i$ in $k[x]$ and an endomorphism $\phi : V \to V$, observe that:

$$p(\phi) \circ q(\phi) = \sum_i r_i \phi^i \circ \left(\sum_j s_j \phi^j \right) = \sum_{i,j} r_i s_j \phi^i \circ \phi^j = \sum_{i,j} r_i s_j \phi^{i+j} = pq(\phi)$$

Similarly, we can define an evaluation map from $k[x]$ to $M_n\ k$, with powers of a square matrix defined in the usual way (self-multiplication). By Proposition 5.1.4, as composition of linear transformations corresponds to matrix multiplication, all of the above still hold for this map.

Proposition 9.1.1

Suppose V is a n-dimensional vector space over k, where $1 \leq n < \infty$, and ϕ is an endomorphism of V with respect to which the only invariant subspaces are the trivial vector space and V. Then there exists a basis B for V such that $M_B(\phi) = \begin{pmatrix} 0 & 0 & ... & 0 & r_1 \\ 1 & 0 & ... & 0 & r_2 \\ 0 & 1 & ... & 0 & r_3 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & ... & 1 & r_n \end{pmatrix}$, where $r_i \in k$.

Proof Recall how we construct $M_B(\phi) = M(\gamma_B \circ \phi \circ \gamma_B^{-1})$ in Proposition 5.1.3. Suppose $B = \{\vec{u_1}, \vec{u_2}, ..., \vec{u_n}\}$. Then the matrix of linear transformation would take such a form if and only if $\vec{u}_{i+1} = \phi(\vec{u_i})$ for $1 \leq i \leq n-1$ and $\phi(\vec{u_n}) = r_1\vec{u_1} + r_2\vec{u_2} + ... + r_n\vec{u_n}$.

We try to construct such a B from the empty set by adding a suitable vector to it one at a time. To start with, choose any non-zero vector in V as $\vec{u_1}$. Expand the set by adding $\phi(\vec{u_m})$, where $m = |B|$, until we get n elements.

To show that B is linearly independent and thus a basis (by the basis theorem), we first assume the opposite: there exists a non-trivial $\{s_i\} \subseteq k$ such that $s_1\vec{u_1} + s_2\vec{u_2} + ... + s_n\vec{u_n} = 0$. This implies $\phi(s_2\vec{u_1} + s_3\vec{u_2} + ... + s_n\vec{u}_{n-1}) = -s_1\vec{u_1}$. Let $S = \mathrm{span}\{\vec{u_1}, \vec{u_2}, ..., \vec{u_m}\}$ such that $1 \leq m \leq n-1$, $s_{m+1} \neq 0$ and $s_{m+2}, ..., s_n = 0$. For every vector $\vec{v}$ in S, $\vec{v} = t_1\vec{u_1} + t_2\vec{u_2} + ... + t_m\vec{u_m}$ for some $t_i \in k$.

$$\begin{aligned} \text{Then } \phi\left(\sum_{i=1}^{m} s_{i+1}\vec{u_i}\right) = -s_1\vec{u_1} &\implies \phi\left(t_m s_{m+1}^{-1} \sum_{i=1}^{m} s_{i+1}\vec{u_i}\right) = -t_m s_{m+1}^{-1} s_1 \vec{u_1} \\ &\implies \phi\left(\sum_{i=1}^{m} t_i\vec{u_i}\right) = -t_m s_{m+1}^{-1} s_1\vec{u_1} + \sum_{i=1}^{m-1}(t_i - t_m s_{m+1}^{-1} s_{i+1})\phi(\vec{u_i}) \\ &\implies \phi(\vec{v}) \in S \ (\because \phi(\vec{u_i}) = \vec{u}_{i+1}) \end{aligned}$$

In other words, as S is a subspace of V (Proposition 3.3.1), ϕ is S-invariant. Moreover, by Proposition 3.8.3, $S \neq V$, which means our assumption has been violated.

Proposition 9.1.2

In continuation of Proposition 9.1.1, the characteristic polynomial $p_\phi(x) = x^n - r_n x^{n-1} - ... - r_2 x - r_1$.

Proof This can be proved by induction. The case when $n = 1$ is trivial. Suppose the statement is true when $\dim V = n - 1$. Then:

$$p_\phi(x) = \det(xI_n - M_B(\phi))$$

$$= \det \begin{pmatrix} x & 0 & ... & 0 & -r_1 \\ -1 & x & ... & 0 & -r_2 \\ 0 & -1 & ... & 0 & -r_3 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & ... & -1 & x - r_n \end{pmatrix}$$

By expanding by the first row,

$$= x \det \begin{pmatrix} x & 0 & ... & 0 & -r_2 \\ -1 & x & ... & 0 & -r_3 \\ 0 & -1 & ... & 0 & -r_4 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & ... & -1 & x - r_n \end{pmatrix} + (-1)^{n+1}(-r_1) \det \begin{pmatrix} -1 & 0 & ... & 0 & 0 \\ x & -1 & ... & 0 & 0 \\ 0 & x & ... & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & ... & x & -1 \end{pmatrix}$$

By assumption and Proposition 7.2.2,

$$= x(x^{n-1} - r_n x^{n-2} - ... - r_3 x - r_2) + (-1)^{n+1}(-r_1)(-1)^{n-1}$$

$$= x^n - r_n x^{n-1} - ... - r_3 x^2 - r_2 x - r_1$$

This means the statement is also true when $\dim V = n$.

Proposition 9.1.3

In continuation of Proposition 9.1.2, ϕ satisfies its characteristic polynomial, i.e. $p_\phi(\phi) = O$, where O is the zero map.

Proof Observe that

$$\begin{aligned} p_\phi(\phi)(\vec{u_1}) &= (\phi^n - r_n \phi^{n-1} - ... - r_2\phi - r_1 \operatorname{id}_V)(\vec{u_1}) \\ &= \phi^n(\vec{u_1}) - r_n\phi^{n-1}(\vec{u_1}) - ... - r_2\phi(\vec{u_1}) - r_1\vec{u_1} \\ &= (r_1\vec{u_1} + r_2\vec{u_2} + ... + r_n\vec{u_n}) - r_n\vec{u_n} - ... - r_2\vec{u_2} - r_1\vec{u_1} = \vec{0} \end{aligned}$$

By linearity, $p_\phi(\phi)(\phi(\vec{u_i})) = \phi(p_\phi(\phi)(\vec{u_i}))$. Hence, by repeated use of this equality, $p_\phi(\phi)(\vec{u_1}) = p_\phi(\phi)(\vec{u_2}) = ... = p_\phi(\phi)(\vec{u_n}) = \vec{0}$. Furthermore, for any $\vec{v} = \sum s_i\vec{u_i} \in V$, where $s_i \in k$, $p_\phi(\phi)(\vec{v}) = p_\phi(\phi)(\sum s_i\vec{u_i}) = s_i \sum p_\phi(\phi)(\vec{u_i}) = \vec{0}$. Therefore, $p_\phi(\phi)$ must be the zero map.

Theorem 9.1.1: Cayley-Hamilton theorem

Suppose V is a n-dimensional vector space over k, where $1 \leq n < \infty$, and ϕ is an endomorphism of V. Then ϕ satisfies its characteristic polynomial, i.e. $p_\phi(\phi) = O$, where O is the zero map.

Proof We will prove this by induction:
When $n = 1$, the only subspaces of V are the trivial vector space and V itself. Therefore, by Proposition 9.1.3, the theorem must be true.

Suppose the theorem is true for all m-dimensional vector spaces, where $m < n$. Find a non-trivial, proper subspace m-dimensional W_1 of V which is ϕ-invariant. If this is impossible, the theorem can already be verified by Proposition 9.1.3. Otherwise, find a basis B_1 for W_1. By Theorem 3.8.1, there exists B_2 such that $B = B_1 \cup B_2$ is a basis for V. Further define W_2 to be the span of B_2. By Proposition 3.3.1, it is also a (non-trivial, proper) subspace of V. It should be evident that $V = W_1 \oplus W_2$.

Consider the matrix of ϕ with respect to B. Note that while W_1 is ϕ-invariant, this is not necessarily true for W_2. Thus, $M_B(\phi)$ should have the following form:

$$\left(\begin{array}{c:c} P & Q \\ \hdashline \mathbf{0} & R \end{array}\right) \tag{1}$$

, where $P \in M_m\ k$, $R \in M_{n-m}\ k$. Furthermore, if R is the matrix of $\phi_R : W_2 \to W_2$ with respect to B_2, by Proposition 7.2.3:

$$p_\phi(x) = \det(xI_n - M_B(\phi)) = \det \left(\begin{array}{c:c} xI_m - P & Q \\ \hdashline \mathbf{0} & xI_{n-m} - R \end{array}\right) = p_{\phi|_{W_1}}(x)p_{\phi_R}(x) \tag{2}$$

For $\vec{w} \in W_1$, the theorem is hence easy to prove. By assumption, $p_{\phi|_{W_1}}(\phi)(\vec{w}) = p_{\phi|_{W_1}}(\phi|_{W_1})(\vec{w}) = \vec{0}$. By (2), $p_\phi(\phi) = p_{\phi_R} p_{\phi|_{W_1}}(\phi) = p_{\phi_R}(\phi) \circ p_{\phi|_{W_1}}(\phi)(\vec{w}) = \vec{0}$.

Now we investigate the relationship between ϕ and ϕ_R. We can deduce by (1) that for any basis vector $\vec{u} \in B_2$, $\phi(\vec{u})$ and $\phi_R(\vec{u})$, when expressed in terms of B, have the same weights with respect to B_2. This implies that for any $\vec{w} \in W_2$:

$$\begin{aligned}
& \phi(\vec{w}) - \phi_R(\vec{w}) \in W_1 && (3)\\
\implies & \phi^2(\vec{w}) - \phi \circ \phi_R(\vec{w}) \in W_1 \text{ } (\because W_1 \text{ is } \phi\text{-invariant}) \\
\implies & \phi^2(\vec{w}) - \phi_R^2(\vec{w}) \in W_1 \text{ (By (3))} \\
\implies & \phi^i(\vec{w}) - \phi_R^i(\vec{w}) \in W_1 \ \forall i \in \mathbb{N} \text{ } (\because W_1 \text{ is } \phi\text{-invariant}) \\
\implies & q(\phi)(\vec{w}) - q(\phi_R)(\vec{w}) \in W_1 \ \forall q \in k[x] \\
\implies & p_\phi(\phi)(\vec{w}) = p_{\phi|_{W_1}}(\phi) \circ p_{\phi_R}(\phi)(\vec{w}) \text{ (By (2))} \\
& \qquad = p_{\phi|_{W_1}}(\phi)(p_{\phi_R}(\phi_R)(\vec{w}) + \vec{w_1}) \text{ for some } \vec{w_1} \in W_1 \\
& \qquad = p_{\phi|_{W_1}}(\phi)(\vec{0} + \vec{w_1}) \text{ (By assumption)} \\
& \qquad = \vec{0} \text{ (By assumption)}
\end{aligned}$$

As any vector $\vec{v}$ in V can be expressed as the sum of two vectors from W_1 and W_2 respectively, $p_\phi(\phi)(\vec{v}) = \vec{0}$. In other words, $p_\phi(\phi)$ is the zero map.

By Proposition 5.1.4, this theorem can easily be applied to square matrices: For any $A \in M_n\ k$, $p_A(A) = \mathbf{0}$, where $\mathbf{0}$ is the $n \times n$ zero matrix.

9.2 Minimal Polynomials

Definition 9.2.1: Minimal Polynomial

Let V be a vector space over k and ϕ be an endomorphism of V. The minimal polynomial $m_\phi \in k[x]$ of ϕ is the smallest-degree monic polynomial such that $m_\phi(\phi) = O$.

Monic polynomials are polynomials with leading coefficients of 1. Such a requirement on minimal polynomials excludes the possibility of them being zero polynomials.

Proposition 9.2.1

For any endomorphism ϕ of a finitely dimensional vector space, m_ϕ exists and is unique.

Proof *Existence* Guaranteed by the Cayley-Hamilton theorem and the well-ordering principle. *Uniqueness* Suppose there exist distinct minimal polynomials m and m' of ϕ with the same degree such that $m(\phi) = m'(\phi) = O$. Then $(m - m')(\phi) = O$, but $m - m' \neq 0$ must be of a lower degree, which contradicts the minimality of m and m'.

Proposition 9.2.2

For any endomorphism ϕ of a vector space V over k, if $p \in k[x]$ and $p(\phi) = O$, p is divisible by m_ϕ.

Proof By polynomial division, we get $p(x) = m_\phi(x)q(x) + r(x)$, where $q, r \in k[x]$ and the degree of r is smaller than m_ϕ. Note that $p(\phi) = O, m_\phi(\phi) = O \implies r(\phi) = O$. By Proposition 9.2.1, the minimal polynomial is unique, so $r = 0$.

A corollary of this proposition is that the minimal polynomial always divides the characteristic polynomial by the Cayley-Hamilton theorem.

Proposition 9.2.3

For any endomorphism ϕ of a vector space, the set of roots of m_ϕ is identical to the set of eigenvalues of ϕ.

Proof By Proposition 9.2.2, factors of m_ϕ must be factors of p_ϕ. Conversely, for any eigenvalue λ of ϕ, let $\vec{v}$ be a non-zero eigenvector of λ. Then:

$$m_\phi(\phi)(\vec{v}) = m_\phi(\lambda)(\vec{v}) = \vec{0} \implies m_\phi(\lambda) = 0$$

Theorem 9.2.1: Primary decomposition theorem

For any endomorphism ϕ of a vector space V over k and $m \in k[x]$, if $m(\phi) = O$ and m can be factorised as $m(x) = p(x)q(x)$, where $p, q \in k[x]$ are co-prime, then $V = \ker p(\phi) \oplus \ker q(\phi)$.

Proof Recall that:

$$\begin{aligned} pq(\phi) = p(\phi) \circ q(\phi) = O &\implies \operatorname{im} q(\phi) \subseteq \ker p(\phi) \\ = qp(\phi) = q(\phi) \circ p(\phi) = O &\implies \operatorname{im} p(\phi) \subseteq \ker q(\phi) \end{aligned}$$

By Bézout's identity, there exists $r, s \in k$ such that $rp(\phi) + sq(\phi) = \mathrm{id}_V$. Then for any $\vec{v} \in V$, $\vec{v} = sq(\phi)(\vec{v}) + rp(\phi)(\vec{v}) \implies V = \ker p(\phi) + \ker q(\phi)$.
Suppose $\vec{v} \in \ker p(\phi) \cup \ker q(\phi)$. Then by the above, $\vec{v} = rp(\phi)(\vec{v}) + sq(\phi)(\vec{v}) = \vec{0}$. By Proposition 3.5.4, the sum is a direct sum.

Proposition 9.2.4

An endomorphism of a vector space is diagonalisable if and only if its minimal polynomial can be factorised into distinct linear factors.

Proof ($\rightarrow$) Let the (distinct) eigenvalues of the endomorphism ϕ be $\lambda_1, \lambda_2, ..., \lambda_m$. By Proposition 9.2.3, $(x - \lambda_i)$ must be a factor of m_ϕ for $i \in \{1, ..., m\}$. By Theorem 8.1.1, there is an eigenbasis $B = \{\vec{v_1}, \vec{v_2}, ..., \vec{v_n}\}$. Let $\pi : \{1, ..., n\} \to \{1, ..., m\}$ be a map such that $\vec{v_i}$ belongs to $\lambda_{\pi(i)}$. Then if

$m = (x-\lambda_1)...(x-\lambda_m)$, for $\vec{v} = \sum r_i\vec{v_i} \in V$, where V is the vector space:

$$m(\phi)(\vec{v}) = m(\phi)\left(\sum_{i=1}^{n} r_i\vec{v_i}\right) = \sum_{i=1}^{n} r_i m(\lambda_{\pi(i)})\vec{v} = \vec{0}$$

Therefore, m is already the minimal polynomial of ϕ, which is the product of distinct linear factors.
(←) Let $m_\phi = (x-\lambda_1)(x-\lambda_2)...(x-\lambda_m)$, where $\lambda_1, \lambda_2, ..., \lambda_m$ are distinct. By the primary decomposition theorem, $V = \ker(\phi-\lambda_1) \oplus ... \oplus \ker(\phi-\lambda_m)$. We collect the bases of the subspaces into $B = \vec{v_1}, \vec{v_2}, ..., \vec{v_n}$. By the unique representation theorem, B is a basis for V. Note that it is also an eigenbasis because if $\vec{v_i} \in \ker(\phi-\lambda_j)$, $(\phi-\lambda_j)\vec{v_i} = \vec{0} \implies \phi(\vec{v_i}) = \lambda_j\vec{v_i}$. By Theorem 8.1.1, $M_B(\phi)$ is a diagonal matrix.

As always, there is a matrix version of this result: Any square matrix is diagonalisable if and only if its minimal polynomial can be factorised into distinct linear factors.

Proposition 9.2.5

If ϕ is an endomorphism of a vector space V and U is an invariant subspace of V with respect to ϕ. Then if ϕ is diagonalisable, $\phi|_U$ is also diagonalisable.

Proof As $m_\phi(\phi|_U) = O$, by Proposition 9.2.2, $m_{\phi|_U}$ is a factor of m_ϕ. By Proposition 9.2.4, m_ϕ and thus $m_{\phi|_U}$ are the products of distinct linear factors. By the same proposition, $\phi|_U$ is diagonalisable.

Theorem 9.2.2: Spectral theorem for commuting operators

Suppose ϕ and ψ are diagonalisable endomorphisms of a vector space V, Then ϕ and ψ commute, i.e. $\phi \circ \psi = \psi \circ \phi$, if and only if ϕ and ψ are simultaneously diagonalisable, i.e. iff there exists a basis B for V such that $M_B(\phi)$ and $M_B(\psi)$ are both diagonal.

Proof (→) Let $\vec{v}$ be an eigenvector of eigenvalue λ with respect to ϕ. Consider that:

$$\phi \circ \psi(\vec{v}) = \psi \circ \phi(\vec{v}) = \lambda\psi(\vec{v}) \implies \psi(\vec{v}) \text{ is an eigenvalue of } \lambda \text{ with respect to } \psi$$

As ϕ is diagonalisable, we can partition V into $V = \Lambda_1 \oplus \Lambda_2 \oplus ... \oplus \Lambda_n$, where Λ_i is an eigenspace of V with respect to ϕ. By the above, Λ_i is invariant with respect to ψ for any $i \in \{1, ..., n\}$, so Proposition 9.2.5 applies. We can thus obtain an eigenbasis for each of the subspace with respect to both ϕ and ψ. By the unique representation theorem, the union of these bases B is a basis for V. Proposition 9.2.4 then tells us $M_B(\phi)$ and $M_B(\psi)$ are both diagonal.
The converse is trivial to show.

A special version of the theorem exists: If in addition, we require the two operators to be normal, the operators are simultaneously unitarily diagonalisable. This can be shown by slightly modifying the above proof. More specifically, an orthonormal basis should now be selected for each Λ_i. By Proposition 8.4.3, such bases are also orthogonal to one another. Hence, their union is our desired basis.

Again, a matrix version of the (two) theorems can be derived:

1. If $A, B \in M_n\ k$ are two digonalisable matrices, and A and B commute, then there exists an (invertible) matrix $P \in GL_n\ k$ such that $P^{-1}AP$ and $P^{-1}BP$ are both diagonal.
2. If $A, B \in M_n\ k$ are two normal matrices, and A and B commute, then there exists an unitary matrix $U \in GL_n\ k$ such that $U^H AU$ and $U^H BU$ are both diagonal.

9.3 Triangulation of Endomorphisms

All mentions of "trigular matrix" in the rest of this chapter refer to upper triangular matrices.

Definition 9.3.1: Triangulation

Let ϕ be an endomorphism of a vector space V. If there exists a basis for V such that $M_B(\phi)$ is a triangular matrix, ϕ can be reduced to a triangular form.

Theorem 9.3.1: Schur triangulation

Suppose V is a n-dimensional vector space over k and ϕ is an endomorphism of V. If the characteristic polynomial p_ϕ can be linearly factorised in k, ϕ can be reduced to a triangular form.

Proof Suppose $p_\phi(x) = \prod_{i=1}^{n}(x - \lambda_i)$. Note that λ_i may not be distinct for every i. By Theorem 8.1.2, λ_1 is an eigenvalue of ϕ. Correspondingly, find a non-zero eigenvector $\vec{v_1} \in V$. By Theorem 3.8.1, we can find a basis B_1 which contains $\vec{v_1}$. Consider the matrix of ϕ with respect to B_1, which should have the following form:

$$M_{B_1}(\phi) = \left(\begin{array}{c:c} \lambda_1 & R \\ \hdashline \mathbf{0} & M_1 \end{array}\right)$$

, where $M_1 \in M_{n-1}\, k$.

Hence, $p_\phi(x) = \det(xI_n - M_{B_1}(\phi)) = \det \left(\begin{array}{c:c} x - \lambda_1 & -R \\ \hdashline \mathbf{0} & xI_{n-1} - M_1 \end{array}\right) = (x - \lambda_1)\det(xI_{n-1} - M_1)$. (Expansion by the first column) This implies $p_{M_1}(x) = \det(xI_{n-1} - M_1) = \prod_{i=2}^{n}(x - \lambda_i)$.

Observe that $V = \operatorname{span} \vec{v_1} \oplus W_1$, where $W_1 = \operatorname{span}(B_1 \setminus \{\vec{v_1}\})$. Let ϕ_{W_1} be the endomorphism of W_1 of which M_1 is the matrix. By considering $p_{M_1}(x)$, we know that λ_2 is an eigenvalue of ϕ_{W_1}. Obtain another non-zero eigenvalue $\vec{v_2} \in W_1$ belonging to λ_2 and construct another basis B_2 which contains $\vec{v_1}$ and $\vec{v_2}$. ($\vec{v_1}$ and $\vec{v_2}$ are linearly independent because $\vec{v_2} \in W_1 \implies \vec{v_2} \notin \operatorname{span} \vec{v_1}$.) As $\phi(\vec{v_2}) = r_1\vec{v_1} + \phi_{M_1}(\vec{v_2}) = r_1\vec{v_1} + \lambda_2\vec{v_2}$ for some $r_1 \in k$, we get another matrix of ϕ:

$$M_{B_2}(\phi) = \left(\begin{array}{c:c:c} \lambda_1 & r_1 & R_1 \\ \hdashline 0 & \lambda_2 & R_2 \\ \hdashline \mathbf{0} & \mathbf{0} & M_2 \end{array}\right)$$

, where $M_2 \in M_{n-2}\, k$.

Hence, by repeating the above steps, we will ultimately get a basis B_n such that $M_{B_n}(\phi)$ is a triangular matrix, with its diagonal consisting of $\lambda_1, \lambda_2, ..., \lambda_n$.

With the change of basis formula, we can easily derive a matrix version of this theorem: For any $A \in M_n\, k$, if the characteristic polynomial of A can be linearly factorised in k, A is similar to a triangular matrix also in $M_n\, k$. (Consider $P_{\mathcal{E},B'}AP_{B',\mathcal{E}}$, where $\mathcal{E}$ is the standard basis and B' is the basis obtained through above.)

Remarks

(The discussion below is not essential to understanding the contents of this section. Uninterested readers may skip to the next result.)

The condition of Theorem 9.3.1 is actually equivalent to "the roots of p_ϕ are all in k", i.e. there is not a larger field k', of which k is a subfield, such that there exists a $x \in k$ which is a root to p_ϕ. A familiar example of such a case is the field of real numbers, with the larger field being the field of complex numbers. Fields in which all roots of any polynomials lie are algebraically closed, of which the field of complex numbers is an instance.

The condition given in Theorem 9.3.1 leads directly to the one introduced above. Hence, it is only necessary to prove the converse (p cannot be linearly factorised in $k \implies$ there exists a larger field $k' \supset k$ in which at least one root of p lies in). To do this, we would construct an arbituary k', taking inspiration from imaginary numbers.

Suppose p is of degree n. Obviously, $n \geq 2$. Introduce the imaginary numbers $i, i^2, ..., i^{n-1}$, which are connected through the identity $\underbrace{i \times ... \times i}_{m \text{ times}} \equiv i^m$. Another identity is given by the equation in question, i.e. $p(i) \equiv 0$, such that i is a root of p. Every element of k' is uniquely expressed in the format: $r_0 + r_1 i + ... + r_{n-1} r^{n-1}$, where $r_i \in k$. k is thus a subfield of k' as $r = r + 0i + ... + 0i^{n-1}$ for any $r \in k$. For any $x = \sum r_j i^j$ and $y = \sum s_j i^j$ in k', addition and multiplication is defined in the following way:

- $x + y = \sum (r_j + s_j) i^j$
- $xy = x \times y = \sum_{m=0}^{2n-2} \sum_{a+b=m} r_a s_b i^m$ (just like what one would get by "expanding brackets"), where i to powers larger than or equal to n is reduced through the second identity.

Finally, the multiplicative inverse of an element x in k' (except 0) is found in the following way:

$xx^{-1} = 1 \iff q_1(i)xx^-1 = q_1(i)$ (Polynomial division is unique)

$\iff -r_1(i)x^{-1} = q_1(i)$, where $q_1(i)x + r_1(i) = p(i) = 0$

Observe that the maximum possible degree of $r_1(i)$ is reduced by 1 to $n-2$.

Hence, by repeated polynomial divisions,

$\iff -r_{n-1}(i)x^{-1} = q_1(i)...q_{n-1}(i)$, where $r_{n-1}(i) \in k$

$\iff x^{-1} = -r_{n-1}^{-1} q_1(i)...q_{n-1}(i)$

Readers are encouraged to check if the above definition agrees with the axioms of a field.

Theorem 9.3.2

Suppose V is a real or complex inner product space and ϕ is an endomorphism of V. If V is defined over $\mathbb{R}$, further suppose p_ϕ can be linearly factorised in $\mathbb{R}$. Then there exists an orthonormal basis B for V such that $M_B(\phi)$ is triangular.

Proof We can employ a near-identical proof as that of Theorem 9.3.1. The only modification needed is that after we choose a basis for V or a subspace of V, we must now also turn it into an orthonormal basis $B = \{\vec{u_1}, \vec{u_2}, ..., \vec{u_m}\}$ through the Gram-Schimdt process. Observe that $(\operatorname{span} \vec{u_1})^\perp = \operatorname{span}\{\vec{u_2}, ..., \vec{u_m}\}$. This ensures the orthogonality of the final product.

Again, there is a matrix version of this theorem: Suppose A is a square matrix in a real or complex inner product space V. If V is defined over $\mathbb{R}$, further suppose p_A can be linearly factorised in $\mathbb{R}$. Then there exists an orthonormal basis B' for V such that $P = P_{B',\mathcal{E}}$ is unitary and $P^{-1}AP$ is triangular, where $\mathcal{E}$ is the standard basis.

9.4 Characteristic Subspaces

Definition 9.4.1: Characteristic subspaces

Suppose V is a vector space over k and ϕ is an endomorphism of V. Furthermore, p_ϕ can be linearly factorised in k. Then we can write:

$$p_\phi(x) = \prod_{i=1}^{n} (x - \lambda_i)^{m_i}$$

, where $\lambda_1, \lambda_2, ..., \lambda_n$ are distinct and m_i is called the (algebraic) multiplicity of λ_i.
Define endomorphisms $\varphi_1, \varphi_2, ..., \varphi_n$ of V by $\varphi_i = (\phi - \lambda_i \operatorname{id}_V)^{m_i}$. Their corresponding kernels $U_1, U_2, ..., U_n$ are the characteristic subspaces of V with respect to ϕ.

Characteristic subspaces have another name, "generalised eigenspaces". Such a name stems from the fact that all eigenspaces of an endomorphism are also characteristic subspaces. As the following theorem would demonstrate, characteristic subspaces is a similar concept to ordinary eigenspaces, while having the advantage of every vector space over algebraically closed fields being able to be completely decomposed into characteristic subspaces.

Theorem 9.4.1

Suppose $U_1, U_2, ..., U_n$ are characteristic subspaces of a finitely dimensional vector space V with respect to ϕ, which is an endomorphism over V. (V is implicitly assumed to meet the conditions given in Definition 9.4.1.) Then the following are true:

1. All of the characteristic subspaces are ϕ-invariant.
2. $V = U_1 \oplus U_2 \oplus ... \oplus U_n$
3. The only eigenvalue of $\phi|_{U_i}$ is λ_i
4. $\dim U_i = m_i$

Proof (1) For any $\vec{v}$ in every U_i:

$$\varphi_i(\phi(\vec{v})) = (\phi - \lambda_i \operatorname{id}_V)^{m_i}(\phi(\vec{v})) = \phi \circ (\phi - \lambda_i \operatorname{id}_V)^{m_i}(\vec{v}) = \phi(\vec{0}) = \vec{0}$$

(2) Given by the primary decomposition theorem.
(3) By definition, $(\phi|_{U_i} - \lambda_i \operatorname{id}_V)^{m_i} = O$ for $i \in \{1, ..., n\}$. By Proposition 9.2.2, $m_{\phi|_{U_i}}(x) = (x - \lambda_i)^p$ for some $p \le m_i$. In other words, the only root of the minimal polynomial is λ_i. Proposition 9.2.3 thus confirms (3).
(4) If we join the respective bases $B_1, B_2, ..., B_n$ of $U_1, U_2, ..., U_n$ into B, we have:

$$M_B(\phi) = \left(\begin{array}{c:c:c:c} M_{B_1}(\phi|_{U_1}) & & & \\ \hdashline & M_{B_2}(\phi|_{U_2}) & & \\ \hdashline & & \ddots & \\ \hdashline & & & M_{B_n}(\phi|_{U_n}) \end{array}\right)$$

(Empty blocks represent zero blocks.)
Proposition 7.2.3 tells us that $p_\phi = p_1 p_2 ... p_n$, where $p_i = p_{\phi|_{U_i}}$. As a result, p_i can be linearly factorised as well for $i \in \{1, ..., n\}$. We can therefore apply Schur triangulation onto each $\phi|_{U_i}$ and get another basis B' such that:

$$M = M_{B'}(\phi) = \left(\begin{array}{c:c:c:c} T_1 & & & \\ \hdashline & T_2 & & \\ \hdashline & & \ddots & \\ \hdashline & & & T_n \end{array}\right), T_i = \begin{pmatrix} \lambda_i & * & ... & * \\ 0 & \lambda_i & ... & * \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & ... & \lambda_i \end{pmatrix}$$

Observe that M is a triangular matrix, by Proposition 7.2.2, if $T_i \in M_{d_i}\ k$, $p_\phi(x) = \prod (x - \lambda_i)^{d_i}$. As $(x - \lambda_i)$ can only come from T_i, $\dim U_i = d_i = m_i$.

9.5 Nilpotent Mappings and the Jordan Normal Form

Definition 9.5.1: Nilpotent endomorphism and matrix

An endomorphism ϕ is nilpotent if there exists an n such that $\phi^n = O$, the zero map.
Similarly, A square matrix A is nilpotent if there exists an n such that $(A)^n = \mathbf{0}$, the zero matrix.
The least such n is called the index of nilpotency.

Definition 9.5.2: Supertriangular matrix

An upper triangular matrix is supertriangular if its diagonal only consists of entries of 0.

Proposition 9.5.1

A supertriangular matrix $A \in M_n\ k$ is nilpotent. More specifically, $(A)^n = \mathbf{0}$.

Proof By Proposition 7.2.2, $p_A(x) = x^n$. Hence, by the Cayley-Hamilton theorem, $A^n = \mathbf{0}$. As $(x - \lambda_i)$ can only come from T_i

Proposition 9.5.2

Suppose ϕ is an endomorphism of a non-zero vector space V and $\vec{w}$ is a vector in V. If n is the smallest natural number such that $\phi^n(\vec{v}) = \vec{0}$, then $\vec{v}, \phi(\vec{v}), ..., \phi^{n-1}(\vec{v})$ are linearly independent.

Proof Suppose there exist r_i such that $r_0\vec{v} + r_1\phi(\vec{v}) + ... + r_{n-1}\phi^{n-1}(\vec{v}) = \vec{0}$.
Then $\phi^{n-1}(\sum_{i=0}^{n-1} r_i\phi^i(\vec{v})) = \sum_{i=0}^{n-1} r_i\phi^{n+i-1}(\vec{v}) = r_{n-1}\phi^{n-1}(\vec{v}) = \vec{0} \implies r_{n-1} = 0$.
Subsequently, by applying $\phi^{n-2}, ..., \phi^0$ on the sum, one can show that $r_i = 0$ for any i, proving linear independence.

This implies that the maximum index of nilpotency of a nilpotent endomorphism of an n-dimensional vector space is n.

Definition 9.5.3: Cyclic subspace

In the context of Proposition 9.5.2, $\operatorname{span}\{\vec{v}, \phi(\vec{v}), ..., \phi^{n-1}(\vec{v})\}$ is the cyclic subspace of V with respect to ϕ. In this case, $\{\vec{v}, \phi(\vec{v}), ..., \phi^{n-1}(\vec{v})\}$ is the cyclic basis and $\vec{v}$ is the root of the cyclic basis.

Proposition 9.5.3

Suppose V is a vector space over k, of which W is a cyclic subspace with the cyclic basis $\{\vec{v}, \phi(\vec{v}), ..., \phi^{n-1}(\vec{v})\}$. Then if $p(\phi) = O$ for some $p \in k[x]$, x^n is a factor of p.

Proof Suppose $p(x) = q(x)x^n + r(x)$, where $q, r \in k[x]$ and the degree of r is smaller than n. Then $p(\phi)(\vec{v}) = q(\phi) \circ \phi^n(\vec{w}) + r(\phi)(\vec{v}) = \vec{0}$. By construction, $\phi^n(\vec{w}) = \vec{0}$, thus $r(\phi)(\vec{v}) = \vec{0}$. By Proposition 9.5.2, this is only possible if $r = 0$. Therefore, x^n is a factor of p.

Proposition 9.5.4

Suppose ϕ is a nilpotent endomorphism of a vector space V and V has a subspace W with the following properties:

1. $\phi(W) = \phi(V)$
2. W is the direct sum of (selected) cyclic subspaces with respect to ϕ.

Then V is likewise a direct sum of (selected) cyclic subspaces with respect to ϕ.

Proof By (1), for any $\vec{v} \in V$, there exists a $\vec{w} \in W$ such that $\phi(\vec{v}) = \phi(\vec{w})$. Note that $\vec{v} - \vec{w} \in \ker\phi$.
Hence, $V = W + \ker\phi$.
Let B_W and B_K be the bases for W and $\ker\phi$ respectively. By construction, $B_W \cup B_K$ spans V.
Hence, by the spanning set theorem, there exists $B'_K \subseteq B_K$ such that $B_W \cup B'_K$ is a basis for V. If $W' = \operatorname{span} B'_K$, $V = W \oplus W'$.
Observe that each basis vector in B'_K individually forms a cyclic basis (with respect to ϕ) already.

In other words, W' is a direct sum of cyclic subspaces. This, combined with (2), shows that V is a direct sum of cyclic subspaces.

Theorem 9.5.1

Suppose ϕ is a nilpotent endomorphism of a non-trivial vector space V over k. Then V is a direct sum of cyclic subspaces with respect to ϕ.

Proof Let $\dim V = n$. We will prove the theorem by induction:
When $n = 1$, Proposition 9.5.2 tells us that the index of nilpotency of ϕ can only be 1. In other words, ϕ can only be the zero map. Hence, for any basis B for V, it is also a cyclic basis and so V is a cyclic subspace already.
Suppose the theorem holds for all $(n-1)$-dimensional vector spaces. First observe that $W = \phi(V) \subsetneq V$, because otherwise $\phi^i(V) = V$ for all i. This means that W is at most $n-1$ dimensional. By assumption, $W = W_1 \oplus W_2 \oplus ... \oplus W_m$ for some cyclic subspaces W_i. Suppose $\vec{w_1}, \vec{w_2}, ..., \vec{w_m}$ are the roots of their cyclic bases respectively. Now find $\vec{v_1}, \vec{v_2}, ..., \vec{v_m}$ such that $\phi(\vec{v_i}) = \vec{w_i}$. $\vec{v_i}$ and the cyclic basis of W_i are still linearly independent due to Proposition 9.5.2. Thus, we can let $W_i' = \operatorname{span} \vec{v_i} \oplus W_i$ and $W' = W_1' + W_2' + ... + W_m'$.
Our next step is to study W'. If W' meets conditions (1) and (2) in Proposition 9.5.4, our proof is complete. Otherwise, for (1), consider that for any $\vec{w'} \in W_i'$, $\vec{w'} = r\vec{v_i} + \vec{w}$ for some $r \in k$ and $\vec{w} \in W_i$. Thus, $\phi(\vec{w'}) = r\vec{w_i} + \phi(\vec{w}) \in W_i \implies \phi(W') \subseteq W$. Also, for any $\vec{w} \in W$, suppose $\vec{w} = \vec{u_1} + \vec{u_2} + ... + \vec{u_m}$, where $\vec{u_i} \in W_i$. For each $\vec{u_i}$, it can be further expanded into $\vec{u_i} = r_0\vec{w_i} + r_1\phi(\vec{w_i}) + ... + r_p\phi^p(\vec{w_i})$, where $r_i \in k$ and $\phi^p(\vec{w_i}) = \vec{0}$. If we let $\vec{u_i'} = r_0\vec{v_i} + r_1\vec{w_i} + r_2\phi(\vec{w_i}) + ... + r_p\phi^{p-1}(\vec{w_i}) \in W_i'$, $\phi(\vec{u_i'}) = \vec{u}$. Therefore, $\phi(W') = W = \phi(V)$.
For (2), suppose $\vec{u_1'} + \vec{u_2'} + ... + \vec{u_m'} = \vec{0}$ for some $\vec{u_i'} \in W'$. By applying ϕ on both sides, we get $\vec{u_1} + \vec{u_2} + ... + \vec{u_m} = \vec{0}$, where $\vec{u_i} = \phi(\vec{u_i'})$. As shown above, $\vec{u_i'} \in W' \implies \vec{u_i} \in W$. Because W is the direct sum of W_i, $\vec{u_1} = \vec{u_2} = ... = \vec{0}$. For every i, suppose $\vec{u_i'} = r_0\vec{v_i} + r_1\vec{w_i} + r_2\phi(\vec{w_i}) + ... + r_p\phi^{p-1}(\vec{w_i})$, where $\phi^p(\vec{w_i}) = \vec{0}$. Then $\phi(\vec{u_i'}) = r_0\vec{w_i} + r_1\phi(\vec{w_i}) + ... + r_{p-1}\phi^{p-1}(\vec{w_i}) = \vec{u_i} = \vec{0}$. As $\{\vec{w_i}, \phi(\vec{w_i}), ..., \phi^{p-1}(\vec{w_i})\}$ is the cyclic basis, $r_0 = r_1 = ... = r_{p-1} = 0$. Consequently, returning to our initial assumption of this paragraph, we have $s_1\phi^{p_1-1}(\vec{w_1}) + s_2\phi^{p_2-1}(\vec{w_2}) + ... + s_m\phi^{p_m-1}(\vec{w_m}) = \vec{0}$, where $s_i = r_p$ and $p_i = p$ for each i in the previous step. Again, because W is the direct sum of W_i, $s_1 = s_2 = ... = s_m = 0$. In conclusion, $\vec{u_1'} = \vec{u_2'} = ... = \vec{u_m'} = \vec{0}$. In other words, by Proposition 3.5.3, $W' = W_1' \oplus W_2' \oplus ... \oplus W_m'$.

Proposition 9.5.5: Reduction to nilpotent Jordan normal form

For every nilpotent endomorphism ϕ on a vector space V, there exists a basis B for V such that $M_B(\phi)$ is in the nilpotent Jordan normal form:

$$M_B(\phi) = \left(\begin{array}{c:c:c:c} O_1 & & & \\ \hdashline & O_2 & & \\ \hdashline & & \ddots & \\ \hdashline & & & O_n \end{array}\right), O_i = \begin{pmatrix} 0 & 1 & & \\ & 0 & \ddots & \\ & & \ddots & 1 \\ & & & 0 \end{pmatrix}$$

(Empty blocks denote zero block and empty entries denote zero entries.)

Proof This proposition is a direct result of Theorem 9.5.1. First decompose V into cyclic subspaces $W_1, W_2, ..., W_m$ with cyclic bases $B_1, B_2, ..., B_m$. Order the bases in the form of $\{\phi^{p-1}(\vec{w_i}), \phi^{p-2}(\vec{w_i}), ..., \vec{w_i}\}$, where $\vec{w_i}$ is the root of B_i and $\phi^p(\vec{w_i}) = \vec{0}$. Also observe that all of the subspaces are ϕ-invariant. Let $B = B_1 \cup B_2 \cup ... \cup B_m$. The proposition should now be obvious.

Theorem 9.5.2: Reduction to Jordan normal form

Suppose V is a vector space over k and ϕ is an endomorphism of V. Further suppose the characteristic polynomial of ϕ can be linearly factorised in k. Let the (distinct) eigenvalues of ϕ be $\lambda_1, \lambda_2, ..., \lambda_n$, and their respective multiplicities be $m_1, m_2, ..., m_n$. Then there exists a basis B for V such that the matrix of ϕ with respect to B is in the Jordan normal form:

$$M_B(\phi) = \begin{pmatrix} A_1 & & & \\ & A_2 & & \\ & & \ddots & \\ & & & A_n \end{pmatrix}, A_i = \begin{pmatrix} J_{i,1} & & & \\ & J_{i,2} & & \\ & & \ddots & \\ & & & J_{i,m} \end{pmatrix}, J_{i,j} = \begin{pmatrix} \lambda_i & 1 & & \\ & \lambda_i & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_i \end{pmatrix}$$

Proof In Theorem 9.4.1, it has been shown that V can be decomposed into characteristic subspaces $U_1, U_2, ..., U_m$, and for each of U_i, there exists a basis B_i for U_i such that M_{B_i} has the following form:

$$M_{B_i}(\phi) = \begin{pmatrix} \lambda_i & * & ... & * \\ 0 & \lambda_i & ... & * \\ 0 & 0 & \ddots & \vdots \\ 0 & 0 & ... & \lambda_i \end{pmatrix} \in M_{m_i}\ k$$

We can further decompose $\phi|_{U_i}$ into the diagonal part D_i and the nilpotent part N_i such that:

$$M_{B_i}(\phi|_{U_i}) = M_{B_i}(D_i) + M_{B_i}(N_i) = \begin{pmatrix} \lambda_i & 0 & ... & 0 \\ 0 & \lambda_i & ... & 0 \\ 0 & 0 & \ddots & \vdots \\ 0 & 0 & ... & \lambda_i \end{pmatrix} + \begin{pmatrix} 0 & * & ... & * \\ 0 & 0 & ... & * \\ 0 & 0 & \ddots & \vdots \\ 0 & 0 & ... & 0 \end{pmatrix}$$

(Nilpotency of N_i comes from Proposition 9.5.1.)

Hence, by Proposition 9.5.5, there exists another basis B_i' for U_i such that $M_{B_i'}(N_i)$ is in the nilpotent Jordan normal form. By the change of basis formula and Proposition 5.6.1:

$$M_{B_i'}(D_i) = P_{B_i,B_i'} M_{B_i}(D_i) P_{B_i',B_i} = P_{B_i,B_i'}(\lambda_i I_{m_i}) P_{B_i',B_i} = \lambda_i P_{B_i,B_i'} P_{B_i',B_i} = \lambda_i I_{m_i} = M_{B_i}(D_i)$$

Therefore, $M_{B_i'}(\phi)$ is in the form of A_i.

Finally, as $V = U_1 \oplus U_2 \oplus ... \oplus U_m$, if we let $B = B_1' \cup B_2' \cup ... \cup B_m'$, the matrix of ϕ (with respect to B) would be in the desired form.

Each $J_{i,j}$ is called a Jordan block with respect to λ_i.

We can easily adapt the theorem in the context of matrices: Suppose $M \in M_n\ k$ and the characteristic polynomial of M can be linearly factorised in k. Let its eigenvalues be $\lambda_1, \lambda_2, ..., \lambda_n$, and their respective multiplicities be $m_1, m_2, ..., m_n$. Then M is similar to another matrix in the Jordan normal form.

Proposition 9.5.6

The Jordan normal form of any endomorphism of a vector space is unique up to the permutation of Jordan blocks.

Proof In Theorem 9.5.2, the size of each A_i is fixed as to preserve the characteristic polynomial of ϕ. As such, we only have to consider the uniqueness of the Jordan normal form of ϕ in each U_i.

Note that if two Jordan normal forms A and A' exist for $\phi|_{U_i}$, by the change of basis formula, they must be similar. Let $N = A - \lambda_i I$ and $N' = A' - \lambda_i I$. Then we have:

$$\begin{aligned} A' = P^{-1}AP &\implies A' - \lambda_i I = P^{-1}AP - \lambda_i P^{-1}P \\ &\implies A' - \lambda_i I = P^{-1}(A - \lambda_i I)P \\ &\implies N' = P^{-1}NP \implies N \sim N' \end{aligned} \tag{1}$$

Observe that if $\mathbf{0}$ denotes a zero block of an appropriate size:

$$O_{n,i} = \left(\begin{array}{c:c} \mathbf{0} & I_{n-i} \\ \hdashline \mathbf{0} & \mathbf{0} \end{array}\right) \implies (O_{n,i})^m = O_{n,i-m+1} \tag{2}$$

Sort the nilpotent Jordan blocks of N and N' by their sizes in decreasing order. Suppose:

$$N = \left(\begin{array}{c:c:c:c} O_1 & & & \\ \hdashline & O_2 & & \\ \hdashline & & \ddots & \\ \hdashline & & & O_p \end{array}\right), N' = \left(\begin{array}{c:c:c:c} O'_1 & & & \\ \hdashline & O'_2 & & \\ \hdashline & & \ddots & \\ \hdashline & & & O'_p \end{array}\right)$$

(To simplify expressions, we write $N = O_1 \oplus O_2 \oplus ... \oplus O_p$ and similarly for N'.)

By (2),

$$\begin{gathered} \dim \operatorname{im}(N)^2 = \dim U_i - p \\ \dim \operatorname{im}(N')^2 = \dim U_i - p' \end{gathered}$$

By (1), for all i:

$$N \sim N' \implies (N)^i \sim (N')^i \implies \dim \operatorname{im}(N)^i = \dim \operatorname{im}(N')^i \implies p = p' \tag{3}$$

In other words, A and A' has the same number of Jordan blocks.

Let $O_i \in M_{S_i}\, k$ and $O'_i \in M_{S'_i}\, k$. By (2),

$$\begin{gathered} (N)^{S_1-1} = (O_1)^{S_1-1} \oplus (O_2)^{S_2-1} \oplus ... \oplus (O_p)^{S_p-1} = O_{S_1,S_1-1} \oplus \mathbf{0} \oplus ... \oplus \mathbf{0} \\ (N')^{S_1-1} = (O'_1)^{S'_1-1} \oplus (O'_2)^{S'_2-1} \oplus ... \oplus (O'_p)^{S'_p-1} = O_{S_1,S_1-1} \oplus \mathbf{0} \oplus ... \oplus \mathbf{0} \end{gathered}$$

As shown in (3), the two matrices must be have the same rank, so $S_1 = S'_1$. (It is possible for $S_1 = S_2 = ... = S_q$ for some q, but the same logic follows.)

By repeating the above procedure for the $S_2, ..., S_p{}^{\text{th}}$ power of N and N', we can show that $S_i = S'_i$ for all i, which leads to $N = N'$ and subsequently $A = A'$.

The Jordan normal form reveals many details about the structure of a linear transformation:

Proposition 9.5.7

Suppose ϕ is an endomorphism of a vector space V over k and its characteristic polynomial can be linearly factorised in k. Let $\lambda_1, \lambda_2, ..., \lambda_n$ be its (distinct) eigenvalues, and $\Lambda_1, \Lambda_2, ..., \Lambda_n$ be the corresponding eigenspaces. Then if M is the Jordan normal form of ϕ, the following hold:

1. The characteristic polynomial $p_\phi(x) = \prod (x - \lambda_i)^{S_i}$, where S_i is the size of A_i.
2. The minimal polynomial $m_\phi(x) = \prod (x - \lambda_i)^{M_i}$, where M_i is the size of the largest Jordan block in A_i.
3. The geometric multiplicity of λ_i, i.e. $\dim \Lambda_i$, is the number of Jordan blocks in A_i.

Proof (1) Trivial.

(2) By (2) in the proof of Proposition 9.5.6.

(3) Let the number of Jordan blocks in A_i be m. By the rank-nullity theorem:

$$\begin{aligned} \dim \Lambda_i &= \dim \ker(M - \lambda_i I) \\ &= \dim V - \dim \operatorname{im}(M - \lambda_i I) \\ &= \sum_{j=1}^{n} S_j - \left(S_i - m + \sum_{j=1, j \neq i}^{n} S_j\right) \\ &= m \end{aligned}$$

10 Additional Topics

10.1 Singular Value Decomposition

In this section, all vectors belong to $\mathbb{C}^n$ equipped with the dot product. Results can be easily applied to $\mathbb{R}^n$ with slight modification.

The spectral decomposition theorem is extremely powerful because it allows us to transform normal matrices into diagonal matrices, which greatly eases computation. However, it has many limitations, two of which are non-square matrices and non-normal matrices. Singular value decomposition is a generalisation of spectral decomposition that can be applied to all (finite) matrices.

Recall that all Hermitian matrices only have real eigenvalues. We start with the following observation:

Proposition 10.1.1

A Hermitian matrix $A \in M_n\,\mathbb{C}$ is positive definite if and only if all of its eigenvalues are positive.

Proof By the spectral decomposition theorem, there is an orthonormal eigenbasis B for $\mathbb{C}^n$ with respect to A. By Theorem 8.1.1, if $P = P_{B,\mathcal{E}}$, $A = PDP^{-1}$, where $\mathcal{E}$ is the standard basis and D is a diagonal matrix consisting of the eigenvalues of A. It is obvious that P is isometric and hence unitary. (Proposition 8.3.1) Hence, by Proposition 8.2.2, $P^{-1} = P^H = \overline{P^T}$.
(→) Suppose the i^{th} eigenvalue in D, λ_i, is non-positive. Then if $\vec{v} = P\vec{e_i}$:

$$\overline{\vec{v}}^T A\vec{v} = (\overline{\vec{e_i}}^T P^H)(PDP^H)(P\vec{e_i}) = \overline{\vec{e_i}}^T D\vec{e_i} = \lambda_i \leq 0$$

(←) For any $\vec{v}$, let $\vec{w} = P^H\vec{v} = \langle w_1, w_2, ..., w_n\rangle \implies \vec{v} = P\vec{w}$. Then:

$$\overline{\vec{v}}^T A\vec{v} = \overline{\vec{w}}^T D\vec{w} = \lambda_1|w_1|^2 + ... + \lambda_n|w_n|^2 > 0$$

Similarly, A is positive semidefinite if and only if all of its eigenvalues are non-negative.

Theorem 10.1.1: Square root of matrix

For any positive definite Hermitian matrix A, there exists a unique positive definite Hermitian "square root" of A, denoted by $\sqrt{A}$, such that $(\sqrt{A})^2 = A$.

Proof *Existence* If $A = PDP^H$, $\sqrt{A} = P\sqrt{D}P^H$, where each diagonal in $\sqrt{D}$ is the square root of that in D. (This must be possible as Proposition 10.1.1 ensures the positivity of eigenvalues.)
Uniqueness Let $\Lambda_1, \Lambda_2, ..., \Lambda_n$ be (distinct) eigenspaces of A, belonging to eigenvalues $\lambda_1, \lambda_2, ..., \lambda_n$ respectively. By the spectral decomposition theorem, their direct sum equal to the entire vector space. First observe that for any $\vec{v} \in \Lambda_i$:

$$A\sqrt{A}\vec{v} = \sqrt{A}A\vec{v} = \lambda_i\sqrt{A}\vec{v} \implies \sqrt{A}\vec{v} \in \Lambda_i$$

In other words, $\sqrt{A}$ is Λ_i-invariant for any i. This means we only have to study its uniqueness for each eigenspace. Again by the spectral decomposition theorem, there exists an orthonormal eigenbasis $\{\vec{w_1}, \vec{w_2}, ..., \vec{w_m}\}$ for Λ_i, belonging to eigenvalues $\{\omega_1, \omega_2, ..., \omega_m\}$ respectively, with respect to $\sqrt{A}$. Note that for any $i \in \{1, ..., m\}$:

$$\sqrt{A}\sqrt{A}\vec{w_j} = A\vec{w_j} = \omega_j^2\vec{w_j} \implies \omega_j^2 = \lambda_i$$

By Proposition 10.1.1, the only eigenvalue of $\sqrt{A}$ in Λ_i is $\sqrt{\lambda_i}$. In other words, $\sqrt{A}\vec{v} = \sqrt{\lambda_i}\vec{v}$ for $\vec{v} \in \Lambda_i$. Therefore, $\sqrt{A}$ is unique.

Similarly, every positive semidefinite Hermitian matrix A has a unique positive semidefinite Hermitian square root $\sqrt{A}$.

Proposition 10.1.2

If a matrix $A \in M_n\ \mathbb{C}$ has a square root, there exists a <u>unique</u> $p \in \mathbb{C}_{n-1}[x]$ such that $p(A) = \sqrt{A}$.

Proof Let $A = PDP^H$ and the diagonal entries of D be $\lambda_1, \lambda_2, ..., \lambda_n$. Using Lagrange interpolation, we can find a $p \in \mathbb{C}_{n-1}[x]$ such that $p(\lambda_i) = \sqrt{\lambda_i}$ for all i, where:

$$p(x) = \sum_{j=0}^{n-1} r_j x^j = \sum_{j=0}^{n-1} \sqrt{\lambda_j} \frac{(x-\lambda_1)...(x-\lambda_{j-1})(x-\lambda_{j+1})...(x-\lambda_n)}{(\lambda_j-\lambda_1)...(\lambda_j-\lambda_{j-1})(\lambda_j-\lambda_{j+1})...(\lambda_j-\lambda_n)}$$

Hence:

$$p(A) = \sum_{j=0}^{n-1} r_j (A)^j = \sum_{j=0}^{n-1} r_j P(D)^j P^H = P\left(\sum_{j=0}^{n-1} (D)^j\right) P^H = P\sqrt{D}P^H = \sqrt{A}$$

Note that p is unique because if there is another such polynomial $q \in \mathbb{C}_{n-1}[x]$, $p - q$ would have $n > n-1$ roots, which is impossible.

A corollary of this proposition is that if a matrix B commutes with A, then B also commutes with $\sqrt{A}$.

Proposition 10.1.3

For any matrix $A \in \mathrm{Mat}_{m\times n}\,\mathbb{C}$, $\overline{A^T}A$ is a positive semidefinite Hermitian matrix. Conversely, any positive semidefinite Hermitian matrix B can be expressed in the form $B = \overline{A^T}A$.

Proof **(→)** Note that $\overline{A^T}A \in M_n\ \mathbb{C}$. By Proposition 8.2.2, $(\overline{A^T}A)^H = \overline{(\overline{A^T}A)^T} = \overline{A^T}A$, which means $\overline{A^T}A$ is Hermitian. For any eigenvalue λ of $\overline{A^T}A$:

$$\overline{A^T}A\vec{v} = \lambda\vec{v} \implies \overline{\vec{v}^T}\,\overline{A^T}A\vec{v} = \lambda\overline{\vec{v}^T}\vec{v} \implies \lambda||\vec{v}||^2 = ||A\vec{v}||^2 \implies \lambda \geq 0$$

By Proposition 10.1.1, $\overline{A^T}A$ is also positive semidefinite.
(←) By letting $A = \sqrt{B}$.

Definition 10.1.1: Absolute value of matrix

For any matrix $A \in \mathrm{Mat}_{m\times n}\,\mathbb{C}$, the absolute value of A is $|A| = \sqrt{\overline{A^T}A}$.

Definition 10.1.2: Singular values

For any matrix $A \in \mathrm{Mat}_{m\times n}\,\mathbb{C}$, the singular values of A are $\sigma_1, \sigma_2, ..., \sigma_n$, where $\{\sigma_i\}$ are (real, non-negative) eigenvalues of $|A|$ in descending order, and each eigenvalue is repeated for a number of times equal to the dimension of its eigenspace.

Proposition 10.1.4

The number of non-zero singular values r of a matrix $A \in \mathrm{Mat}_{m\times n}\,\mathbb{C}$ is equal to its rank.

Proof By definition of singular values, $\sigma_1^2, ..., \sigma_n^2$ are the eigenvalues of $\overline{A^T}A$. Construct an eigenbasis $\{\vec{u_1}, ..., \vec{u_n}\}$ with respect to $\overline{A^T}A$, in which $\vec{u_i}$ belonging to σ_i^2. Apply $\overline{A^T}A$ to each of the vectors: only $\vec{u_1}, ..., \vec{u_r}$ are not eliminated. Hence, the rank of $\overline{A^T}A$ is equal to r.
Recall Proposition 6.4.5, which states that $\ker\overline{A^T} = (\mathrm{im}\,A)^\perp$. As a result, $\ker\overline{A^T}A = \ker A$. By the rank-nullity theorem, $\dim\mathrm{im}\,A = m - \dim\ker\overline{A^T}A = n - (n-r) = r$.

A corollary of this proposition is that $r \leq n$ and $r \leq m$.

Proposition 10.1.5

For any singular value σ of $A \in \text{Mat}_{m\times n}\mathbb{C}$ and a corresponding eigenvector $\vec{v} \in \mathbb{C}^n$, $||A\vec{v}|| = \sigma||\vec{v}||$.

Proof $||A\vec{v}||^2 = \overline{\vec{v}^T A^T} A\vec{v} = \sigma^2 \overline{\vec{v}^T}\vec{v} = \sigma^2||\vec{v}||^2$

Proposition 10.1.6

Suppose $\vec{v}$ and $\vec{w}$ are eigenvectors belonging to distinct singular values of any matrix $A \in \text{Mat}_{m\times n}\mathbb{C}$. If $\vec{v} \perp \vec{w}$, $A\vec{v} \perp A\vec{w}$.

Proof By Propositoin 8.4.3, $A\vec{v} \cdot A\vec{w} = \overline{\vec{v}^T A^T} A\vec{w} = \sigma^2\overline{\vec{v}^T}\vec{w} = 0$.

Theorem 10.1.2: Singular value decomposition (SVD)

Any matrix $A \in \text{Mat}_{m\times n}\mathbb{C}$ can be expressed in the form $A = U\Sigma V^H$, where:

- $U \in M_m\,\mathbb{C}$ is a unitary transformation.
- $V \in M_n\,\mathbb{C}$ is a unitary transformation.
- $\Sigma = \left(\begin{array}{ccc:c} \sigma_1 & & & \\ & \ddots & & \mathbf{0}_{r\times n-r} \\ & & \sigma_r & \\ \hdashline \multicolumn{3}{c:}{\mathbf{0}_{m-r\times r}} & \mathbf{0}_{m-r\times n-r} \end{array}\right) \in \text{Mat}_{m\times n}\mathbb{C}$ is a pseudo-diagonal matrix.
- $\sigma_1, ..., \sigma_r$ are the non-zero singular values of A.
- $\mathbf{0}$ is a block of zero entries of the specified shape.

Proof As in the proof of Proposition 10.1.1, we can find an orthonormal eigenbasis $B_n = \{\vec{v_1}, \vec{v_2}, ..., \vec{v_n}\}$ for $\mathbb{C}^n$ with respect to $|A|$, and $V = P_{B_n,\mathcal{E}}$ would be a unitary matrix, where $\mathcal{E}_n$ is the standard basis of $\mathbb{C}^n$.

Next, let $B'_m = \{\sigma_1^{-1}A\vec{v_1}, \sigma_2^{-1}A\vec{v_2}, ..., \sigma_r^{-1}A\vec{v_r}\}$. Note that this set is normal by Proposition 10.1.5. Therefore, by Proposition 10.1.6 and Proposition 6.1.3, B'_m is orthonormal and hence linear independent. Utilising the basis extension theorem and the Gram-Schimdt process, it can be extended to an orthonormal basis B_m for $\mathbb{C}^m$. Let $U = P_{B_m,\mathcal{E}}$, which is also unitary.

Now we show that $A = U\Sigma V^H$. For $i \in \{1, ..., r\}$:

$$U\Sigma V^H\vec{v_i} = U\Sigma\vec{e_i} = \sigma_i U\vec{e_i} = \sigma_i\sigma_i^{-1}A\vec{v_i} = A\vec{v_i}$$

For $i \in \{r+1, ..., n\}$, note that Proposition 10.1.4 and the basis theorem imply B'_m is a basis for $\text{im}\,A$. Thus, as $A\vec{v_i} \in (\text{im}\,A)^\perp$ (by Proposition 10.1.6) but $A\vec{v_i} \in \text{im}\,A$, $A\vec{v_i} = \vec{0}$. Therefore:

$$U\Sigma V^H\vec{v_i} = \vec{0} = A\vec{v_i}$$

As A and $U\Sigma V^H$ agree on all basis vectors, they also agree on all vectors in $\mathbb{C}^n$. By Theorem 5.1.3, the two matrices are identical.

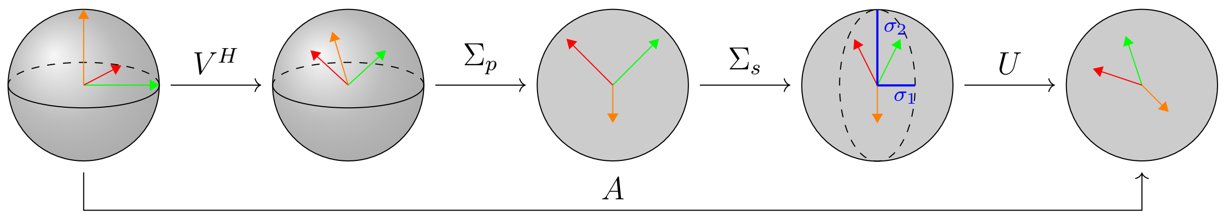

Figure 10.1.1: Geometric interpretation of SVD ($\Sigma = \Sigma_s\Sigma_p$)

In essence, SVD implies that every linear transformation can be seen as a composition of a rotation/reflection transformation, a projection/scaling transformation, and another rotation/reflection transformation.

Theorem 10.1.3: Polar decomposition

Any square matrix $A \in M_n\ \mathbb{C}$ can be expressed in the form $A = UH$, where:

- $U \in M_n\ \mathbb{C}$ is a unitary matrix.
- $H \in M_n\ \mathbb{C}$ is a positive semidefinite Hermitian matrix.

Proof Using SVD, decompose A into $A = U'\Sigma V^H$. The requirement can be met by setting $U = U'V$ and $H = V^H \Sigma V$: U is isometric and hence unitary (Proposition 8.3.1) because it is a composition of isometries, H is positive definite by Proposition 10.1.1.

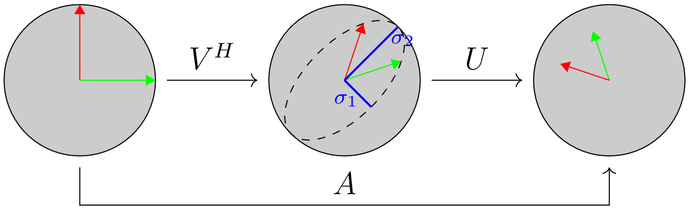

Figure 10.1.2: Geometric interpretation of polar decomposition

Recall that all Hermitian matrices can be eigendecomposed. Hence, they can be seen solely as scaling transformations, in which all scaling factor is non-negative due to the positive definiteness of H. Theorem 10.1.3 is hence a generalisation of polar decomposition of complex numbers:

$$z = re^{i\theta}$$

In this case, U is analogous to θ (the rotation/reflection transformation) and H is analogous to r.

Proposition 10.1.7

Given $A \in M_n\ \mathbb{C}$, all polar decompositions of $A = UH$ have the same H and $H = |A|$.

Proof For any two polar decompositions of A: $A = UH$ and $A = U'H'$, by Proposition 8.2.2:

$$\begin{aligned} A^H &= \overline{A^T} = \overline{H^T U^T} = \overline{(H')^T (U')^T} \\ \implies A^H &= H^H U^H = (H')^H (U')^H \\ \implies A^H &= HU^{-1} = H'(U')^{-1} \\ \implies A^H A &= HU^{-1}UH = H'(U')^{-1}U'H \\ \implies |A|^2 &= H^2 = (H')^2 \end{aligned}$$

By Proposition 10.1.3 and Theorem 10.1.1, $H = H' = |A|$.

A corollary of this proposition is that if A is invertible, the polar decomposition of A is unique and H is positive definite. The latter is obvious by considering determinants, while the former is true because $A = UH = U'H \implies U = U' = AH^{-1}$.

Proposition 10.1.8

Given $A \in M_n\ \mathbb{C}$, let $A = UH$ be its polar decomposition. Then A is normal if and only if U and H commute, or equivalently, iff U and H are simultaneously diagonalisable.

Proof **(→)** By Proposition 10.1.7, $A^H A = H^2$. Similarly, $AA^H = UH^2U^{-1}$. Observe that:

$$AA^H U = UH^2U^{-1}U = UH^2 = UA^H A$$

In other words, U commutes with $A^H A$. By Proposition 10.1.2, U also commutes with $\sqrt{A^H A} = H$.

(←) Note that $UH = HU \implies H = U^{-1}HU$. Hence:

$$AA^H = UH^2U^{-1} = UU^{-1}HUU^{-1}HUU^{-1} = H^2 = A^H A$$

Theorem 10.1.4: Moore-Penrose inverse

For any $A \in \operatorname{Mat}_{m\times n}\mathbb{C}$, there exists a unique pseudoinverse of A, $A^+ \in \operatorname{Mat}_{n\times m}\mathbb{C}$, such that:

1. $(AA^+)A = A$
2. $A^+(AA^+) = A^+$
3. AA^+ is Hermitian.
4. A^+A is Hermitian.

Proof *Existence* Let the SVD of A be $A = U\Sigma V^H$. Define Φ as follows:

$$\Sigma = \begin{pmatrix} D_{r\times r} & \mathbf{0}_{r\times n-r} \\ \mathbf{0}_{m-r\times r} & \mathbf{0}_{m-r\times n-r} \end{pmatrix} \in \operatorname{Mat}_{m\times n}\mathbb{C} \implies \Phi = \begin{pmatrix} D^{-1}_{r\times r} & \mathbf{0}_{r\times m-r} \\ \mathbf{0}_{n-r\times r} & \mathbf{0}_{n-r\times m-r} \end{pmatrix} \in \operatorname{Mat}_{n\times m}\mathbb{C}$$

Observe that:

$$\Sigma\Phi = \begin{pmatrix} I_{r\times r} & \mathbf{0}_{r\times m-r} \\ \mathbf{0}_{m-r\times r} & \mathbf{0}_{m-r\times m-r} \end{pmatrix} \in M_m\,\mathbb{C}$$

$$\Phi\Sigma = \begin{pmatrix} I_{r\times r} & \mathbf{0}_{r\times n-r} \\ \mathbf{0}_{n-r\times r} & \mathbf{0}_{n-r\times n-r} \end{pmatrix} \in M_n\,\mathbb{C}$$

Therefore, if we let $A^+ = V\Phi U^H$, we have:

$$\begin{aligned}
(AA^+)A &= (U\Sigma V^H)(V\Phi U^H)(U\Sigma V^H) = U(\Sigma\Phi)\Sigma V^H = U\Sigma V^H = A \\
A^+(AA^+) &= (V\Phi U^H)(U\Sigma V^H)(V\Phi U^H) = V(\Phi\Sigma)\Phi U^H = V\Sigma U^H = A^+ \\
(AA^+)^H &= ((U\Sigma V^H)(V\Phi U^H))^H = (U(\Sigma\Phi)U^H)^H = U(\Sigma\Phi)U^H = AA^+ \\
(A^+A)^H &= ((V\Phi U^H)(U\Sigma V^H))^H = (V(\Phi\Sigma)V^H)^H = V(\Phi\Sigma)V^H = A^+A
\end{aligned}$$

, where the last two equalities are due to Proposition 8.2.2.

Uniqueness Suppose B and B' are pseudoinverses of A. By Propositipn 8.2.2:

$$B \overset{(2)}{=} BAB \overset{(3)}{=} B\overline{B^TA^T} \overset{(1)}{=} B\overline{B^T(AB'A)^T} \overset{(2)}{=} B\overline{B^TA^T(B')^TA^T} \overset{(3)}{=} BABAB' \overset{(1)}{=} BAB'$$

$$B' \overset{(2)}{=} B'AB' \overset{(4)}{=} \overline{A^T(B')^T}B' \overset{(1)}{=} \overline{(ABA)^T(B')^T}B' \overset{(2)}{=} \overline{A^TB^TA^T(B')^T}B' \overset{(3)}{=} BAB'AB' \overset{(1)}{=} BAB'$$

Hence, $B = B'$, meaning the pseudoinverse must be unique.

Proposition 10.1.9

For any $A \in \operatorname{Mat}_{m\times n}\mathbb{C}$, AA^+ is the matrix of $\operatorname{pr}_{\operatorname{im} A}$ with respect to the standard basis $\mathcal{E}$.

Proof The fact that $(AA^+)A = A$ means that AA^+ maps columns of A to themselves. By first eliminating linearly dependent vectors from the columns of A, then invoking the basis extension theorem and the Gram-Schimdt process, we obtain a basis B for $\mathbb{C}^m$. Proposition 6.4.5 states that $\ker \overline{(AA^+)^T} = \ker AA^+ = (\operatorname{im} A)^\perp$, the first equality coming from Proposition 8.2.2. Therefore, for any $\vec{u} \in B \setminus \{A^j\}$, $AA^+\vec{u} = \vec{0}$. We have hence shown that AA^+ and $\operatorname{pr}_{\operatorname{im} A}$ agree on all basis vectors in B, and thus also all vectors in $\mathbb{C}^m$.

This result hints at a very useful application of SVD: linear regression. Suppose we have determined that a (response) variable (y) depends linearly on some (explanatory) variables $x_1, x_2, ..., x_n$. We wish to find the best estimations for parameters $\beta_1, \beta_2, ..., \beta_n$ such that the equation:

$$y_i = \beta_1 x_{i,1} + \beta_2 x_{i,2} + ... + \beta_n x_{i,n} + \varepsilon_i$$

would result in the least error ε based on a given dataset $\{x_{i,1}, ..., x_{i,n}, y_i \mid i \in \{1, ..., p\}\}$. If the least-square estimation is used, the task would be to minimise $\sum \varepsilon_i^2$.

To see how SVD can be utilised here, we first express the above in matrix form:

$$\vec{y} = X\vec{b} + \vec{\varepsilon} \iff \begin{pmatrix} y_1 \\ \vdots \\ y_p \end{pmatrix} = \begin{pmatrix} x_{1,1} & \dots & x_{1,n} \\ \vdots & \vdots & \vdots \\ x_{p,1} & \dots & x_{p,n} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_n \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_p \end{pmatrix}$$

The objective is to minimise $\sum \varepsilon_i^2 = ||\vec{\varepsilon}||^2$. Note that $\vec{\varepsilon} = \vec{y} - X\vec{b}$. As $X\vec{b} \in \operatorname{im} X$, by Proposition 6.2.7, $||\vec{\varepsilon}||$ is minimised when $X\vec{b} = \operatorname{pr}_{\operatorname{im} X} \vec{y} = XX^+\vec{y}$, which can (obviously) be achieved by setting:

$$\vec{b} = X^+\vec{y}$$

10.2 Quotient Spaces

Vector spaces are not assumed to be finitely dimensional in this section.

Theorem 10.2.1: Quotient space

Let V be a vector space and U be a subspace of V. An equivalence relation $\sim_U$ is defined as follows:

$$\forall \vec{v}, \vec{w} \in V : \vec{v} \sim_U \vec{w} \iff \vec{v} - \vec{w} \in U$$

Furthermore, if the addition and scalar multiplication of equivalence classes (with respect to $\sim_U$) are defined as follows (for $\vec{v}, \vec{w} \in V$ and $r \in k$):

$$[\vec{v}] + [\vec{w}] = [\vec{v} + \vec{w}]$$
$$r[\vec{v}] = [r\vec{v}]$$

, the set of equivalence classes V/U form a vector space, which is named the quotient space.

Proof We first verify that $\sim_U$ is an equivalence relation using the properties of vector spaces. For $\vec{u}, \vec{v}, \vec{w} \in V$:

Reflexivity $\vec{v} - \vec{v} = \vec{0} \in U \implies \vec{v} \sim_U \vec{v}$
Symmetry $\vec{v} \sim_U \vec{w} \implies \vec{v} - \vec{w} \in U \implies \vec{w} - \vec{v} \in U$
Transitivity $\vec{u} \sim_U \vec{v}, \vec{v} \sim_U \vec{w} \implies \vec{u} - \vec{v} \in U, \vec{v} - \vec{w} \in U \implies \vec{u} - \vec{w} \in U \implies \vec{u} \sim_U \vec{w}$

Then we need to show that addition and scalar multiplication of equivalence classes are well-defined. For $x, y \in V/U; \vec{v}, \vec{v'} \in x; \vec{w}, \vec{w'} \in y$:

$$(\vec{v} + \vec{w}) - (\vec{v'} + \vec{w'}) = (\vec{v} - \vec{v'}) + (\vec{w} - \vec{w'}) \in U \implies [\vec{v} + \vec{w}] = [\vec{v'} + \vec{w'}]$$
$$(r\vec{v}) - (r\vec{v'}) = r(\vec{v} - \vec{v'}) \in U \implies [r\vec{v}] = [r\vec{v'}]$$

Finally, due to the definition of addition and scalar multiplication for equivalence classes, they must form a vector space.

Note that equivalence classes themselves are in general not vector spaces. Only $[\vec{0}]$ in V/U is a vector space, and is identical to U.

A quotient space groups vectors by certain properties possessed by the given subspace. For example, in $\mathbb{R}^2$, if $\vec{v}$ is any non-zero vector and $U = \operatorname{span} \vec{v}$, $\mathbb{R}^2/U$ consists of lines that are parallel to $\vec{v}$.

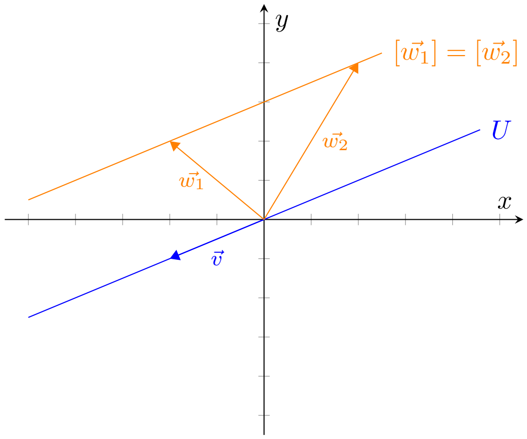

Figure 10.2.1: Visual representation of $\mathbb{R}^2/U$

Proposition 10.2.1: Quotient map

Let V be a vector space and U be its subspace. Then the following map θ_U:

$$V \to V/U$$
$$\vec{v} \mapsto [\vec{v}]$$

is linear, surjective, and has kernel U.

Proof Its linearity and surjectivity follows from the definition of quotient spaces.
For any $\vec{u} \in U$, $\vec{u} - \vec{0} \in U \implies \theta_U(\vec{u}) = [\vec{u}] = [\vec{0}] \implies U \subseteq \ker \theta_U$.
On the other hand, $\theta_U(\vec{v}) = [\vec{0}] \implies \vec{v} - \vec{0} = \vec{v} \in U \implies \ker \theta_U \subseteq U$.

Theorem 10.2.2: First isomorphism theorem

Suppose V and W are vector spaces and $\phi \in \hom(V, W)$, then the following map $\overline{\phi}$ is isomorphic:

$$V/\ker \phi \to \operatorname{im} \phi$$
$$[\vec{v}] \mapsto \phi(\vec{v})$$

Proof Note that $\overline{\phi}$ is well-defined. For $\vec{v_1}, \vec{v_2} \in [\vec{v}]$:

$$\phi(\vec{v_1}) - \phi(\vec{v_2}) = \phi(\vec{v_1} - \vec{v_2} \in \ker \phi) = \vec{0} \implies \phi(\vec{v_1}) = \phi(\vec{v_2})$$

The linearity of $\overline{\phi}$ is trivial. We thus only have to show its bijectivity:
Injectivity By definition, $\ker \overline{\phi} = \{[\vec{v}] \mid \vec{v} \in \ker \phi\} = [\vec{0}]$.
Surjectivity By definition of $\overline{\phi}$.

Theorem 10.2.3: Universal property for quotient

Suppose V and W are vector spaces and $\phi \in \hom(V, W)$. Let U and be a subspace of V. Then $U \subseteq \ker \phi$ if and only if there exists a unique $\overline{\phi}$ such that $\phi = \overline{\phi} \circ \theta_U$.

Proof (→) We can basically reuse $\overline{\phi}$ from the first isomorphism theorem, with the domain of the map changed to V/U. The uniqueness of $\overline{\phi}$ stems from the surjectivity of θ_U (Proposition 10.2.1).
(←) It is impossible for such a map to exist when $U \not\subseteq \ker \phi$. Find a $\vec{u}$ in U but not in $\ker \phi$. Then:

$$\phi(\vec{0}) = \overline{\phi} \circ \theta_U(\vec{0}) = \overline{\phi}([\vec{0}]) = \vec{0}$$
$$\phi(\vec{u}) = \overline{\phi} \circ \theta_U(\vec{u}) = \overline{\phi}([\vec{u}]) \neq \vec{0}$$

However, as $\vec{u} - \vec{0} = \vec{u} \in U$, $\theta_U(\vec{0}) = \theta_U(\vec{u})$. $\overline{\phi}$ is hence not a well-defined map.

Notice that $\overline{\phi}$ may not be injective if $U \neq \ker \phi$.

$$\begin{tikzcd}
V \arrow[rr, "\phi"] \arrow[dr, "\theta_U"'] & & W \\
& V/U \arrow[ur, "\overline{\phi}"'] &
\end{tikzcd}$$

Figure 10.2.2: An illustration of Theorem 10.2.3

We may further generalise this map by setting the codmain to a quotient space as well. Let U' be a subspace of W, then we have a linear map $\overline{\phi}$ induced by ϕ:

$$V/U \to W/U'$$
$$[\vec{v}]_U \mapsto \theta_{U'} \circ \phi(\vec{v})$$

In particular, if ϕ is an endomorphism on V and $U = U'$ is ϕ-invariant, the restriction on U being a subspace of $\ker\phi$ can also be relaxed. To show that $\overline{\phi}$ is still well-defined in this case, consider that for $\vec{v_1}, \vec{v_2} \in [\vec{v}]$:

$$\overline{\phi}(\vec{v_1}) - \overline{\phi}(\vec{v_2}) = \theta_{U'} \circ \phi(\vec{v_1}) - \theta_{U'} \circ \phi(\vec{v_2}) = \theta_{U'} \circ \phi(\vec{v_1} - \vec{v_2} \in U) = \theta_{U'}(\phi(\vec{v_1} - \vec{v_2}) \in U) = [\vec{0}]$$

In this case, $\overline{\phi}$ is an endomorphism on V/U.

Theorem 10.2.4: Splitting of vector space

Suppose V is a vector space and U is its subspace. If V/U is finitely dimensional, then $V \cong U \times V/U$.

Proof Suppose $B = \{[\vec{v_1}], [\vec{v_2}], ..., [\vec{v_n}]\}$ is a basis for V/U. Note that $S = \{\vec{v_1}, ..., \vec{v_n}\}$ is linearly independent. Otherwise there exists a non-trivial $\{r_i\} \subseteq k$ such that:

$$r_1\vec{v_1} + ... + r_n\vec{v_n} = \vec{0} \implies \theta(r_1\vec{v_1} + ... + r_n\vec{v_n}) = \theta(\vec{0}) \implies r_1[\vec{v_1}] + ... + r_n[\vec{v_n}] = [\vec{0}]$$

, which contradicts our assumption.
Define the splitting map χ_U as follows:

$$V \to U \times V/U$$
$$\vec{v} \mapsto (\vec{v} - \gamma_B(\theta(\vec{v})) \cdot S, \theta(\vec{v}))$$

Note that $\vec{v} - \gamma_B([\vec{v}]) \cdot S \in U$ because:

$$\theta(\vec{v} - \gamma_B([\vec{v}]) \cdot S) = \theta(\vec{v}) - \gamma_B([\vec{v}]) \cdot B = [0]$$

Leaving linearity as an exercise to the readers, we show that this map is bijective:
Injectivity If $\theta(\vec{v}) = [\vec{0}]$, $\vec{v} - \gamma_B(\theta(\vec{v})) \cdot S = \vec{v} - \vec{0} = \vec{v}$. Therefore, $\ker\chi_U = \{\vec{0}\}$.
Surjectivity For any $(\vec{u}, x) \in U \times V/U$, let $\vec{v} = \vec{u} + \gamma_B(x) \cdot S$. Then $\chi_U(\vec{v}) = (\vec{u}, x)$.

A corollary of this result is that $\dim V/U = \dim V - \dim U$.
Combined with the first isomorphism theorem, it also provides another proof to the rank-nullity theorem: For any linear transformation ϕ between vector spaces V and W, if $\dim V < \infty$, V can be split into $\ker\phi \times V/\ker\phi$. Then as $V/\ker\phi \cong \operatorname{im}\phi \implies \dim V/\ker\phi = \dim\operatorname{im}\phi$, $\dim V = \dim\ker\phi + \dim\operatorname{im}\phi$.

Theorem 10.2.5: Second isomorphism theorem

Let U and W be subspaces of a vector space V. Then the following map F:

$$U \to (U + W)/W$$
$$\vec{u} \mapsto \theta(\vec{u})$$

is linear, surjective and has kernel $U \cap W$.

Proof We will omit the proof for linearity. To show surjectivity, note that for $x \in (U + W)/W$, by Proposition 10.2.1, there exists a $\vec{v} \in U + W$ such that $\theta(\vec{v}) = x$. Find $\vec{u} \in U$ and $\vec{v} \in V$ such that $\vec{v} = \vec{u} + \vec{w}$. Then $F(\vec{u})$ is exactly x.
It is obvious that $U \cap W \subseteq \ker F$. Conversely, $\theta(\vec{u}) = [\vec{0}] \implies \vec{u} - \vec{0} = \vec{u} \in W \implies \ker F \subseteq U \cap W$.

By the first isomorphism theorem, $U/(U \cap W) \cong (U + W)/W$.

Theorem 10.2.6: Third isomorphism theorem

Let V be a vector space, W be a subspace of V and U be a subspace of W. Observe that $W/U \subseteq V/U$. The following map F:

$$V/U \to V/W$$
$$[\vec{v}]_U \mapsto [\vec{v}]_W$$

is linear, surjective and has kernel W/U.

Proof Note that the map is well-defined. For $\vec{v}, \vec{v'} \in x \in V/U$:

$$F(\vec{v}) - F(\vec{v'}) = [\vec{v}]_W - [\vec{v'}]_W = [\vec{v} - \vec{v'} \in U \subseteq W]_W = [\vec{0}] \implies F(\vec{v}) = F(\vec{v'})$$

We will again skip the proof for linearity. For $x = [\vec{v}]_W \in V/W$, $F([\vec{v}]_V) = x$, showing surjectivity. It is trivial that $W/U \subseteq \ker F$. In reverse:

$$F([\vec{v}]_U) = [\vec{0}]_W \implies \vec{v} - \vec{0} = \vec{v} \in W \implies [\vec{v}]_W \in W/U \implies \ker F \subseteq W/U$$

Again, invoking the first isomorphism theorem, we get $(V/U)/(W/U) \cong V/W$.

Theorem 10.2.7: Fourth isomorphism theorem

If U is a subspace of a vector space V, then there is a bijection between the set of subspaces of V/U and the set of subspaces of V that of which U is a subspace.

Proof For any subspace $X \subseteq V/U$, join the equivalence classes in X. The output must be a subspace of V by the subspace test, while U is given by $[0]$. Name this map f.
The inverse of the above map f^{-1} is to collect the equivalence classes of all vectors in the given subspace of V. Again, by the subspace test, we can verify the output is indeed a subset of V/U.
If $W \supseteq U$ is a subspace of V, for any $\vec{w} \in W$, obviously $\vec{w} \in f \circ f^{-1}(\vec{w}) = [\vec{w}]$, so $W \subseteq f \circ f^{-1}(W)$. On the other hand, $[\vec{w}] = \vec{w} + U \subseteq W$, so $f \circ f^{-1}(W) \subseteq W$. In other words, the two maps are mutually inverse, meaning they are both injections between the two sets.

After isomorphism theorems, we are going to connect quotient spaces with various mathematical objects we have introduced before.

Proposition 10.2.2: Basis extension theorem for quotient

Suppose V is a finitely dimensional vector space with a subspace U and $\{\vec{u_1}, \vec{u_2}, ..., \vec{u_n}\}$ is a basis for U. If $\{[\vec{v_1}], [\vec{v_2}], ..., [\vec{v_m}]\}$ is a basis for V/U, then $B = \{\vec{u_1}, ..., \vec{u_n}, \vec{v_1}, ..., \vec{v_m}\}$ is a basis for V.

Proof If we let $W = \operatorname{span}\{\vec{v_1}, ..., \vec{v_m}\}$, for $\vec{v} = \sum r_i \vec{v_i} \in U \cap W$:

$$\theta(\vec{v}) = \theta\left(\sum r_i \vec{v_i}\right) \implies \sum r_i [\vec{v_i}] = [\vec{0}] \implies r_i = 0$$

By Proposition 3.5.4, $V + W$ is a direct sum. Furthermore, We have shown in the proof of Theorem 10.2.3 that $\vec{v_1}, ..., \vec{v_m}$ are linearly independent. Therefore, by Theorem 10.2.3, the dimensional formula and the basis theorem, $V = U \oplus W$, or equivalently, B is a basis for V.

Proposition 10.2.3: Quotient of inner product space

Let U be a subspace of an inner product space V. Define an inner product $\langle\rangle_{V/U}$ on V/U as follows:

$$\langle [\vec{v}], [\vec{w}] \rangle_{V/U} = \langle \vec{v} - \operatorname{pr}_U(\vec{v}), \vec{w} - \operatorname{pr}_U(\vec{w}) \rangle_V$$

Then $U^\perp$ is isometrically isomorphic to V/U.

Proof It is apparent the required map is the quotient map θ with its domain restricted to $U^\perp$, which obviously preserves the inner product. As $U^\perp \cup U = \{\vec{0}\}$ (Proposition 6.2.5), by Proposition 10.2.1, the map is both surjective and injective.

If we denote the above map by $\tilde{\theta}$, we have the following sequence:

$$\begin{tikzcd} V \arrow[r, leftrightarrow, "\chi"] & U \times V/U \arrow[r, two heads, "\pi_{V/U}"] & V/U \arrow[r, leftrightarrow, "\tilde{\theta}^{-1}"] & U^\perp \arrow[r, hook, "i"] & V \end{tikzcd}$$

($\hookrightarrow$ denotes an injection, $\twoheadrightarrow$ denotes a surjection, $\leftrightarrow$ denotes a bijection.)
, which is just the orthogonal projection map pr_U.

The two results allows us to re-prove Schur triangulation in the language of quotient spaces:
Suppose ϕ is an endomorphism of a vector space V, and $U \subseteq V$ is ϕ-invariant. We try to find a matrix representation of ϕ that involves $\overline{\phi}$, which is an endomorphism of V/U. Let $B = \{\vec{u_1}, \vec{u_2}, ..., \vec{u_n}\}$ and $B' = \{[\vec{v_1}], [\vec{v_2}], ..., [\vec{v_m}]\}$ be bases for U and V/U respectively, and $S = \{\vec{v_1}, ..., \vec{v_m}\}$. By the basis extension theorem, $B \cup S$ is a basis for V. Note that for $\vec{v} \in S$:

$$\phi(\vec{v}) = \sum_{i=1}^{n} r_i\vec{u_i} + \sum_{i=1}^{m} s_i\vec{v_m} \implies \overline{\phi}([\vec{v}]) = \theta\left(\sum_{i=1}^{n} r_i\vec{u_i}\right) + \theta\left(\sum_{i=1}^{m} s_i\vec{v_m}\right) = \sum_{i=1}^{m} s_i[\vec{v_i}]$$

Hence, as U is ϕ-invariant:

$$M_{B\cup S}(\phi) = \left(\begin{array}{c:c} M_B(\phi|_U) & * \\ \hdashline \mathbf{0} & M_{B'}(\overline{\phi}) \end{array}\right)$$

($\mathbf{0}$ denotes a zero block of suitable size.)
Now suppose the characteristic polynomial of ϕ can be linearly factorised. Then the theorem can be proved via induction. As the base case is trivial, we will focus on the inductive case: Due to the extra condition, we must be able to find an eigenvalue λ of ϕ. Let U be the eigenspace of λ, which is clearly ϕ-invariant. By the induction assumption, we can find a basis $B' = \{[\vec{v_1}], [\vec{v_2}], ..., [\vec{v_m}]\}$ for V/U such that $M_{B'}(\overline{\phi})$ is triangular. Find a basis B for U. Then by the above, $B \cup \{\vec{v_1}, ..., \vec{v_m}\}$ is a basis for V and the matrix of ϕ with respect to it is triangular, completing our proof.
If an inner product is defined on V, by Proposition 10.2.3, an inner product can be defined on V/U such that θ is an isometry. Then via the Gram-schimdt process, we can turn $B \cup \{\vec{v_1}, ..., \vec{v_m}\}$ in the above proof to an orthonormal basis. This allows us to prove the unitary version of the theorem (Theorem 9.3.2).

Theorem 10.2.8

Suppose V is a vector space and U is its subspace. Then $(V/U)^* \cong U^0$.

Proof Consider the following map F:

$$(V/U)^* \to U^0$$
$$f \mapsto \overline{f}$$

($\overline{f}$ is defined in the universal property for quotient.)
The linearity of F is trivial. Note that $\overline{f} \in U^0$ because for $\vec{u} \in U$, $\overline{f}(\vec{u}) = f([\vec{u}]) = f([\vec{0}]) = 0$. We now check if F is isomorphic:
Injectivity For $\vec{v} \in V$, $\overline{f}(\vec{v}) = 0 \implies f([\vec{v}]) = 0$. Therefore, by definition, if $\overline{f}$ is the zero map, f is also the zero map. In other words, $\ker F = O$, proving its injectivity.
Surjectivity For any $g \in U^0$, the obvious way to construct a f such that $F(f) = g$ is to define f such that $\overline{f}(\vec{v}) = f([\vec{v}]) = g(\vec{v})$. However, we have to check if such f is well-defined. For $\vec{v_1}, \vec{v_2} \in x \in V/U$:

$$g(\vec{v_1}) - g(\vec{v_2}) = g(\vec{v_1} - \vec{v_2} \in U) = 0 \implies g(\vec{v_1}) = g(\vec{v_2})$$

Therefore, such a construction is valid and F is surjective.

Proposition 10.2.4

Suppose V is a vector space and U is its subspace. Then $U^* \cong V^*/U^0$.

Proof Consider the inclusion map $i : U \to V$, which is an injection. By Proposition 5.3.3, its dual $i^* : V^* \to U^*$ is a surjection. Recall that for $f \in V^*$, $i^*(f) = f \circ i$. It should be apparent then that $\ker i^* = U^0$. Therefore, by the first isomorphism theorem, $V^*/U^0 \cong U^*$ and the natural isomorphism is the induced map of i^*.

Proposition 10.2.5

Suppose W is a subspace of V^*, where V is a finitely dimensional vector space. Then there exists a subspace U of V such that $U^0 = W$.

Proof Recall the evaluation map e. Let $U = e^{-1}(W^0)$. Then by Theorem 5.5.1:

$$\begin{aligned} W^0 &= \{F \mid \forall f \in W : F(f) = 0\} \subseteq V^{**} \\ \cong U &= \{\vec{v} \mid F \in W^0, \forall f \in V^* : F(f) = f(\vec{v})\} \subseteq V \\ &= \{\vec{v} \mid F \in V^{**}, \forall f \in V^* : F(f) = f(\vec{v}), \forall f \in W : f(\vec{v}) = \vec{0}\} \\ &= \{\vec{v} \mid F \in V^{**}, \forall f \in V^* : F(f) = f(\vec{v})\} \cap \{\vec{v} \mid \forall f \in W : f(\vec{v}) = \vec{0}\} \\ &= V \cap \{\vec{v} \mid \forall f \in W : f(\vec{v}) = \vec{0}\} \\ &= \{\vec{v} \mid \forall f \in W : f(\vec{v}) = \vec{0}\} \end{aligned}$$

Therefore, $W \subseteq U^0$. Next, consider that:

$$\begin{aligned} W^* &\cong V^{**}/W^0 \text{ (Proposition 10.2.4)} \\ &\cong V^{**} \times W^0 \text{ (Theorem 10.2.4)} \\ &\cong V \times U \text{ (Theorem 5.5.1)} \\ &\cong V/U \text{ (Theorem 10.2.4)} \\ \implies W &\cong (V/U)^* \cong U^0 \text{ (Proposition 5.4.2, Theorem 10.2.8)} \end{aligned}$$

Hence, $W = U^0$.