Honors in Linear and Abstract Algebra I

Lecture Notes for
MATH 2131

Department of Mathematics
Hong Kong University of Science and Technology

December 18, 2025

Prefaces

These lecture notes were written by a student in the course MATH 2131 – Honors in Linear and Abstract Algebra by Professor Meng Guowu at HKUST in Autumn 2025-26.
All diagrams in these lecture notes are written in LaTeX TikZ code.
The notes reference the textbooks *Linear Algebra* by Friedberg, Insel and Spence, *Abstract Algebra* by Artin and *A First Course in Abstract Algebra* by Fraleigh. Additionally, the notes reference lecture notes from two other professors who taught this course previously: Professor Ivan Ip and Professor Min Yan.

Contents

List of Symbols

Symbols	Meaning
$\mathbb{F}$	a field
U, V, W	vector spaces
α, β	elements in $\mathbb{F}$
$\mathbb{F}^n$	the set of all column matrices with n entries in $\mathbb{F}$
$(\mathbb{F}^n)^*$	the set of all row matrices with n entries in $\mathbb{F}$
$\mathbb{F}[X]$	the polynomial ring
$\mathbb{F}[[X]]$	the formal power series ring
$\mathcal{C}, \mathcal{D}$	categories
$\mathbf{Set}$	the category of sets
$\mathbf{Vec}_{\mathbb{F}}$	the category of vector spaces over a field $\mathbb{F}$
0_V	additive identity of vector space V
1_V	multiplicative identity of vector space V
$\subset$	proper subset
$\subseteq$	subset, i.e. can be equal
$\imath$	Inclusion map
$\hookrightarrow$	Injective arrow
π	Projection map
$\twoheadrightarrow$	Surjective arrow
S, T	Linear maps
A, B	Matrices
$\mathsf{Mor}_{\mathcal{C}}(V, W)$	the set of all morphisms from V to W in category $\mathcal{C}$
$\mathsf{Hom}(V, W)$	Hom-set of V to W
$\mathsf{End}(A)$	Endomorphism ring of A
$\mathsf{M}_{m\times n}(\mathbb{F})$	the set of all $m \times n$ matrices over $\mathbb{F}$
$\vec{x}$	column vector with entries x_i
$\hat{x}$	row vector with entries x_i
$\vec{e}_i$	column vector with only 1 at the i-th row and 0 at other places
$\hat{e}_i$	row vector with only 1 at the i-th column and 0 at other places
$\alpha\cdot$	a map that performs scalar multiplication
$A\cdot$	a map that performs matrix multiplication
δ_x	the Kronecker delta function
δ_X	the set of all Kronecker delta functions
δ_{ij}	the Kronecker delta symbol
$\mathsf{Ker}(T)$	Kernel of linear map T
$\mathsf{Im}(T)$	Image of linear map T
$\mathsf{Coker}(T)$	Cokernel of linear map T
$\mathsf{Coim}(T)$	Coimage of linear map T
$\mathsf{Span}(S)$	Span of a set of vectors S
$\prod$	Product
$\coprod$	Coproduct
$\oplus$	Direct sum
$\otimes$	Tensor product

Symbols	Meaning
$\mathcal{T}^{\bullet}$	Tensor algebra
V^*	Dual space of V
V^{**}	Double dual space of V
$D(V)$	Double
$\mathsf{id}_{\mathcal{C}}$	Identity functor in category $\mathcal{C}$
$\mathbb{F}[-]$	Free vector space functor
$\lvert - \rvert$	Forgetful functor
$(-)^*$	Dual space functor

1. Abstract Linear Spaces

"I assume you have learnt linear algebra."

GUOWU MENG

1.1 Binary Operation

We start with the definition of a binary operation.

Definition 1.1 — Binary Operation. A *binary operation* on a set S is a mapping of the elements of the Cartesian product $S \times S$ to S.

$$\begin{aligned} \cdot : S \times S &\to S \\ (x, y) &\mapsto x \cdot y \end{aligned}$$

For ease of understanding, a binary operation is combining two objects into one. Hence, there is something called unary and ternary operations, corresponding to the action of combining one and three objects into one respectively.

■ **Example 1.1** A common example of a binary operation is addition on the set of natural numbers $\mathbb{N}$.

$$\begin{aligned} + : \mathbb{N} \times \mathbb{N} &\to \mathbb{N} \\ (x, y) &\mapsto x + y \end{aligned} \tag{1.1}$$

■

Definition 1.2 — Associative. A binary operation $\cdot : S \times S \to S$ is said to be *associative* if, for all $x, y, z \in S$,

$$x \cdot (y \cdot z) = (x \cdot y) \cdot z$$

■ **Example 1.2** A common example of an associative (binary) operation is addition on the set of natural numbers $\mathbb{N}$. For all $x, y, z \in \mathbb{N}$, we have $x + (y + z) = (x + y) + z$. ■

Definition 1.3 — Identifiable. A binary operation $\cdot : S \times S \to S$ is said to be *identifiable*, or *unital*, if there exists an element $e \in S$, the *identity* or *unit element*, such that, for all $x \in S$

$$e \cdot x = x = x \cdot e$$

■ **Example 1.3** A common example of an identifiable (binary) operation is multiplication on the set of natural numbers $\mathbb{N}$. The identity element is 1, and for all $x \in \mathbb{N}$, we have $x \cdot 1 = x = 1 \cdot x$. ■

Proposition 1.1 The identity element of an identifiable operation is unique.

Proof. Let e_1 and e_2 be two identity elements for the operation $\cdot$. Then, for any element $x \in S$, we have:

$$e_1 \cdot x = x = x \cdot e_1$$

$$e_2 \cdot x = x = x \cdot e_2$$

Now, consider the element e_1: $e_1 \cdot e_2 = e_1$. But since e_2 is an identity element, we also have: $e_1 \cdot e_2 = e_2$. Therefore, we conclude that $e_1 = e_2$, proving the uniqueness of the identity element. ■

Note that the two-sided identity must be unique, but one-sided identities need not be. The following is an example of it.

■ **Example 1.4** Consider a set $X = \left\{ \begin{bmatrix} 1 & a \\ 0 & 0 \end{bmatrix} \middle| \, a \in \mathbb{R} \right\}$ with the binary operation defined as matrix multiplication. This set has many left identity elements, but no two-sided identity element. ■

Definition 1.4 — Invertible. A binary operation $\cdot : S \times S \to S$ is said to be *invertible* if, for every element $x \in S$, there exists an element $y \in S$, called the two-sided *inverse* of x, denoted as x^{-1}, such that

$$x \cdot y = e = y \cdot x$$

where e is the identity element of the operation.

Remark. An invertible operation must be identifiable, since the identity element is required in the definition of invertibility.

■ **Example 1.5** A common example of an invertible (binary) operation is addition on the set of integers $\mathbb{Z}$. For every integer $x \in \mathbb{Z}$, there exists an integer $y = -x$ such that:

$$x + (-x) = 0 = (-x) + x \tag{1.2}$$

where 0 is the identity element for addition. ■

Proposition 1.2 The inverse element of an invertible operation is unique.

Proof. Let y_1 and y_2 be two inverses of an element $x \in S$. Then, by definition of inverse, we have:

$$x \cdot y_1 = e = y_1 \cdot x$$

$$x \cdot y_2 = e = y_2 \cdot x$$

Now, consider the element y_1: $y_1 \cdot x = e$. But since y_2 is also an inverse of x, we can substitute e with $x \cdot y_2$: $y_1 \cdot x = x \cdot y_2 = e$. By the associativity of the operation, we can rearrange this to:

$$y_1 = y_1 \cdot e = y_1 \cdot (x \cdot y_2) = (y_1 \cdot x) \cdot y_2 = e \cdot y_2 = y_2$$

Thus, the inverse element is unique. ■

The same applies to the inverse; one-sided inverses need not be unique. The example is left as an exercise.

Definition 1.5 — Commutative. A binary operation $\cdot : S \times S \to S$ is said to be *commutative* if, for all $x, y \in S$, the following holds:

$$x \cdot y = y \cdot x$$

■ **Example 1.6** A common example of a commutative operation is addition on the set of integers $\mathbb{Z}$. For all $x, y \in \mathbb{Z}$, we have: $x + y = y + x$ ■

Definition 1.6 — Distributive (Harmonic). A binary operation $\cdot : S \times S \to S$ is said to be *distributive* with respect to another binary operation $+ : S \times S \to S$ if, for all $x, y, z \in S$, the following holds:

$$x \cdot (y + z) = x \cdot y + x \cdot z$$
$$(y + z) \cdot x = y \cdot x + z \cdot x$$

The professor prefers to use the word "harmonic" instead of "distributive". Note that it is important to show that "*which binary operation* is distributive to *which binary operation*". (The two binary operations in this sentence are not commutative.)

■ **Example 1.7** A common example of a distributive operation is multiplication over addition on the set of integers $\mathbb{Z}$. For all $x, y, z \in \mathbb{Z}$, we have:

$$x \cdot (y + z) = x \cdot y + x \cdot z$$
$$(y + z) \cdot x = y \cdot x + z \cdot x$$

■

1.2 Groups, Rings, Fields

With those five properties, we can construct monoids and groups.

Definition 1.7 — Monoid. A *monoid* is a set M equipped with a binary operation $f : M \times M \to M$ having the following properties:

1. *Associative*
2. *Identifiable*

We say (M, f) is a monoid, and f is the *monoid operation* on the set M. A set M with a monoid operation f is the *monoid structure*.

Definition 1.8 — Group. A *group* is a set G equipped with a monoid operation $f : G \times G \to G$ with the additional property that every element has an inverse.

■ **Example 1.8** $(\mathbb{R} \setminus \{0\}, \times)$ is a group, but $(\mathbb{R}, \times)$ is not a group since 0 does not have a multiplicative inverse. ■

Definition 1.9 — Abelian Monoid / Group. A monoid / group (S, f) is said to be an *abelian* if the operation f is commutative.

Definition 1.10 — Unital Ring. A *unital ring* is a set R equipped with two binary operations $f : R \times R \to R$ (addition) and $g : R \times R \to R$ (multiplication) such that the following properties hold:

1. *Additive Group:* (R, f) is an abelian group.
2. *Multiplicative Monoid:* (R, g) is a monoid.
3. *Distributive Property:* g with respect to f.

Definition 1.11 — Commutative Ring. A *commutative ring* is a unital ring R such that the multiplication operation $g : R \times R \to R$ is commutative.

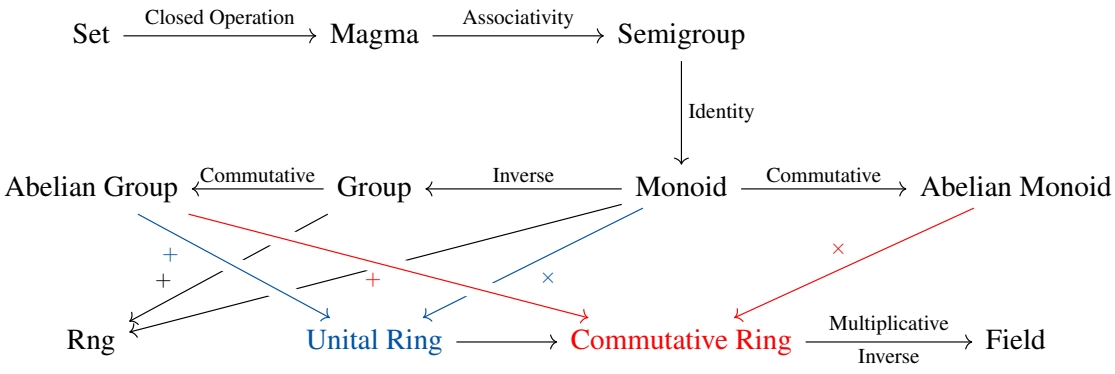
■ **Example 1.9** $(\mathbb{Z}, +, \times)$ is a commutative ring. ■

Definition 1.12 — Field. A *field* is a commutative ring $\mathbb{F}$ such that every non-zero element has a multiplicative inverse.

■ **Example 1.10** $(\mathbb{Q}, +, \times)$, $(\mathbb{R}, +, \times)$ and $(\mathbb{C}, +, \times)$ are fields. ■

■ **Example 1.11 — Finite Field.** $(\mathbb{Z}/2\mathbb{Z}, +, \times)$ is a field, where $\mathbb{Z}/2\mathbb{Z} = \{[0], [1]\}$, $[0]$ is the set of even integers and $[1]$ is the set of odd integers. Note that any $\mathbb{Z}/p\mathbb{Z}$ is a finite field, where p is a prime number. ■

We may draw a diagram for the relationship between the algebraic structures.



1.3 Morphisms

Normally, when we have two sets we can have a set map. What if the two are in the same algebraic structures? They are called the homomorphisms.

Definition 1.13 — Monoid Homomorphism. A *monoid homomorphism* is a morphism between two monoids that preserves the monoid structure. Formally, let $(M_1, \cdot_1)$ and $(M_2, \cdot_2)$ be two monoids with identity elements e_1 and e_2, respectively. A function $f : M_1 \to M_2$ is a monoid homomorphism if:

1. $f(x \cdot_1 y) = f(x) \cdot_2 f(y) \quad \forall x, y \in M_1$
2. $f(e_1) = e_2$

Definition 1.14 — Group Homomorphism. A *group homomorphism* is a morphism between two groups that preserves the group structure. Formally, let $(G_1, \cdot_1)$ and $(G_2, \cdot_2)$ be two groups with identity elements e_1 and e_2, respectively. A function $f : G_1 \to G_2$ is a group homomorphism if:

1. $f(x \cdot_1 y) = f(x) \cdot_2 f(y) \quad \forall x, y \in G_1$
2. $f(e_1) = e_2$
3. $f(x^{-1}) = (f(x))^{-1} \quad \forall x \in G_1$

Proposition 1.3 The second and third properties of a group homomorphism are consequences of the first property.

Proof. Let $f : G_1 \to G_2$ be a group homomorphism satisfying the first property.

Second Property: For any element $x \in G_1$, we have:

$$f(x) = f(x \cdot_1 e_1) = f(x) \cdot_2 f(e_1)$$

So for any $f(x) \in G_2$, this implies that $f(e_1)$ must be the identity element in G_2, i.e., $f(e_1) = e_2$.

Third Property: We have:

$$e_2 = f(e_1) = f(x \cdot_1 x^{-1}) = f(x) \cdot_2 f(x^{-1})$$

This shows that $f(x^{-1})$ is the inverse of $f(x)$ in G_2, i.e., $f(x^{-1}) = (f(x))^{-1}$. ■

For monoid homomorphisms, the second property cannot be derived from the first property. Consider the identity element e_1 in M_1. If we apply the first property, we get $f(e_1 \cdot_1 e_1) = f(e_1) \cdot_2 f(e_1)$. This simplifies to $f(e_1) = f(e_1) \cdot_2 f(e_1)$, which does not necessarily imply that $f(e_1)$ is the identity element in M_2, i.e., $f(e_1) \neq e_2$, but $f(e_1)$ is the idempotent element in M_2. Therefore, the second property must be explicitly stated for monoid homomorphisms.

However in the case of group homomorphisms, the existence of inverses ensures that there is only one element that can be idempotent under the group operation, which is the identity element. Thus, for group homomorphisms, the second property can be derived from the first property.

Definition 1.15 — Idempotent Elements. An element a is said to be *idempotent* if $a = a^2$.

To introduce the vector space, the following two morphisms are required.

Definition 1.16 — Ring Homomorphism. A *ring homomorphism* is a morphism between two rings that preserves both the additive and multiplicative structures. Formally, let $(R_1, +_1, \cdot_1)$ and $(R_2, +_2, \cdot_2)$ be two rings with identity elements 0_1, 1_1 and 0_2, 1_2, respectively. A function $f : R_1 \to R_2$ is a ring homomorphism if:

1. $f(x +_1 y) = f(x) +_2 f(y) \quad \forall x, y \in R_1$
2. $f(x \cdot_1 y) = f(x) \cdot_2 f(y) \quad \forall x, y \in R_1$
3. $f(1_1) = 1_2$

Definition 1.17 — Endomorphism. An *endomorphism* is a morphism from an algebraic structure to itself. Formally, let $(A, \cdot)$ be an algebraic structure. An endomorphism $f : A \to A$ is a set map such that:

$$f(x \cdot y) = f(x) \cdot f(y) \quad \forall x, y \in A$$

The following two sets are the sets of all structure-preserving maps.

Definition 1.18 — Hom-set. The set of all morphisms from an algebraic structure A to another algebraic structure B is called the *hom-set*, denoted by $\mathsf{Hom}(A, B)$.

Definition 1.19 — Endomorphism Ring. The set of all endomorphisms of an abelian group $(A, +)$, denoted by $\mathsf{End}(A)$, forms a (non-commutative) ring under pointwise addition and composition of set maps. The addition and multiplication operations are defined as follows:

$$\begin{aligned} + : \mathsf{End}(A) \times \mathsf{End}(A) &\to \mathsf{End}(A) \\ (f, g) &\mapsto (f + g : x \mapsto f(x) + g(x)) \qquad f + g : A \to A \end{aligned}$$

$$\begin{aligned} \circ : \mathsf{End}(A) \times \mathsf{End}(A) &\to \mathsf{End}(A) \\ (f, g) &\mapsto (f \circ g : x \mapsto f(g(x))) \qquad f \circ g : A \to A \end{aligned}$$

The identity element for addition is the zero endomorphism, which maps every element to the identity element of the group.

$$\begin{aligned} 0 : A &\to A \\ x &\mapsto 0 \end{aligned}$$

The identity element for multiplication is the identity endomorphism, which maps every element to itself.

$$\begin{aligned} 1 : A &\to A \\ x &\mapsto x \end{aligned}$$

Note that all endomorphisms in $\mathsf{End}(A)$ are group homomorphisms and $\mathsf{End}(A) = \mathsf{Hom}(A, A)$.

1.4 Linear Spaces

Then we can define what a linear structure is.

Definition 1.20 — Linear Structure. A *linear structure* over a field $\mathbb{F}$ on a set V is a pair $(+,\cdot)$ where $(V,+)$ is an abelian group with a ring homomorphism $\mathbb{F} \to \mathsf{End}(V)$, where $\mathsf{End}(V)$ is the endomorphism ring of the abelian group $(V,+)$.

$$\begin{aligned} &\cdot : \mathbb{F} \to \mathsf{End}(V) \\ &\quad \alpha \mapsto (\alpha\cdot : \vec{x} \mapsto \alpha\vec{x}) \qquad \alpha\cdot : V \to V \end{aligned}$$

The ring homomorphism is a (ring) action of the field $\mathbb{F}$ on the abelian group $(V,+)$, called *scalar multiplication*. The ring action can be written as a binary operation:

$$\begin{aligned} &\cdot : \mathbb{F} \times V \to V \\ &\quad (\alpha,\vec{x}) \mapsto \alpha\vec{x} \end{aligned}$$

A linear space / vector space is a set with a linear structure over a field on the set. In normal textbook, a linear space will be defined as follows:

Corollary 1.1 — Linear Spaces. A linear space over a field $\mathbb{F}$ is a set V equipped with two operations: vector addition $+ : V \times V \to V$ and scalar multiplication $\cdot : \mathbb{F} \times V \to V$, satisfying the following axioms for all $\vec{u},\vec{v},\vec{w} \in V$ and $\alpha,\beta \in \mathbb{F}$:

Axiom	Statement
1. Associativity of addition	$(\vec{u}+\vec{v})+\vec{w} = \vec{u}+(\vec{v}+\vec{w})$
2. Existence of additive identity	$\exists\vec{0} \in V$ such that $\forall\vec{u} \in V$, $\vec{u}+\vec{0} = \vec{u}$
3. Existence of additive inverses	$\forall\vec{u} \in V$, $\exists -\vec{u} \in V$ such that $\vec{u}+(-\vec{u}) = \vec{0}$
4. Commutativity of addition	$\vec{u}+\vec{v} = \vec{v}+\vec{u}$
5. Distributivity of scalar multiplication with respect to vector addition	$\alpha(\vec{u}+\vec{v}) = \alpha\vec{u}+\alpha\vec{v}$
6. Distributivity of scalar multiplication with respect to field addition	$(\alpha+\beta)\cdot = \alpha\cdot+\beta\cdot$
7. Compatibility of scalar multiplication with field multiplication	$(\alpha\beta)\cdot = (\alpha\cdot)\circ(\beta\cdot)$
8. Identity element of scalar multiplication	$\mathbb{F} \ni 1 \mapsto (1\cdot : x \mapsto x) \in \mathsf{End}(V)$

Remark. The first four axioms ensure that $(V,+)$ is an abelian group, while the fifth axiom describes the distributivity inside $\mathsf{End}(A)$ and the last three axioms describe the ring homomorphism.

■ **Example 1.12** $\mathbb{F}$ is a linear space over itself with the usual addition and multiplication operations.

$$\begin{aligned} &\cdot : \mathbb{F} \times \mathbb{F} \to \mathbb{F} \\ &\quad (\alpha,\beta) \mapsto \alpha\beta \end{aligned}$$

The first $\mathbb{F}$ is the field acting on the second $\mathbb{F}$, which is the abelian group. ■

■ **Example 1.13** Let X be a set and $\mathbb{F}$ be a field. (f is a set map)

$$\begin{aligned} \mathbb{F}[[X]] = \mathsf{Map}(X,\mathbb{F}) &\overset{\text{def}}{=\!=} \text{the set of all } \mathbb{F}\text{-valued functions on } X \\ &=\!= \{f : X \to \mathbb{F}\} \end{aligned}$$

$\mathbb{F}[[X]]$ is a linear space over $\mathbb{F}$ with the following operations defined pointwisely:

$$\begin{aligned} +&: \mathbb{F}[[X]] \times \mathbb{F}[[X]] \to \mathbb{F}[[X]] \\ &\qquad (f,g) \mapsto (f+g : x \mapsto f(x)+g(x)) \qquad f+g : X \to \mathbb{F} \end{aligned}$$

$$\begin{aligned} \cdot&: \mathbb{F} \times \mathbb{F}[[X]] \to \mathbb{F}[[X]] \\ &\qquad (\alpha,f) \mapsto (\alpha f : x \mapsto \alpha f(x)) \qquad \alpha f : X \to \mathbb{F} \end{aligned}$$

■

■ **Example 1.14** Let X be a set and $\mathbb{F}$ be a field.

$$\begin{aligned} \mathbb{F}[X] = \mathrm{Map}_{\mathrm{fin}}(X,\mathbb{F}) &\stackrel{\mathrm{def}}{=\!=} \text{the set of all finitely supported } \mathbb{F}\text{-valued functions on } X \\ &=\!= \{f : X \to \mathbb{F} \mid f \text{ is finitely supported}\} \end{aligned}$$

$\mathbb{F}[X]$ is a linear space over $\mathbb{F}$ as $\mathbb{F}[X] \subseteq \mathbb{F}[[X]]$ and the operations are defined pointwisely as in the previous example.

$f : X \to \mathbb{F}$ is finitely supported if the set $\{x \in X \mid f(x) \neq 0\}$ is finite or $f(x) \neq 0$ for only finitely many $x \in X$. ■

■ **Example 1.15** Let t be a formal variable. Then $\mathbb{F}[[t]] \stackrel{\mathrm{def}}{=\!=} \mathbb{F}[[\{1,t,t^2,\cdots\}]] = \sum_{n=0}^{\infty} a_n t^n$ is the set of all formal power series in t with coefficients in $\mathbb{F}$ and $\mathbb{F}[t] \stackrel{\mathrm{def}}{=\!=} \mathbb{F}[\{1,t,t^2,\cdots\}] = \sum_{n=0}^{N} a_n t^n$ is the set of all polynomials in t with coefficients in $\mathbb{F}$. Both $\mathbb{F}[[t]]$ and $\mathbb{F}[t]$ are linear spaces over $\mathbb{F}$. ■

There are other names for $\mathbb{F}[X]$ and $\mathbb{F}[[X]]$: Polynomial ring and Formal Power Series ring, respectively.

■ **Example 1.16** Let n be a positive integer and $\mathbb{F}$ be a field. Then

$$\mathbb{F}^n \stackrel{\mathrm{def}}{=\!=} \left\{ \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix} \;\middle|\; c_i \in \mathbb{F} \right\}$$

is the set of all *column matrices* with n entries in $\mathbb{F}$. Elements in $\mathbb{F}^n$ are written as $\vec{x}$ and are called *column vectors*. $\mathbb{F}^n$ is a linear space over $\mathbb{F}$ with the following operations defined entrywisely:

$$\begin{aligned} +&: \mathbb{F}^n \times \mathbb{F}^n \to \mathbb{F}^n \\ &\qquad (\vec{a},\vec{b}) \mapsto \vec{a}+\vec{b} = \begin{bmatrix} a_1+b_1 \\ \vdots \\ a_n+b_n \end{bmatrix} \end{aligned}$$

$$\begin{aligned} \cdot&: \mathbb{F} \times \mathbb{F}^n \to \mathbb{F}^n \\ &\qquad (\alpha,\vec{a}) \mapsto \alpha\vec{a} = \begin{bmatrix} \alpha a_1 \\ \vdots \\ \alpha a_n \end{bmatrix} \end{aligned}$$

$\mathbb{F}^n$ is a linear space over $\mathbb{F}$ automatically as $\mathbb{F}$ is a linear space over itself. ■

2. Linear Maps and Matrices

"Linear algebra is the easiest in Mathematics"

GUOWU MENG

2.1 Linear Maps

Linear map, sometimes linear transformation, is a homomorphism preserving linear structure.

Definition 2.1 — Linear Maps. Let V and W be two linear spaces over a field $\mathbb{F}$. A *linear map* is a set map $T : V \to W$ such that for all $u, v \in V$ and $\alpha \in \mathbb{F}$, the following holds:

$$\begin{aligned} T(u+v) &= T(u) + T(v) \\ T(\alpha u) &= \alpha T(u) \end{aligned}$$

The set of all linear maps from V to W is denoted by $\mathsf{Hom}(V,W)$. Some may write $\mathcal{L}(V,W)$.

Definition 2.2 — Linear Combinations. Let V be a linear space over a field $\mathbb{F}$. A *linear combination* of vectors $v_1, v_2, \cdots, v_n \in V$ is a vector of the form:

$$\alpha^1 v_1 + \alpha^2 v_2 + \cdots + \alpha^n v_n$$

where $\alpha^1, \alpha^2, \cdots, \alpha^n \in \mathbb{F}$ are scalars.

The reason for using the superscript for scalars is to avoid confusion with the subscript of vectors. Also, it is due to the concept of dual space, which will be introduced later.

We can combine the two properties of linear maps into one property.

Corollary 2.1 — Linear Maps and Linear Combinations. A set map $f : V \to W$ between two linear spaces over a field $\mathbb{F}$ is a linear map if and only if T respects linear combinations, i.e., for all $v_1, v_2 \in V$ and all scalars $\alpha^1, \alpha^2 \in \mathbb{F}$, the following holds:

$$T(\alpha^1 v_1 + \alpha^2 v_2) = \alpha^1 T(v_1) + \alpha^2 T(v_2)$$

■ **Example 2.1** Let A be an $m \times n$ matrix with entries in a field $\mathbb{F}$. The map $T : \mathbb{F}^n \to \mathbb{F}^m$ defined by

$$Tx = T(x) = Ax$$

where right-hand side is the usual matrix multiplication, is a linear map over $\mathbb{F}$. ■

Proposition 2.1 A linear map $T : \mathbb{F}^n \to \mathbb{F}^m$ is a matrix multiplication by a unique $m \times n$ matrix A with entries in $\mathbb{F}$. The matrix A is called the *standard matrix* of the linear map T.

$$\begin{array}{ccc} \mathsf{Hom}(\mathbb{F}^n, \mathbb{F}^m) & \overset{\text{natural}}{\underset{\text{identification}}{=\!=}} & \mathsf{M}_{m\times n}(\mathbb{F}) \\ T & \longmapsto & A \\ A\cdot & \longleftarrow & A \end{array}$$

where $A\cdot : \vec{x} \mapsto A\vec{x}$ and A can be expressed as follows:

$$A = \begin{bmatrix} | & | & & | \\ T\vec{e}_1 & T\vec{e}_2 & \cdots & T\vec{e}_n \\ | & | & & | \end{bmatrix}$$

The vector $\vec{e}_i$ is the column vector which has only the value 1 at the i-th position and 0 elsewhere.

Proof. Consider a column matrix $x \in \mathbb{F}^n$ with entries $x^1, x^2, \cdots, x^n \in \mathbb{F}$. Then x can be expressed as a linear combination of the vectors $\vec{e}_1, \vec{e}_2, \cdots, \vec{e}_n$:

$$x = x^1\vec{e}_1 + x^2\vec{e}_2 + \cdots + x^n\vec{e}_n = \sum_{i=1}^{n} x^i\vec{e}_i$$

Since T is a linear map, it respects linear combinations. Therefore, we have:

$$Tx = T\left(\sum_{i=1}^{n} x^i\vec{e}_i\right) = \sum_{i=1}^{n} x^i T(\vec{e}_i) = \sum_{i=1}^{n} x^i\vec{a}_i = A\vec{x}$$

where $\vec{a}_i = T\vec{e}_i$ is the i-th column of the matrix $A = \begin{bmatrix} | & | & & | \\ T\vec{e}_1 & T\vec{e}_2 & \cdots & T\vec{e}_n \\ | & | & & | \end{bmatrix}$. Thus, we have $T\vec{x} = A\vec{x}$ for all $\vec{x} \in \mathbb{F}^n$. This shows that T can be represented as a matrix multiplication by the matrix A. ■

There is a simpler way to write $\sum_{i=1}^{n} x^i\vec{e}_i$: The Einstein Summation Convention. When an index variable appears twice in a single term and is not otherwise defined, it implies summation of that term over all the values of the index. Therefore, we can write:

$$x = x^i\vec{e}_i$$

where i is summed from 1 to n.

Definition 2.3 — Linear Functional / Homogeneous Linear Function. A linear map $f : \mathbb{F}^n \to \mathbb{F}$ is called a *homogeneous linear function* or a *linear functional* if for all $\alpha \in \mathbb{F}$ and $x \in \mathbb{F}^n$, the following holds:

$$f(\alpha x) = \alpha f(x)$$

Corollary 2.2 — Standard Matrix of a Linear Map. The standard matrix of a linear map $T : \mathbb{F}^n \to \mathbb{F}^m$ can be written as:

$$A = \begin{bmatrix} — & f_1 & — \\ — & f_2 & — \\ & \vdots & \\ — & f_m & — \end{bmatrix}$$

where $f_i : \mathbb{F}^n \to \mathbb{F}$ is the i-th component function of T, which is a linear functional.

■ **Example 2.2** Let $D : \mathbb{F}[t] \to \mathbb{F}[t]$ be the differentiation operator defined by:

$$D\left(\sum_{n=0}^{N} a_n t^n\right) = \sum_{n=1}^{N} n a_n t^{n-1}$$

for all polynomials $\sum_{n=0}^{N} a_n t^n \in \mathbb{F}[t]$. The differentiation operator D is a linear map over $\mathbb{F}$. The standard matrix of D with respect to the standard basis $\{1, t, t^2, \cdots, t^N\}$ of $\mathbb{F}[t]$ is given by:

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 2 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & N \\ 0 & 0 & 0 & 0 & \cdots & 0 \end{bmatrix}$$

■

Proposition 2.2 Let X be a set and W be a linear space over a field $\mathbb{F}$. Then the set of all set maps from X to W, denoted by $\mathsf{Map}(X, W)$, is a linear space over $\mathbb{F}$ with the following operations defined pointwisely:

$$\begin{aligned} + : \mathsf{Map}(X,W) \times \mathsf{Map}(X,W) &\to \mathsf{Map}(X,W) \\ (f,g) &\mapsto (f+g) : x \mapsto f(x) + g(x) \end{aligned}$$

$$\begin{aligned} \cdot : \mathbb{F} \times \mathsf{Map}(X,W) &\to \mathsf{Map}(X,W) \\ (\alpha, f) &\mapsto (\alpha f) : x \mapsto \alpha f(x) \end{aligned}$$

Proof. The $\mathsf{Map}(X,W)$ is defined pointwisely by $\mathbb{F}$, hence it is trivially a linear map. ■

Proposition 2.3 Let V and W be two linear spaces over a field $\mathbb{F}$. Then $\mathsf{Hom}(V,W)$ is a linear space over $\mathbb{F}$ with the following operations defined pointwisely:

$$\begin{aligned} + : \mathsf{Hom}(V,W) \times \mathsf{Hom}(V,W) &\to \mathsf{Hom}(V,W) \\ (f,g) &\mapsto (f+g) : v \mapsto f(v) + g(v) \end{aligned}$$

$$\begin{aligned} \cdot : \mathbb{F} \times \mathsf{Hom}(V,W) &\to \mathsf{Hom}(V,W) \\ (\alpha, f) &\mapsto (\alpha f) : v \mapsto \alpha f(v) \end{aligned}$$

Proof. Note that $\mathsf{Hom}(V,W) \subseteq \mathsf{Map}(V,W)$. We need to show that the operations defined above are closed in $\mathsf{Hom}(V,W)$, i.e., for all $f, g \in \mathsf{Hom}(V,W)$ and $\alpha \in \mathbb{F}$, $f+g \in \mathsf{Hom}(V,W)$ and $\alpha f \in \mathsf{Hom}(V,W)$ or equivalently, f respects linear combinations.

Let $\vec{u}, \vec{v} \in V$ and $\alpha, \beta \in \mathbb{F}$. Since $f, g \in \mathsf{Hom}(V, W)$, we have:

$$\begin{aligned}(f+g)(\alpha\vec{u}+\beta\vec{v}) &\overset{\text{def}}{=\!=} f(\alpha\vec{u}+\beta\vec{v})+g(\alpha\vec{u}+\beta\vec{v}) \\ &=\!= \alpha f(\vec{u})+\beta f(\vec{v})+\alpha g(\vec{u})+\beta g(\vec{v}) \\ &=\!= \alpha(f(\vec{u})+g(\vec{u}))+\beta(f(\vec{v})+g(\vec{v})) \\ &\overset{\text{def}}{=\!=} \alpha(f+g)(\vec{u})+\beta(f+g)(\vec{v})\end{aligned}$$

where the second equality is due to the linearity of f and g. Thus, $f+g \in \mathsf{Hom}(V, W)$ and $\alpha f \in \mathsf{Hom}(V, W)$. ■

Remark. Note that $\mathsf{End}(V) = \mathsf{Hom}(V, V)$ is a linear space over $\mathbb{F}$ and also a ring with the addition and multiplication operations defined in the previous section. The addition operation is commutative, but the multiplication operation is not necessarily commutative.

Then we can say that

$$\mathsf{Map}(\mathbb{F}^n, \mathbb{F}^m) \supseteq \mathsf{Hom}(\mathbb{F}^n, \mathbb{F}^m) \cong \mathsf{M}_{m\times n}(\mathbb{F})$$

2.2 Injections, Surjections and Isomorphisms

Similar to normal maps, there are injective, surjective and bijective linear maps.

Definition 2.4 — Injective Linear Maps. A linear map $f : V \to W$ between two linear spaces over a field $\mathbb{F}$ is said to be *injective* (or one-to-one) if for all $u, v \in V$, the following holds:

$$f(u) = f(v) \implies u = v$$

Equivalently, f is injective if the only vector in V that maps to the zero vector in W is the zero vector itself:

$$f(u) = 0 \implies u = 0$$

Definition 2.5 — Surjective Linear Maps. A linear map $f : V \to W$ is said to be *surjective* (or onto) if for every $w \in W$, there exists at least one $v \in V$ such that:

$$w = f(v)$$

Definition 2.6 — Invertible Linear Maps / Linear Equivalences. A linear map $T : V \to W$ is said to be *invertible* if T has a unique two-sided inverse S, denoted by T^{-1}, i.e., there exists a linear map $S : W \to V$ such that:

$$TS = 1_W \quad \text{and} \quad ST = 1_V$$

where $1_V : V \to V$ and $1_W : W \to W$ are the identity maps on V and W, respectively. In this case, we say that the linear spaces V and W are *isomorphic* or *linear equivalent*, denoted by $V \cong W$.

Corollary 2.3 — Invertible Linear Maps. A linear map $T : V \to W$ is invertible if and only if T is both injective and surjective, i.e., bijective / one-to-one correspondence.

Proof. ($\Rightarrow$) Assume $T : V \to W$ is invertible. By definition, there exists a linear map $S : W \to V$ such that $TS = 1_W$ and $ST = 1_V$.

To show that T is injective, suppose $T(u) = T(v)$ for some $u, v \in V$. We have:

$$S(T(u)) = S(T(v)) \implies (ST)(u) = (ST)(v) \implies 1_V(u) = 1_V(v) \implies u = v$$

Thus, T is injective. Then, to show that T is surjective, let $w \in W$. Since $TS = 1_W$, we have:

$$T(S(w)) = 1_W(w) = w$$

Then for every $w \in W$, there exists a $v = S(w) \in V$ such that $T(v) = w$. Thus, T is surjective.

($\Leftarrow$) Now assume that $T : V \to W$ is both injective and surjective. We need to show that there exists a linear map $S : W \to V$ such that $TS = 1_W$ and $ST = 1_V$.

Since T is surjective, for each $w \in W$, there exists at least one $v \in V$ such that $T(v) = w$. Define the map $S : W \to V$ by choosing one such preimage for each w:

$$S(w) = \text{a chosen } v \text{ such that } T(v) = w$$

To show that S is well-defined, we need to ensure that if $T(v_1) = T(v_2)$, then $v_1 = v_2$. This follows from the injectivity of T.

Now we verify that $TS = 1_W$: $(TS)(w) = T(S(w)) = w$ for all $w \in W$. Thus, $TS = 1_W$. Next, we verify that $ST = 1_V$: $(ST)(v) = S(T(v)) = v$ for all $v \in V$. Thus, $ST = 1_V$. Then we can check that S is a linear map as follows:

$$T(S(\alpha w_1 + \beta w_2)) = \alpha w_1 + \beta w_2 = \alpha T(S(w_1)) + \beta T(S(w_2)) = T(\alpha S(w_1) + \beta S(w_2))$$

Therefore, T has a two-sided inverse S, and hence T is invertible. ■

Definition 2.7 — Characteristic of a Field. The *characteristic* of a field $\mathbb{F}$ is the smallest positive integer n such that:

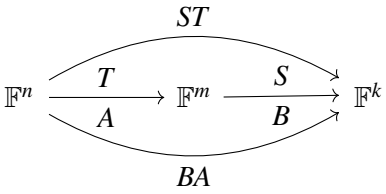
$$\underbrace{1+1+\cdots+1}_{n \text{ times}} = 0$$

If no such positive integer exists, the characteristic of $\mathbb{F}$ is defined to be 0.

■ **Example 2.3** The differentiation operator $D : \mathbb{F}[t] \to \mathbb{F}[t]$ is not an injective linear map as $D(1) = 0 = D(2)$ but is a surjective linear map if $\mathbb{F}$ is a field of characteristic 0. ■

2.3 Matrix Multiplications and Compositions of Linear Maps

We consider two linear maps $T : \mathbb{F}^n \to \mathbb{F}^m$ and $S : \mathbb{F}^m \to \mathbb{F}^k$ with standard matrices A and B, respectively. We want to find the standard matrix of the composition $ST : \mathbb{F}^n \to \mathbb{F}^k$.



Proposition 2.4 The standard matrix of the composition $ST : \mathbb{F}^n \to \mathbb{F}^k$ is the matrix multiplication BA, i.e., for all $x \in \mathbb{F}^n$,

$$(ST)x = B(Ax) = (BA)x$$

Proof. Let $x \in \mathbb{F}^n$ be a column matrix with entries $x^1, x^2, \cdots, x^n \in \mathbb{F}$. Then x can be expressed as a linear combination of the standard basis vectors $\vec{e}_1, \vec{e}_2, \cdots, \vec{e}_n$:

$$x = x^1\vec{e}_1 + x^2\vec{e}_2 + \cdots + x^n\vec{e}_n = x^i\vec{e}_i$$

Consider the j-th column of BA, it is given by:

$$(ST)\vec{e}_j = S(T(\vec{e}_j)) = S(\vec{a}_j) = B\vec{a}_j = (BA)\vec{e}_j$$

for all $j = 1, 2, \cdots, n$. This shows that the standard matrix of the composition ST is indeed the matrix multiplication BA. ■

Remark. Note that B is a $k \times m$ matrix and A is an $m \times n$ matrix, so the matrix multiplication BA is defined and results in a $k \times n$ matrix.

The matrix multiplication BA can be computed as follows:

$$BA = B\begin{bmatrix} | & | & & | \\ \vec{a}_1 & \vec{a}_2 & \cdots & \vec{a}_n \\ | & | & & | \end{bmatrix} = \begin{bmatrix} | & | & & | \\ B\vec{a}_1 & B\vec{a}_2 & \cdots & B\vec{a}_n \\ | & | & & | \end{bmatrix}$$

where $\vec{a}_i = T(\vec{e}_i)$ is the i-th column of the matrix A. Also,

$$Bx = x^1\vec{b}_1 + x^2\vec{b}_2 + \cdots + x^n\vec{b}_n = x^i\vec{b}_i$$

where $\vec{b}_i = B\vec{a}_i$ is the i-th column of the matrix B. Note that B is a $k \times m$ matrix, and $x \in \mathbb{F}^m$. Thus, the matrix multiplication Bx is defined and results in a column matrix in $\mathbb{F}^k$.

2.4 Elementary Row Operations

Definition 2.8 — Elementary Row Operations. Let A be an $m \times n$ matrix over a field $\mathbb{F}$. An *elementary row operation* on A is one of the following operations:

1. Row Interchange: $R_i \leftrightarrow R_j$.
2. Row Multiplication: $R_i \to \alpha R_i$, where $\alpha \in \mathbb{F} \setminus \{0\}$.
3. Row Addition: $R_i \to R_i + \alpha R_j$, where $\alpha \in \mathbb{F}$ and $i \neq j$.

Each elementary row operation can be represented by *left multiplication* of A by an appropriate $m \times m$ matrix over $\mathbb{F}$. Note that all of them are invertible linear maps from $\mathbb{F}^{m\times n}$ to $\mathbb{F}^{m\times n}$.

For ease of notation, we introduce the concept of matrix units, which is similar to the standard basis vectors $\vec{e}_i$.

Definition 2.9 — Matrix Units. Let m and n be two positive integers and $\mathbb{F}$ be a field. The *matrix unit* E_i^j is the $m \times n$ matrix over $\mathbb{F}$ with 1 in the (i,j)-th position and 0 elsewhere, i.e.,

$$(E_i^j)_k^l = \begin{cases} 1 & \text{if } (k,l) = (i,j) \\ 0 & \text{otherwise} \end{cases}$$

for all $1 \leq k \leq m$ and $1 \leq l \leq n$. The (i,j)-th position is the entry in the i-th row and j-th column.

It can also be defined as $E_i^j = \vec{e}_i \hat{e}^j \in \mathsf{M}_{m\times n}(\mathbb{F})$ where $\vec{e}_i \in \mathbb{F}^m$ and $\vec{e}_j^T = \hat{e}^j \in (\mathbb{F}^n)^*$ are the i-th and j-th standard basis vectors, respectively. The $\hat{e}^j$ is the row matrix with 1 in the j-th column and 0 anywhere else.

Remark. Note that for any $m \times n$ matrix A over a field $\mathbb{F}$, we have:

$$A\vec{e}_j = \text{the } j\text{-th column of } A \in \mathbb{F}^n$$
$$\hat{e}^i A = \text{the } i\text{-th row of } A \in (\mathbb{F}^m)^*$$

where $(\mathbb{F}^m)^*$ is the set of all row matrices with m entries in $\mathbb{F}$. $\hat{e}^i$ is an element in $(\mathbb{F}^m)^*$ for any $1 \leq i \leq m$. Note the distinction between superscript and subscript.

$$a_j^i = \hat{e}^i A \vec{e}_j = \text{the } (i,j)\text{-th entry of } A$$

We can write the E_i^j as:

$$E_i^j = \vec{e}_i \hat{e}^j = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \begin{bmatrix} 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \end{bmatrix} = \begin{bmatrix} 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \end{bmatrix}$$

(the j-th column ↓; ← the i-th row)

Then we consider the row operations by using the matrix units.

Proposition 2.5 The row operation $R_i \leftrightarrow R_j$ is a linear map where the standard matrix is $A_{R_i \leftrightarrow R_j} = I - E_i^i - E_j^j + E_i^j + E_j^i$.

Proof. The linear map $T : \mathbb{F}^n \to \mathbb{F}^n$ is defined pointwisely. We can say the map is:

$$\vec{e}_k \mapsto \begin{cases} \vec{e}_j & \text{if } k = i \\ \vec{e}_i & \text{if } k = j \\ \vec{e}_k & \text{if } k \neq i, j \end{cases}$$

Then the standard matrix of T is:

$$A_{R_i \leftrightarrow R_j} = \begin{bmatrix} | & & | & & | & & | \\ \vec{e}_1 & \cdots & \vec{e}_j & \cdots & \vec{e}_i & \cdots & \vec{e}_n \\ | & & | & & | & & | \end{bmatrix} = I - E_i^i - E_j^j + E_i^j + E_j^i$$

where I is the $n \times n$ identity matrix. ■

Proposition 2.6 The row operation $R_i \to \alpha R_i$ where $\alpha \in \mathbb{F}^\times := \mathbb{F} \setminus \{0\}$ is a linear map where the standard matrix is $A_{R_i \to \alpha R_i} = I + (\alpha - 1)E_i^i$.

Proof. The linear map $T : \mathbb{F}^n \to \mathbb{F}^n$ is defined pointwisely. We can say the map is:

$$\vec{e}_k \mapsto \begin{cases} \alpha\vec{e}_i & \text{if } k = i \\ \vec{e}_k & \text{if } k \neq i \end{cases}$$

Then the standard matrix of T is:

$$A_{R_i \to \alpha R_i} = \begin{bmatrix} | & & | & & | \\ \vec{e}_1 & \cdots & \alpha\vec{e}_i & \cdots & \vec{e}_n \\ | & & | & & | \end{bmatrix} = I + (\alpha - 1)E_i^i$$

where I is the $n \times n$ identity matrix. ■

Proposition 2.7 The row operation $R_i \to R_i + \alpha R_j$ where $\alpha \in \mathbb{F}$ and $i \neq j$ is a linear map where the standard matrix is $A_{R_i \to R_i + \alpha R_j} = I + \alpha E_i^j$.

Proof. The linear map $T : \mathbb{F}^n \to \mathbb{F}^n$ is defined pointwisely. We can say the map is:

$$\vec{e}_k \mapsto \begin{cases} \vec{e}_i + \alpha\vec{e}_j & \text{if } k = i \\ \vec{e}_k & \text{if } k \neq i \end{cases}$$

Then the standard matrix of T is:

$$A_{R_i \to R_i + \alpha R_j} = \begin{bmatrix} | & & | & & | & | \\ \vec{e}_1 & \cdots & \vec{e}_i + \alpha\vec{e}_j & \cdots & \vec{e}_n & \\ | & & | & & | & \end{bmatrix} = I + \alpha E_i^j$$

where I is the $n \times n$ identity matrix. ■

All invertible matrices can be written as a product of a finite sequence of elementary row operation matrices.

2.5 Dimensions of Vector Spaces

Definition 2.10 — Finite-Dimensional Vector Spaces. A linear space V over a field $\mathbb{F}$ is said to be *finite-dimensional* if there exists a linear equivalence $T : V \to \mathbb{F}^n$ for some positive integer n. In this case, we say that the dimension of V is n, denoted $\dim_{\mathbb{F}} V = n$ or simply $\dim V = n$.

Definition 2.11 — Infinite-Dimensional Vector Spaces. A linear space V over a field $\mathbb{F}$ is said to be *infinite-dimensional* if V is not finite-dimensional.

We have to prove that the dimension of a finite-dimensional vector space is well-defined.

Proposition 2.8 If there exists two linear equivalences $T : V \to \mathbb{F}^m$ and $S : V \to \mathbb{F}^n$, then $n = m$.

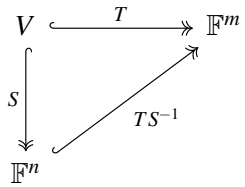
Proof. Since S is linear equivalence, it has a unique two-sided inverses $S^{-1} : \mathbb{F}^n \to V$. Consider the composition of this map:

$$TS^{-1} : \mathbb{F}^n \to \mathbb{F}^m$$

Since TS^{-1} is a composition of linear equivalences, it is also a linear equivalence. Mutatis mutandis, for the opposite direction.

Now, we know that a linear equivalence between two finite-dimensional vector spaces. Then we have $\dim \mathbb{F}^n = \dim \mathbb{F}^m$ or $n = m$. Thus, the dimension of a finite-dimensional vector space is well-defined. ■

Graphically, we have the following commutative diagram:



Remark. In drawing commutative diagram, we can use $\hookrightarrow$ to denote an injective linear map, $\twoheadrightarrow$ to denote a surjective linear map, and $\cong$ or combining the two to denote an invertible linear map.

2.6 Elementary Column Operations, Canonical Form and Rank

Definition 2.12 — Elementary Column Operations. Let A be an $m \times n$ matrix over a field $\mathbb{F}$. An *elementary column operation* on A is one of the following operations:

1. Column Interchange: $C_i \leftrightarrow C_j$.
2. Column Multiplication: $C_i \to \alpha C_i$, where $\alpha \in \mathbb{F} \setminus \{0\}$.
3. Column Addition: $C_i \to C_i + \alpha C_j$, where $\alpha \in \mathbb{F}$ and $i \neq j$.

Each elementary column operation can be represented by *right multiplication* of A by an appropriate $n \times n$ matrix over $\mathbb{F}$. Note that all of them are invertible linear maps from $\mathbb{F}^{m\times n}$ to $\mathbb{F}^{m\times n}$.

Proposition 2.9 Any $m \times n$ matrix A can be transformed into a matrix of the form $\begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}$ by a finite sequence of elementary row and column operations on A, where r is the rank of A.

The following is the commutative diagram of the proposition above, where $B = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}$:

$$\begin{array}{ccc} \mathbb{F}^n & \xrightarrow{\quad A \quad} & \mathbb{F}^m \\ \downarrow{\scriptstyle Q} & & \downarrow{\scriptstyle P} \\ \mathbb{F}^n & \xrightarrow{\quad B \quad} & \mathbb{F}^m \end{array}$$

Note that P is the product of a finite sequence of elementary row operation matrices and Q is the product of a finite sequence of elementary column operation matrices. Both P and Q are elementary and invertible matrices. Thus, we have:

$$\begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} = PAQ^{-1}$$

Definition 2.13 — Canonical Form of a Matrix. The matrix $\begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}$ obtained from an $m \times n$ matrix A by a finite sequence of elementary row and column operations on A is called the *canonical form* of A.

Remark. The canonical form of a matrix defined is also called the *Smith Normal Form* or *Normal Form* of a matrix.

Definition 2.14 — Rank of a Matrix. The *rank* of an $m \times n$ matrix A over a field $\mathbb{F}$, denoted by $\mathsf{Rank}(A)$, is the number of leading 1's in the matrix $\begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}$ obtained from A by a finite sequence of elementary row and column operations on A.

Remark. The value r is uniquely determined by A.

Proposition 2.10 Let A be an $m \times n$ matrix over a field $\mathbb{F}$. Then the following statements are equivalent:

$$A \text{ is invertible} \iff m\left\{\underbrace{\begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}}_{n}\right. \text{ is invertible} \iff \mathsf{Rank}(A) = m = n \iff \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} = I_m = I_n$$

Proof. If A is invertible, then the matrix PAQ^{-1} is also invertible, as P and Q are elementary and invertible matrices, and hence the product is invertible.

If PAQ^{-1} is invertible, and note that $m = n$ is automatically true. As only square matrix is invertible. Without the loss of generality, let say PAQ^{-1} is a $m \times m$ matrix, then we have $\mathsf{Rank}(PAQ^{-1}) = m$. Also note that the rank is invarient under multiplication by invertible matrices, so $\mathsf{Rank}(A) = \mathsf{Rank}(PAQ^{-1})$. Hence, $\mathsf{Rank}(A) = m = n$.

If $\mathsf{Rank}(A) = m = n$, as the canonical matrix remains the $m \times n$ structure, we know that the canonical form is actually a square matrix, let say $m \times m$. Also $r = \mathsf{Rank}(A) = m$. Hence the whole canonical form become an identity matrix I_m.

If the canonical form is an identity matrix I, i.e., it is invertible. Then the matrix $P^{-1}IQ = A$ is also invertible for some elementary and invertible matrices P and Q. ■

Proposition 2.11 Let A be an $m \times n$ matrix over a field $\mathbb{F}$. Then the following statements are equivalent:

$$A \text{ has a left inverse} \iff A \text{ is injective} \iff \mathsf{Rank}(A) = n \iff \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} I_n \\ 0 \end{bmatrix}$$

Proof. If A has a left inverse, let say B, then we have $BA = I_n$. Then for $B(A(x_1)) = B(A(x_2))$, we have $(BA)x_1 = (BA)x_2$, which implies $x_1 = x_2$. Hence it is injective.

If A is injective, we can consider $A = P^{-1}CQ$, where C is the canonical form of the matrix A. Then we consider $P^{-1}CQ\vec{x} = \vec{0}$. Since P^{-1} is invertible, it won't produce non-trivial solutions. We can consider $C(Q\vec{x}) = \vec{0} = C\vec{y}$. Then we have

$$\begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \vec{y}_1 \\ \vec{y}_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

where $\vec{y}_1$ and $\vec{y}_2$ are column vectors with size r and $n-r$ respectively. Then $I_r\vec{y}_1 = 0$, which implies $\vec{y}_1 = 0$, while $\vec{y}_2$ can be anything. As A is invertible, then $A\vec{x} = \vec{0}$ only has one trivial solution $\vec{x} = \vec{0}$. Also, Q is invertible, hence $\vec{y}$ has only one trivial solution $\vec{0}$, i.e., $\vec{y}_2 = \vec{0}$. Hence we have $n - r = 0$ due to the size of $\vec{y}_2$ being 0. Hence the rank of A is n.

If $\mathsf{Rank}(A) = n$, then the canonical form of A is

$$\begin{bmatrix} I_{r\times r} & 0_{r\times(n-r)} \\ 0_{(m-r)\times r} & 0_{(m-r)\times(n-r)} \end{bmatrix} = \begin{bmatrix} I_{n\times n} & 0_{n\times(n-n)} \\ 0_{(m-n)\times n} & 0_{(m-n)\times(n-n)} \end{bmatrix} = \begin{bmatrix} I_{n\times n} \\ 0_{(m-n)\times n} \end{bmatrix} = \begin{bmatrix} I_n \\ 0 \end{bmatrix}$$

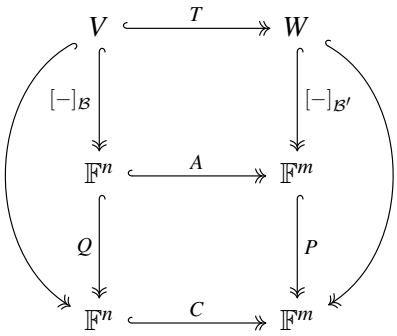
If the canonical form of A is $\begin{bmatrix} I_n \\ 0 \end{bmatrix}$, then we consider $PAQ^{-1} = C$. Also, $A = P^{-1}CQ$. We construct a candidate for left inverse $D = [I_n \quad 0]$. Then we have $DC = [I_n \quad 0]\begin{bmatrix} I_n \\ 0 \end{bmatrix} = I_n$. Then the left inverse of A is $L = QDP^{-1}$. Then we check, $LA = QDP^{-1}A = QDP^{-1}PCQ^{-1} = I_n$. Hence, A indeed has a left inverse. ■

Proposition 2.12 Let A be an $m \times n$ matrix over a field $\mathbb{F}$. Then the following statements are equivalent:

$$A \text{ has a right inverse} \iff A \text{ is surjective} \iff \mathsf{Rank}(A) = m \iff \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} I_m & 0 \end{bmatrix}$$

Proposition 2.13 For every $\vec{b}$, $\begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}\vec{x} = \vec{b}$ has a unique solution.

Linear Algebra is the study of linear map between two finite-dimensional vector spaces.



where $C = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}$, $\dim V = n$ and $\dim W = m$.

The coordinate maps $[-]_{\mathcal{B}}$ and $[-]_{\mathcal{B}'}$ are linear equivalences and they are the trivialisation of V and W, respectively. The matrix A is the standard matrix of the linear map $T : V \to W$ under the bases $\mathcal{B}$ and $\mathcal{B}'$. The matrix C is the canonical form of A. The matrices P and Q are products of finite sequences of elementary row and column operation matrices, respectively. Both P and Q are elementary and invertible matrices.

2.7 Properties of Linear Maps

Let $f : V \to W$ be a linear map between two finite-dimensional vector spaces over $\mathbb{F}$. We have the following properties:

1. f is injective if and only if $\mathsf{Ker}(f) = \{0_V\}$, i.e., the kernel is trivial.
2. f is surjective if and only if $\mathsf{Coker}(f) = \{0_W\}$, i.e., the cokernel is trivial.
3. f is an isomorphism if and only if $\mathsf{Ker}(f) = \{0_V\}$ and $\mathsf{Coker}(f) = \{0_W\}$.
4. f is surjective if and only if for any linear map $g : W \to Z$, $g \circ f = 0$ implies $g = 0$.
5. f is injective if and only if for any linear map $h : U \to V$, $f \circ h = 0$ implies $h = 0$.

Let $f : V \to W$ be a set map between linear spaces. Then the graph of f, $\Gamma_f := \{(v, f(v)) \mid v \in V\}$ is a linear subspace of $V \oplus W$ if and only if f is a linear map. Also, the domain of f is isomorphic to Γ_f.

f is injective if and only if f is an imbedding, i.e., the map $\overline{f} : V \to \mathsf{Im}(f)$ that sends v to $f(v)$ is an isomorphism.

3. Linear Spaces

"Completion is one of the major great ideas in mathematics."

GUOWU MENG

3.1 Linear Subspaces, Kernels and Images

Here, we discuss linear spaces with more in depth terms.

Definition 3.1 — Linear Subspaces. Let W be a linear space over $\mathbb{F}$ and V is a subset of W, denoted as $V \subset W$. V is a *linear subspace* of W if V, with $+$ and $\cdot$ inherited from those of W, is a linear space.

Proposition 3.1 Let $V \subset W$. V is a subspace of W if and only if V is not empty and V is closed under $+$ and $\cdot$.

Proof. If V is a subspace of W, then V is non-empty as a linear space must contain a zero vector by definition, as V is also a linear space. Also, the other two are due to the axioms of linear space.

If V is not empty and closed under $+$ and $\cdot$, we just have to check the each axiom. ■

Definition 3.2 — Kernels. Let $f : V \to W$ be a linear map. The *kernel* of f, denoted as $\mathsf{Ker}(f)$, is defined as

$$\mathsf{Ker}(f) \stackrel{\text{def}}{=\!=} \{v \in V \mid f(v) = 0_W\} = f^{-1}(\{0_W\})$$

■ **Example 3.1** Let $f : V \to W$ be a linear map. $\mathsf{Ker}(f)$ is a subspace of domain of f, i.e., V.

First, we have $0_V \in \mathsf{Ker}(f)$, as $f(0_V) = 0_W$, so $\mathsf{Ker}(f)$ is not empty.

Then we consider $\alpha^1, \alpha^2 \in \mathbb{F}$ and $v_1, v_2 \in \mathsf{Ker}(f)$, we have

$$f(\alpha^1 v_1 + \alpha^2 v_2) = \alpha^1 f(v_1) + \alpha^2 f(v_2) = \alpha^1(0_W) + \alpha^2(0_W) = 0_W$$

The first equality due to the linearity of f and the second is due to $v_i \in \mathsf{Ker}(f)$. ■

Definition 3.3 — Images. Let $f : V \to W$ be a linear map. The *image* of f, denoted by $\mathsf{Im}(f)$, is defined as

$$\mathsf{Im}(f) \stackrel{\text{def}}{=\!=} \{f(v) \mid v \in V\} \subset W$$

■ **Example 3.2** Let $f : V \to W$ be a linear map. $\mathsf{Im}(f)$ is a subspace of codomain of f, i.e., W.
First, we have $f(0_V) = 0_W \in \mathsf{Im}(f)$, so $\mathsf{Im}(f)$ is not empty.
Then we consider $\alpha^1, \alpha^2 \in \mathbb{F}$ and $f(v_1), f(v_2) \in \mathsf{Im}(f)$. We have

$$\alpha^1 f(v_1) + \alpha^2 f(v_2) = f(\alpha^1 v_1 + \alpha^2 v_2) \in \mathsf{Im}(f)$$

The equality is due to the linearity of f. ■

■ **Example 3.3** Let W be a linear space over a field $\mathbb{F}$ and $\{V_\alpha\}_{\alpha \in I}$ be the family of subspaces of W indexed by the element in the index set I. Then $\bigcap_{\alpha \in I} V_\alpha$ is also a subspace of W.
First, we have $0_W \in V_\alpha$ for all $\alpha \in I$, so $0_W \in \bigcap_{\alpha \in I} V_\alpha$. Thus, $\bigcap_{\alpha \in I} V_\alpha$ is not empty.
Then we consider $\alpha^1, \alpha^2 \in \mathbb{F}$ and $v_1, v_2 \in \bigcap_{\alpha \in I} V_\alpha$. We have $v_1, v_2 \in V_\alpha$ for all $\alpha \in I$. Thus, $\alpha^1 v_1 + \alpha^2 v_2 \in V_\alpha$ for all $\alpha \in I$. This shows that $\alpha^1 v_1 + \alpha^2 v_2 \in \bigcap_{\alpha \in I} V_\alpha$. ■

Then we consider the duality of the intersection and union of subspaces. Whether the union of two subspaces is still a subspace? Unfortunately, the answer is no in general case. However, we have the following proposition.

Proposition 3.2 Let W be a linear space over a field $\mathbb{F}$ and consider the family of subspaces $\{V_\alpha\}_{\alpha \in I}$. Then $\overline{\bigcup_{\alpha \in I} V_\alpha}$ is a subspace of W where $\overline{\bigcup_{\alpha \in I} V_\alpha}$ is the completion of $\bigcup_{\alpha \in I} V_\alpha$ under linear combinations. We call $\overline{\bigcup_{\alpha \in I} V_\alpha}$ the *sum* of the subspaces $\{V_\alpha\}_{\alpha \in I}$, denoted by $\sum_{\alpha \in I} V_\alpha$.

3.2 Linear Span and Linear Independence

Definition 3.4 — Linear Span. Let V be a linear space over a field $\mathbb{F}$ and $S \subset V$. The *linear span* of S, denoted by $\mathsf{Span}_{\mathbb{F}}(S)$ or simply $\mathsf{Span}(S)$ or $\overline{S}$ or $\langle S \rangle$, is defined as the completion of S inside V under linear combinations.

Corollary 3.1 The linear span of S can also be defined as the intersection of all subspaces of V containing S, which is the smallest linear subspace of V containing S. It can be written as:

$$\mathsf{Span}(S) = \bigcap_{\alpha \in I} V_\alpha \subset V \quad \text{where } I = \{V_\alpha \subset V \mid V_\alpha \text{ is a subspace of } V \text{ and } S \subset V_\alpha\}$$

Remark. Note that I is not empty as $V \in I$. Thus, $\mathsf{Span}(S)$ is well-defined. V is the largest subspace of itself and $\{0_V\}$ is the smallest subspace of V.

Proposition 3.3 Let W be a linear space over a field $\mathbb{F}$ and $S \subset W$. Then

$$\mathsf{Span}(S) = \left\{ \sum_{i=1}^{n} \alpha^i s_i \mid n \in \mathbb{N}, \alpha^i \in \mathbb{F}, s_i \in S \right\}$$

Note that the summation is a finite summation.

Definition 3.5 — Linear Independences. Let W be a linear space over a field $\mathbb{F}$ and $V_1, \cdots, V_k$ be subspaces of W. The subspaces $V_1, \cdots, V_k$ are said to be *linearly independent* if $V_i \neq \{0_W\}$ for all i and there is one and only one way to split $0_W \in W$ as a sum of vectors from each V_i, i.e., if $v_i \in V_i$ for all i and $\sum_{i=1}^{k} v_i = 0_W$, then $v_i = 0_W$ for all i.

Vectors $v_1, v_2, \cdots, v_k \in W$ are said to be independent if the subspaces $\mathsf{Span}(v_1)$, $\mathsf{Span}(v_2)$, $\cdots$, $\mathsf{Span}(v_k)$ are linearly independent.

Proposition 3.4 $v_1, v_2, \cdots, v_k \in W$ are linearly independent if and only if there is one and only one way to write $0_W \in W$ as the combination of $v_1, \cdots, v_k$ with coefficients in $\mathbb{F}$, i.e., the equation

$$\alpha^1 v_1 + \cdots + \alpha^k v_k = 0_W$$

has only the trivial solution, i.e., $\alpha^i = 0$ for all i.

3.3 Linearly Independent Sets and Spanning Sets

If we consider a set, what does it mean by being linearly independent? Is there any properties for spanning if the set spans the whole codomain?

Definition 3.6 — Linearly Independent Sets. Let V be a linear space over a field $\mathbb{F}$. A subset $S \subseteq V$ is said to be a *linearly independent set* of vectors in V if no elements in S can be expressed as a linear combination of the finitely many other elements in S.

Definition 3.7 — Spanning Sets. Let V be a linear space over a field $\mathbb{F}$. A subset $S \subseteq V$ is said to be a *spanning set* of V if $\mathsf{Span}(S) = V$.

■ **Example 3.4** Let $V = \mathbb{F}^3$ and consider the three vectors $\vec{e}_1$, $\vec{e}_2$ and $\vec{e}_3$.

Then the set $S = \{\vec{e}_1, \vec{e}_2, \vec{e}_1 + \vec{e}_2\}$ is not a spanning set of V as $\mathsf{Span}(S) = \mathsf{Span}\{\vec{e}_1, \vec{e}_2\} \neq V$. If we consider the $\mathsf{Span}\{\vec{e}_1, \vec{e}_2\} = W$, then $\{\vec{e}_1, \vec{e}_2\}$ is a minimal spanning set of W.

The set $S = \{\vec{e}_1, \vec{e}_1 + \vec{e}_2, \vec{e}_1 + \vec{e}_2 + \vec{e}_3\}$ is a spanning set of V. ■

Remark. If we consider the matrix of $\{\vec{e}_1, \vec{e}_2, \vec{e}_1 + \vec{e}_2\}$ with respect to the standard basis of $\mathbb{F}^3$, we have:

$$A = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

Then we have $\mathsf{Rank}(A) = 2 < 3$. Thus, the set is not a spanning set of $\mathbb{F}^3$.

■ **Example 3.5** Consider the subset $S = \{1, t, t^2, \cdots\} \subset \mathbb{F}[[t]]$. Then $\mathsf{Span}(S) = \mathbb{F}[t]$ which is a proper subspace of $\mathbb{F}[[t]]$. As the linear combination of finitely many elements in S is a polynomial, but an element in $\mathbb{F}[[t]]$ can be a power series. ■

Definition 3.8 — Minimal Spanning Sets. Let V be a linear space over a field $\mathbb{F}$. A spanning set $S \subseteq V$ is said to be a *minimal spanning set* of V if no proper subset of S is a spanning set of V, i.e., $S' \subset S \implies \mathsf{Span}(S') \subset \mathsf{Span}(S) = V$ where $\mathsf{Span}(S') \neq V$.

The following is also the equivalence definition of linearly independent sets, spanning sets and minimal spanning sets.

Given a linear space V over a field $\mathbb{F}$. We define the order set $S := \{\vec{v}_1, \vec{v}_2, \cdots, \vec{v}_n\} \subseteq V$. The order set S forms a linear map $\phi_S : \mathbb{F}^n \to V$ defined by:

$$\phi_S(\vec{x}) = \phi_S\left(\begin{bmatrix} x^1 \\ x^2 \\ \vdots \\ x^n \end{bmatrix}\right) = x^1\vec{v}_1 + x^2\vec{v}_2 + \cdots + x^n\vec{v}_n = \sum_{i=1}^{n} x^i\vec{v}_i$$

Proposition 3.5 The order set $S := \{\vec{v}_1, \vec{v}_2, \cdots, \vec{v}_n\} \subseteq V$ is said to be linearly independent if and only if the linear map $\phi_S : \mathbb{F}^n \to V$ defined above is injective.

Proposition 3.6 The order set $S := \{\vec{v}_1, \vec{v}_2, \cdots, \vec{v}_n\} \subseteq V$ is said to be a spanning set of V if and only if the linear map $\phi_S : \mathbb{F}^n \to V$ defined above is surjective.

Proposition 3.7 The order set $S := \{\vec{v}_1, \vec{v}_2, \cdots, \vec{v}_n\} \subseteq V$ is said to be a minimal spanning set of V if and only if the linear map $\phi_S : \mathbb{F}^n \to V$ defined above is bijective.

Remark. A order minimal spanning set is regarded as *basis*.

■ **Example 3.6** Let X be a set, $\mathbb{F}[[X]]$ be the set of all functions $f : X \to \mathbb{F}$ and $\mathbb{F}[X]$ be the set of all finite support functions $f : X \to \mathbb{F}$. For each $x \in X$, we define the Kronecker delta function $\delta_x : X \to \mathbb{F}$ at point x by

$$\delta_x(y) = \begin{cases} 1 & \text{if } y = x \\ 0 & \text{if } y \neq x \end{cases}$$

Clearly, δ_x has finite support, thus $\delta_x \in \mathbb{F}[X]$.

Then we have a set $\delta_X = \{\delta_x \mid x \in X\} \subset \mathbb{F}[X]$. We have $\mathsf{Span}(\delta_X) = \mathbb{F}[X]$ as any finite support function $f : X \to \mathbb{F}$ can be written as a linear combination of finitely many delta functions. Thus, δ_X is a spanning set of $\mathbb{F}[X]$.

Moreover, δ_X is a linearly independent set. Assume that there exists a finite linear combination of other delta functions such that $\delta_x = \sum \alpha^y \delta_y$. Then we have $\delta_x(x) = 1 = \sum \alpha^y \delta_y(x) = 0$. This is a contradiction. Thus, δ_X is a linearly independent set. ■

3.4 Group Actions

Next, we discuss quotient space. However, before introducing quotient space, we have to understand what group actions are.

Definition 3.9 — Group Actions. Let G be a group and X be a set. A *left group action* of G on X is a map $\cdot : G \times X \to X$, $(g,x) \mapsto g \cdot x$, such that for all $g_1, g_2 \in G$ and $x \in X$, the following properties hold:

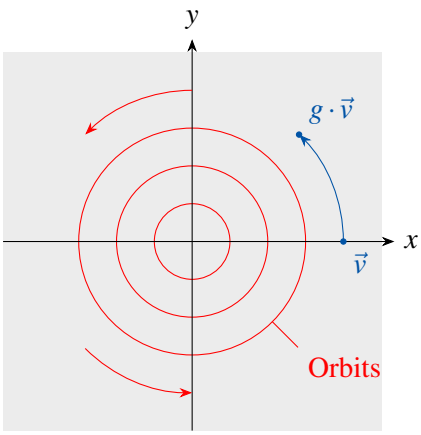
1. Compatibility: $(g_1 g_2) \cdot x = g_1 \cdot (g_2 \cdot x)$.
2. Identity: $e \cdot x = x$ where e is the identity element of G.

Same for the right group action of G on X, just think it dually.

Consider a rotation on a plane. It is a group action of the group $SO(2)$ on the set $\mathbb{R}^2$.

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

Then we have the following group action:



Definition 3.10 — Orbits. Let G be a group acting on a set X. The *orbit* of the action through a point $x \in X$, denoted as $G \cdot x$, is defined as the set of points in X that can be reached from x by the action of elements of G, i.e.,

$$G \cdot x = \{g \cdot x \mid g \in G\}$$

There is only two situation for the orbits, either the origin or a circle.

In the following section, we may regard the orbits $G \cdot x$ as a *coset*.

Definition 3.11 — Partition. A *partition* of a set X is a collection of non-empty, disjoint subsets of X whose union is X. The partition of the set X is the same as an equivalence relation on X.

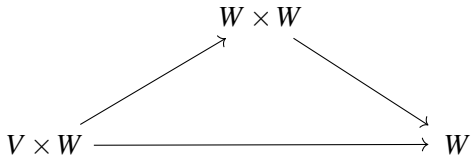
Orbits give a partition of the set X, i.e., X can be expressed as the disjoint union of its orbits. The orbits of the action are the equivalence classes of the equivalence relation.

Let $f : X \to Y$ be a map between two sets X and Y. Then f defines a partition of X by the equivalence relation. The equivalence classes are the preimages of points in Y, i.e., $f^{-1}(y)$ for each $y \in Y$.

3.5 Quotient Spaces

Let V be a subspace of a linear space W over a field $\mathbb{F}$. We know $(V,+)$ is an abelian group. Then we have the group action of V on W defined by: $(v,w) \mapsto v \cdot w$ for all $v \in V, w \in W$. $v \cdot w$ is defined as $v+w$ where $+$ is the addition operation in W. We know that $(v_1+v_2)+w = v_1+(v_2+w)$ and $0_V + w = w$ for all $v_1, v_2 \in V$ and $w \in W$. Thus, it is a group action.

The following commutative diagram illustrates the group action, where the associative and identity properties are inherited from the addition operation in W, i.e., we need not prove the group action as above.



This group action defines the following equivalence relation on W, where V is the acting group:

$$\begin{aligned} w_1 \sim w_2 &\implies \exists v \in V \text{ such that } w_2 = v + w_1 \\ &\iff w_2 - w_1 \in V \end{aligned}$$

Definition 3.12 — Quotient Spaces. Let W be a linear space over a field $\mathbb{F}$ and V be a subspace of W. The *quotient space* of W by V, denoted by W/V, is defined as the set of orbits of the group action of V on W, or the set of V-equivalence classes in W with the equivalence relation defined above, i.e.,

$$W/V = \{V \cdot w \mid w \in W\} = \{w+V \mid w \in W\}$$

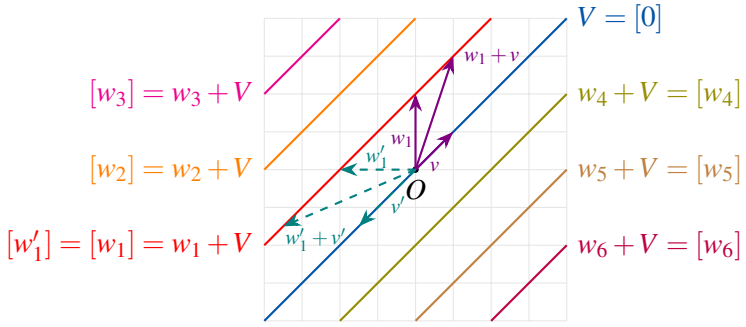
where $V \cdot w = w+V = \{w+v \mid v \in V\}$ is called the *coset* of V in W containing w.

Definition 3.13 — Quotient Map. The natural surjective map $\pi : W \to W/V$ defined by $\pi(w) = w+V$ for all $w \in W$ is called the *quotient map* or *projection map*. Note that $w+V$ can be written as $\overline{w}$ or $[w]$.

In general, if a group G acts on a set X, then the quotient set X/G is defined as the set of orbits of the action, i.e.,

$$X/G = \{G \cdot x \mid x \in X\}$$

Similarly, there is a natural surjective map $\pi : X \to G$ defined by $\pi(x) = G \cdot x$ for all $x \in X$.
The following is a graphical illustration of the quotient space.



We can see that each line parallel to V represents a coset of V in W. The quotient space W/V is the set of all such lines. We may consider each line as an orbit of the group action of V on W. Note that there is not only one unique way to represent the coset $w+V$. Just like the illustration above, w_1 and w_1' are two different representatives of the same coset $w_1+V=w_1'+V$. Note that their difference is an element in V, i.e., $w_1-w_1'\in V$.

Note that we now do not know whether W/V is a linear space or not. We will show that it is indeed a linear space by using the following proposition.

Proposition 3.8 There is a unique linear structure on W/V such that the quotient map $\pi: W\to W/V$ is a linear map.

Proof. Assume that such a linear structure exists. Then for all $w_1,w_2\in W$ and $\alpha_1,\alpha_2\in\mathbb{F}$, we have

$$\pi(\alpha_1 w_1+\alpha_2 w_2)=[\alpha_1 w_1+\alpha_2 w_2]=\alpha_1[w_1]+\alpha_2[w_2]=\alpha_1\pi(w_1)+\alpha_2\pi(w_2)$$

This suggests that $\alpha_1[w_1]+\alpha_2[w_2]$ should be defined as $[\alpha_1 w_1+\alpha_2 w_2]$ if π is linear. As there is only one formula, this proves the uniqueness of the linear structure on W/V.

Then we consider whether the linear combination on W/V is well-defined. Assume that $[w_1]=[w_1']$ and $[w_2]=[w_2']$, i.e., $w_1-w_1'\in V$ and $w_2-w_2'\in V$. Then we have

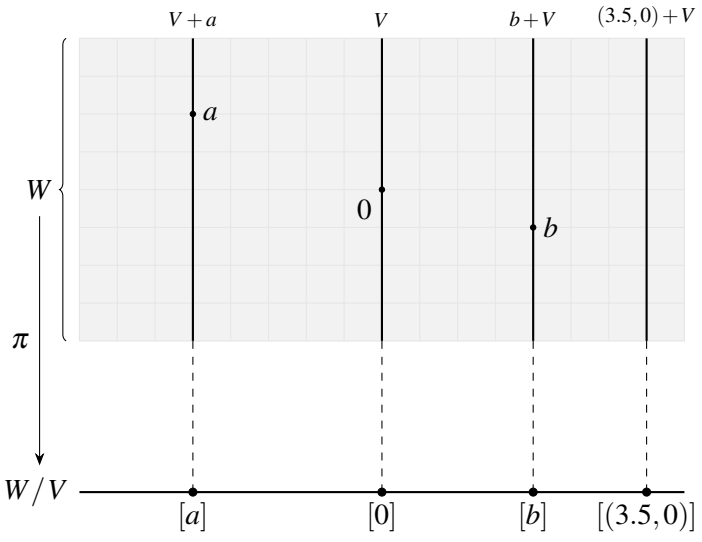
$$(\alpha_1 w_1+\alpha_2 w_2)-(\alpha_1 w_1'+\alpha_2 w_2')=\alpha_1(w_1-w_1')+\alpha_2(w_2-w_2')\in V$$

which means $[\alpha_1 w_1+\alpha_2 w_2]=[\alpha_1 w_1'+\alpha_2 w_2']$. This means that the linear combination is independent of the choice of representatives. Thus, the linear combination is well-defined. ■

In the normal procedure, we first define the operations and then check whether the set is closed under the operations and zero exists. Then we check whether the map preserves the structure and show the uniqueness of the structure. However, in this case, we first assume that such a structure exists and then derive the operations from this assumption. Subsequently, we check whether the operations are well-defined.

In the first part, we show that there is only one possible way to define the operations if the quotient map is linear. Moreover, the definition ensures the preservation of the linear structure. In the second part, we show that the operations on the set W/V are well-defined.

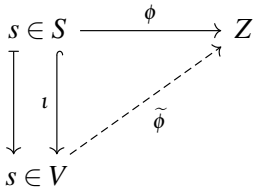
If we consider the graphical representation of the quotient space W/V and the quotient map π, we may use the following diagram:



3.6 Universal Properties

Proposition 3.9 Let V be a linear space over a field $\mathbb{F}$ and S be a minimal spanning set of V. Then for any set map $\phi : S \to Z$, where Z is any linear space over $\mathbb{F}$, there is a unique linear map $\widetilde{\phi} : V \to Z$ such that $\widetilde{\phi}|_S = \phi$.

In other words, the following diagram commutes:



Proof. Assume the existence of such a linear map $\widetilde{\phi}$. Then for all $s \in S$, we have $\widetilde{\phi} \circ \iota(s) = \widetilde{\phi}(s) = \phi(s)$.

Since S is a minimal spanning set of V, for any $v \in V$, we have a unique way to write v as a linear combination of finitely many elements in S, i.e., $v = \sum_{i=1}^{n} \alpha_i s_i$ where $\alpha_i \in \mathbb{F}$ and $s_i \in S$ are distinct. Then we have

$$\widetilde{\phi}(v) = \widetilde{\phi}\left(\sum_{i=1}^{n} \alpha^i s_i\right) = \sum_{i=1}^{n} \alpha^i \widetilde{\phi}(s_i) = \sum_{i=1}^{n} \alpha^i \phi(s_i)$$

This shows the uniqueness of $\widetilde{\phi}$.

Then we claim that the map $\widetilde{\phi}$ defined above is well-defined. Since S is a minimal spanning set of V, there is only one way to write each element in V as a linear combination of elements in S. Thus, the definition of $\widetilde{\phi}$ does not depend on the choice of representation of v. This shows that $\widetilde{\phi}$ is well-defined. ■

Note that we first define the map on the spanning set and then extend it to the whole space. The uniqueness is due to the fact that there is only one way to write each element in V as a linear combination of elements in S and the existence is due to the fact that we can always define the map on V by using the linear combination.

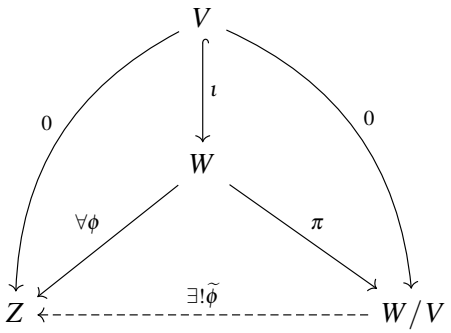
This proposition shows thats a linear space with a minimal spanning set has the following universal property: any set map from the minimal spanning set to another linear space can be uniquely extended to a linear map from the whole space to that linear space.

$$\phi \longmapsto \widetilde{\phi}$$

$$\mathsf{Map}(S,Z) \cong \mathsf{Hom}(V,Z)$$

$$\widetilde{\phi} \circ \iota \longmapsto \widetilde{\phi}$$

Proposition 3.10 Let W be a linear space over a field $\mathbb{F}$ and V be a subspace of W. Then we have the following commutative diagram:



where Z is any linear space over $\mathbb{F}$ and $\phi : W \to Z$ is any linear map such that $\phi(v) = 0_Z$ for all $v \in V$. Then there is a unique linear map $\widetilde{\phi} : W/V \to Z$ such that $\widetilde{\phi} \circ \pi = \phi$.

Proof. Assume the existence of such a linear map $\widetilde{\phi}$. Then for all $w \in W$, we have $\widetilde{\phi}([w]) = \phi(w)$. However, this may not be well-defined. Then, we check whether it is well-defined. Assume that $[w] = [w']$, then we have $\widetilde{\phi}([w']) = \phi(w')$. Note that $w - w' \in V$. Thus, we have $\phi(w' - w) = 0_Z$. This means that $\phi(w') - \phi(w) = 0_Z$, i.e., $\phi(w') = \phi(w)$. This shows that $\widetilde{\phi}([w']) = \widetilde{\phi}([w])$. Thus, $\widetilde{\phi}$ is well-defined.

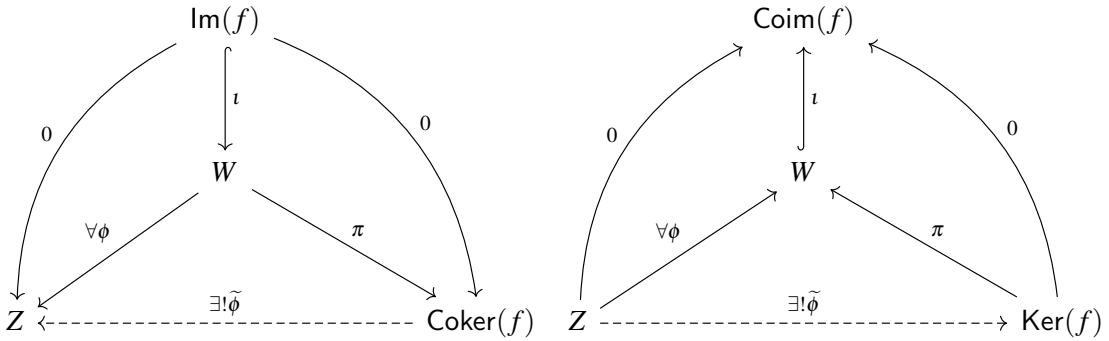
Then we consider the linearity of $\widetilde{\phi}$. For all $[w_1], [w_2] \in W/V$ and $\alpha^1, \alpha^2 \in \mathbb{F}$, we have

$$\begin{aligned}\widetilde{\phi}(\alpha^1[w_1] + \alpha^2[w_2]) &= \widetilde{\phi}([\alpha^1 w_1 + \alpha^2 w_2]) \\ &= \phi(\alpha^1 w_1 + \alpha^2 w_2) \\ &= \alpha^1 \phi(w_1) + \alpha^2 \phi(w_2) \\ &= \alpha^1 \widetilde{\phi}([w_1]) + \alpha^2 \widetilde{\phi}([w_2])\end{aligned}$$

This shows that $\widetilde{\phi}$ is linear. ■

Remark. Note that $[0] = V$. If $v \in V$, then $[v] = v + V = \{v + v' \mid v' \in V\} = \{v'' \mid v'' \in V\} = V = [0]$. Thus, $\pi(v) = [v] = [0]$ for all $v \in V$. So the map from $V \to W/V$ is the zero map. Thus, the triangle commutes. Also, the map from v to Z is defined as the zero map, making the construction of $\widetilde{\phi}$ is possible, as the key step is that $\phi(w' - w) = 0_Z$ for all $w' - w \in V$.

Generally, we may consider the following commutative diagrams, where left is the general case and right is the dual case:



Definition 3.14 — Cokernel. Let $f : V \to W$ be a linear map between two linear spaces over a field $\mathbb{F}$. The *cokernel* of f, denoted by $\mathsf{Coker}(f)$, is defined as the quotient space of W by the image of f, i.e.,

$$\mathsf{Coker}(f) = W/\mathsf{Im}(f) = W/\mathsf{Im}(f)$$

where $\mathsf{Im}(f) = \{f(v) \mid v \in V\}$ is the image of f.

Definition 3.15 — Coimage. Let $f : W \to V$ be a linear map between two linear spaces over a field $\mathbb{F}$. The *coimage* of f, denoted by $\mathsf{Coim}(f)$, is defined as the quotient space of the domain W by the kernel of f, i.e.,

$$\mathsf{Coim}(f) = W/\mathsf{Ker}(f) = W/\mathsf{Ker}(f)$$

where $\mathrm{Ker}(f) = \{w \in W \mid f(w) = 0_V\}$ is the kernel of f.

3.7 Sum and Direct Sum

Definition 3.16 — Sum of Subspaces. Let V_1 and V_2 be two subspaces of a linear space W over a field $\mathbb{F}$. The *sum* of V_1 and V_2, denoted by V_1+V_2, is defined as the set of all possible sums of elements from V_1 and V_2, i.e.,

$$V_1+V_2=\{v_1+v_2 \mid v_1\in V_1, v_2\in V_2\}$$

Proposition 3.11 The sum V_1+V_2 of two subspaces V_1 and V_2 of a linear space W over a field $\mathbb{F}$ is also a subspace of W.

Proposition 3.12 $V_1+V_2=\mathsf{Span}(V_1\cup V_2)$.

Recall the definition of linear independence (Definition 3.5): V_1 and V_2 are said to be *linearly independent* if V_1 and V_2 are non-trivial and $x_1+x_2=0$ for $x_i\in V_i$ implies that $x_1=x_2=0$.

We have the following definition for weakly linear independence.

Definition 3.17 — Weak Linear Independence. Let V_1 and V_2 be two subspaces of a linear space W over a field $\mathbb{F}$. V_1 and V_2 are said to be *weakly linearly independent* if $x_1+x_2=0$ for $x_1\in V_1$ and $x_2\in V_2$ implies that $x_1=x_2=0$. Note that V_1 or V_2 can be trivial.

Then the definition of direct sum is as follows.

Definition 3.18 — Direct Sum of Subspaces. Let V_1 and V_2 be two subspaces of a linear space W over a field $\mathbb{F}$. The *direct sum* of V_1 and V_2, denoted by $V_1\oplus V_2$, is defined as the sum V_1+V_2 when V_1 and V_2 are weakly linearly independent, i.e.,

$$V_1\oplus V_2=V_1+V_2$$

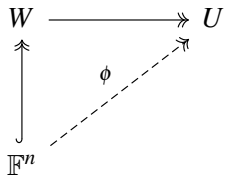
when V_1 and V_2 are weakly linearly independent.

Recall (Definition 2.10) that W is a finite-dimensional if $W\cong\mathbb{F}^n$ for some positive integer n. It is equivalent to saying that W is finitely spanned, i.e., having a finite spanning set.

Proof. If we have a map $\phi:\mathbb{F}^n\to W$, then $W=\mathsf{Span}\{\phi(e_1),\phi(e_2),\cdots,\phi(e_n)\}$. However, the set $\{\phi(e_1),\phi(e_2),\cdots,\phi(e_n)\}$ may not be linearly independent. Thus, we can always find a minimal spanning set of W from it. WLOG, we can say $W=\mathsf{Span}\{\phi(e_1),\phi(e_2),\cdots,\phi(e_k)\}$ for some $k\le n$. Then using (Proposition 3.7), we have a bijective map $\phi_{\{e_1,e_2,\cdots,e_k\}}:\mathbb{F}^k\to W=\mathsf{Span}\{\phi(e_1),\phi(e_2),\cdots,\phi(e_k)\}$. ■

Proposition 3.13 W is finite-dimensional if and only if all its subspaces and quotient spaces are finite-dimensional.

Proof. For subspace $U\subseteq W$ and W is finite-dimensional, we have:



Then the map $\phi:\mathbb{F}^n\to U$ is defined by $x=\alpha^1\vec{e}_1+\cdots+\alpha^n\vec{e}_n\mapsto\phi(x)=\alpha^1\phi(\vec{e}_1)+\cdots+\alpha^n\phi(\vec{e}_n)$. Thus, U is finitely spanned, $U=\mathsf{Span}\{\phi(\vec{e}_1),\phi(\vec{e}_2),\cdots,\phi(\vec{e}_n)\}$.

For quotient space W/V and W is finite-dimensional, we have:

$$V\xhookrightarrow{\iota}W\xrightarrow{\pi}W/V$$

Then we know that $\pi(\vec{e}_1),\pi(\vec{e}_2),\cdots,\pi(\vec{e}_n)$ spans W/V. Thus, W/V is finitely spanned. ■

Proposition 3.14 $\dim(V_1+V_2) \leq \dim V_1 + \dim V_2$. Equality holds if and only if the sum is direct.

Proof. For V_1 and V_2, we can find the minimal spanning sets S_1 and S_2 respectively. Then we claim that $S_1 \cup S_2$ spans V_1+V_2, i.e., $V_1+V_2 = \mathsf{Span}\{S_1 \cup S_2\}$.

This is because for all $v \in V_1+V_2$, we have $v = v_1+v_2$ for some $v_i \in V_i$. Then we can write v_i as a linear combination of finitely many elements in S_i, i.e., $v_i = \sum_{j=1}^{n_i} \alpha_i^j s_i^j$ where $\alpha_i^j \in \mathbb{F}$ and $s_i^j \in S_i$ are distinct. Thus, we have

$$v = v_1+v_2 = \sum_{j=1}^{n_1} \alpha_1^j s_1^j + \sum_{j=1}^{n_2} \alpha_2^j s_2^j \in \mathsf{Span}\{S_1 \cup S_2\}$$

This shows that $V_1+V_2 \subseteq \mathsf{Span}\{S_1 \cup S_2\}$. The other direction is obvious. Thus, we have $V_1+V_2 = \mathsf{Span}\{S_1 \cup S_2\}$.

Then we have $\dim(V_1+V_2) \leq |S_1|+|S_2| = \dim V_1 + \dim V_2$, as $S_1 \cup S_2$ may not be a minimal spanning set. The equality holds if and only if $S_1 \cup S_2$ is a minimal spanning set of V_1+V_2, which is equivalent to saying that V_1 and V_2 are weakly linearly independent. Thus, the equality holds if and only if the sum is direct. ■

3.8 Exact Sequence

Definition 3.19 — Exact and Exact Sequence. A sequence of linear maps between linear spaces over a field $\mathbb{F}$,

$$\cdots \xrightarrow{f_{i-2}} V_{i-1} \xrightarrow{f_{i-1}} V_i \xrightarrow{f_i} V_{i+1} \xrightarrow{f_{i+1}} \cdots$$

is said to be *exact* at V_i if

$$\mathsf{Im}(f_{i-1}) = \mathsf{Ker}(f_i)$$

i.e., the image of the map before V_i is equal to the kernel of the map after V_i.
The sequence is said to be an *exact sequence* if it is exact at every V_i.

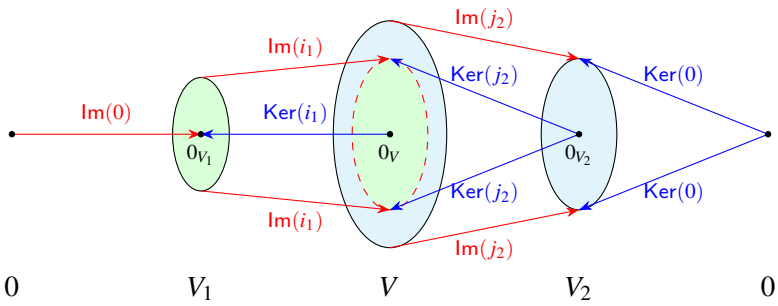
■ **Example 3.7** For the following short exact sequence:

$$0 \longrightarrow V_1 \xrightarrow{i_1} V \xrightarrow{j_2} V_2 \longrightarrow 0$$

for which V_2 is assumed to have a minimal spanning set. Then

- the exactness at V_1 implies that $\{0_{V_1}\} = \mathsf{Im}(0) = \mathsf{Ker}(i_1)$, thus i_1 is injective.
- the exactness at V implies that $\mathsf{Im}(i_1) = \mathsf{Ker}(j_2)$, thus $V_1 \cong \mathsf{Im}(i_1) \subseteq V$.
- the exactness at V_2 implies that $\mathsf{Im}(j_2) = \mathsf{Ker}(0) = V_2$, thus j_2 is surjective.

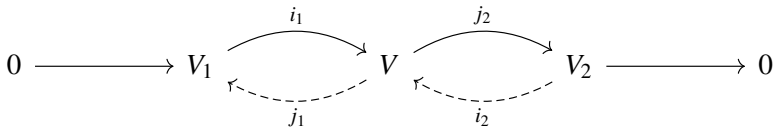
We can draw an Euler diagram to illustrate the situation:



There are some facts about the short exact sequence:

- j_2 has a right inverse, i.e., there exists a linear map $i_2 : V_2 \to V$ such that $j_2 \circ i_2 = \mathsf{id}_{V_2}$.
  This is because V_2 has a minimal spanning set. Thus, for each element in the minimal spanning set of V_2, we can choose one representative in V and define the map on the minimal spanning set. Then we can extend it to the whole space.
- i_1 has a left inverse, i.e., there exists a linear map $j_1 : V \to V_1$ such that $j_1 \circ i_1 = \mathsf{id}_{V_1}$.
  This is because i_1 is injective. Thus, for each element in V_1, we can choose one representative in V and define the map on the whole space by sending all other elements to zero.

The exact sequence becomes:



■

There are some equalities about the composition of the maps in an exact sequence.

- $j_1 \circ i_1 = \mathsf{id}_{V_1}$ because j_1 is a left inverse of i_1.
- $j_2 \circ i_2 = \mathsf{id}_{V_2}$ because i_2 is a right inverse of j_2.
- $j_2 \circ i_1 = 0$ because $\mathsf{Im}(i_1) = \mathsf{Ker}(j_2)$.

- $j_1 \circ i_2 = 0$ because $\mathsf{Im}(i_2) = \mathsf{Ker}(j_1)$.
- $i_1 \circ j_1 + i_2 \circ j_2 = \mathsf{id}_V$ because for all $v \in V$, we have $v = (v - i_2(j_2(v))) + i_2(j_2(v))$ where $v - i_2(j_2(v)) \in \mathsf{Im}(i_1)$ and $i_2(j_2(v)) \in \mathsf{Im}(i_2)$. Also, $\mathsf{Im}(i_1) \cap \mathsf{Im}(i_2) = \{0_V\}$.

There is actually one more fact about the short exact sequence.

Proposition 3.15 $V \cong \mathsf{Im}(i_1) \oplus \mathsf{Im}(i_2)$.

Proof. The meaning of $V \cong \mathsf{Im}(i_1) \oplus \mathsf{Im}(i_2)$ is that for any $x \in V$, it can be uniquely written as $x = x_1 + x_2$ where $x_i \in \mathsf{Im}(i_i)$. Why? Suppose $x = x_1 + x_2 = x_1' + x_2'$ where $x_i, x_i' \in \mathsf{Im}(i_i)$. Then we have $(x_1 - x_1') + (x_2 - x_2') = 0$. Note that $x_1 - x_1' \in \mathsf{Im}(i_1)$ and $x_2 - x_2' \in \mathsf{Im}(i_2)$. Thus, we have $x_1 - x_1' = 0$ and $x_2 - x_2' = 0$. This shows the uniqueness.

Note that all V, V_1 and V_2 are finite-dimensional. Then V_2 has a minimal spanning set, let say S. Then we construct $i_2 : s \mapsto i_2(s)$ where $i_2(s)$ is a choice of element from $j_2^{-1}(s) \neq \emptyset$ for each $s \in S$. Then we extend it to the whole space linearly. Thus, i_2 is injective.

Then we want to prove that $\mathsf{Im}(i_1)$ and $\mathsf{Im}(i_2)$ are weakly independent. Assume that $x_1 + x_2 = 0$ where $x_i \in \mathsf{Im}(i_i)$. Then we have $j_2(x_1 + x_2) = j_2(x_1) + j_2(x_2) = 0$. Note that $j_2(x_1) = 0$ because $x_1 \in \mathsf{Im}(i_1) = \mathsf{Ker}(j_2)$, the exactness of V. Thus, we have $j_2(x_2) = 0$. However, j_2 is injective on $\mathsf{Im}(i_2)$ because $j_2 \circ i_2 = \mathsf{id}_{V_2}$. Thus, we have $x_2 = 0$ and $x_1 = 0$. This shows that $\mathsf{Im}(i_1)$ and $\mathsf{Im}(i_2)$ are weakly independent.

Finally, we want to prove that $\mathsf{Im}(i_1) + \mathsf{Im}(i_2) = V$. For all $x \in V$, we let $x_2 = i_2(j_2(x)) \in \mathsf{Im}(i_2)$ and $x_1 = x - x_2$. Then we have to show that $x_1 \in \mathsf{Im}(i_1) = \mathsf{Ker}(j_2)$. Note that $j_2(x) = j_2(x_1) + j_2(x_2) = j_2(x_1) + j_2 \circ i_2(j_2(x)) = j_2(x_1) + j_2(x)$. This shows that $j_2(x_1) = 0$. Thus, $x_1 \in \mathsf{Ker}(j_2) = \mathsf{Im}(i_1)$. This shows that $\mathsf{Im}(i_1) + \mathsf{Im}(i_2) = V$.

Actually j_1 is the projection from $\mathsf{Im}(i_1) \oplus \mathsf{Im}(i_2)$ to $\mathsf{Im}(i_1)$ and it exists due to the uniqueness of the decomposition. ■

The equalities can be summarized as follows:

$$j_m \circ i_n = \delta_{mn} \mathsf{id}_{V_n}, \quad \sum_{k=1}^{2} i_k \circ j_k = \mathsf{id}_V$$

For the dimension of the spaces, we have:

$$\dim V = \dim \mathsf{Im}(i_1) + \dim \mathsf{Im}(i_2) = \dim V_1 + \dim V_2$$

As $V_1 \cong \mathsf{Im}(i_1)$ and $V_2 \cong \mathsf{Im}(i_2)$. i_1 and i_2 are injective and $V_k \to \mathsf{Im}(i_k)$ are surjective.

Also, we know that $\dim V \geq \dim V_1$ and $\dim V \geq \dim V_2$. Similarly, we have $\dim W \geq \dim V$ and $\dim W \geq \dim W/V$, where V is a subspace of W.

Consider Proposition 3.14, more specifically, we have the following dimension formula:

$$\dim (V_1 + V_2) = \dim V_1 + \dim V_2 - \dim (V_1 \cap V_2)$$

To proof the equality, we can consider the following short exact sequence:

$$0 \longrightarrow V_1 \cap V_2 \xrightarrow{\iota} V_1 \xrightarrow{\pi} (V_1 + V_2)/V_2 \longrightarrow 0$$

Moreover, we have the isomorphism between $(V_1 + V_2)/V_2$ and $V_1/(V_1 \cap V_2)$.

"No problem is difficult in linear algebra.
All problems are trivial."

GUOWU MENG

3.9 Fudan University Problems

Students from Fudan University asked two hard problems but were completely cooked by Professor Guowu Meng

3.9.1 The story behind the two problems

"Well, [in] linear algebra basically, no problem is difficult. All problems are trivial.

"People don't believe me, because many years ago, more than 20 years ago, there were two exchange students from Fudan University, and when they came here, they carry solution manual with some sets of hard linear algebra problems. I told them 'nothing is difficult'.

"They don't believe me, so they dig out one hard problem from that solution book. Well, I told them I haven't seen this problem before, because when I was educated as a physicist engineer, I don't work on hard problems. I just deal with textbook. I don't read anything extra. I don't know but doesn't matter. Let me just write everything on board, and then pretty soon I figured out the answer.

"Ok may be they say that I am lucky. Then the next day they came back with another problem. So again, I said I don't know how to do it but anyway [it] doesn't matter. I put everything on board, then I draw some obvious facts in my mind about linear algebra.

"I say no problems are difficult in linear algebra under the assumption that you know linear algebra inside-out, you know every facts about it. Usually you will say I have seen this type of problems before, and then step 1, step 2 step 3, but this is a very wrong way to do it. This is the way that AI does it, but we are human, we are smarter than machine.

"When I do it, there are some keywords and each keywords remind me of some facts related to it, and keep doing this. Then I see a path from here to there."

— Guowu Meng on the lecture of September 19, 2025.

3.9.2 Introduction to the two problems

Later, we will examine the two problems that were posed by students from Fudan University and solved by Professor Guowu Meng. Before examining the two problems, we need to introduce some basic terminology in standard linear algebra.

Let A be a $m \times n$ matrix. Then we consider the following diagram:

$$\mathsf{Ker}(f) \subseteq \mathbb{F}^n \xrightarrow[A]{f} \mathbb{F}^m \supseteq \mathsf{Im}(f)$$

In normal linear algebra, we have four fundamental concepts: column space, null space, rank and nullity.

Definition 3.20 — Column Space. The *column space* of A, denoted by $\mathsf{Col}(A)$, is defined as the image of the linear map $f : \mathbb{F}^n \to \mathbb{F}^m$ defined by $f(x) = Ax$, i.e.,

$$\mathsf{Col}(A) = \mathsf{Im}(f) = \{Ax \mid x \in \mathbb{F}^n\} \subseteq \mathbb{F}^m$$

Definition 3.21 — Null Space. The *null space* of A, denoted by $\mathsf{Nul}(A)$, is defined as the kernel of the linear map $f : \mathbb{F}^n \to \mathbb{F}^m$ defined by $f(x) = Ax$, i.e.,

$$\mathsf{Nul}(A) = \mathsf{Ker}(f) = \{x \in \mathbb{F}^n \mid Ax = 0\} \subseteq \mathbb{F}^n$$

The alternative, or normal, definition of rank is as follows.

Definition 3.22 — Rank. The *rank* of A, denoted by $\mathsf{Rank}(A)$, is defined as the dimension of the column space of A, i.e.,

$$\mathsf{Rank}(A) = \dim \mathsf{Col}(A) = \dim \mathsf{Im}(f) \leq m$$

Definition 3.23 — Nullity. The *nullity* of A, denoted by $\mathsf{Nullity}(A)$, is defined as the dimension of the null space of A, i.e.,

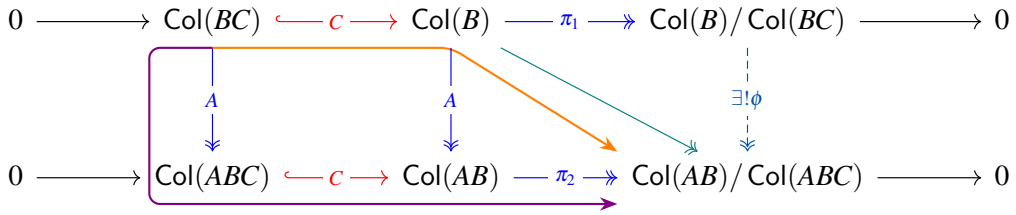
$$\mathsf{Nullity}(A) = \dim \mathsf{Nul}(A) = \dim \mathsf{Ker}(f) \leq n$$

3.9.3 Problem 1

Problem 3.1 Suppose we have three matrices A, B and C. Then prove that

$$\mathsf{Rank}(B) + \mathsf{Rank}(ABC) \geq \mathsf{Rank}(AB) + \mathsf{Rank}(BC)$$

Proof. We consider the following diagram:



We denote the injective map with red color and the surjective map with blue color. Notice that there is a surjective map from $\mathsf{Col}(B)$ to $\mathsf{Col}(AB)/\mathsf{Col}(ABC)$ due to the surjectivity of A and π_2. Then we denote this surjective map with teal color.

Then we have to consider whether the map from $\mathsf{Col}(BC)$ to $\mathsf{Col}(AB)/\mathsf{Col}(ABC)$ is zero. If the map is zero, then we can construct a unique surjective map ϕ from $\mathsf{Col}(B)/\mathsf{Col}(BC)$ to $\mathsf{Col}(AB)/\mathsf{Col}(ABC)$ due to the universal property of quotient space.

Note that the map from $\mathsf{Col}(BC)$ to $\mathsf{Col}(AB)/\mathsf{Col}(ABC)$ is a zero map. As both upper and lower sequences are exact, we have the exactness at $\mathsf{Col}(AB)$, i.e., $\mathsf{Im}(C) = \mathsf{Ker}(\pi_2)$. Thus the composite map $\pi_2 \circ C$ is a zero map. This shows that the map from $\mathsf{Col}(BC)$ to $\mathsf{Col}(AB)/\mathsf{Col}(ABC)$ is a zero map.

Then we can construct a unique surjective map ϕ from $\mathsf{Col}(B)/\mathsf{Col}(BC)$ to $\mathsf{Col}(AB)/\mathsf{Col}(ABC)$ due to the universal property of quotient space.

Finally, we consider the dimensions of the spaces. Note that ϕ is surjective, thus we have

$$\begin{aligned}
\dim \mathsf{Col}(B)/\mathsf{Col}(BC) &\geq \dim \mathsf{Col}(AB)/\mathsf{Col}(ABC) \\
\dim \mathsf{Col}(B) - \dim \mathsf{Col}(BC) &\geq \dim \mathsf{Col}(AB) - \dim \mathsf{Col}(ABC) \\
\dim \mathsf{Col}(B) + \dim \mathsf{Col}(ABC) &\geq \dim \mathsf{Col}(AB) + \dim \mathsf{Col}(BC) \\
\mathsf{Rank}(B) + \mathsf{Rank}(ABC) &\geq \mathsf{Rank}(AB) + \mathsf{Rank}(BC)
\end{aligned}$$

■

3.9.4 Problem 2

Problem 3.2 If A is a $n \times n$ matrix then prove that

$$\mathsf{Rank}(A^n) = \mathsf{Rank}(A^{n+1})$$

Proof. We consider the following diagram:

$$I_n \xrightarrow{A} \mathsf{Im}(A) \xrightarrow{A} \mathsf{Im}(A^2) \xrightarrow{A} \cdots \xrightarrow{A} \mathsf{Im}(A^n) \xrightarrow{A} \cdots$$

As $I_n \supseteq \mathsf{Im}(A) \supseteq \mathsf{Im}(A^2) \supseteq \cdots$, we know that

$$n = \dim I_n \geq r(A) \geq r(A^2) \geq \cdots$$

As the space is finite-dimensional, the sequence will eventually become constant. That means there exists a k such that for all $j \geq k$, we have $r(A^j) = r(A^{j+1})$.

There are two possibilities: either $k \leq n$ or $k > n$. If $k \leq n$, the equality works properly, as for every $j \geq k$, including $j = n$, such that $r(A^j) = r(A^{j+1})$ implies $r(A^n) = r(A^{n+1})$.

For $k > n$, consider the strict inequality, we know that each time the dimension must drop at least 1. Without the loss of generality, we may consider the sequence of dimension as $n, n-1, n-2, \cdots, 1, 0$. This involves n times. So it is impossible to have $k > n$. ■

3.10 Rank-Nullity Theorem

Actually, using short exact sequence, we can easily prove the rank-nullity theorem.

Theorem 3.1 — Rank-Nullity Theorem. For a linear map $f : V \to W$ between finite-dimensional linear spaces over $\mathbb{F}$, we have

$$\mathsf{Rank}(f) + \mathsf{Nullity}(f) = \dim V$$

Proof. Consider the following short exact sequence:

$$0 \longrightarrow \mathsf{Ker}(f) \xhookrightarrow{\iota} V \xrightarrow{f} \mathsf{Im}(f) \longrightarrow 0$$

Then we have $V \cong \mathsf{Ker}(f) \oplus \mathsf{Im}(f)$. Thus, we have $\dim V = \dim \mathsf{Ker}(f) + \dim \mathsf{Im}(f)$. This shows that $\mathsf{Rank}(f) + \mathsf{Nullity}(f) = \dim V$. ■

Moreover, we have the following corollary.

Corollary 3.2 For a linear map $f : V \to W$ between finite-dimensional linear spaces over $\mathbb{F}$, we have

$$\dim W = \mathsf{Rank}(f) + \dim \mathsf{Coker}(f)$$

Proof. Consider the following short exact sequence:

$$0 \longrightarrow \mathsf{Im}(f) \xhookrightarrow{\iota} W \xrightarrow{\pi} \mathsf{Coker}(f) \longrightarrow 0$$

Then we have $W \cong \mathsf{Im}(f) \oplus \mathsf{Coker}(f)$. Thus, we have $\dim W = \dim \mathsf{Im}(f) + \dim \mathsf{Coker}(f)$. This shows that $\dim W = \mathsf{Rank}(f) + \dim \mathsf{Coker}(f)$. ■

Corollary 3.3 For a linear map $f : V \to W$ between finite-dimensional linear spaces over $\mathbb{F}$, we have

$$\dim V = \mathsf{Nullity}(f) + \dim \mathsf{Coim}(f)$$

Proof. Consider the following short exact sequence:

$$0 \longrightarrow \mathsf{Ker}(f) \xhookrightarrow{\iota} V \xrightarrow{\pi} \mathsf{Coim}(f) \longrightarrow 0$$

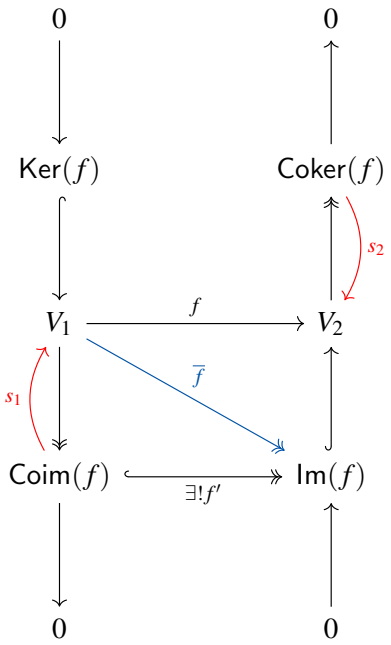
Then we have $V \cong \mathsf{Ker}(f) \oplus \mathsf{Coim}(f)$. Thus, we have $\dim V = \dim \mathsf{Ker}(f) + \dim \mathsf{Coim}(f)$. This shows that $\dim V = \mathsf{Nullity}(f) + \dim \mathsf{Coim}(f)$. ■

Moreover, we have the following properties for rank:

1. The rank of a matrix is invariant under elementary row and column operations.
2. $\mathsf{Rank}(A+B) \leq \mathsf{Rank}(A) + \mathsf{Rank}(B)$
3. $\mathsf{Rank}(AB) \leq \mathsf{Rank}(A)$ and $\mathsf{Rank}(AB) \leq \mathsf{Rank}(B)$

3.11 Canonical Form of Linear Map

First, let $f : V_1 \to V_2$ be a linear map between finite-dimensional linear spaces over $\mathbb{F}$. Recall that $\mathsf{Ker}(f) = f^{-1}(0)$, $\mathsf{Im}(f) = \{f(v_1) \mid v_1 \in V_1\}$, $\mathsf{Coim}(f) = V_1/\mathsf{Ker}(f)$ and $\mathsf{Coker}(f) = V_2/\mathsf{Im}(f)$. We have the following commutative diagram:



Here, each column is an exact sequence, and the square in the middle is commutative, as the lower left triangle and upper right triangle are commutative.

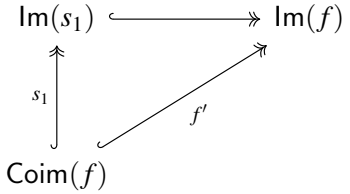
Moreover, the f', the universal property for quotient map, is a linear equivalence. It is injective due to the trivial $\mathsf{Ker}(f')$.

s_1 and s_2 are the *right inverses* or called *sections*.

With respect to the decomposition of V_1 and V_2 into subspaces, i.e., $V_1 = \mathsf{Im}(s_1) \oplus \mathsf{Ker}(f)$ and $V_2 = \mathsf{Im}(f) \oplus \mathsf{Im}(s_2)$, the linear map f is decomposed as follows:

$$\begin{array}{c} f = \begin{bmatrix} \widetilde{f} & 0 \\ 0 & 0 \end{bmatrix} \\ \mathsf{Im}(s_1) \oplus \mathsf{Ker}(f) \longrightarrow \mathsf{Im}(f) \oplus \mathsf{Im}(s_2) \end{array}$$

where $\widetilde{f} : \mathsf{Im}(s_1) \to \mathsf{Im}(f)$ is a linear equivalence, as there are linear equivalences $f' : \mathsf{Coim}(f) \to \mathsf{Im}(f)$ and $s_1 : \mathsf{Coim}(f) \to \mathsf{Im}(s_1)$. Then the graph below commutes:



> *Remark.* The choice of s_1 and s_2 is not unique, so the decomposition of V_1 and V_2, and hence f, is not unique.

The matrix $\begin{bmatrix} \widetilde{f} & 0 \\ 0 & 0 \end{bmatrix}$ is the canonical form of the linear map. Just as the canonical form of a matrix, it reveals the essential structure of the linear map. However, the rank of $\widetilde{f}$ is unique, which is equal to $\mathsf{Rank}(f) = \dim \mathsf{Im}(f)$.

$$\mathbb{F}^r \oplus \mathbb{F}^{n-r} \xrightarrow{\begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}} \mathbb{F}^r \oplus \mathbb{F}^{n-r}$$

Moreover, from the diagram of two exact sequences, we can see that f can be decomposed into two linear maps: $f = \iota \circ \overline{f}$, where $\overline{f} : V_1 \to \mathsf{Coim}(f)$ is a surjective map and $\iota : \mathsf{Coim}(f) \to V_2$ is an injective map. Note that the decomposition is not unique, as we can choose the path from V_1 to $\mathsf{Coim}(f)$ then to V_2.

3.12 Free Vector Space

Let X be a set and $\delta_X = \{\delta_x \mid x \in X\}$. Here $\delta_x : X \to \mathbb{F}$ is the δ-function at x.

Proposition 3.16 δ_X is a linearly independent set of $\mathbb{F}[[X]]$ = the linear space of $\mathbb{F}$-valued functions on X.

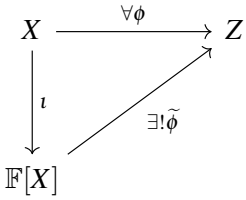
Proposition 3.17 $\mathsf{Span}(\delta_X) = \mathbb{F}[X]$

Thus, δ_X is a minimal spanning set for $\mathbb{F}[X]$.

Proposition 3.18 There is a natural set isomorphism $X \to \delta_X$ which maps x to δ_x.

Then we have an injective set map $\iota : X \equiv \delta_X \to \mathbb{F}[X]$ which maps x to δ_x. This is a set mapping to a linear space.

Among all set maps from X to a linear space over $\mathbb{F}$, the set map $\iota : X \to \mathbb{F}[X]$ is universal in the following sense:

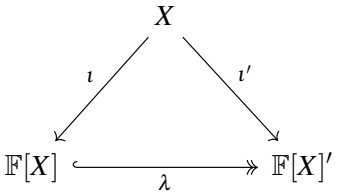


For any set map $\phi : X \to Z$, there exists a unique linear map $\widetilde{\phi} : \mathbb{F}[X] \to Z$ such that $\widetilde{\phi} \circ \iota = \phi$.

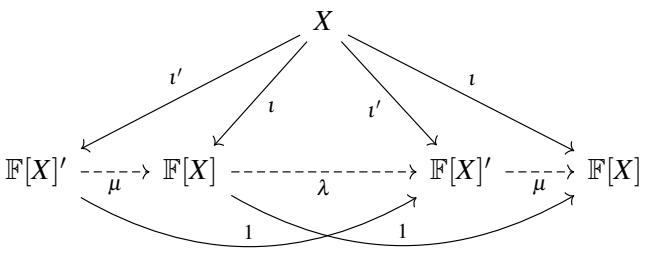
Proof. Assume the existence of such $\widetilde{\phi}$, then $\widetilde{\phi} \circ \iota(x) = \phi(x)$ for all $x \in X$, i.e., $\widetilde{\phi}(\delta_x) = \phi(x)$ for all $x \in X$. As $\{\delta_x \mid x \in X\}$ is a minimal spanning set for $\mathbb{F}[X]$, $\widetilde{\phi}$ must be the linear map such that $\widetilde{\phi}(\delta_x) = \phi(x)$, thus unique. Existence of $\widetilde{\phi}$ is also proved. ■

Via the natural identification of $\delta_X \equiv X$ ($\delta_x \equiv x$), an element $\sum \alpha_x \delta_x \in F[X]$, where the sum is finite and $\alpha_x \in \mathbb{F}$, is naturally identified with $\sum \alpha_x x$, which is called a *formal linear combination* of elements in X. Hereafter, we always use this natural identification, so $\mathbb{F}[X]$ is now defined as *the set of formal linear combinations of elements in the set* X. Then $\iota : X \to \mathbb{F}[X]$ is just the inclusion map : $x \mapsto x$.

The universal map is unique in the following sense: suppose that $\iota' : X \to \mathbb{F}[X]'$ is another inclusion map, then there is a unique linear equivalence λ in the commutative triangle:

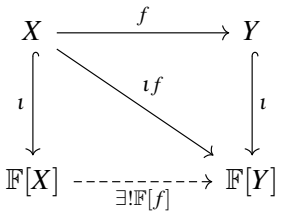


This can be seen from the following diagram:

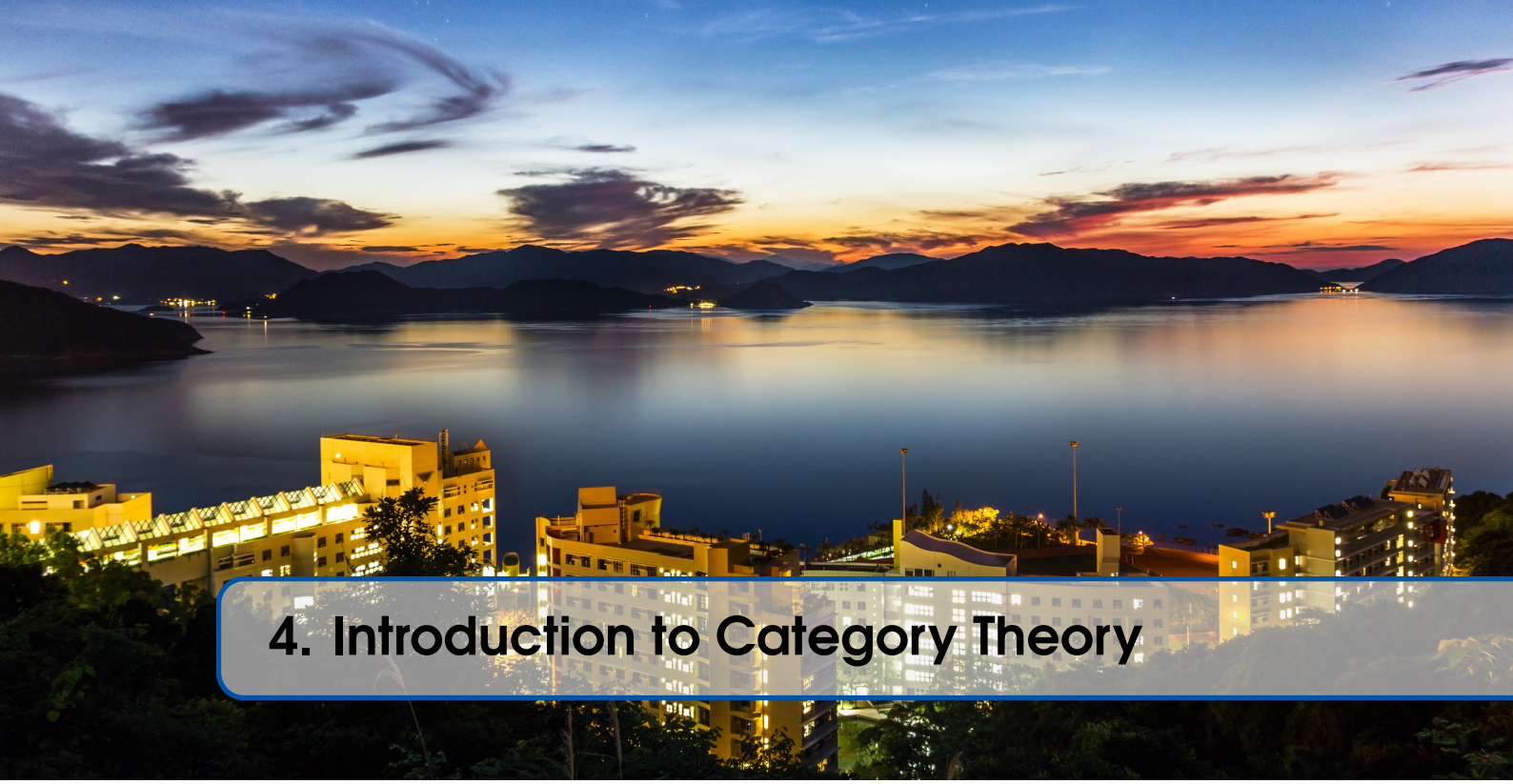


λ exists because ι is universal, and μ exists because ι' is universal. $\lambda\mu = 1$ because ι' is universal, same for $\mu\lambda = 1$. Then λ is isomorphism.

The universal property implies an assignment of a linear map $\mathbb{F}[f] : \mathbb{F}[X] \to \mathbb{F}[Y]$ to any set map $f : X \to Y$. Indeed,



Moreover, $\mathbb{F}[1_X] = 1_{\mathbb{F}[X]}$ or simply $\mathbb{F}[1] = 1$ for all X, and $\mathbb{F}[fg] = \mathbb{F}[f]\mathbb{F}[g]$ for all $f : Y \to Z$ and $g : X \to Y$.



4. Introduction to Category Theory

"In linear algebra, all the proofs should be straight-forward. There is no trick. If you think it's very hard, there is something wrong"

GUOWU MENG

4.1 Categories and Functors

The collection of set maps is denoted by **Set** and the collection of linear maps over $\mathbb{F}$ is denoted by $\mathbf{Vec}_{\mathbb{F}}$. There is a diagram below:

$$\begin{array}{c} \mathbf{Set} \\ \downarrow{\scriptstyle \mathbb{F}[-]} \\ \mathbf{Vec}_{\mathbb{F}} \end{array}$$

where $\mathbb{F}[-]$ sends set map $f : X \to Y$ to a linear map $\mathbb{F}[f] : \mathbb{F}[X] \to \mathbb{F}[Y]$.

$\mathbb{F}[-]$ is an example of functors.

Monoid homomorphisms are another example of functors: in particular group homomorphisms

$$\begin{array}{c} M_1 \\ \downarrow{\scriptstyle \phi} \\ M_2 \end{array}$$

An element $a \in M_1$ is viewed as an arrow, or morphism, that sends $*$ to $*$, i.e., $a : * \to *$. Then ab is viewed as the composition of arrows:

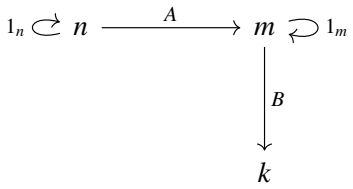
$$* \xrightarrow{b} * \xrightarrow{a} *, \qquad * \xrightarrow{ab} *$$

Recall that a monoid M is a set, which is called a small collection of objects, together with a binary operation, which is also called composition, on M with both the associtivity law and identity law satisfied.

By relaxing the condition on binary operation, allowing the composition being partially defined, we end up with the notion of *small category*.

Being partially defined means that the composition may not always be defined. For example, take $f : X \to Y$ and $g : W \to Z$, then gf is not defined. But for normal, $f : X \to Y$ and $g : Y \to Z$, then gf is defined. In monoid, as we may suggest there is only one element $*$, then the composition is always defined.

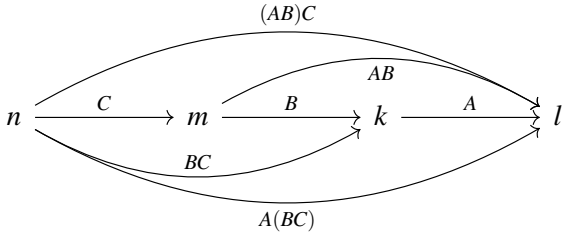
An example of a small category: the collection of all matrices over $\mathbb{F}$. We may consider any $m \times n$ matrix as an arrow that sends n to m: $A : n \to m$. If we have a $k \times m$ matrix B that sends m to k, then we have the composition $BA : n \to k$. Note that $I_n : n \to n$ is the identity, which is not unique, there can be I_m and I_k. We have



Note that $A1_n = A = 1_m A$ and $B1_m = B$.

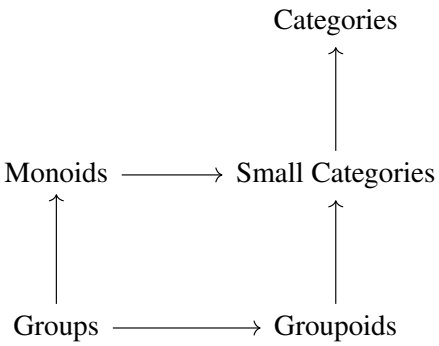
> *Remark.* The identity elements are not unique unlike the case of monoid.

The following shows the associativity law:



Hence, the set of all matrices form a small category.

Consider the set of all invertible matrices over $\mathbb{F}$, it is also a small category, in fact, it is a *groupoid*. Groupoid is defined as a small category such that every morphism is invertible.



The graph above shows the relation, the arrows show the subsets relation. The arrow head is the larger set and arrow tail is the subset.

4.2 Small Categories

Definition 4.1 — Small Categories. A small category is a set $\mathcal{C}$ together with a subset $\mathcal{C}_0$ of $\mathcal{C}$, two surjective maps $s,t : \mathcal{C} \to \mathcal{C}_0$ and a composition map $\mathcal{C} \times_{(s,t)} \mathcal{C} \to \mathcal{C}$ that sends (f,g) to fg which satisfies the identity law and associativity law.

Here $\mathcal{C} \times_{s,t} \mathcal{C}$ is defined as the pullback of the diagram below:

$$\begin{array}{ccc} \mathcal{C} \times_{s,t} \mathcal{C} & \xrightarrow{p_1} & \mathcal{C} \\ \downarrow{\scriptstyle p_2} & \ulcorner & \downarrow{\scriptstyle t} \\ \mathcal{C} & \xrightarrow{s} & \mathcal{C}_0 \end{array}$$

where the set $\mathcal{C} \times_{s,t} \mathcal{C} = \{(x,y) \in \mathcal{C} \times \mathcal{C} \mid s(x) = t(y)\}$. Intuitively, the pullback is to filter out the mappings that can do composition, such as $f,g \in \mathcal{C} \times_{(s,t)} \mathcal{C}$ where $A \xrightarrow{f} B \xrightarrow{g} C$.

The s and t are called the *source map* and *target map* respectively. We can picture the composition graphically as follows:

$$\begin{array}{ccccccc} * & \xleftarrow{f} & * & \quad * & \xleftarrow{g} & * & \quad * & \xleftarrow{fg} & * \\ t(f) & & s(f) = t(g) & & s(g) & & t(f) & & s(g) \end{array}$$

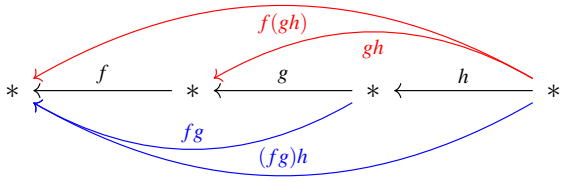
The left diagram is the equivalent to the right one.

We may draw the identity law this way:

$$\begin{array}{ccccccc} 1_{t(f)} \circlearrowright * & \xleftarrow{f} & * & \quad * & \xleftarrow{f} & * & \quad * & \xleftarrow{f} & * \circlearrowright 1_{s(f)} \\ t(f) & & s(f) & & t(f) & & s(f) & & t(f) & & s(f) \end{array}$$

The three diagrams are equivalent.

We may draw the associativity law this way:



■ **Example 4.1** In the small category of matrices over $\mathbb{F}$, we have

$$\mathcal{C} = \{\mathsf{M}_{m\times n}(\mathbb{F}) \mid m,n \in \mathbb{N}\}$$
$$\mathcal{C}_0 = \{I_n \mid n \in \mathbb{N}\} \equiv \mathbb{N}$$

If $A \in \mathcal{C}$ is an $m \times n$ matrix, then $s(A) = I_n \equiv n$ and $t(A) = I_m \equiv m$. We can represent A as follows:

$$\begin{array}{ccc} * & \xleftarrow{A} & * \\ m & & n \end{array}$$

Note that $(A,B) \in \mathcal{C} \times_{s,t} \mathcal{C}$, where the composition of A and B defined as the matrix multiplication AB, means for some positive integer m,n and k:

$$\begin{array}{ccccc} * & \xleftarrow{A} & * & \xleftarrow{B} & * \\ m & & n & & k \end{array}$$

■

Remark. Elements in $\mathcal{C}$ are *morphisms* or *arrows*, and elements in $\mathcal{C}_0$ are *identity morphisms*. A morphism f is viewed as an arrow from $s(f) \in C_0$ to $t(f) \in C_0$, i.e., $f : s(f) \to t(f)$. An identity morphism is drawn in the following way with X being called the *object*:

$$\begin{array}{ccc} * & \xleftarrow{1_X} & * \\ X & & X \end{array}$$

In the last example, I_n is the identity morphism at n. So $\mathcal{C}_0$ is also called the set of objects. Then a morphism f is viewed as an arrow from object $X \equiv 1_X = s(f)$ to object $Y \equiv 1_Y = t(f)$, i.e., $f : X \to Y$.

So, normally, we denote a small category as $\mathcal{C}$ and its set of objects as $\mathcal{C}_0$.

Remark. The set of morphisms from object X to object Y is denoted by $\mathsf{Mor}(X,Y)$. In the last example, $\mathsf{Mor}(m,n) = \mathsf{M}_{m\times n}(\mathbb{F})$, the set of all $m \times n$ matrices over $\mathbb{F}$. Note that $1_X \in \mathsf{Mor}(X,X)$, so $\mathsf{Mor}(X,X) \neq \emptyset$ for all $X \in \mathcal{C}_0$.

Then $\mathcal{C}$ is the disjoint union of all $\mathsf{Mor}(X,Y)$ for all pairs of objects (X,Y):

$$\mathcal{C} = \bigsqcup_{X,Y\in\mathcal{C}_0} \mathsf{Mor}(X,Y)$$

Remark. The composition can be written as follows:

$$\begin{array}{ccc} \mathsf{Mor}(Y,Z) \times \mathsf{Mor}(X,Y) & \longrightarrow & \mathsf{Mor}(X,Z) \\ (Z \xleftarrow{f} Y, X \xleftarrow{g} Y) & \longmapsto & X \xleftarrow{fg} Z \end{array}$$

Then the following is the second definition of small category, which is also the normal definition of a small category.

Definition 4.2 — Small Categories. A small category $\mathcal{C}$ is a collection of the following data:

1. A set of objects $\mathcal{C}_0$;
2. A set of morphisms $\mathsf{Mor}(X,Y)$ for each pair of objects (X,Y);
3. A composition map $\mathsf{Mor}(Y,Z) \times \mathsf{Mor}(X,Y) \to \mathsf{Mor}(X,Z)$ that sends (f,g) to fg for each triple of objects (X,Y,Z);
4. An identity morphism $1_X \in \mathsf{Mor}(X,X)$ for each object X;

Moreover, these data satisfies the following conditions:

(a) (Identity Law) For all $f \in \mathsf{Mor}(X,Y)$, we have $f1_X = f = 1_Y f$;
(b) (Associativity Law) For all appropriate morphisms f,g,h, we have $(fg)h = f(gh)$.

For a small category $\mathcal{C}$, the set of objects is denoted by $\mathsf{Ob}(\mathcal{C})$ and the set of morphisms for any pair of objects (X,Y) is denoted by $\mathsf{Mor}(X,Y)$, $\mathsf{Mor}_{\mathcal{C}}(X,Y)$, $\mathrm{Hom}_{\mathcal{C}}(X,Y)$ or simply $\mathcal{C}(X,Y)$.

If we allow $\mathsf{Ob}(\mathcal{C})$ and $\mathsf{Mor}_{\mathcal{C}}(X,Y)$ for any pair of objects (X,Y) being a *class*, (a larger collection than set), we end up with the definition of *category*.

We say a morphism is *isomorphic* or *invertible* if it has a two-sided inverse. A category such that every morphism is isomorphic is called a *groupoid*.

■ **Example 4.2** The collection of all sets and set maps, denoted by **Set**, is a category. ■

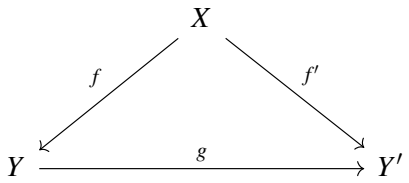
■ **Example 4.3** The collection of all linear spaces over $\mathbb{F}$ and linear maps, denoted by $\mathbf{Vec}_{\mathbb{F}}$, is a category. ■

■ **Example 4.4** If $\mathcal{C}$ and $\mathcal{D}$ are two categories, then we have the product category $\mathcal{C} \times \mathcal{D}$ with objects (X,Y) and morphisms (f,g), where $X \in \mathsf{Ob}(\mathcal{C})$, $Y \in \mathsf{Ob}(\mathcal{D})$, $f \in \mathsf{Mor}_{\mathcal{C}}(X,X')$ and $g \in \mathsf{Mor}_{\mathcal{D}}(Y,Y')$.
■

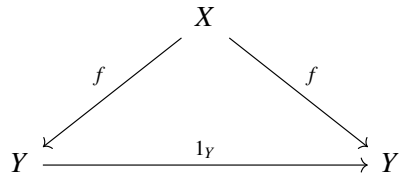
■ **Example 4.5** The category of set maps between finite sets, denoted by **FinSet**, is a subcategory of **Set**. ■

■ **Example 4.6** Fix an object X in a category $\mathcal{C}$. Then the collection of all morphisms with source X, denoted by $\mathcal{C}(X,-)$, is a new category:

- Objects: all morphisms $f : X \to Y$ in $\mathcal{C}$ for all $Y \in \mathsf{Ob}(\mathcal{C})$;
- Morphisms: commutative triangles in $\mathcal{C}$:



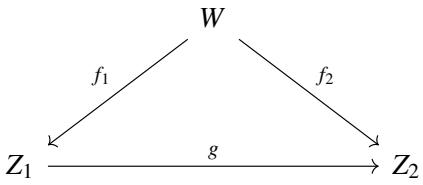
- The identity morphism at object $f : X \to Y$ is the commutative triangle in $\mathcal{C}$:



■

■ **Example 4.7** Let V be a subspace of the linear space W over $\mathbb{F}$. Then we have a category:

- Objects: all morphisms $f : W \to Z$ in $\mathbf{Vec}_{\mathbb{F}}$ such that $f|_V = 0$;
- Morphisms: commutative triangles in $\mathbf{Vec}_{\mathbb{F}}$:



■

Definition 4.3 — Terminal Object and Initial Object. Let $\mathcal{C}$ be a category. An object $T \in \mathsf{Ob}(\mathcal{C})$ is called a *terminal object* if for all object X, there exists a unique morphism from X to T, i.e., $|\mathcal{C}(X,T)| = 1$. An object $I \in \mathsf{Ob}(\mathcal{C})$ is called an *initial object* if for all object X, there exists a unique morphism from I to X, i.e., $|\mathcal{C}(I,X)| = 1$.

Corollary 4.1 A terminal object or an initial object is unique up to isomorphism.

■ **Example 4.8** In the last example of category, the quotient map $\pi : W \to W/V$ is an initial object and the zero map $0 : W \to 0$ is a terminal object. ■

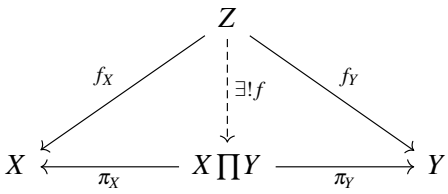
■ **Example 4.9** In **Set**, any singleton set is a terminal object, and the empty set is an initial object. ■

■ **Example 4.10** In $\mathbf{Vec}_{\mathbb{F}}$, the zero vector space is both a terminal object and an initial object. ■

4.3 Products and Coproducts

4.3.1 Products

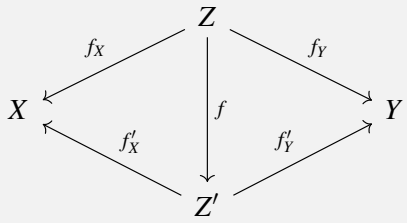
Definition 4.4 — Products. Let $\mathcal{C}$ be a category and $X, Y \in \mathsf{Ob}(\mathcal{C})$. The *product* of X and Y is an object $X \prod Y$ together with two morphisms $\pi_X : X \prod Y \to X$ and $\pi_Y : X \prod Y \to Y$ such that for any object Z and any two morphisms $f_X : Z \to X$ and $f_Y : Z \to Y$, there exists a unique morphism $f : Z \to X \prod Y$ such that the following diagram commutes:



Remark. The product is unique up to isomorphism if it exists.

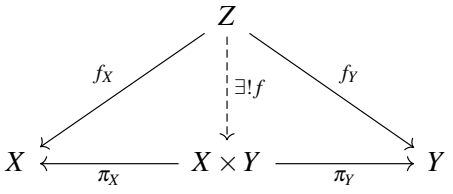
Corollary 4.2 Let $\mathcal{C}$ be a category and $X, Y \in \mathsf{Ob}(\mathcal{C})$. Consider the following new category:

- Objects: all morphisms $X \xleftarrow{f_X} Z \xrightarrow{f_Y} Y$ in $\mathcal{C}$ for all $Z \in \mathsf{Ob}(\mathcal{C})$;
- Morphisms: commutative diagrams in $\mathcal{C}$:



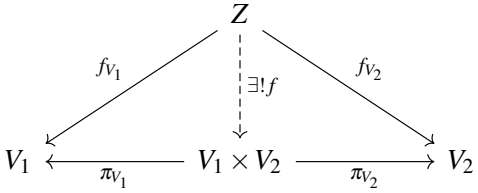
Then the product of X and Y is a terminal object in this new category.

■ **Example 4.11** In **Set**, the product of two sets X and Y is the Cartesian product $X \times Y = \{(x,y) \mid x \in X, y \in Y\}$ with the projection maps $\pi_X(x,y) = x$ and $\pi_Y(x,y) = y$. Then with $f(z) = (f_X(z), f_Y(z))$ for all $z \in Z$, we have the following commutative diagram:



■

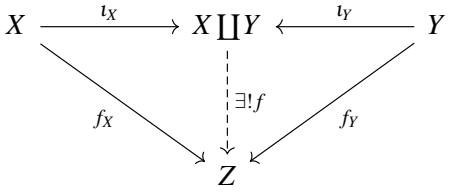
■ **Example 4.12** In $\mathbf{Vec}_{\mathbb{F}}$, the product of two linear spaces V_1 and V_2 over $\mathbb{F}$ is the direct product $V_1 \times V_2 = \{(v_1, v_2) \mid v_1 \in V_1, v_2 \in V_2\}$ with the projection maps $\pi_{V_1}(v_1, v_2) = v_1$ and $\pi_{V_2}(v_1, v_2) = v_2$. Then with $f(z) = (f_{V_1}(z), f_{V_2}(z))$ for all $z \in Z$, we have the following commutative diagram:



■

4.3.2 Coproducts

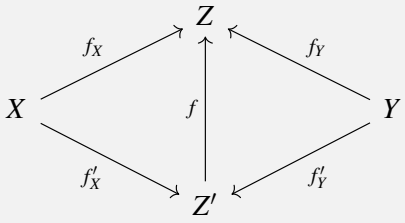
Definition 4.5 — Coproducts. Let $\mathcal{C}$ be a category and $X, Y \in \mathsf{Ob}(\mathcal{C})$. The *coproduct* of X and Y is an object $X \coprod Y$ together with two morphisms $\iota_X : X \to X \coprod Y$ and $\iota_Y : Y \to X \coprod Y$ such that for any object Z and any two morphisms $f_X : X \to Z$ and $f_Y : Y \to Z$, there exists a unique morphism $f : X \coprod Y \to Z$ such that the following diagram commutes:



Remark. The coproduct is unique up to isomorphism if it exists.

Corollary 4.3 Let $\mathcal{C}$ be a category and $X, Y \in \mathsf{Ob}(\mathcal{C})$. The *coproduct* of X and Y is the initial object in the new category:

- Objects: all morphisms $X \xrightarrow{f_X} Z \xleftarrow{f_Y} Y$ in $\mathcal{C}$ for all $Z \in \mathsf{Ob}(\mathcal{C})$;
- Morphisms: commutative diagrams in $\mathcal{C}$:

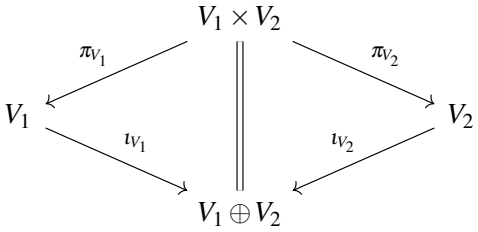


■ **Example 4.13** In **Set**, the coproduct of two sets X and Y is the disjoint union $X \sqcup Y = \{(x,1) \mid x \in X\} \cup \{(y,2) \mid y \in Y\}$. ■

■ **Example 4.14** In $\mathbf{Vec}_{\mathbb{F}}$, the coproduct of two linear spaces V_1 and V_2 over $\mathbb{F}$ is the direct sum $V_1 \oplus V_2 = \{(v_1, v_2) \mid v_1 \in V_1, v_2 \in V_2\}$. ■

4.3.3 Biproducts

In $\mathbf{Vec}_{\mathbb{F}}$, the product and coproduct are the same, i.e., $V_1 \times V_2 \cong V_1 \oplus V_2$. Then we will say the *biproduct* of V_1 and V_2 and denote it by $V_1 \oplus V_2$. The following diagram commutes:



Definition 4.6 — Biproducts. The *biproduct* of two objects X and Y in a category $\mathcal{C}$ is an object $X \oplus Y$ that is both the product and coproduct of X and Y.

Remark. The biproduct exists if and only if the product and coproduct exist and are isomorphic, or if the initial object and the terminal object exist and are isomorphic.

■ **Example 4.15** In $\mathbf{Vec}_{\mathbb{F}}$, the zero vector space is both a terminal object and an initial object, so the biproduct exists. ■

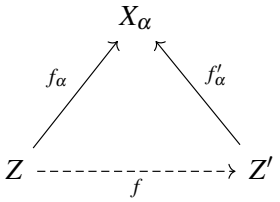
However, in **Set**, the empty set is an initial object but the terminal object is any singleton set, so the biproduct does not exist.

4.3.4 Products and Coproducts of a Family of Objects

In general, we may have the product or coproduct of a family of objects.

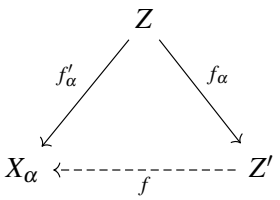
Let $\mathcal{C}$ be a category and $\{X_\alpha\}_{\alpha\in I}$ be a collection of objects in $\mathcal{C}$ indexed by a set I, called the *indexing set*. The *product* of $\{X_\alpha\}_{\alpha\in I}$ is the terminal object in the new category:

- Objects: all collections of morphisms $\{f_\alpha : Z \to X_\alpha\}_{\alpha\in I}$ in $\mathcal{C}$ for all $Z \in \mathsf{Ob}(\mathcal{C})$;
- Morphisms: for all $\alpha \in I$, commutative diagrams in $\mathcal{C}$:

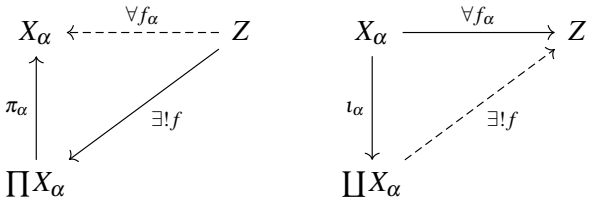


The *coproduct* of $\{X_\alpha\}_{\alpha\in I}$ is the initial object in the new category:

- Objects: all collections of morphisms $\{f_\alpha : X_\alpha \to Z\}_{\alpha\in I}$ in $\mathcal{C}$ for all $Z \in \mathsf{Ob}(\mathcal{C})$;
- Morphisms: for all $\alpha \in I$, commutative diagrams in $\mathcal{C}$:



Then the product and coproduct have the following universal properties respectively:



The elements in the product of a family of objects in $\mathbf{Vec}_{\mathbb{F}}$ can be written as ordered tuples: $(v_\alpha)_{\alpha\in I}$. The product can be defined as follows:

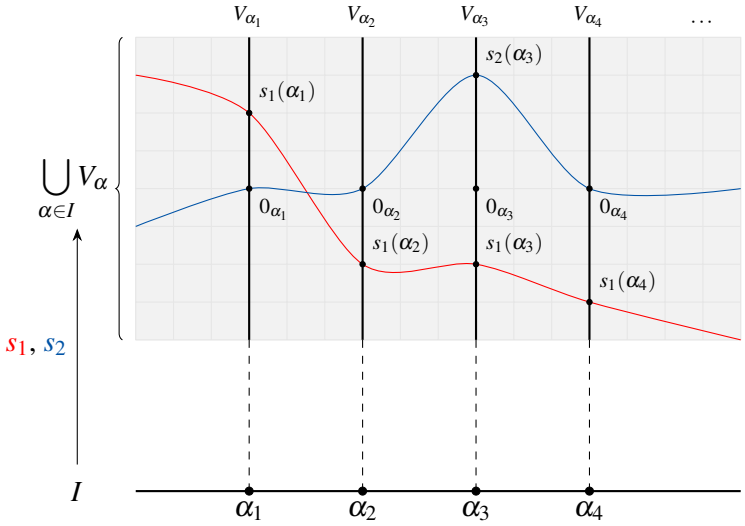
$$\prod_{\alpha\in I} V_\alpha = \{(v_\alpha)_{\alpha\in I} \mid v_\alpha \in V_\alpha\}$$

Then the coproduct can be defined as follows:

$$\bigoplus_{\alpha\in I} V_\alpha = \{(v_\alpha) \in \prod_{\alpha\in I} V_\alpha \mid v_\alpha \text{ is finitely supported}\} \subseteq \prod_{\alpha\in I} V_\alpha$$

Remark. In general, the product is not equal to the coproduct. They are equal if and only if the indexing set I is finite.

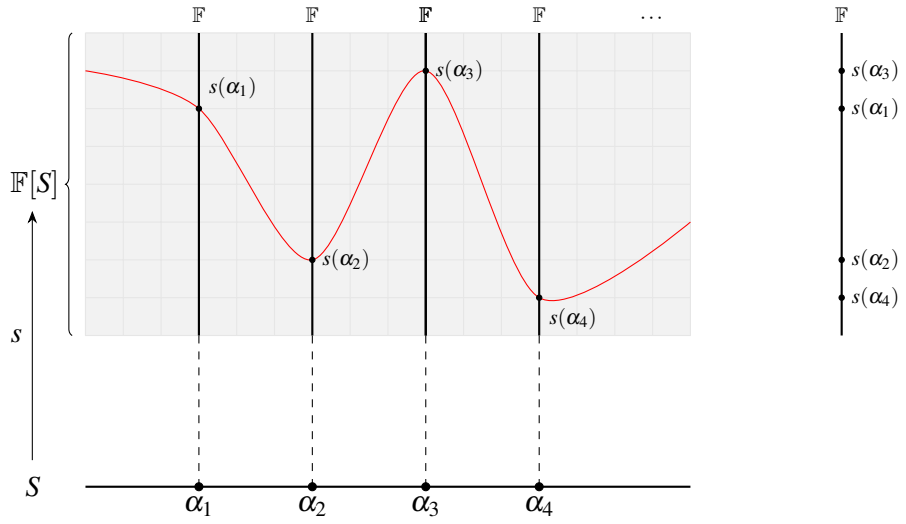
Consider the following diagram:



Remark. The right sections s_1 and s_2 are two elements in the product $\prod V_\alpha$. Note that s_2 is likely to be "finitely supported" since it is zero in almost all components shown in the diagram. However, if I is an infinite set, then s_2 may not be finitely supported since there may be infinitely many non-zero components not shown in the diagram. So s_2 may not be an element in the coproduct $\bigoplus V_\alpha$ if I is an infinite set, but most likely to be.

So the product $\prod V_\alpha$ contains all possible sections $s : I \to \bigcup V_\alpha$, so it is called the *space of sections*. The coproduct $\bigoplus V_\alpha$ contains all finitely supported sections, so it is called the *space of sections with finite support*. The elements in the coproduct $\bigoplus V_\alpha$ written as ordered tuples $(v_\alpha)_{\alpha\in I}$ can also be written as finite sums $\sum_{\alpha\in I} v_\alpha$ since only finitely many v_α are non-zero.

Actually, the product and coproduct are the generalisation of the polynomial ring and the formal power series ring respectively. We can consider the following diagrams:



The left shows the diagram in generalised version, but it can be squeezed to the right since all fibres are the same. So we can consider the set map as $s : S \to \mathbb{F}$ as shown on the right.

4.4 Functors

Definition 4.7 — Functors. Let $\mathcal{C}$ and $\mathcal{D}$ be two categories. A *functor* $F : \mathcal{C} \to \mathcal{D}$ consists of the following data:

- A map $F : \mathsf{Ob}(\mathcal{C}) \to \mathsf{Ob}(\mathcal{D})$;
- A map $F : \mathsf{Mor}_{\mathcal{C}}(X,Y) \to \mathsf{Mor}_{\mathcal{D}}(F(X),F(Y))$ for all $X,Y \in \mathsf{Ob}(\mathcal{C})$;

such that the following conditions are satisfied:

(a) For all $X \in \mathsf{Ob}(\mathcal{C})$, we have $F(1_X) = 1_{F(X)}$;
(b) For all appropriate morphisms f,g in $\mathcal{C}$, we have $F(fg) = F(f)F(g)$.

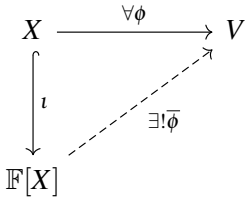
■ **Example 4.16** There are two functors from **Set** to $\mathbf{Vec}_{\mathbb{F}}$:

$$\mathbf{Set} \underset{|-|}{\overset{\mathbb{F}[-]}{\rightleftarrows}} \mathbf{Vec}_{\mathbb{F}}$$

where $\mathbb{F}[-]$ sends set X to the free vector space $\mathbb{F}[X]$ generated by X, and a set map $f : X \to Y$ to the linear map $\mathbb{F}[f] : \mathbb{F}[X] \to \mathbb{F}[Y]$ induced by f. The functor $|-|$ sends a vector space V to its underlying set $|V|$, and a linear map $\phi : V \to W$ to the set map $|\phi| : |V| \to |W|$ induced by ϕ.

The functor $\mathbb{F}[-]$ is called the *free functor*, specifically the *free vector space functor*. The functor $|-|$ is called the *underlying functor* or *forgetful functor*. ■

For some set X and any vector space V, we can consider the following diagram:



This is called the *universal property of free vector space over a set*. Here $\iota : X \to \mathbb{F}[X]$ is the inclusion map, $\phi : X \to V$ is any set map, and $\overline{\phi} : \mathbb{F}[X] \to V$ is the unique linear map induced by ϕ.

Remark. The universal property of free vector space over a set can be rephrased as follows: for any set X and any vector space V, there is a natural identification:

$$\mathbf{Set}(X,|V|) \equiv \mathbf{Vec}_{\mathbb{F}}(\mathbb{F}[X],V)$$

where $\mathbf{Set}(X,|V|)$ is the set of all set maps from X to the underlying set of V, and $\mathbf{Vec}_{\mathbb{F}}(\mathbb{F}[X],V)$ is the set of all linear maps from the free vector space $\mathbb{F}[X]$ to V.

If we consider $\phi : X \to |V|$ as an element in $\mathbf{Set}(X,|V|)$, then the corresponding element in $\mathbf{Vec}_{\mathbb{F}}(\mathbb{F}[X],V)$ is the unique linear map $\overline{\phi} : \mathbb{F}[X] \to V$ induced by ϕ.

Note that $\iota \equiv 1_{\mathbb{F}[X]}$ is the identity element in $\mathbf{Vec}_{\mathbb{F}}(\mathbb{F}[X],\mathbb{F}[X])$, so it corresponds to an element in $\mathbf{Set}(X,|\mathbb{F}[X]|)$, which is exactly the inclusion map $\iota : X \to |\mathbb{F}[X]|$.

Definition 4.8 — Adjoint Functors. Let $\mathcal{C}$ and $\mathcal{D}$ be two categories. A functor $F : \mathcal{C} \to \mathcal{D}$ is called a *left adjoint* of a functor $G : \mathcal{D} \to \mathcal{C}$, and G is called a *right adjoint* of F, if there is a natural identification:

$$\mathcal{D}(F(X),Y) \equiv \mathcal{C}(X,G(Y))$$

for all $X \in \mathsf{Ob}(\mathcal{C})$ and $Y \in \mathsf{Ob}(\mathcal{D})$.

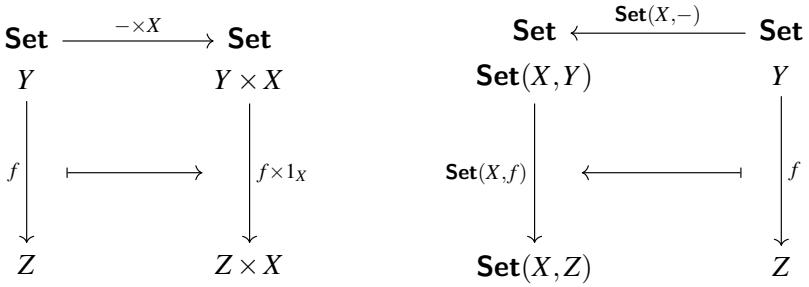
■ **Example 4.17** The free functor $\mathbb{F}[-] : \mathbf{Set} \to \mathbf{Vec}_{\mathbb{F}}$ is a left adjoint of the underlying functor $|-| : \mathbf{Vec}_{\mathbb{F}} \to \mathbf{Set}$. This is exactly the universal property of free vector space over a set. ■

Definition 4.9 — Endofunctors. An *endofunctor* is a functor $F : \mathcal{C} \to \mathcal{C}$ that maps a category to itself.

■ **Example 4.18** Let X be a set. Then we have an adjoint pair of functors:

$$\mathbf{Set} \underset{\mathbf{Set}(X,-)}{\overset{-\times X}{\rightleftarrows}} \mathbf{Set}$$

On the left is the endofunctor $- \times X$ and on the right is the endofunctor $\mathbf{Set}(X,-)$.



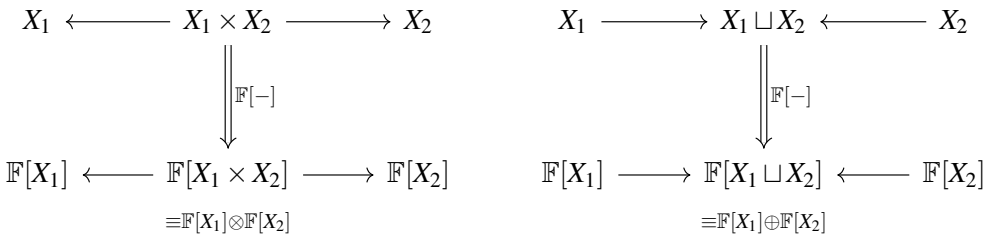
Consider an element $g \in \mathbf{Set}(X,Y)$, which is a set map $g : X \to Y$. Then the corresponding element in $\mathbf{Set}(X,Z)$ is $\mathbf{Set}(X,f)(g) = fg : X \to Z$.

Then we can write the natural identification as follows:

$$\mathbf{Set}(Y \times X, Z) \equiv \mathbf{Set}(Y, \mathbf{Set}(X,Z))$$

for all sets Y and Z. This means that a set map $F : Y \times X \to Z$ corresponds to a set map $F_\natural : Y \to \mathbf{Set}(X,Z)$ such that a $y \in Y$ is mapped to a set map $F_\natural(y) : X \to Z$ defined by $F_\natural(y)(x) = F(y,x)$ for all $x \in X$. ■

Consider the following two diagrams:



The left diagram shows that the free functor sends the product of two sets to the tensor product of two vector spaces. The right diagram shows that the free functor sends the coproduct of two sets to the direct sum of two vector spaces, i.e., the coproduct of two vector spaces. Note that the tensor product of two vector spaces is *not* the product of two vector spaces, as the dimension of the tensor product is $\dim(V_1 \otimes V_2) = \dim(V_1) \cdot \dim(V_2)$ while the dimension of the product is $\dim(V_1 \oplus V_2) = \dim(V_1) + \dim(V_2)$. There is a unique but not isomorphic linear map $\phi : V_1 \otimes V_2 \to V_1 \oplus V_2$.

Remark. The left adjoint functor preserves coproducts, and the right adjoint functor preserves products. This is the consequences of the *adjoint functor theorem*.

Similarly, we have the following natural identifications:

$$\mathbf{Vec}_\mathbb{F}(X \otimes Y, Z) \equiv \mathbf{Vec}_\mathbb{F}(Y, \mathbf{Vec}_\mathbb{F}(X,Z))$$

Note that $\mathbf{Vec}_\mathbb{F}(X,Z)$ is a vector space over $\mathbb{F}$, as $\mathbf{Vec}_\mathbb{F} \equiv \mathsf{Hom}_\mathbb{F}$. Then, we have the following adjoint pair of endofunctors on $\mathbf{Vec}_\mathbb{F}$:

$$\mathbf{Vec}_\mathbb{F} \underset{\mathsf{Hom}_\mathbb{F}(X,-)}{\overset{-\otimes X}{\rightleftarrows}} \mathbf{Vec}_\mathbb{F}$$

4.5 Dual Spaces and Dual Bases

Let V be a finite-dimensional linear space over $\mathbb{F}$. The *dual space* of V is the vector space $V^* = \mathsf{Hom}_{\mathbb{F}}(V,\mathbb{F})$, the set of all linear functionals from V to $\mathbb{F}$, or *covectors*.

Proposition 4.1 Let V be a finite-dimensional linear space over $\mathbb{F}$. Then $\dim(V^*) = \dim(V)$. So, V^* is isomorphic to V but not naturally isomorphic to V.

Proof. Without the loss of generality, we may assume $\dim V = n$ and $V = \mathbb{F}^n$. Then $V^* = \mathsf{Hom}_{\mathbb{F}}(\mathbb{F}^n,\mathbb{F}) \equiv \mathsf{M}_{1\times n}(\mathbb{F})$, the linear space of row matrices with n entries. The linear space is the span of n standard basis row matrices: $\hat{e}^1,\hat{e}^2,\cdots,\hat{e}^n$. So $\dim(V^*) = n = \dim(V)$. We can say $V^* \cong V$. ∎

We have a map $\phi_s : \mathbb{F}^n \to (\mathbb{F}^n)^* \supset S = \{\hat{e}^1,\hat{e}^2,\cdots,\hat{e}^n\}$ defined by $\phi_s(\vec{x}) = \sum_{i=1}^n x_i\hat{e}^i$. This is a vector space isomorphism but not a natural isomorphism, as it depends on the choice of S.

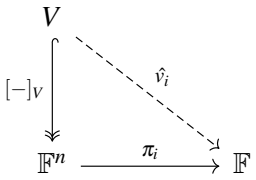
Definition 4.10 — Bases. A *basis* of a linear space V over $\mathbb{F}$ is the minimal spanning set of V with an order. The set of all bases of V is denoted by $\mathcal{B}_V$.

Proposition 4.2 $\mathcal{B}_V$ and $\mathcal{B}_{V^*}$ are naturally isomorphic in **Set**, i.e., the following natural identification holds:

$$\begin{aligned} \mathcal{B}_V &\equiv \mathcal{B}_{V^*} \\ v = (\vec{v}_1,\vec{v}_2,\cdots,\vec{v}_n) &\equiv (\hat{v}^1,\hat{v}^2,\cdots,\hat{v}^n) = v^* \end{aligned}$$

where $\hat{v}^i \in V^*$ is defined by $\hat{v}^i(\vec{v}_j) = \delta^i_j$ for all $1 \le i,j \le n$.

Proof. Consider the following commutative diagram:



The projection map π_i is a linear functional in $\mathbb{F}^n$ that sends $\vec{x} = (x_1,x_2,\cdots,x_n)$ to x_i. It is actually $\hat{e}^i$. Note that $[-]_V : V \to \mathbb{F}^n$ is a coordinate map defined by a basis $v = (\vec{v}_1,\vec{v}_2,\cdots,\vec{v}_n) \in \mathcal{B}_V$ such that $[\vec{v}_j]_V = \vec{e}_j$ for all $1 \le j \le n$. It is a unique linear map which identify $\vec{v}_i$ with $\vec{e}_i$. It can be done by trivialisation of V with respect to the basis v. Then we define $\hat{v}^i(\vec{v}_j) = \delta^i_j$ for all $1 \le i,j \le n$.

Then we have to consider whether $(\hat{v}^1,\hat{v}^2,\cdots,\hat{v}^n)$ is a basis of V^*. As $\dim V^* = n$, we only need to show that $(\hat{v}^1,\hat{v}^2,\cdots,\hat{v}^n)$ is a spanning set of V^* or linearly independent. We have to check whether the equation $\sum_{i=1}^n x_i\hat{v}^i = 0$ for some $x_i \in \mathbb{F}$ has only the trivial solution. Applying it to $\vec{v}_j$ for all $1 \le j \le n$, we have $0 = \sum_{i=1}^n x_i\hat{v}^i(\vec{v}_j) = \sum_{i=1}^n x_i\delta^i_j = x_j$. So $x_j = 0$ for all $1 \le j \le n$. This means that $(\hat{v}^1,\hat{v}^2,\cdots,\hat{v}^n)$ is linearly independent, and hence it is a basis of V^*. We call it the *dual basis* of the basis $\vec{v} = (v_1,v_2,\cdots,v_n)$ and denote it by $\vec{v}^* = (\hat{v}^1,\hat{v}^2,\cdots,\hat{v}^n)$.

Then we have to show that there is a unique basis in V^* that satisfies $\hat{v}^i(\vec{v}_j) = \delta^i_j$. Let $V = \mathbb{F}^n$ and $v = (\vec{v}_1,\vec{v}_2,\cdots,\vec{v}_n)$ be a basis of V. Then $A = [\vec{v}_1 \quad \vec{v}_2 \quad \cdots \quad \vec{v}_n]$ is an invertible matrix. Let $(\alpha^1,\alpha^2,\cdots,\alpha^n)$ be a basis of V^*. Then we have the following equations:

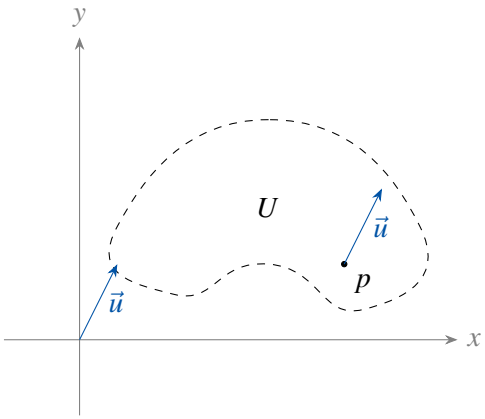
$$[\delta^i_j] = \begin{bmatrix} — & \alpha^1 & — \\ & \vdots & \\ — & \alpha^n & — \end{bmatrix} \begin{bmatrix} | & & | \\ \vec{v}_1 & \cdots & \vec{v}_n \\ | & & | \end{bmatrix} = I_n$$

Then $(\alpha_1,\alpha_2,\cdots,\alpha_n) = A^{-1}$. So the dual basis is unique.

Finally, we have the natural identification: ∎

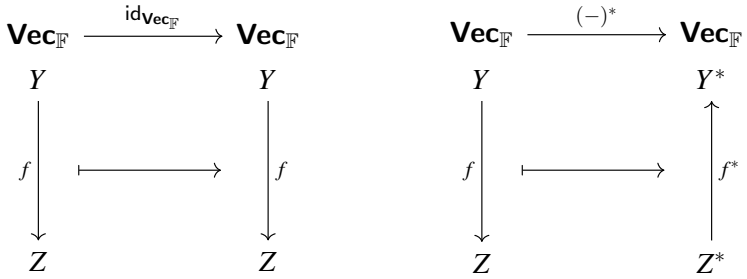
Remark. $V \cong V^*$ but $\mathcal{B}_V \equiv \mathcal{B}_{V^*}$. The isomorphism $V \cong V^*$ depends on the choice of a basis in $\mathcal{B}_V$, while the natural isomorphism $\mathcal{B}_V \equiv \mathcal{B}_{V^*}$ does not depend on any choice.

■ **Example 4.19** Consider the following open subset U of $\mathbb{R}^2$:



Consider the cotangent vector df_p at point p for some smooth function $f : U \to \mathbb{R}$. It is a linear functional $\mathrm{d}f_p : \mathrm{T}_pU \to \mathbb{R}$ defined by $\mathrm{d}f_p(\vec{u}) = \nabla f(p) \cdot \vec{u}$ for all $\vec{u} \in \mathrm{T}_pU$. Here T_pU is the tangent space of U at point p, which is a vector space over $\mathbb{R}$. Note that both $\vec{u}$ and $\nabla f(p)$ are depending on the choice of a coordinate system. However, $\mathrm{d}f_p$ is independent of any choice of coordinate system. In normal calculus, $\mathrm{d}f_p$ is called the *first partial derivative* of f at point p, and normally we write it as $\frac{\partial f}{\partial x}(p)$ and $\frac{\partial f}{\partial y}(p)$. ■

The dual functor is not naturally isomorphic to the identity functor on $\mathbf{Vec}_\mathbb{F}$, as $(-)^*$ is a contravariant functor, while the identity is a contravariant functor, so there is no natural transformation from $\mathrm{id}_{\mathbf{Vec}_\mathbb{F}}$ to $(-)^*$.



4.6 Double Dual Spaces and Doubles

Consider the endofunctors on $\mathbf{Vec}_{\mathbb{F}}$:

$$\mathbf{Vec}_{\mathbb{F}} \underset{\mathsf{id}_{\mathbf{Vec}_{\mathbb{F}}}}{\overset{(-)^{**}}{\rightrightarrows}} \mathbf{Vec}_{\mathbb{F}}$$

There is a natural transformation from $\mathsf{id}_{\mathbf{Vec}_{\mathbb{F}}}$ to $(-)^{**}$ defined by the natural identification: $V \equiv V^{**}$. As $\mathsf{id}_{\mathbf{Vec}_{\mathbb{F}}}$ and $(-)^{**}$ are covariant functors, there is a natural transformation between them.

$$\begin{array}{ccc} \mathbf{Vec}_{\mathbb{F}} & \xrightarrow{\mathsf{id}_{\mathbf{Vec}_{\mathbb{F}}}} & \mathbf{Vec}_{\mathbb{F}} \\ Y & & Y \\ \downarrow f & \longmapsto & \downarrow f \\ Z & & Z \end{array} \qquad \begin{array}{ccc} \mathbf{Vec}_{\mathbb{F}} & \xrightarrow{(-)^{**}} & \mathbf{Vec}_{\mathbb{F}} \\ Y & & Y^{**} \\ \downarrow f & \longmapsto & \downarrow f^{**} \\ Z & & Z^{**} \end{array}$$

Let $\langle -,- \rangle : V^* \times V \to \mathbb{F}$ be the natural pairing defined by $\langle \alpha, u \rangle = \alpha(u)$ where $\alpha : V \to \mathbb{F}$ that sends $u \to \alpha u$. It is the pairing of a covector with a vector and the map is bilinear.

Definition 4.11 — Bilinear Maps. A map $B : U \times V \to W$ is called *bilinear* if for all $u \in U$, the map $B(u,-) : V \to W$ is linear, and for all $v \in V$, the map $B(-,v) : U \to W$ is linear.

We have the following natural identification:

$$\begin{array}{ccccc} V^* \times V & \xrightarrow{\langle -,- \rangle} & \mathbb{F} & \equiv & V^* \xrightarrow{1_{V^*}} \mathsf{Hom}_{\mathbb{F}}(V, \mathbb{F}) \\ & & \| & & \| \\ V \times V^* & \longrightarrow & \mathbb{F} & \equiv & V \xrightarrow{\iota_V} \mathsf{Hom}_{\mathbb{F}}(V^*, \mathbb{F}) \\ \downarrow & \nearrow_{\langle -,- \rangle} & & & \\ V^* \times V & & & & \end{array}$$

where $\iota_V : V \to V^{**}$ is defined by $\iota_V(u) = \check{u}$ such that $\check{u}(\alpha) = \alpha(u)$. Then $V^{**} = \mathsf{Hom}_{\mathbb{F}}(V^*, \mathbb{F}) \equiv V$.

Definition 4.12 — Doubles. Let V be a linear space over $\mathbb{F}$. The *double* of V, denoted by $D(V)$, is defined as follows:

$$D(V) = V \oplus V^*$$

As V is naturally isomorphic to V^{**}, we have the following natural identification:

$$D(V) = V \oplus V^* \equiv V^* \oplus V^{**} = D(V^*)$$

The matrix representation of the isomorphism between $D(V)$ and $D(V^*)$ is

$$\begin{bmatrix} 0 & -\iota_V \\ 1 & 0 \end{bmatrix}$$

where $\iota_V : V \to V^{**}$ is the natural isomorphism defined above. The negative sign is used to make the isomorphism a symplectic isomorphism, which will be discussed in the later chapters.

4.7 Natural Transformation and Natural Equivalences

Definition 4.13 — Natural Transformations. Let $F, G : \mathcal{C} \to \mathcal{D}$ be two functors. A *natural transformation* $\eta : F \to G$ is a collection of morphisms $\eta_X : F(X) \to G(X)$ in $\mathcal{D}$ for all objects X in $\mathcal{C}$, such that for all morphisms $f : X \to Y$ in $\mathcal{C}$, the following diagram commutes:

$$\begin{array}{ccc} F(X) & \xrightarrow{F(f)} & F(Y) \\ \downarrow{\scriptstyle \eta_X} & & \downarrow{\scriptstyle \eta_Y} \\ G(X) & \xrightarrow{G(f)} & G(Y) \end{array}$$

Definition 4.14 — Natural Equivalences. A *natural equivalence* from functor F to functor G is a natural transformation $\eta : F \to G$ which has a two-sided inverse natural transformation $\eta^{-1} : G \to F$ such that $\eta\eta^{-1} = 1_G$ and $\eta^{-1}\eta = 1_F$. In this case, we say that F and G are *naturally equivalent*, denoted by $F \equiv G$.

■ **Example 4.20** Consider the endofunctors on $\mathbf{Vec}_{\mathbb{F}}$:

$$\mathbf{Vec}_{\mathbb{F}} \underset{\mathsf{id}_{\mathbf{Vec}_{\mathbb{F}}}}{\overset{(-)^{**}}{\rightrightarrows}} \mathbf{Vec}_{\mathbb{F}}$$

We have the following natural transformation:

$$\begin{array}{ccccccc} (-)^{**} & V_1 & \xrightarrow{f} & V_2 & \longmapsto & V_1^{**} & \xrightarrow{f^{**}} & V_2^{**} \\ \| & & & & & \downarrow{\scriptstyle \eta_{V_1}} \cong & & \cong \downarrow{\scriptstyle \eta_{V_2}} \\ \mathsf{id}_{\mathbf{Vec}_{\mathbb{F}}} & V_1 & \xrightarrow{f} & V_2 & \longmapsto & V_1 & \xrightarrow{f} & V_2 \end{array}$$

Then we have the natural equivalence: $(-)^{**} \equiv \mathsf{id}_{\mathbf{Vec}_{\mathbb{F}}}$. ■

■ **Example 4.21** We have the following natural equivalence:

$$\mathsf{Map}^{\mathsf{BL}}(U \times V, -) \equiv \mathsf{Hom}_{\mathbb{F}}(U, \mathsf{Hom}_{\mathbb{F}}(V, -))$$

where both are endofunctors on $\mathbf{Vec}_{\mathbb{F}}$. For any linear space Z over $\mathbb{F}$, we have the natural isomorphism:

$$\natural_Z : \mathsf{Map}^{\mathsf{BL}}(U \times V, Z) \to \mathsf{Hom}_{\mathbb{F}}(U, \mathsf{Hom}_{\mathbb{F}}(V, Z))$$

■

■ **Example 4.22** We have the following natural equivalence:

$$\mathbb{F} \otimes - \equiv \mathsf{id}_{\mathbf{Vec}_{\mathbb{F}}} \equiv - \otimes \mathbb{F} \equiv \mathsf{Hom}_{\mathbb{F}}(\mathbb{F}, -) \equiv (-)^{**}$$

■

4.8 Exact Functors

Definition 4.15 — Covariant Exact Functors. Let $\mathcal{C}$ and $\mathcal{D}$ be two abelian categories. A covariant functor $F : \mathcal{C} \to \mathcal{D}$ is called:

- *left exact* if whenever $0 \to A \to B \to C \to 0$ is exact then $0 \to F(A) \to F(B) \to F(C)$ is exact in $\mathcal{D}$, i.e., it preserves all finite limits;
- *right exact* if whenever $0 \to A \to B \to C \to 0$ is exact then $F(A) \to F(B) \to F(C) \to 0$ is exact in $\mathcal{D}$, i.e., it preserves all finite colimits;
- *exact* if it is both left exact and right exact.

Definition 4.16 — Contravariant Exact Functors. Let $\mathcal{C}$ and $\mathcal{D}$ be two abelian categories. A contravariant functor $G : \mathcal{C} \to \mathcal{D}$, it is called:

- *contravariant left exact* if whenever $0 \to A \to B \to C \to 0$ is exact then $0 \to G(C) \to G(B) \to G(A)$ is exact in $\mathcal{D}$;
- *contravariant right exact* if whenever $0 \to A \to B \to C \to 0$ is exact then $G(C) \to G(B) \to G(A) \to 0$ is exact in $\mathcal{D}$;
- *contravariant exact* if it is both contravariant left exact and contravariant right exact.

■ **Example 4.23** The dual functor $(-)^* : \mathbf{Vec}_{\mathbb{F}} \to \mathbf{Vec}_{\mathbb{F}}$ is a contravariant left exact functor, as it sends a short exact sequence $0 \to U \to V \to W \to 0$ to a left exact sequence $0 \to W^* \to V^* \to U^*$. Moreover, $U \to V \to W$ is exact if and only if $W^* \to V^* \to U^*$ is exact. Also, the map $U \to V$ is injective if and only if the map $V^* \to U^*$ is surjective; the map $U \to V$ is surjective if and only if the map $V^* \to U^*$ is injective. This can be shown by considering the following two exact sequences: $0 \to U \to V$ and $U \to V \to 0$. ■

In general, the hom-set functor $\mathsf{Hom}_{\mathcal{C}}(X, -) : \mathcal{C} \to \mathbf{Set}$ is a covariant left exact functor for any object X in an abelian category $\mathcal{C}$, and the hom-set functor $\mathsf{Hom}_{\mathcal{C}}(-, X) : \mathcal{C} \to \mathbf{Set}$ is a contravariant left exact functor for any object X in an abelian category $\mathcal{C}$.

■ **Example 4.24** The tensor product functor $- \otimes V$ is a covariant right exact functor, as it sends a short exact sequence $0 \to U \to V \to W \to 0$ to a right exact sequence $U \otimes V \to V \otimes V \to W \otimes V \to 0$. ■

Note that the tensor product functor is a left adjoint functor, and left adjoint functors are right exact in general, while the $\mathbf{Vec}_{\mathbb{F}}(V, -)$ functor is a right adjoint functor, and right adjoint functors are left exact in general.

5. Tensor Algebra

In high level universities, students will blame themselves if they don't understand the content, but in low level universities, students will blame the professors.

GUOWU MENG

5.1 Tensor Products

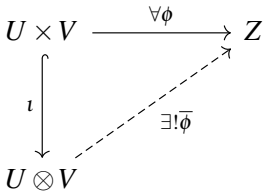
Let U and V be two fixed linear spaces over $\mathbb{F}$ and Z be any linear space over $\mathbb{F}$. Consider the set of all bilinear maps from $U \times V$ to Z, denoted by $\mathsf{Map}^{\mathsf{BL}}(U \times V, Z)$. It is a vector space over $\mathbb{F}$ as it is a subset of $\mathsf{Map}(U \times V, Z)$, the set of all maps from $U \times V$ to Z.

By the universal property of tensor product, we have a natural identification:

$$\mathsf{Map}^{\mathsf{BL}}(U \times V, Z) \equiv \mathsf{Hom}_{\mathbb{F}}(U \otimes V, Z)$$

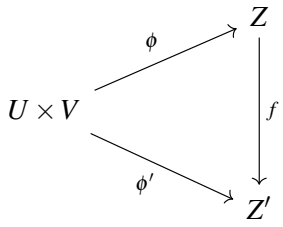
Note that both are naturally identical to $\mathsf{Hom}_{\mathbb{F}}(U, \mathsf{Hom}_{\mathbb{F}}(V, Z))$. Also note that $\mathsf{Hom}(- \otimes V, Z) \equiv \mathsf{Hom}(-, \mathsf{Hom}(V, Z))$ is a *tensor-hom adjunction*.

The natural identification is the universal property of tensor product. Consider the following commutative diagram:



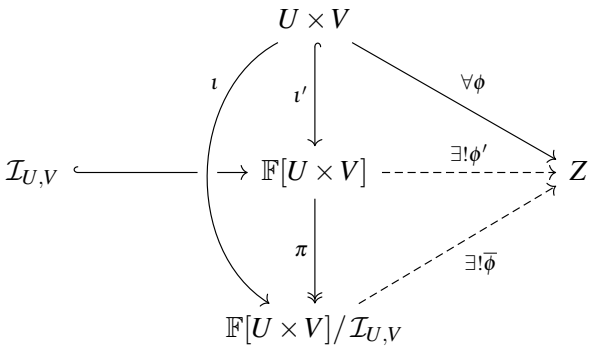
Note that the map ι and ϕ are bilinear maps, and the existence of the unique linear map $\overline{\phi}$ follows from the universal property of the tensor product. We can also consider it as the initial object in a new category:

- Objects: all bilinear maps $\phi : U \times V \to Z$ for all $Z \in \mathsf{Ob}(\mathbf{Vec}_{\mathbb{F}})$;
- Morphisms: commutative diagrams in $\mathbf{Vec}_{\mathbb{F}}$:



The existence of tensor product follows from the existence of free vector space over a set and the existence of quotient spaces.

Consider the following commutative diagram:



where $\mathcal{I}_{U,V}$ is the subspace of $\mathbb{F}[U \times V]$ generated by the following elements for all $u, u_1, u_2 \in U$, $v, v_1, v_2 \in V$ and $\alpha, \beta \in \mathbb{F}$:

- $(\alpha u_1 + \beta u_2, v) - \alpha(u_1, v) - \beta(u_2, v)$;
- $(u, \alpha v_1 + \beta v_2) - \alpha(u, v_1) - \beta(u, v_2)$;

Why the construction of $\mathcal{I}_{U,V}$ is like this? This is because we want ι to be a bilinear map. Then $\iota(\alpha u_1 + \beta u_2, v) = \alpha\iota(u_1, v) + \beta\iota(u_2, v)$ and $\iota(u, \alpha v_1 + \beta v_2) = \alpha\iota(u, v_1) + \beta\iota(u, v_2)$. This means that the elements in $\mathcal{I}_{U,V}$ should be mapped to 0 by ι. So we have to quotient $\mathbb{F}[U \times V]$ by $\mathcal{I}_{U,V}$ to make ι a bilinear map.

We define $U \otimes V = \mathbb{F}[U \times V]/\mathcal{I}_{U,V}$ and this shows the existence of tensor product.

> *Remark.* The inclusion map $\iota : U \times V \to U \otimes V$ is 'surjective' in the sense that the image of ι spans $U \otimes V$, i.e. $\mathsf{Span}(\mathsf{Im}(\iota)) = U \otimes V$. To know $\overline{\phi}$, it suffices to know $\overline{\phi}(u \otimes v) = \phi(u, v)$ for all $u \in U$ and $v \in V$.

We can talk about the tensor product of k linear spaces with $k \geq 2$. Moreover, the tensor product is associative and commutative up to isomorphism, i.e., $V_1 \otimes V_2 \otimes V_3 \cong (V_1 \otimes V_2) \otimes V_3 \cong V_1 \otimes (V_2 \otimes V_3)$ and $V_1 \otimes V_2 \cong V_2 \otimes V_1$. Both of them are natural isomorphisms.

$$\begin{array}{ccc} V_1 \times V_2 \times V_3 & \longrightarrow & V_1 \otimes V_2 \otimes V_3 \\ \downarrow & & \| \\ (V_1 \otimes V_2) \times V_3 & \longrightarrow & (V_1 \otimes V_2) \otimes V_3 \end{array} \qquad \begin{array}{ccc} V_1 \times V_2 & \longrightarrow & V_1 \otimes V_2 \\ \updownarrow & & \| \\ V_2 \times V_1 & \longrightarrow & V_2 \otimes V_1 \end{array}$$

We have a natural equivalence:

$$\mathsf{Hom}(U, V \otimes W) \equiv \mathsf{Hom}(U, V) \otimes W$$

Then we can prove that $\mathsf{Hom}(V_1, V_2) \equiv V_1^* \otimes V_2$ and $(V_1 \otimes V_2)^* \equiv V_1^* \otimes V_2^*$. Also, we have the following equation, by considering $V_1 \otimes V_2 \equiv \mathsf{Hom}(V_1^*, V_2)$:

$$\mathsf{dim}\,(V \otimes W) = \mathsf{dim}\,(V) \cdot \mathsf{dim}\,(W)$$

If e is a minimal spanning set of V_1 and f is a minimal spanning set of V_2, then $e \otimes f$ is a minimal spanning set of $V_1 \otimes V_2$. Moreover, we have $\mathsf{End}(V) \equiv (\mathsf{End}(V))^*$ and the identity map 1_V corresponds to the *trace* map $\mathtt{tr} : \mathsf{End}(V) \to \mathbb{F}$ under this identification.

We also have the distribution of tensor product over direct sum: $V_1 \otimes (V_2 \oplus V_3) \equiv (V_1 \otimes V_2) \oplus (V_1 \otimes V_3)$. Moreover, $\mathsf{Hom}(V_1, V_2 \oplus V_3) \equiv \mathsf{Hom}(V_1, V_2) \oplus \mathsf{Hom}(V_1, V_3)$ and $\mathsf{Hom}(V_1 \oplus V_2, V_3) \equiv \mathsf{Hom}(V_1, V_3) \times \mathsf{Hom}(V_2, V_3)$.

5.2 Algebras

Definition 5.1 — Algebras. An *algebra* over a field $\mathbb{F}$ is a linear space A over $\mathbb{F}$ equipped with a bilinear product map $A \times A \to A$, or equivalently a linear map $A \otimes A \to A$.

■ **Example 5.1** The set of all polynomials in t with coefficients in $\mathbb{F}$, denoted $\mathbb{F}[t]$, is an algebra over $\mathbb{F}$. As $\mathbb{F}[t] \times \mathbb{F}[t] \to \mathbb{F}[t]$ defined by $(f,g) \mapsto fg$ is a bilinear map. Moreover, $\mathbb{F}[t]$ has a multiplicative identity $1 \in \mathbb{F}[t]$, $fg = gf$ for all $f,g \in \mathbb{F}[t]$, and $(fg)h = f(gh)$ for all $f,g,h \in \mathbb{F}[t]$. So $\mathbb{F}[t]$ is a unital commutative associative algebra over $\mathbb{F}$. ■

■ **Example 5.2** The set of all square matrices with order n over $\mathbb{F}$, denoted by $\mathsf{M}_{n\times n}(\mathbb{F})$, is an algebra over $\mathbb{F}$. As $\mathsf{M}_{n\times n}(\mathbb{F}) \times \mathsf{M}_{n\times n}(\mathbb{F}) \to \mathsf{M}_{n\times n}(\mathbb{F})$ defined by $(A,B) \mapsto AB$ is a bilinear map. Moreover, $\mathsf{M}_{n\times n}(\mathbb{F})$ has a multiplicative identity $I_n \in \mathsf{M}_{n\times n}(\mathbb{F})$, $(AB)C = A(BC)$ for all $A,B,C \in \mathsf{M}_{n\times n}(\mathbb{F})$. However, in general $AB \neq BA$ for some $A,B \in \mathsf{M}_{n\times n}(\mathbb{F})$. So $\mathsf{M}_{n\times n}(\mathbb{F})$ is a unital associative algebra but it is a non-commutative algebra over $\mathbb{F}$. ■

■ **Example 5.3** The 3-dimensional Euclidean space $\mathbb{R}^3$ with the cross product $\times : \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}^3$ is an algebra over $\mathbb{R}$. As the cross product is bilinear. However, it does not have a multiplicative identity, not associative and not commutative. So $\mathbb{R}^3$ with the cross product is a non-unital non-associative non-commutative algebra over $\mathbb{R}$. ■

Remark. $(\mathbb{R}^3, \times)$ is an example of a simple real lie algebra. It is the lie algebra of the lie group SO(3), the special orthogonal group in dimension 3, i.e., the 3-dimensional rotations. $(\mathbb{R}^3, \times)$ is denoted by $\mathfrak{so}(3)$. Also, it is the lie algebra of the infinitesimal symmetries of a pointed 3-dimensional Euclidean space.

Definition 5.2 — Lie Algebras. An algebra is $\mathfrak{g}$ over a field $\mathbb{F}$ is called a *lie algebra* if the *lie bracket* or lie product $[-,-] : \mathfrak{g} \times \mathfrak{g} \to \mathfrak{g}$ satisfies the following two conditions:

- Skew-symmetry: $[x,x] = 0$ for all $x \in \mathfrak{g}$, i.e., $[x,y] = -[y,x]$ for all $x,y \in \mathfrak{g}$ if $\mathsf{char}(\mathbb{F}) \neq 2$;
- Jacobi Identity: $[x,[y,z]] + [y,[z,x]] + [z,[x,y]] = 0$ for all $x,y,z \in \mathfrak{g}$.

Definition 5.3 — Graded Linear Space. A linear space $V_\bullet$ over $\mathbb{F}$ is called a $\mathbb{Z}_{\geq 0}$-*graded linear space* or *graded vector space* if it is a direct sum of linear subspaces V_n for all $n \in \mathbb{Z}_{\geq 0}$:

$$V_\bullet = \bigoplus_{n=0}^{\infty} V_n$$

The elements in V_n are called *homogeneous elements* of degree n. If $v \in V_n$ is a homogeneous element, we write $\deg(v) = n$.

Definition 5.4 — Graded Linear Maps. A linear map $\phi : V_\bullet \to W_\bullet$ is called a *graded linear map* with graded degree $k \geq 0$ if $\phi(V_n) \subseteq W_{n+k}$ for all $n \in \mathbb{Z}_{\geq 0}$.

5.3 Tensor Algebras

Let V be a finite-dimensional linear space over $\mathbb{F}$. We define a new notation:

$$V^{\otimes k} = \underbrace{V \otimes V \otimes \cdots \otimes V}_{k \text{ times}}$$

for all $k \geq 0$. Note that $V^{\otimes 0} = \mathbb{F}$. Also, $\dim(V^{\otimes k}) = (\dim V)^k$ for all $k \geq 0$.

We define the *tensor algebra* of V over $\mathbb{F}$, denoted by $\mathcal{T}^\bullet V$, as follows:

$$\mathcal{T}^\bullet V = \bigoplus_{k=0}^{\infty} V^{\otimes k} = \mathbb{F} \oplus V \oplus (V \otimes V) \oplus (V \otimes V \otimes V) \oplus \cdots$$

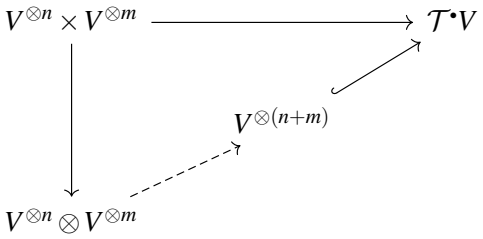
The tensor algebra $\mathcal{T}^\bullet V$ is an algebra over $\mathbb{F}$ with the bilinear product map defined by the tensor product:

$$\otimes : \mathcal{T}^\bullet V \times \mathcal{T}^\bullet V \to \mathcal{T}^\bullet V$$

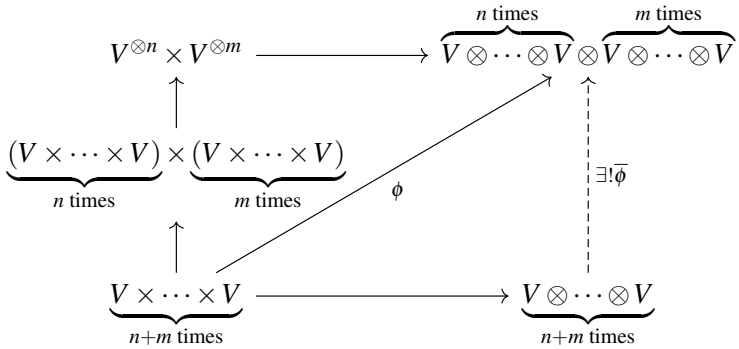
which sends $(\sum_n u_n, \sum_m v_m)$ to $\sum_{n,m}(u_n \otimes v_m)$.

> *Remark.* As the algebra product is bilinear, it suffices to know the product of two homogeneous elements, i.e., $V^{\otimes n} \times V^{\otimes m} \to \mathcal{T}^\bullet V$ for all $n, m \geq 0$. So $\mathcal{T}^\bullet V$ is a $\mathbb{Z}_{\geq 0}$-graded algebra over $\mathbb{F}$. As the tensor algebra is bi-additive, we have the following equality:
>
> $$\sum_n u_n \otimes \sum_m v_m = \sum_n (u_n \otimes \sum_m v_m) = \sum_n \sum_m (u_n \otimes v_m) = \sum_{n,m} (u_n \otimes v_m)$$

Then to define the bilinear product above, we have to define the tensor product of two homogeneous elements:



We have to prove the existence of the bilinear map $V^{\otimes n} \times V^{\otimes m} \to V^{\otimes (n+m)}$ for all $n, m \geq 0$. We can prove it by the following commutative diagram:

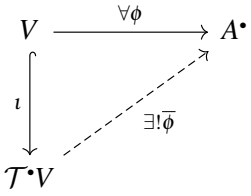


The proof used a lot of universal properties of tensor products. Note that the map ϕ is a multilinear map and $\overline{\phi}$ is a linear equivalence.

So we have proved the existence of the bilinear product map $\otimes : \mathcal{T}^\bullet V \times \mathcal{T}^\bullet V \to \mathcal{T}^\bullet V$. Then $\mathcal{T}^\bullet V$ is an algebra over $\mathbb{F}$.

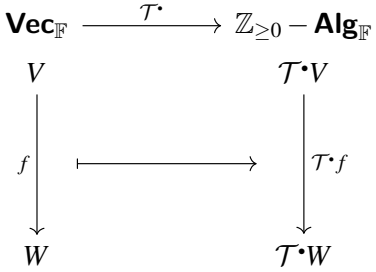
Remark. The tensor algebra $(\mathcal{T}^\bullet V, \otimes)$ is a graded unital associative algebra over $\mathbb{F}$. It is graded, as it is degree additive, i.e., $V^{\otimes n} \times V^{\otimes m} \to V^{\otimes(n+m)}$ for all $n, m \geq 0$. It is unital, as $V^{\otimes 0} - \mathbb{F} \times V^{\otimes m} \to V^{\otimes m}$ and the reverse. The multiplicative identity is $1 \in \mathbb{F}$. It is associative, as $(u \otimes v) \otimes w = u \otimes (v \otimes w)$ for all $u, v, w \in V$ and the associativity can be extended to all homogeneous elements by bi-additivity. However, in general it is not commutative, as $u \otimes v \neq v \otimes u$ for some $u, v \in V$.

There is a universal property of tensor algebras. Consider the following commutative diagram:



Note that $V = V^{\otimes 1} = 0 \oplus V \oplus 0 \oplus \cdots \subseteq \mathcal{T}^\bullet V$ and ι is the inclusion map. Here $A^\bullet$ is any graded unital associative algebra over $\mathbb{F}$ and $\phi : V \to A^\bullet$ is a graded linear map with graded degree 0. Then there exists a unique graded algebra homomorphism with graded degree 0 $\overline{\phi} : \mathcal{T}^\bullet V \to A^\bullet$ such that $\overline{\phi} \circ \iota = \phi$. This shows the universal property of tensor algebras. More specifically, the map ϕ is a map from V to the degree 1 part of $A^\bullet$, i.e., $\phi : V \to A_1$, then with an inclusion map.

The tensor algebra construction is actually a functor from $\mathbf{Vec}_\mathbb{F}$ to the category of graded unital associative algebras over $\mathbb{F}$, denoted by $\mathbb{Z}_{\geq 0} - \mathbf{Alg}_\mathbb{F}$:



where $\mathcal{T}^\bullet f : \mathcal{T}^\bullet V \to \mathcal{T}^\bullet W$ is the unique graded algebra homomorphism with graded degree 0 such that $\mathcal{T}^\bullet f \circ \iota_V = \iota_W \circ f$. Here $\iota_V : V \to \mathcal{T}^\bullet V$ and $\iota_W : W \to \mathcal{T}^\bullet W$ are the inclusion maps.

The existence of the functor $\mathcal{T}^\bullet$ follows from the universal property of tensor algebras. It is called the *free graded algebra functor*, normally the "unital" and "associative" will be omitted.

5.4 Quotient Algebras

In this section, we will discuss three quotient algebras associated with a vector space V: the symmetric algebra, exterior algebra, and universal enveloping algebra.

Before that we need to introduce the concept of ideals in algebras.

Definition 5.5 — Ideals of Algebras. An *ideal* of an algebra A over a field $\mathbb{F}$ is a non-empty subset I of A which is closed under linear combinations and algebra multiplications by elements in A. That is, for all $x, y \in I$, $\alpha, \beta \in \mathbb{F}$ and $a \in A$, we have:

- $\alpha x + \beta y \in I$;
- $ax \in I$ and $xa \in I$.

Simply speaking, an ideal is a generalisation of a rule to an algebra.

The following is an example of an ideal in a ring, which is an example of ideal in a more general concept.

■ **Example 5.4** Consider the ring of integers, $\mathbb{Z}$. The set of all n-multiples, denoted by $n\mathbb{Z}$, is an ideal of $\mathbb{Z}$ for all $n \in \mathbb{Z}$. As it is closed under addition and multiplication by any integer. ■

5.4.1 Symmetric Algebras

The *symmetric algebra* of a vector space V over a field $\mathbb{F}$, denoted by $\mathcal{S}^\bullet V$, is defined as the quotient algebra of the tensor algebra $\mathcal{T}^\bullet V$ by the ideal of $\mathcal{T}^\bullet V$ generated by elements of the form $u \otimes v - v \otimes u$ for all $u, v \in V$:

$$\mathcal{S}^\bullet V = \mathcal{T}^\bullet V / \mathcal{I}_{\mathcal{S}^\bullet} = \mathcal{T}^\bullet V / \langle u \otimes v - v \otimes u \mid u, v \in V \rangle$$

The $\mathcal{I}_{\mathcal{S}^\bullet}$ is called the *symmetrising ideal* of $\mathcal{T}^\bullet V$. It is actually the *ideal completion* of the relation $u \otimes v = v \otimes u$ for all $u, v \in V$. We use $\langle - \rangle$ to denote the ideal generated by a set.

Then the elements in $\mathcal{S}^\bullet V$ are equivalence classes of elements in $\mathcal{T}^\bullet V$. We have $uv \in \mathcal{S}^\bullet V$ as the equivalence class of $u \otimes v \in \mathcal{T}^\bullet V$ denoted by $[u \otimes v]$. Note that $uv = [u \otimes v] = [v \otimes u] = vu$ in $\mathcal{S}^\bullet V$, as $[u \otimes v - v \otimes u] = 0$. So the product in $\mathcal{S}^\bullet V$ is commutative.

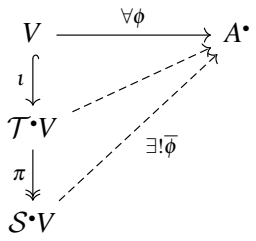
Remark. Symmetric algebra is still a graded algebra. As the ideal $\mathcal{I}_{\mathcal{S}^\bullet}$ is a graded ideal, i.e., $\mathcal{I}_{\mathcal{S}^\bullet} = \bigoplus_{k=0}^{\infty} (\mathcal{I}_{\mathcal{S}^\bullet} \cap V^{\otimes k})$.

Similar to tensor algebras, we have the following expression:

$$\mathcal{S}^\bullet V = \bigoplus_{k=0}^{\infty} \mathcal{S}^k V$$

where $\mathcal{S}^k V$ is the k-th symmetric power of V.

We also have the following universal property of symmetric algebras. Consider the following commutative diagram:



Here $A^\bullet$ is any graded unital commutative associative algebra over $\mathbb{F}$.

Similarly, $\mathcal{S}^\bullet V$ is the *free graded commutative algebra functor* from $\mathbf{Vec}_\mathbb{F}$ to the category of graded unital commutative associative algebras over $\mathbb{F}$, denoted by $\mathbb{Z}_{\geq 0} - \mathbf{CAlg}_\mathbb{F}$:

5.4.2 Exterior Algebras

The *exterior algebra* of a vector space V over a field $\mathbb{F}$, denoted by $\bigwedge^\bullet V$, is defined as the quotient algebra of the tensor algebra $\mathcal{T}^\bullet V$ by the ideal of $\mathcal{T}^\bullet V$ generated by elements of the form $v \otimes v$ for all $v \in V$:

$$\bigwedge^\bullet V = \mathcal{T}^\bullet V / \mathcal{I}_{\bigwedge^\bullet} = \mathcal{T}^\bullet V / \langle v \otimes v \mid v \in V \rangle = \mathcal{T}^\bullet V / \langle v \otimes w + w \otimes v \mid v, w \in V \rangle$$

The $\mathcal{I}_{\bigwedge^\bullet}$ is called the *alternating ideal* of $\mathcal{T}^\bullet V$. It is actually the *ideal completion* of the relation $v \otimes v = 0$ for all $v \in V$, or equivalently $v \otimes w = -w \otimes v$ for all $v, w \in V$. Sometimes the exterior algebra is also called the *skew-symmetric algebra*. Note that the characteristic of the field $\mathbb{F}$ should not be 2, i.e., $\mathsf{char}(\mathbb{F}) \neq 2$, otherwise $v \otimes w = -w \otimes v$ implies that $v \otimes w = w \otimes v$.

Then the product in $\bigwedge^\bullet V$ is called the *exterior product* or *wedge product*, denoted by $\wedge$. We have $u \wedge v = -v \wedge u$ in $\bigwedge^\bullet V$ for all $u, v \in V$. So the product in $\bigwedge^\bullet V$ is skew-commutative.

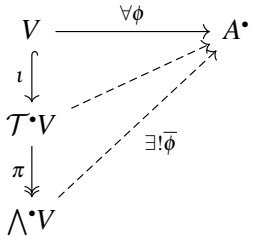
> *Remark.* Exterior algebra is still a graded algebra. As the ideal $\mathcal{I}_{\bigwedge^\bullet}$ is a graded ideal, i.e., $\mathcal{I}_{\bigwedge^\bullet} = \bigoplus_{k=0}^{\infty} (\mathcal{I}_{\bigwedge^\bullet} \cap V^{\otimes k})$.

Then we have the following expression:

$$\bigwedge^\bullet V = \bigoplus_{k=0}^{\infty} \bigwedge^k V$$

where $\bigwedge^k V$ is the k-th exterior power of V.

We also have the following universal property of exterior algebras. Consider the following commutative diagram:



Here $A^\bullet$ is any graded unital associative skew-commutative algebra over $\mathbb{F}$.

Similarly, $\bigwedge^\bullet V$ is the *free graded skew-commutative algebra functor* from $\mathbf{Vec}_\mathbb{F}$ to the category of graded unital associative skew-commutative algebras over $\mathbb{F}$, denoted by $\mathbb{Z}_{\geq 0} - \mathbf{SAlg}_\mathbb{F}$:

5.4.3 Universal Enveloping Algebras

Let $\mathfrak{g}$ be a lie algebra over a field $\mathbb{F}$. The *universal enveloping algebra* of $\mathfrak{g}$ over $\mathbb{F}$, denoted by $\mathcal{U}\mathfrak{g}$, is defined as the quotient algebra of the tensor algebra $\mathcal{T}^\bullet \mathfrak{g}$ by the ideal of $\mathcal{T}^\bullet \mathfrak{g}$ generated by elements of the form $x \otimes y - y \otimes x - [x, y]$ for all $x, y \in \mathfrak{g}$:

$$\mathcal{U}\mathfrak{g} = \mathcal{T}^\bullet \mathfrak{g} / \mathcal{I}_\mathcal{U} = \mathcal{T}^\bullet \mathfrak{g} / \langle x \otimes y - y \otimes x - [x, y] \mid x, y \in \mathfrak{g} \rangle$$

The $\mathcal{I}_\mathcal{U}$ is called the *lie ideal* of $\mathcal{T}^\bullet \mathfrak{g}$. It is actually the *ideal completion* of the relation $xy - yx = [x, y]$ for all $x, y \in \mathfrak{g}$.

> *Remark.* However, the universal enveloping algebra is not a graded algebra. As the ideal $\mathcal{I}_\mathcal{U}$ is not a graded ideal. The $x \otimes y - y \otimes x$ is in $\mathfrak{g}^{\otimes 2}$ but $[x, y]$ is in $\mathfrak{g}^{\otimes 1}$.

5.5 Hilbert-Poincaré Series

Let $V_\bullet = \bigoplus_{i\geq 0} V_i$ be a $\mathbb{Z}_{\geq 0}$-graded finite-dimensional linear space over a field $\mathbb{F}$. The *Hilbert-Poincaré series* of $V_\bullet$ is defined as the following formal power series:

$$P_{V_\bullet}(t) = \sum_{i=0}^{\infty} \dim\,(V_i)\; t^i$$

■ **Example 5.5** The Hilbert-Poincaré series of the tensor algebra $\mathcal{T}^\bullet V$ is:

$$P_{\mathcal{T}^\bullet V}(t) = \sum_{i=0}^{\infty} \dim\,(V^{\otimes i})\; t^i = \sum_{i=0}^{\infty} (\dim V)^i\; t^i = \frac{1}{1-\dim V\; t}$$

■

■ **Example 5.6** The Hilbert-Poincaré series of the symmetric algebra $\mathcal{S}^\bullet V$ is:

$$P_{\mathcal{S}^\bullet V}(t) = \sum_{i=0}^{\infty} \dim\,(\mathcal{S}^i V)\; t^i = \sum_{i=0}^{\infty} \binom{\dim V + i - 1}{i} t^i = \frac{1}{(1-t)^{\dim V}}$$

■

■ **Example 5.7** The Hilbert-Poincaré series of the exterior algebra $\bigwedge^\bullet V$ is:

$$P_{\bigwedge^\bullet V}(t) = \sum_{i=0}^{\infty} \dim\,(\bigwedge^i V)\; t^i = \sum_{i=0}^{\infty} \binom{\dim V}{i} t^i = (1+t)^{\dim V}$$

■

As the Hilbert-Poincaré series of the exterior algebra $\bigwedge^\bullet V$ is a polynomial of degree $\dim V$, we have $\bigwedge^k V = 0$ for all $k > \dim V$. Especially, if $\dim V = n$, then $\bigwedge^n V$ is 1-dimensional and $\bigwedge^{n+1} V = 0$. This is because any $(n+1)$ vectors in an n-dimensional vector space are linearly dependent, so the exterior product of them is 0. Moreover, $\dim\,(\bigwedge^k V) = \dim\,(\bigwedge^{n-k} V)$ for all $0 \leq k \leq n$.

Any one-dimensional linear space is called a *line*.

6. Determinants

"If you are willing to prove it, you can prove it. There is no trick."

GUOWU MENG

6.1 Determinant Lines

We have known that the top exterior power $\bigwedge^n V$ of an n-dimensional vector space V over a field $\mathbb{F}$ is 1-dimensional. So we can define the following:

Definition 6.1 — Determinant Lines. The *determinant line* of an n-dimensional vector space V over a field $\mathbb{F}$ is defined as the top exterior power of V:

$$\det V = \bigwedge^n V = \bigwedge^{\dim V} V$$

Note that the $\det = \bigwedge^k$ is a functor from the category of vector spaces with n-dimensions $\mathbf{Vec}_{\mathbb{F}}^n$ to the category of vector spaces with 1-dimensional, i.e., the category of lines, $\mathbf{Vec}_{\mathbb{F}}^1$ for all $k \geq 0$:

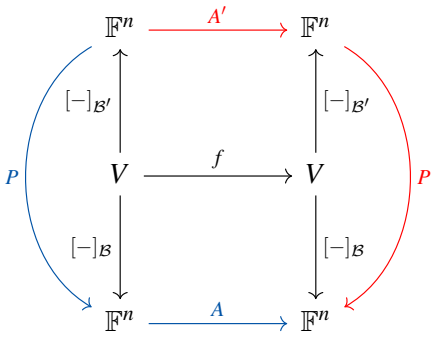
$$\begin{array}{ccc} \mathbf{Vec}_{\mathbb{F}}^n & \xrightarrow{\bigwedge^n = \det} & \mathbf{Vec}_{\mathbb{F}}^1 \\ V_1 & & \bigwedge^n V_1 = \det V_1 \\ \Big\downarrow{\scriptstyle f} & \longmapsto & \Big\downarrow{\scriptstyle \bigwedge^n f = \det f} \\ V_2 & & \bigwedge^n V_2 = \det V_2 \end{array}$$

As $\det$ is a functor, we have the following two properties:

$$\det \mathrm{id}_V = \mathrm{id}_{\det V}, \qquad \det fg = \det f \cdot \det g$$

In particular, if $f \in \mathrm{End}(V)$, then $\det f : \det V \to \det V$ is a multiplication by a scalar in $\mathbb{F}$. So we can identify $\det f$ with a scalar in $\mathbb{F}$. This scalar is called the *determinant* of f and is denoted by $\det f$.

Consider the following commutative diagram:



where V is an n-dimensional vector space over $\mathbb{F}$, $\mathcal{B}$ and $\mathcal{B}'$ are two bases of V, $f \in \mathsf{End}(V)$, A and A' are the matrix representations of f under the bases $\mathcal{B}$ and $\mathcal{B}'$ respectively, and P is the change of basis matrix from $\mathcal{B}$ to $\mathcal{B}'$. Then by focusing the red and blue commutative square, we have:

$$AP = PA', \qquad A = PA'P^{-1}$$

Then we have:

$$\det A \overset{\text{def}}{=\!=} \det f \overset{\text{def}}{=\!=} \det A'$$

In ordinary linear algebra, A and A' are called *similar matrices*, i.e., $A \sim A'$. This means they represent the same endomorphism, so they have the same determinant.

6.2 Permutation Groups

Before we derive the explicit formula of determinants, we need to introduce the concept of permutation groups.

Definition 6.2 — Automorphisms. An *automorphism* is an isomorphism from a mathematical object to itself.

Definition 6.3 — Automorphism Groups. The set of all automorphisms on a mathematical object X forms a group under the composition of functions, denoted by $\mathsf{Aut}(X)$.

■ **Example 6.1** The general linear group $\mathsf{GL}(V)$ of a vector space V over $\mathbb{F}$ is the group of all invertible linear maps from V to V, i.e., $\mathsf{GL}(V) = \mathsf{Aut}(V)$. The group operation is the composition of functions. ■

■ **Example 6.2** The general linear group $\mathsf{GL}_n(\mathbb{F})$ of degree n over $\mathbb{F}$ is the group of all invertible $n \times n$ matrices over $\mathbb{F}$, i.e., $\mathsf{GL}_n(\mathbb{F}) = \mathsf{Aut}(\mathbb{F}^n)$. The group operation is the matrix multiplication. Note that $\mathsf{GL}_n(\mathbb{F}) \cong \mathsf{GL}(\mathbb{F}^n)$. Also note that the group is not abelian if $n \geq 2$. ■

Definition 6.4 — Permutation Groups. A *permutation group* S_n on a set $\underline{n} := \{1, 2, \cdots, n\}$ is the group of all bijections from $\underline{n}$ to itself, i.e., $S_n = \mathsf{Aut}(\underline{n})$. It is called the *symmetric group* on n elements. The group operation is the composition of functions.

Then the order of S_n, denoted by $|S_n|$, is $n!$.

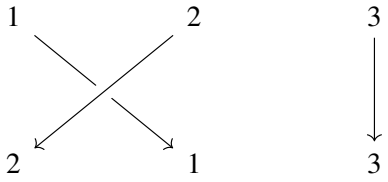
■ **Example 6.3** The permutation group S_2 has two elements: the identity permutation 1 and the transposition σ_1 defined by $\sigma_1(1) = 2$ and $\sigma_1(2) = 1$. ■

Instead of writing $S_2 = \{1, \sigma_1\}$, we can write $S_2 = \langle \sigma_1 \mid \sigma_1^2 = 1 \rangle$, where σ_1 is called the *generator* of S_2 and $\sigma_1^2 = 1$ is called the *relation* of S_2. This is called the *presentation* of S_2.

In general, the generator σ_i of S_n is defined by:

$$\sigma_i(j) = \begin{cases} j+1, & j = i \\ j-1, & j = i+1 \\ j, & \text{otherwise} \end{cases} = (i \quad i+1)$$

■ **Example 6.4** The generator σ_1 of S_3 can be represented by the following diagram:



It can also be written as $\sigma_1 = (1 \quad 2)$ or $(1 \quad 2)(3)$ or $\begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix}$. ■

Moreover, we have a cycle with 3 elements denoted as $(1 \quad 2 \quad 3)$ defined by the $\begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}$.

Then the presentation of S_3 is:

$$S_3 = \langle \sigma_1, \sigma_2 \mid \sigma_1^2 = 1, \sigma_2^2 = 1, \sigma_1\sigma_2\sigma_1 = \sigma_2\sigma_1\sigma_2 \rangle$$

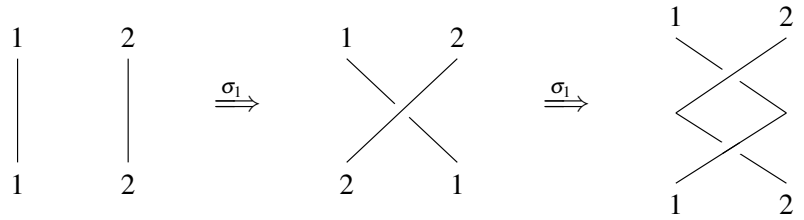
In general, the presentation of S_n is:

$$S_n = \langle \sigma_1, \sigma_2, \cdots, \sigma_{n-1} \mid \sigma_i^2 = 1, \sigma_i\sigma_j = \sigma_j\sigma_i \ (|i-j| > 1), \sigma_i\sigma_{i+1}\sigma_i = \sigma_{i+1}\sigma_i\sigma_{i+1} \rangle$$

The last two relations are called the *braid relations*:

- *Far commutativity:* $\sigma_i\sigma_j = \sigma_j\sigma_i$ for all $|i-j| > 1$;
- *Braid relation:* $\sigma_i\sigma_{i+1}\sigma_i = \sigma_{i+1}\sigma_i\sigma_{i+1}$.

The permutation group S_n is generated by quotienting the braid group B_n by the relations $\sigma_i^2 = 1$ for all $1 \leq i \leq n-1$. We call B_n the *braid group* on n strands. A simple way to visualise the braid group is to think about braiding n strands of hair. The braid group B_n has the same presentation as S_n except that there is no relation $\sigma_i^2 = 1$ for all $1 \leq i \leq n-1$. Consider the following diagrams:



Consider the following exact sequence:

$$1 \longrightarrow A_n \hookrightarrow S_n \xrightarrow{\mathsf{Sgn}} \{\pm 1\} \longrightarrow 1$$

where A_n is the *alternating group* on n elements, i.e., the subgroup of S_n consisting of all even permutations, and $\mathsf{Sgn} : S_n \to \{\pm 1\}$, the *sign homomorphism*, is the unique group homomorphism such that $\mathsf{Sgn}(\sigma_i) = -1$ for all $1 \leq i \leq n-1$. Note that $\mathsf{Ker}(\mathsf{Sgn}) = A_n$ and $\mathsf{Im}(\mathsf{Sgn}) = \{\pm 1\}$.

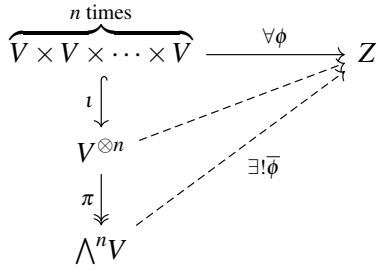
> *Remark.* A_n is simple for all $n \geq 5$. This means that A_n has no non-trivial normal subgroups for all $n \geq 5$.

Then we have two properties of the sign homomorphism:

- $\mathsf{Sgn}(1) = 1$;
- $\mathsf{Sgn}(\sigma\tau) = \mathsf{Sgn}(\sigma) \cdot \mathsf{Sgn}(\tau)$ for all $\sigma, \tau \in S_n$.

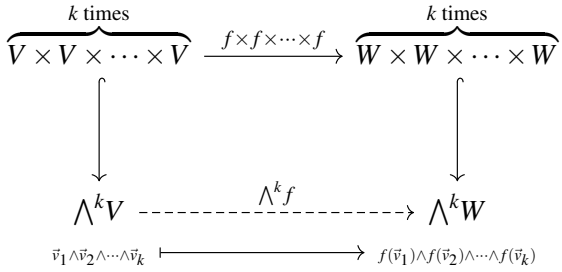
6.3 Universal Property of Exterior Powers

We have known that the k-th exterior power $\bigwedge^k V$ of a vector space V over a field $\mathbb{F}$ is the quotient of the k-th tensor power $V^{\otimes k}$ by the alternating ideal. So, consider $\dim V = n$, we have the following commutative diagram:



Here Z is any vector space over $\mathbb{F}$ and $\phi : V \times V \times \cdots \times V \to Z$ is an alternating (skew-symmetric) multilinear map, i.e., $\phi(v_1, v_2, \cdots, v_n) = 0$ if $v_i = v_j$ for some $i \neq j$. Then there exists a unique linear map $\overline{\phi} : \bigwedge^n V \to Z$ such that $\overline{\phi} \circ \pi \circ \iota = \phi$. This shows the universal property of exterior powers.

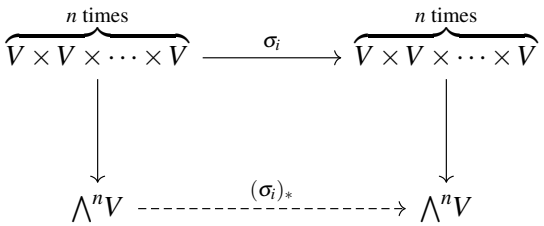
Also, we can consider the $\bigwedge^k$ as a functor applied to the map $f : V \to W$. Then we have $\bigwedge^k f : \bigwedge^k V \to \bigwedge^k W$. Then the following diagram commutes:



Note that the permutation group S_n acts on $\overbrace{V \times V \times \cdots \times V}^{n \text{ times}}$ by:

$$\sigma_i : (v_1, v_2, \cdots, v_n) \mapsto (v_1, v_2, \cdots, v_{i-1}, v_{i+1}, v_i, v_{i+2}, \cdots, v_n)$$

By the universal property of exterior powers, we have:



Consider that $a \wedge b = -b \wedge a$. Then in general, we have:

$$P \wedge Q = (-1)^{pq} Q \wedge P$$

where $P \in \bigwedge^p V$ and $Q \in \bigwedge^q V$. This is called the *graded commutativity* of exterior algebras.

6.4 Determinants and Duals

Let V be an n-dimensional vector spaces over $\mathbb{F}$ and $\mathcal{B}_V = \{v_1, v_2, \cdots, v_n\}$ be a basis of V.

As $\det V$ is a 1-dimensional vector space, so there is a basis. So the basis of $\det V$ is actually equivalent to $\det V \setminus \{0\}$ Then we have a map from $\mathcal{B}_V$ to $\mathcal{B}_{\det V}$ defined by:

$$\vec{v} = (v_1, v_2, \cdots, v_n) \mapsto v_1 \wedge v_2 \wedge \cdots \wedge v_n = \det \vec{v} \in \det V$$

Then we have the following commutative diagram:

$$\begin{array}{ccc} \mathcal{B}_{V^*} & \equiv & \mathcal{B}_V \\ \downarrow & & \downarrow \\ \mathcal{B}_{\det V^*} & \equiv & \mathcal{B}_{\det V} \end{array}$$

Note that $(\det v)^* \equiv \det v$ where $v \in \mathcal{B}_V$. So we have the following equivalence:

$$\det v^* \equiv \det v \equiv (\det v)^*$$

The first equivalence is because of the commutative diagram above, and the second equivalence is because of the definition of dual basis.

Consider L be a line over $\mathbb{F}$ and L^n defined as $\overbrace{L \otimes L \otimes \cdots \otimes L}^{n \text{ times}}$. Also, L^0 is defined as $\mathbb{F}$. Normally, we have $L^* \otimes L \to \mathbb{F}$. However, as L is 1-dimensional, we have the following isomorphism:

$$L^* \otimes L \equiv \mathbb{F}$$

Then L^* is regarded as L^{-1}, and they from a group under the tensor product operation, $(\{L^k\}, \otimes)$ where $k \in \mathbb{Z}$.

Consider V_1 and V_2 are two n-dimensional vector spaces over $\mathbb{F}$. Then we have the following diagram:

$$\begin{array}{ccccccc} V_1 & \xrightarrow{\quad f \quad} & V_2 & \xRightarrow{(-)^*} & V_1^* & \xleftarrow{\quad f^* \quad} & V_2^* \\ & \det \Big\downarrow & & & & \Big\downarrow \det & \\ \det V_1 & \xrightarrow{\quad \det f \quad} & \det V_2 & \dashrightarrow & \det V_1^* & \xleftarrow{\quad \det f^* \quad} & \det V_2^* \end{array}$$

Then we consider the left part, we have:

$$\det f \in \mathrm{Hom}(\det V_1, \det V_2) \equiv (\det V_1)^* \otimes \det V_2$$

Similarly, for the right part, we have:

$$\det f^* \in (\det V_2^*)^* \otimes \det V_1^* \equiv (\det V_2)^* * \otimes \det V_1^* \equiv \det V_2 \otimes (\det V_1)^*$$

Note that the first equivalence is due to $\det V^* \equiv (\det V)^*$. As the tensor product is commutative, we have:

$$\det f^* \equiv \det f$$

6.5 Determinant Formula

Consider the following diagram:

$$\begin{array}{ccc} \mathbb{F}^n & \xrightarrow{\quad A \quad} & \mathbb{F}^n \\ & \Big\Downarrow \det & \\ \det \mathbb{F}^n & \xrightarrow{\quad \det A \quad} & \det \mathbb{F}^n \end{array}$$

Given the standard basis $\mathcal{B} = \{\vec{e}_1, \vec{e}_2, \cdots, \vec{e}_n\}$ of $\mathbb{F}^n$, we have:

$$\det \mathcal{B} = \vec{e}_1 \wedge \vec{e}_2 \wedge \cdots \wedge \vec{e}_n$$

Note that

$$A = \begin{bmatrix} | & | & & | \\ \vec{a}_1 & \vec{a}_2 & \cdots & \vec{a}_n \\ | & | & & | \end{bmatrix}$$

Consider the map $\det A : \det \mathcal{B} \mapsto \det A \cdot \det \mathcal{B}$ where $\det A \in \mathbb{F}$ is a scalar, we have

$$\det A \cdot \det \mathcal{B} = A\vec{e}_1 \wedge A\vec{e}_2 \wedge \cdots \wedge A\vec{e}_n = \vec{a}_1 \wedge \vec{a}_2 \wedge \cdots \wedge \vec{a}_n$$

So, we know that $\det A$ is multilinear and alternating in the columns of A. Also, $\det I = 1$.

Consider the elements of A as $\vec{a}_j = \sum_{i_j=1}^{n} a_j^{i_j} \vec{e}_{i_j}$ for all $1 \le j \le n$. Then we have:

$$\vec{a}_1 \wedge \cdots \wedge \vec{a}_n = \sum_{i_1=1}^{n} a_1^{i_1} \vec{e}_{i_1} \wedge \cdots \wedge \sum_{i_n=1}^{n} a_n^{i_n} \vec{e}_{i_n} = \sum_{i_1,\cdots,i_n=1}^{n} a_1^{i_1} \cdots a_n^{i_n} \left(\vec{e}_{i_1} \wedge \cdots \wedge \vec{e}_{i_n}\right)$$

We assume that $\vec{e}_{i_k}$ are mutually distinct for all $1 \le k \le n$. Otherwise, the term is 0 because of the alternating property of exterior products. So there exists a unique permutation $\sigma \in S_n$ such that $i_k = \sigma(k)$ for all $1 \le k \le n$. Then we have:

$$\vec{a}_1 \wedge \cdots \wedge \vec{a}_n = \sum_{\sigma \in S_n} a_1^{\sigma(1)} \cdots a_n^{\sigma(n)} \left(\vec{e}_{\sigma(1)} \wedge \cdots \wedge \vec{e}_{\sigma(n)}\right) = \sum_{\sigma \in S_n} a_1^{\sigma(1)} \cdots a_n^{\sigma(n)} \operatorname{Sgn}(\sigma) \left(\vec{e}_1 \wedge \cdots \wedge \vec{e}_n\right)$$

Hence, we have the formula of determinants:

$$\det A = \sum_{\sigma \in S_n} \operatorname{Sgn}(\sigma)\, a_1^{\sigma(1)} a_2^{\sigma(2)} \cdots a_n^{\sigma(n)}$$

> *Remark.* For the magnitude part in the formula, $a_1^{\sigma(1)} a_2^{\sigma(2)} \cdots a_n^{\sigma(n)}$, they are in distinct rows and in distinct columns. They are in distinct columns because of the subscript of $a_j^{\sigma(j)}$ is j for all $1 \le j \le n$. They are in distinct rows due to the σ, otherwise it will be zero because of the alternating property of exterior products.

6.6 Properties of Determinants

The $\det A$ has the following properties:

- Linear in each column: for all $1 \leq j \leq n$;
- Alternating (skew-symmetric): $\cdots \vec{a}_i \cdots \vec{a}_j \cdots = -\cdots \vec{a}_j \cdots \vec{a}_i \cdots$ for all $i < j$;
- $\det I = 1$;

For the alternating property, we have the following evaluation from the original definition of wedge products (we assumed that $\mathrm{char}(\mathbb{F}) \neq 2$):

$$\begin{aligned}\cdots \vec{a}_i \overbrace{\cdots}^{k \text{ times}} \vec{a}_j \cdots &= (-1)^k \cdots\cdots \vec{a}_i \vec{a}_j \cdots \\ &= (-1)^{k+1} \cdots\cdots \vec{a}_j \vec{a}_i \cdots \\ &= -\cdots \vec{a}_j \cdots \vec{a}_i \cdots\end{aligned}$$

Moreover, the three properties above uniquely determine the determinant function.

> *Remark.* The first two properties can be defined on the rows of A as well and they still hold. This is because the determinant of a matrix is equal to the determinant of its transpose, which is the matrix part of $\det f^* \equiv \det f$ shown in the previous section.

If we drop the last property, then the function is called the *alternating multilinear form*. Suppose that $\phi : \mathsf{M}_{n\times n}(\mathbb{F}) \to \mathbb{F}$ is an alternating multilinear form, then we have:

$$\phi(A) = \det A \phi(I_n)$$

Proposition 6.1 The following equality holds:

$$\det \begin{bmatrix} A_1 & * \\ 0 & A_2 \end{bmatrix} = \det A_1 \cdot \det A_2$$

Proof. Consider the part on the left-hand side, we know that it is multilinear in the columns and alternating. Then we have the following evaluation:

$$\begin{aligned}\det \begin{bmatrix} A_1 & * \\ 0 & A_2 \end{bmatrix} &= \det A_1 \cdot \det \begin{bmatrix} I_{n_1} & * \\ 0 & A_2 \end{bmatrix} \\ &= \det A_1 \cdot \det A_2 \cdot \det \begin{bmatrix} I_{n_1} & * \\ 0 & I_{n_2} \end{bmatrix} \\ &= \det A_1 \cdot \det A_2 \cdot \det \begin{bmatrix} I_{n_1} & 0 \\ 0 & I_{n_2} \end{bmatrix} \\ &= \det A_1 \cdot \det A_2 \cdot \det I_{n_1+n_2} = \det A_1 \cdot \det A_2\end{aligned}$$

For the last equality, as we know the following property:

$$\cdots \vec{a}_i \cdots (k\vec{a}_i + \vec{a}_j) \cdots = k \cdots \vec{a}_i \cdots \vec{a}_j \cdots + \cdots \vec{a}_i \cdots \vec{a}_j \cdots = \cdots \vec{a}_i \cdots \vec{a}_j \cdots$$

Note that k can be 0 as well. Therefore, we can eliminate all the $*$ in the matrix by using the above property without changing the determinant value. ■

Instead of writing $\det$, we can use two pipes to denote the determinant. Concretely, we have the following determinants:

$$\left|\begin{array}{ccc|ccc} 1 & & & * & * & * \\ & 1 & & * & * & * \\ \hline & & & 1 & & \\ & & & & 1 & \\ & & & & & 1 \end{array}\right| = \left|\begin{array}{ccc|ccc} 1 & & & 0 & 0 & 0 \\ & 1 & & * & * & * \\ \hline & & & 1 & & \\ & & & & 1 & \\ & & & & & 1 \end{array}\right| = \left|\begin{array}{ccc|ccc} 1 & & & 0 & 0 & 0 \\ & 1 & & 0 & 0 & 0 \\ \hline & & & 1 & & \\ & & & & 1 & \\ & & & & & 1 \end{array}\right| = |I_5| = 1$$

For the first equality, we eliminated the first row's $*$ by using the first row. For the second equality, we eliminated the second row's $*$ by using the second row.

So, for block upper-triangular matrices, its determinant is equal to the product of the determinants of the diagonal blocks. Same for the block lower-triangular matrices.

In particular, we have the following equation:

$$\begin{vmatrix} a_{11} & \cdots & * \\ \vdots & \ddots & \vdots \\ 0 & \cdots & a_{nn} \end{vmatrix} = a_{11} \cdots a_{nn}$$

Also, $\det [a] = a \det [1] = a$.

> *Remark.* In determinant, we prefer to use a_{ij} to denote the element in the i-th row and j-th column instead of using superscript and subscript like a^i_j. This is because in determinants, we usually consider the rows and columns instead of vectors.

Consider the following determinant:

$$\begin{array}{l} \\ \text{the } i\text{-th row} \longrightarrow \end{array} \begin{vmatrix} & * & & 0 & & * & \\ a_{i,1} & \cdots & a_{i,j-1} & 1 & a_{i,j+1} & \cdots & a_{i,n} \\ & * & & 0 & & * & \end{vmatrix} = (-1)^{i-1} \begin{vmatrix} a_{i,1} & \cdots & a_{i,j-1} & 1 & a_{i,j+1} & \cdots & a_{i,n} \\ & * & & 0 & & * & \end{vmatrix}$$

$$= (-1)^{i-1+j-1} \begin{vmatrix} 1 & a_{i,1} & \cdots & \widehat{a_{i,j}} & \cdots & a_{i,n} \\ 0 & & & A^i_j & & \end{vmatrix}$$

$$= (-1)^{i+j} \det A^i_j$$

Here, $\widehat{a_{i,j}}$ means that the element $a_{i,j}$ is omitted, and A^i_j is the submatrix obtained by deleting the i-th row and j-th column of A.

Then we can consider general matrix A, for any j, we have:

$$\begin{aligned} \det A &= \det [\cdots \quad \vec{a}_j \quad \cdots] \\ &= \sum_{i=1}^{n} a^i_j \det [\cdots \quad \vec{e}_i \quad \cdots] \\ &= \sum_{i=1}^{n} a^i_j (-1)^{i+j} \det A^i_j \end{aligned}$$

This is called the *cofactor expansion* or *Laplace expansion* along the j-th column. Similarly, we can have the cofactor expansion along the i-th row.

Then we have the definition of *adjoint* of a matrix.

Definition 6.5 — Adjoint Matrices. The *adjoint matrix* of A, denoted by $\operatorname{Adj} A$, is defined as the matrix whose (i, j)-th entry is $(-1)^{i+j} \det A^j_i$.

Remark. Be aware of the notation difference between A_i^j and A_j^i. The former means deleting the j-th row and i-th column, while the latter means deleting the i-th row and j-th column. Also note that the notation of $\vec{e}_i$ means that the i-th row is 1 and other rows are 0 (standard basis vector), which is different from the notation in A_i^j and A_j^i. To conclude, the subscript is for columns and the superscript is for rows, except they are in the notation of standard basis vectors.

Proposition 6.2 The following equality holds:

$$A \cdot \operatorname{Adj} A = \operatorname{Adj} A \cdot A = \det A I_n$$

In particular, if $\det A \neq 0$, then $A^{-1} = \frac{1}{\det A} \operatorname{Adj} A$.

Proof. In particular, we just have to show

$$\sum_{k=1}^{n} a_j^k (\operatorname{Adj} A)_k^i = \det A \delta_j^i$$

From the previous Laplace expansion, we know:

$$\det A = \sum_{i=1}^{n} a_j^i (-1)^{i+j} \det A_j^i = \sum_{i=1}^{n} a_j^i (\operatorname{Adj} A)_i^j = (A \cdot \operatorname{Adj} A)_j^j$$

Then we know that for $i = j$, the equality holds. If $i \neq j$, then we can consider the following determinant:

$$\det \begin{vmatrix} \cdots & \vec{a}_j & \cdots & \vec{a}_j & \cdots \end{vmatrix} = 0$$

(the j-th column: the second $\vec{a}_j$; the i-th column: the first $\vec{a}_j$)

This means that originally, there are two same columns in the determinant, so its value is zero. Then by the Laplace expansion along the j-th column, we have:

$$0 = \sum_{k=1}^{n} a_j^k (-1)^{k+j} \det A_i^k = \sum_{k=1}^{n} a_j^k (\operatorname{Adj} A)_k^i = (A \cdot \operatorname{Adj} A)_j^i$$

■

To better understand the reason why the equality holds when $i \neq j$, we can consider the following explanation [1]. Consider the 3×3 case:

$$\underbrace{\begin{bmatrix} A_1^1 & -A_1^2 & A_1^3 \\ -A_2^1 & A_2^2 & -A_2^3 \\ A_3^1 & -A_3^2 & A_3^3 \end{bmatrix}}_{\operatorname{Adj} A} \cdot \underbrace{\begin{bmatrix} a_1^1 & a_2^1 & a_3^1 \\ a_1^2 & a_2^2 & a_3^2 \\ a_1^3 & a_2^3 & a_3^3 \end{bmatrix}}_{A}$$

If we multiply the first row of $\operatorname{Adj} A$ with the first column of A, we have the same result as the Laplace expansion along the first column:

$$a_1^1 A_1^1 - a_1^2 A_1^2 + a_1^3 A_1^3 = \begin{vmatrix} a_1^1 & a_2^1 & a_3^1 \\ a_1^2 & a_2^2 & a_3^2 \\ a_1^3 & a_2^3 & a_3^3 \end{vmatrix} = \det A = \sum_{k=1}^{3} a_1^k A_1^k = \sum_{k=1}^{3} a_1^k (\operatorname{Adj} A)_k^1$$

If we multiply the first row of $\operatorname{Adj} A$ with the second column of A, we have:

$$a_2^1 A_1^1 - a_2^2 A_1^2 + a_2^3 A_1^3 = \begin{vmatrix} a_2^1 & a_2^1 & a_3^1 \\ a_2^2 & a_2^2 & a_3^2 \\ a_2^3 & a_2^3 & a_3^3 \end{vmatrix} = 0 = \sum_{k=1}^{3} a_2^k A_1^k = \sum_{k=1}^{3} a_2^k (\operatorname{Adj} A)_k^1$$

6.7 Vandermonde Determinant

Consider the following determinant; here, the superscript means the power:

$$\det V_n = \begin{vmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \\ \vdots & \vdots & \ddots & \vdots \\ x_1^{n-1} & x_2^{n-1} & \cdots & x_n^{n-1} \end{vmatrix}$$

Then we consider $x_1, x_2, \cdots, x_{n-1}$ are fixed and we consider the determinant as a polynomial of x_n. Note that the degree of x_n is $n-1$, and the polynomial is:

$$\det V_n = (-1)^{n+1}|\cdots| + (-1)^{n+2}x_n|\cdots| + \cdots + (-1)^{n+n}x_n^{n-1}\begin{vmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ x_1^{n-1} & x_2^{n-1} & \cdots & x_{n-1}^{n-1} \end{vmatrix}$$

Also note that if $x_n = x_i$ for some $1 \le i \le n-1$, let say $i = n-1$, then the determinant becomes:

$$\begin{vmatrix} 1 & 1 & \cdots & 1 & 1 \\ x_1 & x_2 & \cdots & x_{n-1} & x_{n-1} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ x_1^{n-1} & x_2^{n-1} & \cdots & x_{n-1}^{n-1} & x_{n-1}^{n-1} \end{vmatrix} = 0$$

This means that $x_n - x_i$ is a factor of the polynomial. Therefore, by the fundamental theorem of algebra, we have:

$$\det V_n = C\overbrace{(x_n - x_1)(x_n - x_2)\cdots(x_n - x_{n-1})}^{n-1 \text{ factors}}$$

Here C is a constant that does not depend on x_n. To find C, we can consider the coefficient of x_n^{n-1}. Note that the coefficient of x_n^{n-1} in the above polynomial expansion is $\det V_{n-1}$. So $C = \det V_{n-1}$. Then by induction, we have:

$$\det V_n = \prod_{1 \le i < j \le n} (x_j - x_i)$$

6.8 Feynman Diagram Formula

Consider the case where $\mathrm{char}(\mathbb{F}) = 0$. Let A be a $n \times n$ matrix and I be the identity matrix of order n. Then we have the following formula:

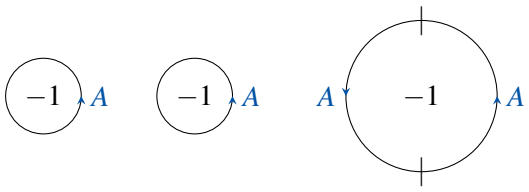
$$\det(I+tA) = 1 - \operatorname{tr} A\, t + \left(\frac{(\operatorname{tr} A)^2}{2!} - \frac{\operatorname{tr} A^2}{2} \right) t^2 - \cdots + (-1)^n \det A\, t^n$$

This is called the *Feynman diagram formula*, as it is inspired by Feynman diagrams in quantum field theory. From this formula, the determinant can be expressed by traces.

It is hard to remember the coefficients in the formula. However, we can use the following method to derive them. Consider the following diagram for t^1 term:

Here the circle means a trace operation, and the arrow means A. So the coefficient is $-\operatorname{tr} A$.

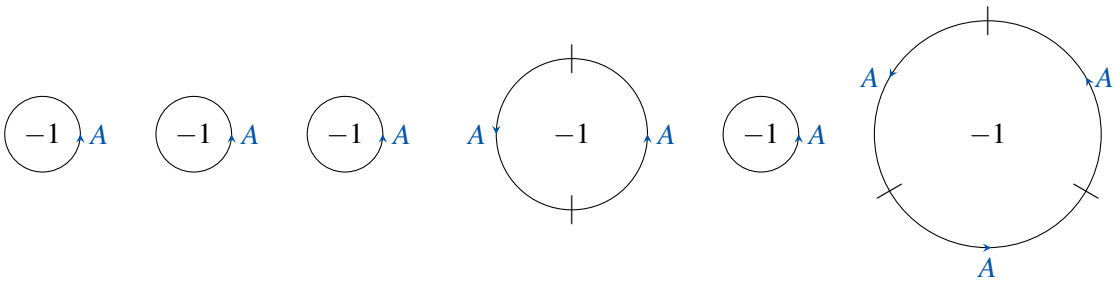
For t^2 term, we have diagram:



The left two circles mean $(-\operatorname{tr} A)^2$, and we have to divide by 2! because of the symmetry of the two identical circles. The right circle means $-\operatorname{tr} A^2$, but this is a cyclic group of order 2, so we have to divide by 2. Therefore, the total term for t^2 is:

$$\frac{(-\operatorname{tr} A)^2}{2!} - \frac{\operatorname{tr} A^2}{2} = \frac{(\operatorname{tr} A)^2}{2!} - \frac{\operatorname{tr} A^2}{2}$$

For t^3 term, we have diagram:



The left three circles mean $(-\operatorname{tr} A)^3$, and we have to divide by 3! because of the symmetry of the three identical circles. The second diagram means $(-\operatorname{tr} A)(-\operatorname{tr} A^2)$, and we have to divide by 2 because of the cyclic group of order 2 on the bigger circle. The last diagram means $-\operatorname{tr} A^3$, and this is a cyclic group of order 3, so we have to divide by 3. Therefore, the total term for t^3 is:

$$\frac{(-\operatorname{tr} A)^3}{3!} + \frac{(-\operatorname{tr} A)(-\operatorname{tr} A^2)}{2} - \frac{\operatorname{tr} A^3}{3} = -\frac{(\operatorname{tr} A)^3}{3!} + \frac{(\operatorname{tr} A)(\operatorname{tr} A^2)}{2} - \frac{\operatorname{tr} A^3}{3}$$

7. Canonical Forms of Endomorphisms

> "Babies have to survive, so they have the strong desire to learn stuffs. You think you are not good at math because you don't have the strong desire to learn math."
>
> GUOWU MENG

7.1 Diagonal Forms

Before, we have studied the canonical matrix representation of linear maps between two different dimension vector spaces. It is natural to ask what is the canonical form of linear maps from a vector space to itself, i.e. endomorphisms. Consider the following diagram:

$$\begin{array}{ccc} V & \xrightarrow{T} & V \\ {\scriptstyle [-]_{\mathcal{B}}}\downarrow & & \downarrow{\scriptstyle [-]_{\mathcal{B}}} \\ \mathbb{F}^n & \xrightarrow{A} & \mathbb{F}^n \\ \downarrow & & \downarrow \\ \mathbb{F}^n & \xrightarrow{\overline{A}} & \mathbb{F}^n \end{array}$$

As both the domain and codomain are the same vector space, both basis $\mathcal{B}$ are the same. So the matrix representation of T is much more restricted. The $\overline{A}$ is simplest looking matrix repsentation of T, but what does it look like?

Generically, we have the following form:

$$\overline{A} = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix}$$

where empty places are filled with zeros. It is called the *diagonal matrix*. Here λ_i are the *eigenvalues* of T. If such form exists, we say that T is *completely reducible*, or normally say that T is *diagonalisable*. If T is not completely reducible, then we have to consider more complicated forms, which will be discussed later.

Then we have the diagram:

$$\begin{array}{ccc} \mathbb{F}^n & \xrightarrow{A} & \mathbb{F}^n \\ {\scriptstyle P^{-1}}\downarrow\cong & & \cong\downarrow{\scriptstyle P^{-1}} \\ \mathbb{F}^n & \xrightarrow{D} & \mathbb{F}^n \\ & \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} & \end{array}$$

Here P is the change of basis matrix from the basis that gives A to the basis that gives D. Then we have:

$$A = PDP^{-1}$$

We have $A \sim D$, i.e. A is similar to D.

Then we have two questions:

1. How do we know whether T is completely reducible?
2. If T is completely reducible, how can we find P and D?

Assume that $D = \begin{bmatrix} \lambda_1 I_{n_1} & & \\ & \ddots & \\ & & \lambda_k I_{n_k} \end{bmatrix}$, where $\lambda_i \in \mathbb{F}$ are distinct eigenvalues and I_{n_i} are identity matrices of order n_i, $n_i > 0$ and $\sum_{i=1}^{k} n_i = n$. For example, we have:

$$D = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

where $\lambda_1 = 1$, $\lambda_2 = 2$, $n_1 = 2$ and $n_2 = 1$.

Then we have the decomposition of V:

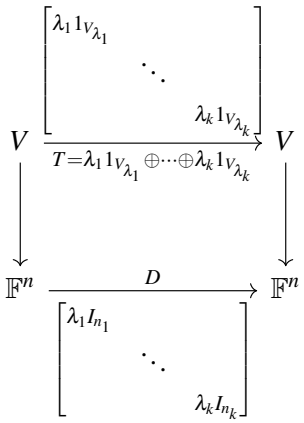
$$V = V_{\lambda_1} \oplus V_{\lambda_2} \oplus \cdots \oplus V_{\lambda_k}$$

where $V_i = \mathsf{Ker}(T - \lambda_i 1_V)$ are the *eigenspaces* of T corresponding to eigenvalues λ_i. Moreover, we have the decomposition of $\mathbb{F}^n$:

$$\mathbb{F}^n = \mathsf{Span}(e_1, \cdots, e_{n_1}) \oplus \mathsf{Span}(e_{n_1+1}, \cdots, e_{n_1+n_2}) \oplus \cdots \oplus \mathsf{Span}(e_{n_1+\cdots+n_{k-1}+1}, \cdots, e_{n_1+\cdots+n_k})$$

Note that $\dim V_{\lambda_i} = n_i$ and $\sum_{i=1}^{k} n_i = n$.

Then we have the following commutative diagram:



In other words, if T is completely reducible, then there are distinct numbers $\lambda_1, \cdots, \lambda_k \in \mathbb{F}$ and a non-trivial decomposition $V = V_{\lambda_1} \oplus \cdots \oplus V_{\lambda_k}$ such that $T|_{V_{\lambda_i}} = \lambda_i 1_{V_{\lambda_i}}$ for each $1 \leq i \leq k$, and $T = \lambda_1 1_{V_{\lambda_1}} \oplus \cdots \oplus \lambda_k 1_{V_{\lambda_k}}$. Each non-zero vector v_i in V_{λ_i} is an *eigenvector* of T corresponding to eigenvalue λ_i. This answered the first question.

Then how to find the eigenvalues and eigenspaces? We can consider the following linear map:

$$\lambda_i 1_{V_{\lambda_i}} : V_{\lambda_i} \to V_{\lambda_i}, \quad x \mapsto \lambda_i x$$

Then we have the following equation:

$$Tx = \lambda_i x \iff (\lambda_i 1_V - T)x = 0 \iff x \in \mathsf{Ker}(\lambda_i 1_V - T)$$

As x is non-zero, then $(\lambda_i 1_V - T)$ is not injective, i.e. not invertible. Therefore, we have:

$$\det(\lambda_i 1_V - T) = 0$$

So the eigenvalues λ_i are exactly the roots of the polynomial $\det(\lambda 1_V - T)$, which is called the *characteristic polynomial* of T. Note that $p_T(\lambda) = \det(\lambda 1_V - T)$ is a polynomial of degree $n = \dim V$. Similarly, we can define the characteristic polynomial of a matrix A as $p_A(\lambda) = \det(\lambda I_n - A)$.

For example, consider the following matrix:

$$A = \begin{bmatrix} 1 & 3 \\ 0 & 2 \end{bmatrix}, \quad \lambda I - A = \begin{bmatrix} \lambda - 1 & -3 \\ 0 & \lambda - 2 \end{bmatrix}, \quad p_A(\lambda) = (\lambda - 1)(\lambda - 2)$$

The roots of $p_A(\lambda)$ are 1 and 2, so the eigenvalues of A are 1 and 2. Then we can find the eigenspaces:

$$V_{\lambda=1} = \mathsf{Nul}(1 \cdot I - A) = \mathsf{Nul} \begin{bmatrix} 0 & -3 \\ 0 & -1 \end{bmatrix} = \mathsf{Nul} \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} = \mathsf{Span} \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$
$$V_{\lambda=2} = \mathsf{Nul}(2 \cdot I - A) = \mathsf{Nul} \begin{bmatrix} 1 & -3 \\ 0 & 0 \end{bmatrix} = \mathsf{Span} \begin{bmatrix} 3 \\ 1 \end{bmatrix}$$

Then we have:

$$A = \begin{bmatrix} 1 & 3 \\ 0 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 3 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 1 & 3 \\ 0 & 1 \end{bmatrix}^{-1} = PDP^{-1}$$

Remark. To find the null space, we first use row operations to reduce the matrix to its row echelon form. Then we consider the number of free variables to find the number of basis vectors in the null space. Then we can let one free variable as 1 and other free variables as 0 to find the value of each pivot variable. Repeating this process for each free variable, we can find all basis vectors of the null space.

For example, for the first matrix above, we have: $0\cdot 1+1\cdot x_2=0 \implies x_2=0$. So the null space is $\mathsf{Span}\begin{bmatrix}1\\0\end{bmatrix}$. For the second matrix above, we have: $1\cdot x_1-3\cdot 1=0 \implies x_1=3$. So the null space is $\mathsf{Span}\begin{bmatrix}3\\1\end{bmatrix}$.

In matrix, we have:

$$\begin{bmatrix}A\vec{p}_1 & \cdots & A\vec{p}_n\end{bmatrix}=AP=PD=\begin{bmatrix}\lambda_1\vec{p}_1 & \cdots & \lambda_n\vec{p}_n\end{bmatrix} \iff A\vec{p}_i=\lambda_i\vec{p}_i$$

Proposition 7.1 The following are equivalent:

1. T is completely reducible.
2. $T=\lambda_1 1_{V_{\lambda_1}}\oplus\cdots\oplus\lambda_k 1_{V_{\lambda_k}}$ for some distinct eigenvalues $\lambda_1,\cdots,\lambda_k$ and non-trivial decomposition $V=V_{\lambda_1}\oplus\cdots\oplus V_{\lambda_k}$.
3. V has an eigenvector basis of T, i.e. there exists a basis of V consisting of eigenvectors of T.
4. $\dim V=\sum_{i=1}^k \dim E_{\lambda_i}(T)=\sum_{i=1}^k \dim V_{\lambda_i}$, where $\lambda_1,\cdots,\lambda_k$ are the distinct eigenvalues of T and $V_{\lambda_i}=E_{\lambda_i}(T)$ are the eigenspaces of T.

Consider the following example:

■ **Example 7.1** $A=\begin{bmatrix}0&1\\0&0\end{bmatrix}$ is not completely reducible. The $p_A(\lambda)=\lambda^2$, so the only eigenvalue is 0. Then we have:

$$V_{\lambda=0}=\mathsf{Nul}(0\cdot I-A)=\mathsf{Nul}\begin{bmatrix}0&-1\\0&0\end{bmatrix}=\mathsf{Span}\begin{bmatrix}1\\0\end{bmatrix}$$

So there does not exist a eigenvector basis of A, as choosing any two vectors in $V_{\lambda=0}$ will be linearly dependent. Therefore, A is not completely reducible. ■

Proposition 7.2 $E_{\lambda_1}+\cdots+E_{\lambda_k}$ is a direct sum.

Proof. We just need to check if $x_1+\cdots+x_k=0$ with $x_i\in E_{\lambda_i}$, then each $x_i=0$. We can use induction on k. For $k=1$, we have $x_1=0 \implies x_1=0$. Assume that the statement holds for $k-1$. Then we have:

$$\begin{cases}x_1+\cdots+x_k=0\\ Tx_1+\cdots+Tx_k=\lambda_1x_1+\cdots+\lambda_kx_k=0\end{cases}$$

Then we subtract λ_k times the first equation from the second equation, we have:

$$(\lambda_1-\lambda_k)x_1+\cdots+(\lambda_{k-1}-\lambda_k)x_{k-1}=0$$

Given that λ_i are distinct, by the induction hypothesis, we have $(\lambda_i-\lambda_k)x_i=0 \implies E_{\lambda_i}\ni x_i=0$ for each $1\le i\le k-1$. Then by the first equation, we have $x_k=0$. This completed the induction. ■

Then we know that the sum of eigenspaces is a direct sum, i.e. $E_{\lambda_1}\oplus\cdots\oplus E_{\lambda_k}$. Then we have:

$$\dim V=\sum\dim E_{\lambda_i}(T)$$

■ **Example 7.2** Let $A = \begin{bmatrix} 1 & 0 & 4 \\ 0 & 1 & 3 \\ 0 & 0 & 2 \end{bmatrix}$. Then we have $p_A(\lambda) = (\lambda-1)^2(\lambda-2)$. The eigenvalues are 1 and 2, where $\lambda = 1$ has algebraic multiplicity 2 and $\lambda = 2$ has algebraic multiplicity 1. Then we can find the eigenspaces:

$$E_{\lambda=1}(A) = \mathsf{Nul}(1\cdot I - A) = \mathsf{Nul}\begin{bmatrix} 0 & 0 & -4 \\ 0 & 0 & -3 \\ 0 & 0 & -1 \end{bmatrix} = \mathsf{Nul}\begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \mathsf{Span}\left\{\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}\right\}$$

$$E_{\lambda=2}(A) = \mathsf{Nul}(2\cdot I - A) = \mathsf{Nul}\begin{bmatrix} 1 & 0 & -4 \\ 0 & 1 & -3 \\ 0 & 0 & 0 \end{bmatrix} = \mathsf{Span}\begin{bmatrix} 4 \\ 3 \\ 1 \end{bmatrix}$$

Then we have $\dim E_{\lambda=1} + \dim E_{\lambda=2} = 2+1 = 3 = \dim V$. Therefore, A is completely reducible. Then we can find the diagonalisation:

$$\begin{bmatrix} 1 & 0 & 4 \\ 0 & 1 & 3 \\ 0 & 0 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 4 \\ 0 & 1 & 3 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix}\begin{bmatrix} 1 & 0 & 4 \\ 0 & 1 & 3 \\ 0 & 0 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 4 & 1 & 0 \\ 3 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 4 & 1 & 0 \\ 3 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix}^{-1}$$

■

Completely reducible matrix representations are "the" simplest forms of endomorphisms. Note that it is not unique, it is unique up to isomorphism, unless the field is ordered. However, not all endomorphisms are completely reducible. Then we have another term called *semisimple*. These two terms are borrowed from representation theory of lie algebras.

Definition 7.1 — Completely Reducible. We say T is a completely reducible if there exists a matrix representation of T of the following form:

$$\begin{bmatrix} \lambda_1 I_{n_1} & & \\ & \ddots & \\ & & \lambda_k I_{n_k} \end{bmatrix}$$

Equivalently, T is completely reducible if V has a non-trivial decomposition $V = V_{\lambda_1} \oplus \cdots \oplus V_{\lambda_k}$ with respect to which $T = \lambda_1 1_{V_{\lambda_1}} \oplus \cdots \oplus \lambda_k 1_{V_{\lambda_k}}$ for some distinct eigenvalues $\lambda_1, \cdots, \lambda_k$.

Definition 7.2 — Semisimple. We say T is semisimple if $T \otimes_{\mathbb{F}} \overline{\mathbb{F}} : V \otimes_{\mathbb{F}} \overline{\mathbb{F}} \to V \otimes_{\mathbb{F}} \overline{\mathbb{F}}$ is completely reducible, where $\overline{\mathbb{F}}$ is the algebraic closure of $\mathbb{F}$ and $V \otimes_{\mathbb{F}} \overline{\mathbb{F}}$ is linear space over $\overline{\mathbb{F}}$.

Remark. We can take $\mathbb{F} = \mathbb{R}$, then $\overline{\mathbb{F}} = \mathbb{C}$. Algebraic closure means that every polynomial in $\overline{\mathbb{F}}[x]$ has a root in $\overline{\mathbb{F}}$. For example, x^2+1 has no root in $\mathbb{R}$, but it has roots $\pm i$ in $\mathbb{C}$.

Note that $- \otimes \mathbb{F} \equiv \mathsf{id}_{\mathbb{F}}$, so if we change it to $- \otimes_{\mathbb{F}} \overline{\mathbb{F}}$, then we are just changing the field from $\mathbb{F}$ to $\overline{\mathbb{F}}$ without changing the values inside. For example, 1 can be viewed as an element in $\mathbb{R}$ or $\mathbb{C}$.

In general, T is not semisimple, but it can be decomposed into a semisimple part and a *nilpotent* part. Moreover, this decomposition is unique.

We can consider the $\mathsf{End}(V) \equiv \mathsf{M}_{n\times n}(\mathbb{F}) \equiv \mathbb{F}^{n^2}$ as a vector space. Then $T \in \mathbb{F}^{n^2}$ is a vector. Then such the set of containing such T forms a dense open subset of $\mathsf{End}(V) = \mathbb{F}^{n^2}$. The dense open subset is in the Zariski topology. More precisely, the set of all completely reducible endomorphisms with distinct eigenvalues forms a dense open subset of $\mathsf{End}(V)$. We will study Zariski topology next section.

Once we know that completely reducible endomorphisms are dense in $\mathsf{End}(V)$, then if we want to prove some identity, it suffices to prove it for completely reducible endormophisms. One of the example is the Cayley-Hamilton theorem.

Theorem 7.1 — Cayley-Hamilton Theorem. Let $T : V \to V$ be an endomorphism of a finite-dimensional vector space V over $\mathbb{F}$. Then T satisfies its own characteristic polynomial, i.e. $p_T(\lambda)|_{\lambda=T} = 0$.

Remark. $p_T(\lambda) = \det(\lambda 1_V - T) = \lambda^n + \cdots + (-1)^n \det(T)\lambda^0$, where $\lambda^0 = 1$ and $T^0 = 1_V$.

Proof. As $p_T(\lambda)|_{\lambda=T}$ is a polynomial in T, it suffices to verify the theorem on a dense set.

Let $T = \lambda_1 1_{V_{\lambda_1}} \oplus \cdots \oplus \lambda_k 1_{V_{\lambda_k}}$ be a completely reducible endomorphism with distinct eigenvalues $\lambda_1, \cdots, \lambda_k$ and non-trivial decomposition $V = V_{\lambda_1} \oplus \cdots \oplus V_{\lambda_k}$. Then we have $1_V = 1_{V_{\lambda_1}} \oplus \cdots \oplus 1_{V_{\lambda_k}}$. Therefore, we have:

$$\lambda 1_V - T = (\lambda - \lambda_1)1_{V_{\lambda_1}} \oplus \cdots \oplus (\lambda - \lambda_k)1_{V_{\lambda_k}}$$

Then the characteristic polynomial is:

$$p_T(\lambda) = \det(\lambda 1_V - T) = (\lambda - \lambda_1)^{\dim V_{\lambda_1}} \cdots (\lambda - \lambda_k)^{\dim V_{\lambda_k}} = \prod_{i=1}^{k} (\lambda - \lambda_i)^{n_i}$$

where $n_i = \dim V_{\lambda_i}$. Note that $\lambda_i 1_{V_{\lambda_i}} - T = 0$ on V_{λ_i}, as $T|_{V_{\lambda_i}} = \lambda_i 1_{V_{\lambda_i}}$. Therefore, we have:

$$p_T(\lambda)|_{\lambda=T} = \prod_{i=1}^{k} (\lambda_i 1_V - T)^{n_i} = 0$$

As for any $v \in V$, we can write $v = v_1 + \cdots + v_k$ with $v_i \in V_{\lambda_i}$, then we have:

$$(\lambda_i 1_{V_{\lambda_i}} - T)^{n_i}(v_i) = 0 \quad \forall i \implies p_T(\lambda)|_{\lambda=T}(v) = 0$$

This completed the proof. ■

If T is completely reducible, then

$$n_i = \dim V_{\lambda_i}$$

where n_i is the algebraic multiplicity of eigenvalue λ_i and $\dim V_{\lambda_i}$ is the geometric multiplicity of eigenvalue λ_i. In general, we have $n_i \geq \dim V_{\lambda_i}$. Then $\{\lambda_1, \cdots, \lambda_k\}$ is the set of roots of $p_T(\lambda)$ and $V_{\lambda_i} = \mathsf{Ker}(\lambda_i 1_V - T)$.

Then for any T, if the set of roots of $p_T(\lambda)$ in $\mathbb{F}$ is $\{\lambda_1, \cdots, \lambda_k\}$, then we can define the *generalised eigenspaces*:

$$V_{\lambda_i} = \mathsf{Ker}(\lambda_i 1_V - T) \quad \forall 1 \leq i \leq k$$

Then we check whether $\dim V = \sum_{i=1}^{k} \dim V_{\lambda_i}$. If it holds, then T is completely reducible. If not, then T is not. So this characterise completely reducible endomorphisms.

7.2 Zariski Topology

Before studying Zariski topology, we first introduce *affine spaces*.

Definition 7.3 — Affine Spaces. A set $\mathbb{A}$ is called an *affine space* over a field $\mathbb{F}$ if it is a principal $(\mathbb{F}^n,+)$-set, i.e. there is a free and transitive action of the additive group $(\mathbb{F}^n,+)$ on $\mathbb{A}$:

$$+ : \mathbb{A}\times\mathbb{F}^n \to \mathbb{A}, \quad (P,\vec{v}) \mapsto P+\vec{v}$$

Each element $P \in \mathbb{A}$ is called a *point* in $\mathbb{A}$.

Principal means that the group action is free and transitive. Free means that if g is not the identity element, then $g\cdot x \neq x$ for any x in the set. Transitivity means that any two elements x,y in the set are related by some action of the group, g, such that $g\cdot x = y$.

For example, consider the $SO(2)$ action on the plane $\mathbb{R}^2$. The action is not free and not transitive. It is not free because rotating a point on the plane by 0 degree (the identity element) keeps the point unchanged, but rotating it by any other angle will change the point. It is not transitive because there is no rotation that can map a point to another point with a different distance from the origin. However, if we consider the orbits of the action, i.e. circles centered at the origin, then the action is transitive on each orbit and free except for the origin.

Then we introduce what topology is.

Definition 7.4 — Topology. Let X be a set. A *topology* on X is a collection τ of subsets of X such that:

1. $\phi, X \in \tau$;
2. the union of any collection of sets in τ is also in τ;
3. the intersection of any finite number of sets in τ is also in τ.

The pair (X,τ) is called a *topological space*. Each set in τ is called an *open set* in X.

We can define *closed sets* in X as the complements of open sets in X. Then we have the following equivalent definition of topology.

Definition 7.5 — Topology (Closed Set Version). Let X be a set. A *topology* on X is a collection τ of subsets of X such that:

1. $\phi, X \in \tau$;
2. the intersection of any collection of sets in τ is also in τ;
3. the union of any finite number of sets in τ is also in τ.

The pair (X,τ) is called a *topological space*. Each set in τ is called an *closed set* in X.

Then Zariski topology is defined as follows.

Definition 7.6 — Zariski Topology. Let $\mathbb{A}$ be an affine space over a field $\mathbb{F}$. The *Zariski topology* on $\mathbb{A}$ is defined by taking the closed sets to be the zero loci of sets of polynomials in $\mathbb{F}[x_1,\cdots,x_n]$. More precisely, for any set of polynomials $S \subseteq \mathbb{F}[x_1,\cdots,x_n]$, the corresponding closed set is:

$$V(S) = \{P \in \mathbb{A} : f(P) = 0 \quad \forall f \in S\} = \bigcap_{\alpha} \{f_\alpha = 0\}$$

The pair $(\mathbb{A},\tau_{Zar})$ is called a *Zariski topological space*, where τ_{Zar} is the Zariski topology on $\mathbb{A}$.

Then the $A \in \mathbb{F}^{n^2} \equiv \mathbb{A}^{n^2}_{\mathbb{F}}$ can be viewed as a point in the affine space $\mathbb{A}^{n^2}_{\mathbb{F}}$ over $\mathbb{F}$. Then the set of all completely reducible endomorphisms with distinct eigenvalues forms a dense open subset of $\mathsf{End}(V) = \mathbb{F}^{n^2}$ in the Zariski topology. Dense means that its closure is the whole space. Open means that its complement is a closed set, i.e. the zero locus of some set of polynomials in $\mathbb{F}[x_1,\cdots,x_{n^2}]$.

7.3 Ring Theory

Before studying the canonical forms of not completely reducible endormorphisms, we need to introduce some concepts in ring theory.

Definition 7.7 — Domain. A *domain* is a non-trivial commutative ring R with unity $1_R \neq 0_R$ if non-zero elements $a, b \in R$ satisfy $ab \neq 0_R$.

■ **Example 7.3** $\mathbb{Z}$ is a domain. Given any two non-zero integers $a, b \in \mathbb{Z}$, we have $ab \neq 0$. ■

■ **Example 7.4** $\mathbb{Z}/6$ is not a domain. For example, $2, 3 \in \mathbb{Z}/6$ are non-zero elements, but $2 \cdot 3 = 0$ in $\mathbb{Z}/6$. ■

Definition 7.8 — Module. A module over R is an abelian group $(M, +)$ together with a ring action of R on $(M, +)$.

■ **Example 7.5** R itself is a module over R with the ring action being the multiplication in R. ■

Definition 7.9 — Submodule. A *submodule* N of a module M over a ring R is a subgroup of $(M, +)$ that is closed under the ring action of R on M, i.e. for any $r \in R$ and $n \in N$, we have $r \cdot n \in N$.

Definition 7.10 — Ideal. An *ideal* I of a ring R is a submodule of the module R over itself.

■ **Example 7.6** Consider $\mathbb{F}$ over itself. Then the only ideals are $\{0\}$ and $\mathbb{F}$ itself. So the ideal of a field is trivial. ■

■ **Example 7.7** Consider $\mathbb{Z}$ over itself. Then the ideals are all of the form $(n) = n\mathbb{Z} = \{nk : k \in \mathbb{Z}\}$ for some $n \in \mathbb{Z}$. So the ideals of $\mathbb{Z}$ are non-trivial. For example, $(2) = \{0, \pm 2, \pm 4, \cdots\}$. ■

Definition 7.11 — Principal Ideal Domain. A *principal ideal domain* (PID) is a domain R such that every ideal of R is of the form $(a) = aR$ for some $a \in R$.

■ **Example 7.8** $\mathbb{Z}$ is a principal ideal domain, as every ideal of $\mathbb{Z}$ is of the form $(n) = n\mathbb{Z}$ for some $n \in \mathbb{Z}$. ■

■ **Example 7.9** $\mathbb{F}[x]$ is a principal ideal domain, as every ideal of $\mathbb{F}[x]$ is of the form $(f(x)) = f(x)\mathbb{F}[x]$ for some $f(x) \in \mathbb{F}[x]$. It can be proved using the division algorithm of polynomials. ■

Definition 7.12 — Finitely Generated Module. A module M over a ring R is called *finitely generated* if M is the span of a finite set of elements in M, i.e., $M = \langle m_1, m_2, \cdots, m_k \rangle$ for some $m_1, m_2, \cdots, m_k \in M$. It may not be unique.

Note that we do not use the definition of the finite-dimensional vector space here, as a module over a ring may not have a basis. There exists something called the torsion module that prevents the existence of basis. We will discuss it later.

Then we introduce the following theorem which can derive Jordan canonical form.

Theorem 7.2 — Classification Theorem of Finitely Generated Modules over a PID. Let R be a principal ideal domain and M be a finitely generated module over R. Then M is isomorphic to a finite direct sum of cyclic modules of the form:

$$M \cong R^r \oplus \bigoplus_{i=1}^{m} R/(a_i) = R^r \oplus R/(a_1) \oplus \cdots \oplus R/(a_m)$$

with $a_i \in R \setminus \{0\}$ and $a_i | a_{i+1}$ for each $1 \leq i \leq m-1$.

Remark. Note that $a|b$ means that there exists some $c \in R$ such that $b = ac$.

Here R^r is the free part of M and $\bigoplus_{i=1}^{m} R/(a_i)$ is the torsion part of M. The torsion part prevents the existence of basis of M. If the torsion part is trivial, i.e., $m = 0$, then M is a free module and has a basis. Moreover, r is the rank of M and is unique. a_i are called the invariant factors of M and are unique up to multiplication by units in R. This is called the invariant factor decomposition of M. There is another decomposition called primary decomposition, or elementary divisor decomposition, or Chinese Remainder decomposition.

Theorem 7.3 — Classification Theorem of Finitely Generated Modules over a PID (Primary Decomposition). Let R be a principal ideal domain and M be a finitely generated module over R. Then M is isomorphic to a finite direct sum of cyclic modules of the form:

$$M \cong R^r \oplus \bigoplus_{i=1}^{m} R/(p_i^{e_i}) = R^r \oplus R/(p_1^{e_1}) \oplus \cdots \oplus R/(p_m^{e_m})$$

with p_i being prime or irreducible elements in R and $e_i \in \mathbb{Z}^+$ for each $1 \leq i \leq m$.

Remark. As R is a PID, so every ideal is principal. Therefore, every ideal generated by a prime or irreducible element is a prime ideal. This is why we call it primary decomposition.

For any ring R, we can decomposite as follows:

$$R = \{0\} \cup R^\times \cup S$$

where $R^\times$ is the set of units in R and S is the set of non-units and non-zero elements in R. Then any $u \in R$ is called a unit if there exists some $v \in R$ such that $uv = vu = 1_R$. For example, in $\mathbb{Z}$, the units are ± 1. In $\mathbb{F}[x]$, the units are all non-zero constant polynomials.

Then the set of all prime elements and the set of all irreducible elements in R are subsests of S. In general, they are not the same. The set of all prime elements is a subset of the set of all irreducible elements. However, in a principal ideal domain they are the same. Irreducible elements are those elements that cannot be factored into the product of two non-unit elements, i.e., if $x \neq 0$ and $x \notin R^\times$, then whenever $x = yz$, then y or z must be a unit.

7.4 Jordan Canonical Form

Let V be a finite-dimensional linear space over an algebraically closed field $\mathbb{F}$, e.g. $\mathbb{C}$. Then $\mathbb{F}[x]$ is a principal ideal domain and $x-\lambda_i$ are the prime or irreducible elements in $\mathbb{F}[x]$ for each $\lambda_i \in \mathbb{F}$.

> *Remark.* If we take non-zero $\alpha \in \mathbb{F}$, then $\alpha(x-\lambda_i)$ is also an irreducible element in $\mathbb{F}[x]$, as α is a unit in $\mathbb{F}[x]$ and we have $(x-\lambda_i) = (\alpha(x-\lambda_i))$. Therefore, the irreducible elements are only unique up to multiplication by units. We can just choose monic polynomials as the irreducible elements.

Then for any endomorphism $T : V \to V$. It is equivalent to consider V as a module over $\mathbb{F}[x]$ with the ring action defined as:

$$F[x] \times V \to V, \quad (p(x), v) \mapsto p(T)v$$

■ **Example 7.10** Let $p(x) = 2x^2 + 3x - 1 \in \mathbb{F}[x]$ and $T \in \mathsf{End}(V)$. Then for any $v \in V$, we have $p(T)v = 2T^2v + 3Tv - v$. ■

V is the finite-dimensional linear space over $\mathbb{F}$, so it is a finitely generated module over $\mathbb{F}[x]$ with rank 0. It is the torsion part only. Therefore, by the classification theorem of finitely generated modules over a PID, we have:

$$V \cong \bigoplus_{i=1}^{m} \frac{\mathbb{F}[x]}{(x-\lambda_i)^{e_i}} = \frac{\mathbb{F}[x]}{(x-\lambda_1)^{e_1}} \oplus \cdots \oplus \frac{\mathbb{F}[x]}{(x-\lambda_m)^{e_m}}$$

Note that T is the same as the multiplication by x in the module, i.e., $x\cdot : V \to V$ defined as $v \mapsto xv$. Then for each cyclic module $\dfrac{\mathbb{F}[x]}{(x-\lambda_i)^{e_i}}$, we have the dimension being e_i. Therefore, we have the basis on $\dfrac{\mathbb{F}[x]}{(x-\lambda_i)^{e_i}}$ as:

$$\mathcal{B}_i = \{1, (x-\lambda_i), (x-\lambda_i)^2, \cdots, (x-\lambda_i)^{e_i-1}\}$$

Then we consider the following diagram:

$$\begin{array}{ccc} \frac{\mathbb{F}[x]}{(x-\lambda_i)^{e_i}} & \xrightarrow[T_i]{x\cdot} & \frac{\mathbb{F}[x]}{(x-\lambda_i)^{e_i}} \\ \Big\downarrow{\scriptstyle [-]_{\mathcal{B}_i}} & & \Big\downarrow{\scriptstyle [-]_{\mathcal{B}_i}} \\ \mathbb{F}^{e_i} & \xrightarrow[J_{e_i}(\lambda_i)]{} & \mathbb{F}^{e_i} \end{array}$$

Then what is $J_{e_i}(\lambda_i)$? We have:

$$x \cdot 1 = x = 1 \cdot (x-\lambda_i) + \lambda_i \cdot 1$$

So the first column of $J_{e_i}(\lambda_i)$ is $[\lambda_i \quad 1 \quad 0 \quad \cdots \quad 0]^T$. Similarly, we have:

$$\begin{aligned} x \cdot (x-\lambda_i) &= 1 \cdot (x-\lambda_i)^2 + \lambda_i \cdot (x-\lambda_i) \\ x \cdot (x-\lambda_i)^{e_i-1} &= 1 \cdot (x-\lambda_i)^{e_i} + \lambda_i \cdot (x-\lambda_i)^{e_i-1} = \lambda_i \cdot (x-\lambda_i)^{e_i-1} \end{aligned}$$

So the matrix representation of $x\cdot$ on $\dfrac{\mathbb{F}[x]}{(x-\lambda_i)^{e_i}}$ with respect to the basis $\mathcal{B}_i$ is:

$$J_{e_i}(\lambda_i) = \begin{bmatrix} \lambda_i & & & & \\ 1 & \lambda_i & & & \\ & 1 & \lambda_i & & \\ & & \ddots & \ddots & \\ & & & 1 & \lambda_i \end{bmatrix}$$

We can switch the order of basis elements in $\mathcal{B}_i$ to get the following equivalent representation:

$$J_{e_i}(\lambda_i) = \begin{bmatrix} \lambda_i & 1 & & & \\ & \lambda_i & 1 & & \\ & & \lambda_i & \ddots & \\ & & & \ddots & 1 \\ & & & & \lambda_i \end{bmatrix}$$

This is called a *Jordan block* of size e_i with eigenvalue λ_i.

Then the matrix representation of T on V with respect to the basis $\mathcal{B} = \mathcal{B}_1 \cup \mathcal{B}_2 \cup \cdots \cup \mathcal{B}_m$ is:

$$J = \begin{bmatrix} J_{e_1}(\lambda_1) & & & \\ & J_{e_2}(\lambda_2) & & \\ & & \ddots & \\ & & & J_{e_m}(\lambda_m) \end{bmatrix}$$

8. Euclidean Spaces

"The idea of representation is one of the few great ideas in Mathematics."

GUOWU MENG

Before studying Euclidean spaces, we first review tensors and then introduce inner products.

8.1 Tensor

Let V be a finite-dimensional vector space over a field $\mathbb{F}$. Then we have the following definitions.

Definition 8.1 — k-form. A *k-form* on V is a multilinear map:

$$\underbrace{V \times V \times \cdots \times V}_{k \text{ times}} \to \mathbb{F}$$

which is linear in each argument. It is an element in $(V^*)^{\otimes k}$.

More concretely, for 1-form, it is a linear functional on V, i.e. an element in V^*. It is also called *covector*. For 2-form, it is a bilinear map on V, i.e. an element in $V^* \otimes V^*$. To prove that the set of all 2-forms on V is isomorphic to $V^* \otimes V^*$, we can consider the following diagram:

$$\begin{array}{ccc} \mathsf{Map}^{\mathsf{ML}}(V \times V, \mathbb{F}) & \equiv & \mathsf{Hom}(V, V^*) \\ \vdots & & \| \\ V^* \otimes V^* & \equiv & \mathsf{Hom}(V, \mathbb{F}) \otimes V^* \end{array}$$

Remember that $\mathsf{Hom}(V_1, V_2 \otimes V_3) \equiv \mathsf{Hom}(V_1, V_2) \otimes V_3$.

Moreover, we have the following two special types of 2-forms which are the elements inside the symmetric and exterior powers of V^*.

Definition 8.2 — Symmetric and Skew-symmetric 2-forms. A 2-form $\omega : V \times V \to \mathbb{F}$ is

called *symmetric* if

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

for all $u,v \in V$. It is an element in $\mathcal{S}^2V^*$. The 2-form ω is called *skew-symmetric*, or antisymmetric, if

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

for all $u,v \in V$. It is an element in $\bigwedge^2 V^*$.

Then we define the tensor spaces.

Definition 8.3 — Tensor Spaces. Let V be a finite-dimensional vector space over a field $\mathbb{F}$. The *tensor space of type* (r,s) on V is defined as:

$$\mathcal{T}^{r,s}V = \underbrace{V \otimes V \otimes \cdots \otimes V}_{r \text{ times}} \otimes \underbrace{V^* \otimes V^* \otimes \cdots \otimes V^*}_{s \text{ times}}$$

Elements in $\mathcal{T}^{r,s}V$ are called *tensors of type* (r,s) on V, which is a mixed type if $r,s \neq 0$.

If a tensor of type $(r,0)$, then it is called a *contravariant tensor* or simply a *tensor*. If a tensor of type $(0,s)$, then it is called a *covariant tensor* or simply a *form*. For $\mathcal{T}^{0,0}V$, it is defined as $\mathbb{F}$ itself. Any elements in $\mathcal{T}^{0,0}V$ are *scalar type* tensor on V, or simply *scalars*.

Then we know that $\mathsf{End}(V) \equiv V \otimes V^* \equiv \mathcal{T}^{1,1}V$. Therefore, any endomorphism on V can be viewed as a tensor of type $(1,1)$ on V, represented by a^i_j with respect to a basis $\mathcal{B}_V = \{\vec{v}_1, \vec{v}_2, \cdots, \vec{v}_n\}$ of V. Here the upper index i represents the contravariant part and the lower index j represents the covariant part. To know that what a^i_j means, we can consider the following diagram:

$$\begin{array}{ccc} V & \xrightarrow{\quad T \quad} & V \\ {\scriptstyle [-]_{\mathcal{B}_V}}\downarrow & & \downarrow{\scriptstyle [-]_{\mathcal{B}_V}} \\ \mathbb{F}^n & \xrightarrow[A=[a^i_j]_{\mathcal{B}_V}]{} & \mathbb{F}^n \end{array}$$

Then how to get the matrix representation $A = [a^i_j]_{\mathcal{B}_V}$ of T with respect to the basis $\mathcal{B}_V$? We have:

$$\vec{a}_j = A\vec{e}_j, \qquad a^i_j = \vec{e}^T_i A \vec{e}_j = \hat{e}^i A \vec{e}_j = \langle \hat{e}^i, A\vec{e}_j \rangle.$$

So we have $[a^i_j] = \langle \hat{v}^i, T\vec{v}_j \rangle$. We can have an identification between $\mathsf{End}(V)$ and $\mathcal{T}^{1,1}V$ as follows:

$$T \leftrightarrow T\vec{v}_j \otimes \hat{v}^j$$

For covariant and contravariant, we have the following table:

Object	**Transformation Type**
Standard Basis Vector ($\vec{e}_i$)	Covariant
Dual Basis Vector ($\hat{e}^i$)	Contravariant
Component of a Vector (v^i)	Contravariant
Component Basis Vector (v_i)	Covariant

An object is considered as covariant if it transform in the same way as the basis vectors of the original vector space. If you cannot understand it, make up some examples of scaling the vector spaces.

In general, an element $t \in \mathcal{T}^{r,s}V$ can be represented as:

$$t^{i_1 i_2 \cdots i_r}_{j_1 j_2 \cdots j_s} \vec{v}_{i_1} \otimes \vec{v}_{i_2} \otimes \cdots \otimes \vec{v}_{i_r} \otimes \hat{v}^{j_1} \otimes \hat{v}^{j_2} \otimes \cdots \otimes \hat{v}^{j_s}$$

Note that the representation depends on the choice of basis $\mathcal{B}_V$ of V, i.e., the following two represents the same tensor with respect to different bases:

$$\left[t^{i_1 i_2 \cdots i_r}_{j_1 j_2 \cdots j_s}\right]_{\mathcal{B}_V} \sim \left[\tilde{t}^{\tilde{i}_1 \tilde{i}_2 \cdots \tilde{i}_r}_{\tilde{j}_1 \tilde{j}_2 \cdots \tilde{j}_s}\right]_{\widetilde{\mathcal{B}_V}}$$

The two representations are related by the base change matrices::

$$(\tilde{v}_1, \tilde{v}_2, \cdots, \tilde{v}_n) = (v_1, v_2, \cdots, v_n)A, \quad A = [a^{i}_{\tilde{j}}]^{\widetilde{\mathcal{B}_V}}_{\mathcal{B}_V} \in \mathsf{GL}(V)$$

Remark. It is actually the right action of $\mathsf{GL}(V)$ on the set of all bases of V, $\mathcal{B}_V$:

$$\mathcal{B}_V \times \mathsf{GL}(V) \to \mathcal{B}_V, \quad (v, A) \mapsto vA = \tilde{v}$$

Then we have the following equation:

$$\tilde{v}_{\tilde{j}} = v_i a^{i}_{\tilde{j}}$$

For $A^{-1} = [b^{\tilde{i}}_{j}]^{\mathcal{B}_V}_{\widetilde{\mathcal{B}_V}}$, we have $a^{i}_{\tilde{j}} b^{\tilde{k}}_{j} = \delta^{\tilde{k}}_{\tilde{j}}$ and $b^{\tilde{i}}_{j} a^{j}_{\tilde{k}} = \delta^{\tilde{i}}_{\tilde{k}}$. Therefore, we have:

$$v_k = \tilde{v}_{\tilde{j}} b^{\tilde{j}}_{k}$$

Remark. For easier memorisation, we use the calculus operators:

$$\frac{\partial \tilde{v}_{\tilde{j}}}{\partial v_i} = a^{i}_{\tilde{j}}, \quad \frac{\partial v_k}{\partial \tilde{v}_{\tilde{j}}} = b^{\tilde{j}}_{k}$$

To memorise it, we consider the lower indices in denominators (lower) will flip to the upper indices in numerators. (As lower twice, so flip to upper)

Then we can use the chain rule to verify the two equations of A and A^{-1}:

$$\frac{\partial \tilde{v}_{\tilde{j}}}{\partial v_i} \frac{\partial v_k}{\partial \tilde{v}_{\tilde{j}}} = \delta^{i}_{k}$$

Then we have the transformation rule for the representation of $t \in \mathcal{T}^{r,s}V$ under the base change from $\mathcal{B}_V$ to $\widetilde{\mathcal{B}_V}$:

$$\tilde{t}^{\tilde{i}_1 \tilde{i}_2 \cdots \tilde{i}_r}_{\tilde{j}_1 \tilde{j}_2 \cdots \tilde{j}_s} = \left(b^{\tilde{i}_1}_{i_1} b^{\tilde{i}_2}_{i_2} \cdots b^{\tilde{i}_r}_{i_r}\right) t^{i_1 i_2 \cdots i_r}_{j_1 j_2 \cdots j_s} \left(a^{j_1}_{\tilde{j}_1} a^{j_2}_{\tilde{j}_2} \cdots a^{j_s}_{\tilde{j}_s}\right)$$

Given that $\mathcal{B}_V = \{\vec{v}_1, \cdots, \vec{v}_n\}$ is a basis of V, then we can define a basis of $\mathcal{T}^{r,s}V$ as follows:

$$\mathcal{B}_{\mathcal{T}^{r,s}V} = \{\vec{v}_{i_1} \otimes \vec{v}_{i_2} \otimes \cdots \otimes \vec{v}_{i_r} \otimes \hat{v}^{j_1} \otimes \hat{v}^{j_2} \otimes \cdots \otimes \hat{v}^{j_s} : 1 \le i_1, i_2, \cdots, i_r, j_1, j_2, \cdots, j_s \le n\}$$

Then for symmetric and skew-symmetric k-forms, we have:

$$\mathcal{B}_{\mathcal{S}^k V} = \{\vec{v}_{i_1} \vec{v}_{i_2} \cdots \vec{v}_{i_k} : 1 \le i_1, i_2, \cdots, i_k \le n\}$$
$$\mathcal{B}_{\bigwedge^k V} = \{\vec{v}_{i_1} \wedge \vec{v}_{i_2} \wedge \cdots \wedge \vec{v}_{i_k} : 1 \le i_1, i_2, \cdots, i_k \le n\}$$

Then "honest" definition of symmetric basis is:

$$\{\vec{v}_{i_1} \vec{v}_{i_2} \cdots \vec{v}_{i_k} : 1 \le i_1 \le i_2 \le \cdots \le i_k \le n\}$$

but it is redundant. We just have to make sure that the representation of any symmetric k-form is unique for a given basis. For example, in 2-form case with the basis $\{\vec{e}_i \otimes \vec{e}_j\}$, we originally have to write:

$$t = \sum_{1 \le i \le j \le n} t_{ij} \vec{e}_i \otimes \vec{e}_j$$

but this is ugly, so we just write:

$$t = t^{ij} \vec{e}_i \vec{e}_j$$

with $t^{ij} = t^{ji}$. If we ignored the condition on t^{ij}, then we have $a^{ij} = -a^{ji}$ such that:

$$t = t^{ij} \vec{e}_i \wedge \vec{e}_j + a^{ij} \vec{e}_i \wedge \vec{e}_j = (t^{ij} + a^{ij}) \vec{e}_i \wedge \vec{e}_j = 0$$

As $a^{ij} = a^{ji} = -a^{ij}$.

Then for skew-symmetric basis, let say $t \in \mathcal{B}_{\bigwedge^k V}$, then we have:

$$t = t^{\mathcal{I}} \vec{v}_{\mathcal{I}} = t^{i_1 i_2 \cdots i_k} \vec{v}_{i_1} \wedge \vec{v}_{i_2} \wedge \cdots \wedge \vec{v}_{i_k}$$

with $\mathcal{I} = (i_1, i_2, \cdots, i_k)$ being an ordered index set with $1 \le i_1 < i_2 < \cdots < i_k \le n$. Then for any permutation $\sigma \in S_k$, to make sure it is unique, we require:

$$t^{\sigma(\mathcal{I})} = \mathsf{Sgn}(\sigma) t^{\mathcal{I}}$$

where $\sigma(\mathcal{I}) = (i_{\sigma(1)}, i_{\sigma(2)}, \cdots, i_{\sigma(k)})$.

In conclusion, we have to make sure that the representation of any symmetric or skew-symmetric k-form is unique for a given basis by the following conditions respectively:

Symmetric: $t^{i_1 i_2 \cdots i_k} = t^{i_{\sigma(1)} i_{\sigma(2)} \cdots i_{\sigma(k)}}$

Skew-symmetric: $t^{i_1 i_2 \cdots i_k} = \mathsf{Sgn}(\sigma) t^{i_{\sigma(1)} i_{\sigma(2)} \cdots i_{\sigma(k)}}$

8.2 Inner Product

Let V be a finite-dimensional real linear space. Then we have the following definitions.

Definition 8.4 — Inner Product. An inner product on V is a map $\langle -,- \rangle : V \times V \to \mathbb{R}$ such that
1. *Bilinearity:* $\langle -, u \rangle$ and $\langle u, - \rangle$ are linear functionals on V for all $u \in V$;
2. *Symmetry:* $\langle u, v \rangle = \langle v, u \rangle$ for all $u, v \in V$;
3. *Positive-definiteness:* $\langle v, v \rangle \geq 0$ for all $v \in V$ with equality if and only if $v = 0$.

Note that an inner product on V is a positive-definite symmetric 2-form on V.

Definition 8.5 — Pseudo Inner Product. A pseudo inner product on V is a non-degenerate symmetric bilinear form on V, i.e., an element $\langle -,- \rangle \in \mathcal{S}^2 V^*$ such that $\langle -,- \rangle\flat : V \to V^*$ is isomorphic.

Then a real linear space V with an inner product $\langle -,- \rangle$ is called a *Euclidean space*, denoted by $(V, \langle -,- \rangle)$.

Definition 8.6 — Metric Space. A metric space is a non-empty set X together with a metric structure, i.e., a distance function $d : X \times X \to \mathbb{R}$ that sends (x, y) to $d(x, y)$ such that
1. *Positivity:* $d(x, y) \geq 0$ for all $x, y \in X$ with equality if and only if $x = y$;
2. *Symmetry:* $d(x, y) = d(y, x)$ for all $x, y \in X$;
3. *Triangle Inequality:* $d(x, z) \leq d(x, y) + d(y, z)$ for all $x, y, z \in X$.

If we want to combine the metric structure with the linear structure on V, we have to make sure that the distance function $d : V \times V \to \mathbb{R}$ satisfies the two additional properties in order to be compatible with the linear structure. We would say the properties are *harmonic* with the linear structure.

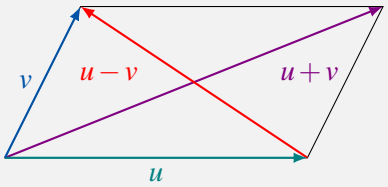
Definition 8.7 — Normed Linear Space. A real normed linear space is a real linear space V together with a compatible metric structure or a normed structure, i.e., a distance function $d : V \times V \to \mathbb{R}$ such that
1. *Translation Invariance:* $d(u + w, v + w) = d(u, v)$ for all $u, v, w \in V$;
2. *Homogeneity:* $d(\alpha u, \alpha v) = |\alpha| d(u, v)$ for all $u, v \in V$ and $\alpha \in \mathbb{R}$.

Then we can define the norm on V as $\|v\| = d(v, 0)$ for all $v \in V$.

Then a function $\| - \| : V \to \mathbb{R}$ that sends v to $\|v\|$ is called a norm on V if it satisfies:
1. *Positive-definiteness:* $\|v\| \geq 0$ for all $v \in V$ with equality if and only if $v = 0$;
2. *Homogeneity:* $\|\alpha v\| = |\alpha| \|v\|$ for all $v \in V$ and $\alpha \in \mathbb{R}$;
3. *Triangle Inequality:* $\|u + v\| \leq \|u\| + \|v\|$ for all $u, v \in V$.

We can use the norm with the properties above to define the distance function by $d(x, y) = \|x - y\|$.

Theorem 8.1 — Parallelogram Law. The parallelogram law states that the sum of squares of the lengths of the four sides of a parallelogram equals the sum of squares of the lengths of the two diagonals, i.e., with the following figure:



we have:

$$\|u + v\|^2 + \|u - v\|^2 = 2\|u\|^2 + 2\|v\|^2$$

Proposition 8.1 An inner product on V is equivalence to a norm structure on V which satisfies the parallelogram law.

Proof. ($\Rightarrow$) Let $(V, \langle -, - \rangle)$ be a Euclidean space. Then we can define the norm on V as $\|v\| = \sqrt{\langle v, v \rangle}$ for all $v \in V$. Then we have:

1. *Positive-definiteness:* $\|v\| = \sqrt{\langle v, v \rangle} \geq 0$ for all $v \in V$ with equality if and only if $v = 0$;
2. *Homogeneity:* $\|\alpha v\| = \sqrt{\langle \alpha v, \alpha v \rangle} = \sqrt{\alpha^2 \langle v, v \rangle} = |\alpha| \|v\|$ for all $v \in V$ and $\alpha \in \mathbb{R}$;
3. *Triangle Inequality:* By Cauchy-Schwarz inequality, we have:

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

   Therefore, the triangle inequality holds.
4. *Parallelogram Law:* We have:

$$\begin{aligned} \|u+v\|^2 + \|u-v\|^2 &= \langle u+v, u+v \rangle + \langle u-v, u-v \rangle \\ &= \langle u, u \rangle + \langle v, v \rangle + \langle u, v \rangle + \langle v, u \rangle + \langle u, u \rangle + \langle v, v \rangle - \langle u, v \rangle - \langle v, u \rangle \\ &= 2\langle u, u \rangle + 2\langle v, v \rangle = 2\|u\|^2 + 2\|v\|^2 \end{aligned}$$

($\Leftarrow$) We define the inner product for all $u, v \in V$ as follows:

$$\langle u, v \rangle = \frac{1}{2} \left(\|u+v\|^2 - \|u\|^2 - \|v\|^2 \right)$$

Then we check the three properties of inner product:

1. *Bilinearity:* For all $u, v, w \in V$, we have to show that $\langle u+w, v \rangle = \langle u, v \rangle + \langle w, v \rangle$, which is equivalent to show that:

$$\begin{aligned} & \|u+w+v\|^2 - \|u+w\|^2 - \|v\|^2 = \|u+v\|^2 - \|u\|^2 - \|v\|^2 + \|w+v\|^2 - \|w\|^2 - \|v\|^2 \\ \iff & \|u+w+v\|^2 + \|u\|^2 + \|w\|^2 + \|v\|^2 = \|u+w\|^2 + \|u+v\|^2 + \|w+v\|^2 \end{aligned}$$

   Then we may consider $x = u+w$ and $y = v+w$, and $x' = u+v+w$ and $y' = w$, and we have

$$\begin{aligned} \|u+v+2w\|^2 + \|u-v\|^2 &= 2\|u+w\|^2 + 2\|v+w\|^2 \\ \|u+v+2w\|^2 + \|u+v\|^2 &= 2\|u+v+w\|^2 + 2\|w\|^2 \end{aligned}$$

   Then we have

$$\|u-v\|^2 - \|u+v\|^2 = 2\|u+w\|^2 + 2\|v+w\|^2 - 2\|u+v+w\|^2 - 2\|w\|^2$$

   Moreover, by the parallelogram law on $u-v$, we have

$$\begin{aligned} & 2\|u\|^2 + 2\|v\|^2 - 2\|u+v\|^2 = 2\|u+w\|^2 + 2\|v+w\|^2 - 2\|u+v+w\|^2 - 2\|w\|^2 \\ \iff & \|u+v+w\|^2 + \|u\|^2 + \|v\|^2 + \|w\|^2 = \|u+w\|^2 + \|v+w\|^2 + \|u+v\|^2 \end{aligned}$$

   Hence, additivity in the first argument holds. We can show the additivity in the second argument similarly. For homogeneity, we may consider the following steps:
   - Prove natural number homogeneity
   - Prove reciprocal of natural number homogeneity

- Prove Cauchy-Schwarz inequality
- Prove that for any $\lambda \in \mathbb{R}$, every $r \in \mathbb{Q}$, we have:

$$|\lambda\langle u,v\rangle - \langle \lambda u,v\rangle| = |(\lambda - r)\langle u,v\rangle - \langle (\lambda - r)u,v\rangle| \leq 2|\lambda - r|\|u\|\|v\|$$

- Hence, prove real number homogeneity by taking limit on both sides as $r \to \lambda$.

2. *Symmetry:* For all $u, v \in V$, we have:

$$\begin{aligned}\langle u,v\rangle &= \frac{1}{2}\left(\|u+v\|^2 - \|u\|^2 - \|v\|^2\right) \\ &= \frac{1}{2}\left(\|v+u\|^2 - \|v\|^2 - \|u\|^2\right) = \langle v,u\rangle\end{aligned}$$

3. *Positive-definiteness:* For all $v \in V$, we have:

$$\langle v,v\rangle = \frac{1}{2}\left(\|v+v\|^2 - \|v\|^2 - \|v\|^2\right) = \frac{1}{2}(4\|v\|^2 - 2\|v\|^2) = \|v\|^2 \geq 0$$

Thus, $\langle -,-\rangle$ is an inner product on V. ■

Theorem 8.2 — Cauchy-Schwarz Inequality. Let $(V, \langle -,-\rangle)$ be a Euclidean space. Then for all $u, v \in V$, we have:

$$|\langle u,v\rangle| \leq \|u\|\|v\|$$

with equality if and only if u and v are linearly dependent.

Proof. Let $f(t) = \|tu+v\|^2 = \langle tu+v, tu+v\rangle = t^2\|u\|^2 + 2t\langle u,v\rangle + \|v\|^2 \geq 0$ for all $t \in \mathbb{R}$. Then we have $f(t) \geq 0$ for all $t \in \mathbb{R}$. For $u = 0$, the inequality holds trivially. For $u \neq 0$, the quadratic function $f(t)$ has at most one real root, so its discriminant is less than or equal to zero:

$$\Delta = 4\langle u,v\rangle^2 - 4\|u\|^2\|v\|^2 \leq 0 \implies \langle u,v\rangle^2 \leq \|u\|^2\|v\|^2$$

■

Definition 8.8 If both $u, v \in V$ are non-zero vectors in a Euclidean space $(V, \langle -,-\rangle)$, then the angle θ between u and v is defined as:

$$\theta = \arccos\left(\frac{\langle u,v\rangle}{\|u\|\|v\|}\right)$$

Moreover, if $\langle u,v\rangle = 0$, then we say that u and v are orthogonal.

8.3 Orthogonality

Let V be a Euclidean space with inner product $\langle -,- \rangle$ and $W \subseteq V$ is a subspace of V. Then we claim that W inherits an Euclidean structure from $\langle -,- \rangle$ in V. We can simply restrict the inner product $\langle -,- \rangle$ on V to W:

$$\begin{array}{ccccc} W \times W & \hookrightarrow & V \times V & \xrightarrow{\langle -,- \rangle} & \mathbb{R} \\ & \searrow & & \nearrow & \\ & & \langle -,- \rangle & & \end{array}$$

Note that the restriction $\langle -,- \rangle$ is still an inner product on W. Also, the positive-definiteness of $\langle -,- \rangle$ implies that $\langle -,- \rangle$ is non-degenerate, i.e., the map $\langle -,- \rangle^\natural : W \to W^*$ is isomorphism. Note that W and W^* have the same dimension and it has a trivial kernel: $\langle u, - \rangle_W = 0$ implies $\langle u,u \rangle_W = 0$ implies $u = 0$. Now, suppose $w = (w_1, \cdots, w_k)$ is a basis of W and $w^* = (w_1^*, \cdots, w_k^*)$ is the dual basis of W^*, then we have the following diagram:

$$\begin{array}{ccccccccc} 0 & \longrightarrow & \mathsf{Ker}(\lambda_w) & \longrightarrow & V & \underset{\lambda_w}{\overset{s}{\rightleftarrows}} & \mathbb{R}^k & \longrightarrow & 0 \\ & & & & \uparrow & & \uparrow{\scriptstyle [-]_{w^*}} & & \\ & & & & W & \xrightarrow{\langle -,- \rangle^\natural} & W^* & & \\ & & & & w_i & \longmapsto & \langle w_i, - \rangle & & \end{array}$$

where $\lambda_w = \begin{bmatrix} \langle w_1, - \rangle \\ \vdots \\ \langle w_k, - \rangle \end{bmatrix}$, and s is a section of λ_w with image W. Then we have the decomposition:

$$V = \mathsf{Im}(s) \oplus \mathsf{Ker}(\lambda_w) = W \oplus \mathsf{Ker}(\lambda_w)$$

Note that it is an internal direct sum. Then we define the orthogonal complement of W in V as follows.

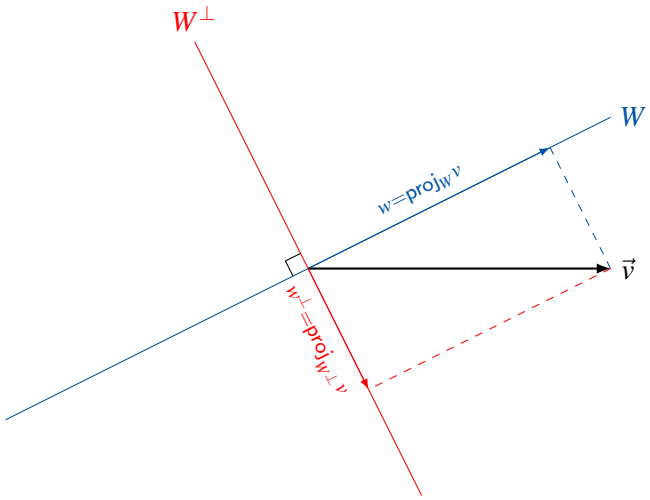
Definition 8.9 — Orthogonal Complement. The orthogonal complement of W in V, denoted by $W^\perp$, is defined as:

$$W^\perp = \{v \in V \mid \langle v, w \rangle = 0 \text{ for all } w \in W\} = \{v \in V \mid \langle v, w_i \rangle = 0 \text{ for all basis } w_i \in W\}$$

Then we have the decomposition:

$$V = W \oplus W^\perp$$

Then any vector $v \in V$ can be uniquely decomposed as $v = w + w^\perp$ with $w = \mathsf{proj}_W(v) \in W$ and $w^\perp = \mathsf{proj}_{W^\perp}(v) \in W^\perp$. The map $\mathsf{proj}_W : V \to W$ is called the orthogonal projection onto W along $W^\perp$. Take a look at the following figure:



Then we have the following properties of the orthogonal projection:

1. $(\text{proj}_W)^2 = \text{proj}_W$;
2. $\text{Im}(\text{proj}_W) = W$;
3. $\text{Ker}(\text{proj}_W) = W^\perp$;
4. $\text{proj}_W + \text{proj}_{W^\perp} = \text{id}_V$.

Definition 8.10 — Orthonormal Basis. A basis v is orthogonal if $\langle v_i, v_j \rangle = 0$ for all $i \neq j$. An orthogonal basis is orthonormal if $\|v_i\| = 1$ for all i.

Then we have the following proposition.

Proposition 8.2 For any Euclidean space V with inner product, there exists an orthonormal basis of V. Moreover, there exists a linear isometric isomorphism between V and $\mathbb{R}^n$ with the standard inner product, the dot product.

Note that $(\mathbb{R}^n, \cdot)$ is up to isomorphism the only Euclidean space with dimension n, where $\cdot$ denotes the standard dot product.

Moreover, if $w = (w_1, w_2, \cdots, w_k)$ is an orthonormal basis of W, then

$$\text{proj}_W u = \sum_{i=1}^{k} \langle w_i, u \rangle w_i$$

for all $u \in V$. In case w is orthogonal but not orthonormal, then we have:

$$\text{proj}_W u = \sum_{i=1}^{k} \frac{\langle w_i, u \rangle}{\langle w_i, w_i \rangle} w_i$$

8.4 Gram-Schmidt Process

Let $w = (w_1, w_2, \cdots, w_k)$ be an orthonormal basis of $W \subseteq V$. Then we have:

$$x = \underbrace{\sum_{i=1}^{k} \langle w_i, x\rangle w_i}_{\in W} + \underbrace{x - \sum_{i=1}^{k} \langle w_i, x\rangle w_i}_{\in W^\perp} = \mathsf{proj}_W x + \mathsf{proj}_{W^\perp} x.$$

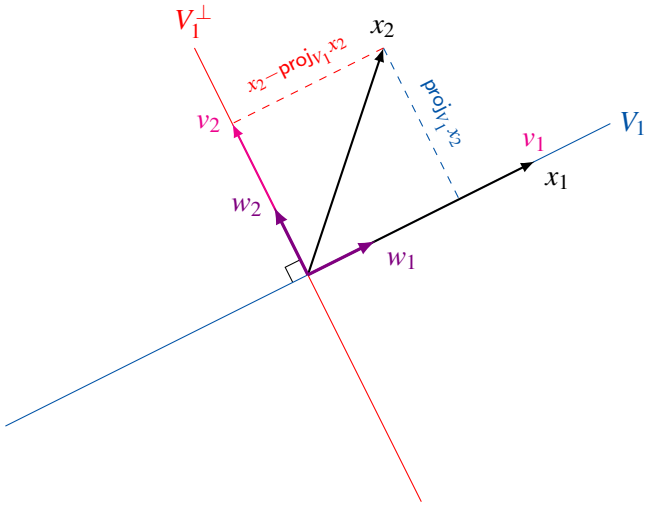
To show that $\mathsf{proj}_{W^\perp} x \in W^\perp$, it suffices to show that $\langle w_j, \mathsf{proj}_{W^\perp} x\rangle = 0$ for all $1 \leq j \leq k$:

$$\begin{aligned}
\langle w_j, \mathsf{proj}_{W^\perp} x\rangle &= \langle w_j, x - \sum_{i=1}^{k} \langle w_i, x\rangle w_i\rangle \\
&= \langle w_j, x\rangle - \sum_{i=1}^{k} \langle w_i, x\rangle \langle w_j, w_i\rangle \\
&= \langle w_j, x\rangle - \langle w_j, x\rangle = 0
\end{aligned}$$

Note that the key step is to use the bilinearity of the inner product and the orthonormality of w.

Now, given any basis $x = (x_1, x_2, \cdots, x_n)$ of V, we can use the Gram-Schmidt process to construct an orthonormal basis $w = (w_1, w_2, \cdots, w_n)$ of V by inductive argument. The idea is: We have $V_n \supset V_{n-1} \supset \cdots \supset V_2 \supset V_1 \supset V_0 = \{0\}$ with the dimension $n, n-1, \cdots, 2, 1, 0$ respectively. Then we have w_1 as the orthonormal basis of V_1, then we can extend it to w_1, w_2 as the orthonormal basis of V_2, and so on and so forth until we reach $V_n = V$.

Then we consider the first two cases to illustrate the idea. Let $v_1 = u_1$. Then we have $w_1 = \frac{v_1}{\|v_1\|}$ as the orthonormal basis of $V_1 = \mathsf{Span}\{u_1\}$. Then we want to find the w_2 such that w_1, w_2 is the orthonormal basis of $V_2 = \mathsf{Span}\{u_1, u_2\}$. We can consider the following diagram:



Then $v_2 = x_2 - \mathsf{proj}_{V_1} x_2 = x_2 - \langle w_1, x_2\rangle w_1$ is orthogonal to w_1. Note that w_1 is normalised. Then we can normalise v_2 to get $w_2 = \frac{v_2}{\|v_2\|}$. Therefore, w_1, w_2 is the orthonormal basis of V_2. Then for general k-th step, we have:

$$v_k = x_k - \sum_{i=1}^{k-1} \langle w_i, x_k\rangle w_i = x_k - \sum_{i=1}^{k-1} \frac{\langle v_i, x_k\rangle}{\langle v_i, v_i\rangle} v_i, \quad w_k = \frac{v_k}{\|v_k\|}$$

given that $w_1, w_2, \cdots, w_{k-1}$ is the orthonormal basis of $V_{k-1} = \mathsf{Span}\{x_1, x_2, \cdots, x_{k-1}\}$ and the orthogonal basis of V_{k-1}, $v_1, v_2, \cdots, v_{k-1}$.

Then there is a useful corollary of the Gram-Schmidt process, the *QR* Decomposition.

Let V be a Euclidean space. We can interpret it as $(\mathbb{R}^n, \cdot)$ up to isomorphism. Then we have a basis $(\vec{x}_1, \vec{x}_2, \cdots, \vec{x}_n)$ of V and we can form an invertible matrix A whose columns are the vectors $\vec{x}_1, \vec{x}_2, \cdots, \vec{x}_n$, i.e.,

$$A = \begin{bmatrix} | & | & & | \\ \vec{x}_1 & \vec{x}_2 & \cdots & \vec{x}_n \\ | & | & & | \end{bmatrix}$$

Then we have an orthogonal basis $(\vec{v}_1, \cdots, \vec{v}_n)$ of V and an orthonormal basis $(\vec{w}_1, \cdots, \vec{w}_n)$ obtained by the Gram-Schmidt process. Then we should have an invertible matrix to convert between bases. Then what is the matrix to convert from the original basis to the orthonormal basis?

Note that each $\vec{x}_k$ can be expressed as a linear combination of $\vec{w}_1, \cdots, \vec{w}_k$:

$$\vec{x}_k = \vec{v}_k + \sum_{i=1}^{k-1} \frac{\langle \vec{v}_i, \vec{x}_k \rangle}{\langle \vec{v}_i, \vec{v}_i \rangle} \vec{v}_i = \|\vec{v}_k\| \vec{w}_k + \sum_{i=1}^{k-1} \langle \vec{w}_i, \vec{x}_k \rangle \vec{w}_i$$

Also, we can express $\vec{x}_k$ as follows:

$$\vec{x}_k = \begin{bmatrix} | & | & & | \\ \vec{w}_1 & \vec{w}_2 & \cdots & \vec{w}_n \\ | & | & & | \end{bmatrix} \begin{bmatrix} \langle \vec{w}_1, \vec{x}_k \rangle \\ \langle \vec{w}_2, \vec{x}_k \rangle \\ \vdots \\ \langle \vec{w}_{k-1}, \vec{x}_k \rangle \\ \|\vec{v}_k\| \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Then we have the matrix equation:

$$\underbrace{\begin{bmatrix} | & | & & | \\ \vec{x}_1 & \vec{x}_2 & \cdots & \vec{x}_n \\ | & | & & | \end{bmatrix}}_{A} = \underbrace{\begin{bmatrix} | & | & & | \\ \vec{w}_1 & \vec{w}_2 & \cdots & \vec{w}_n \\ | & | & & | \end{bmatrix}}_{Q} \underbrace{\begin{bmatrix} \langle \vec{w}_1, \vec{x}_1 \rangle & \langle \vec{w}_1, \vec{x}_2 \rangle & \cdots & \langle \vec{w}_1, \vec{x}_n \rangle \\ 0 & \langle \vec{w}_2, \vec{x}_2 \rangle & \cdots & \langle \vec{w}_2, \vec{x}_n \rangle \\ 0 & 0 & \ddots & \vdots \\ 0 & 0 & 0 & \langle \vec{w}_n, \vec{x}_n \rangle \end{bmatrix}}_{R}$$

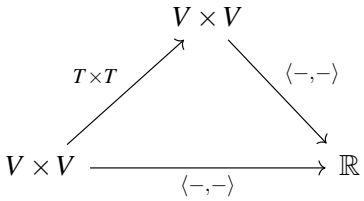
which is called the *QR Decomposition* of A, where Q is an orthogonal matrix and R is an upper-triangular matrix with positive diagonal entries. However, normally we denote the orthogonal matrix by O instead of Q and an upper-triangular matrix by U instead of R.

8.5 Orthogonal Group and Special Orthogonal Group

Let V be a Euclidean space with inner product $\langle -,- \rangle$. Then we view V as a linear space, and we have $\mathsf{Aut}(V) = \mathsf{GL}(V)$. If we view V as a Euclidean space, then we have $\mathsf{Aut}(V) = \mathsf{O}(V) \subseteq \mathsf{GL}(V)$, where $\mathsf{O}(V)$ is the subgroup of $\mathsf{GL}(V)$ that respects the Euclidean structure, i.e., for all $T \in \mathsf{O}(V)$, we have:

$$\langle T(u), T(v) \rangle = \langle u, v \rangle$$

for all $u, v \in V$, so length and angles are preserved under T. Or equivalently, the following diagram commutes:



We can also define the orthogonal group $\mathsf{O}(n)$ using this property. Let $V = \mathbb{R}^n$ with the dot product. Then for any $A \in \mathsf{GL}_n(\mathbb{R})$, $A \in \mathsf{O}(n)$ if and only if A satisfies:

$$\langle \vec{a}_i, \vec{a}_j \rangle = \langle A\vec{e}_i, A\vec{e}_j \rangle = \langle \vec{e}_i, \vec{e}_j \rangle = \delta_{ij}$$

It is equivalent to say that $A^T A = I_n$, i.e., $A^T = A^{-1}$. Therefore, we have:

$$\mathsf{O}(n) = \{A \in \mathsf{GL}_n(\mathbb{R}) \mid A^T A = I_n\}$$

Note that $\det(A^T) = \det(A)^T = \det(A)$. Therefore, we have $\det(A)^2 = 1$ for all $A \in \mathsf{O}(n)$, i.e., $\det(A) = \pm 1$.

Then consider the following exact sequence:

$$1 \longrightarrow \mathsf{SL}(V) \hookrightarrow \mathsf{GL}(V) \xrightarrow{\det} \mathbb{R}^\times \longrightarrow 1$$

where $\mathbb{R}^\times = \mathsf{GL}_1(\mathbb{R}) = \mathbb{R} \setminus \{0\}$ is the multiplicative group of non-zero real numbers. As for any automorphism $A \in \mathsf{GL}(V)$, we have a determinant $\det A \in \mathbb{R}^\times$, which is surjective. $\mathsf{SL}(V)$ is defined as the kernel of the determinant map, i.e., $\mathsf{SL}(V) = \{A \in \mathsf{GL}(V) \mid \det A = 1\}$.

Similarly, we have the special orthogonal group $\mathsf{SO}(V)$ as the subgroup of $\mathsf{O}(V)$ with determinant 1:

$$\mathsf{SO}(V) = \{A \in \mathsf{O}(V) \mid \det A = 1\}$$

8.6 Matrix Representation of Inner Products

Let V be a Euclidean space with inner product $\langle -, - \rangle$. Then we can choose a basis $v = (v_1, v_2, \cdots, v_n)$ of V. Then we have

$$x = x^i v_i = \begin{bmatrix} x^1 \\ x^2 \\ \vdots \\ x^n \end{bmatrix}, \quad y = y^i v_i = \begin{bmatrix} y^1 \\ y^2 \\ \vdots \\ y^n \end{bmatrix}$$

Then the inner product $\langle x, y \rangle$ can be represented as:

$$\langle x, y \rangle = x^i y^j \langle v_i, v_j \rangle = x^T \omega y = x \cdot (\omega y)$$

where we let $\omega = [\langle v_i, v_j \rangle]$ be the matrix representation of the inner product with respect to the basis v. Then $\langle -, - \rangle = \cdot \omega -$. To find the canonical form of the inner product, we left it to the next chapter.

Proposition 8.3 — Spectral Theorem for Real Symmetric Matrices. Let A be a $n \times n$ real symmetric matrix. Then there exists an orthogonal matrix O and a diagonal matrix D such that:

$$A = ODO^{-1} = ODO^T$$

where the entries of D are the eigenvalues of A. Or equivalently, there exists an orthonormal basis of $\mathbb{R}^n$ consisting of eigenvectors of A.

To prove this proposition, we would use the result in Hermitian spaces, so we leave the proof to the next chapter.

9. Hermitian Spaces

"In Mathematics, one of the great ideas is anytime you are interested in vector space over real numbers, but real number is not as nice as complex numbers. So you should turn the problem into complex case, then use the result there to do it in real case."

GUOWU MENG

9.1 Hermitian Forms and Unitary Groups

9.1.1 Hermitian Forms

Similar to the definitions in Euclidean spaces, we can define Hermitian forms and Hermitian spaces as follows.

Definition 9.1 — Hermitian Form. Let V be a complex vector space. A *Hermitian form* or *Hermitian product* on V is a map $\langle -,- \rangle : V \times V \to \mathbb{C}$ such that the following properties hold:

1. *Sesquilinearity:* For all $u, v \in V$ and $\alpha \in \mathbb{C}$, we have:
   a. Biadditivity
   b. $\langle u, \alpha v \rangle = \alpha \langle u, v \rangle$
   c. $\langle \alpha u, v \rangle = \overline{\alpha} \langle u, v \rangle$
2. *Conjugate Symmetry:* For all $u, v \in V$, we have:
   $$\langle u, v \rangle = \overline{\langle v, u \rangle} = \langle u, v \rangle^\dagger$$
   The dagger symbol † is defined as $\langle u, v \rangle^\dagger = \overline{\langle v, u \rangle}$.
3. *Positive-Definiteness:* For all $v \in V$, we have:
   $$\langle v, v \rangle \geq 0$$

When the positive-definiteness property becomes non-degeneracy, i.e., $\langle v, v \rangle = 0$ implies $v = 0$, then the Hermitian form is called a *pseudo Hermitian form*.

We can also define the norm of a vector $v \in V$ as:

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

The other four properties of norm is the same as in Euclidean spaces. Moreover the Cauchy-Schwarz inequality is as follows:

$$|\langle u, v \rangle| \leq \|u\| \|v\|$$

for all $u, v \in V$, with equality if and only if u and v are linearly dependent.

Proof. Let $f(t) = \|tu+v\|^2 = \langle tu+v, tu+v \rangle = t^2\|u\|^2 + 2\Re(\langle u, v \rangle)t + \|v\|^2 \geq 0$ for all $t \in \mathbb{R}$. Then we have $f(t) \geq 0$ for all $t \in \mathbb{R}$. For $u = 0$, the inequality holds trivially. For $u \neq 0$, the quadratic function $f(t)$ has at most one real root, so its discriminant is less than or equal to zero:

$$\Delta = 4(\Re(\langle u, v \rangle))^2 - 4\|u\|^2\|v\|^2 \leq 0 \implies (\Re(\langle u, v \rangle))^2 \leq \|u\|^2\|v\|^2 \implies |\Re(\langle u, v \rangle)| \leq \|u\|\|v\|$$

Note that $\langle u, v \rangle = |\langle u, v \rangle| e^{i\theta}$ for some $\theta \in \mathbb{R}$. Then we have:

$$\langle e^{-i\theta} u, v \rangle = e^{-i\theta} \langle u, v \rangle = |\langle u, v \rangle|$$

Therefore, we have:

$$|\langle u, v \rangle| = |\Re(\langle e^{-i\theta} u, v \rangle)| \leq \|e^{-i\theta} u\| \|v\| = \|u\| \|v\|$$

■

The sesquilinear map $\langle -, - \rangle$ can be defined as a bilinear map $\overline{V} \times V \to \mathbb{C}$, where $\overline{V}$ is the complex conjugate vector space of V, or linear map $\overline{V} \otimes V \to \mathbb{C}$. The complex conjuage vector space $\overline{V}$ is defined as the same set as V with the same addition operation, but the scalar multiplication is defined as:

$$\mathbb{C} \times \overline{V} \to \overline{V}, \quad (\alpha, v) \mapsto \overline{\alpha} v$$

Then we have the following examples:

■ **Example 9.1** We define the standard Hermitian form on $\mathbb{C}^n$ as:

$$\langle \vec{u}, \vec{v} \rangle = \vec{u}^\dagger \vec{v} = \overline{\vec{u}}^T \vec{v}$$

for all $\vec{u}, \vec{v} \in \mathbb{C}^n$. It is straightforward to verify that it satisfies all the properties of Hermitian forms. For example, the positive-definiteness property holds since:

$$\vec{u}^\dagger \vec{u} = \sum_{i=1}^{n} \overline{u_i} u_i = \sum_{i=1}^{n} |u_i|^2 \geq 0$$

■

Then a complex linear space V with an Hermitian form $\langle -, - \rangle$ is called a *Hermitian space*. Also, the model / standard Hermitian space is $(\mathbb{C}^n, \langle -, - \rangle)$ with the standard Hermitian form, that is, the inner product defined above.

Let V be a Hermitian space with Hermitian form $\langle -, - \rangle$. Then we say $u, v \in V$ are orthogonal if $\langle u, v \rangle = 0$. Similar to the Euclidean case, we can define orthogonal complement, orthogonal projection, orthonormal basis, and Gram-Schmidt process in Hermitian spaces. We also have the decomposition $V = W^\perp \oplus W$ for any subspace $W \subseteq V$.

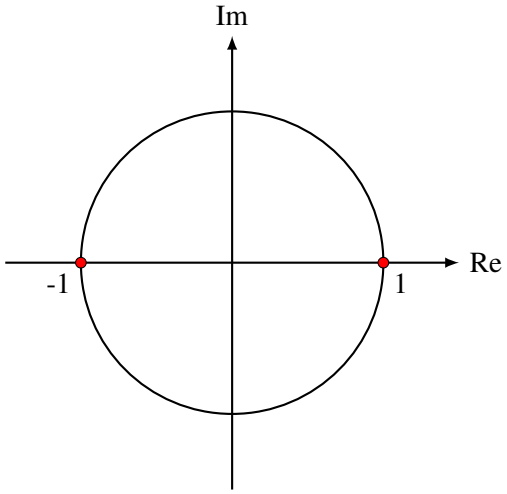
Similarly, there is only one Hermitian space up to isomorphism with dimension n, that is, $(\mathbb{C}^n, \langle -, - \rangle)$ with the standard Hermitian form, i.e., for any Hermitian space V with dimension n, there exists a linear isometric isomorphism between V and $(\mathbb{C}^n, \langle -, - \rangle)$.

9.1.2 Unitary Groups

Similar to the orthogonal groups in Euclidean spaces, we can define unitary groups in Hermitian spaces as the automorphism groups that respect the Hermitian structure. Then we have

$$\mathsf{U}(n) = \{A \in \mathsf{GL}_n(\mathbb{C}) \mid A^\dagger A = I\}$$

where $A^\dagger = \overline{A}^T$ is the conjugate transpose of A. Note that $\det(A^\dagger) = \overline{\det(A)}$. Therefore, we have $|\det(A)|^2 = 1$ for all $A \in \mathsf{U}(V)$, i.e., $|\det(A)| = 1$. This means $\mathsf{U}(1) = \{z \in \mathbb{C} \mid |z| = 1\}$ is the unit circle in the complex plane. Graphically we have:



where the unit circle represents $\mathsf{U}(1)$ in the complex plane. Also in orthogonal group, the determinant of any orthogonal matrix is either 1 or -1. This is the special case of unitary group when the entries are real numbers. Also we have the special unitary group $\mathsf{SU}(n)$ as the subgroup of $\mathsf{U}(n)$ with determinant 1.

Then we have the following definition similar to orthogonal matrices:

Definition 9.2 — Unitary Matrix. A matrix $A \in \mathsf{GL}_n(\mathbb{C})$ is called a *unitary matrix* if $A^\dagger A = I_n$, i.e., $A^{-1} = A^\dagger$.

Using similar Gram-Schmidt process in Euclidean spaces, we get the following QR decomposition in Hermitian spaces:

$$A = QR$$

where Q is a unitary matrix and R is an upper-triangular matrix with positive real diagonal entries. However, normally we denote the unitary matrix by U instead of Q. One reason why others use QR instead is to distinguish the same notation on unitary and upper-triangular matrices in Hermitian spaces and orthogonal and upper-triangular matrices in Euclidean spaces.

9.1.3 Matrix representation of Hermitian forms

Then we have the matrix representation of Hermitian forms as follows.

Let V be a Hermitian space with Hermitian form $\langle -, - \rangle$. Then we can choose a basis $v = (v_1, v_2, \cdots, v_n)$ of V. Then we have

$$\omega = [\langle v_i, v_j \rangle]$$

Note that ω is a Hermitian matrix, i.e., $\omega^\dagger = \omega$. Then we claim that if A and $\widetilde{A}$ are two matrix representations of the Hermitian form $\langle -, - \rangle$ with respect to two different bases v and $\widetilde{v}$ respectively, then there exists an invertible matrix $P \in \mathsf{GL}_n(\mathbb{C})$ such that:

$$\widetilde{A} = P^\dagger A P$$

where P is the change-of-basis matrix from v to $\widetilde{v}$. Or equivalently,

$$\mathsf{H}_n(\mathbb{C}) \times \mathsf{GL}_n(\mathbb{C}) \to \mathsf{H}_n(\mathbb{C}), \quad (A, P) \mapsto P^\dagger AP$$

where $\mathsf{H}_n(\mathbb{C})$ is the real linear space of Hermitian matrix of order n. The reason why it is real, as it is not closed under multiplication by complex numbers. Take $n = 1$, then $\mathsf{H}_1(\mathbb{C}) = \mathbb{R}$, which is not closed under multiplication by complex numbers.

9.2 Self-Adjoint Operators and Unitary Operators

Let V be a Hermitian space with Hermitian form $\langle -, - \rangle$. Then we have the following definitions.

Definition 9.3 — Self-Adjoint Operator. A linear operator $T : V \to V$ is called a *self-adjoint operator* or *Hermitian operator* if:

$$\langle Tu, v \rangle = \langle u, Tv \rangle$$

for all $u, v \in V$. Or equivalently, $T = T^\dagger$, where $T^\dagger$ is the adjoint operator of T defined as the unique operator satisfying:

$$\langle Tu, v \rangle = \langle u, T^\dagger v \rangle$$

Definition 9.4 — Unitary Operator. A linear operator $U : V \to V$ is called a *unitary operator* if:

$$\langle Uu, Uv \rangle = \langle u, v \rangle$$

for all $u, v \in V$. Or equivalently, $U^\dagger = U^{-1}$.

Definition 9.5 — Normal Operator. A linear operator $N : V \to V$ is called a *normal operator* if:

$$N^\dagger N = N N^\dagger$$

Proposition 9.1 For $T : V \to W$ a linear operator between two Hermitian spaces V and W, there also exists a unique adjoint operator $T^\dagger : W \to V$ satisfying:

$$\langle Tu, w \rangle_W = \langle u, T^\dagger w \rangle_V$$

Proof. We can reduce the problem to $\mathbb{C}^n$ and $\mathbb{C}^m$ with standard Hermitian forms by choosing orthonormal bases of V and W. Then we have T represented by a matrix $A \in \mathsf{M}_{m \times n}(\mathbb{C})$. Then we propose there is a matrix $B \in \mathsf{M}_{n \times m}(\mathbb{C})$ such that for all $\vec{e}_i \in \mathbb{C}^n$ and $\vec{f}_j \in \mathbb{C}^m$, we have:

$$\langle A\vec{e}_i, \vec{f}_j \rangle = (A\vec{e}_i)^\dagger \vec{f}_j = \vec{e}_i^\dagger A^\dagger \vec{f}_j = \vec{e}_i^T A^\dagger \vec{f}_j$$

which is the (i, j)-th entry of $A^\dagger$. On the other hand, we have:

$$\langle \vec{e}_i, B\vec{f}_j \rangle = \vec{e}_i^\dagger (B\vec{f}_j) = \vec{e}_i^T B \vec{f}_j$$

which is the (i, j)-th entry of B. Therefore, we have $B = A^\dagger$. This proves the existence of the adjoint operator. The uniqueness is straightforward. ■

Proposition 9.2 Let T be a self-adjoint operator on a Hermitian space V. Then we have the following properties:

1. All eigenvalues of T are real numbers.
2. Eigenspaces of T are mutually orthogonal, i.e., if u and v are eigenvectors of T corresponding to distinct eigenvalues, then $\langle u, v \rangle = 0$.
3. V is the direct sum of the eigenspaces of T.

So T is completely reducible.

Proof. Given that $T^\dagger = T$, we have:

1. Let $\lambda \neq 0$ be an eigenvalue of T, so there exists a non-zero eigenvector v such that $Tv = \lambda v$. Then we have:

   $$\langle Tv, v\rangle = \langle v, T^\dagger v\rangle = \langle v, Tv\rangle$$

   which implies that:

   $$\lambda\langle v, v\rangle = \overline{\lambda}\langle v, v\rangle$$

   Since $v \neq 0$, we have $\langle v, v\rangle > 0$. Therefore, we have $\lambda = \overline{\lambda}$, i.e., λ is a real number.
2. Let λ_1 and λ_2 be two distinct eigenvalues of T with corresponding eigenvectors v_1 and v_2. Then we have:

   $$\langle Tv_1, v_2\rangle = \langle v_1, T^\dagger v_2\rangle$$

   which implies that:

   $$\lambda_1\langle v_1, v_2\rangle = \overline{\lambda_2}\langle v_1, v_2\rangle$$

   Since $\lambda_1 \neq \lambda_2$, we have $\langle v_1, v_2\rangle = 0$.
3. We know that $V_{\lambda_1}(T) \oplus \cdots \oplus V_{\lambda_k}(T) \subseteq V$, where the spectrum of T, $\sigma(T) = \{\lambda_1, \lambda_2, \cdots, \lambda_k\}$. To show the equality, we let $W = V_{\lambda_1}(T) \oplus \cdots \oplus V_{\lambda_k}(T)$ and consider the orthogonal complement $W^\perp$. Since T is self-adjoint, we have $W^\perp$ is T-invariant, i.e., for all $w^\perp \in W^\perp$, we have $Tw^\perp \in W^\perp$. As for all $w \in W$ and $w^\perp \in W^\perp$, we have:

   $$\langle Tw^\perp, w\rangle = \langle w^\perp, T^\dagger w\rangle = \langle w^\perp, Tw\rangle = 0$$

   where $Tw \in W$ since W is T-invariant. Then we claim that $W^\perp = \{0\}$. If not, then we have an eigenvector $w^\perp \in W^\perp$ with eigenvalue λ, such that there exists a map $\widetilde{T} : W^\perp \to W^\perp$ defined by $\widetilde{T}(w^\perp) = T(w^\perp)$. Then $\widetilde{T}w^\perp = \lambda w^\perp$ and $\widetilde{T}w^\perp = Tw^\perp$ by definition. So we know that λ is an eigenvalue of T, i.e., $\lambda \in \sigma(T)$. Say $\lambda = \lambda_1$. Then we have $w^\perp \in V_{\lambda_1}(T) \subseteq W$, which contradicts the assumption that $w^\perp \in W^\perp$. Therefore, we have $W^\perp = \{0\}$, which implies that $V = W$.

■

Proposition 9.3 Let T be a unitary operator on a Hermitian space V. Then we have the following properties:

1. All eigenvalues of T are complex numbers with absolute value 1.
2. Eigenspaces of T are mutually orthogonal, i.e., if u and v are eigenvectors of T corresponding to distinct eigenvalues, then $\langle u, v\rangle = 0$.
3. V is the direct sum of the eigenspaces of T.

So T is completely reducible.

Proof. Given that $T^\dagger T = TT^\dagger = 1_V$, we have:

1. Let $\lambda \neq 0$ be an eigenvalue of T, so there exists a non-zero eigenvector v such that $Tv = \lambda v$. Then we have:

   $$\langle Tv, v\rangle = \langle v, T^\dagger v\rangle$$

   which implies that:

   $$\lambda\langle v, v\rangle = \overline{\lambda}^{-1}\langle v, v\rangle \implies (\lambda \cdot \overline{\lambda} - 1)\langle v, v\rangle = 0$$

   Since $v \neq 0$, we have $\langle v, v\rangle > 0$. Therefore, we have $\lambda \cdot \overline{\lambda} = |\lambda|^2 = 1$, i.e., $|\lambda| = 1$.

2. Let λ_1 and λ_2 be two distinct eigenvalues of T with corresponding eigenvectors v_1 and v_2. Then we have:

$$\langle Tv_1, v_2\rangle = \langle v_1, T^\dagger v_2\rangle$$

which implies that:

$$\lambda_1\langle v_1, v_2\rangle = \overline{\lambda_2}^{-1}\langle v_1, v_2\rangle \implies (\lambda_1\overline{\lambda_2} - 1)\langle v_1, v_2\rangle = 0$$

Since $\lambda_1 \neq \lambda_2$, we have $\langle v_1, v_2\rangle = 0$.

3. We know that $V_{\lambda_1}(T) \oplus \cdots \oplus V_{\lambda_k}(T) \subseteq V$, where the spectrum of T, $\sigma(T) = \{\lambda_1, \lambda_2, \cdots, \lambda_k\}$. To show the equality, we let $W = V_{\lambda_1}(T) \oplus \cdots \oplus V_{\lambda_k}(T)$ and consider the orthogonal complement $W^\perp$. Since T is unitary, we have $W^\perp$ is T-invariant, i.e., for all $w^\perp \in W^\perp$, we have $Tw^\perp \in W^\perp$. As for all $w \in W$ and $w^\perp \in W^\perp$, we have $w' = Tw \in W$ and:

$$\langle Tw^\perp, w'\rangle = \langle Tw^\perp, Tw\rangle = \langle w^\perp, w\rangle = 0$$

where the second equality holds since T is unitary. Then we claim that $W^\perp = \{0\}$. If not, then we have an eigenvector $w^\perp \in W^\perp$ with eigenvalue λ, such that there exists a map $\widetilde{T} : W^\perp \to W^\perp$ defined by $\widetilde{T}(w^\perp) = T(w^\perp)$. Then $\widetilde{T}w^\perp = \lambda w^\perp$ and $\widetilde{T}w^\perp = Tw^\perp$ by definition. So we know that λ is an eigenvalue of T, i.e., $\lambda \in \sigma(T)$. Say $\lambda = \lambda_1$. Then we have $w^\perp \in V_{\lambda_1}(T) \subseteq W$, which contradicts the assumption that $w^\perp \in W^\perp$. Therefore, we have $W^\perp = \{0\}$, which implies that $V = W$.

■

9.3 Spectral Theorem

The canonical matrix representation of self-adjoint operator is a real diagonal matrix, and the canonical matrix representation of unitary operator is a diagonal matrix with entries on the unit circle in the complex plane. This is stated in the following spectral theorem.

Theorem 9.1 — Spectral Theorem. A Hermitian or unitary operator T on a Hermitian space V is "diagonalisable" by a unitary matrix in the following sense: Choose an orthonormal basis of V such that T is represented by a Hermitian matrix A. Then there is a unitary matrix U and a diagonal matrix D such that:

$$A = UDU^{-1} = UDU^{\dagger}$$

Note that the diagonal entries of D are all real numbers if T is Hermitian, and the diagonal entries of D are all complex numbers with modulus 1 if T is unitary. Moreover, D is the cononical form of T, i.e., there exists a set of distinct complex eigenvalues $\{\lambda_1, \lambda_2, \cdots, \lambda_k\}$ and a set of non-trivial complex linear subspaces $\{V_{\lambda_1}, V_{\lambda_2}, \cdots, V_{\lambda_k}\}$ such that:

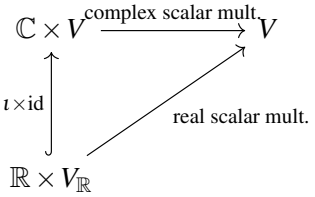
$$V = V_{\lambda_1} \oplus V_{\lambda_2} \oplus \cdots \oplus V_{\lambda_k}$$

with respect to which T has the decomposition:

$$T = \lambda_1 1_{V_{\lambda_1}} + \lambda_2 1_{V_{\lambda_2}} + \cdots + \lambda_k 1_{V_{\lambda_k}}$$

If U is a unitary matrix, then the columns of U form an orthonormal basis of $\mathbb{C}^n$. Moreover, the columns of U are eigenvectors of A corresponding to the eigenvalues on the diagonal of D. As $\mathbb{C}^n = \bigoplus_i E_{\lambda_i}(A)$, where λ_i are the eigenvalues of A, we have found an orthonormal basis consisting of eigenvectors of A.

If V is a complex linear space, then V is a real linear space with dimension doubled and we write $V_{\mathbb{R}}$ for the underlying real linear space of V. Then we losed some information from V to $V_{\mathbb{R}}$. Then we add an extra structure $\mathcal{J} : V_{\mathbb{R}} \to V_{\mathbb{R}}$ defined by $\mathcal{J}(v) = iv$ for all $v \in V$. Then we have $\mathcal{J}^2 = -1_{V_{\mathbb{R}}}$. Such a structure is called an *complex structure* on $V_{\mathbb{R}}$. Moreover, we have the following commutative diagram:



For example, we can write $(a+bi)v = av + b\mathcal{J}(v)$ for all $a+bi \in \mathbb{C}$ and $v \in V$. Note that as $(\det \mathcal{J})^2 = (-1)^{\dim_{\mathbb{R}} V}$, we have $\dim_{\mathbb{R}} V$ is even. The dimension doubled as we consider $v = (v_1, v_2, \cdots, v_n) \in V$ as $v_{\mathbb{R}} = (v_1, v_2, \cdots, v_n, \mathcal{J}v_1, \mathcal{J}v_2, \cdots, \mathcal{J}v_n) \in V_{\mathbb{R}}$.

Then we can do the reverse process. Let W be a real linear space. The complexification of W, denoted by $W_{\mathbb{C}}$, is defined to be the following complex linear space:

$$W \otimes_{\mathbb{R}} \mathbb{C}$$

Then we have the following natural identification:

$$W \subseteq W \otimes_{\mathbb{R}} \mathbb{C} = W_{\mathbb{C}}$$
$$w \mapsto w \otimes_{\mathbb{R}} 1$$

Then W is a real linear subspace of $W_{\mathbb{C}}$. Note that $\dim_{\mathbb{C}} W_{\mathbb{C}} = \dim_{\mathbb{R}} W$.

There are two corollories of the spectral theorem as follows.

Corollary 9.1 If A is a real symmetric matrix, then A can be diagonalised by an orthogonal matrix, i.e., there is an orthogonal matrix O and a diagonal matrix D such that:

$$A = ODO^{-1} = ODO^T$$

Proof. As real symmetric matrices are Hermitian matrices, by the spectral theorem for Hermitian matrices, we know that any real symmetric matrix can be diagonalised by a unitary matrix, i.e., $A = U^\dagger DU$ for some unitary matrix U and real diagonal matrix D. Note that the entries of U are complex numbers in general. Then we should try to find an orthogonal matrix O such that $A = O^T DO$. Note that for any real eigenvalue λ of A, the system $(A-\lambda I)\vec{v} = 0$ has real coefficients. Then if $\vec{v} = \vec{x} + i\vec{y}$ is a complex solution, then we have:

$$(A-\lambda I)\vec{v} = (A-\lambda I)\vec{x} + i(A-\lambda I)\vec{y} = 0$$

which implies that both $\vec{x}$ and $\vec{y}$ are real solutions. Therefore, we can always find a real eigenvector corresponding to each real eigenvalue of A. Then we can choose an orthonormal basis of $\mathbb{R}^n$ consisting of real eigenvectors of A by Gram-Schmidt process. Let O be the matrix whose columns are the orthonormal basis of real eigenvectors. Then we have O is an orthogonal matrix and $A = O^T DO$. Therefore, we conclude that any real symmetric matrix A can be diagonalised by an orthogonal matrix. ■

Corollary 9.2 The canonical form of a orthogonal matrix O of order n is of the following form:

$$\begin{bmatrix} R_{\theta_1} J_q & & & & \\ & R_{\theta_2} & & & \\ & & \ddots & & \\ & & & R_{\theta_k} & \\ & & & & I_p \end{bmatrix}$$

where $R_{\theta_i} = \begin{bmatrix} \cos\theta_i & -\sin\theta_i \\ \sin\theta_i & \cos\theta_i \end{bmatrix}$ is the rotation matrix of angle θ_i, $p = 1$ if n is odd and $p = 0$ if n is even, with $n = 2k + p$, and J_q is I_2 if $\det O = 1$ and $\begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$ if $\det O = -1$.

Corollary 9.3 The matrix representation H of the Hermitian form on a complex vector space V with respect to a basis v is a Hermitian matrix. Moreover, there exists a unitary matrix U and a real diagonal matrix D such that:

$$H = UDU^\dagger$$

Then the Hermitian form can be represented as:

$$\langle x, y\rangle = x^\dagger Hy = x^\dagger UDU^\dagger y = (U^\dagger x)^\dagger D(U^\dagger y)$$

Moreover, D can be expressed as:

$$D = \begin{bmatrix} \lambda & & \\ & -\mu & \\ & & 0 \end{bmatrix}$$

where $\lambda = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_r \end{bmatrix}$ and $\mu = \begin{bmatrix} \mu_1 & & \\ & \ddots & \\ & & \mu_s \end{bmatrix}$ with $\lambda_i, \mu_j > 0$ for all i, j. The pair (r,s) is called the *signature* of the Hermitian form. We may further decompose the Hermitian form as:

$$D = \begin{bmatrix} \sqrt{\lambda} & & \\ & -\sqrt{\mu} & \\ & & 0 \end{bmatrix} \begin{bmatrix} I_r & & \\ & -I_s & \\ & & 0 \end{bmatrix} \begin{bmatrix} \sqrt{\lambda} & & \\ & -\sqrt{\mu} & \\ & & 0 \end{bmatrix} = U' I_{r,s} U'^{\dagger}$$

where $\sqrt{\lambda} = \begin{bmatrix} \sqrt{\lambda_1} & & \\ & \ddots & \\ & & \sqrt{\lambda_r} \end{bmatrix}$ and $\sqrt{\mu} = \begin{bmatrix} \sqrt{\mu_1} & & \\ & \ddots & \\ & & \sqrt{\mu_s} \end{bmatrix}$.

So the Hermitian form can be represented as:

$$\langle x, y \rangle = (U'^{\dagger} U^{\dagger} x)^{\dagger} I_{r,s} (U'^{\dagger} U^{\dagger} y)$$

In summary,

- Any Hermitian form on a complex vector space can be represented by a Hermitian matrix.
- The canonical representation of Hermitian form is $I_{r,s}$ up to a unitary change of basis. If the Hermitian form is positive-definite, then the canonical representation is I_n.
- Any symmetric 2-form ω on a real vector space can be represented by a real symmetric matrix.
- The canonical representation of symmetric 2-form is $\begin{bmatrix} I_r & & \\ & -I_s & \\ & & 0 \end{bmatrix}$ up to an orthogonal change of basis. If the symmetric 2-form is positive-definite, then the canonical representation is I_n.
- The canonical representation of pseudo inner product is $\begin{bmatrix} I_p & \\ & -I_q \end{bmatrix}$ up to an orthogonal change of basis, with $n = p + q$. Then we call (p,q) the signature of the pseudo inner product.
- V is a real vector space of dimension n. Then up to isomorphism, there are $n+1$ different pseudo inner products on V, corresponding to the signatures $(n,0), (n-1,1), \cdots, (1,n-1), (0,n)$.
- Any pseudo inner product V is isomorphic to $(\mathbb{R}^n, I_{p,q}) = \mathbb{R}^{p,q}$. As it sends $(x,y) \to x_1 y_1 + \cdots + x_p y_p - x_{p+1} y_{p+1} - \cdots - x_n y_n$.

The set of inner products on a real vector space V of dimension n is isomorphic to the orbit space of the right action of group $\mathsf{O}(n)$ on $\mathsf{GL}_n(\mathbb{R})$ $\mathsf{GL}_n(\mathbb{R})/\mathsf{O}(n)$, where $\mathsf{O}(n)$ is the orthogonal group of order n.

$$\mathsf{GL}_n(\mathbb{R}) \times \mathsf{O}(n) \to \mathsf{GL}_n(\mathbb{R}), \quad (X, g) \mapsto X \cdot g$$

As $\mathsf{GL}_n(\mathbb{R})$ and $\mathsf{O}(n)$ have the same homotopy type, the orbit space $\mathsf{GL}_n(\mathbb{R})/\mathsf{O}(n)$ is trivially contractible. We may consider the following example:

$$\mathsf{GL}_1(\mathbb{R}) = \mathbb{R}^{\times} \quad \mathsf{O}(1) = \{-1, 1\}$$

Then we have:

$$\mathsf{GL}_1(\mathbb{R})/\mathsf{O}(1) \cong \mathbb{R}_{>0}$$

Similarly, the set of Hermitian forms on a complex vector space V of dimension n is isomorphic to the orbit space $\mathsf{GL}_n(\mathbb{C})/\mathsf{U}(n)$, where $\mathsf{U}(n)$ is the unitary group of order n. Again, it is contractible.

We have a simple introduction to the Lorentz inner product on $\mathbb{R}^4$. It sends $(x,y) \in \mathbb{R}^4 \times \mathbb{R}^4$ to $x \cdot y = x_0 y_0 - \vec{x} \cdot \vec{y}$, where $x = \begin{bmatrix} x_0 \\ \vec{x} \end{bmatrix}$ and $y = \begin{bmatrix} y_0 \\ \vec{y} \end{bmatrix}$.

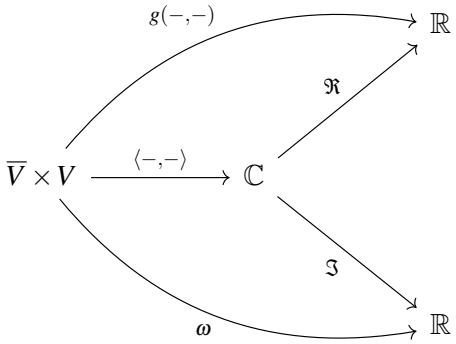
10. Symplectic Vector Spaces

By now, you should feel comfortable switching between the 2 pictures. One is the abstract picture. Another is a concrete presentation.

GUOWU MENG

10.1 Symplectic Forms

Let $(V, \langle -, - \rangle)$ be a Hermitian space. Then we have:



where $g(-,-)$ is the real part of the Hermitian product and ω is the imaginary part of the Hermitian product. Both of them are 2-forms on $V_{\mathbb{R}}$. ω is called a *symplectic form* on V.

Definition 10.1 — Symplectic form. A *symplectic form* on a real vector space V is a non-degenerate, skew-symmetric 2-form $\omega : V \times V \to \mathbb{R}$.

A symplectic vector space is a pair (V, ω).

We have $\mathcal{J} \in \mathrm{End}(V_{\mathbb{R}})$ defined as the scalar multiplication by i on $V_{\mathbb{R}}$, such that $\mathcal{J}^2 = -1_{V_{\mathbb{R}}}$. Note that we have three structures on $V_{\mathbb{R}}$:

- **Complex structure:** $\mathcal{J} : V_{\mathbb{R}} \to V_{\mathbb{R}}$ with $\mathcal{J}^2 = -1_{V_{\mathbb{R}}}$;
- **Symplectic structure:** $\omega : V_{\mathbb{R}} \times V_{\mathbb{R}} \to \mathbb{R}$ is a non-degenerate, skew-symmetric bilinear form;
- **Riemannian structure:** $g : V_{\mathbb{R}} \times V_{\mathbb{R}} \to \mathbb{R}$ is a positive-definite, symmetric bilinear form.

Then we have the following equation:

$$\langle x,y\rangle = g(x,y) + i\omega(x,y)$$

for all $x,y \in V_{\mathbb{R}}$. Moreover, we have:

$$\langle ix,y\rangle = -i\langle x,y\rangle \implies g(\mathcal{J}x,y) + i\omega(\mathcal{J}x,y) = \omega(x,y) - ig(x,y)$$

for all $x,y \in V_{\mathbb{R}}$. This implies that:

$$\omega(x,y) = g(\mathcal{J}x,y), \quad g(x,y) = -\omega(x,\mathcal{J}y)$$

Consider the following equation:

$$\langle ix,iy\rangle = \langle x,y\rangle \implies g(\mathcal{J}x,\mathcal{J}y) + i\omega(\mathcal{J}x,\mathcal{J}y) = g(x,y) + i\omega(x,y)$$

for all $x,y \in V_{\mathbb{R}}$. This implies that:

$$g(\mathcal{J}x,\mathcal{J}y) = g(x,y), \quad \omega(\mathcal{J}x,\mathcal{J}y) = \omega(x,y)$$

for all $x,y \in V_{\mathbb{R}}$. Or equivalently, we have $\mathcal{J}^* g = g$ and $\mathcal{J}^*\omega = \omega$.

Note that the Hermitian product is positive-definite, so we have

$$\langle x,x\rangle > 0 \implies g(x,x) > 0, \omega(x,x) = 0$$

for all $x \in V_{\mathbb{R}} \setminus \{0\}$. If $x = 0$, then we have $\langle 0,0\rangle = 0$, $g(0,0) = 0$ and $\omega(0,0) = 0$. Also, we have

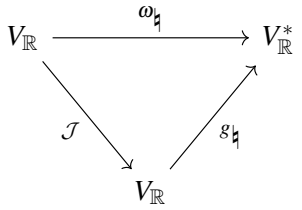
$$\overline{\langle y,x\rangle} = \langle x,y\rangle \implies g(y,x) - i\omega(y,x) = g(x,y) + i\omega(x,y)$$

for all $x,y \in V_{\mathbb{R}}$. This implies that:

$$g(x,y) = g(y,x), \quad \omega(x,y) = -\omega(y,x)$$

for all $x,y \in V_{\mathbb{R}}$, i.e., g is symmetric and ω is skew-symmetric.

As $\omega(x,y) = g(\mathcal{J}x,y)$ for all $x,y \in V_{\mathbb{R}}$, so ω is non-degenerate if g is non-degenerate. Then we have the following commutative diagram:



As $\omega_\natural(x) = g_\natural(\mathcal{J}x)$ for all $x \in V_{\mathbb{R}}$.

Then we can recover a Hermitian space from a real vector space with these structures. Let V be a real vector space. If any two of the above three structures are given and compatible, the third will be determined. Moreover, we have a Hermitian product on V on the complex linear space $(V,\mathcal{J})$ where $iv = \mathcal{J}v$ for all $v \in V$.

The meaning of being compatible pair:

- $(g,\mathcal{J})$ are compatible if $\mathcal{J}^* g = g$, i.e., $J \in \mathsf{Aut}(W,g) = \mathsf{O}(W,g)$; Then we can define $\omega(x,y) = g(\mathcal{J}x,y)$ and $\langle -,-\rangle = g + i\omega$. We can check that ω is skew-symmetric and non-degenerate, and $\langle -,-\rangle$ is a Hermitian product:

$$\omega(y,x) = g(\mathcal{J}y,x) = g(\mathcal{J}\mathcal{J}y,\mathcal{J}x) = g(-y,\mathcal{J}x) = -g(\mathcal{J}x,y) = -\omega(x,y)$$

Also, if $\omega(x,y)=0$ for all $y\in V$, then we have $g(\mathcal{J}x,y)=0$ for all $y\in V$, which implies that $\mathcal{J}x=0$ as g is non-degenerate, i.e., $x=0$. Therefore, ω is non-degenerate. As for the Hermitian product, the sesquilinearity is shown as follows:

$$\langle ix,y\rangle = g(\mathcal{J}x,y)+i\omega(\mathcal{J}x,y)=\omega(x,y)-ig(x,y)=-i(g(x,y)+i\omega(x,y))=-i\langle x,y\rangle$$

For the conjugate symmetry, we have:

$$\langle y,x\rangle = g(y,x)+i\omega(y,x)=g(x,y)-i\omega(x,y)=\overline{\langle x,y\rangle}$$

for all $x,y\in V$. Also, we have:

$$\langle x,x\rangle = g(x,x)+i\omega(x,x)=g(x,x)>0$$

- $(\omega,\mathcal{J})$ are compatible if $\mathcal{J}^*\omega=\omega$, i.e., $\mathcal{J}\in\mathsf{Aut}(W,\omega)=\mathsf{Sp}(W,\omega)$ and $-\omega(\mathcal{J}x,x)\geq 0$ and equality holds if and only if $x=0$. Then we can define $g(x,y)=-\omega(x,\mathcal{J}y)$ and $\langle -,-\rangle = g+i\omega$. We can check that g is symmetric and positive-definite, and $\langle -,-\rangle$ is a Hermitian product:

  $$g(y,x)=-\omega(y,\mathcal{J}x)=-\omega(\mathcal{J}y,\mathcal{J}\mathcal{J}x)=-\omega(\mathcal{J}y,-x)=-\omega(x,\mathcal{J}y)=g(x,y)$$

  Also, as $-\omega(\mathcal{J}x,x)\geq 0$ for all $x\in V$ and equality holds if and only if $x=0$, we have $g(x,x)\geq 0$ for all $x\in V$ and equality holds if and only if $x=0$. Therefore, g is positive-definite. For the Hermitian product, we may use the similar proof as above.
- (g,ω) are compatible if $\omega(x,y)=g(Ax,y)$ for some $A\in\mathsf{End}(V)$. If $A^2=-1$, then $\mathcal{J}=A$. In general, A is skew-symmetric, i.e., $g(Ax,y)=g(x,-Ay)$, as ω is skew-symmetric. Since $AA^\dagger$ is symmetric and positive-definite, we can define $\mathcal{J}=\sqrt{AA^\dagger}^{-1}A$, which satisfies that $\mathcal{J}^2=-1$, as $\mathcal{J}$ commutes with A and $\sqrt{AA^\dagger}$. Then we have $A=\sqrt{AA^\dagger}\mathcal{J}$ and let $P=\sqrt{AA^\dagger}$. Therefore, we have:

  $$\omega(\mathcal{J}x,\mathcal{J}y)=g(A\mathcal{J}x,\mathcal{J}y)=g(P\mathcal{J}\mathcal{J}x,\mathcal{J}y)=-g(Px,\mathcal{J}y)=g(\mathcal{J}Px,y)=g(Ax,y)=\omega(x,y).$$

  Also, we have:

  $$-\omega(\mathcal{J}x,x)=-g(A\mathcal{J}x,x)=-g(P\mathcal{J}\mathcal{J}x,x)=g(Px,x)>0$$

  for all $x,y\in V$ and $x\neq 0$. Then we can define $\langle -,-\rangle = g+i\omega$. We can check that $\langle -,-\rangle$ is a Hermitian product by the similar proof as above.

Let V be a vector space over $\mathbb{F}$ where $\mathsf{char}(\mathbb{F})\neq 2$. We define the double $D(V)=V\oplus V^*$. Then we have a natural symplectic form on $D(V)$ defined as:

$$\omega((u,\alpha),(v,\beta))=\alpha(v)-\beta(u)$$

Also $D(V)$ is called the *canonical symplectic vector space* associated to V. Then when we choose a basis $\{\vec{e}_1,\vec{e}_2,\cdots,\vec{e}_n\}$ of V and the dual basis $\{\hat{e}^1,\hat{e}^2,\cdots,\hat{e}^n\}$ of V^*, we have the matrix representation of ω on $D(V)$ as:

$$\begin{bmatrix}\omega(\vec{e}_i,\vec{e}_j) & \omega(\vec{e}_i,\hat{e}^j)\\ \omega(\hat{e}^i,\vec{e}_j) & \omega(\hat{e}^i,\hat{e}^j)\end{bmatrix}=\begin{bmatrix}0 & I_n\\ -I_n & 0\end{bmatrix}$$

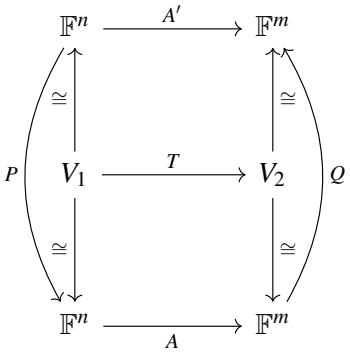
Also the basis $\{\vec{e}_1,\cdots,\vec{e}_n,\hat{e}^1,\cdots,\hat{e}^n\}$ is called a *symplectic basis* of $D(V)$.

10.2 Matrix Representation and Canonical Form

We may revise all the canonical forms we have learned before as follows.

10.2.1 Linear Maps

Consider a linear map $T : V_1 \to V_2$ between two vector spaces V_1 and V_2 of dimensions n and m respectively. Then we have:



where A and A' are the matrix representations of T with respect to different bases of V_1 and V_2 and $P \in \mathsf{GL}_n(\mathbb{F})$ and $Q \in \mathsf{GL}_m(\mathbb{F})$ are the change-of-basis matrices. P represents the column operations on A and Q represents the row operations on A. Then we have:

$$AP = QA', \quad A = QA'P^{-1}$$

Then we have the left group action of $\mathsf{GL}_m(\mathbb{F}) \times \mathsf{GL}_n(\mathbb{F})$ on the set of $m \times n$ matrices $\mathsf{M}_{m\times n}(\mathbb{F})$ defined as:

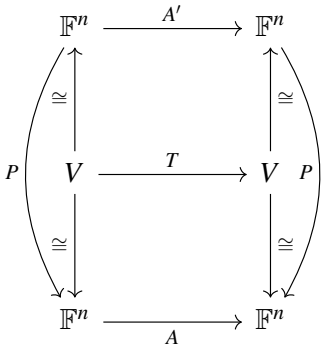
$$(Q,P) \cdot A = QAP^{-1}$$

The canonical form of A under this group action is:

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

10.2.2 Linear Endomorphisms

Consider a linear endomorphism $T : V \to V$ on a vector space V of dimension n. Then we have:



where A and A' are the matrix representations of T with respect to different bases of V and $P \in \mathsf{GL}_n(\mathbb{F})$ is the change-of-basis matrix. Then we have:

$$AP = PA', \quad A = PA'P^{-1}$$

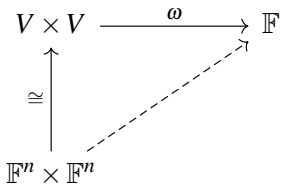
Then we have the left group action of $\mathsf{GL}_n(\mathbb{F})$ on the set of $n \times n$ matrices $\mathsf{M}_{n\times n}(\mathbb{F})$ defined as:

$$P \cdot A = PAP^{-1}$$

The actual canonical form of A is complicated (Rational Canonical Form), but in generic case, they are diagonal matrix.

10.2.3 2-Forms

Consider a 2-form $\omega : V \times V \to \mathbb{F}$ on a vector space V of dimension n. Then we have:



Then $[\omega]_v = [\omega(v_i, v_j)]$ is the matrix representation of ω with respect to the basis $v = \{v_1, v_2, \cdots, v_n\}$ of V. If we change the basis of V to u, then there is a unique invertible matrix, $P \in \mathsf{GL}_n(\mathbb{F})$, such that $u_j = \sum_i v_i P^i_j$ for all j. Then we have:

$$\begin{aligned}
[\omega]_u = [\omega(u_i, u_j)] &= [\omega(\sum_k v_k P^k_i, \sum_l v_l P^l_j)] \\
&= [\sum_{k,l} P^k_i \omega(v_k, v_l) P^l_j] \\
&= [\sum_{k,l} (P^T)^i_k \omega(v_k, v_l) P^l_j] \\
&= P^T [\omega(v_k, v_l)] P
\end{aligned}$$

So we have the right group action of $\mathsf{GL}_n(\mathbb{F})$ on the set of $n \times n$ matrices $\mathsf{M}_{n\times n}(\mathbb{F})$ defined as:

$$A \cdot P = P^T AP$$

We may check that $(A \cdot P_1) \cdot P_2 = A \cdot (P_1 P_2)$ for all $A \in \mathsf{M}_{n\times n}(\mathbb{F})$ and $P_1, P_2 \in \mathsf{GL}_n(\mathbb{F})$.

Note that the right action leaves the symmetric and skew-symmetric properties invariant, i.e., if $A^T = A$ (or $A^T = -A$), then we have $(P^T AP)^T = P^T AP$ (or $(P^T AP)^T = -P^T AP$) for all $P \in \mathsf{GL}_n(\mathbb{F})$. For symmetric 2-forms, as $(P^T AP)^T = P^T A^T (P^T)^T = P^T A^T P$, where $A^T = A$, so we have $(P^T AP)^T = P^T AP$. For skew-symmetric 2-forms, as $(P^T AP)^T = P^T A^T (P^T)^T = P^T (-A) P$, where $A^T = -A$, so we have $(P^T AP)^T = -P^T AP$.

When $\mathbb{F} = \mathbb{R}$, then the ω being symmetric or skew-symmetric corresponds to the matrix representation being real symmetric or real skew-symmetric respectively. If ω is symmetric, then the representation A is a Hermitian matrix, and we have $A = ODO^T$ for some orthogonal matrix O and diagonal matrix D. Then the canonical form of symmetric 2-form is:

$$\begin{bmatrix} I_r & & \\ & -I_s & \\ & & 0 \end{bmatrix}$$

where $r + s \leq n$ and $r + s + t = n$. If ω is skew-symmetric, then the iA is a Hermitian matrix, and we have $iA = UDU^\dagger$ for some unitary matrix U and real diagonal matrix D. Then the canonical form of skew-symmetric 2-form is:

$$\begin{bmatrix} J_2 & & & \\ & \ddots & & \\ & & J_2 & \\ & & & 0 \end{bmatrix}$$

where $J_2 = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ and the canonical form can be represented by $J_2 \oplus J_2 \oplus \cdots \oplus J_2 \oplus 0$. Note that $J_2^2 = -I_2$.

The canonical form of a pseudo inner product on a real linear space of dimension n is $I_{p,q} = \begin{bmatrix} I_p & 0 \\ 0 & -I_q \end{bmatrix}$ where $p+q=n$. The basis inside the canonical representation is called *pseudo-orthonormal* basis.

A pseudo Euclidean space is isomorphic to $\mathbb{R}^{p,q} := (\mathbb{R}^n, (\vec{x},\vec{y}) \mapsto \vec{x} \cdot I_{p,q} \vec{y})$. In case the dimension of V is n, then up to isomorphism, there are $n+1$ pseudo Euclidean structures on V, namely, $\mathbb{R}^{0,n}, \mathbb{R}^{1,n-1}, \cdots, \mathbb{R}^{n,0}$. Note that $(v_i, v_j) = \delta_{ij}$ for $1 \leq i, j \leq p$, $(v_i, v_j) = -\delta_{ij}$ for $p+1 \leq i, j \leq n$ and $(v_i, v_j) = 0$ otherwise.

Up to isomorphism, there is only one real symplectic vector space of dimension $2n$, i.e., $D(\mathbb{R}^n) := \mathbb{R}^n \oplus (\mathbb{R}^n)^*$ with the canonical symplectic form. The representation of the symplectic form is

$$\begin{bmatrix} 0 & I \\ -I & 0 \end{bmatrix}$$

with respect to the symplectic basis: $(x_1, \cdots, x_n, x^1, \cdots, x^n)$, where $\{x_1, \cdots, x_n\}$ is the standard basis of $\mathbb{R}^n$ and $\{x^1, \cdots, x^n\}$ is the dual basis of $(\mathbb{R}^n)^*$. Also, $\omega(x_i, x_j) = \omega(x^i, x^j) = 0$ and $\omega(x_i, x^j) = \delta_i^j = -\omega(x^j, x_i)$ for all i, j.

Note that we have $A^T = -A$ where A is the representation of a symplectic form. As $\det A^T = \det A = (-1)^n \det A$, we know that n has to be even. Moreover, if we consider the a non-degenerate skew-symmetric 2-form on a real vector space of dimension $2n$, then its canonical form is:

$$\begin{bmatrix} J_2 & & \\ & \ddots & \\ & & J_2 \end{bmatrix}$$

where $J_2 = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$. Note that this is similar to the canonical form of symplectic forms mentioned above.

11. Further Topics

When you studying higher maths, whether algebra, geometry or anything, you realised that the hard part is the language. It takes time. People are impatient. If you are impatient, you cannot learn mathematics. But if you are patient enough, you learn the language, you understand the basic facts. No tricks, tricks are useless. And then towards the end, you enjoy the fruit, that means, everything become so easy. Just do a simple calculation. You can get many result. The center of mathematics is always like that.

GUOWU MENG

11.1 Polar Decomposition and Singular Value Decomposition

11.1.1 Polar Decomposition

If $z \neq 0$, then $z = \rho e^{i\theta}$ for a unique $\rho > 0$ and $e^{i\theta}$ being a complex number of modulus 1. This is called the polar decomposition of z. Then we have the following isomorphism:

$$\mathsf{GL}_1(\mathbb{C}) = \mathsf{U}(1) \cdot \mathsf{H}_1^{>0}(\mathbb{C}), \quad [z] \mapsto [e^{i\theta}] \cdot [\rho]$$

where $\mathsf{H}_1^{>0}(\mathbb{C})$ is the set of positive Hermitian 1×1 matrices, i.e., positive real numbers, and $\mathsf{U}(1)$ is the set of complex numbers of modulus 1.

Then we may generalise this to matrices, i.e.,

$$\mathsf{GL}_n(\mathbb{C}) = \mathsf{U}(n) \cdot \mathsf{H}_n^{>0}(\mathbb{C}), \quad [A] \mapsto [U] \cdot [P]$$

where $\mathsf{H}_n^{>0}(\mathbb{C})$ is the set of positive Hermitian $n \times n$ matrices and $\mathsf{U}(n)$ is the unitary group of order n. Then we claim that any invertible matrix A can be uniquely decomposed as $A = PU$ for some $P \in \mathsf{H}_n^{>0}(\mathbb{C})$ and $U \in \mathsf{U}(n)$, and we call this the *polar decomposition* of A.

Proof. Assume the existance, if $A = UP$ then $A^\dagger = PU^\dagger$. Then we have:

$$A^\dagger A = PU^\dagger UP = P^2$$

As A is invertible, so is $A^\dagger A$. Therefore, $P = \sqrt{A^\dagger A}$ is a positive Hermitian matrix. Then we have $U = AP^{-1}$. Also, we have:

$$(A^\dagger A)^\dagger = A^\dagger A \implies P^\dagger P^\dagger = P^2 \implies P^\dagger = P$$

and

$$\vec{z}^\dagger A^\dagger A\vec{z} = (A\vec{z})^\dagger (A\vec{z}) > 0 \implies \|P\vec{z}\| \geq 0$$

for all $\vec{z}$ and equal to 0 if and only if $\vec{z} = 0$ as $A \in \mathsf{GL}_n(\mathbb{C})$. Therefore, P and $A^\dagger A$ are positive Hermitian. Then we know that $A^\dagger A = U'DU'^\dagger$ where $U' \in \mathsf{U}(n)$ and D is a diagonal matrix with positive real numbers on the diagonal. Then we have $P = U'\sqrt{D}U'^\dagger$. Also, we have:

$$P^2 = U'\sqrt{D}U'^\dagger U'\sqrt{D}U'^\dagger = U'DU'^\dagger = A^\dagger A$$

Then we have:

$$U^\dagger U = P^{-1}A^\dagger AP^{-1} = P^{-1}P^2P^{-1} = I_n$$

Therefore, $U \in \mathsf{U}(n)$. ■

If it is real number, then we have the similar polar decomposition:

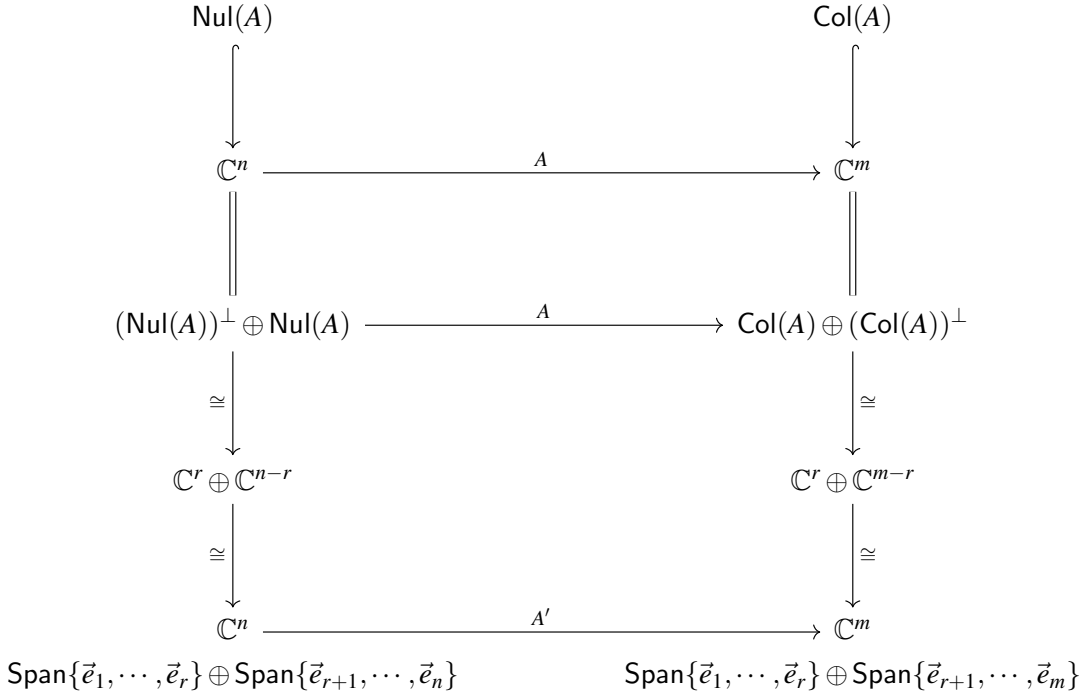
$$\mathsf{GL}_n(\mathbb{R}) = \mathsf{O}(n) \cdot \mathsf{S}_n^{>0}(\mathbb{R}), \quad [A] \mapsto [O] \cdot [S]$$

where $\mathsf{S}_n^{>0}(\mathbb{R})$ is the set of positive symmetric $n \times n$ matrices and $\mathsf{O}(n)$ is the orthogonal group of order n.

11.1.2 Singlular Value Decomposition

The corollary of polar decomposition is the singular value decomposition.

We consider the following commutative diagram:



where $A' = \begin{bmatrix} \overline{A} & 0 \\ 0 & 0 \end{bmatrix}$ with $\overline{A} \in \mathsf{GL}_r(\mathbb{C})$. Moreover, the direct sum in $(\mathsf{Nul}(A))^\perp \oplus \mathsf{Nul}(A)$ and $\mathsf{Col}(A) \oplus (\mathsf{Col}(A))^\perp$ are orthogonal direct sums; the direct sum in $\mathbb{C}^r \oplus \mathbb{C}^{n-r}$ and $\mathbb{C}^r \oplus \mathbb{C}^{m-r}$ are external direct sums; the direct sum in $\mathsf{Span}\{\vec{e}_1, \cdots, \vec{e}_r\} \oplus \mathsf{Span}\{\vec{e}_{r+1}, \cdots, \vec{e}_n\}$ and $\mathsf{Span}\{\vec{e}_1, \cdots, \vec{e}_r\} \oplus \mathsf{Span}\{\vec{e}_{r+1}, \cdots, \vec{e}_m\}$ are internal direct sums. Note that all the isomorphisms in the diagram are of Hermitian spaces. Then we may simplify the diagram as follows:

$$\begin{array}{ccc} \mathbb{C}^n & \xrightarrow{A} & \mathbb{C}^m \\ {\scriptstyle U_1}\downarrow & & \downarrow{\scriptstyle U_2} \\ \mathbb{C}^n & \xrightarrow{A'} & \mathbb{C}^m \end{array}$$

As $\overline{A} \in \mathsf{GL}_r(\mathbb{C})$, we have the polar decomposition $\overline{A} = U_3 P$ for some $P \in \mathsf{H}_r^{>0}(\mathbb{C})$ and $U_3 \in \mathsf{U}(r)$. Moreover, we may further decompose P as $P = U_4 D_\lambda U_4^\dagger$ for some $U_4 \in \mathsf{U}(r)$ and D_λ being a diagonal matrix with positive real numbers on the diagonal. Then we have:

$$\begin{aligned} U_2 A &= \begin{bmatrix} \overline{A} & 0 \\ 0 & 0 \end{bmatrix} U_1 \\ A &= U_2^\dagger \begin{bmatrix} U_3 U_4 D_\lambda U_4^\dagger & 0 \\ 0 & 0 \end{bmatrix} U_1 \\ &= \left(U_2^\dagger \begin{bmatrix} U_3 U_4 & 0 \\ 0 & I_{m-r} \end{bmatrix} \right) \begin{bmatrix} D_\lambda & 0 \\ 0 & 0 \end{bmatrix} \left(\begin{bmatrix} U_4^\dagger & 0 \\ 0 & I_{n-r} \end{bmatrix} U_1 \right) \end{aligned}$$

Then we have the singular value decomposition of A:

$$A = U \Sigma V^\dagger$$

where $U = U_2^\dagger \begin{bmatrix} U_3 U_4 & 0 \\ 0 & I_{m-r} \end{bmatrix}$, $\Sigma = \begin{bmatrix} D_\lambda & 0 \\ 0 & 0 \end{bmatrix}$, $V = U_1^\dagger \begin{bmatrix} U_4 & 0 \\ 0 & I_{n-r} \end{bmatrix}$.

Theorem 11.1 — Singular Value Decomposition. For any $A \in \mathsf{M}_{m \times n}(\mathbb{C})$, there exist unitary matrices $U \in \mathsf{U}(m)$, $V \in \mathsf{U}(n)$ and a set of positive numbers $\{\lambda_1, \cdots, \lambda_r\}$ such that:

$$A = U \Sigma V^\dagger, \quad \Sigma = \begin{bmatrix} \lambda_1 & & & & \\ & \lambda_2 & & & \\ & & \ddots & & \\ & & & \lambda_r & \\ & & & & 0 \end{bmatrix}$$

11.2 Simultaneous Diagonalisation Theorem

Theorem 11.2 — Simultaneous Diagonalisation Theorem. Suppose that $A_1, \cdots, A_k$ are mutually commuting Hermitian matrices of order n, i.e., $A_i \in \mathsf{H}_n(\mathbb{C})$ and $[A_i, A_j] := A_iA_j - A_jA_i = 0$ for all $1 \leq i, j \leq k$, where $[A_i, A_j]$ is called the commutator of A_i and A_j. Then there is a set of distinct vectors $\vec{\lambda_\alpha} \in \mathbb{R}^k$ for $\alpha = 1, 2, \cdots, l$ and an orthogonal decomposition of $\mathbb{C}^n$ into non-trivial subspaces:

$$\mathbb{C}^n = \bigoplus_{\alpha=1}^{l} E_{\vec{\lambda_\alpha}}$$

such that for all $\vec{z} \in E_{\vec{\lambda_\alpha}}$ and $A_i\vec{z} = \lambda_\alpha(i)\vec{z}$ for all $1 \leq i \leq k$. In particular, there is a unitary matrix $U \in \mathsf{U}(n)$ such that:

$$A_i = UD_iU^\dagger, \quad D_i = \begin{bmatrix} d_1(i) & & \\ & \ddots & \\ & & d_n(i) \end{bmatrix} \in \mathsf{M}_{n\times n}(\mathbb{R})$$

for all $1 \leq i \leq k$ and $d_j(i)$ are distinct.

Proof. We may induct on k or prove the case $k = 2$. For $k = 2$, as A_1, A_2 are Hermitian, we have $A_1A_2 = A_2A_1$. Then we have A_1 acts on $\mathbb{C}^n = E_{\lambda_1}(A_1) \oplus E_{\lambda_2}(A_1) \oplus \cdots \oplus E_{\lambda_k}(A_1)$ where $\lambda_1, \lambda_2, \cdots, \lambda_k$ are the distinct eigenvalues of A_1. Then we also consider A_2 acts on $\mathbb{C}^n$. We have the following claim: The action of A_2 on $\mathbb{C}^n$ leaves each eigenspace of A_1 invariant. For any $\vec{z} \in E_{\lambda_i}(A_1)$, we have:

$$A_1(A_2\vec{z}) = A_2(A_1\vec{z}) = A_2(\lambda_i\vec{z}) = \lambda_i(A_2\vec{z})$$

Hence, $A_2\vec{z} \in E_{\lambda_i}(A_1)$. Then, we have $A_2 = A_2^1 \oplus A_2^2 \oplus \cdots \oplus A_2^k$. We claim that each A_2^i is Hermitian on $E_{\lambda_i}(A_1)$. For any $\vec{x}, \vec{y} \in E_{\lambda_i}(A_1)$, we have:

$$\langle \vec{x}, A_2^i\vec{y}\rangle = \langle \vec{x}, A_2\vec{y}\rangle = \langle A_2\vec{x}, \vec{y}\rangle = \langle A_2^i\vec{x}, \vec{y}\rangle$$

So, A_2^i is diagonalisable on $E_{\lambda_i}(A_1)$ with an orthonormal eigenbasis and distinct eigenvalues μ_j. Therefore, we have:

$$E_{\lambda_i}(A_1) = \bigoplus_j E_{\lambda_i,\mu_j}(A_1, A_2).$$

Then we have:

$$\mathbb{C}^n = \bigoplus_{i,j} E_{\lambda_i,\mu_j}(A_1, A_2).$$

We may also write λ_i, μ_j as a vector in $\mathbb{R}^2$, i.e., $\vec{\lambda_{i,j}} = (\lambda_i, \mu_j)$. ■

We can use the simultaneous diagonalisation theorem to prove the spectral theorem for normal operators.

Theorem 11.3 A complex square matrix can be diagonalised by a unitary matrix if and only if it is normal.

Proof. ($\Rightarrow$) Assume that A can be diagonalised by a unitary matrix, i.e., there is a unitary matrix U such that $A = UDU^\dagger$ where D is a diagonal matrix. Then we have:

$$A^\dagger = UD^\dagger U^\dagger$$

where $D^\dagger$ is also a diagonal matrix. Then we have:

$$AA^\dagger = UDU^\dagger UD^\dagger U^\dagger = UDD^\dagger U^\dagger = UD^\dagger DU^\dagger = A^\dagger A$$

$DD^\dagger = D^\dagger D$ as we have the following equality:

$$\begin{bmatrix} d_1 & & \\ & \ddots & \\ & & d_n \end{bmatrix} \begin{bmatrix} \overline{d_1} & & \\ & \ddots & \\ & & \overline{d_n} \end{bmatrix} = \begin{bmatrix} |d_1|^2 & & \\ & \ddots & \\ & & |d_n|^2 \end{bmatrix}.$$

Therefore, A is normal.

($\Leftarrow$) Assume that A is normal, i.e., $AA^\dagger = A^\dagger A$. Then we write $A = B + iC$ where $B = \frac{A+A^\dagger}{2}$ and $C = \frac{A-A^\dagger}{2i}$. Then we claim that $[B,C] = 0$ if and only if A is normal. We have:

$$AA^\dagger = (B+iC)(B-iC) = B^2 + C^2 - i[B,C]$$
$$A^\dagger A = (B-iC)(B+iC) = B^2 + C^2 + i[B,C]$$

Therefore, $AA^\dagger = A^\dagger A$ if and only if $[B,C] = 0$. Also, we may check that B and C are Hermitian:

$$B^\dagger = \left(\frac{A+A^\dagger}{2}\right)^\dagger = \frac{A^\dagger + A}{2} = B, \quad C^\dagger = \left(\frac{A-A^\dagger}{2i}\right)^\dagger = \frac{A^\dagger - A}{-2i} = C.$$

Then, by the simultaneous diagonalisation theorem, there is a unitary matrix U such that:

$$B = UD_B U^\dagger, \quad C = UD_C U^\dagger$$

where D_B and D_C are diagonal matrices. Therefore, we have:

$$A = B + iC = UD_B U^\dagger + iUD_C U^\dagger = U(D_B + iD_C)U^\dagger = UD_A U^\dagger$$

where $D_A = D_B + iD_C$ is also a diagonal matrix. Hence, A can be diagonalised by a unitary matrix. ■

The following chapters are not in the exam syllabus, but for your reference.

11.3 Affine Spaces

A line or a plane can be regarded as an affine space. An affine space differs from a vector space in that it does not have a distinguished origin. We may say that $\mathcal{T}_O\mathbb{A}$ is the tangent space of an affine space $\mathbb{A}$ at a point $O \in \mathbb{A}$. We also have symmetric spaces, which can be a sphere.

Let $\mathbb{F}$ be a field. An affine space of dimension n over $\mathbb{F}$, $\mathbb{A}$, is a principal $(\mathbb{F}^n, +)$-set. A G-set, the set on which G acts, is called principal G-set if the action is principal, i.e., transitive and free.

■ **Example 11.1** $\mathbb{F}^n$ is an affine space of dimension n over $\mathbb{F}$ with the usual addition action of $(\mathbb{F}^n, +)$ on itself.

$$\begin{aligned}(\mathbb{F}^n, +) \times \mathbb{F}^n &\to \mathbb{F}^n \\ (\vec{v}, \vec{x}) &\mapsto \vec{v} + \vec{x}\end{aligned}$$

For any $\vec{x}, \vec{y} \in \mathbb{F}^n$, there is a unique $\vec{v} = \vec{y} - \vec{x} \in \mathbb{F}^n$ such that $\vec{v} + \vec{x} = \vec{y}$. Therefore, the action is transitive and free. ■

In fact, any $\mathbb{F}$-linear space is a $\mathbb{F}$-affine space.

Problem 11.1 Any set with 2 elements is an affine space over $\mathbb{Z}_2$ in the unique way. However, for 3 elements, there does not have a unique affine space structure over $\mathbb{Z}_3$.

The model one of the $\mathbb{A}$ is $\mathbb{A}^n_{\mathbb{F}} := \{(x_1, \cdots, x_n) \mid x_i \in \mathbb{F}\}$. Then the group action is:

$$\begin{aligned}(\mathbb{F}^n, +) \times \mathbb{A}^n_{\mathbb{F}} &\to \mathbb{A}^n_{\mathbb{F}} \\ (\vec{v}, \vec{x}) &\mapsto \vec{v} + \vec{x} := (v_1 + x_1, \cdots, v_n + x_n)\end{aligned}$$

for all $\vec{v} = (v_1, \cdots, v_n) \in \mathbb{F}^n$ and $\vec{x} = (x_1, \cdots, x_n) \in \mathbb{A}^n_{\mathbb{F}}$. Moreover, up to isomorphism, there is only one affine space of dimension n over $\mathbb{F}$.

Similarly, we have the following conversion table:

Vector Space	Affine Space
Linear Combinations	Affine Combinations
Basis	Affine Frame
Span	Affine Span/Hull
Subspace	Affine Subspace
Linear Map	Affine Map
Linear Independence	Affine Independence
Vectors	Points

For the affine combinations, we have:

$$p_0, p_1, \cdots, p_k \in \mathbb{A}, \quad c^0, c^1, \cdots, c^k \in \mathbb{F}$$

with $\sum_i c_i = 1$, then the affine combination is defined as:

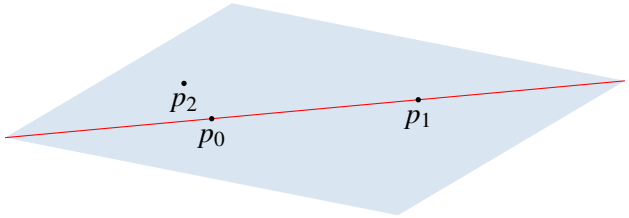
$$\sum_i c^i p_i := O + \sum_i c^i (p_i - O)$$

for some $O \in \mathbb{A}$ and $c^i(p_i - O)$ is the linear combination in the vector space $\mathcal{T}_O\mathbb{A}$. Note that we may have different O, let say O'. We may check the independence of the choice of O: We know

that $O' = O + (O' - O)$, then we have:

$$\begin{aligned} c^i p_i &= O' + \sum_i c^i (p_i - O') \\ &= O + (O' - O) + \sum_i c^i (p_i - O') \\ &= O + \sum_i c^i (O' - O) + \sum_i c^i (p_i - O') \\ &= O + \sum_i c^i ((O' - O) + (p_i - O')) \\ &= O + \sum_i c^i (p_i - O) \\ &= c^i p_i \end{aligned}$$

For affine subspaces and spans, we consider the following diagram:



The red line is the smallest affine subspace containing p_0 and $p_1 2$, i.e., the affine span of p_0 and p_1. We may write $\mathsf{Span}\{p_0, p_1\} := \{c^0 p_0 + c^1 p_1 \mid c^0 + c^1 = 1, c^i \in \mathbb{R}\} = \{t p_0 + (1-t) p_1 \mid t \in \mathbb{R}\}$. Note that $\overline{p_0 p_1} = \{t p_0 + (1-t) p_1 \mid t \in [0,1]\}$ is a subset of the affine span.

For the affine frame, we may consider the same picture above. Then $\{p_0, p_1, p_2\}$ is an affine frame of the affine space (the plane) as no point is in the affine span of the other two points.

For the representation of the affine map, we have the following commutative diagram:

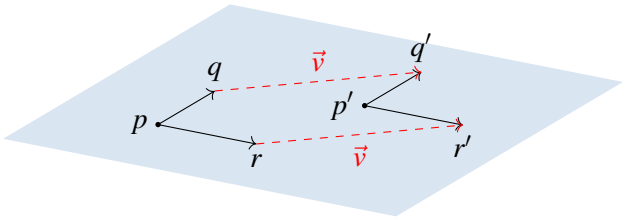
$$\begin{array}{ccc} \mathbb{A}_1 & \xrightarrow{\phi} & \mathbb{A}_2 \\ \downarrow & & \downarrow \\ \mathbb{A}^n_{\mathbb{F}} & \longrightarrow & \mathbb{A}^m_{\mathbb{F}} \\ \downarrow & & \downarrow \\ \mathbb{F}^n & \xrightarrow{A} & \mathbb{F}^m \\ \vec{x} = x - 0 & \longmapsto & A\vec{x} + \vec{b} \end{array}$$

where $A \in \mathsf{M}_{m \times n}(\mathbb{F})$ and $\vec{b} \in \mathbb{F}^m$. Note that the representation of ϕ depends on the choice of origins in $\mathbb{A}_1$ and $\mathbb{A}_2$.

A Euclidean space is a finite-dimensional real affine space with a Euclidean structure on its tangent space. The Euclidean structure means the translation invariant assignment of inner product to each tangent space of $\mathbb{A}$. Let $\mathbb{A}$ be an n-dimensional real affine space. Take $p \in \mathbb{A}$. Then the pointed affine space $(\mathbb{A}, p)$ is isomorphic to the vector space $\mathcal{T}_p \mathbb{A}$. Moreover, it is equivalent to $\mathbb{R}^n$ with the standard inner product, and $q \in (\mathbb{A}, p)$ corresponds to the vector $\vec{v} = q - p \in \mathbb{R}^n$. Then we have:

$$\alpha_1 q_1 + \alpha_2 q_2 = p + \alpha_1 (q_1 - p) + \alpha_2 (q_2 - p)$$

Note that $\alpha_1 + \alpha_2$ need not be 1 here, as it is linear combination. Then the translation invariant means that the length and angle remains unchanged in the inner product after translation, i.e., $\langle \vec{pq}, \vec{pr} \rangle = \langle \vec{p'q'}, \vec{p'r'} \rangle$.



Then $q = q' + \vec{v}$ and $r = r' + \vec{v}$. Note that $\mathcal{T}_p \mathbb{A}$ is different from $\mathcal{T}_{p'} \mathbb{A}$ as they are tangent spaces at different points, but they are isomorphic via translation by $\vec{v}$. We may consider the tangent line on the circle at different points as an example.

Up to isomorphism, there is only one Euclidean space of dimension n, denoted by $\mathbb{E}^n := (\mathbb{A}^n_{\mathbb{R}}, \langle \cdot, \cdot \rangle)$ where $\langle \cdot, \cdot \rangle$ is:

$$\langle \vec{pq}, \vec{pr} \rangle = (q - p) \cdot (r - p)$$

where the $\cdot$ is the standard dot product on $\mathbb{R}^n$. This is equivalent to say that an orthogonal frame exists, i.e., the rectangular coordinate system.

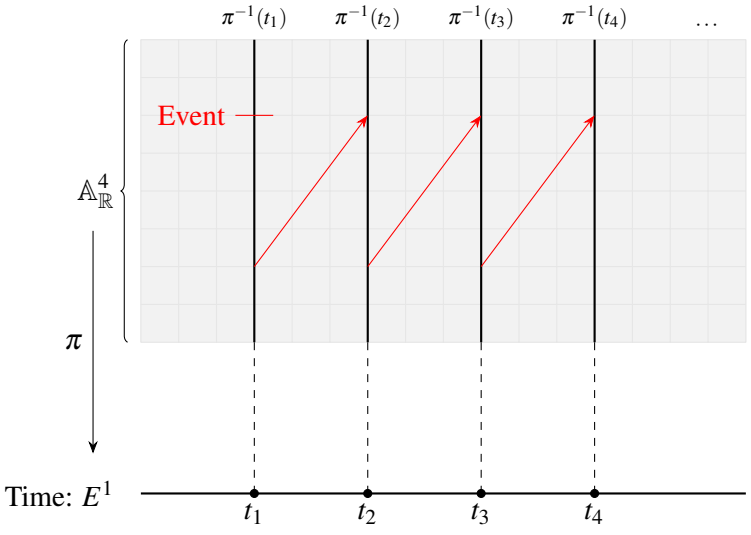
For an affine map $\phi : \mathbb{A}_1 \to \mathbb{A}_2$ between two affine spaces, we say that ϕ is injective implies that $\dim \mathbb{A}_1 \leq \dim \mathbb{A}_2$. The proof is by picking a point $p_1 \in \mathbb{A}_1$ and take $p_2 = \phi(p_1)$. Then we have the following commutative diagram:

$$\begin{array}{ccc} (\mathbb{A}_1, p_1) & \xrightarrow{\mathcal{T}_{p_1}\phi} & (\mathbb{A}_2, p_2) \\ \cong \downarrow & & \downarrow \cong \\ \mathbb{A}_1 & \xrightarrow{\phi} & \mathbb{A}_2 \end{array}$$

We have two space-time affine space in Physics, namely Minkowski and Galilean.

The Minkowski space-time $\mathbb{M}$ is a 4-dimensional real affine space $\mathbb{A}^4_{\mathbb{R}}$ with a Lorentz structure. Take a point $p \in \mathbb{A}^4_{\mathbb{R}}$ and $u = (u_0, \vec{u}), v = (v_0, \vec{v}) \in \mathbb{R}^4$. Then the Lorentzian inner product is $\langle u, v \rangle_p = u_0 v_0 - \vec{u} \cdot \vec{v}$.

The Galilean space-time $\mathbb{G}$ is a 4-dimensional real affine space $\mathbb{A}^4_{\mathbb{R}}$ with a Galilean structure. It is the Minkowski space-time taking the limit of light speed $c \to \infty$. We have the following diagram:



11.4 Quadratic Form and Clifford Algebra

Let V be a vector space over a field $\mathbb{F}$. A quadratic form on V is a map $q : V \to \mathbb{F}$ such that:

- $q(\alpha v) = \alpha^2 q(v)$ for all $\alpha \in \mathbb{F}$ and $v \in V$;
- The map $B : V \times V \to \mathbb{F}$ defined by $B(u,v) = q(u+v) - q(u) - q(v)$ is bilinear.

In case $\mathsf{char}(\mathbb{F}) \neq 2$, the set of all quadratic forms on V is equivalent to the set of all symmetric 2-forms on V. A quadratic form q can define a symmetric 2-form as $B(u,v) = \frac{1}{2}(q(u+v) - q(u) - q(v))$; a symmetric 2-form B can define a quadratic form $q(u) := B(u,u)$. We have the matrix representation of symmetric 2-form with respects to a basis. So we can also have the matrix representation of quadratic form, which is the symmetric matrices over $\mathbb{F}$ of order $\dim V = n$. Moreover, (V,q) forms a quadratic space.

> *Remark.* When $\mathsf{char}(\mathbb{F}) = 2$, we may define a symmetric bilinear form $B(u,v) = q(u+v) - q(u) - q(v)$. However, the quadratic form cannot be recovered from the symmetric bilinear form as $B(u,u) = 0$ for all $u \in V$, and so it is alternating. However, we can use a new bilinear form B', may not be symmetric, or even not unique, such that $q(u) = B'(u,u)$ for all $u \in V$.

A Clifford algebra $\mathsf{Cl}(V,q) := \mathcal{T}^\bullet V / I_q$ is an associative algebra over $\mathbb{F}$ generated by $v \otimes v - q(v)1$ for all $v \in V$. The ideal is equivalent to the ideal generated by $u \otimes v + v \otimes u - 2B(u,v)1$ for all $u, v \in V$. Note that $\mathsf{Cl}(V,q)$ is $\mathbb{Z}/2$ graded algebra.

We have the following isomorphisms:

- $\mathsf{Cl}(\mathbb{R}^{0,1}) \cong \mathbb{C}$ as $\mathbb{R}$-algebras, where elements in $\mathsf{Cl}(\mathbb{R}^{0,1})$ are of the form $a + be_1$ with $e_1^2 = -1$;
- $\mathsf{Cl}(\mathbb{R}^{1,0}) \cong \mathbb{R} \oplus \mathbb{R}$, the split-complex number, where elements in $\mathsf{Cl}(\mathbb{R}^{1,0})$ are of the form $a + be_1$ with $e_1^2 = 1$;
- $\mathsf{Cl}(\mathbb{R}^{0,2}) \cong \mathbb{H}$, the quaternion, as $\mathbb{R}$-algebras, where elements in $\mathsf{Cl}(\mathbb{R}^{0,2})$ are of the form $a + be_1 + ce_2 + de_1e_2$ with $e_1^2 = e_2^2 = -1$ and $e_1e_2 = -e_2e_1$;
- $\mathsf{Cl}(\mathbb{R}^{1,1}) \cong \mathsf{M}_{2\times 2}(\mathbb{R})$, the split-quaternion, as $\mathbb{R}$-algebras, where elements in $\mathsf{Cl}(\mathbb{R}^{1,1})$ are of the form $a + be_1 + ce_2 + de_1e_2$ with $e_1^2 = 1$, $e_2^2 = -1$ and $e_1e_2 = -e_2e_1$;
- $\mathsf{Cl}(\mathbb{R}^{2,0}) \cong \mathsf{M}_{2\times 2}(\mathbb{R})$, the split-quaternion, as $\mathbb{R}$-algebras.

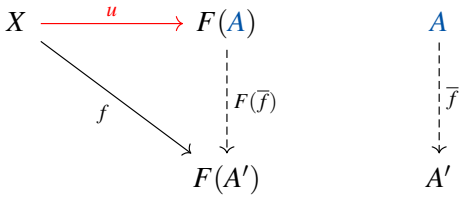
The $\mathbb{R}$, $\mathbb{C}$ and $\mathbb{H}$ are called the associative real division algebras.

This is the end of the main content. Thank you for your support! I hope you have enjoyed my notes!! By the way, I would like to remake this notes later after the final. If you are interested, find me through Discord `@stupidbenz`, or Instagram `@stupid.benz.0621`.

A. Universal Properties

We first state the formal definition of universal properties.

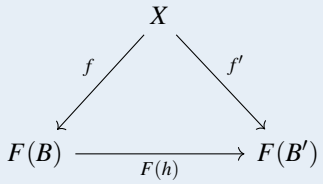
Definition A.1 — Universal Properties. Let $F : \mathcal{C} \to \mathcal{D}$ be a functor between two categories $\mathcal{C}$ and $\mathcal{D}$. A universal morphism from an object $X \in \mathsf{Ob}(\mathbb{D})$ to the functor F is a unique pair $(A, u : X \to F(A))$ in $\mathcal{D}$ such that for any morphism $f : X \to F(A')$ in $\mathcal{D}$, there exists a unique morphism $\overline{f} : A \to A'$ in $\mathcal{C}$ such that the following diagram commutes:



Such a property is called the *universal property* of the object A. Note that the dual version of universal morphism from F to X can be defined similarly.

Remark. Such an object A is an initial object in a new category:

- Objects: all pairs $(B, f : X \to F(B))$ for all $B \in \mathsf{Ob}(\mathcal{C})$;
- Morphisms: commutative diagrams in $\mathcal{C}$:



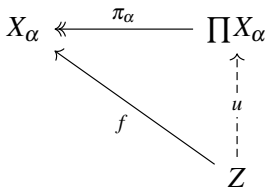
The initial object in this category is exactly the object A with the universal property. This type of construction is called the *comma category* and is denoted by $(X \downarrow F)$.

Similarly, the dual version of terminal object can be defined for the dual version of universal morphism, and the category is denoted by $(F \downarrow X)$.

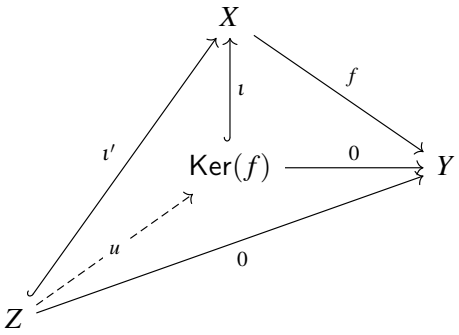
A.1 Universal Properties of Limits

The following are the universal properties of some common limits in category theory.

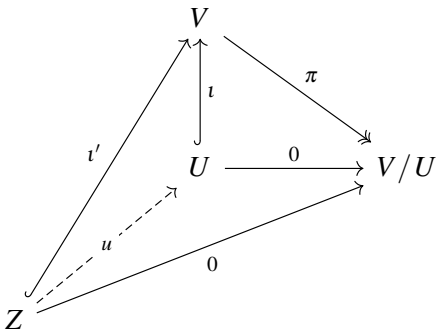
- **Products:**



- **Kernel:**



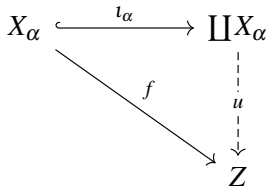
- **Subspaces:**



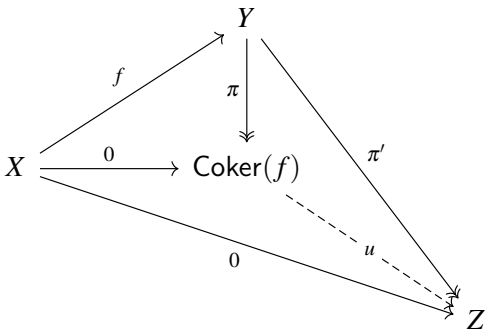
A.2 Universal Properties of Colimits

The following are the universal properties of some common colimits in category theory.

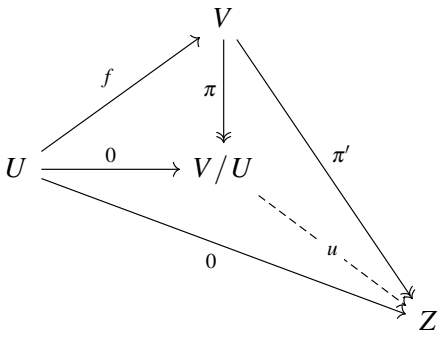
- **Coproducts:**



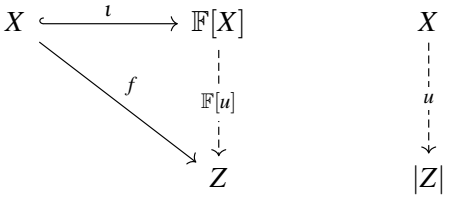
- **Cokernel:**



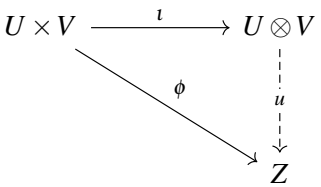
- **Quotient Spaces:**



- **Free Vector Spaces:**



- **Tensor Products:**



Bibliography

Websites

[1] A.Γ. *Why in the proof of* $A \cdot \mathrm{Adj}\,(A) = \det(A) \cdot I_n$ *entries not on the diagonal are zero?* 2015. URL: https://math.stackexchange.com/q/1404250 (cited on page 90).