Honors in Linear and Abstract Algebra I

Lecture Notes for
MATH 2131

Department of Mathematics
Hong Kong University of Science and Technology

January 17, 2026

Copyright Notice

Contents

Preface

This book is written by a student in the course MATH 2131 — Honors in Linear and Abstract Algebra I at The Hong Kong University of Science and Technology (HKUST) taught by Professor MENG Guowu during the Fall Semester of the Academic Year 2025–2026.

This book is designed to provide an abstract perspective on linear algebra. The book aims to give rigorous proofs of fundamental theorems in linear algebra while emphasizing the underlying structures and concepts. The book covers topics such as vector spaces, linear transformations, eigenvalues and eigenvectors, inner product spaces and more.

The target audience of this book includes undergraduate students studying linear algebra in a rigorous manner, as well as anyone interested in deepening their understanding of linear algebra from an abstract viewpoint. A solid foundation in basic linear algebra and mathematical proof techniques is recommended for readers.

CHAPTER 1

Linear Spaces

1.1. Introduction

Linear algebra originally arose from the study of systems of linear equations. Over time, it has evolved into a fundamental area of mathematics with applications in various fields such as physics, computer science, and economics. In this chapter, we will explore the concept of linear spaces, also known as vector spaces, which provide a framework for understanding linear combinations, subspaces, and linear transformations.

1.2. Operations and Structures

Before delving into linear spaces, it is essential to understand the basic operations and structures that underpin them.

1.2.1. Operations on Sets. There are several types of operations that can be performed on set S, including:

Definition 1.1 — Unary Operation.
A *unary operation* on a set S is a map

$$\begin{aligned} f : S &\to S \\ a &\mapsto f(a) \end{aligned}$$

Example 1.2.1. Common examples of unary operations include:

- Logical negation operation $\neg$ on the set $\{\text{true}, \text{false}\}$;
- Numeric negation operation $-$ on the set of real numbers $\mathbb{R}$;
- Complex conjugation operation $\overline{z}$ on the set of complex numbers $\mathbb{C}$.

Definition 1.2 — Binary Operation.
A *binary operation* on a set S is a map

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

Example 1.2.2. A common example of a binary operation is the addition operation $+$ on the set of natural numbers $\mathbb{N}$ which assigns to each pair of natural numbers (a, b) their sum $a + b$.

1.2.2. Properties of Binary Operations. There are several properties that binary operations may satisfy:

Definition 1.3 — Associative.
A binary operation $\cdot$ on a set S is *associative* if for all $a, b, c \in S$, we have
$$(a \cdot b) \cdot c = a \cdot (b \cdot c)$$

Example 1.2.3. The addition operation $+$ on the set of natural numbers $\mathbb{N}$ is associative since for all $a, b, c \in \mathbb{N}$, we have
$$(a + b) + c = a + (b + c)$$

Definition 1.4 — Unital.
A binary operation $\cdot$ on a set S is *unital* if there exists an element $e \in S$ such that for all $a \in S$, we have
$$e \cdot a = a = a \cdot e$$

Example 1.2.4. The multiplication operation $\cdot$ on the set of natural numbers $\mathbb{N}$ is unital with the identity element 1 since for all $a \in \mathbb{N}$, we have
$$1 \cdot a = a = a \cdot 1$$

Remark. Such an element e must be unique if it exists and is called the two-sided *identity element* of the operation. To see why, suppose there are two identity elements e and e'. Then we have
$$e = e \cdot e' = e'$$
Note that one-sided identity elements (left or right) may not be unique.

Definition 1.5 — Invertible.
A binary operation $\cdot$ on a set S with identity element e is *invertible* if for each $a \in S$, there exists an element $b \in S$ such that
$$a \cdot b = e = b \cdot a$$

Remark. Note that invertibility requires the existence of an identity element.

Example 1.2.5. The addition operation $+$ on the set of integers $\mathbb{Z}$ is invertible since for each integer $a \in \mathbb{Z}$, there exists an integer $-a \in \mathbb{Z}$ such that
$$a + (-a) = 0 = (-a) + a$$

Remark. Such an element b must be unique if it exists and is called the two-sided *inverse* of the element a, denoted by a^{-1}. To see why, suppose there are two inverses b and b'. Then we have
$$b = e \cdot b = (a \cdot b') \cdot b = a \cdot (b' \cdot b) = a \cdot e = b'$$
Note that one-sided inverses (left or right) may not be unique.

Definition 1.6 — Commutative.
A binary operation $+$ on a set S is *commutative* if for all $a, b \in S$, we have
$$a + b = b + a$$

Example 1.2.6. The addition operation $+$ on the set of natural numbers $\mathbb{N}$ is commutative since for all $a, b \in \mathbb{N}$, we have
$$a + b = b + a$$

Definition 1.7 — Distributive.
A binary operation $\cdot$ on a set S is *distributive* over another binary operation $+$ on S if for all $a, b, c \in S$, we have

$$a \cdot (b + c) = (a \cdot b) + (a \cdot c)$$

and

$$(a + b) \cdot c = (a \cdot c) + (b \cdot c)$$

The professor prefers to use the term "harmonic" instead of "distributive". Note that it is important to specify the order of the operations when discussing distributivity, as the two operations may not be commutative with each other.

Example 1.2.7. The multiplication operation $\cdot$ on the set of integers $\mathbb{Z}$ is distributive over the addition operation $+$ since for all $a, b, c \in \mathbb{Z}$, we have

$$a \cdot (b + c) = (a \cdot b) + (a \cdot c)$$

and

$$(a + b) \cdot c = (a \cdot c) + (b \cdot c)$$

1.2.3. Algebraic Structures. Most objects in mathematics can be described with the following template.

A _____ *is a set with a* _____ *structure on it.*

Some common algebraic structures include:

Definition 1.8 — Monoidic Structure.
A *monoidic structure* on a set M is a binary operation $\cdot$ that is associative and unital. The pair $(M, \cdot)$ is called a *monoid*.

Definition 1.9 — Groupic Structure.
A *groupic structure* on a set G is a binary operation $\cdot$ that is associative, unital, and invertible. The pair $(G, \cdot)$ is called a *group*.

Example 1.2.8. The pair $(\mathbb{R} \setminus \{0\}, \times)$, where $\times$ is the multiplication operation on real numbers, forms a group since multiplication is associative, unital (with identity element 1), and invertible (with inverse element $a^{-1} = \frac{1}{a}$ for each $a \in \mathbb{R} \setminus \{0\}$). Note that $(\mathbb{R}, \times)$ is not a group since 0 does not have an inverse.

Definition 1.10 — Abelian Structure.
An *abelian structure* on a monoid or group $(A, +)$ is a binary operation $+$ that is also commutative. The pair $(A, +)$ is called an *abelian monoid* or *abelian group* respectively.

Example 1.2.9. The pair $(\mathbb{Z}, +)$, where $+$ is the addition operation on integers, forms an abelian group since addition is associative, unital (with identity element 0), invertible (with inverse element $-a$ for each $a \in \mathbb{Z}$), and commutative.

Definition 1.11 — Ringic Structure.
A *ringic structure* on a set R is two binary operations $+$ and $\cdot$ such that

- $(R, +)$ is an abelian group;

- $(R, \cdot)$ is a monoid; and
- the operation $\cdot$ is distributive over the operation $+$.

The triple $(R, +, \cdot)$ is called a *ring*.

Remark. In this book, we will only consider unital rings and refer to them simply as "rings".

Definition 1.12 — Commutative Ring.
A *commutative ring* is a ring $(R, +, \cdot)$ where the operation $\cdot$ is also commutative.

Definition 1.13 — Field.
A *field* is a commutative ring $(F, +, \cdot)$ where the operation $\cdot$ is also invertible on $F \setminus \{0\}$.

Example 1.2.10. The triples $(\mathbb{Q}, +, \times)$, $(\mathbb{R}, +, \times)$, and $(\mathbb{C}, +, \times)$, where $+$ is the addition operation and $\times$ is the multiplication operation on rational numbers, real numbers, and complex numbers respectively, all form fields.

Example 1.2.11 — Finite Field. The set $\mathbb{F}_2 = \mathbb{Z}/2\mathbb{Z} = \{0, 1\}$ with XOR as addition and AND as multiplication forms a field. More generally, for any prime number p, the set $\mathbb{F}_p = \mathbb{Z}/p\mathbb{Z} = \{0, 1, 2, \ldots, p-1\}$ with addition and multiplication defined modulo p forms a field.

1.3. Homomorphisms

In mathematics, a *homomorphism* is a structure-preserving map between two algebraic structures of the same type.

Definition 1.14 — Monoid Homomorphism.
A *monoid homomorphism* is a set map $\phi : M_1 \to M_2$ between two monoids $(M_1, \cdot)$ and $(M_2, \odot)$ which respects the monoidic structure, i.e., for all $a, b \in M_1$, we have

- $\phi(a \cdot b) = \phi(a) * \phi(b)$;
- $\phi(e_1) = e_2$, where e_1 and e_2 are the identity elements of M_1 and M_2 respectively.

Definition 1.15 — Group Homomorphism.
A *group homomorphism* is a set map $\phi : G_1 \to G_2$ between two groups $(G_1, \cdot)$ and $(G_2, \odot)$ which respects the groupic structure, i.e., for all $a, b \in G_1$, we have

- $\phi(a \cdot b) = \phi(a) * \phi(b)$;
- $\phi(e_1) = e_2$, where e_1 and e_2 are the identity elements of G_1 and G_2 respectively;
- $\phi(a^{-1}) = (\phi(a))^{-1}$.

Proposition 1.3.1. The second and third properties in the definition of group homomorphism are consequences of the first property.

Proof. Let $\phi : G_1 \to G_2$ be a group homomorphism satisfying the first property. For any $a \in G_1$, we have

$$\phi(a) = \phi(a \cdot e_1) = \phi(a) * \phi(e_1).$$

So $\phi(e_1)$ is the identity element of G_2, i.e., $\phi(e_1) = e_2$. Similarly, we have

$$e_2 = \phi(e_1) = \phi(a \cdot a^{-1}) = \phi(a) * \phi(a^{-1}).$$

Thus, $\phi(a^{-1})$ is the inverse of $\phi(a)$, i.e., $\phi(a^{-1}) = (\phi(a))^{-1}$. □

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 $\phi(e_1 \cdot e_1) = \phi(e_1) * \phi(e_1)$. This simplifies to $\phi(e_1) = \phi(e_1) * \phi(e_1)$, which does not necessarily imply that $\phi(e_1)$ is the identity element in M_2, i.e., $\phi(e_1) \neq e_2$. 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 1.16 — Ring Homomorphism.

A *ring homomorphism* is a set map $\phi : R_1 \to R_2$ between two rings $(R_1, +, \cdot)$ and $(R_2, \oplus, \odot)$ which respects the ringic structure, i.e., for all $a, b \in R_1$, we have

- $\phi(a + b) = \phi(a) \oplus \phi(b)$;
- $\phi(a \cdot b) = \phi(a) \odot \phi(b)$;
- $\phi(\mathrm{id}_{R_1}) = \mathrm{id}_{R_2}$, where id_{R_1} and id_{R_2} are the multiplicative identity elements of R_1 and R_2 respectively.

Remark. Originally, there are 6 properties in the definition of ring homomorphism, including the preservation of additive identity, additive inverses and commutative property. However, it can be shown that these properties are consequences of the first property. Also, we do not include the trivial ring homomorphism, as it does not preserve the multiplicative identity.

On top of homomorphisms, we have special types of homomorphisms.

Definition 1.17 — Endomorphism.

An *endomorphism* is a homomorphism $\phi : A \to A$ from an algebraic structure to itself.

Definition 1.18 — Isomorphisms.

An *isomorphism* is a homomorphism $\phi : A \to B$ between two algebraic structures that has an inverse homomorphism $\phi^{-1} : B \to A$ such that $\phi \circ \phi^{-1} = \mathrm{id}_B$ and $\phi^{-1} \circ \phi = \mathrm{id}_A$.

Definition 1.19 — Automorphism.

An *automorphism* is an isomorphism $\phi : A \to A$ from an algebraic structure to itself.

Several maps can form a set as below.

Definition 1.20 — Homomorphism Set.

Given two algebraic structures A and B of the same type, the *homomorphism set* from A to B, denoted by $\mathrm{Hom}(A, B)$, is the set of all homomorphisms from A to B.

Definition 1.21 — Endomorphism Ring.
Given an abelian group $(G,+)$, the *endomorphism ring* of G, denoted by $\mathrm{End}(G)$, is the set of all endomorphisms from G to itself, equipped with the pointwise addition and composition of functions as the two binary operations. The two operations are defined as follows:

$$\begin{aligned} +\colon \mathrm{End}\, G \times \mathrm{End}\, G &\to \mathrm{End}\, G \\ (\phi,\psi) &\mapsto (\phi+\psi)\colon G \to G, \quad (\phi+\psi)(a) = \phi(a)+\psi(a) \\ \circ\colon \mathrm{End}\, G \times \mathrm{End}\, G &\to \mathrm{End}\, G \\ (\phi,\psi) &\mapsto (\phi\circ\psi)\colon G \to G, \quad (\phi\circ\psi)(a) = \phi(\psi(a)) \end{aligned}$$

The identity element for the addition operation is the zero map $0\colon G \to G$ defined by $0(a) = 0_G$ for all $a \in G$, where 0_G is the identity element of the group $(G,+)$. The identity element for the composition operation is the identity map $\mathrm{id}_G\colon G \to G$ defined by $\mathrm{id}_G(a) = a$ for all $a \in G$.

Remark. Endomorphisms in $\mathrm{End}(G)$ are group homomorphisms since $(G,+)$ is an abelian group. So $\mathrm{End}(G) = \mathrm{Hom}(G,G)$.

1.4. Linear Spaces

A linear space, or vector space, is a set with a linear structure defined over a field. We then need to define what a linear structure is.

Definition 1.22 — Linear Structure.
A *linear structure* on a set V over a field F is a pair of binary operations $(+,\cdot)$ where $(V,+)$ is an abelian group with a ring action $\cdot$ of F on $(V,+)$. A ring action of F on $(V,+)$ is equivalent to a ring homomorphism

$$\begin{aligned} \cdot\colon F &\to \mathrm{End}(V) \\ \alpha &\mapsto \alpha\cdot\colon V \to V, \quad (\alpha\cdot)(v) = \alpha\cdot v \end{aligned}$$

Remark. The actual definition of a ring action of F over $(V,+)$ is a map

$$\begin{aligned} \cdot\colon F \times V &\to V \\ (\alpha, v) &\mapsto \alpha\cdot v \end{aligned}$$

such that it satisfies the following four properties for all $\alpha,\beta \in F$ and $u,v \in V$:

- Distributivity over vector addition: $\alpha\cdot(u+v) = \alpha\cdot u + \alpha\cdot v$;
- Distributivity over field addition: $(\alpha+\beta)\cdot v = \alpha\cdot v + \beta\cdot v$;
- Compatibility: $(\alpha\beta)\cdot v = \alpha\cdot(\beta\cdot v)$;
- Unital: $1_F\cdot v = v$, where 1_F is the multiplicative identity element of the field F.

In usual textbooks, there are 8 axioms in the definition of linear structure. For all $\alpha,\beta \in F$ and $u,v \in V$:

1. Addition is associative: $(u+v)+w = u+(v+w)$;
2. Addition is unital: there exists an element $0_V \in V$ such that $0_V + v = v = v + 0_V$;
3. Addition is invertible: for each $v \in V$, there exists an element $-v \in V$ such that $v+(-v) = 0_V = (-v)+v$;
4. Addition is commutative: $u+v = v+u$;

5. Distributivity over vector addition: $\alpha \cdot (u+v) = \alpha \cdot u + \alpha \cdot v$;
6. Distributivity over field addition: $(\alpha+\beta)\cdot v = \alpha \cdot v + \beta \cdot v$;
7. Compatibility of scalar multiplication: $(\alpha\beta)\cdot v = \alpha \cdot (\beta \cdot v)$;
8. Identity element of scalar multiplication: $1_F \cdot v = v$, where 1_F is the multiplicative identity element of the field F.

The first four axioms ensure that $(V,+)$ is an abelian group. The fifth axiom describes the distributivity inside $\mathsf{End}(V)$, while the last three axioms corresponds to the properties of ring homomorphism from F to $\mathsf{End}(V)$. Thus, the 8 axioms can be reduced to the 2 conditions in the definition of linear structure.

Example 1.4.1. The field F itself can be considered as a linear space over F with the usual addition and multiplication operations. Here, the set V is F, the addition operation $+$ is the field addition, and the scalar multiplication $\cdot$ is the field multiplication.

Example 1.4.2. The set of all F-valued functions defined on a non-empty set X, i.e., $\{f : X \to F\}$, denoted by $\mathsf{Map}(X,F)$ or F^X, forms a linear space over F with the following operations:

$$\begin{aligned} +\colon \mathsf{Map}(X,F)\times\mathsf{Map}(X,F) &\to \mathsf{Map}(X,F) \\ (f,g) &\mapsto (f+g)\colon X\to F, \quad (f+g)(x) = f(x)+g(x) \\ \cdot\colon F\times\mathsf{Map}(X,F) &\to \mathsf{Map}(X,F) \\ (\alpha,f) &\mapsto (\alpha\cdot f)\colon X\to F, \quad (\alpha\cdot f)(x) = \alpha\cdot f(x) \end{aligned}$$

Remark. In fact, as long as the codomain is a linear space, the set of all functions from a non-empty set to that codomain forms a linear space with pointwise addition and scalar multiplication.

Example 1.4.3. The set of all finitely supported F-valued functions defined on a non-empty set X, i.e., $\{f : X \to F \mid f(x) \neq 0_F \text{ for only finitely many } x \in X\}$, denoted by $F[X]$ or $\mathsf{Map}_{\mathsf{fin}}(X,F)$ or $F^{(X)}$, forms a linear space over F with the same operations as in the previous example.

Example 1.4.4. The formal power series ring $F[[x]]$ over F forms a linear space over F with the usual addition and multiplication operations on formal power series. Formal means that we treat the elements as symbols without considering their convergence.

Example 1.4.5. The polynomial ring $F[x]$ over F forms a linear space over F with the usual addition and multiplication operations on polynomials.

Example 1.4.6. The set of all *column vectors* with n entries from F, denoted by F^n, forms a linear space over F with the operations defined entrywisely.

$$\begin{aligned} +\colon F^n\times F^n &\to F^n \\ (\vec{u},\vec{v}) &\mapsto \vec{u}+\vec{v} \\ \left(\begin{bmatrix}u^1\\ \vdots\\ u^n\end{bmatrix},\begin{bmatrix}v^1\\ \vdots\\ v^n\end{bmatrix}\right) &\mapsto \begin{bmatrix}u^1+v^1\\ \vdots\\ u^n+v^n\end{bmatrix} \end{aligned} \qquad \begin{aligned} \cdot\colon F\times F^n &\to F^n \\ (\alpha,\vec{v}) &\mapsto \alpha\cdot\vec{v} \\ \left(\alpha,\begin{bmatrix}v^1\\ \vdots\\ v^n\end{bmatrix}\right) &\mapsto \begin{bmatrix}\alpha\cdot v^1\\ \vdots\\ \alpha\cdot v^n\end{bmatrix} \end{aligned}$$

Remark. Here, we use superscripts to denote the entries of a column matrix due to the elements in vectors are *contravariant*. That is, when we change the basis, the coordinates of the vectors change in

the opposite way compared to the basis transformation. This is in contrast to *covariant* elements, such as the entries of row matrices (or covectors), which change in the same way as the basis transformation. We will discuss covariance and contravariance in Chapter 4.

Example 1.4.7. The set of all *matrices* with m rows and n columns from F, denoted by $\mathrm{Mat}_{m\times n}(F)$, forms a linear space over F with the operations defined entrywisely.

1.5. Linear Subspaces, Linear Combinations and Linear Span

Definition 1.23 — Linear Subspace.
A *linear subspace* of a linear space $(V,+,\cdot)$ over F is a non-empty subset $W \subseteq V$ with the operations $+$ and $\cdot$ inherited from V such that $(W,+,\cdot)$ is also a linear space over F.

Proposition 1.5.1. W is a linear subspace of V if and only if W is non-empty and closed under the operations $+$ and $\cdot$, i.e., for all $u, v \in W$ and $\alpha \in F$, we have

- $u + v \in W$;
- $\alpha \cdot v \in W$.

Proof. If W is a linear subspace of V, then by definition V is non-empty, as it contains the zero vector. Also, since $(W,+,\cdot)$ is a linear space, it must be closed under the operations $+$ and $\cdot$.
If W is non-empty and closed under the operations $+$ and $\cdot$, then we can easily verify that $(W,+,\cdot)$ satisfies all the axioms of a linear space over F. It is left as an exercise to the reader to check the axioms. □

We can actually combine two propoties into one by considering linear combinations.

Definition 1.24 — Linear Combination.
A *linear combination* of vectors $v_1, v_2, \ldots, v_n$ in a linear space V over F is any vector of the form

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

where $\alpha^1, \alpha^2, \ldots, \alpha^n$ are scalars in F.

To use linear combinations showing the condition for linear subspaces, we can consider the following example. We normally use $n = 2$ to proof the condition, and the general case can be proved by induction.

Proposition 1.5.2. The intersection of any collection of linear subspaces of a linear space V over F is also a linear subspace of V.

Proof. Let $\{W_i\}_{i\in I}$ be a collection of linear subspaces of V, where I is an index set. Define

$$W = \bigcap_{i\in I} W_i.$$

Then we have to show that W is a linear subspace of V. For any $i \in I$, we have $0_V \in W_i$ since W_i is a linear space. Thus, $0_V \in W$, so W is non-empty. Then, for any $u, v \in W$ and $\alpha, \beta \in F$, we have $u, v \in W_i$ for all $i \in I$. Since each W_i is a linear space, we have $\alpha u + \beta v \in W_i$ for all $i \in I$. Thus, $\alpha u + \beta v \in W$. Therefore, W is closed under the operations $+$ and $\cdot$. By the previous proposition, W is a linear subspace of V. □

Then it is natural to ask: the union of any collection of linear subspaces of a linear space V over F is also a linear subspace of V? The answer is no in general. However, if we perform "completion", or technically taking the *linear span*, we can get a linear subspace again and it is called the *sum* of those linear subspaces.

Definition 1.25 — Linear Span.

The *linear span* of a subset S of a linear space V over F, denoted by $\text{span}_F(S)$ or simply $\text{span}(S)$, $\overline{S}$ or $\langle S\rangle$, is the completion of S inside V under linear combinations, which is

$$\text{span}(S) = \left\{ \sum_{i=1}^{|S|} \alpha^i s_i \;\middle|\; \alpha^i \in F, s_i \in S \right\}$$

where $|S|$ is the cardinality of the set S (if S is infinite, we only consider finite linear combinations). Equivalently, the linear span of S is the smallest linear subspace of V that contains S. It can be written as

$$\text{span}(S) = \bigcap_{i\in I} W_i \subseteq V,$$

where $\{W_i\}_{i\in I}$ is the collection of all linear subspaces of W that contain S.

Definition 1.26 — Linear Spanning Set.

A subset S of a linear space V over F is a *linear spanning set*, or *linear generating set*, of V if its linear span is equal to V, i.e., $\text{span}\, S = V$.

For simplicity, we may omit the word "linear" when there is no ambiguity.

Example 1.5.1. Consider the linear space F^3 with vectors $\vec{e}_1$, $\vec{e}_2$, and $\vec{e}_3$. The set $S = \{\vec{e}_1, \vec{e}_2, \vec{e}_1+\vec{e}_2\}$ is not a spanning set of F^3 since $\text{span}(S)$ is the same as $\text{span}\{\vec{e}_1, \vec{e}_2\}$. However, the set $T = \{\vec{e}_1, \vec{e}_1+\vec{e}_2, \vec{e}_1+\vec{e}_2, \vec{e}_3\}$ is a spanning set of F^3 since $\text{span}(T) = F^3$.

Remark. If you have learnt linear algebra before, consider the matrix whose columns are the vectors in a spanning set, then the matrix must have full row rank.

Example 1.5.2. Consider the subset $S = \{1, x, x^2, \cdots\} \subset F[[x]]$. The linear span of S is the polynomial ring $F[x]$, i.e., $\text{span}(S) = F[x]$. The reason is that any polynomial can be expressed as a finite linear combination of the elements in S, while any formal power series that is not a polynomial cannot be expressed as such.

Definition 1.27 — Minimal Spanning Set.

A minimal spanning set of a linear space V over F is a spanning set S of V such that for any proper subset $S' \subset S$, we have $\text{span}(S') \subset \text{span}(S) = V$.

Remark. An ordered minimal spanning set is called a *basis*, which will use a round bracket notation, e.g., $(v_1, v_2, \ldots, v_n)$.

1.6. Linear Independence

Definition 1.28 — Linear Independence.

The non-trivial subspaces $W_1, W_2, \ldots, W_n$ of a linear space V over F are *linearly independent* if there is one and only one way to express the zero vector 0_V as a linear combination of vectors from these subspaces, i.e., if

$$w_1 + w_2 + \cdots + w_n = 0_V,$$

where $w_i \in W_i$ for each $i = 1, 2, \ldots, n$, then we must have $w_1 = w_2 = \cdots = w_n = 0_V$.

We also have a slightly weaker version of linear independence for future discussions.

Definition 1.29 — Weakly Linear Independence.
The subspaces $W_1, W_2, \ldots, W_n$ of a linear space V over F are *weakly linearly independent* if the only way to express the zero vector 0_V as a linear combination of vectors from these subspaces is the trivial way, i.e., if

$$w_1 + w_2 + \cdots + w_n = 0_V,$$

where $w_i \in W_i$ for each $i = 1, 2, \ldots, n$, then we must have $w_1 = w_2 = \cdots = w_n = 0_V$.

Remark. Weak linear independence allows subspaces to be trivial, i.e., equal to $\{0_V\}$.

Definition 1.30 — Linearly Independent Set.
A subset S of a linear space V over F is linearly independent if and only if there is only one way to express the zero vector 0_V as a linear combination of vectors from S, i.e., if

$$\alpha^1 s_1 + \alpha^2 s_2 + \cdots + \alpha^n s_n = 0_V,$$

where $s_i \in S$ and $\alpha^i \in F$ for each $i = 1, 2, \ldots, n$, then we must have $\alpha^1 = \alpha^2 = \cdots = \alpha^n = 0_F$.

Remark. Equivalently, a subset S of a linear space V over F is linearly independent if no elements in S can be expressed as a linear combination of other elements in S.

Also, similar to minimal spanning sets, we have the following definition.

Definition 1.31 — Maximal Linearly Independent Set.
A maximal linearly independent set of a linear space V over F is a linearly independent set S of V such that for any proper superset $S' \supset S$, we have S' is not linearly independent.

Remark. A minimal spanning set and a maximal linearly independent set describe the same concept. We will use minimal spanning sets in this book.

Example 1.6.1. Let X be a non-empty set. For each $x \in X$, define the Kronecker delta function $\delta_x : X \to F$ by

$$\delta_x(t) = \begin{cases} 1, & \text{if } t = x; \\ 0, & \text{if } t \neq x. \end{cases} \tag{1}$$

Clearly, δ_x is in $F[X]$ since it is finitely supported. The set ${}^{\delta}\mathrm{X} = \{\delta_x \mid x \in X\}$ is a linearly independent set in the linear space $F[X]$ over F. To show this, assume there exists a finite linear combination of other delta functions such that $\delta_x = \sum \alpha^t \delta_t$. Then we have $\delta_x(x) = 1$ and $\delta_x(x) = \sum \alpha^t \delta_t(x) = 0$, which shows a contradition. Moreover, it is a minimal spanning set of $F[X]$ since any finitely-supported function can be expressed as a finite linear combination of the functions in this set.

1.7. Sum and Direct Sum of Linear Subspaces

As we have mentioned before, the linear span of the union of several linear subspaces is again a linear subspace, which is called the sum of those linear subspaces.

Definition 1.32 — Sum of Linear Subspaces.
The *sum* of the linear subspaces $W_1, W_2, \ldots, W_n$ of a linear space V over F, denoted by $W_1 + W_2 + \cdots + W_n$,

is the linear span of their union, i.e.,

$$W_1 + W_2 + \cdots + W_n = \text{span}(W_1 \cup W_2 \cup \cdots \cup W_n).$$

Equivalently, the sum can be expressed as

$$W_1 + W_2 + \cdots + W_n = \{w_1 + w_2 + \cdots + w_n \mid w_i \in W_i, i = 1, 2, \ldots, n\}.$$

Definition 1.33 — Internal Direct Sum of Linear Subspaces.

The *internal direct sum* of the linear subspaces $W_1, W_2, \ldots, W_n$ of a linear space V over F, denoted by $W_1 \oplus W_2 \oplus \cdots \oplus W_n$, is their sum $W_1 + W_2 + \cdots + W_n$ provided that the subspaces are weakly linearly independent, i.e.,

$$W_1 \oplus W_2 \oplus \cdots \oplus W_n = W_1 + W_2 + \cdots + W_n,$$

and for any $w \in W_1 \oplus W_2 \oplus \cdots \oplus W_n$, there exist unique vectors $w_i \in W_i$ for each $i = 1, 2, \ldots, n$ such that

$$w = w_1 + w_2 + \cdots + w_n.$$

The equivalent definition of internal direct sum is that the intersection of any subspace with the sum of the other subspaces is trivial, i.e., for each $i = 1, 2, \ldots, n$,

$$W_i \cap \left(\sum_{j \neq i} W_j \right) = \{0_V\}.$$

There is also an *external* version of direct sum which constructs a new linear space from several linear spaces. We normally use the symbol $=$ to denote internal direct sum decomposition such as $V = W_1 \oplus W_2 \oplus \cdots \oplus W_n$ and use the symbol $\cong$ to denote the external direct sum isomorphic to a linear space such as $V \cong W_1 \oplus W_2 \oplus \cdots \oplus W_n$.

Definition 1.34 — External Direct Sum of Linear Spaces.

The *external direct sum* of the linear spaces $W_1, W_2, \ldots, W_n$ over a field F, denoted by $W_1 \oplus W_2 \oplus \cdots \oplus W_n$, is the linear space defined as

$$W_1 \oplus W_2 \oplus \cdots \oplus W_n = \{(w_1, w_2, \ldots, w_n) \mid w_i \in W_i, i = 1, 2, \ldots, n\},$$

with the vector addition and scalar multiplication defined componentwisely, i.e., for any $(w_1, w_2, \ldots, w_n), (w_1', w_2', \ldots, w_n') \in W_1 \oplus W_2 \oplus \cdots \oplus W_n$ and $\alpha \in F$,

$$(w_1, w_2, \ldots, w_n) + (w_1', w_2', \ldots, w_n') = (w_1 + w_1', w_2 + w_2', \ldots, w_n + w_n'),$$
$$\alpha \cdot (w_1, w_2, \ldots, w_n) = (\alpha \cdot w_1, \alpha \cdot w_2, \ldots, \alpha \cdot w_n).$$

For internal direct sum, $W_1 \oplus W_2 = \{w_1 + w_2 \mid w_i \in W_i, i = 1, 2\}$ with the vector addition and scalar multiplication inherited from V. For external direct sum, $W_1 \oplus W_2 = \{(w_1, w_2) \mid w_i \in W_i, i = 1, 2\}$ with the vector addition and scalar multiplication defined componentwisely.

1.8. Exercise

Problem 1.1. On the logic set $X = \{\text{true}, \text{false}\}$, we have two binary operations: one is "OR", denoted by $\vee$, and the other is "AND",denoted by $\wedge$. If we use 1 to represent "true" and 0 to represent "false", then

$$1 \vee 0 = 0 \vee 1 = 0 \vee 0 = 0, \qquad 1 \vee 1 = 1$$
$$1 \wedge 1 = 1 \wedge 0 = 0 \wedge 1 = 0, \qquad 0 \wedge 0 = 0$$

(a) Show that both $\vee$ and $\wedge$ are abelian monoid structures on X.

(b) Show that $\vee$ distributes with respect to $\wedge$.

(c) Show that $\wedge$ distributes with respect to $\vee$.

(d) Is $(X, \vee, \wedge)$ a ring? If not, can you modify $\vee$ to arrive at a new binary operation $\vee'$ such that $(X, \vee', \wedge)$ is a commutative ring with unity? If yes, is this ring a field?

Problem 1.2. Find a non-empty subset X of 2×2 matrices over $\mathbb{R}$ such that

– the set X is closed under matrix multiplication, and

– there are many left-identities, but there is no two-sided identity.

Problem 1.3. Let F be a field and X be a non-empty set. Recall that $\text{Map}(X, F)$ is the set of F-valued functions on X and $F[X]$ is the set of finitely-supported F-valued functions on X. Both $\text{Map}(X, F)$ and $F[X]$ are linear spaces over the field F.
Let $T : X \to Y$ be a set map and $T_* : F[X] \to F[Y]$ be the map such that

$$T_*(f)(y) = \sum_{x \in T^{-1}(y)} f(x), \qquad \forall y \in Y.$$

In case $T^{-1}(y)$ is the empty set $\emptyset$, the sum is assumed to be 0. Please check that the sum above is well-defined and $T_*(f)$ has a finite-support.

(a) Show that $(1_X)_* = 1_{F[X]}$ for all non-empty set X.

(b) Show that $(TS)_* = T_* S_*$ for all set maps T and S such that the composition TS is defined.

(c) For any set map $T : X \to Y$, we have an induced map $T^* : \text{Map}(Y, F) \to \text{Map}(X, F)$ via the formula $T^* f = f T$. Show that $1_X^* = 1_{\text{Map}(X,F)}$ and $(TS)^* = S^* T^*$.

(d) Can we get a natural map from $\text{Map}(X, F)$ to $\text{Map}(Y, F)$ or from $F[Y]$ to $F[X]$ for any set map $T : X \to Y$ between two infinite sets X and Y?

Problem 1.4. Let V be a linear space and S be a spanning set for V. Show that S is a minimal spanning set for $V \iff S$ is a linearly independent set. Note: S here is not required to be finite.

CHAPTER 2

Linear Maps and Matrices

Linear maps are fundamental objects in linear algebra. In this chapter, we will explore their definitions and properties.

2.1. Linear Maps, Kernel, and Image

2.1.1. Linear Maps. We begin by defining linear maps between linear spaces.

Definition 2.1 — Linear Map.
A *linear map*, or *linear transformation*, between two linear spaces V and W over the same field F is a set map $T : V \to W$ that respects the linear structure; that is, for all $u, v \in V$ and all scalars $\alpha \in F$, the following properties hold:

- $T(u+v) = T(u) + T(v)$;
- $T(\alpha \cdot u) = \alpha \cdot T(u)$.

Equivalently, for all $u, v \in V$ and all scalars $\alpha, \beta \in F$, we have

$$T(\alpha \cdot u + \beta \cdot v) = \alpha \cdot T(u) + \beta \cdot T(v).$$

Remark. Originally, linear maps required 8 properties to be satisfied. However, it can be shown easily that these two properties imply the rest.

For simplicity, we often write Tu instead of $T(u)$ for the image of a vector u under the linear map T. The set of all linear maps from V to W is denoted by $\mathrm{Hom}_F(V, W)$ or simply $\mathrm{Hom}(V, W)$ when the field is clear from context. Some author use $\mathcal{L}(V, W)$ instead.

From Example 1.4.2, we know that $\mathrm{Map}(V, W)$ forms a linear space over F with pointwise addition and scalar multiplication. Then $\mathrm{Hom}(V, W)$ is a subset of $\mathrm{Map}(V, W)$. Moreover $\mathrm{Hom}(V, W)$ is actually a linear subspace of $\mathrm{Map}(V, W)$.

Proposition 2.1.1. The set $\mathrm{Hom}(V, W)$ of all linear maps from V to W forms a linear space over F with pointwise addition and scalar multiplication.

Proof. We need to show that $\mathrm{Hom}(V, W)$ is closed under pointwise addition and scalar multiplication. Let T and S be two linear maps from V to W. For all $u, v \in V$ and all $\alpha, \beta \in F$, we have

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

so $T + S$ is a linear map from V to W. □

2.1.2. Kernel and Image. In this section, we introduce two important concepts associated with linear maps: the kernel and the image.

Definition 2.2 — Kernel.
The *kernel* of a linear map $T: V \to W$ is the set of all vectors in V that are mapped to the zero vector in W:
$$\ker(T) = \{v \in V \mid T(v) = 0\}.$$

Proposition 2.1.2. The kernel of a linear map $T: V \to W$ is a linear subspace of V.

Proof. We need to show that $\ker(T)$ is closed under vector addition and scalar multiplication.
Let $u, v \in \ker(T)$. Then we have $T(u) = 0$ and $T(v) = 0$. For any scalar $\alpha \in F$, we have
$$T(u+v) = T(u) + T(v) = 0 + 0 = 0,$$
$$T(\alpha \cdot u) = \alpha \cdot T(u) = \alpha \cdot 0 = 0.$$
Therefore, $u + v \in \ker(T)$ and $\alpha \cdot u \in \ker(T)$, and so $\ker(T)$ is a linear subspace of V. □

Definition 2.3 — Image.
The *image* of a linear map $T: V \to W$ is the set of all vectors in W that can be expressed as $T(v)$ for some vector v in V:
$$\operatorname{im}(T) = \{w \in W \mid w = T(v) \text{ for some } v \in V\}.$$

Proposition 2.1.3. The image of a linear map $T: V \to W$ is a linear subspace of W.

Proof. We need to show that $\operatorname{im}(T)$ is closed under vector addition and scalar multiplication.
Let $w_1, w_2 \in \operatorname{im}(T)$. Then there exist vectors $v_1, v_2 \in V$ such that $w_1 = T(v_1)$ and $w_2 = T(v_2)$. For any scalar $\alpha \in F$, we have
$$w_1 + w_2 = T(v_1) + T(v_2) = T(v_1 + v_2),$$
$$\alpha \cdot w_1 = \alpha \cdot T(v_1) = T(\alpha \cdot v_1).$$
Therefore, $w_1 + w_2 \in \operatorname{im}(T)$ and $\alpha \cdot w_1 \in \operatorname{im}(T)$, and so $\operatorname{im}(T)$ is a linear subspace of W. □

2.2. Injection, Surjection, and Isomorphism

Definition 2.4 — Injective Linear Map.
A linear map $T: V \to W$ is *injective*, or a *monomorphism*, if for any $u, v \in V$, $T(u) = T(v)$ implies that $u = v$.

Exercise 2.2.1. Show that a linear map $T: V \to W$ is injective if and only if $\ker(T) = \{0\}$.

Definition 2.5 — Surjective Linear Map.
A linear map $T: V \to W$ is *surjective*, or an *epimorphism*, if for any $w \in W$, there exists a vector $v \in V$ such that $T(v) = w$.

Exercise 2.2.2. Show that a linear map $T: V \to W$ is surjective if and only if $\operatorname{im}(T) = W$.

Definition 2.6 — Linear Isomorphism.
A linear map $T: V \to W$ is a *linear isomorphism*, or an *isomorphism*, if T has an inverse map $T^{-1}: W \to V$ that is also a linear map, i.e., there exists a linear map $T^{-1}: W \to V$ such that $T^{-1} \circ T = \mathrm{id}_V$ and $T \circ T^{-1} = \mathrm{id}_W$. Then the linear spaces V and W are said to be *isomorphic*, denoted by $V \cong W$.

Remark. The professor prefers to use the term "linear equivalence" instead of "linear isomorphism", probably influenced by terminology in category theory and homotopy theory. However, the term "isomorphism" is more widely used in the literature, so we will stick to that in this book.

Proposition 2.2.1. A linear map $T: V \to W$ is a linear isomorphism if and only if it is both injective and surjective.

Proof.

($\Rightarrow$) : If T is a linear isomorphism, then there exists a linear map $T^{-1}: W \to V$ such that $T^{-1} \circ T = \mathrm{id}_V$ and $T \circ T^{-1} = \mathrm{id}_W$.

Injective : $T(u) = T(v) \implies T^{-1}(T(u)) = T^{-1}(T(v)) \implies u = v$ for any $u, v \in V$.

Surjective : For any $w \in W$, let $v = T^{-1}(w)$. Then we have $T(v) = T(T^{-1}(w)) = w$.

($\Leftarrow$) : If T is both injective and surjective, we can define the inverse map $T^{-1}: W \to V$ as follows: for any $w \in W$, since T is surjective, there exists a vector $v \in V$ such that $T(v) = w$. We define $T^{-1}(w) = v$. To show that T^{-1} is well-defined, suppose there are two vectors $v_1, v_2 \in V$ such that $T(v_1) = w$ and $T(v_2) = w$. Then we have $T(v_1) = T(v_2)$, which implies that $v_1 = v_2$ since T is injective. Therefore, T^{-1} is well-defined.

Invertible : for all $v \in V$, $w \in W$, we have $T^{-1}(T(v)) = v$ and $T(T^{-1}(w)) = w$.

Linearity : For any $w_1, w_2 \in W$ and any scalars $\alpha, \beta \in F$, let $v_1 = T^{-1}(w_1)$ and $v_2 = T^{-1}(w_2)$. Then we have

$$\begin{aligned} T^{-1}(\alpha \cdot w_1 + \beta \cdot w_2) &= T^{-1}(\alpha \cdot T(v_1) + \beta \cdot T(v_2)) \\ &= T^{-1}(T(\alpha \cdot v_1 + \beta \cdot v_2)) \\ &= \alpha \cdot v_1 + \beta \cdot v_2 \\ &= \alpha \cdot T^{-1}(w_1) + \beta \cdot T^{-1}(w_2). \end{aligned}$$

□

Example 2.2.1. The differential operator $D: F[x] \to F[x]$ is not an injective linear map as $D(1) = 0 = D(2)$ but it is a surjective linear map for F is a field of characteristic 0.

Definition 2.7 — Characteristic of a Field.

The *characteristic* of a field F is the smallest positive integer n such that

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

where 1 is the multiplicative identity in F. If no such positive integer exists, the characteristic of F is defined to be 0.

2.3. Dimension of Linear Spaces

Definition 2.8 — Finite-Dimensional Linear Space.

A linear space V over F is *finite-dimensional* if there exists an isomorphism $T: V \to F^n$ for some positive integer n. The integer n is the *dimension* of the linear space V, denoted by $\dim_F(V)$ or simply $\dim(V)$.

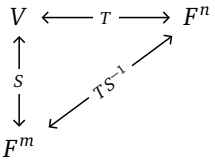
If a linear space is not finite-dimensional, it is called *infinite-dimensional*. Then we have to show that the dimension is well-defined.

Proposition 2.3.1. If there exists isomorphisms $T: V \to F^n$ and $S: V \to F^m$, then $n = m$.

Proof. Since S is an isomorphism, it has an inverse map $S^{-1}: F^m \to V$ that is also a linear map. Then the composition $TS^{-1}: F^m \to F^n$ is also a linear isomorphism. Mutatis mutandis for the opposite direction. Therefore, it suffices to show that if there exists a linear isomorphism $L: F^m \to F^n$, then $m = n$. □

Remark. Mutatis mutandis means "the necessary changes having been made" in Latin. Here it means that the argument for one direction is similar to the other direction with necessary changes.

This proposition also shows a key result in linear algebra: up to isomorphism, there is only one linear space of dimension n over a field F, which is F^n. We can also interpret the proposition by a commutative diagram:



Remark. In commutative diagrams, we use $V \hookrightarrow W$ to denote an injective map, $V \twoheadrightarrow W$ to denote a surjective map. In this book, we would use $V \xleftrightarrow{} W$ to denote an isomorphism.

There is a equivalent way to characterise linearly independent sets, spanning sets and minimal spanning sets using linear maps.

Exercise 2.3.1. Let V be an n-dimensional linear space, and $S = (v_1, \dots, v_k)$ be an ordered set of k vectors in V. Let $\phi_S: \mathbb{F}^k \to V$ be the linear map that sends $\vec{x} \in \mathbb{F}^k$ to $x^1 v_1 + \cdots + x^k v_k$. Show that

(1) S is a linearly independent set $\iff$ ϕ_S is injective.

(2) S is a spanning set for V $\iff$ ϕ_S is surjective.

(3) S is a minimal spanning set for V $\iff$ ϕ_S is invertible. Note: a minimal order spanning set is called a basis.

In case S is a basis, the inverse ϕ_S^{-1} is written as $[-]_S$.

Moreover, equivalently, we can characterise finite-dimensional linear spaces using spanning sets.

Proposition 2.3.2. A linear space V over F is finite-dimensional if and only if V is finitely generated, i.e., there is a finite spanning set for V.

Proof. If V is finite-dimensional, there exists an isomorphism $T: F^n \to V$ for some positive integer n. Then the set $\{T(\vec{e}_1), T(\vec{e}_2), \dots, T(\vec{e}_n)\}$, where $\vec{e}_i$ is the column vector with 1 in the i-th entry and 0 elsewhere, is a finite spanning set for V. However, it may not be linearly independent. Fortunately, we can always extract a minimal spanning set of V from it. Then, without the loss of generality, we can say $\{T(\vec{e}_1), T(\vec{e}_2), \dots, T(\vec{e}_k)\}$ for some $k \leq n$ is a minimal spanning set for V. Then by Exercise 2.3.1, the linear map $\phi_S: F^k \to V$ defined by $\phi_S(\vec{x}) = x^1 T(\vec{e}_1) + x^2 T(\vec{e}_2) + \cdots + x^k T(\vec{e}_k)$ is an isomorphism. Therefore, V is finitely generated. □

Moreover, we have the following dimension inequality.

Proposition 2.3.3. $\dim(V_1 + V_2) \leq \dim(V_1) + \dim(V_2)$ for any two finite-dimensional linear subspaces V_1 and V_2 of a linear space V. 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 $S_1 \cup S_2$ is a spanning set for $V_1 + V_2$. Indeed, for any vector $v \in V_1 + V_2$, there exist vectors $v_1 \in V_1$ and $v_2 \in V_2$ such that $v = v_1 + v_2$. Then we can express v_1 and v_2 as linear combinations of the vectors in S_1 and S_2 respectively. Therefore, v can be expressed as a linear combination of the vectors in $S_1 \cup S_2$. This shows that $V_1 + V_2 \subseteq \operatorname{span}(S_1 \cup S_2)$. The converse inclusion is trivial. Thus, we have $V_1 + V_2 = \operatorname{span}(S_1 \cup S_2)$.
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 linearly independent. Equality holds if and only if $S_1 \cup S_2$ is linearly independent, which is equivalent to the sum being direct. □

Then we can define the rank and nullity of a linear map.

Definition 2.9 — Rank.
The *rank* of a linear map $T : V \to W$ is the dimension of its image:

$$\operatorname{rank}(T) = \dim(\operatorname{im}(T)).$$

Definition 2.10 — Nullity.
The *nullity* of a linear map $T : V \to W$ is the dimension of its kernel:

$$\operatorname{nullity}(T) = \dim(\ker(T)).$$

2.4. Matrices

2.4.1. Matrix Representation of Linear Maps. Matrices provide a convenient way to represent linear maps between finite-dimensional linear spaces. Let A be an $m \times n$ matrix with entries from F. Then the map

$$F^n \to F^m$$
$$\vec{x} \mapsto A\vec{x}$$

is a linear map over F.

Proposition 2.4.1. Every linear map $T : F^n \to F^m$ can be represented as multiplication by a unique $m \times n$ matrix A over F. The matrix A is called the *standard matrix*, or the *matrix representation*, of the linear map T. There is an isomorphism between two linear spaces $\operatorname{Hom}(F^n, F^m)$ and $\operatorname{Mat}_{m\times n}(F)$. Then we have

– The standard matrix of the linear map T is given by

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

where $\vec{e}_i$ is the column vector with 1 in the i-th entry and 0 elsewhere.

– For any matrix A in $\operatorname{Mat}_{m\times n}(F)$, the corresponding linear map $T_A : F^n \to F^m$ is given by

$$T_A\vec{x} = A\vec{x}, \quad \text{for all } \vec{x} \in F^n.$$

Proof. Let $T : F^n \to F^m$ be a linear map. Define the matrix A as above. For any vector $\vec{x} \in F^n$, we can express $\vec{x}$ as a linear combination of $\vec{e}_1, \vec{e}_2, \ldots, \vec{e}_n$:

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

Then, using the linearity of T, we have

$$\begin{aligned} T\vec{x} &= T(x^1\vec{e}_1 + x^2\vec{e}_2 + \cdots + x^n\vec{e}_n) \\ &= x^1 T\vec{e}_1 + x^2 T\vec{e}_2 + \cdots + x^n T\vec{e}_n \\ &= A\vec{x}. \end{aligned}$$

This shows that $T\vec{x}$ can be computed as the matrix-vector product $A\vec{x}$. Conversely, given a matrix A in $\mathrm{Mat}_{m\times n}(F)$, we can define a linear map $T_A\colon F^n \to F^m$ by $T_A\vec{x} = A\vec{x}$. The linearity of T_A follows from the properties of matrix multiplication. □

Remark. Although it is an isomorphism, the correspondence between linear maps and their standard matrices depends on the choice of bases for the domain and codomain. So it is not a natural isomorphism. Natural means that the isomorphism does not depend on any choices.

As the linear combinations of vectors is clumsy to write, there is a simpler way to write it — Einstein summation notation. In this notation, we use an index to represent the components of a vector. For example, a vector $\vec{v}$ in F^n can be represented as v^i, where i runs from 1 to n. Then the linear combination

$$\sum_{i=1}^{n} v^i \vec{e}_i$$

can be written simply as $v^i\vec{e}_i$, where the summation over the repeated index i is implied.

The columns of the standard matrix A are vectors in F^m. Dually, we can also consider the rows of A as vectors in F^n. Let A be an $m \times n$ matrix with rows $\hat{a}^1, \hat{a}^2, \ldots, \hat{a}^m$ in $(F^n)^*$. Each row vector $\hat{a}^j$ is a *linear functional* on F^n, which is a linear map from F^n to F. Then the matrix-vector product $A\vec{x}$ can be expressed in terms of these linear functionals as

$$A\vec{x} = \begin{bmatrix} \hat{a}^1(\vec{x}) \\ \hat{a}^2(\vec{x}) \\ \vdots \\ \hat{a}^m(\vec{x}) \end{bmatrix}.$$

Definition 2.11 — Linear Functional.
A *linear functional*, or *covector*, on a linear space V over F is a linear map from V to F.

Example 2.4.1. Consider the differential operator $D\colon F[x] \to F[x]$ defined by

$$D\left(a_0 + a_1x + a_2x^2 + \cdots + a_nx^n\right) = a_1 + 2a_2x + 3a_3x^2 + \cdots + na_nx^{n-1}.$$

The standard matrix of D with respect to $\{1, x, x^2, \ldots, x^n\}$ is given by

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

The following definitions correspond to the definitions of kernel, image, rank and nullity of linear maps and isomorphisms respectively.

Definition 2.12 — Null Space.

The *null space* of an $m \times n$ matrix A over F is the set of all vectors in F^n that are mapped to the zero vector in F^m:

$$\text{null}(A) = \{\vec{x} \in F^n \mid A\vec{x} = 0\}.$$

Definition 2.13 — Column Space.

The *column space* of an $m \times n$ matrix A over F is the set of all vectors in F^m that can be expressed as $A\vec{x}$ for some vector $\vec{x}$ in F^n:

$$\text{col}(A) = \{\vec{y} \in F^m \mid \vec{y} = A\vec{x} \text{ for some } \vec{x} \in F^n\}.$$

Definition 2.14 — Rank.

The *rank* of an $m \times n$ matrix A over F is the dimension of its column space:

$$\text{rank}(A) = \dim(\text{col}(A)).$$

Definition 2.15 — Nullity.

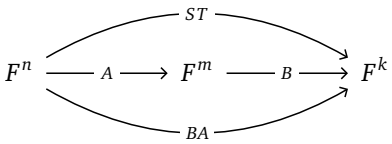
The *nullity* of an $m \times n$ matrix A over F is the dimension of its null space:

$$\text{nullity}(A) = \dim(\text{null}(A)).$$

Definition 2.16 — Invertible Matrix.

An $n \times n$ matrix A over F is *invertible*, or *nonsingular*, if A has an inverse matrix A^{-1} such that $AA^{-1} = I_n$ and $A^{-1}A = I_n$, where I_n is the $n \times n$ identity matrix.

2.4.2. Composition of Linear Maps and Matrix Multiplication. Consider two linear maps $T : F^n \to F^m$ and $S : F^m \to F^k$ with standard matrices A and B, respectively. Then we want to find the standard matrix of the composition $ST : F^n \to F^k$.



Proposition 2.4.2. The standard matrix of the composition $ST : F^n \to F^k$ is given by the matrix product BA, i.e., for any vector $\vec{x} \in F^n$, we have

$$(ST)(\vec{x}) = B(A\vec{x}) = (BA)\vec{x}.$$

Proof. For any vector $\vec{x} \in F^n$ with entries $x^1, x^2, \ldots, x^n$, we have

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

The j-th column of BA is given by

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

Therefore, the standard matrix of ST is BA. □

Remark. B is a $k \times m$ matrix and A is a $m \times n$ matrix. So the matrix product BA is defined and results in a $k \times n$ matrix.

The matrix multiplication BA can be computed as follows.

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

2.4.3. Elementary Row Operations and Elementary Column Operations. Elementary row operations are operations that can be performed on the rows of a matrix to transform it into a different form. There are three types of elementary row operations:

– Row swapping: $R_i \longleftrightarrow R_j$ (swap row i and row j)
– Row scaling: $R_i \leftarrow \alpha R_i$ (multiply row i by a non-zero scalar α)
– Row addition: $R_i \leftarrow R_i + \alpha R_j$ (add α times row j to row i)

Each elementary row operation is a *left multiplication* by an *elementary matrix*. An elementary matrix is obtained by performing a single elementary row operation or elementary column operation on an identity matrix. Moreover, every elementary matrix is invertible, and its inverse is also an elementary matrix.

We introduce the concept of *matrix units* for convenience. A matrix unit E_{ij} is a matrix with a 1 in the (i,j)-th position and 0s elsewhere. The (i,j)-th entry of a matrix is the entry located in the i-th row and j-th column.

> **Remark.** Be careful the distinction between superscripts and subscripts in matrix units. As $E_i^j = \vec{e}_i\hat{e}^j$ is a matrix, while $a_j^i = \hat{e}^i A\vec{e}_j$ is the (i,j)-th entry of a matrix A, which is a scalar. In this book, we always use i for row index and j for column index.

Proposition 2.4.3. The row operation $R_i \longleftrightarrow R_j$ is equivalent to left multiplication by the elementary matrix $E = I - E_i^i - E_j^j + E_i^j + E_j^i$.

Proof. The linear map corresponding to the elementary matrix E is given by

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

Therefore, the matrix E is

$$E = \begin{bmatrix} | & & | & & | & & | \\ \vec{e}_1 & \cdots & \vec{e}_j & \cdots & \vec{e}_i & \cdots & \vec{e}_n \\ | & & | & & | & & | \end{bmatrix} = I - E_i^i - E_j^j + E_i^j + E_j^i. \qquad \square$$

Exercise 2.4.1. Show that the row operation $R_i \leftarrow \alpha R_i$ is equivalent to left multiplication by the elementary matrix $E = I + (\alpha - 1)E_i^i$.

Exercise 2.4.2. Show that the row operation $R_i \leftarrow R_i + \alpha R_j$ is equivalent to left multiplication by the elementary matrix $E = I + \alpha E_i^j$.

Similarly, elementary column operations are operations that can be performed on the columns of a matrix. There are three types of elementary column operations:

– Column swapping: $C_i \longleftrightarrow C_j$ (swap column i and column j)
– Column scaling: $C_i \leftarrow \alpha C_i$ (multiply column i by a non-zero scalar α)
– Column addition: $C_i \leftarrow C_i + \alpha C_j$ (add α times column j to column i)

Each elementary column operation is a *right multiplication* by an *elementary matrix*. Moreover, every elementary matrix is invertible, and its inverse is also an elementary matrix.

2.4.4. Canonical Forms of Matrices and Trivialisation. Using elementary row and column operations, we can transform any matrix into a simpler form called the *canonical form*. One common canonical form is the *row echelon form* (REF) and the *reduced row echelon form* (RREF) which are useful for solving systems of linear equations. However, for the purpose of understanding the structure of linear maps, we focus on the *Smith normal form* or *normal form* of a matrix.

Proposition 2.4.4. Any matrix A in $\mathrm{Mat}_{m\times n}(F)$ can be transformed into a normal form

$$N = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}$$

by a finite sequence of elementary row and column operations, where r is the rank of the matrix A.

Proof. Consider the following commutative diagram:

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

Here, P and Q are invertible matrices obtained by performing finite sequence of row operations and column operations on the identity matrices of appropriate sizes respectively. Thus, we have $N = PAQ$. □

Remark. The rank is uniquely determined by the matrix A and does not depend on the sequence of elementary row and column operations used to transform A into its normal form.

Proposition 2.4.5. Let A be an $m \times n$ matrix over F. The following statements are equivalent:

(1) A is invertible;

(2) the normal form of A is invertible;

(3) $\mathrm{rank}(A) = n = m$;

(4) the normal form of A is I_n.

Proof.

(1) $\Longrightarrow$ (2) : If A is invertible, then PAQ^{-1} is also invertible for any elementary matrices P and Q. Thus, the normal form of A is invertible.

(2) $\Longrightarrow$ (3) : If the normal form of A is invertible, then it must be a square matrix with full rank since the matrix is surjective and dimension of column space is n. Therefore, $\mathrm{rank}(PAQ^{-1}) = n$. Moreover, as rank is invariant under multiplication by invertible matrices, we have $\mathrm{rank}(A) = \mathrm{rank}(PAQ^{-1}) = n$. Since PAQ^{-1} is an $m \times n$ invertible matrix, we must have $m = n$.

(3) $\Longrightarrow$ (4) : If $\mathrm{rank}(A) = r = n = m$, then the normal form of A must be I_n.

(4) $\Longrightarrow$ (1) : If the normal form of A is I_n, then we have $I_n = PAQ$ for some invertible matrices P and Q. Thus, we have $A = P^{-1}I_nQ^{-1} = P^{-1}Q^{-1}$, which shows that A is invertible. □

Exercise 2.4.3. Show that the following statements are equivalent for an $m \times n$ matrix A over F:

(1) A has a left inverse, i.e., there exists an $n \times m$ matrix B such that $BA = I_n$;

(2) A is injective;

(3) $\mathrm{rank}(A) = n$;

(4) the normal form of A is $\begin{bmatrix} I_n \\ 0 \end{bmatrix}$.

Exercise 2.4.4. Show that the following statements are equivalent for an $m \times n$ matrix A over F:

(1) A has a right inverse, i.e., there exists an $n \times m$ matrix C such that $AC = I_m$;

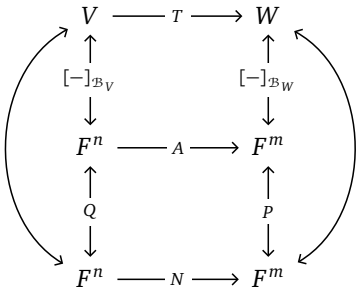
(2) A is surjective;

(3) $\operatorname{rank}(A) = m$;

(4) the normal form of A is $\begin{bmatrix} I_m & 0 \end{bmatrix}$.

Remark. From the exercises above, for any algebraic structure, having a left inverse is equivalent to being injective, while having a right inverse is equivalent to being surjective. However, having both a left inverse and a right inverse is equivalent to being invertible only in the case of linear maps between finite-dimensional linear spaces.
The definition of monomorphism is a left-cancellative morphism, or equivalently, there is a *retraction* that is a left inverse. The definition of epimorphism is a right-cancellative morphism, or equivalently, there is a *section* that is a right inverse.
In the category of finite-dimensional linear spaces over a field F, monomorphisms are exactly injective linear maps, and epimorphisms are exactly surjective linear maps. However, in general categories, monomorphisms are not necessarily injective, and epimorphisms are not necessarily surjective.

Any linear map $T : V \to W$ between finite-dimensional linear spaces can be represented by a matrix once we choose bases for V and W. The process of representing a linear map by a matrix is called *trivialisation*. Consider the following commutative diagram:



Here, $\mathcal{B}_V$ and $\mathcal{B}_W$ are bases for V and W respectively, A is the standard matrix of the linear map T with respect to the chosen bases, and N is the normal form of the matrix A. The coordinate maps $[-]_{\mathcal{B}_V}$ and $[-]_{\mathcal{B}_W}$ are the isomorphisms that map vectors in V and W to their coordinate representations in F^n and F^m respectively. The matrices P and Q are invertible matrices corresponding to the elementary row and column operations used to transform A into its normal form N.

2.5. Quotient Spaces

2.5.1. Group Actions, Orbits, and Stabilisers. Before studying quotient spaces, we introduce the concept of (left) group actions.

Definition 2.17 — Left Group Action.
A *left group action* of a group G on a set X is a map $\cdot : G \times X \to X$ such that for all $g, h \in G$ and all $x \in X$, we have

- Compatibility: $g \cdot (h \cdot x) = (gh) \cdot x$;
- Unital: $e \cdot x = x$, where e is the identity element of G.

Remark. A right group action of a group G on a set X is defined similarly, with the action map $\cdot : X \times G \to X$ satisfying the compatibility condition $(x \cdot g) \cdot h = x \cdot (gh)$ and the unital condition $x \cdot e = x$ for all $g, h \in G$ and all $x \in X$.

A rotation on a plane is a group action of the group $\mathrm{SO}(2)$ on the set of points in the plane. Each element of $\mathrm{SO}(2)$ can be represented by a 2×2 matrix of the form

$$\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix},$$

where θ is the angle of rotation. The group action is defined by matrix multiplication, where each point in the plane is represented as a vector in $\mathbb{R}^2$. We will explore more about the group $\mathrm{SO}(n)$ in later chapters. We can visualise the group action as shown in Figure 1.

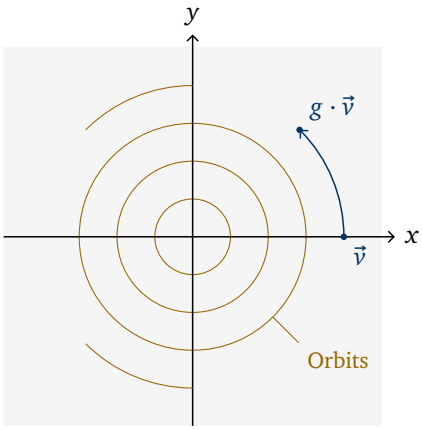


Figure 1. A group action of $\mathrm{SO}(2)$ on the plane.

Definition 2.18 — Orbits.
Let G be a group acting on a set X. The *orbit* of an element $x \in X$ under the action of G is the set

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

Definition 2.19 — Stabiliser.
Let G be a group acting on a set X. The *stabiliser* of an element $x \in X$ under the action of G is the set

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

In the example of rotation on a plane, the orbits are circles centered at the origin, and the stabiliser of any non-zero point is the trivial group containing only the identity element.

Definition 2.20 — Partition.
A *partition* of a set X is a collection of non-empty disjoint subsets $\{X_i\}_{i \in I}$ of X such that their union is X:

$$X = \bigsqcup_{i \in I} X_i.$$

The set of orbits of a group action forms a partition of the set being acted upon and we denote the set of orbits by $X \,/\, G = \{G \cdot x \mid x \in X\}$. Then there is a natural surjective map $\pi : X \to X \,/\, G$ that sends

each element $x \in X$ to its corresponding orbit $G \cdot x$ in the set of orbits X / G. This map is called the *quotient map*.

2.5.2. Quotient Spaces. Consider a linear space V and a subspace W of V. Note that $(W, +)$ is an abelian group under vector addition. We can define a group action of $(W, +)$ on V as follows:

$$W \times V \to V$$
$$(w, v) \mapsto v + w.$$

It is straightforward to verify that this map satisfies the compatibility and unital conditions of a group action. One way is to check all conditions directly. Another way is to observe that the condtions follow from the properties of vector addition in V with the commutative diagram:

$$W \times V \hookrightarrow{\iota \times \mathrm{id}_V} V \times V \xrightarrow{\quad + \quad} V$$

The orbits of this group action are the sets of the form $v + W = \{v + w \mid w \in W\}$ for each $v \in V$. These orbits partition the linear space V into disjoint subsets. Algebraically, such subsets are called *cosets* of W in V. The cosets can be written as $[v]$ or $\overline{v}$ for simplicity. In this book, we use the notation $[v]$ for cosets.

Definition 2.21 — Linear Quotient Space.

The *linear quotient space* of a linear space V by a subspace W is the set of orbits of the group action of $(W, +)$ on V:

$$V / W = \{v + W \mid v \in V\}.$$

Another way to view the quotient space V / W is to consider the equivalence relation $\sim$ on V defined by $v_1 \sim v_2$ if and only if $v_1 - v_2 \in W$. The equivalence classes under this relation are precisely the cosets of W in V. Thus, the quotient space V / W can be identified with the set of equivalence classes of V under the relation $\sim$. Graphically, we can visualise the quotient space V / W as shown in Figure 2. Here,

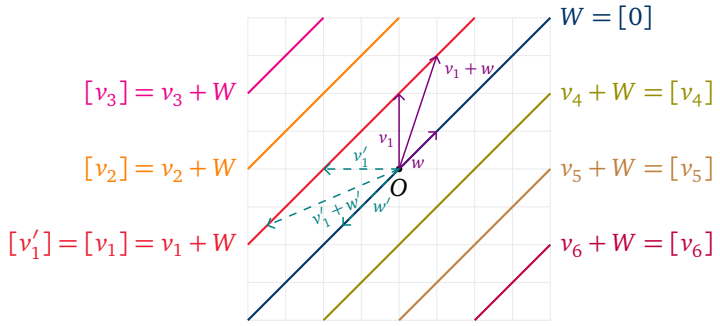


Figure 2. A graphical representation of the quotient space V / W.

the blue line represents the subspace W, and each colored line represents a distinct coset in the quotient space V / W. The vectors w and w' belong to the same coset if they differ by an element of W.

Similarly, there is a natural surjective map from the linear space V to the quotient space V / W that sends each vector to its corresponding coset.

Definition 2.22 — Linear Quotient Map.

The *linear quotient map* $\pi : V \to V / W$ is the map that sends each vector $v \in V$ to its corresponding coset $v + W$ in the quotient space V / W:

$$\pi(v) = v + W = [v].$$

Currently, the quotient space V / W is only defined as a set. To show that V / W is indeed a linear space, we consider the following proposition.

Proposition 2.5.1. There is a unique linear structure on the quotient space V / W such that the quotient map $\pi: V \to V / W$ is a linear map.

Proof. If such a linear structure exists, then for any $v_1, v_2 \in V$ and any scalar $\alpha, \beta \in F$, we must have

$$\pi(\alpha v_1 + \beta v_2) = \alpha\pi(v_1) + \beta\pi(v_2).$$

This suggests the unique way to define the linear combination in V / W is

$$\alpha[v_1] + \beta[v_2] = [\alpha v_1 + \beta v_2].$$

We need to verify that this definition is well-defined. Suppose $[v_1] = [v_1']$ and $[v_2] = [v_2']$, i.e., $v_1' - v_1 \in W$ and $v_2' - v_2 \in W$. Then,

$$(\alpha v_1' + \beta v_2') - (\alpha v_1 + \beta v_2) = \alpha(v_1' - v_1) + \beta(v_2' - v_2) \in W,$$

which implies that $[\alpha v_1' + \beta v_2'] = [\alpha v_1 + \beta v_2]$. Thus, the linear combination is well-defined. □

Remark. In normal procedure, we first define the operations on a set and then verify the set is closed under these operations and zero vector exists. Then we check the map preserves these operations. However, in this case, we define the operations on the quotient space V / W by requiring the quotient map π to be a linear map. Then we verify that the operations are well-defined. This approach is often used in abstract algebra.

If we want to visualise the graphical representation of the quotient space V / W and the quotient map $\pi: V \to V / W$, we can refer to Figure 3.

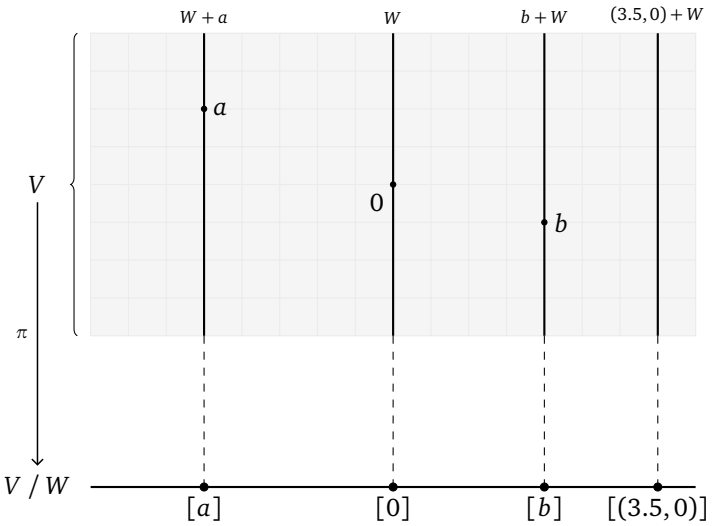
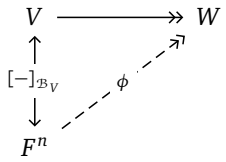


Figure 3. A graphical representation of the quotient map $\pi: V \to V / W$.

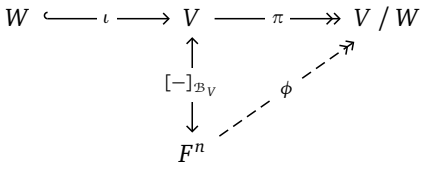
We also have the following properties about finite-dimensional linear spaces.

Proposition 2.5.2. V is finite-dimensional if and only if all of its subspaces and quotient spaces are finite-dimensional.

Proof. If V is finite-dimensional and W is a subspace of V, then we have the following commutative diagram:



Here, the map $\phi : F^n \to W$ is a surjective linear map from a finite-dimensional linear space F^n to W. Thus, W is finitely generated.
Similarly, consider the following commutative diagram:



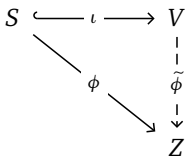
Then we know that $\phi : F^n \to V / W$ is a surjective linear map from a finite-dimensional linear space F^n to V / W. Thus, V / W is finitely generated. □

2.6. Universal Properties in Linear Spaces

Universal properties provide a powerful and abstract way to characterise mathematical objects based on their relationships with other objects. They are often used to define and study various constructions in category theory, algebra, and topology. Starting here, we should change our perspective to a more categorical viewpoint: instead of focusing on the elements of sets or spaces, we focus on the morphisms (maps) between objects and how these morphisms interact with each other.

We first start with a simple example: the universal property of minimal spanning set.

Proposition 2.6.1 — Universal Property of Minimal Spanning Set. Let S be a minimal spanning set of a linear space V. For any linear space Z and any set map $\phi : S \to Z$, there exists a unique linear map $\widetilde{\phi} : V \to Z$ such that the following diagram commutes:



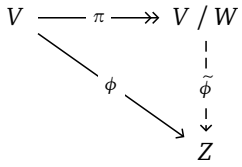
Proof. If such a linear map $\widetilde{\phi}$ exists, then for any $s \in S$, we must have $\widetilde{\phi} \circ \iota(s) = \phi(s)$, which suggests that $\widetilde{\phi}$ is defined by extending ϕ linearly to the whole space V. Specifically, for any $v \in V$, we can express v as a linear combination of elements in S, i.e., $v = \sum_{i=1}^{k} \alpha_i s_i$ for some $s_i \in S$ and $\alpha_i \in F$. Then, we define

$$\widetilde{\phi}(v) = \widetilde{\phi}\left(\sum_{i=1}^{k} \alpha_i s_i\right) = \sum_{i=1}^{k} \alpha_i \widetilde{\phi}(s_i) = \sum_{i=1}^{k} \alpha_i \phi(s_i).$$

As S is a minimal spanning set, there is only one way to express v as a linear combination of elements in S. So, the definition of $\widetilde{\phi}$ is well-defined, i.e., does not depend on the choice of representation of v. □

This proposition shows that any set map from a minimal spanning set S to another linear space Z can be uniquely extended to a linear map from the entire space V, i.e., $\mathrm{Map}(S, Z) \simeq \mathrm{Hom}(V, Z)$.

Proposition 2.6.2 — Universal Property of Quotient Space. Let W be a subspace of a linear space V. For any linear space Z and any linear map $\phi : V \to Z$ such that $W \subseteq \ker(\phi)$, there exists a unique linear map $\widetilde{\phi} : V / W \to Z$ such that the following diagram commutes:



Proof. If such a linear map $\widetilde{\phi}$ exists, then for any $v \in V$, we must have $\widetilde{\phi} \circ \pi(v) = \phi(v)$, which suggests that $\widetilde{\phi}$ is defined by

$$\widetilde{\phi}([v]) = \phi(v).$$

We need to verify that this definition is well-defined. Suppose $[v] = [v']$, i.e., $v' - v \in W$. Then,

$$\widetilde{\phi}([v']) - \widetilde{\phi}([v]) = \phi(v') - \phi(v) = \phi(v' - v) = 0,$$

which implies that $\widetilde{\phi}([v']) = \widetilde{\phi}([v])$. Thus, the definition of $\widetilde{\phi}$ is well-defined. Then we consider the linearity of $\widetilde{\phi}$: for any $[v_1], [v_2] \in V / W$ and any scalars $\alpha, \beta \in F$, we have

$$\begin{aligned}\widetilde{\phi}(\alpha[v_1] + \beta[v_2]) &= \widetilde{\phi}([\alpha v_1 + \beta v_2]) = \phi(\alpha v_1 + \beta v_2) \\ &= \alpha\phi(v_1) + \beta\phi(v_2) = \alpha\widetilde{\phi}([v_1]) + \beta\widetilde{\phi}([v_2]).\end{aligned}$$

□

Remark. Note that $[0] = W$ in the quotient space V / W. Thus, the map from W to V / W is the zero map. This is consistent with the condition that $W \subseteq \ker(\phi)$, which implies that the restriction of ϕ to W is also the zero map.

This proposition shows that any linear map from V to another linear space Z that vanishes on the subspace W can be uniquely *factored* through the quotient space V/W, i.e., $\mathrm{Hom}(V, Z)_W \cong \mathrm{Hom}(V/W, Z)$, where $\mathrm{Hom}(V, Z)_W$ denotes the set of linear maps from V to Z that vanish on W.

There are two terms that is "dual" to the kernel and image of a linear map: the *cokernel* and *coimage*.

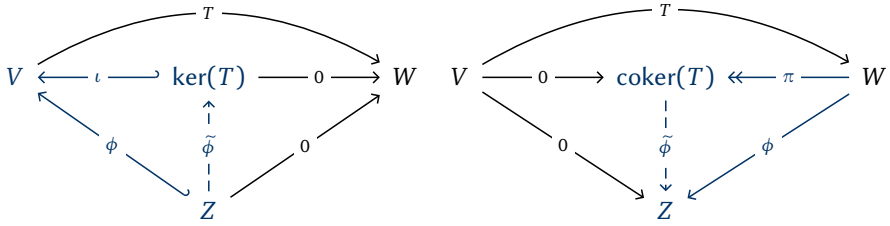
Definition 2.23 — Cokernel.
The *cokernel* of a linear map $T : V \to W$ is the quotient space of W by the image of T:

$$\mathrm{coker}(T) = W / \mathrm{im}(T).$$

Definition 2.24 — Coimage.
The *coimage* of a linear map $T : V \to W$ is the quotient space of V by the kernel of T:

$$\mathrm{coim}(T) = V / \ker(T).$$

We also have universal properties for kernel and cokernel with the following commutative diagrams:



We will explore more universal properties when we introduce category theory in later chapters.

2.7. Exact Sequences

Exact sequences are useful tools in linear algebra and homological algebra to study the relationships between linear spaces and linear maps.

Definition 2.25 — Exact Sequence.

A sequence of linear maps between linear spaces over F

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

is *exact* at V_i if the image of f_{i-1} is equal to the kernel of f_i:

$$\operatorname{im}(f_{i-1}) = \ker(f_i).$$

The sequence is called an *exact sequence* if it is exact at every V_i.

Example 2.7.1. Consider the following *short exact sequence* of linear spaces:

$$0 \longrightarrow V_1 \xrightarrow{\iota_1} V \xrightarrow{\pi_2} V_2 \longrightarrow 0$$

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

- the exactness at V_1 implies that $\{0_{V_1}\} = \operatorname{im}(0) = \ker(\iota_1)$, thus ι_1 is injective.
- the exactness at V implies that $\operatorname{im}(\iota_1) = \ker(\pi_2)$, thus $V_1 \cong \operatorname{im}(\iota_1) \subseteq V$.
- the exactness at V_2 implies that $\operatorname{im}(\pi_2) = \ker(0) = V_2$, thus π_2 is surjective.

There are some facts about the short exact sequence:

- π_2 has a right inverse, or a section, i.e., there exists a linear map $\iota_2 : V_2 \to V$ such that $\pi_2 \circ \iota_2 = \mathrm{id}_{V_2}$. 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.
- ι_1 has a left inverse, or a retraction, i.e., there exists a linear map $\pi_1 : V \to V_1$ such that $\pi_1 \circ \iota_1 = \mathrm{id}_{V_1}$. This is because ι_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:

$$0 \longrightarrow V_1 \underset{\pi_1}{\overset{\iota_1}{\rightleftarrows}} V \underset{\pi_2}{\overset{\iota_2}{\rightleftarrows}} V_2 \longrightarrow 0$$

We can draw an Euler diagram to illustrate the situation as shown in Figure 4.

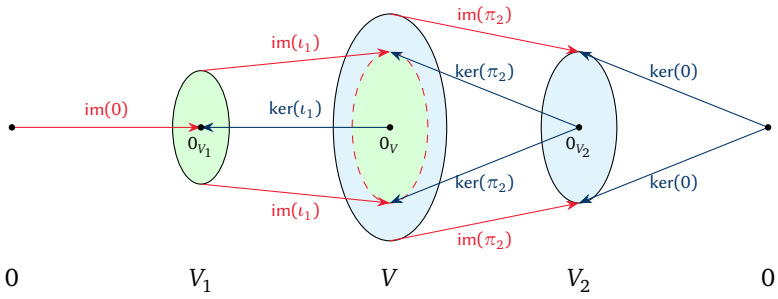


Figure 4. An Euler diagram illustrating a short exact sequence of linear spaces.

Consider the same short exact sequence as above, we have the following equalities:

– $\pi_1 \circ \iota_1 = \mathrm{id}_{V_1}$ because π_1 is a left inverse of ι_1.

– $\pi_2 \circ \iota_2 = \mathrm{id}_{V_2}$ because π_2 is a right inverse of ι_2.

– $\pi_2 \circ \iota_1 = 0$ because $\mathrm{im}(\iota_1) = \ker(\pi_2)$.

– $\pi_1 \circ \iota_2 = 0$ because $\mathrm{im}(\iota_2) = \ker(\pi_1)$.

– $\iota_1 \circ \pi_1 + \iota_2 \circ \pi_2 = \mathrm{id}_V$ because for all $v \in V$, we have $v = (v - \iota_2(\pi_2(v))) + \iota_2(\pi_2(v))$ where $v - \iota_2(\pi_2(v)) \in \mathrm{im}(\iota_1)$ and $\iota_2(\pi_2(v)) \in \mathrm{im}(\iota_2)$. Also, $\mathrm{im}(\iota_1) \cap \mathrm{im}(\iota_2) = \{0_V\}$.

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

Proposition 2.7.1. The linear space V is isomorphic to the internal direct sum of the images of ι_1 and ι_2:

$$V = \mathrm{im}(\iota_1) \oplus \mathrm{im}(\iota_2).$$

Proof. The meaning of $V \cong \mathrm{im}(\iota_1) \oplus \mathrm{im}(\iota_2)$ is that for any $x \in V$, it can be uniquely written as $x = x_1 + x_2$ where $x_i \in \mathrm{im}(\iota_i)$. Why? Suppose $x = x_1 + x_2 = x_1' + x_2'$ where $x_i, x_i' \in \mathrm{im}(\iota_i)$. Then we have $(x_1 - x_1') + (x_2 - x_2') = 0$. Note that $x_1 - x_1' \in \mathrm{im}(\iota_1)$ and $x_2 - x_2' \in \mathrm{im}(\iota_2)$. 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 $\iota_2 : s \mapsto \iota_2(s)$ where $\iota_2(s)$ is a choice of element from $\pi_2^{-1}(s) \neq \emptyset$ for each $s \in S$. Then we extend it to the whole space linearly. Thus, ι_2 is injective.
Then we want to prove that $\mathrm{im}(\iota_1)$ and $\mathrm{im}(\iota_2)$ are weakly independent. Assume that $x_1 + x_2 = 0$ where $x_i \in \mathrm{im}(\iota_i)$. Then we have $\pi_2(x_1 + x_2) = \pi_2(x_1) + \pi_2(x_2) = 0$. Note that $\pi_2(x_1) = 0$ because $x_1 \in \mathrm{im}(\iota_1) = \ker(\pi_2)$, the exactness of V. Thus, we have $\pi_2(x_2) = 0$. However, π_2 is injective on $\mathrm{im}(\iota_2)$ because $\pi_2 \circ \iota_2 = \mathrm{id}_{V_2}$. Thus, we have $x_2 = 0$ and $x_1 = 0$. This shows that $\mathrm{im}(\iota_1)$ and $\mathrm{im}(\iota_2)$ are weakly independent.
Finally, we want to prove that $\mathrm{im}(\iota_1) + \mathrm{im}(\iota_2) = V$. For all $x \in V$, we let $x_2 = \iota_2(\pi_2(x)) \in \mathrm{im}(\iota_2)$ and $x_1 = x - x_2$. Then we have to show that $x_1 \in \mathrm{im}(\iota_1) = \ker(\pi_2)$. Note that $\pi_2(x) = \pi_2(x_1) + \pi_2(x_2) = \pi_2(x_1) + \pi_2 \circ \iota_2(\pi_2(x)) = \pi_2(x_1) + \pi_2(x)$. This shows that $\pi_2(x_1) = 0$. Thus, $x_1 \in \ker(\pi_2) = \mathrm{im}(\iota_1)$. This shows that $\mathrm{im}(\iota_1) + \mathrm{im}(\iota_2) = V$.
Actually π_1 is the projection from $\mathrm{im}(\iota_1) \oplus \mathrm{im}(\iota_2)$ to $\mathrm{im}(\iota_1)$ and it exists due to the uniqueness of the decomposition. □

The equalities can be summarized as follows:

$$\pi_m \circ \iota_n = \delta_{mn}\, \mathrm{id}_{V_n}, \quad \sum_{k=1}^{2} \iota_k \circ \pi_k = \mathrm{id}_V$$

For the dimension of the spaces, we have:

$$\dim(V) = \dim(\mathrm{im}(\iota_1)) + \dim(\mathrm{im}(\iota_2)) = \dim(V_1) + \dim(V_2)$$

As $V_1 \cong \mathrm{im}(\iota_1)$ and $V_2 \cong \mathrm{im}(\iota_2)$. ι_1 and ι_2 are injective and $V_k \to \mathrm{im}\, i_k$ 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.

Using the exact sequence, we can prove the dimension formula for linear spaces easily.

Exercise 2.7.1 — Dimension Formula. Show that the following short sequence of linear spaces is exact:

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

Then we can establish the natural isomorphism:

$$V_1 \,/\, (V_1 \cap V_2) \simeq (V_1 + V_2) \,/\, V_2,$$

and the dimension formula:

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

[Hints: You may use the universal property of quotient space to construct the isomorphism.]

There are two special questions which can be solved using exact sequences easily. Please refer to the Appendix 9.4 to check the story about these two questions.

Exact sequences can also be used to prove the Rank-Nullity Theorem.

Theorem 2.1 — Rank-Nullity Theorem.

For a linear map $T : V \to W$ between finite-dimensional linear spaces, we have:

$$\dim(V) = \operatorname{rank}(T) + \operatorname{nullity}(T). \tag{2}$$

Proof. Consider the following short exact sequence:

$$0 \longrightarrow \ker(T) \xhookrightarrow{\iota} V \xrightarrow{T} \operatorname{im}(T) \longrightarrow 0$$

Then we have the internal direct sum decomposition $V = \ker(T) \oplus \operatorname{im}(T)$. Thus, we have $\dim(V) = \dim(\ker(T)) + \dim(\operatorname{im}(T))$. This shows that $\operatorname{rank}(T) + \operatorname{nullity}(T) = \dim(V)$. □

Moreover, we have the following corollary.

Corollary 2.1.

For a linear map $T : V \to W$ between finite-dimensional linear spaces, we have:

$$\dim(W) = \operatorname{rank}(T) + \dim(\operatorname{coker}(T)).$$

Proof. Consider the following short exact sequence:

$$0 \longrightarrow \operatorname{im}(T) \xhookrightarrow{\iota} W \xrightarrow{\pi} \operatorname{coker}(T) \longrightarrow 0$$

Then we have the external direct sum decomposition $W \cong \operatorname{im}(T) \oplus \operatorname{coker}(T)$. Thus, we have $\dim(W) = \dim(\operatorname{im}(T)) + \dim(\operatorname{coker}(T))$. This shows that $\operatorname{rank}(T) + \dim(\operatorname{coker}(T)) = \dim(W)$. □

Corollary 2.2.

For a linear map $T : V \to W$ between finite-dimensional linear spaces, we have:

$$\dim(V) = \operatorname{nullity}(T) + \dim(\operatorname{coim}(T)).$$

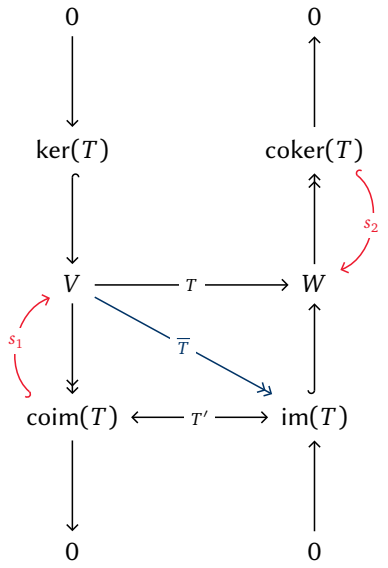
Proof. Consider the following short exact sequence:

$$0 \longrightarrow \ker(T) \xhookrightarrow{\iota} V \xrightarrow{\pi} \operatorname{coim}(T) \longrightarrow 0$$

Then we have $V \cong \ker(T) \oplus \operatorname{coim}(T)$. Thus, we have $\dim(V) = \dim(\ker(T)) + \dim(\operatorname{coim}(T))$. This shows that $\operatorname{nullity}(T) + \dim(\operatorname{coim}(T)) = \dim(V)$. □

2.8. Canonical Form of Linear Maps

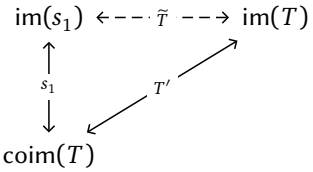
We have already known that any linear map $T : V \to W$ between finite-dimensional linear spaces can be represented by a matrix once we choose bases for V and W. Moreover, we can choose appropriate coordinate maps such that the matrix representation of T is in canonical form. How about the abstract form of the linear map without choosing any bases or coordinate maps? We can see it using exact sequences. Consider the following commutative diagram:



Here, each column is a short exact sequence and the square in the middle commutes. Also, $\overline{T}$ and T' are isomorphisms. Moreover, we have the sections s_1 and s_2 that are the right inverses of the projections from V to $\mathsf{coim}(T)$ and from W to $\mathsf{coker}(T)$ respectively. Thus, we can decompose V and W into $V = \mathsf{im}(s_1) \oplus \ker(T)$ and $W = \mathsf{im}(s_2) \oplus \mathsf{im}(T)$ respectively. Then, with respect to these decompositions, the linear map $T : V \to W$ can be represented as:

$$\mathsf{im}(s_1) \oplus \ker(T) \xrightarrow{\begin{bmatrix} \widetilde{T} & 0 \\ 0 & 0 \end{bmatrix}} \mathsf{im}\, T \oplus \mathsf{im}(s_2)$$

where $\widetilde{T} : \mathsf{im}(s_1) \to \mathsf{im}(T)$ is an isomorphism, as there are isomorphisms $T' : \mathsf{coim}(T) \to \mathsf{im}(T)$ and $\mathsf{im}(s_1) \cong \mathsf{coim}(T)$. Then the graph below commutes:



This shows the canonical form of a linear map without choosing any bases or coordinate maps.

> **Remark.** Note that the choice of the sections s_1 and s_2 is not unique. Thus, the internal direct sum decompositions of V and W are not unique. However, up to isomorphisms, the decompositions are unique. This is similar to the situation of choosing complements of subspaces.

The rank of the isomorphism $\widetilde{T}$ is unique and it is equal to the rank of the original linear map T. Moreover, after trivialisation, we have the following matrix representation of T:

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

2.9. Exercises

Problem 2.1. Show that

(a) for any $m \times n$ matrix A, the map $(F^m)^* \to (F^n)^*$ that sends α to αA is a linear map;

(b) any linear map $\phi : (F^m)^* \to (F^n)^*$ is of the form $\phi(\alpha) = \alpha A$ for a unique matrix A;

(c) the i-th row of A is the row matrix $\hat{e}^i A$;

(d) the (i, j)-th entry of A is $a^i_j = \hat{e}^i A \vec{e}_j$;

(e) $A = \sum_{1 \le i \le m, 1 \le j \le n} a^i_j E^j_i$ where $E^j_i = \vec{e}_i \hat{e}^j$.

Problem 2.2. Show that an elementary matrix E that corresponds to an elementary row operation is also an elementary matrix F that corresponds to an elementary column operation. Prove by induction that any matrix can be turned into a matrix of the block form

$$\begin{bmatrix} I_r & O \\ O & O \end{bmatrix}$$

by finitely many elementary row or column operation. Here, I_r denotes the identity matrix of order r and matrices O denote the zero matrices.

Problem 2.3. Let $r \le s \le n$ be non-negative integers. Denote by A_r the square matrix of order n of the block form

$$\begin{bmatrix} I_r & O \\ O & O \end{bmatrix}.$$

Show that, if there are invertible matrices P and Q such that $PA_rQ^{-1} = A_s$, then $r = s$.

Problem 2.4. With reference to Problem 1.3, show that both T_* and T^* are linear maps. Also show that, if T is a bijection, then both T_* and T^* are linear isomorphisms.

Problem 2.5. Let $f : V \to W$ be a linear map. Show that

(a) f is injective $\iff$ the kernel of f is trivial (i.e., $\{0\}$);

(b) f is surjective $\iff$ the cokernel of f is trivial;

(c) f is isomorphism $\iff$ both kernel and cokernel of f are trivial;

(d) f is surjective $\iff$ for any linear map $g : W \to Z$, $gf = 0 \implies g = 0$;

(e) f is injective $\iff$ for any linear map $h : U \to V$, $fh = 0 \implies h = 0$.

Now we assume that V and W are finite-dimensional, say $V = \mathbb{F}^n$ and $W = \mathbb{F}^m$, then f is the multiplication by an $m \times n$ matrix A.

(f) Please translate the five statements above into the corresponding statements about matrix A.

Problem 2.6. Let $f : V \to W$ be a set map between linear spaces. Show that

(a) its graph $\Gamma_f := \{(v, f(v)) \mid v \in V\}$ is a linear subspace of the product linear space $V \times W \iff f$ is a linear map.

(b) in case f is linear, its domain is naturally linear isomorphic to its graph: domain $f = \Gamma_f$.

Problem 2.7. We say a linear map $f : V \to W$ is *imbedding* if the map $\overline{f} : V \to \mathrm{im}(f)$ that sends v to $f(v)$ is a linear isomorphism. Show that f is imbedding $\iff$ f is one-to-one.

An optional exercise: We say a topological map, i.e., continuous map, $f : X \to Y$ is imbedding if the map $\overline{f} : X \to \operatorname{im} f$ that sends x to $f(x)$ is a topological equivalence, i.e., homeomorphism. Show that f is imbedding implies that f is one-to-one, but the converse is not true.

Problem 2.8. Let W be a linear subspace of V and $\sim$ be the equivalence relation on V:

$$v \sim v' \iff v - v' \in W.$$

We let $V \,/\, W$ denote the set of equivalence classes.

(a) Show that there is a unique linear structure on $V \,/\, W$ such that the quotient map $q : V \to V \,/\, W$ is a linear map.

(b) Show that, for any linear map $\phi : V \to Z$ such that $\phi(v) = 0$ for any $v \in W$, there *is* a *unique* linear map $\overline{\phi} : V \,/\, W \to Z$ such that

$$\overline{\phi} \circ q = \phi.$$

Remark: $V \,/\, W$ is called the quotient space of V by the subspace W and is also called the algebraic normal space of V in W. It is a fact that $\dim(V \,/\, W) = \dim(V) - \dim(W)$.

(c) Let W be a linear subspace of V. Then the inclusion map $W \xrightarrow{\ \iota\ } V$ is a linear map with image inside V. Please formulate and prove the universal property for the inclusion map ι.

Problem 2.9. Consider an exact sequence

$$0 \longrightarrow V_1 \xrightarrow{\ \iota_1\ } V \xrightarrow{\ \pi_2\ } V_2 \longrightarrow 0$$

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

(a) $\pi_2 \iota_1 = 0$ and $V_1 \cong \operatorname{im} \iota_1$;

(b) π_2 has a right inverse. Let us fix a right inverse ι_2;

(c) $V = \operatorname{im}(\iota_1) \oplus \operatorname{im}(\iota_2)$, i.e., any v in V can be uniquely split into the sum of two, one is of the form $\iota_1(v_1)$ and the other is of the form $\iota_2(v_2)$;

(d) the splitting in part (c) defines two maps, one is from V to V_1 and is denoted by π_1 and the other is $\pi_2 : V \to V_2$;

(e) π_1 is linear and the sequence

$$0 \longleftarrow V_1 \xleftarrow{\ \pi_1\ } V \xleftarrow{\ \iota_2\ } V_2 \longleftarrow 0$$

is exact;

(f) $j_k i_l = \delta_{kl}$, and $\iota_1 \pi_1 + \iota_2 \pi_2 = 1$ (i.e., 1_V);

(g) both V_1 and V_2 are finite-dimensional $\iff$ V is finite-dimensional. In case V is finite-dimensional, we have $\dim(V) = \dim(V_1) + \dim(V_2)$, thus $\dim(V_i) \le \dim(V)$;

(h) for any finite-dimensional linear space, none of its subspaces or quotient spaces has a bigger dimension.

Problem 2.10. Let A be an $m \times n$-matrix, then the multiplication by A defines a linear map $f : \mathbb{F}^n \to \mathbb{F}^m$. The rank of A, denoted by $\operatorname{rank} A$, is defined to be the rank of the linear map f. Note: $\operatorname{im}(f) = \operatorname{col}(A)$ — the span of columns of A.

(a) Show that the rank of a matrix is unchanged under both row operations and column operations;

(b) Show that $\operatorname{rank}(A + B) \le \operatorname{rank}(A) + \operatorname{rank}(B)$ provided that the matrix addition is defined here;

(c) Show that $\operatorname{rank}(AB) \le \operatorname{rank}(A)$ and $\operatorname{rank}(AB) \le \operatorname{rank}(B)$ provided that the matrix multiplication is defined here.

Problem 2.11. (a) Show that any matrix A can be expressed as a product BC of two matrices, where the columns of B are linearly independent and the rows of C are linearly independent. Is this decomposition $A = BC$ unique?

(b) Let V be an n-dimensional linear space over F, and let S be a set of linearly independent vectors in V. Prove that $|S| \leq n$ and that S can be expanded to form a minimal spanning set $\tilde{S}$ of V.

CHAPTER 3

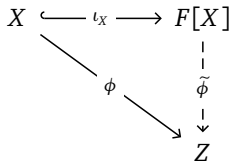
Introduction to Category Theory

Category theory is a branch of mathematics that deals with abstract structures and relationships between them. It provides a unifying framework for understanding various mathematical concepts by focusing on the relationships (morphisms) between objects rather than the objects themselves.

3.1. Free Vector Spaces

Before delving into category theory, it is helpful to understand the concept of *free vector spaces*, or *free linear space*. Let X be a set and ${}^{\delta}\mathrm{X} = \{\delta_x \mid x \in X\}$. Here $\delta_x : X \to F$ is the Kronecker delta function at x, defined in Equation (1). We have already shown that ${}^{\delta}\mathrm{X}$ is a minimal spanning set for $F[X]$. Moreover, there is a natural bijection between X and ${}^{\delta}\mathrm{X}$ given by $x \mapsto \delta_x$. The natural isomorphism, or *natural equivalence*, of X with ${}^{\delta}\mathrm{X}$ allows us to write this map as $\iota_X : X \to F[X]$ where $\iota_X(x) = \delta_x$ for all $x \in X$. This motivates the following universal property of free vector spaces.

Proposition 3.1.1 — Universal Property of Free Vector Space. Let X be a set. For any linear space Z and any set map $\phi : X \to Z$, there exists a unique linear map $\widetilde{\phi} : F[X] \to Z$ such that the following diagram commutes:



Proof. If such a linear map $\widetilde{\phi}$ exists, then for any $x \in X$ we must have

$$\widetilde{\phi}(\delta_x) = \phi(x).$$

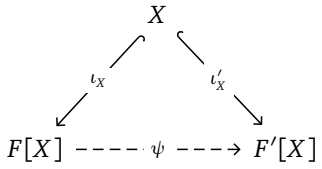
Since ${}^{\delta}\mathrm{X}$ is a spanning set for $F[X]$, this completely determines $\widetilde{\phi}$. □

Via the natural equivalence of X with ${}^{\delta}\mathrm{X}$, denoted by $X \simeq {}^{\delta}\mathrm{X}$, an element in $F[X]$ can be expressed as a finite linear combination of elements in X:

$$\sum \alpha^x \delta_x \Longleftrightarrow \sum \alpha^x x.$$

This is called the *formal linear combination* of elements in X with coefficients in F. Hereafter, we always identify X with ${}^{\delta}\mathrm{X}$ in this way and $F[X]$ is referred to the *set of formal linear combinations* of elements in X with coefficients in F or simply the free vector space on X.

The uniqueness of the universal property is in the following sense: if there is another inclusion map ι'_X into a linear space $F'[X]$ satisfying the same universal property, then there exists a unique isomorphism of linear spaces $\psi : F[X] \to F'[X]$ such that the diagram commutes:



The universal property implies an assignment of a linear map $T_*: F[X] \to F[Y]$, or you may write the map as $F[T]$ instead, to each set map $T: X \to Y$ defined by the following commutative diagram:

$$\begin{array}{ccc} X & \xrightarrow{\quad T \quad} & Y \\ \downarrow{\scriptstyle \iota_X} & & \downarrow{\scriptstyle \iota_Y} \\ F[X] & \xrightarrow{\quad T_* \quad} & F[Y] \end{array}$$

Moreover, this assignment preserves identities and composition:

– For the identity map $\mathrm{id}_X : X \to X$, we have $(\mathrm{id}_X)_* = \mathrm{id}_{F[X]}$.

– For set maps $T: X \to Y$ and $S: Y \to Z$, we have $(ST)_* = S_* T_*$.

You may refer to Problem 1.3 for elementary proofs of these properties.

3.2. Introduction to Categories and Functors

The structure described in the previous section can be abstracted into the notion of a category. We denote a collection of set maps as **Set** and a collection of linear maps over F as $\mathbf{Vec}_F$. Then $F[-]$ can be viewed as a map from the collection of sets to the collection of linear spaces over F that assigns to each set S the free vector space $F[S]$ and to each set map $T: S \to T$ the linear map $T_*: F[S] \to F[T]$:

$$\mathbf{Set} \xrightarrow{\quad F[-] \quad} \mathbf{Vec}_F$$

This map preserves identities and composition, as described above. Such a structure is called a *functor* from the category **Set** to the category $\mathbf{Vec}_F$.

Any monoids can be viewed as *categories* with a single object $*$ where the elements of the monoid are the *morphisms*, or *arrows*, from $*$ to $*$. The composition of morphisms is given by the multiplication in the monoid and the identity morphism is given by the identity element of the monoid. Consider the following composition of morphisms:

$$* \overset{ab}{\frown} * \xrightarrow{\ a\ } * \xrightarrow{\ b\ } *$$

Here a and b are morphisms from $*$ to $*$. The composition ab is also a morphism from $*$ to $*$.

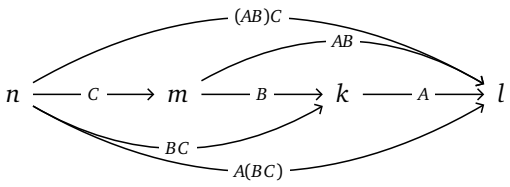
Recall that a monoid is a set M, which is also called a *small collection of objects*, together with a binary operation on M, which is also called a *compositon of morphisms*, with associative property and unital property being satisfied. By relaxing the condition on binary operation, allowing the composition being only partially defined, we end up with the definition 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 $f: X \to Y$ 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.

Example 3.2.1. The collection of all matrices over F is a small category. 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

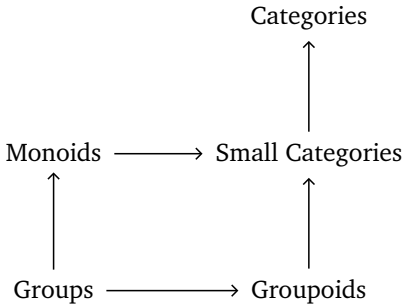
$$\begin{array}{ccccc} {\scriptstyle I_n}\circlearrowright\ n & \xrightarrow{\quad A \quad} & m & \circlearrowright{\scriptstyle I_m} \\ & & \downarrow{\scriptstyle B} & \\ & & k & \end{array}$$

Note that $AI_n = A = I_m A$ and $BI_m = B = I_k B$. The following shows the associativity law:



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

Consider the set of all invertible matrices over 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.

3.3. Small Categories

Definition 3.1 — Small Category.
A *small category* is a set $\mathcal{C}$ together with a subset $\mathcal{C}_0$ of $\mathcal{C}$, two surjective maps $s, t\colon \mathcal{C} \to \mathcal{C}_0$ called the *source* and *target* maps respectively, and a composition map, or a binary operation, $\circ\colon \mathcal{C} \times_{(s,t)} \mathcal{C} \to \mathcal{C}$ that assigns to each pair (f, g) with $s(f) = t(g)$ an element $f \circ g$ in $\mathcal{C}$ satisfying the associative and unital properties.

Here $\mathcal{C} \times_{(s,t)} \mathcal{C} = \{(f, g) \in \mathcal{C} \times \mathcal{C} \mid s(f) = t(g)\}$ is the *fibre product*, or *pullback*, of $\mathcal{C}$ with itself via the maps s and t. The following digram illustrates the fibre product:

$$\begin{array}{ccc} \mathcal{C} \times_{s,t} \mathcal{C} & \xrightarrow{\pi_1} & \mathcal{C} \\ \downarrow{\scriptstyle \pi_2} & \ulcorner & \downarrow{\scriptstyle t} \\ \mathcal{C} & \xrightarrow{s} & \mathcal{C}_0 \end{array}$$

Intuitively, the fibre product $\mathcal{C} \times_{(s,t)} \mathcal{C}$ is to filter out the pairs (f, g) in $\mathcal{C} \times \mathcal{C}$ such that the source of f is the target of g, so that the composition $f \circ g$ is defined.

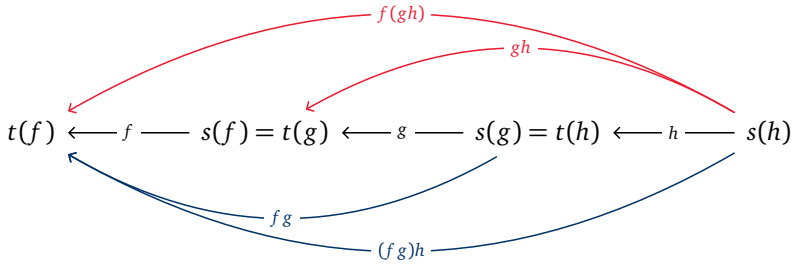
We can picture the composition $f \circ g$ as the following diagram:

$$t(f) \xleftarrow{f} s(f) \qquad t(g) \xleftarrow{g} s(g) \qquad t(f) \xleftarrow{f \circ g} s(g)$$

The left diagram is the composition of f and g, and the right diagram is the resulting morphism $f \circ g$. We can also show the unital property as follows:

$$\overset{\mathrm{id}_{t(f)}}{\curvearrowright} t(f) \xleftarrow{f} s(f) \qquad t(f) \xleftarrow{f} s(f) \qquad t(f) \xleftarrow{f} s(f) \overset{\mathrm{id}_{s(f)}}{\curvearrowleft}$$

The left diagram shows the composition of f with the identity morphism at $t(f)$, the middle diagram is the resulting morphism f, and the right diagram shows the composition of f with the identity morphism at $s(f)$. We can also illustrate the associative property as follows:



Example 3.3.1. In the small category of matrices over F, we have the following:

$$\mathcal{C} = \{\mathrm{Mat}_{m\times n}(F) \mid m, n \in \mathbb{N}\},$$
$$\mathcal{C}_0 = \{I_n \mid n \in \mathbb{N}\} \cong \mathbb{N}.$$

If $A \in \mathcal{C}$ is an $m \times n$ matrix, then $s(A) = I_n$, which is naturally identified with n, and $t(A) = I_m$, which is naturally identified with m. Then A can be viewed as a morphism from n to m: $A\colon n \to m$. The composition is given by the matrix multiplication.

Remark. Elements in $\mathcal{C}$ are morphisms or arrows, and elements in $\mathcal{C}_0$ are identity morphisms which can be identified as objects. Thus, $\mathcal{C}_0$ is also called the *set of objects*. Then a morphism f in $\mathcal{C}$ can be viewed as an arrow from the object $X \simeq \mathrm{id}_X = s(f)$ to the object $Y \simeq \mathrm{id}_Y = t(f)$, denoted by $f : X \to Y$.

We generalise the concept of homomorphisms in algebraic structures to the concept of morphisms in categories.

Definition 3.2 — Morphism.

Let $\mathcal{C}$ be a small category. A *morphism* f in $\mathcal{C}$ is a structure-preserving map from an object X to an object Y, denoted by $f : X \to Y$, where $X, Y \in \mathcal{C}_0$. The object X is called the *source* of f and the object Y is called the *target* of f.

The set of morphisms from an object X to an object Y in a small category $\mathcal{C}$ is denoted by $\mathrm{Hom}_{\mathcal{C}}(X, Y)$ or $\mathcal{C}(X, Y)$ or $\mathrm{Hom}(X, Y)$ if the category is clear from the context. Then $\mathcal{C}$ can be viewed as the disjoint union of all $\mathrm{Hom}(X, Y)$ for all pairs of objects (X, Y):

$$\mathcal{C} = \bigsqcup_{X,Y\in\mathcal{C}_0} \mathrm{Hom}(X, Y).$$

The composition map can be written as follows:

$$\mathrm{Hom}(Y, Z) \times \mathrm{Hom}(X, Y) \longrightarrow \mathrm{Hom}(X, Z)$$
$$(Z \xleftarrow{f} Y, Y \xleftarrow{g} X) \longmapsto (Z \xleftarrow{fg} X)$$

The following is the normal definition of category, which is equivalent to Definition 3.1.

Definition 3.3 — Small Category – Alternative Definition.

A *small category* $\mathcal{C}$ consists of the following data:

– A set of objects $\mathcal{C}_0 = \mathrm{Ob}(\mathcal{C})$;

- A set of morphisms $\mathcal{C}_1 = \mathrm{Hom}(\mathcal{C})$ containing morphisms between each pair of objects in $\mathcal{C}_0$;
- A binary operation, called *composition of morphisms*, $\circ\colon \mathrm{Hom}(Y,Z) \times \mathrm{Hom}(X,Y) \to \mathrm{Hom}(X,Z)$ that assigns to each pair (f,g) with $f : Y \to Z$ and $g : X \to Y$ a morphism $f \circ g : X \to Z$;
- An identity morphism $\mathrm{id}_X : X \to X$ for each object X in $\mathcal{C}_0$;

satisfying the associative and unital properties.

Definition 3.4 — Category.
A *category* is a collection of objects and morphisms between these objects satisfying the same conditions as in Definition 3.3, except that the collection of objects and morphisms may be proper classes instead of sets.

Example 3.3.2 — Common Categories. Some common categories in mathematics are:

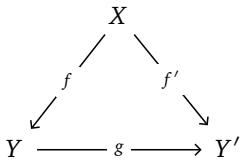
- **Set**: the category of sets and set maps;
- $\mathbf{Vec}_F$: the category of linear spaces over a field F and linear maps;
- **Grp**: the category of groups and group homomorphisms;
- **Rng**: the category of rings and ring homomorphisms;
- **Top**: the category of topological spaces and continuous maps;
- $\mathbf{Mat}_F$: the category of matrices over a field F as described above.

We also have the categories of categories, denoted by **Cat**, and small categories, denoted by **SmallCat**.

Example 3.3.3. If $\mathcal{C}$ and $\mathcal{D}$ are two categories, then their *product category* $\mathcal{C} \times \mathcal{D}$ with objects (X,Y) for $X \in \mathcal{C}_0$ and $Y \in \mathcal{D}_0$, and morphisms (f,g) for $f \in \mathrm{Hom}_{\mathcal{C}}(X,X')$ and $g \in \mathrm{Hom}_{\mathcal{D}}(Y,Y')$, is also a category.

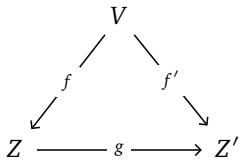
Example 3.3.4. The category of finite sets and set maps, denoted by **FinSet**, is a subcategory of **Set**.

Example 3.3.5. Fix an object X in a category $\mathcal{C}$. The *coslice category* under X, or *under category* of X, denoted by $X \,/\, \mathcal{C}$ or $X \downarrow \mathcal{C}$, has objects that are morphisms with source X: $\{f : X \to Y \mid Y \in \mathcal{C}_0\}$, and morphisms that are commutative triangles:



Moreover, the identity morphism at an object $f : X \to Y$ is given by $\mathrm{id}_Y : Y \to Y$, and the composition of morphisms is given by the composition in $\mathcal{C}$.

Example 3.3.6. Let W be a subspace of a linear space V over F. The coslice category under $V \,/\, W$, denoted by $(V \,/\, W) \,/\, \mathbf{Vec}_F$ or $(V \,/\, W) \downarrow \mathbf{Vec}_F$, has objects that are linear maps $\overline{f} : V \,/\, W \to Z$ for some linear space Z over F, or linear maps $f : V \to Z$ that factor through the quotient map $V \to V \,/\, W$, i.e., $f|_W = 0$, and morphisms that are commutative triangles:



Definition 3.5 — Initial Object.
An object I in a category $\mathcal{C}$ is called an *initial object* if for every object X in $\mathcal{C}$, up to isomorphism, there exists a unique morphism from I to X, i.e., $\mathrm{Hom}_{\mathcal{C}}(I,X)$ is a singleton set or equivalently $|I \,/\, X| = 1$.

Definition 3.6 — Terminal Object.
An object T in a category $\mathcal{C}$ is called a *terminal object* if for every object X in $\mathcal{C}$, up to isomorphism, there exists a unique morphism from X to T, i.e., $\mathrm{Hom}_{\mathcal{C}}(X,T)$ is a singleton set or equivalently $|X \,/\, T| = 1$.

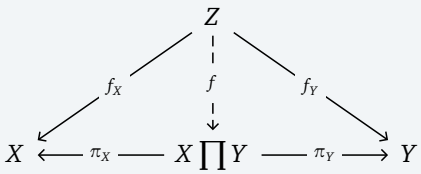
Example 3.3.7. In the category $(V \,/\, W) \,/\, \mathbf{Vec}_F$, the quotient map $\pi : V \to V \,/\, W$ is an initial object and the zero map $0 : V \to 0$ is a terminal object.

Example 3.3.8. In the category **Set**, the empty set $\varnothing$ is an initial object and any singleton set $\{*\}$ is a terminal object.

Example 3.3.9. In the category $\mathbf{Vec}_F$, the zero linear space $\{0\}$ is both an initial object and a terminal object, hence it is a *zero object*.

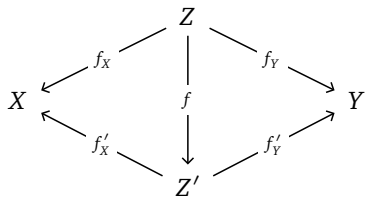
3.4. Products and Coproducts

Definition 3.7 — Product.
Let X and Y be two objects in a category $\mathcal{C}$. A *product* of X and Y is an object $X \prod Y$ in $\mathcal{C}$ 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 in $\mathcal{C}$ with two morphisms $f_X : Z \to X$ and $f_Y : Z \to Y$, there exists a unique morphism $f : Z \to X \prod Y$ making the following diagram commute:



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

There is another way to view the product. Let X and Y be two objects in a category $\mathcal{C}$. Consider the *category of span* from X and Y, denoted by $\mathbf{Span}(X,Y)$, whose objects are triples (Z, f_X, f_Y) such that $X \xleftarrow{f_X} Z \xrightarrow{f_Y} Y$ for any Z, and whose morphisms from (Z, f_X, f_Y) to (Z', f_X', f_Y') are morphisms $f : Z \to Z'$ in $\mathcal{C}$ such that the following diagram commutes:



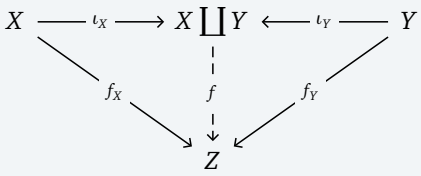
Then the product of X and Y is the terminal object in the category $\mathbf{Span}(X, Y)$.

Example 3.4.1. In the category **Set**, the product of two sets X and Y is their Cartesian product $X \times Y$ with the projection maps $\pi_X : X \times Y \to X$ and $\pi_Y : X \times Y \to Y$. Then for any set Z with two set maps $f_X : Z \to X$ and $f_Y : Z \to Y$, there exists a unique set map $f : Z \to X \times Y$ defined by $f(z) = (f_X(z), f_Y(z))$ for all $z \in Z$ such that the diagram commutes.

Example 3.4.2. In the category $\mathbf{Vec}_F$, the product of two linear spaces V and W over F is their *direct product* $V \times W$ defined by the Cartesian product with the projection maps $\pi_V : V \oplus W \to V$ and $\pi_W : V \oplus W \to W$. Then for any linear space Z over F with two linear maps $f_V : Z \to V$ and $f_W : Z \to W$, there exists a unique linear map $f : Z \to V \oplus W$ defined by $f(z) = (f_V(z), f_W(z))$ for all $z \in Z$ such that the diagram commutes.

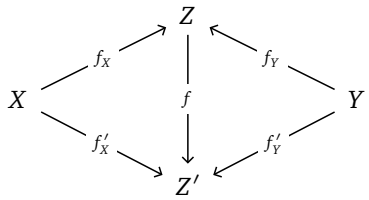
Definition 3.8 — Coproduct.

Let X and Y be two objects in a category $\mathcal{C}$. A *coproduct* of X and Y is an object $X \coprod Y$ in $\mathcal{C}$ 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 in $\mathcal{C}$ with 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.

Similarly, we can view the coproduct in another way. Let X and Y be two objects in a category $\mathcal{C}$. Consider the *category of cospan* from X and Y, denoted by $\mathbf{Cospan}(X, Y)$, whose objects are triples (Z, f_X, f_Y) such that $X \xrightarrow{f_X} Z \xleftarrow{f_Y} Y$ for any Z, and whose morphisms from (Z, f_X, f_Y) to (Z', f'_X, f'_Y) are morphisms $f : Z \to Z'$ in $\mathcal{C}$ such that the following diagram commutes:



Then the coproduct of X and Y is the initial object in the category $\mathbf{Cospan}(X, Y)$.

Example 3.4.3. In the category **Set**, the coproduct of two sets X and Y is their *disjoint union* $X \sqcup Y$ with the inclusion maps $\iota_X : X \to X \sqcup Y$ and $\iota_Y : Y \to X \sqcup Y$. Then for any set Z with two set maps $f_X : X \to Z$

and $f_Y : Y \to Z$, there exists a unique set map $f : X \sqcup Y \to Z$ defined by

$$f(a) = \begin{cases} f_X(a), & \text{if } a \in X, \\ f_Y(a), & \text{if } a \in Y, \end{cases}$$

for all $a \in X \sqcup Y$ such that the diagram commutes.

Example 3.4.4. In the category $\mathbf{Vec}_F$, the coproduct of two linear spaces V and W over F is their external direct sum $V \oplus W$ with the inclusion maps $\iota_V : V \to V \oplus W$ and $\iota_W : W \to V \oplus W$. Then for any linear space Z over F with two linear maps $f_V : V \to Z$ and $f_W : W \to Z$, there exists a unique linear map $f : V \oplus W \to Z$ defined by $f(v, w) = f_V(v) + f_W(w)$ for all $(v, w) \in V \oplus W$ such that the diagram commutes.

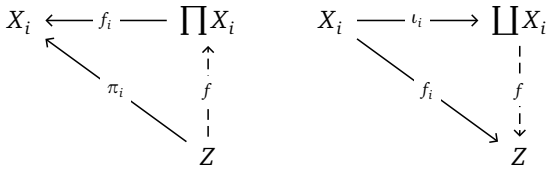
Note that in the category $\mathbf{Vec}_F$, the product and coproduct of two linear spaces V and W over F are isomorphic: $V \sqcap W \cong V \sqcup W \cong V \oplus W$. In this case, we will say the *biproduct* of V and W is $V \oplus W$.

Definition 3.9 — Biproduct.
Let X and Y be two objects in a category $\mathcal{C}$. A *biproduct* of X and Y is an object $X \oplus Y$ in $\mathcal{C}$ that is both a product and a coproduct of X and Y.

Remark. The biproduct exists if and only if both the product and coproduct exist, and it is unique up to isomorphism if it exists.

Example 3.4.5. In the category **Set**, the product and coproduct of two sets X and Y are not isomorphic unless one of them is the empty set: $X \times Y \ncong X \sqcup Y$ if $X \neq \varnothing$ and $Y \neq \varnothing$. So the biproduct does not exist in **Set** in general.

In general, we can define the product, coproduct, and biproduct of a finite collection of objects in a category similarly by using the universal properties or the category of *multi-span* and *multi-cospan*. The following are the commutative diagrams for the universal properties of product and coproduct of multiple objects:



The elements in the product can be expressed as an ordered tuples: $(v_i)_{i \in I}$. The product and coproduct is defined as follows:

$$\prod_{i \in I} V_i = \left\{ (v_i)_{i \in I} \mid v_i \in V_i \text{ for all } i \in I \right\},$$

$$\bigoplus_{i \in I} V_i = \left\{ (v_i)_{i \in I} \in \prod_{i \in I} V_i \;\middle|\; v_i \text{ has finite support} \right\} \subseteq \prod_{i \in I} V_i,$$

where I is a finite index set. Hence, the product and coproduct coincide for finite collections of objects in $\mathbf{Vec}_F$, but not in infinite cases. Consider the following Figure 5 that illustrates the product and coproduct of an infinite collection of linear spaces $\{V_i\}_{i \in I}$ over a field F.

Remark. The right sections s_1 and s_2 are two elements in the product $\prod V_i$. 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

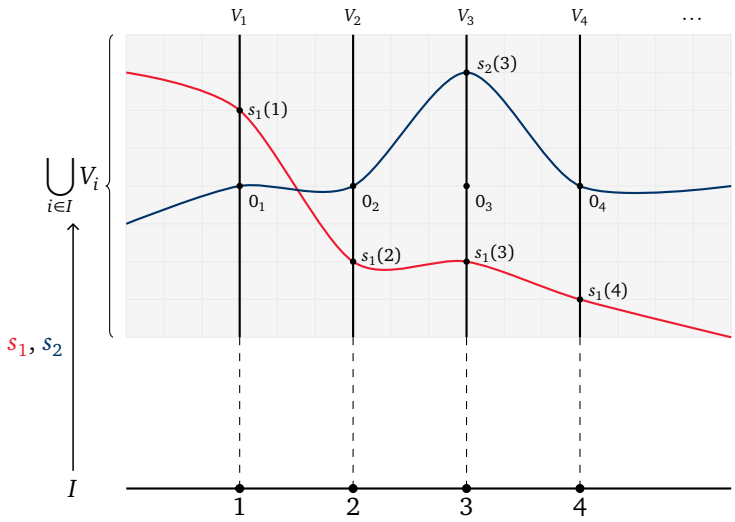


Figure 5. The product and coproduct of an infinite collection of linear spaces $\{V_i\}_{i\in I}$ over a field F. The left section s_1 represents an element in the product $\prod V_i$, and the right section s_2 represents an element in the coproduct $\bigoplus V_i$.

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 $\bigoplus V_i$ if I is an infinite set, but most likely to be.

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

Actually, the product and coproduct can be regarded as the generalisation of $\mathrm{Map}(X, F)$ and $F[X]$ respectively. We can consider the following Figure 6 that illustrates the set map $s : X \to F$.

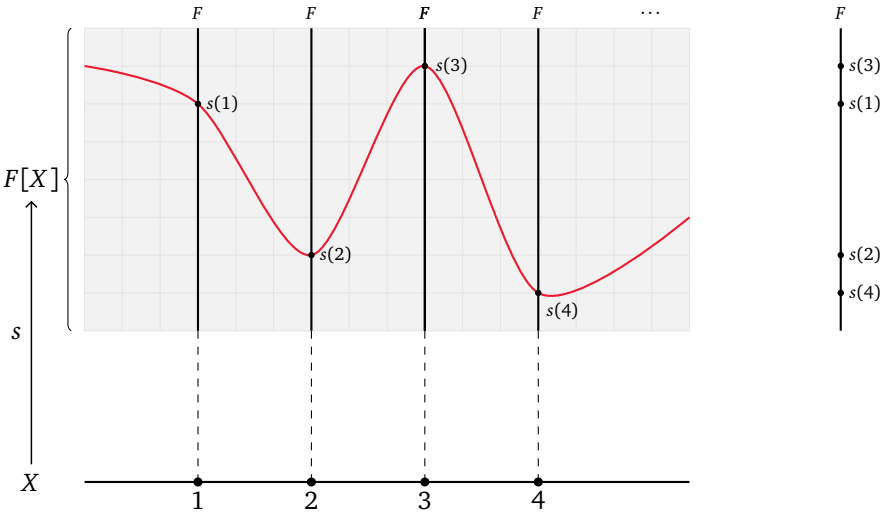


Figure 6. The set map $s : X \to F$ illustrated as sections over fibres.

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 $s: X \to F$ as shown on the right.

3.5. Functors

Definition 3.10 — Functor.
Let $\mathcal{C}$ and $\mathcal{D}$ be two categories. A *functor* $F: \mathcal{C} \to \mathcal{D}$ consists of the following data:

- A map $F_0: \mathcal{C}_0 \to \mathcal{D}_0$ that assigns to each object X in $\mathcal{C}$ an object $F(X)$ in $\mathcal{D}$;
- A map $F_1: \mathcal{C}(X,Y) \to \mathcal{D}(F(X),F(Y))$ that assigns to each morphism $f: X \to Y$ in $\mathcal{C}$ a morphism $F(f): F(X) \to F(Y)$ in $\mathcal{D}$;

satisfying the following properties:

- For any objects X, Y, Z in $\mathcal{C}$ and morphisms $f: Y \to Z$, $g: X \to Y$, we have
$$F(f \circ g) = F(f) \circ F(g);$$
- For any object X in $\mathcal{C}$, we have
$$F(\mathrm{id}_X) = \mathrm{id}_{F(X)}.$$

Example 3.5.1. There are two functors from the category **Set** to the category $\mathbf{Vec}_F$:
$$\mathbf{Set} \underset{|-|}{\overset{F[-]}{\rightleftarrows}} \mathbf{Vec}_F$$
The *free vector space functor* $F[-]: \mathbf{Set} \to \mathbf{Vec}_F$ assigns to each set X the free vector space $F[X]$ over F generated by X, and to each set map $f: X \to Y$ the linear map $F[f]: F[X] \to F[Y]$ induced by f. The *underlying set functor*, or *forgetful functor*, $|-|: \mathbf{Vec}_F \to \mathbf{Set}$ assigns to each linear space V over F its underlying set $|V|$, and to each linear map $g: V \to W$ the set map $|g|: |V| \to |W|$ induced by g.

Remark. The universal property of free vector space over a set can be rephrased as follows: for any set X and any linear space V over F, there is a natural identification:
$$\mathbf{Set}(X, |V|) \simeq \mathbf{Vec}_F(F[X], V)$$
where $\mathbf{Set}(X,|V|)$ is the set of all set maps from X to the underlying set of V, and $\mathbf{Vec}_F(F[X],V)$ is the set of all linear maps from the free vector space $F[X]$ to V.
If we consider $\phi: X \to |V|$ as an element in $\mathbf{Set}(X,|V|)$, then the corresponding element in $\mathbf{Vec}_F(F[X],V)$ is the unique linear map $\bar{\phi}: F[X] \to V$ induced by ϕ.
Note that $\iota \equiv \mathrm{id}_{F[X]}$ is the identity element in $\mathbf{Vec}_F(F[X],F[X])$, so it corresponds to an element in $\mathbf{Set}(X,|F[X]|)$, which is exactly the inclusion map $\iota: X \to |F[X]|$.

Definition 3.11 — Adjoint Functor.
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, denoted by $F \dashv G$, if for any object X in $\mathcal{C}$ and any object Y in $\mathcal{D}$, there exists a natural isomorphism:
$$\mathcal{D}(F(X), Y) \simeq \mathcal{C}(X, G(Y)).$$

Example 3.5.2. The free vector space functor $F[-]: \mathbf{Set} \to \mathbf{Vec}_F$ is a left adjoint of the underlying set functor $|-|: \mathbf{Vec}_F \to \mathbf{Set}$, i.e., $F[-] \dashv |-|$. This follows from the natural isomorphism:
$$\mathbf{Vec}_F(F[X], V) \simeq \mathbf{Set}(X, |V|)$$
for any set X and any linear space V over F.

Definition 3.12 — Endofunctor.
An *endofunctor* is a functor from a category to itself, i.e., $F : \mathcal{C} \to \mathcal{C}$.

Example 3.5.3. 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 $\mathbf{Set}(X, -)$.

$$\begin{array}{ccc} \mathbf{Set} & \xrightarrow{-\times X} & \mathbf{Set} \\ Y & \longmapsto & Y \times X \\ \downarrow f & & \downarrow f \times \mathrm{id}_X \\ Z & & Z \times X \end{array} \qquad \begin{array}{ccc} \mathbf{Set} & \xleftarrow{\mathbf{Set}(X,-)} & \mathbf{Set} \\ \mathbf{Set}(X,Y) & \longmapsto & Y \\ \downarrow \mathbf{Set}(X,f) & & \downarrow f \\ \mathbf{Set}(X,Z) & & Z \end{array}$$

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 $\mathbf{Set}(X, f)(g) = f\,g : X \to Z$.
Then we have the natural isomorphism:

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

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 $x \in X$.s

Remark. In the definition of linear structure, we said that a ring action of F on $(V, +)$ is equivalent to a ring homomorphism from F to the endomorphism ring $\mathsf{End}(V)$. The reason behind this is similar to this example. We have the natural isomorphism:

$$\mathbf{Rng}(F \times V, V) \simeq \mathbf{Rng}(F, \mathsf{End}(V)).$$

Such a process is called *currying*. This kind of natural isomorphism is very common in category theory.

Consider the following two diagrams:

$$\begin{array}{ccccc} X & \longleftarrow & X \times Y & \longrightarrow & Y \\ & & \| & & \\ & & {\scriptstyle F[-]} \Downarrow & & \\ F[X] & \longleftarrow & F[X \times Y] & \longrightarrow & F[Y] \end{array} \qquad \begin{array}{ccccc} X & \longrightarrow & X \sqcup Y & \longleftarrow & Y \\ & & \| & & \\ & & {\scriptstyle F[-]} \Downarrow & & \\ F[X] & \longrightarrow & F[X \sqcup Y] & \longleftarrow & F[Y] \end{array}$$

Moreover, we have the following natural isomorphisms:

$$F[X \times Y] \simeq F[X] \otimes F[Y], \quad F[X \sqcup Y] \simeq F[X] \oplus F[Y].$$

The left shows that the free functor $F[-]$ sends the product in $\mathbf{Set}$ to the tensor product in $\mathbf{Vec}_F$, and the right shows that it sends the coproduct in $\mathbf{Set}$ to the direct sum in $\mathbf{Vec}_F$, i.e., the coproduct in $\mathbf{Vec}_F$. Note that the tensor product $\otimes$ is *not* the product in $\mathbf{Vec}_F$, as the dimension does not match: $\dim(F[X] \otimes F[Y]) = |X| \cdot |Y|$ while $\dim(F[X] \oplus F[Y]) = |X| + |Y|$. There is a unique but not isomorphic linear map $\phi : V \otimes W \to V \oplus W$ defined by $\phi(v \otimes w) = (v, w)$ for all $v \in V$ and $w \in W$.

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

3.6. Dual Spaces and Dual Bases

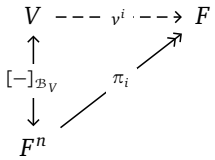
Definition 3.13 — Dual Space.
Let V be a linear space over F. The *dual space* of V, denoted by V^*, is the linear space $\mathrm{Hom}(V,F)$, or $\mathrm{Hom}(V,F)$, consisting of all linear functionals from V to F.

Proposition 3.6.1. Let V be a finite-dimensional linear space over F. Then $\dim(V^*) = \dim(V)$. So V^* is also finite-dimensional.

Proof. Without the loss of generality, we may assume $\dim(V) = n$ and $V = F^n$. Then $V^* = \mathrm{Hom}(F^n, F) \cong \mathrm{Mat}_{1\times n}(F)$, the linear space of row matrices with n entries. The linear space is the span of n standard basis row matrices: $\hat{e}^1, \hat{e}^2, \dots, \hat{e}^n$. So $\dim(V^*) = n = \dim(V)$. □

Proposition 3.6.2. Let V be a finite-dimensional linear space over F. The basis of V and the basis of V^* are in one-to-one correspondence. More precisely, if $\{v_1, v_2, \dots, v_n\}$ is a basis of V, then there exists a unique basis $\{v^1, v^2, \dots, v^n\}$ of V^*, called the *dual basis*, such that $v^i(v_j) = \delta_{ij}$ for all $1 \le i, j \le n$, where δ^i_j is the Kronecker delta.

Proof. Consider the following commutative diagram:



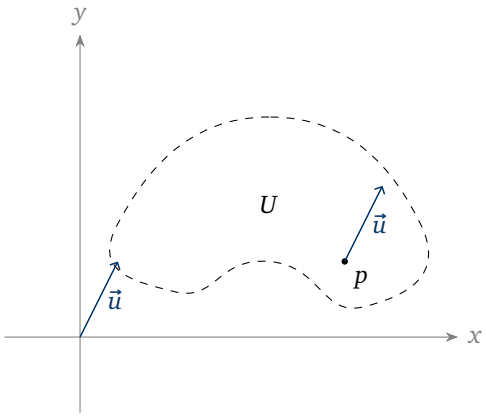
The projection map π_i is a linear functional in F^n that sends $\vec{x} = (x^1, x^2, \dots, x^n)$ to x^i. The projection map π_i is actually $\hat{e}^i$. Note that $[-]_{\mathcal{B}_V} : V \to F^n$ is a coordinate map defined by a basis $\mathcal{B}_V = (v_1, v_2, \dots, v_n)$ such that $[v_j]_{\mathcal{B}_V} = \vec{e}_j$ for all $1 \le j \le n$. It is a unique linear map which identify v_j with $\vec{e}_j$. It can be done by trivialisation of V with respect to the basis v. Then we define $v^i(v_j) = \delta^i_j$ for all $1 \le i, j \le n$.
Then we have to consider whether $(v^1, v^2, \dots, v^n)$ is a basis of V^*. As $\dim(V^*) = n$, we only need to show that $(v^1, v^2, \dots, v^n)$ is a spanning set of V^* or linearly independent. We choose to check if it is linearly independent. We have to check whether the equation $\sum_{i=1}^n x_i v^i = 0$ for some $x_i \in F$ has only the trivial solution. Applying it to v_j for all $1 \le j \le n$, we have $0 = \sum_{i=1}^n x_i v^i(v_j) = \sum_{i=1}^n x_i \delta^i_j = x_j$. So $x_j = 0$ for all $1 \le j \le n$. This means that $(v^1, v^2, \dots, v^n)$ is linearly independent, and hence it is a basis of V^*. We call it the *dual basis* of the basis $\mathcal{B}_V = (v_1, v_2, \dots, v_n)$ and denote it by $\mathcal{B}_{V^*} = (v^1, v^2, \dots, v^n)$. Then we have to show that there is a unique basis in V^* that satisfies $v^i(v_j) = \delta^i_j$. Let $V = F^n$ and $\mathcal{B}_V = (v_1, v_2, \dots, v_n)$ be a basis of V. Then $A = [v_1 \quad v_2 \quad \cdots \quad v_n]$ is an invertible matrix. Let $(v^1, v^2, \dots, v^n)$ be a basis of V^*. Then we have the following equations:

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

Then $(v^1, v^2, \dots, v^n) = A^{-1}$. So the dual basis is unique. □

Remark. There is a bijection between the set of all bases of V and the set of all bases of V^*. However, this bijection is not natural, as it depends on the choice of basis of V. In other words, there is no natural isomorphism between V and V^* but only isomorphic.

Example 3.6.1. Consider the following open subset U of $\mathbb{R}^2$:



Consider the cotangent vector $\mathrm{d}f_p$ at point p for some smooth function $f : U \to \mathbb{R}$. It is a linear functional $\mathrm{d}f_p : \mathsf{T}_pU \to \mathbb{R}$ defined by $\mathrm{d}f_p(\vec{u}) = J_f(p)\vec{u}$ for all $\vec{u} \in \mathsf{T}_pU$. Here T_pU is the tangent space of U at point p, which is a linear space over $\mathbb{R}$. Note that both $\vec{u}$ and $J_f(p)$ 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 $f_x(p)$ and $f_y(p)$.

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

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

Recall the universal property of minimal spanning set in Proposition 2.6.1 and the relation between product and $\mathsf{Map}(X, F)$. We have the following natural isomorphism:

$$\begin{aligned} V^* &= \mathsf{Hom}(V, F) \\ &= \mathbf{Vec}_F(V, F) \\ &\simeq \mathbf{Set}(S, |F|) \quad \text{where } S \text{ is a minimal spanning set of } V \\ &= \mathsf{Map}(S, F) \\ &\simeq \prod_{s \in S} F \end{aligned}$$

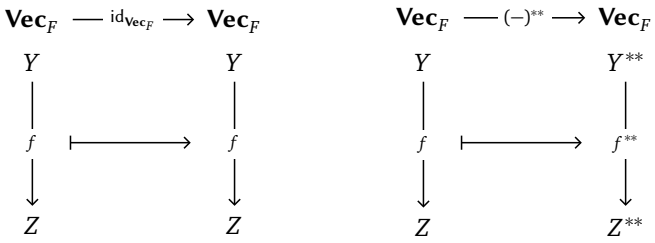
Moreover, V is ismorphic to the coproduct $\bigoplus_{s\in S} F$ since S is a minimal spanning set of V. Hence, if V is infinite-dimensional, then $\dim(V) < \dim(V^*)$.

3.7. Double Dual Spaces and Doubles

Consider the endofunctors on $\mathbf{Vec}_F$:

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

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



Definition 3.14 — Bilinear Map.

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

Definition 3.15 — Natural Pairing.

Let V be a linear space over F. A *natural pairing*, or *evaluation map* on V is a bilinear map $\langle -,-\rangle : V^* \times V \to F$ defined by $\langle f,u\rangle = f(u)$ where $f \in V^*$ and $u \in V$, i.e., a pairing between a covector and a vector.

We have the following natural isomorphism between V and V^{**} induced by the natural pairing:

$$\begin{array}{ccccccc} V^* \times V & \xrightarrow{\langle -,-\rangle} & F & \Longleftrightarrow & V^* & \xrightarrow{1_{V^*}} & \mathrm{Hom}(V,F) \\ \uparrow \\ \mathrm{flip} \\ \downarrow \\ V \times V^* & \longrightarrow & F & \Longleftrightarrow & V & \xrightarrow{\iota_V} & \mathrm{Hom}(V^*,F) \end{array}$$

where $\iota_V : V \to V^{**}$ is defined by $\iota_V(u) = \check{u}$ such that $\check{u}(f) = f(u)$. Then $V^{**} = \mathrm{Hom}(V^*,F) \simeq V$.

Unfortunately, there is no natural isomorphism between V and V^*. Then the professor introduced the concept of *doubles* to tackle this problem.

Definition 3.16 — Double.

Let V be a linear space over 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:

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

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.

3.8. Natural Transformations and Natural Isomorphisms

Definition 3.17 — Natural Transformation.

Let $F, G\colon \mathcal{C} \to \mathcal{D}$ be two functors. A *natural transformation* $\eta\colon F \to G$ 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 3.18 — Natural Isomorphism.

A *natural isomorphism*, or *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 F and G are *naturally isomorphic* or *naturally equivalent*, denoted by $F \simeq G$.

Example 3.8.1. Consider the endofunctors on $\mathbf{Vec}_F$:

$$\mathbf{Vec}_F \underset{\mathrm{id}_{\mathbf{Vec}_F}}{\overset{(-)^{**}}{\rightrightarrows}} \mathbf{Vec}_F$$

We have the following natural transformation:

$$\begin{array}{ccccccc} (-)^{**} & & V_1 & \xrightarrow{f} & V_2 & \longmapsto & V_1^{**} \xrightarrow{f^{**}} V_2^{**} \\ \updownarrow & & & & & & \uparrow{\scriptstyle \eta_{V_1}} \quad \uparrow{\scriptstyle \eta_{V_2}} \\ \mathrm{id}_{\mathbf{Vec}_F} & & V_1 & \xrightarrow{f} & V_2 & \longmapsto & V_1 \xrightarrow{f} V_2 \end{array}$$

Then we have the natural isomorphism: $(-)^{**} \simeq \mathrm{id}_{\mathbf{Vec}_F}$.

Example 3.8.2. We have the natural isomorphism between the following two endofunctors on $\mathbf{Vec}_F$:

$$\mathrm{Bil}(V \times W, -) \simeq \mathrm{Hom}(V, \mathrm{Hom}(W, -))$$

where $\mathrm{Bil}(V \times W, -)$ is the functor that sends a linear space Z to the set of bilinear maps with the source $V \times W$. For any linear space Z over F, we have the natural isomorphism:

$$\natural : \mathrm{Bil}(V \times W, Z) \to \mathrm{Hom}(V, \mathrm{Hom}(W, Z))$$

3.9. Exercises

Problem 3.1. Show that

(a) the endofunctor $(-)^{**}$ on $\mathbf{Vec}_F^{\text{f.d.}}$ that sends T to $T^{**} := (T^*)^*$ is a category isomorphism;

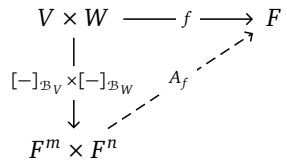
(b) the set of all bilinear maps from $V \times W$ to Z, denoted by $\mathsf{Bil}(V \times W, Z)$, is a linear space;

(c) a bilinear map $f : V \times W \to Z$ naturally defines a linear map $f_\natural : V \to \mathsf{Hom}(W, Z)$;

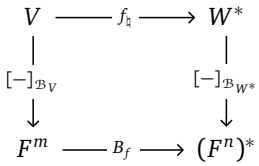
(d) Show that the natural map $\natural : \mathsf{Bil}(V \times W, Z) \to \mathsf{Hom}(V, \mathsf{Hom}(W, Z))$ that sends F to $F_\natural$ is a linear isomorphism. So we write

$$\mathsf{Bil}(V \times W, Z) \simeq \mathsf{Hom}(V, \mathsf{Hom}(W, Z))$$

A bilinear map $f : V \times W \to F$ can be represented by a matrix A_f uniquely defined via the commutative diagram



such that the dashed map is $(x, y) \mapsto x^T A_f y$. Let B_f be the matrix that represents $f_\natural$ with respect to the bases $\mathcal{B}_V$ and $\mathcal{B}_{W^*}$ via the commutative diagram



We say that a bilinear map $f : V \times W \to F$ is *non-degenerate* if $f_\natural : V \to W^*$ is a linear isomorphism.

(e) What is the relationship between A_f and B_f?

(f) Show that a bilinear map $f : V \times W \to F$ is non-degenerate if and only if A_f is an invertible matrix.

Problem 3.2. Let V be a n-dimensional F-linear space and V^* be its dual. Suppose that there are n elements $v_1, \dots, v_n$ in V and n elements $v^1, \dots, v^n$ in V^* such that $\langle v^i, v_j \rangle = \delta^i_j$. Show that $(v_1, \dots, v_n)$ is a basis of V and $(v^1, \dots, v^n)$ is a basis of V^*.

Problem 3.3. Let $T : V_1 \to V_2$ be a linear map and $\mathcal{B}_i$ be a basis of linear space V_i. Denote by $\mathcal{B}_i^*$ the dual basis of $\mathcal{B}_i$ for linear space V_i^*, by $T^* : V_2^* \to V_1^*$ the dual map of T, i.e., the map that sends f to $f \circ T$.

(a) Show that the sequence $V_1 \to V_2 \to V_3$ is exact if and only if its dual sequence $V_1^* \leftarrow V_2^* \leftarrow V_3^*$ is exact, and then conclude that T is injective if and only if T^* is surjective, and T is surjective if and only if T^* is injective.

(b) Let A be the matrix representation of T with respect to bases $\mathcal{B}_1, \mathcal{B}_2$ and A^* be the matrix representation of T^* with respect to bases $\mathcal{B}_2^*, \mathcal{B}_1^*$. Show that A^* and A are transposes of each other.

Problem 3.4. The set of real numbers $\mathbb{R}$ under the order $\leq$ is a category $\mathcal{R}$: objects are real numbers, the morphism set $\mathcal{R}(a, b)$ is the empty set $\emptyset$ if $a > b$ and is $\{a \leq b\}$ other wise. The composition is $a \leq b \circ b \leq c = a \leq c$ and the identity morphisms 1_a are $a \leq a$. Let S be a bounded sest of real numbers. Please find the coproduct and product for the family of objects: $\{s\}_{s \in S}$.

CHAPTER 4

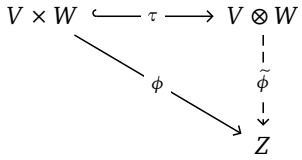
Multilinear Algebras

4.1. Tensor Products

We have learnt what is a bilinear map between linear spaces. However, bilinear maps are not linear maps, so we cannot directly apply the tools we have developed for linear maps to bilinear maps. To fix this problem, we introduce the notion of *tensor products*, which "linearises" bilinear maps. There are several ways to characterise tensor products.

4.1.1. Characterisation via Universal Property. We start with the following universal property that characterises tensor products.

Proposition 4.1.1 — Universal Property of Tensor Product. Let V and W be linear spaces over F. For any linear space Z and any bilinear map $\phi : V \times W \to Z$, there exists a unique linear map $\widetilde{\phi} : V \otimes W \to Z$ such that the following diagram commutes:



Proof. If such a map $\widetilde{\phi}$ exists, then for any $v \in V$ and $w \in W$, we must have

$$\widetilde{\phi}(\tau(v, w)) = \widetilde{\phi}(v \otimes w) = \phi(v, w).$$

Since elements of the form $v \otimes w$ span $V \otimes W$, this determines $\widetilde{\phi}$ uniquely. To show existence, we define $\widetilde{\phi}$ on the basis elements by $\widetilde{\phi}(v \otimes w) := \phi(v, w)$ for all $v \in V, w \in W$, and extend linearly to all of $V \otimes W$. It is straightforward to verify that $\widetilde{\phi}$ is linear and makes the diagram commute. □

This proposition shows that any bilinear map from $V \times W$ to another linear space Z can be uniquely factored through the tensor product $V \otimes W$, i.e., $\mathsf{Bil}(V \times W, Z) \simeq \mathsf{Hom}(V \otimes W, Z)$. Moreover, we have the natural isomorphism between $\mathsf{Bil}(V \times W, -)$ and $\mathsf{Hom}(V, \mathsf{Hom}(W, -))$ in Example 3.8.2. Thus, we have the following natural isomorphism:

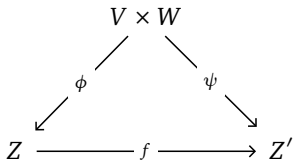
$$\mathsf{Hom}(V \otimes W, -) \simeq \mathsf{Hom}(V, \mathsf{Hom}(W, -)).$$

This shows that the tensor product functor $- \otimes W$ is left adjoint to the Hom functor $\mathsf{Hom}(W, -)$, i.e., we have the following adjoint pair of endofunctors on $\mathbf{Vec}_F$:

$$\mathbf{Vec}_F \underset{\mathsf{Hom}(W,-)}{\overset{-\otimes W}{\rightleftarrows}} \mathbf{Vec}_F$$

Such an adjunction is called a *tensor-Hom adjunction*.

4.1.2. Categorical Characterisation. There is actually another way to characterise tensor products using categories. Consider the category $\mathbf{Bilin}(V, W)$ whose objects are bilinear maps from $V \times W$ to linear spaces in $\mathbf{Vec}_F$, and whose morphisms are commutative diagrams of the following form:



where $\phi : V \times W \to Z$ and $\psi : V \times W \to Z'$ are bilinear maps, and $f : Z \to Z'$ is a linear map such that $f \circ \phi = \psi$. Then the tensor product $V \otimes W$ together with the bilinear map $\tau : V \times W \to V \otimes W$ is a *initial object* in the category $\mathbf{Bilin}(V, W)$, i.e., for any bilinear map $\phi : V \times W \to Z$, there exists a unique morphism from τ to ϕ in $\mathbf{Bilin}(V, W)$.

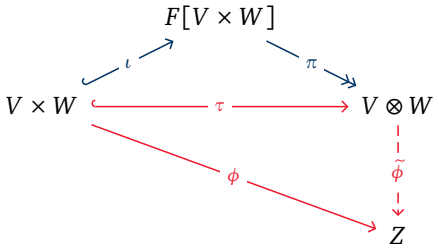
4.1.3. Construction of Tensor Products via Quotient Spaces. Recall that we can construct free vector spaces using sets. We can use this idea to construct tensor products. Consider the free vector space $F[V \times W]$ generated by the set $V \times W$. Then elements of $F[V \times W]$ are finite linear combinations of elements of the form (v, w) for $v \in V, w \in W$. However, $F[V \times W]$ is too large to be the tensor product $V \otimes W$, as it does not satisfy the bilinearity conditions. To fix this, we take the quotient of $F[V \times W]$ by the subspace R spanned by all elements of the following forms:

$$\begin{aligned}
&(v_1 + v_2, w) - (v_1, w) - (v_2, w),\\
&(v, w_1 + w_2) - (v, w_1) - (v, w_2),\\
&(\alpha v, w) - \alpha(v, w),\\
&(v, \alpha w) - \alpha(v, w),
\end{aligned}$$

for all $v, v_1, v_2 \in V$, $w, w_1, w_2 \in W$, and $\alpha \in F$. The reason for taking this quotient is to enforce the bilinearity conditions. Then we define the tensor product $V \otimes W$ as the quotient space:

$$V \otimes W := F[V \times W]/R.$$

The bilinear map $\tau : V \times W \to V \otimes W$ is defined by $\tau(v, w) = [(v, w)]$, the equivalence class of (v, w) in the quotient space. It is straightforward to verify that this construction satisfies the universal property of tensor products. We can conclude this with the following diagram:



Here, the blue arrows represent the construction of the tensor product, where ι is the inclusion map and π is the quotient map. The red arrows represent the universal property of the tensor product, where ϕ is any bilinear map from $V \times W$ to Z, and $\widetilde{\phi}$ is the unique linear map from $V \otimes W$ to Z that makes the diagram commute. The dashed blue arrow $\overline{\phi}$ represents the composition of ϕ with ι, which factors through the quotient map π to give $\widetilde{\phi}$.

Moreover, the inclusion map τ is 'surjective' in the sense that the images of elements of the form $\tau(v, w) = v \otimes w$ span the entire tensor product $V \otimes W$.

4.1.4. Tensor Products of k Linear Spaces. The tensor product can be generalised to more than two linear spaces. Given k linear spaces $V_1, V_2, \ldots, V_k$ over a field F, their tensor product $V_1 \otimes V_2 \otimes \cdots \otimes V_k$ is defined similarly using the universal property. Moreover, the tensor product is associative and commutative up to natural isomorphisms, i.e.,

$$V_1 \otimes V_2 \otimes V_3 \simeq (V_1 \otimes V_2) \otimes V_3 \simeq V_1 \otimes (V_2 \otimes V_3), \quad V_1 \otimes V_2 \simeq V_2 \otimes V_1.$$

We can consider the following diagram to illustrate the associativity and commutativity of tensor products:

$$\begin{array}{ccc} V_1 \times V_2 \times V_3 & \xhookrightarrow{\tau_1} & (V_1 \otimes V_2) \times V_3 \\ \downarrow{\scriptstyle \tau_2} & & \downarrow{\scriptstyle \tau_3} \\ V_1 \times (V_2 \otimes V_3) & \xhookrightarrow{\tau_4} & V_1 \otimes V_2 \otimes V_3 \end{array} \qquad \begin{array}{ccc} V_1 \times V_2 & \xhookrightarrow{\tau} & V_1 \otimes V_2 \\ \downarrow{\scriptstyle \sigma} & & \uparrow{\scriptstyle \widetilde{\sigma}} \\ V_2 \times V_1 & \xhookrightarrow{\tau'} & V_2 \otimes V_1 \end{array}$$

Here σ is the swap map that sends (v_1, v_2) to (v_2, v_1). The $\widetilde{\sigma}$ is the induced isomorphism between $V_1 \otimes V_2$ and $V_2 \otimes V_1$ that makes the diagram commute.

4.1.5. Properties of Tensor Products. We have the following natural isomorphisms for tensor products of linear spaces:

$$- \otimes F \simeq \mathrm{id}_{\mathbf{Vec}_F} \simeq F \otimes - \simeq \mathrm{Hom}(F, -) \simeq (-)^{**}.$$

We have the following natural isomorphism for tensor products of finite-dimensional linear spaces:

$$\mathrm{Hom}(V, W \otimes Z) \simeq \mathrm{Hom}(V, W) \otimes Z.$$

This shows that the Hom functor $\mathrm{Hom}(V, -)$ commutes with tensor products. However, in general, the tensor product functor $- \otimes Z$ does not commute with the Hom functor $\mathrm{Hom}(V, -)$. One reason is that the tensor product functor is a left adjoint, while the Hom functor is a right adjoint; left adjoints do not generally commute with right adjoints. Another reason is that the tensor product is defined via quotienting by certain *finite* linear combinations, while the Hom functor may involve *infinite* linear combinations when the domain is infinite-dimensional.

This natural isomorphism can be used to prove $\mathrm{Hom}(V, W) \simeq V^* \otimes W$ for finite-dimensional linear spaces V and W. We can see this by setting $Z = F$ in the above natural isomorphism:

$$\mathrm{Hom}(V, W \otimes F) \simeq \mathrm{Hom}(V, W) \otimes F.$$

Since $W \otimes F \simeq W$ and $\mathrm{Hom}(V, W) \otimes F \simeq \mathrm{Hom}(V, W)$, we have

$$\mathrm{Hom}(V, W) \simeq V^* \otimes W.$$

Another one is $(V \otimes W)^* \simeq V^* \otimes W^*$ for finite-dimensional linear spaces V and W. We can see this by setting $Z = F$ in the tensor-Hom adjunction:

$$\mathrm{Hom}(V \otimes W, F) \simeq \mathrm{Hom}(V, \mathrm{Hom}(W, F)).$$

Since $\mathrm{Hom}(W, F) \simeq W^*$ and $\mathrm{Hom}(V, W^*) \simeq V^* \otimes W^*$, we have

$$(V \otimes W)^* \simeq V^* \otimes W^*.$$

By considering $V \otimes W \simeq \mathrm{Hom}(V^*, W)$, we can deduce the dimension formula for tensor products:

$$\dim(V \otimes W) = \dim(\mathrm{Hom}(V^*, W)) = \dim(V^*) \cdot \dim(W) = \dim(V) \cdot \dim(W).$$

Under the natural isomorphism $\mathrm{End}(V) \simeq (\mathrm{End}(V))^*$, we can identify the trace map $\mathrm{tr}\colon \mathrm{End}(V) \to F$ with the identity map id_V in $\mathrm{End}(V)$. This gives us a categorical interpretation of the trace map as the evaluation of the identity endomorphism.

We also have the distribution of tensor products and Hom-set over direct sums:

$$V \otimes (W_1 \oplus W_2) \simeq (V \otimes W_1) \oplus (V \otimes W_2).$$
$$\mathrm{Hom}(V, W_1 \oplus W_2) \simeq \mathrm{Hom}(V, W_1) \oplus \mathrm{Hom}(V, W_2),$$
$$\mathrm{Hom}(V_1 \oplus V_2, W) \simeq \mathrm{Hom}(V_1, W) \oplus \mathrm{Hom}(V_2, W).$$

In categorical terms, consider the category $\mathbf{Vec}_F$ as the category of finite-dimensional linear spaces over F and $\mathbf{Vec}_F^{\mathrm{op}}$ as its opposite category. Then the first isomorphism shows that there is a natural isomorphism between functors from $\mathbf{Vec}_F \times \mathbf{Vec}_F \times \mathbf{Vec}_F$ to $\mathbf{Vec}_F$. The second isomorphism shows that there is a natural isomorphism between functors from $\mathbf{Vec}_F^{\mathrm{op}} \times \mathbf{Vec}_F \times \mathbf{Vec}_F$ to $\mathbf{Vec}_F$. The third isomorphism shows that there is a natural isomorphism between functors from $\mathbf{Vec}_F^{\mathrm{op}} \times \mathbf{Vec}_F^{\mathrm{op}} \times \mathbf{Vec}_F$ to $\mathbf{Vec}_F$.

4.2. Tensors

In this section, we introduce tensors as multilinear maps and multilinear forms, and explore the change of basis for tensors. We also define tensor spaces and discuss some special types of tensors, such as symmetric and skew-symmetric tensors.

Definition 4.1 — Multilinear Map.

A *multilinear map* between finite-dimensional F-linear spaces $V_1, V_2, \dots, V_k$ and W is a map $\phi : V_1 \times V_2 \times \cdots \times V_k \to W$ (k times) that is linear in each argument when the other arguments are held fixed; that is, for each $1 \le i \le k$, and for all $v_j \in V_j$ ($j \neq i$), the map

$$V_i \to W, \quad v_i \mapsto \phi(v_1, v_2, \dots, v_k)$$

is a linear map.

Definition 4.2 — Multilinear Form.

A *multilinear form*, or *covariant tensor* or *k-tensor* or *k-linear form*, on a finite-dimensional F-linear space V is a multilinear map $\phi : V^k \to F$ (k times). It is an element of the k-th tensor power of the dual space V^*, i.e., $\phi \in (V^*)^{\otimes k}$.

Specifically, a linear form is simply a linear functional on V, i.e., an element of the dual space V^*. A bilinear form is an element of $V^* \otimes V^*$. To show that the set of all bilinear forms on V is isomorphic to $V^* \otimes V^*$, we consider the following diagram:

$$\mathrm{Bil}(V \times V, F) \longleftrightarrow \mathrm{Hom}(V, V^*) \longleftrightarrow V^* \otimes V^*$$

Here, the first isomorphism is given by the currying process, and the second isomorphism is given by the natural isomorphism $\mathrm{Hom}(V, W) \simeq V^* \otimes W$ for finite-dimensional F-linear spaces V and W.

Moreover, we have the following two important special cases of 2-linear forms.

Definition 4.3 — Symmetric Bilinear Form.

A *symmetric bilinear form* on a finite-dimensional F-linear space V is a multilinear form $\phi : V \times V \to F$ such that $\phi(v, w) = \phi(w, v)$ for all $v, w \in V$. It is an element of the second symmetric power of the dual space V^*, i.e., $\phi \in S^2(V^*)$.

Definition 4.4 — Skew-symmetric Bilinear Form.

A *skew-symmetric bilinear form* on a finite-dimensional F-linear space V is a multilinear form $\phi : V \times V \to F$ such that $\phi(v, w) = -\phi(w, v)$ for all $v, w \in V$. It is an element of the second exterior power of the dual space V^*, i.e., $\phi \in \Lambda^2(V^*)$.

Remark. Here, we assume that the characteristic of the field F is not 2, so that the relation $\phi(v, w) = -\phi(w, v)$ is equivalent to $\phi(v, v) = 0$ for all $v, w \in V$. If the characteristic of F is 2, then the skew-symmetric bilinear form would be defined by the relation $\phi(v, v) = 0$ for all $v \in V$.

Then we can define tensor spaces.

Definition 4.5 — Tensor Space.

The *tensor space* of type (k, l) on a finite-dimensional F-linear space V, denoted by $T^{k,l}(V)$, is defined as

the tensor product of the k-th tensor power of V and the l-th tensor power of the dual space V^*:

$$T^{k,l}(V) = V^{\otimes k} \otimes (V^*)^{\otimes l}.$$

Elements of $T^{k,l}(V)$ are called *tensors of type* (k,l) on V, where k is the number of *contravariant* indices and l is the number of *covariant* indices.

If a tensor is of type $(k,0)$, then it is called a *contravariant tensor* or simply a *k-tensor*, and it is an element of the k-th tensor power of V, i.e., $\mathcal{T}^{k,0}(V) = V^{\otimes k}$. If a tensor is of type $(0,l)$, then it is called a *covariant tensor* or simply an *l-form*, and it is an element of the l-th tensor power of the dual space V^*, i.e., $\mathcal{T}^{0,l}(V) = (V^*)^{\otimes l}$.

Given that $\mathrm{End}(V) \simeq V \otimes V^* = \mathcal{T}^{1,1}(V)$, we can see that tensors of type $(1,1)$ can be interpreted as linear operators on V and represented by a^i_j in a basis. Here, the contravariant index i represents the row index, and the covariant index j represents the column index. To get the matrix representation of the linear operator with respect to a basis, we have

$$a^i_j = \langle \hat{e}^i, A(\vec{e}_j) \rangle,$$

where $\{\vec{e}_j\}$ is a basis of F^n and $\{\hat{e}^i\}$ is the dual basis of F^n. Then we can identify the linear operator T with tensor of type $(1,1)$ as follows:

$$T \simeq T^i_j v_i \otimes v^j.$$

Here, the $\{v_i\}$ form a basis of V, and the $\{v^j\}$ form the dual basis of V^*.

An object is considered as *covariant* if it transforms like the basis vectors under a change of basis, and it is considered as *contravariant* if it transforms oppositely to the basis vectors under a change of basis. We have the following transformation tables for covariant and contravariant objects under a change of basis.

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 of a Covector (v_i)	Covariant

In general, an element $t \in \mathcal{T}^{r,s}(V)$ can be represented in a basis as follows:

$$t = t^{i_1 i_2 \cdots i_r}_{j_1 j_2 \cdots j_s} v_{i_1} \otimes v_{i_2} \otimes \cdots \otimes v_{i_r} \otimes v^{j_1} \otimes v^{j_2} \otimes \cdots \otimes v^{j_s},$$

where the $\{v_i\}$ form a basis of V, and the $\{v^j\}$ form the dual basis of V^*. The representation depends on the choice of basis, i.e., the following two representations are equivalent under a change of basis:

$$\left[t^{i_1 i_2 \cdots i_r}_{j_1 j_2 \cdots j_s} \right]_{\mathcal{B}_V} \simeq \left[\tilde{t}^{\tilde{i}_1 \tilde{i}_2 \cdots \tilde{i}_r}_{\tilde{j}_1 \tilde{j}_2 \cdots \tilde{j}_s} \right]_{\widetilde{\mathcal{B}_V}}$$

The two representations are related by the change of basis matrices as follows:

$$(\tilde{v}_1, \tilde{v}_2, \cdots, \tilde{v}_n) = (v_1, v_2, \cdots, v_n) A,$$

where A is the change of basis matrix, $A = [a^i_{\tilde{j}}]^{\widetilde{\mathcal{B}_V}}_{\mathcal{B}_V}$, which is in the general linear group of V, $\mathrm{GL}(V)$. It is an automorphism group of V that consists of all invertible linear maps from V to itself. We will learn more later.

Remark. It is the right action of $\mathrm{GL}(V)$ on the set of bases of V:

$$\mathcal{B}_V \times \mathrm{GL}(V) \to \mathcal{B}_V, \quad (\mathcal{B}_V, A) \mapsto \mathcal{B}_V A.$$

Then we have the following transformation rule for the components of the tensor under the change of basis:

$$\widetilde{v}_{\widetilde{j}} = v_i a^i_{\widetilde{j}}, \quad v_k = \widetilde{v}_{\widetilde{j}} b^{\widetilde{j}}_k,$$

where B is the inverse of A, i.e., $B = A^{-1}$, and $B = [b^{\widetilde{i}}_j]^{\mathcal{B}_V}_{\widetilde{\mathcal{B}_V}}$. We have $a^i_{\widetilde{j}} b^{\widetilde{k}}_j = \delta^{\widetilde{k}}_{\widetilde{j}}$ and $b^{\widetilde{i}}_j a^j_{\widetilde{k}} = \delta^{\widetilde{i}}_{\widetilde{k}}$.

Remark. For easier memorisation, our professor suggests using the following partial derivative notation to represent the change of basis matrices:

$$\frac{\partial \widetilde{v}_{\widetilde{j}}}{\partial v_i} = a^i_{\widetilde{j}}, \quad \frac{\partial v_k}{\partial \widetilde{v}_{\widetilde{j}}} = b^{\widetilde{j}}_k$$

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}:

$$\frac{\partial \widetilde{v}_{\widetilde{j}}}{\partial v_i} \frac{\partial v_k}{\partial \widetilde{v}_{\widetilde{j}}} = \delta^i_k$$

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 $\widetilde{\mathcal{B}_V}$:

$$\widetilde{t}^{\widetilde{i}_1 \widetilde{i}_2 \cdots \widetilde{i}_r}_{\widetilde{j}_1 \widetilde{j}_2 \cdots \widetilde{j}_s} = \left(b^{\widetilde{i}_1}_{i_1} b^{\widetilde{i}_2}_{i_2} \cdots b^{\widetilde{i}_r}_{i_r} \right) t^{i_1 i_2 \cdots i_r}_{j_1 j_2 \cdots j_s} \left(a^{j_1}_{\widetilde{j}_1} a^{j_2}_{\widetilde{j}_2} \cdots a^{j_s}_{\widetilde{j}_s} \right)$$

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:

$$\mathcal{B}_{\mathcal{T}^{r,s}V} = \{\vec{v}_{i_1} \otimes \vec{v}_{i_2} \otimes \cdots \otimes \vec{v}_{i_r} \otimes \hat{v}^{j_1} \otimes \hat{v}^{j_2} \otimes \cdots \otimes \hat{v}^{j_s} : 1 \le i_1, i_2, \cdots, i_r, j_1, j_2, \cdots, j_s \le n\}$$

For symmetric and skew-symmetric tensors, we have:

$$\mathcal{B}_{\mathcal{S}^k V} = \{v_{i_1} v_{i_2} \cdots v_{i_k} \mid 1 \le i_1 \le i_2 \le \cdots \le i_k \le n\},$$
$$\mathcal{B}_{\Lambda^k V} = \{v_{i_1} \wedge v_{i_2} \wedge \cdots \wedge v_{i_k} \mid 1 \le i_1 < i_2 < \cdots < i_k \le n\}.$$

However, there is a more elegant way to define the basis of symmetric and skew-symmetric tensors. We just need to ensure that the representation of any symmetric tensor is unique for a given basis. For example, for symmetric bilinear forms, we can add the condition that the components remain the same under the swap of any two indices:

$$t = t^{ij} v_i v_j = t^{ji} v_j v_i \implies t^{ij} = t^{ji}.$$

Thus, we can define the basis of symmetric bilinear forms as follows:

$$\mathcal{B}_{\mathcal{S}^2 V} = \{v_i v_j \mid 1 \le i, j \le n\}.$$

Similarly, for skew-symmetric tensors, we can add the condition that the components change sign under the swap of any two indices:

$$t = t^{ij} v_i \wedge v_j = -t^{ji} v_j \wedge v_i \implies t^{ij} = -t^{ji}.$$

Thus, we can define the basis of skew-symmetric bilinear forms as follows:

$$\mathcal{B}_{\Lambda^2 V} = \{v_i \wedge v_j \mid 1 \le i < j \le n\}.$$

In conclusion, we have to make sure that the representation of any symmetric or skew-symmetric multilinear form is unique for a given basis by the following conditions respectively:

$$\text{Symmetric:} \quad t^{i_1 i_2 \cdots i_k} = t^{i_{\sigma(1)} i_{\sigma(2)} \cdots i_{\sigma(k)}}$$
$$\text{Skew-symmetric:} \quad t^{i_1 i_2 \cdots i_k} = \operatorname{sgn}(\sigma) t^{i_{\sigma(1)} i_{\sigma(2)} \cdots i_{\sigma(k)}}$$

The σ is any permutation in the symmetric group S_k. We will learn more about symmetric groups later.

4.3. Multilinear Algebras

4.3.1. Algebras. Before introducing tensor algebras, we first define algebras over fields.

Definition 4.6 — Algebraic Structure.
An *algebraic structure* on a F-linear space A is a bilinear map $\mu: A \times A \to A$, called the *multiplication* on A; or equivalently, a linear map $\mu: A \otimes A \to A$. The pair (A, μ) is called a *F-algebra*.

Example 4.3.1. The set of all polynomials with coefficients in F, denoted by $F[x]$, forms an algebra over F with the usual polynomial multiplication. The multiplication map $\mu: F[x] \times F[x] \to F[x]$ is defined by $\mu(p(x), q(x)) = p(x) \cdot q(x)$ for all $p(x), q(x) \in F[x]$. Moreover, $F[x]$ is a unital commutative associative algebra over F, where the multiplicative identity is the constant polynomial 1.

Example 4.3.2. The set of all square matrices of size n with entries in F, denoted by $\mathrm{Mat}_n(F)$, forms an algebra over F with the usual matrix multiplication. The multiplication map $\mu: M_n(F) \times M_n(F) \to M_n(F)$ is defined by $\mu(A, B) = AB$ for all $A, B \in M_n(F)$. Moreover, $M_n(F)$ is a unital non-commutative associative algebra over F, where the multiplicative identity is the identity matrix I_n. However, $M_n(F)$ is generally not commutative for $n \geq 2$.

Example 4.3.3. The 3-dimensional Euclidean space $\mathbb{R}^3$ forms an algebra over $\mathbb{R}$ with the cross product $\times: \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}^3$ as multiplication. The multiplication map $\mu: \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}^3$ is defined by $\mu(\mathbf{u}, \mathbf{v}) = \mathbf{u} \times \mathbf{v}$ for all $\mathbf{u}, \mathbf{v} \in \mathbb{R}^3$. Moreover, $\mathbb{R}^3$ is a non-unital non-commutative non-associative 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 $\mathfrak{so}(3)$. It is the Lie algebra of the infinitesimal symmetries of a pointed 3-dimensional Euclidean space.

Definition 4.7 — Lie Algebra.
A *Lie algebra* over F is an algebra $\mathfrak{g}$ over F with the *Lie bracket* $[\cdot, \cdot]: \mathfrak{g} \times \mathfrak{g} \to \mathfrak{g}$ as the multiplication satisfying the following properties for all $x, y, z \in \mathfrak{g}$:

(a) (Skew-symmetry) $[x, x] = 0$;

(b) (Jacobi Identity) $[x, [y, z]] + [y, [z, x]] + [z, [x, y]] = 0$.

The skew-symmetry property implies that $[x, y] = -[y, x]$ for all $x, y \in \mathfrak{g}$ if the characteristic of F is not 2.

Definition 4.8 — Algebra Homomorphism.
An *algebra homomorphism* between two algebras (A, μ) and (B, ν) over the same field F is a linear map $\phi: A \to B$ that respects the algebraic structure; that is, for all $u, v \in A$, the following property holds:

$$\phi(\mu(u, v)) = \nu(\phi(u), \phi(v)).$$

Definition 4.9 — Graded Linear Space.
A *$\mathbb{Z}_{\geq 0}$-graded linear space* over F is a linear space V over F together with a decomposition of V into a direct sum of linear subspaces indexed by the integers:

$$V_\bullet = \bigoplus_{n \in \mathbb{Z}_{\geq 0}} V_n,$$

where each V_n is called the *degree n component* of V. An element $v \in V_n$ is said to be *homogeneous of degree n*.

Definition 4.10 — Graded Linear Map.
A *graded linear map* with graded degree k $\phi : V_\bullet \to W_\bullet$ is a linear map if $\phi(V_n) \subseteq W_{n+k}$ for all $n \in \mathbb{Z}_{\geq 0}$.

4.3.2. Tensor Algebras. We are now ready to define tensor algebras.

Definition 4.11 — Tensor Power.
The *k-th tensor power* of a finite-dimensional F-linear space V, denoted by $V^{\otimes k}$, is defined as follows:

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

for all $k \geq 0$. By convention, we define $V^{\otimes 0} = F$.

The dimension of the k-th tensor power is given by the following formula:

$$\dim(V^{\otimes k}) = (\dim(V))^k$$

Definition 4.12 — Tensor Algebra.
The *tensor algebra* of a finite-dimensional F-linear space V is the $\mathbb{Z}_{\geq 0}$-graded linear space over F defined as the external direct sum of all tensor powers of V:

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

equipped with the natural multiplication defined by the tensor product, making $\mathcal{T}^\bullet(V)$ an algebra over F. The multiplication map is defined as follows:

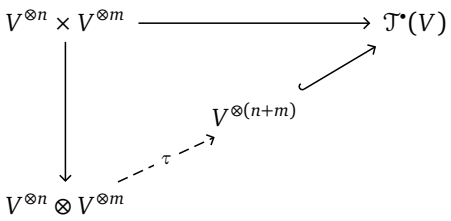
$$\otimes \colon \mathcal{T}^\bullet(V) \times \mathcal{T}^\bullet(V) \to \mathcal{T}^\bullet(V)$$
$$\left(\sum_n u_n, \sum_m v_m\right) \mapsto \sum_{n,m} (u_n \otimes v_m)$$

where $u_n \in V^{\otimes n}$ and $v_m \in V^{\otimes m}$ for all $n, m \geq 0$.

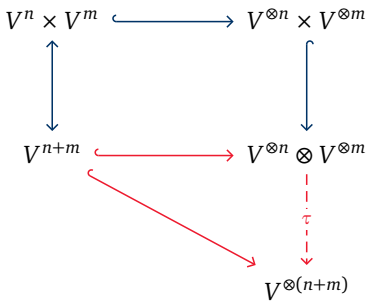
Remark. As the algebra product is bilinear, it suffices to know the product of two homogeneous elements, i.e., $V^{\otimes n} \times V^{\otimes m} \to \mathcal{T}^\bullet(V)$ for all $n, m \geq 0$. So $\mathcal{T}^\bullet(V)$ is a $\mathbb{Z}_{\geq 0}$-graded algebra over F. As the tensor algebra is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

$$\sum_n u_n \otimes \sum_m v_m = \sum_n (u_n \otimes \sum_m v_m) = \sum_n \sum_m (u_n \otimes v_m) = \sum_{n,m} (u_n \otimes v_m)$$

To show that the multiplication map $\otimes \colon \mathcal{T}^\bullet(V) \times \mathcal{T}^\bullet(V) \to \mathcal{T}^\bullet(V)$ is well-defined, we consider the following diagram:



Then, we have to show the existence of τ. Again, we can consider the following diagram:



Here, the blue arrows represent the construction of the tensor products $V^{\otimes n} \otimes V^{\otimes m}$. The red arrows represent the universal property of the tensor product $V^{\otimes(n+m)}$. The dashed red arrow τ is the unique bilinear map that makes the diagram commute. Then we have shown the well-definedness of the multiplication map $\otimes\colon \mathcal{T}^\bullet(V) \times \mathcal{T}^\bullet(V) \to \mathcal{T}^\bullet(V)$.

The tensor algebra $(\mathcal{T}^\bullet(V), \otimes)$ is a graded unital associative algebra over 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 $n, m \geq 0$;

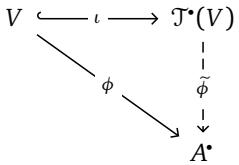
– It is unital, as the multiplicative identity is 1 in F, since $V^{\otimes 0} = F$;

– 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;

– It is non-commutative, as $u \otimes v \neq v \otimes u$ for some $u, v \in V$.

Similarly, we can characterise tensor algebras via a universal property.

Proposition 4.3.1 — Universal Property of Tensor Algebra. Let V be a finite-dimensional F-linear space. For any graded unital associative F-algebra $(A^\bullet, \cdot)$ and any graded linear map with graded degree zero $\phi\colon V \to A^\bullet$, there exists a unique graded algebra homomorphism with graded degree zero $\widetilde{\phi}\colon \mathcal{T}^\bullet(V) \to A^\bullet$ such that the following diagram commutes:



More specifically, here ι is a map that includes V into the degree 1 part of $\mathcal{T}^\bullet(V)$, i.e., $\iota\colon V \to V^{\otimes 1} = V$, and ϕ is a map from V to the degree 1 part of $A^\bullet$, i.e., $\phi\colon V \to A_1$, as we can consider V as a graded linear space concentrated in degree 1.

Proof. If such a map $\widetilde{\phi}$ exists, then for any $v \in V$, we must have

$$\widetilde{\phi}(\iota(v)) = \widetilde{\phi}(0, v, 0, 0, \ldots) = (0, \phi(v), 0, 0, \ldots) = \phi(v).$$

Since elements of the form $v \in V$ generate $\mathcal{T}^\bullet(V)$ as an algebra, this determines $\widetilde{\phi}$ uniquely. To show existence, we define $\widetilde{\phi}$ on the basis elements by $\widetilde{\phi}(v_1 \otimes v_2 \otimes \cdots \otimes v_k) := \phi(v_1) \cdot \phi(v_2) \cdots \phi(v_k)$ for all $v_1, v_2, \ldots, v_k \in V$, and extend linearly to all of $\mathcal{T}^\bullet(V)$. It is straightforward to verify that $\widetilde{\phi}$ is a graded algebra homomorphism with graded degree zero and makes the diagram commute. □

This proposition shows that any graded linear map with graded degree zero from V to a graded unital associative F-algebra $A^\bullet$ can be uniquely factored through the tensor algebra $\mathcal{T}^\bullet(V)$, i.e., $\mathbf{Vec}_F(V, A^\bullet) \simeq \mathbf{Alg}(\mathcal{T}^\bullet(V), A^\bullet)$. Moreover, we have the natural isomorphism between $\mathbf{Alg}(\mathcal{T}^\bullet(V), -)$ and $\mathbf{Vec}_F(V, -)$:

$$\mathbf{Vec}_F(V, |-|) \simeq \mathbf{Alg}(\mathcal{T}^\bullet(V), -).$$

This shows that the tensor algebra functor $\mathcal{T}^\bullet$ is left adjoint to the forgetful functor from the category of graded unital associative algebras over F to the category of linear spaces over F:

$$\mathbf{Vec}_F \underset{|-|}{\overset{\mathcal{T}^\bullet}{\rightleftarrows}} \mathbb{Z}_{\geq 0}-\mathbf{Alg}_F$$

Such an adjunction is called a *free-forgetful adjunction*. We can also see the tensor algebra as the *free graded unital associative algebra* generated by the linear space V:

$$\begin{array}{ccc} \mathbf{Vec}_F & \xrightarrow{\quad \mathcal{T}^\bullet \quad} & \mathbb{Z}_{\geq 0}-\mathbf{Alg}_F \\ V & & \mathcal{T}^\bullet(V) \\ \Big| & & \Big| \\ f & \longmapsto & \mathcal{T}^\bullet(f) \\ \Big\downarrow & & \Big\downarrow \\ W & & \mathcal{T}^\bullet(W) \end{array}$$

Here, $f : V \to W$ is a linear map between linear spaces, and $\mathcal{T}^\bullet(f)\colon \mathcal{T}^\bullet(V) \to \mathcal{T}^\bullet(W)$ is the induced graded algebra homomorphism with graded degree zero between tensor algebras.

4.3.3. Quotient Algebras. We can construct new algebras from existing algebras using quotienting. We will discuss three types of quotient algebras: the symmetric algebra, the exterior algebra, and the universal enveloping algebra. Before that, we first define ideals in algebras.

Definition 4.13 — Ideal.

An *ideal* I in an algebra (A, μ) over F is a linear subspace of A that respects the algebraic structure; that is, for all $a \in A$ and $x \in I$, the following properties hold:

$$\mu(a, x) \in I, \quad \mu(x, a) \in I.$$

This ideal definition is the restriction of the general ideal definition in rings to algebras over fields. We can consider the following example for general ideals in rings.

Example 4.3.4. In the ring of integers $\mathbb{Z}$, the set of all n-multiples $n\mathbb{Z}$ forms an ideal for any integer n. This ideal respects the ringic structure, as for any integer a and any $x \in n\mathbb{Z}$, we have $a \cdot x \in n\mathbb{Z}$ and $x \cdot a \in n\mathbb{Z}$.

4.3.3.1. *Symmetric Algebras.* As the name suggests, symmetric algebras are algebras that are 'commutative' in some sense.

Definition 4.14 — Symmetric Algebra.

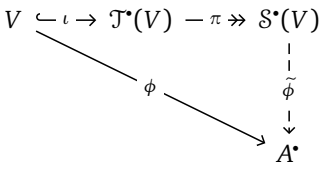
The *symmetric algebra* of a finite-dimensional F-linear space V, denoted by $\mathcal{S}^\bullet(V)$, is the quotient algebra of the tensor algebra $\mathcal{T}^\bullet(V)$ by the *symmetrising ideal* $I_{\mathcal{S}^\bullet}$ generated by all elements of the form $v \otimes w - w \otimes v$ for all $v, w \in V$:

$$\mathcal{S}^\bullet(V) = \mathcal{T}^\bullet(V) \,/\, I_{\mathcal{S}^\bullet} = \mathcal{T}^\bullet(V) \,/\, \langle v \otimes w - w \otimes v \mid v, w \in V \rangle.$$

The symmetrising ideal is the ideal completion of the relation $v \otimes w = w \otimes v$ for all $v, w \in V$, which enforces the commutativity condition in the symmetric algebra. Thus, the symmetric algebra is a graded unital commutative associative algebra over F. We can also characterise symmetric algebras via a universal property.

Proposition 4.3.2 — Universal Property of Symmetric Algebra. Let V be a finite-dimensional F-linear space. For any graded unital commutative associative F-algebra $(A^\bullet, \cdot)$ and any graded linear map with

graded degree zero $\phi : V \to A^\bullet$, there exists a unique graded algebra homomorphism with graded degree zero $\widetilde{\phi} : \mathcal{S}^\bullet(V) \to A^\bullet$ such that the following diagram commutes:



More specifically, here ι is a map that includes V into the degree 1 part of $\mathcal{S}^\bullet(V)$, i.e., $\iota : V \to V^{\otimes 1} = V$, and ϕ is a map from V to the degree 1 part of $A^\bullet$, i.e., $\phi : V \to A_1$, as we can consider V as a graded linear space concentrated in degree 1.

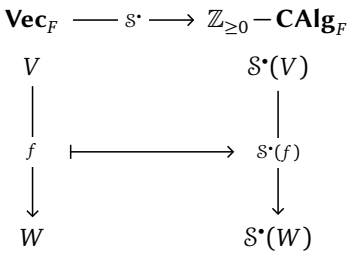
This proposition shows that any graded linear map with graded degree zero from V to a graded unital commutative associative F-algebra $A^\bullet$ can be uniquely factored through the symmetric algebra $\mathcal{S}^\bullet(V)$, i.e., $\mathbf{Vec}_F(V, A^\bullet) \simeq \mathbf{CAlg}(\mathcal{S}^\bullet(V), A^\bullet)$. Moreover, we have the natural isomorphism between $\mathbf{CAlg}(\mathcal{S}^\bullet(V), -)$ and $\mathbf{Vec}_F(V, -)$:

$$\mathbf{Vec}_F(V, |-|) \simeq \mathbf{CAlg}(\mathcal{S}^\bullet(V), -).$$

This shows that the symmetric algebra functor $\mathcal{S}^\bullet$ is left adjoint to the forgetful functor from the category of graded unital commutative associative algebras over F to the category of linear spaces over F:

$$\mathbf{Vec}_F \underset{|-|}{\overset{\mathcal{S}^\bullet}{\rightleftarrows}} \mathbb{Z}_{\geq 0}\text{-}\mathbf{CAlg}_F$$

We can also see the symmetric algebra as the *free graded unital commutative associative algebra* generated by the linear space V:



Here, $f : V \to W$ is a linear map between linear spaces, and $\mathcal{S}^\bullet(f) : \mathcal{S}^\bullet(V) \to \mathcal{S}^\bullet(W)$ is the induced graded algebra homomorphism with graded degree zero between symmetric algebras.

Definition 4.15 — Symmetric Power.

The *k-th symmetric power* of a finite-dimensional F-linear space V, denoted by $\mathcal{S}^k(V)$, is defined as the degree k component of the symmetric algebra $\mathcal{S}^\bullet(V)$:

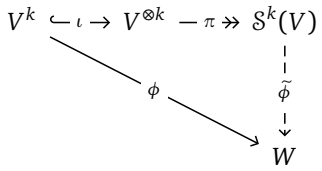
$$\mathcal{S}^\bullet(V) = \bigoplus_{k=0}^{\infty} \mathcal{S}^k(V).$$

The dimension of the k-th symmetric power is given by the following formula:

$$\dim(\mathcal{S}^k(V)) = \binom{\dim(V) + k - 1}{k}.$$

Similarly, we can characterise symmetric powers via a universal property.

Proposition 4.3.3 — Universal Property of Symmetric Power. Let V be a finite-dimensional F-linear space. For any finite-dimensional F-linear space W and any symmetric bilinear map $\phi : V^k \to W$ (k times), there exists a unique linear map $\widetilde{\phi} : \mathcal{S}^k(V) \to W$ such that the following diagram commutes:



4.3.3.2. *Exterior Algebras.* Exterior algebras are algebras that are 'skew-symmetric' in some sense.

Definition 4.16 — Exterior Algebra.

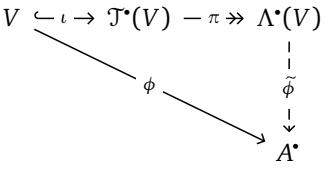
The *exterior algebra* of a finite-dimensional F-linear space V, denoted by $\Lambda^\bullet(V)$, is the quotient algebra of the tensor algebra $\mathcal{T}^\bullet(V)$ by the *antisymmetrising ideal* $I_{\Lambda^\bullet}$ generated by all elements of the form $v\otimes w+w\otimes v$ for all $v,w\in V$:

$$\Lambda^\bullet(V)=\mathcal{T}^\bullet(V)\,/\,I_{\Lambda^\bullet}=\mathcal{T}^\bullet(V)\,/\,\langle v\otimes w+w\otimes v\mid v,w\in V\rangle.$$

Remark. Here, we assume that the characteristic of the field F is not 2, so that the relation $v\otimes w+w\otimes v=0$ is equivalent to $v\otimes w=-(w\otimes v)$ for all $v,w\in V$. If the characteristic of F is 2, then the antisymmetrising ideal would be generated by all elements of the form $v\otimes v$ for all $v\in V$.

The antisymmetrising ideal is the ideal completion of the relation $v\otimes w=-(w\otimes v)$ for all $v,w\in V$, which enforces the skew-symmetry condition in the exterior algebra. Thus, the exterior algebra is a graded unital associative algebra over F. We can also characterise exterior algebras via a universal property.

Proposition 4.3.4 — Universal Property of Exterior Algebra. Let V be a finite-dimensional F-linear space. For any graded unital associative F-algebra $(A^\bullet,\cdot)$ and any graded linear map with graded degree zero $\phi: V\to A^\bullet$ satisfying $\phi(v)\cdot\phi(v)=0$ for all $v\in V$, there exists a unique graded algebra homomorphism with graded degree zero $\widetilde{\phi}:\ \Lambda^\bullet(V)\to A^\bullet$ such that the following diagram commutes:



More specifically, here ι is a map that includes V into the degree 1 part of $\Lambda^\bullet(V)$, i.e., $\iota: V\to V^{\otimes 1}=V$, and ϕ is a map from V to the degree 1 part of $A^\bullet$, i.e., $\phi: V\to A_1$, as we can consider V as a graded linear space concentrated in degree 1.

This proposition shows that any graded linear map with graded degree zero from V to a graded unital associative F-algebra $A^\bullet$ that satisfies the condition $\phi(v)\cdot\phi(v)=0$ for all $v\in V$ can be uniquely factored through the exterior algebra $\Lambda^\bullet(V)$, i.e., $\mathbf{Vec}_F(V,A^\bullet)\simeq\mathbf{Alg}(\Lambda^\bullet(V),A^\bullet)$. Moreover, we have the natural isomorphism between $\mathbf{SAlg}(\Lambda^\bullet(V),-)$ and $\mathbf{Vec}_F(V,-)$:

$$\mathbf{Vec}_F(V,|-|)\simeq\mathbf{SAlg}(\Lambda^\bullet(V),-).$$

This shows that the exterior algebra functor $\Lambda^\bullet$ is left adjoint to the forgetful functor from the category of graded unital associative algebras over F to the category of linear spaces over F:

$$\mathbf{Vec}_F\ \underset{|-|}{\overset{\Lambda^\bullet}{\rightleftarrows}}\ \mathbb{Z}_{\geq 0}-\mathbf{SAlg}_F$$

We can also see the exterior algebra as the *free graded unital skew-symmetric associative algebra* generated by the linear space V:

$$\begin{array}{ccc} \mathbf{Vec}_F & \xrightarrow{\ \Lambda^\bullet\ } & \mathbb{Z}_{\geq 0}\text{-}\mathbf{SAlg}_F \\ V & & \Lambda^\bullet(V) \\ \downarrow f & \longmapsto & \downarrow \Lambda^\bullet(f) \\ W & & \Lambda^\bullet(W) \end{array}$$

Here, $f : V \to W$ is a linear map between linear spaces, and $\Lambda^\bullet(f) \colon \Lambda^\bullet(V) \to \Lambda^\bullet(W)$ is the induced graded algebra homomorphism with graded degree zero between exterior algebras.

Definition 4.17 — Exterior Power.

The *k-th exterior power* of a finite-dimensional F-linear space V, denoted by $\Lambda^k(V)$, is defined as the degree k component of the exterior algebra $\Lambda^\bullet(V)$:

$$\Lambda^\bullet(V) = \bigoplus_{k=0}^{\infty} \Lambda^k(V).$$

The dimension of the k-th exterior power is given by the following formula:

$$\dim(\Lambda^k(V)) = \binom{\dim(V)}{k}$$

Similarly, we can characterise exterior powers via a universal property.

Proposition 4.3.5 — Universal Property of Exterior Power. Let V be a finite-dimensional F-linear space. For any finite-dimensional F-linear space W and any skew-symmetric bilinear map $\phi : V^k \to W$ (k times), there exists a unique linear map $\widetilde{\phi} : \Lambda^k(V) \to W$ such that the following diagram commutes:

$$\begin{array}{ccccc} V^k & \xhookrightarrow{\ \iota\ } & V^{\otimes k} & \xtwoheadrightarrow{\ \pi\ } & \Lambda^k(V) \\ & \searrow^{\phi} & & & \downarrow \widetilde{\phi} \\ & & & & W \end{array}$$

4.3.3.3. *Universal Enveloping Algebras.* Universal enveloping algebras are algebras that 'envelop' Lie algebras in some sense.

Definition 4.18 — Universal Enveloping Algebra.

The *universal enveloping algebra* of a finite-dimensional Lie algebra $\mathfrak{g}$ over F, denoted by $\mathcal{U}(\mathfrak{g})$, is the quotient algebra of the tensor algebra $\mathcal{T}^\bullet(\mathfrak{g})$ by the *Lie ideal* $I_{\mathcal{U}}$ generated by all elements of the form $x \otimes y - y \otimes x - [x, y]$ for all $x, y \in \mathfrak{g}$:

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

The Lie ideal is the ideal completion of the relation $x \otimes y - y \otimes x = [x, y]$ for all $x, y \in \mathfrak{g}$, which enforces the Lie bracket condition in the universal enveloping algebra. Thus, the universal enveloping algebra is a unital associative algebra over F. However, it is not graded, as the ideal is not homogeneous with respect to the grading of the tensor algebra: $x \otimes y - y \otimes x$ is in degree 2, while $[x, y]$ is in degree 1.

4.3.4. Hilbert-Poincaré Series. We use Hilbert-Poincaré series to encode the dimension information of graded linear spaces.

Definition 4.19 — Hilbert-Poincaré Series.

Let $V_\bullet = \bigoplus_{n\in\mathbb{Z}_{\geq 0}} V_n$ be a $\mathbb{Z}_{\geq 0}$-graded linear space over F with each degree component V_n being finite-dimensional. The *Hilbert-Poincaré series* of $V_\bullet$ is the formal power series defined as follows:

$$P_{V_\bullet}(t) = \sum_{n=0}^{\infty} (\dim(V_n)) t^n.$$

Example 4.3.5. The Hilbert-Poincaré series of the tensor algebra $\mathcal{T}^\bullet(V)$ of a finite-dimensional F-linear space V is given by the following formula:

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

Example 4.3.6. The Hilbert-Poincaré series of the symmetric algebra $\mathcal{S}^\bullet(V)$ of a finite-dimensional F-linear space V is given by the following formula:

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

Example 4.3.7. The Hilbert-Poincaré series of the exterior algebra $\Lambda^\bullet(V)$ of a finite-dimensional F-linear space V is given by the following formula:

$$P_{\Lambda^\bullet(V)}(t) = \sum_{n=0}^{\infty} (\dim(\Lambda^n(V))) = \sum_{n=0}^{\infty} \binom{\dim(V)}{n} t^n = (1+t)^{\dim(V)}.$$

As the Hilbert-Poincaré series of the exterior algebra is a polynomial of degree $\dim(V)$, we have $\Lambda^k(V) = 0$ for all $k > \dim(V)$. Moreover, if $\dim(V) = n$, then $\Lambda^n(V)$ is a 1-dimensional linear space called the *top exterior power* of V. The reason is that any $n+1$ vectors in an n-dimensional linear space are linearly dependent, so their exterior product is zero. Also, $\dim(\Lambda^k(V)) = \dim(\Lambda^{n-k}(V))$ for all $0 \leq k \leq n$, as $\binom{n}{k} = \binom{n}{n-k}$.

Definition 4.20 — Line.

A *line* over F is a 1-dimensional linear space over F.

4.4. Exercises

Problem 4.1. Show that there is a functor $\otimes$ from the product category $\mathbf{Vec}_F^{\text{f.d.}} \times \mathbf{Vec}_F^{\text{f.d.}}$ to the category $\mathbf{Vec}_F^{\text{f.d.}}$ that sends an object (V_1, V_2) to $V_1 \otimes V_2$ and a morphism (f, g) to $f \otimes g$. Show that $f \otimes g$ is bilinear, meaning it is linear in both f and g. Finally, show that the functors $- \otimes V$, $\mathrm{Hom}(V, -)$ and $\mathrm{Hom}(-, V)$ preserve exactness: If $A \to B \to C$ is exact, then the sequence $A \otimes V \to B \otimes V \to C \otimes V$, $\mathrm{Hom}(V, A) \to \mathrm{Hom}(V, B) \to \mathrm{Hom}(V, C)$ and $\mathrm{Hom}(A, V) \leftarrow \mathrm{Hom}(B, V) \leftarrow \mathrm{Hom}(C, V)$ are exact.

Problem 4.2. (a) Show that functors $(-)^{**}$, $- \otimes F$, $F \otimes -$, $\mathrm{Hom}(F, -)$ and the identity functor 1 are all naturally equivalent endofunctors on the category $\mathbf{Vec}_F^{\text{f.d.}}$. A simpler way to record these facts is to write

$$V^{**} \equiv V \otimes F \equiv F \otimes V \equiv \mathrm{Hom}(F, V) \equiv V$$

(b) Show that

$$\mathrm{Hom}(V_1, V_2 \otimes V_3) \equiv \mathrm{Hom}(V_1, V_2) \otimes V_3.$$

Consequently, $\mathrm{Hom}(V_1, V_2) \equiv V_1^* \otimes V_2$ and $(V_1 \otimes V_2)^* \equiv V_1^* \otimes V_2^*$.

(c) $\dim V_1 \otimes V_2 = \dim V_1 \cdot \dim V_2$. Moreover, if e_i is a minimal spanning set of V_1 and f_j is a minimal spanning set of V_2, then $e_i \otimes f_j$ is a minimal spanning set of $V_1 \otimes V_2$.

(d) Show that

$$V_1 \otimes V_2 \equiv V_2 \otimes V_1, \qquad \mathrm{End}\, V \equiv (\mathrm{End}\, V)^*$$

(e) Under the natural identification $\mathrm{End}\, V \equiv (\mathrm{End}\, V)^*$, id_V is identified with a linear map $\mathrm{tr} : \mathrm{End}\, V \to F$. Show that tr is cyclic, i.e. $\mathrm{tr}(TS) = \mathrm{tr}(ST)$, and $\mathrm{tr}\, \mathrm{id}_V = \dim V$. This map is called the trace map.

Problem 4.3. Let the category $\mathbf{Vec}_F^{\text{f.d.}}$ be denoted by $\mathcal{V}$. Denote by $\mathcal{V}^{\text{op}}$ be the opposite category of $\mathcal{V}$. Show that

(a) $V_1 \otimes (V_2 \oplus V_3) \equiv (V_1 \otimes V_2) \oplus (V_1 \otimes V_3)$. This is a natural equivalene of two functors from $\mathcal{V} \times \mathcal{V} \times \mathcal{V}$ to $\mathcal{V}$;

(b) $\mathrm{Hom}(V_1, V_2 \oplus V_3) \equiv \mathrm{Hom}(V_1, V_2) \oplus \mathrm{Hom}(V_1, V_3)$. This is a natural equivalene of two functors from $\mathcal{V}^{\text{op}} \times \mathcal{V} \times \mathcal{V}$ to $\mathcal{V}$;

(c) $\mathrm{Hom}(V_1 \oplus V_2, V_3) \equiv \mathrm{Hom}(V_1, V_3) \times \mathrm{Hom}(V_2, V_3)$. This is a natural equivalene of two functors from $\mathcal{V}^{\text{op}} \times \mathcal{V}^{\text{op}} \times \mathcal{V}$ to $\mathcal{V}$.

CHAPTER 5

Determinants

5.1. Determinant Lines

Recall that for a n-dimensional F-linear space V, the top exterior power $\Lambda^n V$ is a one-dimensional F-linear space. Such a one-dimensional linear space is also called a *line*. Then we have the following definition.

Definition 5.1 — Determinant Line.

The *determinant line* of a n-dimensional F-linear space V, denoted $\det(V)$, is defined to be the top exterior power of V:

$$\det(V) := \Lambda^n V.$$

Note that the det operator is the same as $\Lambda^{\dim}$ operator, i.e., det is a functor the category of n-dimensional linear spaces $\mathbf{Vec}_F^n$ to the category of lines $\mathbf{Vec}_F^1$:

$$\begin{array}{ccc} \mathbf{Vec}_F^n & \xrightarrow{\ \det\ } & \mathbf{Vec}_F^1 \\ V & & \det(V) \\ \downarrow f & \longmapsto & \downarrow \det(f) \\ W & & \det(W) \end{array}$$

Here, for a linear map $f : V \to W$, the induced map $\det(f)\colon\ \det(V) \to \det(W)$ is defined by

$$\det(f)(v_1 \wedge v_2 \wedge \cdots \wedge v_n) := f(v_1) \wedge f(v_2) \wedge \cdots \wedge f(v_n).$$

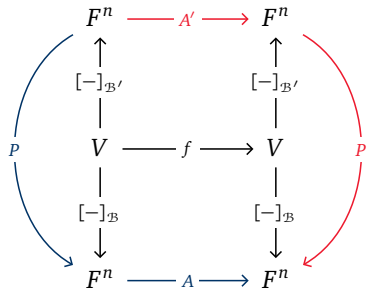
Moreover, as det is a functor, it preserves composition and identities, i.e., for linear maps $f : V_1 \to V_2$ and $g : V_2 \to V_3$, we have

$$\det(f \circ g) = \det(f) \circ \det(g), \quad \text{and} \quad \det(\mathrm{id}_V) = \mathrm{id}_{\det(V)}.$$

If f is an endomorphism on V, i.e., $f : V \to V$, then $\det(f)$ is an endomorphism on the line $\det(V)$. Since any endomorphism on a one-dimensional space is just a scalar multiplication, there exists a unique scalar $\lambda \in F$ such that

$$\det(f)(\omega) = \lambda\omega, \quad \text{for all } \omega \in \det(V).$$

Then we can identify $\det(f)$ with this scalar λ and called it the *determinant* of f. Recall that we can trivialise a linear space by choosing a basis. Then consider the following diagram:



where V is an n-dimensional F-linear space, $\mathcal{B}$ and $\mathcal{B}'$ are two bases of V, A and A' are the matrix representations of f with respect to the bases $\mathcal{B}$ and $\mathcal{B}'$ respectively, and P is the change-of-basis matrix from $\mathcal{B}$ to $\mathcal{B}'$. Then we have

$$AP = PA', \quad \text{or equivalently,} \quad A = PA'P^{-1}.$$

Moreover, the $\det(A)$ is defined to be the determinant of the corresponding linear map f, i.e., $\det(A) := \det(f)$, which is the same as in the ordinary linear algebra. Also, A and A' are *similar* matrices in ordinary linear algebra, meaning they represent the same endomorphism under different bases. Thus, they have the same determinant. Hence, the determinant of a matrix is independent of the choice of basis.

5.2. Permutation Groups

Before we proceed to derive the explicit formula for the determinant of a linear map, we need to introduce the concept of automorphism groups and permutation groups.

Definition 5.2 — Automorphism Group.

The *automorphism group* of a set X, denoted $\mathrm{Aut}(X)$, is the set of all automorphisms of X that forms a group under the composition of functions.

Example 5.2.1. The general linear group of V, denoted by $\mathrm{GL}(V)$, is the automorphism group of the F-linear space V:

$$\mathrm{GL}(V) = \mathrm{Aut}(V).$$

That is the set of all invertible linear maps from V to itself forms a group under the composition of functions.

Example 5.2.2. The general linear group over F of degree n, denoted by $\mathrm{GL}_n(F)$, is the automorphism group of the n-dimensional F-linear space F^n:

$$\mathrm{GL}_n(F) = \mathrm{Aut}(F^n).$$

That is the set of all invertible $n \times n$ matrices with entries in F forms a group under the matrix multiplication.

Definition 5.3 — Permutation Group.

The *permutation group* on a set X, denoted by S_X or $\mathrm{Aut}(X)$, is the automorphism group of X when X is a finite set. If $X = \{1, 2, \ldots, n\}$ for some $n \in \mathbb{N}$, then we denote the permutation group on X by S_n.

The order of the permutation group S_n, denoted by $|S_n|$, is $n!$ since there are $n!$ possible bijections from the set $\{1, 2, \ldots, n\}$ to itself.

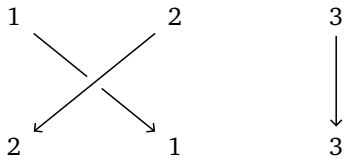
Example 5.2.3. The permutation group S_2 has two elements: the identity permutation 1 and the transposition σ_1 defined by $\sigma_1(1) = 2$ and $\sigma_1(2) = 1$.

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:

$$\sigma_i(j) = \begin{cases} j+1, & j = i \\ j-1, & j = i+1 \quad = (i, i+1) \\ j, & \text{otherwise} \end{cases}$$

Example 5.2.4. The generator σ_1 of S_3 can be represented by the following diagram:



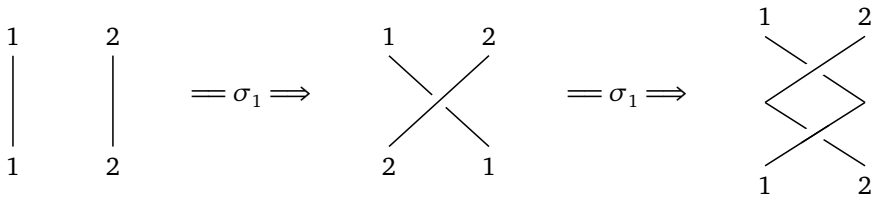
It can also be written as $\sigma_1 = (12)$ or $(12)(3)$ or $\begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix}$. Moreover, we have a cycle with 3 elements denoted as (123) defined by the $\begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}$. Then the presentation of S_3 is:

$$S_3 = \langle \sigma_1, \sigma_2 \mid \sigma_1^2 = 1, \sigma_2^2 = 1, \sigma_1\sigma_2\sigma_1 = \sigma_2\sigma_1\sigma_2 \rangle$$

In general, the presentation of S_n has generators $\sigma_1, \sigma_2, \ldots, \sigma_{n-1}$ and relations:

- *Involution relations*: $\sigma_i^2 = 1$ for all $1 \leq i \leq n-1$;
- *Braid relations*: $\sigma_i\sigma_{i+1}\sigma_i = \sigma_{i+1}\sigma_i\sigma_{i+1}$ for all $1 \leq i \leq n-2$;
- *Commutation relations*: $\sigma_i\sigma_j = \sigma_j\sigma_i$ for all $|i-j| \geq 2$.

The permutation group S_n is generated by quotienting the braid group B_n by the involution relations. 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:

$$1 \longrightarrow A_n \hookrightarrow S_n \xrightarrow{\text{sgn}} \mathbb{Z}/2\mathbb{Z} \longrightarrow 1$$

where A_n is the *alternating group* on n elements, i.e., the subgroup of S_n consisting of all even permutations, and $\text{sgn}\colon S_n \to \mathbb{Z}/2\mathbb{Z} = \{\pm 1\}$, the *sign homomorphism*, is the unique group homomorphism such that $\text{sgn}(\sigma_i) = -1$ for all $1 \leq i \leq n-1$. Note that $\ker(\text{sgn}) = A_n$ and $\text{im}(\text{sgn}) = \mathbb{Z}/2\mathbb{Z}$.

Remark. A_n is simple for all $n \geq 5$, i.e., A_n has no non-trivial normal subgroups for all $n \geq 5$.

Then we have two properties of the sign homomorphism:

- $\text{sgn}(1) = 1$;
- $\text{sgn}(\sigma\tau) = \text{sgn}(\sigma) \cdot \text{sgn}(\tau)$ for all $\sigma, \tau \in S_n$.

5.3. Determinant Formula

The permutation group S_n acts on V^n by permuting the factors:

$$\sigma : (v_1, v_2, \dots, v_n) \mapsto (v_{\sigma(1)}, v_{\sigma(2)}, \dots, v_{\sigma(n)}).$$

Recall the universal property of the exterior power in Proposition 4.3.5, we have the following commutative diagram:

$$\begin{array}{ccc} V^n & \xrightarrow{\quad \sigma \quad} & V^n \\ \downarrow{\scriptstyle \phi} & & \downarrow{\scriptstyle \phi} \\ \Lambda^n V & \xrightarrow{\quad \Lambda^n\sigma \quad} & \Lambda^n V \end{array}$$

The induced map $\Lambda^n\sigma : \Lambda^n V \to \Lambda^n V$ is defined by

$$\Lambda^n\sigma(v_1 \wedge v_2 \wedge \cdots \wedge v_n) := (v_{\sigma(1)} \wedge v_{\sigma(2)} \wedge \cdots \wedge v_{\sigma(n)}) = \operatorname{sgn}(\sigma)(v_1 \wedge v_2 \wedge \cdots \wedge v_n).$$

Now, we are ready to derive the explicit formula for the determinant of a linear map. Consider the following diagram:

$$\begin{array}{ccc} \mathbf{Vec}_F^n & \xrightarrow{\quad \det \quad} & \mathbf{Vec}_F^1 \\ F^n & & \det(F^n) \\ \downarrow{\scriptstyle A} & & \downarrow{\scriptstyle \det(A)} \\ F^n & & \det(F^n) \end{array}$$

Here, $A \in \mathrm{Mat}_n(F)$ is an invertible matrix representing a linear map $f : F^n \to F^n$. Let $\{\vec{e}_1, \vec{e}_2, \dots, \vec{e}_n\}$ be the standard basis of F^n. Then $\{\vec{e}_1 \wedge \cdots \wedge \vec{e}_n\}$ is a basis of the line $\det(F^n)$. To find the scalar $\det(A)$, we compute the action of $\det(A)$ on the basis element $\vec{e}_1 \wedge \cdots \wedge \vec{e}_n$ as follows:

$$\begin{aligned} \det(A)(\vec{e}_1 \wedge \cdots \wedge \vec{e}_n) &= (A\vec{e}_1) \wedge \cdots \wedge (A\vec{e}_n) \\ &= \left(\sum_{i_1=1}^{n} a_1^{i_1}\vec{e}_{i_1}\right) \wedge \cdots \wedge \left(\sum_{i_n=1}^{n} a_n^{i_n}\vec{e}_{i_n}\right) \\ &= \sum_{i_1,i_2,\dots,i_n=1}^{n} a_1^{i_1} \cdots a_n^{i_n}(\vec{e}_{i_1} \wedge \cdots \wedge \vec{e}_{i_n}) \\ &= \sum_{\sigma \in S_n} a_1^{\sigma(1)} \cdots a_n^{\sigma(n)}(\vec{e}_{\sigma(1)} \wedge \cdots \wedge \vec{e}_{\sigma(n)}) \\ &= \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) a_1^{\sigma(1)} \cdots a_n^{\sigma(n)}(\vec{e}_1 \wedge \cdots \wedge \vec{e}_n). \end{aligned}$$

Hence, we have derived the explicit formula for the determinant of the matrix A:

$$\det(A) = \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) a_1^{\sigma(1)} \cdots a_n^{\sigma(n)}.$$

This is also known as the *Leibniz formula* for the determinant.

5.4. Properties of Determinants

5.4.1. Determinant and Dual Basis. We now consider the relationship between the determinant and the dual basis. Let V be an n-dimensional F-linear space with basis $\mathcal{B}_V = \{v_1, v_2, \dots, v_n\}$. As $\det(V)$ is a one-dimensional F-linear space, it has a basis. Moreover, the basis of $\det(V)$, $\mathcal{B}_{\det(V)}$, can be

identified with $\det(V) \setminus \{0\}$. Then we have a map from the basis $\mathcal{B}_V$ of V to the basis $\mathcal{B}_{\det(V)}$ of $\det(V)$ defined by

$$(v_1, v_2, \ldots, v_n) \mapsto v_1 \wedge v_2 \wedge \cdots \wedge v_n.$$

Recall that we have a natural bijection between the basis of V and the basis of the dual space V^* in Proposition 3.6.2. Then we have the following commutative diagram:

$$\begin{array}{ccc} \mathcal{B}_V & \longleftrightarrow & \mathcal{B}_{V^*} \\ \downarrow & & \downarrow \\ \mathcal{B}_{\det(V)} & \dashleftarrow\dashrightarrow & \mathcal{B}_{\det(V^*)} \end{array}$$

Then we can define a map from the basis $\mathcal{B}_{\det(V)}$ of $\det(V)$ to the basis $\mathcal{B}_{\det(V^*)}$ of $\det(V^*)$ induced by the above diagram. This map is defined by

$$(v_1 \wedge v_2 \wedge \cdots \wedge v_n) \mapsto (v_1^* \wedge v_2^* \wedge \cdots \wedge v_n^*),$$

where $\{v_1^*, v_2^*, \ldots, v_n^*\}$ is the dual basis of $\mathcal{B}_V$. Note that this map is an isomorphism between the lines $\det(V)$ and $\det(V^*)$. Moreover, we have the following isomorphisms for the basis elements:

$$\det(v^*) \cong \det(v) \cong \det(v)^*.$$

The first isomorphism is given by the above map, and the second isomorphism is given by the natural isomorphism between a linear space and its dual space.

5.4.2. Determinant and Dual Map. Consider V and W be two n-dimensional F-linear spaces, and let $T : V \to W$ be a linear map. Then we have the following commutative diagram:

$$\begin{array}{ccccccccc} V & \xrightarrow{\quad T \quad} & W & \Longrightarrow^{(-)^*} & V^* & \xleftarrow{\quad T^* \quad} & W^* \\ & \det \Big\downarrow & & & & \det \Big\downarrow & \\ \det(V) & \xrightarrow{\quad \det(T) \quad} & \det(W) & \dashrightarrow & \det(V^*) & \xleftarrow{\quad \det(T^*) \quad} & \det(W^*) \end{array}$$

Consider the left part of the diagram. The map $\det(T)\colon \det(V) \to \det(W)$ is in $\mathrm{Hom}(\det(V), \det(W)) \simeq (\det(V))^* \otimes \det(W)$. While the right part of the diagram, the map $\det(T^*)\colon \det(W^*) \to \det(V^*)$ is in $\mathrm{Hom}(\det(W^*), \det(V^*)) \simeq (\det(W^*))^* \otimes \det(V^*) \simeq \det(W) \otimes \det(V)^*$. Hence, we have the following isomorphism:

$$\det(T) \cong \det(T^*).$$

5.4.3. Properties of Determinants. From the definition of the determinant, we can derive several important properties of determinants as follows:

– Linear in each column or row;

– Alternating in columns or rows;

– $\det(I) = 1$.

These properties uniquely characterise the determinant function up to a scalar multiple. For the alternating property, assume the characteristic of the field F is not 2. Then swapping two rows or two columns of a matrix changes the sign of the determinant. To see this, consider swapping the i-th and j-th vectors $\vec{a}_i$ and $\vec{a}_j$ in the wedge product $\cdots \wedge \vec{a}_i \wedge \cdots \wedge \vec{a}_j \wedge \cdots$. There are k vectors between $\vec{a}_i$ and $\vec{a}_j$. Then we have

$$\cdots \vec{a}_i \underbrace{\cdots}_{k \text{ times}} \vec{a}_j \cdots = (-1)^k \cdots \vec{a}_i \vec{a}_j \cdots = (-1)^{k+1} \cdots \vec{a}_j \vec{a}_i \cdots = - \cdots \vec{a}_j \cdots \vec{a}_i \cdots$$

If we dropped the property $\det(I) = 1$, then the function is called the *alternating multilinear form*. Suppose $\phi : \mathrm{Mat}_n(F) \to F$ is an alternating multilinear form. Then for any matrix $A \in \mathrm{Mat}_n(F)$, we have

$$\phi(A) = \det(A) \cdot \phi(I).$$

That is, any alternating multilinear form is a scalar multiple of the determinant function.

Instead of writing det as a function from matrices to scalars, we can use two pipes to denote the determinant of a matrix.

Proposition 5.4.1. We have the following property of determinants:

$$\begin{vmatrix} A_1 & * \\ 0 & A_2 \end{vmatrix} = \det(A_1) \cdot \det(A_2).$$

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:

$$\begin{aligned} \begin{vmatrix} A_1 & * \\ 0 & A_2 \end{vmatrix} &= \det(A_1) \cdot \begin{vmatrix} I_{n_1} & * \\ 0 & A_2 \end{vmatrix} \\ &= \det(A_1) \cdot \det(A_2) \cdot \begin{vmatrix} I_{n_1} & * \\ 0 & I_{n_2} \end{vmatrix} \\ &= \det(A_1) \cdot \det(A_2) \cdot \begin{vmatrix} I_{n_1} & 0 \\ 0 & I_{n_2} \end{vmatrix} \\ &= \det(A_1) \cdot \det(A_2) \cdot \det(I_{n_1+n_2}) \\ &= \det(A_1) \cdot \det(A_2). \end{aligned}$$

Note that in the third equality, we eliminated the $*$ by using the rows below. □

Concretely, we have the following determinants:

$$\left|\begin{array}{ccc|ccc} 1 & & & * & * & * \\ & 1 & & * & * & * \\ \hline & & & 1 & & \\ & & & & 1 & \\ & & & & & 1 \end{array}\right| = \left|\begin{array}{ccc|ccc} 1 & & & 0 & 0 & 0 \\ & 1 & & * & * & * \\ \hline & & & 1 & & \\ & & & & 1 & \\ & & & & & 1 \end{array}\right| = \left|\begin{array}{ccc|ccc} 1 & & & 0 & 0 & 0 \\ & 1 & & 0 & 0 & 0 \\ \hline & & & 1 & & \\ & & & & 1 & \\ & & & & & 1 \end{array}\right| = \det(I_5) = 1$$

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 an upper-triangular matrix, the determinant is the product of the diagonal entries. Similarly, for a lower-triangular matrix, the determinant is also the product of the diagonal entries. In particular, we have the following equation:

$$\begin{vmatrix} a_{11} & & * \\ & \ddots & \\ 0 & & a_{nn} \end{vmatrix} = a_{11} \cdots a_{nn}$$

This also holds for diagonal matrices as they are both upper-triangular and lower-triangular. Moreover, $\det[a] = a \cdot \det(I) = a$.

> **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.

5.4.4. Cofactor Expansion. We have a famous formula called the *cofactor expansion* or *Laplace expansion* for determinants. It is a recursive formula that expresses the determinant of a matrix in terms of the determinants of its smaller submatrices.

Consider the following determinant:

$$\begin{vmatrix} & * & & 0 & & * & \\ a_{i,1} & \cdots & a_{i,j-1} & 1 & a_{i,j+1} & \cdots & a_{i,n} \\ & * & & 0 & & * & \end{vmatrix} = (-1)^{i-1} \begin{vmatrix} a_{i,1} & \cdots & a_{i,j-1} & 1 & a_{i,j+1} & \cdots & a_{i,n} \\ & * & & 0 & & * & \end{vmatrix}$$

$$= (-1)^{i-1+j-1} \begin{vmatrix} 1 & a_{i,1} & \cdots & \widehat{a_{i,j}} & \cdots & a_{i,n} \\ 0 & & & A_j^i & & \end{vmatrix}$$

$$= (-1)^{i+j} \det A_j^i$$

Here, $\widehat{a_{i,j}}$ means that the element $a_{i,j}$ is omitted, and A_j^i is the submatrix obtained by deleting the i-th row and j-th column of A and is called the *minor* of $a_{i,j}$. Then we have the following cofactor expansion along the i-th row:

$$\det(A) = |\cdots \quad a_j \quad \cdots| = \sum_{j=1}^{n} a_j^i |\cdots \quad \vec{e}_i \quad \cdots| = \sum_{j=1}^{n} a_j^i (-1)^{i+j} \det(A_j^i).$$

Similarly, we have the cofactor expansion along the j-th column:

$$\det(A) = \sum_{i=1}^{n} a_j^i (-1)^{i+j} \det(A_j^i).$$

We have the following definition.

Definition 5.4 — Adjugate Matrix.

The *adjugate matrix*, or *classical adjoint*, of a matrix $A \in \mathrm{Mat}_n(F)$, denoted by $\mathrm{adj}(A)$, is defined to be the transpose of the cofactor matrix of A:

$$\mathrm{adj}(A) := (-1)^{i+j} \det(A_i^j).$$

Some literature defines this as the *adjoint matrix*, but to avoid confusion with the adjoint of a linear operator in inner product spaces, we use the term *adjugate matrix* here.

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 5.4.2. We have the following property of the adjugate matrix:

$$A \cdot \mathrm{adj}(A) = \mathrm{adj}(A) \cdot A = \det(A) I_n.$$

In particular, if $\det A \neq 0$, then $A^{-1} = \frac{1}{\det A} \mathrm{adj}(A)$.

Proof. In particular, we just have to show

$$\sum_{k=1}^{n} a_j^k (\text{adj}(A))_k^i = \det(A)\delta_j^i$$

From the previous Laplace expansion, we know:

$$\det(A) = \sum_{i=1}^{n} a_j^i (-1)^{i+j} \det(A_j^i) = \sum_{i=1}^{n} a_j^i (\text{adj}(A))_i^j = (A \cdot \text{adj}(A))_j^j$$

Then we know that for $i = j$, the equality holds. If $i \neq j$, then we can consider the following determinant:

$$\det \begin{vmatrix} \cdots & \underset{\uparrow}{\vec{a}_j} & \cdots & \underset{\downarrow}{\vec{a}_j} & \cdots \end{vmatrix} = 0$$

the j-th column (above $\vec{a}_j$), the i-th column (below $\vec{a}_j$)

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:

$$0 = \sum_{k=1}^{n} a_j^k (-1)^{k+j} \det(A_i^k) = \sum_{k=1}^{n} a_j^k (\text{adj}(A))_k^i = (A \cdot \text{adj}(A))_j^i$$

□

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:

$$\underbrace{\begin{bmatrix} A_1^1 & -A_1^2 & A_1^3 \\ -A_2^1 & A_2^2 & -A_2^3 \\ A_3^1 & -A_3^2 & A_3^3 \end{bmatrix}}_{\text{adj}(A)} \cdot \underbrace{\begin{bmatrix} a_1^1 & a_2^1 & a_3^1 \\ a_1^2 & a_2^2 & a_3^2 \\ a_1^3 & a_2^3 & a_3^3 \end{bmatrix}}_{A}$$

If we multiply the first row of adj(A) with the first column of A, we have the same result as the Laplace expansion along the first column:

$$a_1^1 A_1^1 - a_1^2 A_1^2 + a_1^3 A_1^3 = \begin{vmatrix} a_1^1 & a_2^1 & a_3^1 \\ a_1^2 & a_2^2 & a_3^2 \\ a_1^3 & a_2^3 & a_3^3 \end{vmatrix} = \det(A) = \sum_{k=1}^{3} a_1^k A_1^k = \sum_{k=1}^{3} a_1^k (\text{adj}(A))_k^1$$

If we multiply the first row of adj(A) with the second column of A, we have:

$$a_2^1 A_1^1 - a_2^2 A_1^2 + a_2^3 A_1^3 = \begin{vmatrix} a_2^1 & a_2^1 & a_3^1 \\ a_2^2 & a_2^2 & a_3^2 \\ a_2^3 & a_2^3 & a_3^3 \end{vmatrix} = 0 = \sum_{k=1}^{3} a_2^k A_1^k = \sum_{k=1}^{3} a_2^k (\text{adj}(A))_k^1$$

5.4.5. Vandermonde Determinant. Consider the following determinant where superscipts denote powers:

$$\det(V_n) = \begin{vmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \\ \vdots & \vdots & \ddots & \vdots \\ x_1^{n-1} & x_2^{n-1} & \cdots & x_n^{n-1} \end{vmatrix}$$

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 compare the leading coefficients of both sides. The leading coefficient of the right-hand side is Cx_n^{n-1}. The leading coefficient of the left-hand side can be found by considering the term with the highest power of x_n, which is obtained by taking x_n^{n-1} from the last column and multiplying it with the determinant of the remaining $(n-1)\times(n-1)$ matrix:

$$\begin{vmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ x_1^{n-2} & x_2^{n-2} & \cdots & x_{n-1}^{n-2} \end{vmatrix} = \det(V_{n-1})$$

By induction, we can assume that $\det(V_{n-1}) = \prod_{1\le i<j\le n-1}(x_j - x_i)$. Therefore, we have:

$$C = \det(V_{n-1}) = \prod_{1\le i<j\le n-1}(x_j - x_i).$$

Thus, we have derived the formula for the Vandermonde determinant:

$$\det(V_n) = \prod_{1\le i<j\le n}(x_j - x_i)$$

5.5. Feynman Diagram Formula

This Formula is discovered by Professor Guowu Meng in the Hong Kong University of Science and Technology. It is inspired by Feynman diagrams in quantum field theory.

Consider the case where the characteristic of the field is 0. Let A be a $n \times n$ matrix and I be the identity matrix of order n. Then we have the following formula:

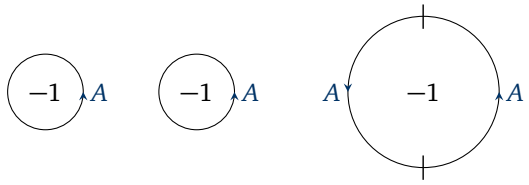
$$\det(I + tA) = 1 - \mathrm{tr}A\, t + \left(\frac{(\mathrm{tr}A)^2}{2!} - \frac{\mathrm{tr}A^2}{2}\right)t^2 - \cdots + (-1)^n \det A\, t^n$$

This is called the *Feynman diagram formula*. 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 $-\operatorname{tr} A$.

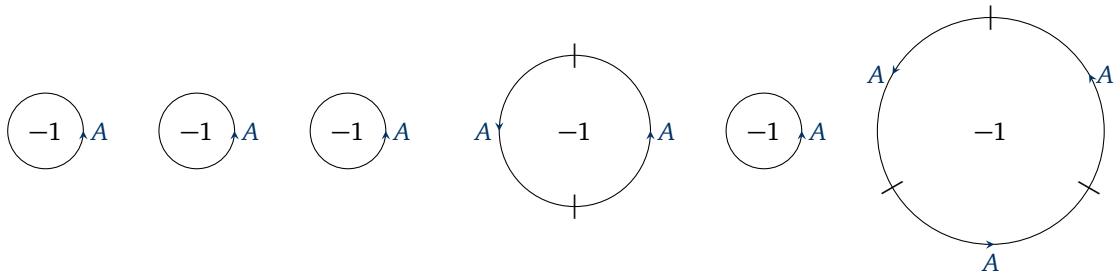
For t^2 term, we have diagram:



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:

$$\frac{(-\operatorname{tr}(A))^3}{3!} + \frac{(-\operatorname{tr}(A))(-\operatorname{tr}(A^2))}{2} - \frac{\operatorname{tr}(A^3)}{3} = -\frac{(\operatorname{tr}(A))^3}{3!} + \frac{(\operatorname{tr}(A))(\operatorname{tr}(A^2))}{2} - \frac{\operatorname{tr}(A^3)}{3}$$

Continuing this process, we can derive the coefficients for higher powers of t in the expansion of $\det(I + tA)$. Each term corresponds to a specific arrangement of traces and powers of A, with appropriate combinatorial factors accounting for symmetries in the diagrams.

5.6. Exercises

Problem 5.1. Let $f : \mathrm{Mat}_n(F) \to F$ be a function of matrix variable which is linear in each columns and skew-symmetric in columns. Show that $f(A) = \det(A) f(I)$.

Problem 5.2. Let V be a f.d. linear space and V^* be its dual space. The pairing $\langle -, - \rangle : V^* \times V \to F$ that sends (α, v) to $\alpha(v)$ can be extended to the pairing

$$\langle -, - \rangle : \Lambda^k V^* \times \Lambda^k V \to F$$

By definition, this is the unique bilinear map such that

$$\langle \alpha_1 \wedge \cdots \wedge \alpha_k, v_1 \wedge \cdots \wedge v_k \rangle = \det[\langle \alpha_i, v_j \rangle]$$

(a) For any basis $v = (v_1, \ldots, v_n)$ of V and any ordered set $I_k = (i_1, \ldots, i_k)$ of k numbers with $1 \le i_1 < i_2 < \cdots < i_k \le n$, we let $v_{I_k} = v_{i_1} \wedge \cdots \wedge v_{i_k}$. Show that $\{v_{I_k}\}$ is a minimal spanning set of $\Lambda^k V$.

(b) For each $k \ge 1$, show that the pairing is non-degenerate, i.e., the resulting map $\Lambda^k V^* \to (\Lambda^k V)^*$ is an isomorphism.

(c) Show that $\Lambda^k V^* \simeq (\Lambda^k V)^*$ in the sense that the two functors from $(\mathbf{Vec}_F^{\mathrm{f.d.}})^{\mathrm{op}}$ to $\mathbf{Vec}_F^{\mathrm{f.d.}}$ are equivalent. In particular, we have $\det V^* \simeq (\det V)^*$.

(d) Denote by $\mathcal{B}_U$ the set of bases of U. Show that the diagram

$$\begin{array}{ccccc} \mathcal{B}_V & & \longleftrightarrow & & \mathcal{B}_{V^*} \\ \downarrow & & & & \downarrow \\ \mathcal{B}_{\det V} & \longleftrightarrow & \mathcal{B}_{(\det V)^*} & \longleftrightarrow & \mathcal{B}_{\det V^*} \end{array}$$

commutes. Here a vertical map always sends basis $u = (u_1, \ldots, u_n)$ to its determinant $\det u := u_1 \wedge \cdots \wedge u_n$, the horizontal arrows map either sends a basis to its dual basis or is the isomorphism in part (b).

Problem 5.3. Let $\phi : A \to B$ be a linear map between finite-dimensional linear spaces of equal dimension. Denoted by ϕ^* the dual linear map of ϕ. Since $\det(\phi) \in \mathrm{Hom}(\det A, \det B) \simeq (\det A)^* \otimes \det B$ and

$$\det(\phi^*) \in \mathrm{Hom}(\det(B^*), \det(A^*)) \simeq (\det(B^*))^* \otimes \det(A^*) \simeq (\det(A))^* \otimes \det(B)$$

it makes sense that $\det(\phi^*) \simeq \det(\phi)$. Show that this is indeed the case.

Problem 5.4. In class we introduced the adjoint matrix $\mathrm{adj} A$ for any square matrix A of order n and showed that $A\,\mathrm{adj}(A) = \mathrm{adj}(A)A = \det(A) I$. Assuming A is invertible, which is equivalent to say that equation $A\vec{x} = \vec{b}$ has a unique solution for any vector $\vec{b} \in F^n$. Indeed, the unique solution is $\vec{x} = A^{-1}\vec{b}$.

(a) Show that the unique solution is $\vec{x} = \frac{1}{\det(A)} \mathrm{adj}(A)\vec{b}$.

(b) If x_i denotes the i-th entry of the solution $\vec{x}$, show that

$$x_i = \frac{\Delta_i}{\Delta}$$

where $\Delta = \det A$ and $\Delta_i = \det A_i$ with A_i being the matrix obtained from A by replacing its i-th column by the vector $\vec{b}$.

Problem 5.5. Let A be a $\mathrm{Mat}_n(\mathbb{R})$-valued function of one real variable t and A' be its (entry-wise) derivative with respect to t. Show that

$$\frac{d}{dt} \det(A) = \mathrm{tr}(A' \,\mathrm{adj}(A))$$

Problem 5.6. Suppose that A and B are square matrices of order 3 over a field of characteristic 0. Please derive a formula for $\det(A+B)$ of the form

$$\det(A+B) = \det(A) + \det(B) + \cdots .$$

[Hints: You may use the Feynman diagram formula introduced in this chapter.]

CHAPTER 6

Structure Theory of Linear Operators

Before, we have studied the canonical matrix representation of linear maps between two different dimension linear spaces. It is natural to ask what is the canonical form of linear maps from a linear space to itself, i.e., endomorphisms. There is another name for endomorphisms: linear operators.

Definition 6.1 — Linear Operator.
A linear map $T : V \to V$ is called a *linear operator* on a F-linear space V, or an *endomorphism* of V.

Consider the following diagram:

$$\begin{array}{ccc} V & \xrightarrow{\quad T \quad} & V \\ \uparrow{\scriptstyle [-]_{\mathcal{B}}} & & \uparrow{\scriptstyle [-]_{\mathcal{B}}} \\ F^n & \xrightarrow{\quad A \quad} & F^n \\ \updownarrow & & \updownarrow \\ F^n & \xrightarrow{\quad N \quad} & F^n \end{array}$$

As both the domain and codomain are the same linear space, both basis $\mathcal{B}$ are the same. So the matrix representation of T is much more restricted. The N is simplest looking matrix repsentation of T, but what does it look like? There are various canonical forms, depending on the field F and the linear operator T.

6.1. Diagonalisation

Generically, the simplest form of a linear operator is a *diagonal matrix*. That is, there exists a basis of V such that the matrix representation of T with respect to this basis is of the form:

$$\begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix}$$

where empty places are filled with zeros. Here λ_i are the *eigenvalues* of T. Then the set of all eigenvalues of T is the *spectrum* of T denoted by $\sigma(T)$. If such form exists, we say that T is *diagonalisable*, or *completely reducible*. If T is not diagonalisable, then we have to consider more complicated forms, which will be discussed later. Then we have the diagram:

$$\begin{array}{ccc} F^n & \xrightarrow{\quad A \quad} & F^n \\ \uparrow{\scriptstyle P^{-1}} & & \uparrow{\scriptstyle P^{-1}} \\ F^n & \xrightarrow{\quad D \quad} & F^n \end{array}$$

Here, D is the diagonal matrix and 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 raise two questions:

– How do we know whether T is diagonalisable?

– If T is diagonalisable, how can we find P and D?

If such D exists and in the form below:

$$D = \begin{bmatrix} \lambda_1 I_{n_1} & & \\ & \ddots & \\ & & \lambda_k I_{n_k} \end{bmatrix},$$

where $\lambda_i \in F$ are distinct eigenvalues and I_{n_i} are identity matrices of order n_i, $n_i > 0$ and $\sum_{i=1}^{k} n_i = n$. Then we have the internal direct sum decomposition of V:

$$V = V_{\lambda_1} \oplus V_{\lambda_2} \oplus \cdots \oplus V_{\lambda_k},$$

where $V_i = \ker(T - \lambda_i \operatorname{id}_V)$ are the *eigenspaces* of T corresponding to eigenvalues λ_i and $\dim(V_{\lambda_i}) = n_i$. Moreover, we have the external direct sum decomposition of F^n:

$$F^n = \operatorname{span}\{\vec{e}_1, \ldots, \vec{e}_{n_1}\} \oplus \operatorname{span}\{\vec{e}_{n_1+1}, \ldots, \vec{e}_{n_1+n_2}\} \oplus \cdots \oplus \operatorname{span}\{\vec{e}_{n_1+\ldots+n_{k-1}+1}, \ldots, \vec{e}_{n_1+\ldots+n_k}\},$$

and the decomposition of T:

$$T = \lambda_1 \operatorname{id}_{V_{\lambda_1}} \oplus \lambda_2 \operatorname{id}_{V_{\lambda_2}} \oplus \cdots \oplus \lambda_k \operatorname{id}_{V_{\lambda_k}}$$

Example 6.1.1. For the following diagonal matrix:

$$D = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix},$$

we have $\lambda_1 = 1$, $\lambda_2 = 2$, $n_1 = 2$ and $n_2 = 1$. Then we have the decomposition of V:

$$V = V_{\lambda_1} \oplus V_{\lambda_2}$$

where $V_{\lambda_1} = \ker(T - 1 \cdot \operatorname{id}_V)$ and $V_{\lambda_2} = \ker(T - 2 \cdot \operatorname{id}_V)$.

If T is diagonalisable, there are distinct numbers $\lambda_1, \cdots, \lambda_k \in F$, a non-trivial decomposition $V = V_{\lambda_1} \oplus \cdots \oplus V_{\lambda_k}$ such that $T\big|_{V_{\lambda_i}} = \lambda_i \operatorname{id}_{V_{\lambda_i}}$ for each $1 \le i \le k$, and $T = \lambda_1 \operatorname{id}_{V_{\lambda_1}} \oplus \cdots \oplus \lambda_k \operatorname{id}_{V_{\lambda_k}}$. 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? Consider the following linear map:

$$\lambda_i \operatorname{id}_{V_{\lambda_i}} : V_{\lambda_i} \to V_{\lambda_i}, \quad x \mapsto \lambda_i x$$

Then we have the following equation:

$$T(x) = \lambda_i(x) \iff (\lambda_i \operatorname{id}_V - T)(x) = 0 \iff x \in \ker(\lambda_i \operatorname{id}_V - T)$$

As x is non-zero, then $(\lambda_i \operatorname{id}_V - T)$ is not injective, i.e., not invertible. Therefore, we have:

$$\det(\lambda_i \operatorname{id}_V - T) = 0$$

So the eigenvalues λ_i are exactly the roots of the polynomial $\det(\lambda \operatorname{id}_V - T)$, which is called the *characteristic polynomial* of T. Note that $p_T(\lambda) = \det(\lambda \operatorname{id}_V - T)$ is a polynomial of degree $n = \dim V$. Similarly, we can define the characteristic polynomial of a matrix A as $p_A(\lambda) = \det(\lambda I_n - A)$.

Example 6.1.2. 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:

$$V_{\lambda=1}=\operatorname{null}(1\cdot I-A)=\operatorname{null}\begin{bmatrix}0 & -3\\ 0 & -1\end{bmatrix}=\operatorname{null}\begin{bmatrix}0 & 1\\ 0 & 0\end{bmatrix}=\operatorname{span}\begin{bmatrix}1\\ 0\end{bmatrix}$$

$$V_{\lambda=2}=\operatorname{null}(2\cdot I-A)=\operatorname{null}\begin{bmatrix}1 & -3\\ 0 & 0\end{bmatrix}=\operatorname{span}\begin{bmatrix}3\\ 1\end{bmatrix}$$

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 the first matrix in the example, we have: $0\cdot 1+1\cdot x_2=0\implies x_2=0$. So the null space is $\operatorname{span}\begin{bmatrix}1\\ 0\end{bmatrix}$.

For the second one, we have: $1\cdot x_1-3\cdot 1=0\implies x_1=3$. So the null space is $\operatorname{span}\begin{bmatrix}3\\ 1\end{bmatrix}$.

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 6.1.1. The following are equivalent:

(1) T is diagonalisable.

(2) $T=\lambda_1\operatorname{id}_{V_{\lambda_1}}\oplus\cdots\oplus\lambda_k\operatorname{id}_{V_{\lambda_k}}$ 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.

Example 6.1.3. Consider the matrix:

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

The $p_A(\lambda)=\lambda^2$, so the only eigenvalue is 0. Then we have:

$$V_{\lambda=0}=\operatorname{null}0\cdot I-A=\operatorname{null}\begin{bmatrix}0 & -1\\ 0 & 0\end{bmatrix}=\operatorname{span}\begin{bmatrix}1\\ 0\end{bmatrix}$$

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 diagonalisable.

Proposition 6.1.2. $E_{\lambda_1}+\cdots+E_{\lambda_k}$ 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:

$$\begin{cases} x_1 + \cdots + x_k = 0 \\ Tx_1 + \cdots + Tx_k = \lambda_1 x_1 + \cdots + \lambda_k x_k = 0 \end{cases}$$

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 $1 \leq i \leq k-1$. 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 6.1.4. Consider the matrix:

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

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:

$$E_{\lambda=1}(A) = \operatorname{null} 1 \cdot I - A = \operatorname{null} \begin{bmatrix} 0 & 0 & -4 \\ 0 & 0 & -3 \\ 0 & 0 & -1 \end{bmatrix} = \operatorname{null} \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \operatorname{span}\{\vec{e}_1, \vec{e}_2\}$$

$$E_{\lambda=2}(A) = \operatorname{null} 2 \cdot I - A = \operatorname{null} \begin{bmatrix} 1 & 0 & -4 \\ 0 & 1 & -3 \\ 0 & 0 & 0 \end{bmatrix} = \operatorname{span} \begin{bmatrix} 4 \\ 3 \\ 1 \end{bmatrix}$$

Then we have $\dim(E_{\lambda=1}) + \dim(E_{\lambda=2}) = 2 + 1 = 3 = \dim(V)$. Therefore, A is diagonalisable. Then we can find the diagonalisation:

$$\begin{bmatrix} 1 & 0 & 4 \\ 0 & 1 & 3 \\ 0 & 0 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 4 \\ 0 & 1 & 3 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} 1 & 0 & 4 \\ 0 & 1 & 3 \\ 0 & 0 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 4 & 1 & 0 \\ 3 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 4 & 1 & 0 \\ 3 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix}^{-1}$$

Diagonalisable 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 diagonalisable. Then we have another term called *semisimple*. This term and the term "completely reducible" are borrowed from representation theory of lie algebras.

Definition 6.2 — Diagonalisable.

A linear operator T is *diagonalisable*, or *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 diagonalisable if V has a non-trivial decomposition $V = V_{\lambda_1} \oplus \cdots \oplus V_{\lambda_k}$ with respect to which $T = \lambda_1 \operatorname{id}_{V_{\lambda_1}} \oplus \cdots \oplus \lambda_k \operatorname{id}_{V_{\lambda_k}}$ for some distinct eigenvalues $\lambda_1, \cdots, \lambda_k$.

Definition 6.3 — Semisimple.
A linear operator T is *semisimple* if $T \otimes_F \overline{F} : V \otimes_F \overline{F} \to V \otimes_F \overline{F}$ is diagonalisable, where $\overline{F}$ is the algebraic closure of F and $V \otimes_F \overline{F}$ is linear space over $\overline{F}$.

Remark. We can take $F = \mathbb{R}$, then $\overline{F} = \mathbb{C}$. Algebraic closure means that every polynomial in $\overline{F}[x]$ has a root in $\overline{F}$. For example, $x^2 + 1$ has no root in $\mathbb{R}$, but it has roots $\pm i$ in $\mathbb{C}$.
Note that $- \otimes F \simeq \mathrm{id}_F$, so if we change it to $- \otimes_F \overline{F}$, then we are just changing the field from F to $\overline{F}$ 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. Such decomposition is called the *Jordan-Chevalley decomposition*. However, it is too complicated that will not be discussed.

We can consider the $\mathrm{End}(V) \simeq \mathrm{Mat}_n(F) \simeq F^{n^2}$ as a linear space. Then $T \in F^{n^2}$ is a vector and such the set of containing such T forms a dense open subset of $\mathrm{End}(V) = F^{n^2}$. The dense open subset is in the Zariski topology. More precisely, the set of all diagonalisable endomorphisms with distinct eigenvalues forms a dense open subset of $\mathrm{End}(V)$. The study of Zariski topology is in the appendix 9.4. A subset of a space is *dense* and *open* in Zariski Topology if the only polynomial that vanishes on the subset is the zero polynomial.

Once we know that diagonalisable endomorphisms are dense in $\mathrm{End}(V)$, then if we want to prove some identity, it suffices to prove it for diagonalisable endormophisms. One of the example is the Cayley-Hamilton theorem.

Theorem 6.1 — Cayley-Hamilton Theorem.
Let T be a linear operator on a finite-dimensional linear space V over a field F. Let $p_T(\lambda) = \det(\lambda \,\mathrm{id}_V - T)$ be the characteristic polynomial of T. Then we have:

$$p_T(\lambda)\big|_{\lambda=T} = 0$$

Remark. $p_T(\lambda) = \det(\lambda \,\mathrm{id}_V - T) = \lambda^n + \cdots + (-1)^n \det(T)\lambda^0$, where $\lambda^0 = 1$ and $T^0 = \mathrm{id}_V$.

Proof. As $p_T(\lambda)\big|_{\lambda=T}$ is a polynomial in T, it suffices to verify the theorem on a dense set. Let $T = \lambda_1 \,\mathrm{id}_{V_{\lambda_1}} \oplus \cdots \oplus \lambda_k \,\mathrm{id}_{V_{\lambda_k}}$ be a diagonalisable operator 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 $\mathrm{id}_V = \mathrm{id}_{V_{\lambda_1}} \oplus \cdots \oplus \mathrm{id}_{V_{\lambda_k}}$. Hence, we have:

$$\lambda \,\mathrm{id}_V - T = (\lambda - \lambda_1)\,\mathrm{id}_{V_{\lambda_1}} \oplus \cdots \oplus (\lambda - \lambda_k)\,\mathrm{id}_{V_{\lambda_k}}$$

Then the characteristic polynomial is:

$$p_T(\lambda) = \det(\lambda \,\mathrm{id}_V - T) = (\lambda - \lambda_1)^{\dim(V_{\lambda_1})} \cdots (\lambda - \lambda_k)^{\dim(V_{\lambda_k})} = \prod_{i=1}^{k} (\lambda - \lambda_i)^{n_i}$$

where $n_i = \dim(V_{\lambda_i})$. Note that $\lambda_i \,\mathrm{id}_{V_{\lambda_i}} - T = 0$ on V_{λ_i} as $T\big|_{V_{\lambda_i}} = \lambda_i \,\mathrm{id}_{V_{\lambda_i}}$. As for any $v \in V$, we can write $v = v_1 + \cdots + v_k$ with $v_i \in V_{\lambda_i}$, then for all $1 \le i \le k$, we have:

$$(\lambda_i \,\mathrm{id}_{V_{\lambda_i}} - T)^{n_i}(v_i) = 0.$$

Therefore, we have:

$$p_T(\lambda)\big|_{\lambda=T}(v) = \prod_{i=1}^{k} (\lambda_i \,\mathrm{id}_V - T)^{n_i}(v) = \prod_{i=1}^{k} (\lambda_i \,\mathrm{id}_V - T)^{n_i}(v_i) = 0$$

This completed the proof. □

If T is diagonalisable, 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 $V_{\lambda_i} = \ker \lambda_i \operatorname{id}_V - T$.

Then for any T, if the set of roots of $p_T(\lambda)$ in F is $\{\lambda_1, \cdots, \lambda_k\}$, then we can define the *generalised eigenspaces*:

$$V_{\lambda_i} = \ker \lambda_i \operatorname{id}_V - T \quad \forall 1 \leq i \leq k$$

Then we check whether $\dim V = \sum_{i=1}^{k} \dim V_{\lambda_i}$. If it holds, then T is diagonalisable. If not, then T is not. So this characterise diagonalisable endomorphisms.

6.2. Ring Theory

Before studying the canonical forms of not diagonalisable operators, we need to introduce some concepts in ring theory.

Definition 6.4 — Domain.
A *domain* is a non-trivial commutative ring R with unity $\mathrm{id}_R \neq 0_R$ if non-zero elements $a, b \in R$ satisfy $ab \neq 0_R$.

Example 6.2.1. $\mathbb{Z}$ is a domain. Given any two non-zero integers $a, b \in \mathbb{Z}$, we have $ab \neq 0$.

Example 6.2.2. $\mathbb{Z}/6$ 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 6.5 — Module.
A *module* over a ring R is an abelian group $(M, +)$ together with a ring action of R on $(M, +)$.

Example 6.2.3. R itself is a module over R with the ring action being the multiplication in R.

Definition 6.6 — 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 6.7 — Ideal.
An *ideal* I of a ring R is a submodule of the module R over itself.

Example 6.2.4. The only ideals of F over itself are $\{0\}$ and F itself. So the ideal of a field is trivial.

Example 6.2.5. The ideals of $\mathbb{Z}$ over itself 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 6.8 — 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 6.2.6. $\mathbb{Z}$ 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 6.2.7. $F[x]$ is a principal ideal domain, as every ideal of $F[x]$ is of the form $(f(x)) = f(x)F[x]$ for some $f(x) \in F[x]$. It can be proved using the division algorithm of polynomials.

Definition 6.9 — Cyclic Module.
A module M over R is *cyclic* if there exists some $m \in M$ such that $M = \langle m \rangle = R \cdot m = \{r \cdot m : r \in R\}$.

Example 6.2.8. R itself is a cyclic module over R, as $R = \langle \mathrm{id}_R \rangle$.

Definition 6.10 — Finitely Generated Module.
A module M over R is *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 the definition here is different from finite-dimensional linear space, 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 6.2 — Classification Theorem of Finitely Generated Modules over a Principal Ideal Domain (Invariant Factor Decomposition).
A finitely generated module M over a principal ideal domain R 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 $1 \le i \le m-1$.

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 6.3 — Classification Theorem of Finitely Generated Modules over a Principal Ideal Domain (Primary Decomposition).
A finitely generated module M over a principal ideal domain R 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 $1 \le i \le m$.

Remark. As R is a principal ideal domain, 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 $uv = vu = \mathrm{id}_R$. For example, in $\mathbb{Z}$, the units are ± 1. In $F[x]$, 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.

6.3. Jordan Canonical Form

Let V be a finite-dimensional linear space over an algebraically closed field F, e.g. $\mathbb{C}$. Then $F[x]$ is a principal ideal domain and $x - \lambda_i$ are the prime or irreducible elements in $F[x]$ for some $\lambda_i \in F$.

Remark. If we take non-zero $\alpha \in F$, then $\alpha(x - \lambda_i)$ is also an irreducible element in $F[x]$, as α is a unit in $F[x]$ 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.

Suppose T is a linear operator on V, i.e., $T \in \mathrm{End}(V)$. Then it is equivalent to consider V as a module over $F[x]$. Moreover, V is a finitely generated module over $F[x]$ with rank 0. The ring action of $F[x]$ on V is defined as follows:

$$p(x) \cdot v = p(T)(v) = (a_n T^n + a_{n-1} T^{n-1} + \cdots + a_1 T + a_0 \,\mathrm{id}_V)(v).$$

Example 6.3.1. Let $p(x) = 2x^2 + 3x - 1 \in F[x]$ and $T \in \mathrm{End}(V)$. Then for any $v \in V$, we have $p(T)v = 2T^2(v) + 3T(v) - v$.

Therefore, by the classification theorem of finitely generated modules over a principal ideal domain, we have:

$$V \cong \bigoplus_{i=1}^{m} \frac{F[x]}{(x - \lambda_i)^{e_i}} = \frac{F[x]}{(x - \lambda_1)^{e_1}} \oplus \cdots \oplus \frac{F[x]}{(x - \lambda_m)^{e_m}}$$

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 $\dfrac{F[x]}{(x - \lambda_i)^{e_i}}$, we have the dimension being e_i. Therefore, we have the basis on $\dfrac{F[x]}{(x - \lambda_i)^{e_i}}$ 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:

$$\begin{array}{ccc} \frac{F[x]}{(x-\lambda_i)^{e_i}} & \xrightarrow{\quad T_i \quad} & \frac{F[x]}{(x-\lambda_i)^{e_i}} \\ \Big\downarrow{\scriptstyle [-]_{\mathcal{B}_i}} & & \Big\downarrow{\scriptstyle [-]_{\mathcal{B}_i}} \\ F^{e_i} & \xrightarrow{\quad J_{e_i}(\lambda_i) \quad} & F^{e_i} \end{array}$$

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 $\begin{bmatrix}\lambda_i & 1 & 0 & \cdots & 0\end{bmatrix}^T$. Similarly, we have:

$$x \cdot (x - \lambda_i) = 1 \cdot (x - \lambda_i)^2 + \lambda_i \cdot (x - \lambda_i)$$
$$x \cdot (x - \lambda_i)^{e_i - 1} = 1 \cdot (x - \lambda_i)^{e_i} + \lambda_i \cdot (x - \lambda_i)^{e_i - 1} = \lambda_i \cdot (x - \lambda_i)^{e_i - 1}$$

So the matrix representation of $x\cdot$ on $\dfrac{F[x]}{(x-\lambda_i)^{e_i}}$ 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:

$$J = \begin{bmatrix} J_{e_1}(\lambda_1) & & & \\ & J_{e_2}(\lambda_2) & & \\ & & \ddots & \\ & & & J_{e_m}(\lambda_m) \end{bmatrix}$$

This is called the *Jordan canonical form* of T. However, this is for algebraically closed fields. For non-algebraically closed fields, it is more complicated in general that will not be discussed here. Interested readers may find more information on Jordan-Chevalley decomposition and rational canonical form, or try the field, $\mathrm{Gal}(\mathbb{C}/\mathbb{R})$ which is of order 2. The irreducible polynomials in $\mathbb{R}[x]$ are all of degree 1 or 2, i.e., $x - \lambda$ for some $\lambda \in \mathbb{R}$ or $x^2 + ax + b$ with $a, b \in \mathbb{R}$ and $a^2 - 4b < 0$.

6.4. Exercises

Problem 6.1. We say that A is diagonalisable if there is an invertible matrix P, written as $[v_1, \ldots, v_n]$, and a diagonal matrix D, with diagonal entries d_{ii} denoted by λ_i, such that $AP = PD$, i.e., $Av_i = \lambda_i v_i$ for each $i = 1, \ldots, n$.

(a) Show that A is diagonalisable $\iff$ F^n has a basis consisting of eigenvectors of A.

(b) Show that eigenspaces are linearly independent, i.e., they are all non-trivial and there is only one way to write 0 as a finite sum of vectors, one from each of them.

(c) Let $\sigma(A) = \{\lambda_1, \ldots, \lambda_k\}$. Show that A is diagonalisable $\iff$ $F^n = \bigoplus_i E_{\lambda_i}(A)$ $\iff$ $n = \sum_i \dim E_{\lambda_i}(A)$.

(d) Show that A is diagonalisable if $|\sigma(A)| = n$, i.e., A has n distinct eigenvalues.

(e) Find a non-diagonalisable square matrix A of order 2.

Remark. The goal of this exercise is to give a road map for a sketchy proof of Cayley-Hamilton Theorem.

Problem 6.2. Let $f \in F[x]$ be a monic polynomial of degree $n \geq 1$. Over the algebraic closure $\overline{F}$ of F, we can factorise f as $(x - x_1) \cdots (x - x_n)$ with $x_i \in \overline{F}$. The discriminant of f, denoted by $\mathrm{Disc}(f)$, is defined to be $\prod_{i<j} (x_i - x_j)^2$. Being symmetric in $x_1, \ldots, x_n$, $\mathrm{Disc}(f)$ must be a polynomial in the coefficients of f.

(a) Show that the discriminant of the quadratic polynomial $x^2 + bx + c$ is $b^2 - 4c$. How about the discriminant of the cubic polynomial $x^3 + px + q$?

(b) Show that f has n-distinct roots in $\overline{F}$ $\iff$ $\mathrm{Disc}(f) \neq 0$.

(c) Show that $P_A(\lambda)\big|_{\lambda=A} = 0$ $\iff$ $P_A(\lambda)\big|_{\lambda=A} = 0$ when A is viewed as a square matrix over $\overline{F}$.

Thus, to prove $P_A(\lambda)\big|_{\lambda=A} = 0$, without loss of generality, we shall assume in the following that the field F is algebraically closed.

(d) Show that $P_A(\lambda)\big|_{\lambda=A} = 0$ if A is a diagonal matrix.

(e) Show that $P_A(\lambda)\big|_{\lambda=A} = 0$ if A is a diagonalisable matrix.

(f) Show that $P_A(\lambda)\big|_{\lambda=A} = 0$ if $\mathrm{Disc}(P_A) \neq 0$.

(g) Show that the map $A \mapsto \mathrm{Disc}(P_A)$ is a polynomial map f from the affine space $\mathrm{End}\, F^n$ to F. Note that the $\mathrm{End}\, F^n$ is isomorphic to the affine space $\mathbb{A}_F^{n^2}$ of dimension n^2.

We need the following two facts from Topology:

1. In Zariski Topology, any finite-dimensional affine space over an algebraically closed field is an irreducible topological space.
2. Any non-empty open set U in an irreducible topological space X must be dense, i.e., X is equal to the topological closure $\overline{U}$ of U.

(h) Assume these facts and let f be the polynomial f in part (g) and f_{ij} be the polynomial map from affine space $\mathrm{End}\, F^n$ to F that sends A to the (i, j)-entry of the matrix $P_A(\lambda)\big|_{\lambda=A}$. Show that $[f_{ij}(A)] = P_A(\lambda)\big|_{\lambda=A} = 0$ for all A on the non-empty Zariski open set $\{f \neq 0\}$ of $\mathrm{End}\, F^n$.

(i) Show that all $P_A(\lambda)\big|_{\lambda=A} = 0$ for all A in the affine space $\mathrm{End}\, F^n$.

Remark. The goal of this exercise is to outline an elementary proof of Jordan Canonical Form.

Problem 6.3. Assume the field F is algebraically closed and $\sigma(A) = \{\lambda_1, \ldots, \lambda_k\}$. Then $P_A(\lambda) = \prod_{i=1}^{k} (\lambda - \lambda_i)^{n_i}$ where each integer n_i is positive. Cayley-Hamilton Theorem says that

$$\prod_{i=1}^{k} (A - \lambda_i I)^{n_i} = 0. \tag{3}$$

(a) Show that, for each i, we have a sequence

$$\mathsf{null}(A - \lambda_i I) \subseteq \mathsf{null}\,(A - \lambda_i I)^2 \subseteq \cdots$$

that will eventually stabilise.

The generalised eigenspace of A with eigenvalue λ_i, denoted by $\widetilde{E}_{\lambda_i}(A)$, is defined to be the increasing union $\bigcup_{k\geq 1} \mathsf{null}\,(A - \lambda_i I)^k$. By definition, any non-zero vector v in $\widetilde{E}_{\lambda_i}(A)$ is called a generalised eigenvector of A with eigenvalue λ_i. Let v be a generalised eigenvector of A with eigenvalue λ.

(b) Show that there is an integer $m \geq 0$ such that $v_m := (A - \lambda I)^m v$ is an eigenvector of A with eigenvalue λ.

(c) Show that v is never a generalised eigenvector of A with eigenvalue $\mu \neq \lambda$.

(d) Show that, for any $k \geq 0$ and scalar $\mu \neq \lambda$, $(A - \mu I)^k v$ is always a generalised eigenvector of A with eigenvalue λ. Consequently, $(A - \mu I)^k$ maps $\widetilde{E}_{\lambda}(A)$ isomorphically onto $\widetilde{E}_{\lambda}(A)$.

(e) Show that the algebraic multiplicity of the eigenvalue λ_i is bigger than or equal to the geometric multiplicity, i.e., $\dim E_{\lambda_i}(A)$, of the eigenvalue λ_i.

(f) Show that the generalised eigenspaces of A are linearly independent and their direct sum is the entire linear space F^n.

(g) Show that, with respect to the decomposition $F^n = \widetilde{E}_{\lambda_1}(A) \oplus \cdots \oplus \widetilde{E}_{\lambda_k}(A)$, we have the decomposition $A = (\lambda_1 I_{n_1} + N_1) \oplus \cdots \oplus (\lambda_k I_{n_k} + N_k)$ where each N_i is a nilpotent matrix of order n_i.

CHAPTER 7

Inner Product Spaces

As of now, we have studied linear spaces and linear maps between linear spaces. In this chapter, we will introduce some geometric structures on linear spaces, which will lead to the definition of inner product spaces. Inner product spaces are very important in both pure and applied mathematics, as they provide a way to measure angles and lengths in linear spaces. We will explore the properties of inner product spaces, including orthogonality, projections, and orthonormal bases.

7.1. Inner Products and Euclidean Spaces

Inner products allow us to define notions of length and angle in more abstract linear spaces. We will start with the field of real numbers, $\mathbb{R}$, and later extend the definition to complex numbers, $\mathbb{C}$.

Definition 7.1 — Inner Product.

An *inner product* on a real linear space V is a map $\langle -, - \rangle : V \times V \to \mathbb{R}$ such that the following properties hold:

Bilinearity : $\langle -, v \rangle$ and $\langle u, - \rangle$ are linear maps for all $u, v \in V$;

Symmetry : $\langle u, v \rangle = \langle v, u \rangle$ for all $u, v \in V$;

Positive-definiteness : $\langle v, v \rangle \geq 0$ for all $v \in V$ with equality if and only if $v = 0$.

In short, an inner product is a positive-definite symmetric bilinear form on V.

Remark. The notation $\langle -, - \rangle$ right here and afterwards is used to denote inner products, which is different from the natural pairing notation used in Definition 3.15. These two look similar but are different concepts, the inner product is a tensor of type $(0, 2)$ while the natural pairing is a tensor of type $(1, 1)$.

We can further more relax a bit the positive-definiteness condition to get the following definition.

Definition 7.2 — Pseudo Inner Product.

A *pseudo inner product* on a real linear space V is a *non-degenerate* symmetric bilinear form on V, i.e., an element $\langle -, - \rangle \in \mathcal{S}^2 V^*$ such that $\langle -, - \rangle_\natural : V \to V^*$ is an isomorphism.

Definition 7.3 — Euclidean Structure.

A *Euclidean structure* on a real linear space V is an inner product $\langle -, - \rangle : V \times V \to \mathbb{R}$. The pair $(V, \langle -, - \rangle)$ is called a *Euclidean space*.

We have mentioned that inner products can define lengths and angles in linear spaces. There is another way to define lengths in linear spaces, which turns out to be equivalent to inner products.

Definition 7.4 — Metric Structure.
A *metric structure* on a non-empty set X is a map $d : X \times X \to \mathbb{R}$ such that the following properties hold:

Positive-definiteness : $d(x,y) \geq 0$ for all $x, y \in X$ with equality if and only if $x = y$;

Symmetry : $d(x,y) = d(y,x)$ for all $x, y \in X$;

Triangle inequality : $d(x,z) \leq d(x,y) + d(y,z)$ for all $x, y, z \in X$.

The pair (X,d) is called a *metric space*.

To make a metric structure on a linear space V, it is required that the metric structure is compatible with the linear structure on V in the following sense:

Translation invariance : $d(u+v, u+w) = d(v,w)$ for all $u, v, w \in V$;

Homogeneity : $d(\alpha v, \alpha w) = |\alpha| d(v,w)$ for all $\alpha \in \mathbb{R}$ and $v, w \in V$.

There is actually the third way to define lengths in linear spaces, which is called norms.

Definition 7.5 — Normed Structure.
A *normed structure* on a real linear space V is a map $\|-\| : V \to \mathbb{R}$ such that the following properties hold:

Positive-definiteness : $\|v\| \geq 0$ for all $v \in V$ with equality if and only if $v = 0$;

Homogeneity : $\|\alpha v\| = |\alpha| \|v\|$ for all $\alpha \in \mathbb{R}$ and $v \in V$;

Triangle inequality : $\|u+v\| \leq \|u\| + \|v\|$ for all $u, v \in V$.

The pair $(V, \|-\|)$ is called a *normed linear space*.

Proposition 7.1.1 — Norm Induced by Metric. Let (V,d) be a metric linear space. Then the map $\|-\| : V \to \mathbb{R}$ defined as $\|v\| = d(v,0)$ for all $v \in V$ is a norm on V.

It is straightforward to check the three properties in Definition 7.5.

Proposition 7.1.2 — Metric Induced by Norm. Let $(V, \|-\|)$ be a normed linear space. Then the map $d : V \times V \to \mathbb{R}$ defined as $d(u,v) = \|u - v\|$ for all $u, v \in V$ is a metric on V.

It is straightforward to check the three properties in Definition 7.4

To relate inner products with norms, we need the following one important inequality and one important identity.

Proposition 7.1.3 — Cauchy-Schwarz Inequality. Let $(V, \langle -,- \rangle)$ be a Euclidean space. Then for all $u, v \in V$, we have:

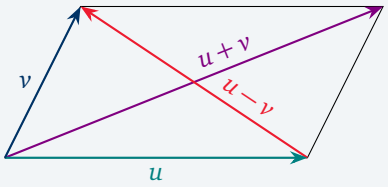
$$|\langle u, v \rangle| \leq \|u\| \|v\|$$

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:

$$\Delta = 4\langle u, v \rangle^2 - 4\|u\|^2 \|v\|^2 \leq 0 \implies \langle u, v \rangle^2 \leq \|u\|^2 \|v\|^2$$ □

Theorem 7.1 — 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 the following identity for all $u, v \in V$:

$$\|u+v\|^2 + \|u-v\|^2 = 2\|u\|^2 + 2\|v\|^2$$

Proposition 7.1.4. An Euclidean structure on a real linear space V is equivalent to a norm structure on V satisfying the parallelogram law.

Proof. Let $(V, \langle -, - \rangle)$ be a Euclidean space.

($\Rightarrow$) : We can define the norm on V as $\|v\| = \sqrt{\langle v, v \rangle}$ for all $v \in V$. Then we have:

Positive-definiteness : $\|v\| = \sqrt{\langle v, v \rangle} \geq 0$ for all $v \in V$ with equality if and only if $v = 0$;

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 $\alpha \in \mathbb{R}$;

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.

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:

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:

$$\|u+w+v\|^2 - \|u+w\|^2 - \|v\|^2 = \|u+v\|^2 - \|u\|^2 - \|v\|^2 + \|w+v\|^2 - \|w\|^2 - \|v\|^2$$

$$\|u+w+v\|^2 + \|u\|^2 + \|w\|^2 + \|v\|^2 = \|u+w\|^2 + \|u+v\|^2 + \|w+v\|^2$$

Then we may consider $x = u + w$ and $y = v + w$, and $x' = u + v + w$ and $y' = w$, and we have

$$\|u + v + 2w\|^2 + \|u - v\|^2 = 2\|u + w\|^2 + 2\|v + w\|^2$$

$$\|u + v + 2w\|^2 + \|u + v\|^2 = 2\|u + v + w\|^2 + 2\|w\|^2$$

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

$$2\|u\|^2 + 2\|v\|^2 - 2\|u + v\|^2 = 2\|u + w\|^2 + 2\|v + w\|^2 - 2\|u + v + w\|^2 - 2\|w\|^2$$

$$\|u + v + w\|^2 + \|u\|^2 + \|v\|^2 + \|w\|^2 = \|u + w\|^2 + \|v + w\|^2 + \|u + v\|^2$$

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 $r \to \lambda$.

Symmetry : For all $u, v \in V$, we have:

$$\begin{aligned}\langle u, v\rangle &= \frac{1}{2}\left(\|u + v\|^2 - \|u\|^2 - \|v\|^2\right) \\ &= \frac{1}{2}\left(\|v + u\|^2 - \|v\|^2 - \|u\|^2\right) = \langle v, u\rangle\end{aligned}$$

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. □

Definition 7.6 — Length.

If $v \in V$ is a vector in a Euclidean space $(V, \langle -, -\rangle)$, then the *length*, or *norm* of v is defined as:

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

Definition 7.7 — Angle.

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.

7.2. Orthogonality

7.2.1. Orthogonal Subspaces and Orthogonal Complements. We have already defined the notion of orthogonality between two vectors in Definition 7.7. We can further more extend the notion of orthogonality to subspaces in a Euclidean space.

Let V be a Euclidean space with inner product $\langle -,- \rangle$ and $W \subseteq V$ is a subspace of V. Then W inherits an Euclidean structure from $\langle -,- \rangle$ in V. We restrict the inner product $\langle -,- \rangle$ on V to W:

$$W \times W \hookrightarrow V \times V \xrightarrow{\langle -,- \rangle} \mathbb{R}$$

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:

$$\begin{array}{ccccccccc} 0 & \longrightarrow & \ker(\lambda_w) & \longrightarrow & V & \overset{\lambda_w}{\twoheadrightarrow} & \mathbb{R}^k & \longrightarrow & 0 \\ & & & & \uparrow & & \uparrow{\scriptstyle [-]_{w^*}} & & \\ & & & & W & \xleftarrow{\langle -,- \rangle^\natural} & W^* & & \\ & & & & w_i & \longmapsto & \langle w_i, - \rangle & & \end{array}$$

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:

$$V = \operatorname{im}(s) \oplus \ker(\lambda_w) = W \oplus \ker(\lambda_w)$$

Note that it is an internal direct sum. Then we define the orthogonal complement of W in V as follows.

Definition 7.8 — 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 $w = \operatorname{proj}_W(v) \in W$ and $w^\perp = \operatorname{proj}_{W^\perp}(v) \in W^\perp$. The map $\operatorname{proj}_W : V \to W$ is called the orthogonal projection onto W along $W^\perp$. For visualization, see Figure 7.

Then we have the following properties of the orthogonal projection:

- $(\operatorname{proj}_W)^2 = \operatorname{proj}_W$;
- $\operatorname{im} \operatorname{proj}_W = W$;
- $\ker \operatorname{proj}_W = W^\perp$;
- $\operatorname{proj}_W + \operatorname{proj}_{W^\perp} = \mathrm{id}_V$.

Definition 7.9 — Orthogonal Basis and Orthonormal Basis.

An *orthogonal basis* of a Euclidean space $(V, \langle -,- \rangle)$ is a basis $w = (w_1, w_2, \cdots, w_n)$ of V such that $\langle w_i, w_j \rangle = 0$ for all $i \neq j$. An *orthonormal basis* is an orthogonal basis $w = (w_1, w_2, \cdots, w_n)$ such that $\langle w_i, w_i \rangle = 1$ for all $1 \leq i \leq n$.

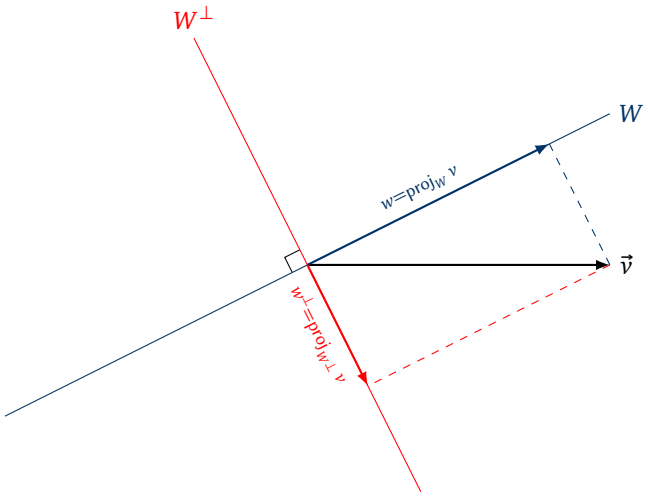


Figure 7. Orthogonal Projection onto Subspace W along $W^\perp$

Proposition 7.2.1. 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. Up to isomorphism, there is a unique Euclidean space with dimension n, i.e., $(\mathbb{R}^n, \cdot)$.

If $w = (w_1, w_2, \cdots, w_k)$ is an orthonormal basis of W, then for all $u \in V$, we have the following formula for the orthogonal projection of u onto W:

$$\text{proj}_W u = \sum_{i=1}^{k} \langle w_i, u \rangle w_i$$

In case w is orthogonal but not orthonormal, then we have the following formula:

$$\text{proj}_W u = \sum_{i=1}^{k} \frac{\langle w_i, u \rangle}{\langle w_i, w_i \rangle} w_i$$

7.2.2. Gram-Schmidt Process. We will now show how to construct an orthonormal basis of a Euclidean space V given any basis of V. The key idea is to use the orthogonal projection to iteratively construct orthonormal vectors.

Let $w = (w_1, w_2, \cdots, w_k)$ be an orthonormal basis of $W \subseteq V$. Then we have:

$$x = \underbrace{\sum_{i=1}^{k} \langle w_i, x \rangle w_i}_{\in W} + \underbrace{x - \sum_{i=1}^{k} \langle w_i, x \rangle w_i}_{\in W^\perp} = \text{proj}_W x + \text{proj}_{W^\perp} x.$$

To show that $\text{proj}_{W^\perp} x \in W^\perp$, it suffices to show that $\langle w_j, \text{proj}_{W^\perp} x \rangle = 0$ for all $1 \leq j \leq k$:

$$\begin{aligned} \langle w_j, \text{proj}_{W^\perp} x \rangle &= \langle w_j, x - \sum_{i=1}^{k} \langle w_i, x \rangle w_i \rangle \\ &= \langle w_j, x \rangle - \sum_{i=1}^{k} \langle w_i, x \rangle \langle w_j, w_i \rangle \\ &= \langle w_j, x \rangle - \langle w_j, x \rangle = 0 \end{aligned}$$

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 $V_1 = \text{span}\{u_1\}$. Then we want to find the w_2 such that w_1, w_2 is the orthonormal basis of $V_2 = \text{span}\{u_1, u_2\}$. For visualization, see the Figure 8.

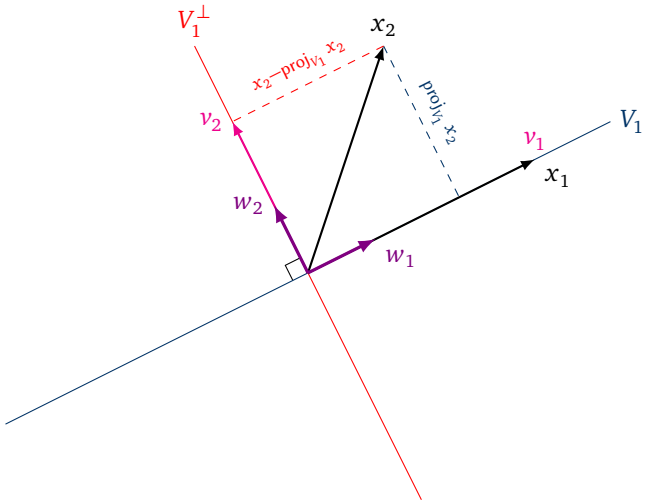


Figure 8. Gram-Schmidt Process for Constructing Orthonormal Basis

Then $v_2 = x_2 - \text{proj}_{V_1} x_2 = x_2 - \langle w_1, x_2 \rangle w_1$ 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 $V_{k-1} = \text{span}\{x_1, x_2, \cdots, x_{k-1}\}$ 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:

$$\underbrace{\begin{bmatrix} | & | & & | \\ \vec{x}_1 & \vec{x}_2 & \cdots & \vec{x}_n \\ | & | & & | \end{bmatrix}}_{A} = \underbrace{\begin{bmatrix} | & | & & | \\ \vec{w}_1 & \vec{w}_2 & \cdots & \vec{w}_n \\ | & | & & | \end{bmatrix}}_{Q} \underbrace{\begin{bmatrix} \langle \vec{w}_1, \vec{x}_1 \rangle & \langle \vec{w}_1, \vec{x}_2 \rangle & \cdots & \langle \vec{w}_1, \vec{x}_n \rangle \\ 0 & \langle \vec{w}_2, \vec{x}_2 \rangle & \cdots & \langle \vec{w}_2, \vec{x}_n \rangle \\ & & \ddots & \vdots \\ 0 & 0 & 0 & \langle \vec{w}_n, \vec{x}_n \rangle \end{bmatrix}}_{R}$$

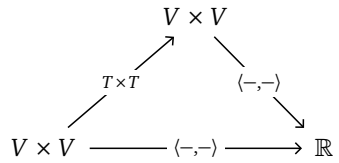
which is called the *QR Decomposition* of A, where Q is an orthogonal matrix and R is an upper-triangular matrix with positive diagonal entries. However, normally we denote the orthogonal matrix by O instead of Q and an upper-triangular matrix by U instead of R.

7.2.3. Orthogonal Group and Special Orthogonal Group. Similar to the automorphism group of a linear space, we can define the automorphism group of a Euclidean space which is called the orthogonal group.

Let V be a Euclidean space with inner product $\langle -, - \rangle$. Then we view V as a linear space, and we have $\mathrm{Aut}(V) = \mathrm{GL}(V)$. If we view V as a Euclidean space, then we have $\mathrm{Aut}(V) = \mathrm{O}(V) \subseteq \mathrm{GL}(V)$, where $\mathrm{O}(V)$ is the subgroup of $\mathrm{GL}(V)$ that respects the Euclidean structure, i.e., for all $T \in \mathrm{O}(V)$, 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 $\mathrm{O}(n)$ using this property. Let $V = \mathbb{R}^n$ with the dot product. Then for any $A \in \mathrm{GL}_n(\mathbb{R})$, $A \in \mathrm{O}(n)$ 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 $O^T O = I_n$, i.e., $O^T = O^{-1}$. Therefore, we have:

$$\mathrm{O}(n) = \{O \in \mathrm{GL}_n(\mathbb{R}) \mid O^T O = I_n\}$$

Note that $\det(O^T) = \det(O)^T = \det(O)$. Therefore, we have $\det(O)^2 = 1$ for all $O \in \mathrm{O}(n)$, i.e., $\det(O) = \pm 1$.

Then consider the following exact sequence:

$$1 \longrightarrow \mathrm{SL}(V) \hookrightarrow \mathrm{GL}(V) \xrightarrow{\det} \mathbb{R}^\times \longrightarrow 1$$

where $\mathbb{R}^\times = \mathrm{GL}_1(\mathbb{R}) = \mathbb{R} \setminus \{0\}$ is the multiplicative group of non-zero real numbers. As for any automorphism $A \in \mathrm{GL}(V)$, we have a determinant $\det A \in \mathbb{R}^\times$, which is surjective. $\mathrm{SL}(V)$ is defined as the kernel

of the determinant map, i.e., $\mathrm{SL}(V) = \{A \in \mathrm{GL}(V) \mid \det A = 1\}$. Similarly, we have the special orthogonal group $\mathrm{SO}(V)$ as the subgroup of $\mathrm{O}(V)$ with determinant 1:

$$\mathrm{SO}(V) = \{A \in \mathrm{O}(V) \mid \det A = 1\}$$

7.2.4. 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 will discuss in Section 7.5. But the inner product matrix is a real symmetric matrix.

Definition 7.10 — Real Symmetric Matrix.
A real matrix A of order n is called a *real symmetric matrix* if $A^T = A$.

There is a key property of real symmetric matrices that is the spectral theorem, which will be discussed in Section 7.5.

7.3. Hermitian Inner Products and Unitary Groups

7.3.1. Hermitian Inner Products. We can generalise the notion of inner products to complex linear spaces. The generalisation is called Hermitian inner products.

Definition 7.11 — Hermitian Inner Products.
A *Hermitian inner product*, or *Hermitian product*, is a *Hermitian form* that is also *positive-definite*, i.e., for any $v \in V$, $\langle v, v \rangle \geq 0$ with equality if and only if $v = 0$. A Hermitian form on a complex linear space V is a map $\langle -, - \rangle : V \times V \to \mathbb{C}$ such that the following properties hold:

Sesquilinearity : For any $u, v \in V$ and $\alpha \in \mathbb{C}$, we have:

Biadditivity: $\langle u + v, w \rangle = \langle u, w \rangle + \langle v, w \rangle$ and $\langle u, v + w \rangle = \langle u, v \rangle + \langle u, w \rangle$;

Linear in the first arguement: $\langle \alpha u, v \rangle = \alpha \langle u, v \rangle$;

Conjugate-linear in the second arguement: $\langle u, \alpha v \rangle = \overline{\alpha} \langle u, v \rangle$.

Conjugate Symmetry : For any $u, v \in V$, we have:

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

The asterisk symbol * is defined as $\langle u, v \rangle^* = \overline{\langle v, u \rangle}$.

A *pseudo-Hermitian inner product* is a Hermitian form that is *non-degenerate*, i.e., for all $u \in V$, if $\langle u, v \rangle = 0$ for all $v \in V$, then $u = 0$.

Remark. In physics, the Hermitian product is written with Dirac's bra-ket notation:

$$\langle x | y \rangle := \langle y, x \rangle$$

With this notation, the inner product is linear in the second argument and conjugate-linear in the first argument. Moreover, for the conjugate of the inner product, the dagger symbol * is used:

$$\langle x|y\rangle^* := \overline{\langle y|x\rangle} = \langle x|y\rangle$$

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:

$$|\langle u, v\rangle| = |\Re(\langle e^{-i\theta}u, v\rangle)| \leq \|e^{-i\theta}u\|\|v\| = \|u\|\|v\| \qquad \square$$

The sesquilinear map $\langle -, -\rangle$ can be defined as a bilinear map $V \times \overline{V} \to \mathbb{C}$, where $\overline{V}$ is the complex-conjugate linear space of V, or linear map $V \otimes \overline{V} \to \mathbb{C}$. The complex-conjugate linear space $\overline{V}$ is defined as the same set as V with the same addition operation, but the scalar multiplication is defined as:

$$\mathbb{C} \times \overline{V} \to \overline{V}, \quad (\alpha, v) \mapsto \overline{\alpha}v$$

Example 7.3.1. For any $\vec{z}$ and $\vec{w}$ in $\mathbb{C}^n$, we can define the *standard Hermitian inner product* as:

$$\langle \vec{z}, \vec{w}\rangle = \vec{w}^*\vec{z} = \overline{\vec{w}}^T\vec{z}.$$

It is straightforward to verify that it satisfies all the properties of Hermitian products. For example, the positive-definiteness property holds since:

$$\vec{z}^*\vec{z} = \sum_{i=1}^{n} \overline{z^i} z^i = \sum_{i=1}^{n} |z^i|^2 \geq 0$$

Then a complex linear space V with an Hermitian product $\langle -, -\rangle$ is called a *Hermitian space*. Also, the model / standard Hermitian space is $(\mathbb{C}^n, \langle -, -\rangle)$ with the standard Hermitian product, that is, the inner product defined above.

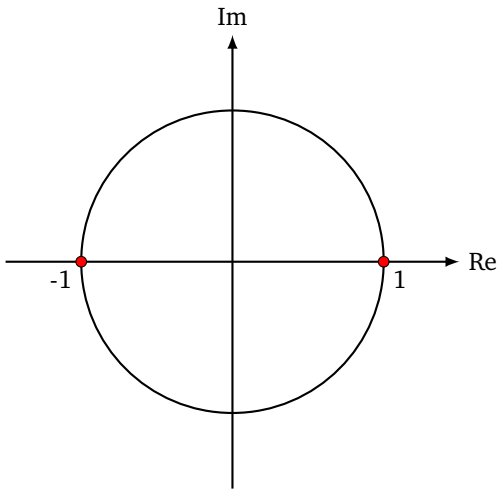
Let V be a Hermitian space with Hermitian product $\langle -, -\rangle$. Then we say $u, v \in V$ are orthogonal if $\langle u, v\rangle = 0$. Similar to the Euclidean case, we can define orthogonal complement, orthogonal projection, orthonormal basis, and Gram-Schmidt process in Hermitian spaces. We 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 product, i.e., for any Hermitian space V with dimension n, there exists a linear isomorphism between V and $(\mathbb{C}^n, \langle -, -\rangle)$.

7.3.2. Unitary Group. 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

$$\mathrm{U}(n) = \{U \in \mathrm{GL}_n(\mathbb{C}) \mid U^*U = I\}$$

where $U^* = \overline{U}^T$ is the conjugate transpose of U. Note that $\det(U^*) = \overline{\det(U)}$. Therefore, we have $|\det(U)|^2 = 1$ for all $U \in \mathrm{U}(V)$, i.e., $|\det(U)| = 1$. This means $\mathrm{U}(1) = \{z \in \mathbb{C} \mid |z| = 1\}$ is the unit circle in the complex plane. Graphically we have:



where the unit circle represents $\mathrm{U}(1)$ 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 $\mathrm{SU}(n)$ as the subgroup of $\mathrm{U}(n)$ with determinant 1.

Then we have the following definition similar to orthogonal matrices:

Definition 7.12 — Unitary Matrix.
An invertible complex matrix U of order n is a *unitary matrix* if $U^*U = I_n$, i.e., $U^{-1} = U^*$.

Similarly, we have the following definition similar to symmetric matrices:

Definition 7.13 — Hermitian Matrix.
A complex matrix A of order n is called a *Hermitian matrix* if $A^* = A$.

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.

7.3.3. Matrix representation of Hermitian products. Then we have the matrix representation of Hermitian products as follows.

Let V be a Hermitian space with Hermitian product $\langle -, - \rangle$. 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., $\omega^* = \omega$. Then we claim that if A and $\widetilde{A}$ are two matrix representations of the Hermitian product $\langle -, - \rangle$ with respect to two different bases v and $\widetilde{v}$ respectively, then there exists an invertible matrix $P \in \mathrm{GL}_n(\mathbb{C})$ such that:

$$\widetilde{A} = P^* A P$$

where P is the change-of-basis matrix from v to $\widetilde{v}$. Or equivalently,

$$\mathsf{H}_n(\mathbb{C}) \times \mathsf{GL}_n(\mathbb{C}) \to \mathsf{H}_n(\mathbb{C}), \quad (A, P) \mapsto P^*AP$$

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.

7.4. Self-Adjoint Operators and Unitary Operators

Let V be a Hermitian space with Hermitian product $\langle -, - \rangle$. We have the following equality for any linear operator $T : V \to V$:

$$\langle Tu, v \rangle = \langle u, T^*v \rangle.$$

The linear operator $T^* : V \to V$ is called the *adjoint operator* of T. It can be shown that the adjoint operator T^* exists and is unique. The following definitions are the abstract generalisations of Hermitian matrices and unitary matrices.

Definition 7.14 — Hermitian Operator.
A linear operator $T : V \to V$ is a *Hermitian operator* or *self-adjoint operator* if for any $u, v \in V$, we have:

$$\langle Tu, v \rangle = \langle u, Tv \rangle.$$

Definition 7.15 — Unitary Operator.
A linear operator $U : V \to V$ is a *unitary operator* if for any $u, v \in V$, we have:

$$\langle Uu, Uv \rangle = \langle u, v \rangle.$$

There is a generalisation of the above two operators.

Definition 7.16 — Normal Operator.
A linear operator $N : V \to V$ is a *normal operator* if it commutes with its adjoint operator, i.e., $NN^* = N^*N$.

Proposition 7.4.1. For any linear map $T : V \to W$ between two Hermitian spaces V and W, there also exists a unique adjoint map $T^* : W \to V$ satisfying:

$$\langle Tu, w \rangle_W = \langle u, T^*w \rangle_V$$

Proof. We can reduce the problem to $\mathbb{C}^n$ and $\mathbb{C}^m$ with standard Hermitian products by choosing orthonormal bases of V and W. Then we have T represented by a matrix $A \in \mathsf{Mat}_{m\times n}(\mathbb{C})$. Then we propose there is a matrix $B \in \mathsf{Mat}_{n\times m}(\mathbb{C})$ such that for all $\vec{e}_i \in \mathbb{C}^n$ and $\vec{f}_j \in \mathbb{C}^m$, we have:

$$\langle A\vec{e}_i, \vec{f}_j \rangle = (A\vec{e}_i)^* \vec{f}_j = \vec{e}_i^* A^* \vec{f}_j = \vec{e}_i^T A^* \vec{f}_j$$

which is the (i, j)-th entry of A^*. On the other hand, we have:

$$\langle \vec{e}_i, B\vec{f}_j \rangle = \vec{e}_i^* (B\vec{f}_j) = \vec{e}_i^T B \vec{f}_j$$

which is the (i, j)-th entry of B. Therefore, we have $B = A^*$. This proves the existence of the adjoint operator. The uniqueness is straightforward. □

Proposition 7.4.2. 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 diagonalisable.

Proof. Given that $T^* = T$, 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:
$$\langle Tv, v\rangle = \langle v, T^*v\rangle = \langle v, Tv\rangle$$
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:
$$\langle Tv_1, v_2\rangle = \langle v_1, T^*v_2\rangle$$
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:
$$\langle Tw^\perp, w\rangle = \langle w^\perp, T^*w\rangle = \langle w^\perp, Tw\rangle = 0$$
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 7.4.3. 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 diagonalisable.

Proof. Given that $T^*T = TT^* = 1_V$, 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:
$$\langle Tv, v\rangle = \langle v, T^*v\rangle$$
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:

$$\langle Tv_1, v_2\rangle = \langle v_1, T^* v_2\rangle$$

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 $V = W$. □

7.5. 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 7.2 — Spectral Theorem for Hermitian Matrices.

A Hermitian operator A on a finite-dimensional Hermitian space V can be diagonalised by a unitary matrix, i.e., for some orthonormal basis of V, there exists a unitary matrix U and a diagonal matrix D such that:

$$A = UDU^{-1} = UDU^*.$$

The diagonal entries of D are the eigenvalues of A, which are all real numbers.

Theorem 7.3 — Spectral Theorem for Unitary Matrices.

A unitary operator A on a finite-dimensional Hermitian space V can be diagonalised by a unitary matrix, i.e., for some orthonormal basis of V, there exists a unitary matrix U and a diagonal matrix D such that:

$$A = UDU^{-1} = UDU^*.$$

The diagonal entries of D are the eigenvalues of A, which are all complex numbers with absolute value 1. Moreover, the columns of U form an orthonormal basis of V consisting of eigenvectors of A.

Then we have the following corollary.

Corollary 7.1 — Spectral Theorem for Real Symmetric Matrices.

A real symmetric matrix A can be diagonalised by an orthogonal matrix, i.e., there exists an orthogonal matrix O and a real diagonal matrix D such that:

$$A = ODO^T.$$

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., $A = U^*DU$ 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. □

Proposition 7.5.1 — Canonical form of orthogonal matrices. 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 $\det(O) = -1$.

The proof is left as an exercise.

Proposition 7.5.2 — Canonical form of Hermitian forms. The matrix representation H of the Hermitian form on a complex linear space V with respect to a basis $\mathcal{B}$ is a Hermitian matrix. So, there exists a unitary matrix U and a real diagonal matrix D such that:

$$H = UDU^*$$

Then the Hermitian form can be represented as:

$$\langle x, y\rangle = x^* Hy = x^* UDU^* y = (U^* x)^* D(U^* y)$$

Moreover, D can be expressed as:

$$D = \begin{bmatrix} \lambda & & \\ & -\mu & \\ & & 0 \end{bmatrix}, \text{ where } \lambda = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_k \end{bmatrix} \text{ and } \mu = \begin{bmatrix} \mu_1 & & \\ & \ddots & \\ & & \mu_l \end{bmatrix}$$

with $\lambda_i, \mu_j > 0$ for all i, j. The pair (k, l) is called the *signature* of the Hermitian form. We may further decompose the Hermitian form as:

$$D = \begin{bmatrix} \sqrt{\lambda} & & \\ & -\sqrt{\mu} & \\ & & 0 \end{bmatrix} \begin{bmatrix} I_k & & \\ & -I_l & \\ & & 0 \end{bmatrix} \begin{bmatrix} \sqrt{\lambda} & & \\ & -\sqrt{\mu} & \\ & & 0 \end{bmatrix} = U' I_{k,l} U'^*$$

where $\sqrt{\lambda} = \begin{bmatrix} \sqrt{\lambda_1} & & \\ & \ddots & \\ & & \sqrt{\lambda_k} \end{bmatrix}$ and $\sqrt{\mu} = \begin{bmatrix} \sqrt{\mu_1} & & \\ & \ddots & \\ & & \sqrt{\mu_l} \end{bmatrix}$.

So the Hermitian form can be represented as:

$$\langle x, y\rangle = (U'^* U^* x)^* I_{k,l} (U'^* U^* y)$$

Then for Hermitian products, as they are positive-definite, the canonical representation is simply I_n. For pseudo Hermitian products, the canonical representation is $I_{p,q}$ where $p+q=n$, i.e., there is no zero block. Up to isomorphism, there are $n+1$ different pseudo inner products on a complex vector space of dimension n, corresponding to the signatures $(n,0),(n-1,1),\cdots,(1,n-1),(0,n)$. Any pseudo Hermitian product space V is isomorphic to $(\mathbb{C}^n, I_{p,q}) = \mathbb{C}^{p,q}$. As it sends $(x,y) \to x_1\overline{y_1} + \cdots + x_p\overline{y_p} - x_{p+1}\overline{y_{p+1}} - \cdots - x_n\overline{y_n}$. In real vector spaces, the results are almost the same.

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 $\mathrm{O}(n)$ on $\mathrm{GL}_n(\mathbb{R})$ $\mathrm{GL}_n(\mathbb{R})/\mathrm{O}(n)$, where $\mathrm{O}(n)$ is the orthogonal group of order n.

$$\mathrm{GL}_n(\mathbb{R}) \times \mathrm{O}(n) \to \mathrm{GL}_n(\mathbb{R}), \quad (X,g) \mapsto X \cdot g$$

As $\mathrm{GL}_n(\mathbb{R})$ and $\mathrm{O}(n)$ have the same homotopy type, the orbit space $\mathrm{GL}_n(\mathbb{R})/\mathrm{O}(n)$ is trivially contractible. We may consider the following example:

$$\mathrm{GL}_1(\mathbb{R}) = \mathbb{R}^\times \quad \mathrm{O}(1) = \{-1, 1\}$$

Then we have:

$$\mathrm{GL}_1(\mathbb{R})/\mathrm{O}(1) \cong \mathbb{R}_{>0}$$

Similarly, the set of Hermitian forms on a complex vector space V of dimension n is isomorphic to the orbit space $\mathrm{GL}_n(\mathbb{C}) \,/\, \mathrm{U}(n)$, where $\mathrm{U}(n)$ is the unitary group of order n. Again, it is contractible.

7.6. Exercises

Problem 7.1. Let A be a real symmetric $n \times n$ matrix and A_i be the matrix obtained from A by deleting its last $n-i$ rows and last $n-i$ columns. Show that

$$A > 0 \iff \det A_i > 0 \text{ for each } i$$

Problem 7.2. Let A be a real symmetric $n \times n$ matrix. Assume that $A \geq 0$. It is clear that $A \geq 0 \iff A + tI > 0$ for any $t > 0$. Based on this observation and the result in the previous exercise to derive a necessary and sufficient condition for $A \geq 0$.

Problem 7.3. Let V be a n-dimensional linear space over a field F. The pairing $V^* \times V \to F$ yields the multilinear map $V^* \times V^* \times V \times V \to F$ that sends $(\alpha_1, \alpha_2, v_1, v_2)$ to $\alpha_1(v_1)\alpha_2(v_2)$. Thus we have a linear map

$$(V^* \otimes V^*) \otimes (V \otimes V) \to F$$

or equivalently a linear map $\iota : V^* \otimes V^* \to (V \otimes V)^*$.

(a) Show that the linear map ι is a linear equivalence. In fact a natural one in the language of category. So we shall write $V^* \otimes V^* \equiv (V \otimes V)^*$.

(b) The quotient linear map $V \otimes V \to S^2V$ yields the injective map $(S^2V)^* \to (V \otimes V)^*$. So we have the following composition map

$$(S^2V)^* \hookrightarrow (V \otimes V)^* \qquad \equiv \qquad V^* \otimes V^* \twoheadrightarrow S^2V^*$$

Show that this natural map $(S^2V)^* \to S^2V^*$ is a linear equivalence if and only if the characteristic of the field F is not 2.

(c) Assume that the characteristic of the field F is not 2, please find the inverse of the natural map in part (b).

(d) Assume that the characteristic of the field F is not 2, show that there is a natural linear equivalence $\bigwedge^2 V^* \equiv (\bigwedge^2 V)^*$.

Problem 7.4. Let ω be a bilinear form on V, i.e., a bilinear map $V \times V \to F$. Then $\omega_\natural : V \to V^*$ is the linear map tha sends v to $\omega(v, -)$. Let v_i be a basis of V and the resulting dual basis of V^* be denoted by $\hat{v}^i$. Let A be the matrix representation of ω with respect to the basis v_i, i.e., $A = [\omega(v_i, v_j)]$. Let A' be the matrix representation of $\omega_\natural$ with respect to the bases v_i and $\hat{v}^i$.

(a) Show that A' is the transpose of A. Thus ω is non-degenerate means that its any matrix representation is invertible.

(b) If $\tilde{v}_{\tilde{i}}$ is another basis of V, then there is a unique invertible matrix $S = [s^i_{\tilde{i}}]$ such that $\tilde{v}_{\tilde{i}} = v_i s^i_{\tilde{i}}$. Let $\tilde{A}$ be the matrix representation of ω with respect to the basis $\tilde{v}_{\tilde{i}}$. Show that $\tilde{A} = S^T A S$.

(c) Show that the map that sends (A, S) to $S^T A S$ is a right action of $\mathrm{GL}_n(F)$ on $(V \otimes V)^*$.

(d) The quotient map $V \otimes V \to S^2V$ yields injective linear map $(S^2V)^* \to (V \otimes V)^*$, that is not a surprise because any symmetric bilinear form is a bilinear form. Similarly, $(\bigwedge^2 V)^*$ is a linear subsubspace of $(V \otimes V)^*$ as well. Show that $\mathrm{GL}(V)$ acts on $(V \otimes V)^*$ and leaves invariant both subspaces $(S^2V)^*$ and $(\bigwedge^2 V)^*$.

(e) Show that, under the identification $(V \otimes V)^* \equiv V^* \otimes V^*$, we have

$$\omega = \omega(v_i, v_j)\hat{v}^i \otimes \hat{v}^j$$

(f) Assume that the characteristic of the field F is not 2, then we have the natural identification $(S^2V)^* \equiv S^2V^*$ found in the last question. Show that, if ω is a symmetric bilinear form on V, then $\omega = \omega(v_i, v_j)\hat{v}^i\hat{v}^j$. Similarly, if ω is a skew-symmetric bilinear form on V, then $\omega = \omega(v_i, v_j)\hat{v}^i \wedge \hat{v}^j$.

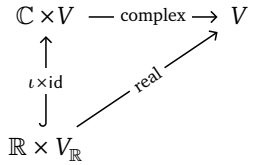
CHAPTER 8

Symplectic Linear Spaces

8.1. Complex Structures

Before we study symplectic linear spaces, we need to introduce some basic concepts about complex linear spaces. The following discussion is to show how a complex linear space can be viewed as a real linear space with some extra structure.

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. We losed some information from V to $V_{\mathbb{R}}$. If we want to recover back to the complex linear space, we need to add an extra structure $J : V_{\mathbb{R}} \to V_{\mathbb{R}}$ defined by $J^2 = -\mathrm{id}_{V_{\mathbb{R}}}$. It acts as the imaginary number i on $V_{\mathbb{R}}$ and simulate the action of multiplying the imaginary number i. Such a structure is called an *complex structure* on $V_{\mathbb{R}}$. The scalar multiplication on V and $V_{\mathbb{R}}$ are related by the following commutative diagram:



For example, we can write $(a+bi)v = av + bJ(v)$ for any complex number $a+bi$ and v in V. Note that as $(\det J)^2 = (-1)^{\dim_{\mathbb{R}}(V)}$, we have $\dim_{\mathbb{R}}(V)$ is even. The dimension doubled as we consider the basis of V as $\mathcal{B}_V = (v_1, v_2, \cdots, v_n)$ and basis of $V_{\mathbb{R}}$ as $\mathcal{B}_{V_{\mathbb{R}}} = (v_1, v_2, \cdots, v_n, J(v_1), J(v_2), \cdots, J(v_n)) \in V_{\mathbb{R}}$.

Definition 8.1 — Complex Structure.

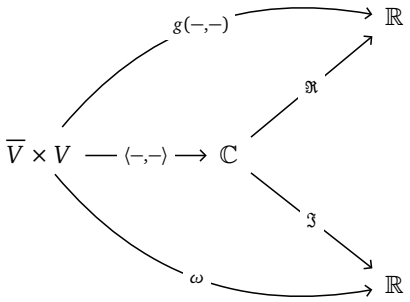
A *complex structure* on a real linear space W is a linear map $J : W \to W$ such that $J^2 = -\mathrm{id}_W$. A *complex linear space* is a pair (W, J) where W is a real linear space and J is a complex structure on W.

Actually, a real linear space is a complex linear space with a complex conjugation structure. The proof is left as an exercise in the end of this chapter.

8.2. Symplectic Structures

A symplectic structure is closely related to a Hermitian structure. We have learnt that complex structures are to recover complex linear spaces from real linear spaces. Similarly, symplectic structures with another structure, Riemannian structures, are to recover Hermitian linear spaces from real linear spaces.

Consider the Hermitian space V with Hermitian inner product $\langle -, - \rangle$. 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 8.2 — Symplectic Structure.

A *symplectic structure* on a real linear space V is a non-degenerate, skew-symmetric bilinear form $\omega : V \times V \to \mathbb{R}$. A *symplectic linear space* is a pair (V, ω) where V is a real linear space and ω is a symplectic structure on V.

Note that we have three structures on $V_{\mathbb{R}}$:

Complex Structure : $J : V_{\mathbb{R}} \to V_{\mathbb{R}}$ with $J^2 = -1_{V_{\mathbb{R}}}$;

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 for any $x, y \in V_{\mathbb{R}}$, we have:

$$\langle x, y\rangle = g(x, y) + i\omega(x, y) \tag{4}$$

Moreover, for any $x, y \in V_{\mathbb{R}}$, from $\langle x, iy\rangle = -i\langle x, y\rangle$, we have:

$$g(x, J(y)) + i\omega(x, J(y)) = \omega(x, y) - ig(x, y)$$

which implies that:

$$\omega(x, y) = g(x, J(y)) \tag{5}$$

$$g(x, y) = -\omega(x, J(y)) \tag{6}$$

For the compatibility between these three structures, for any $x, y \in V_{\mathbb{R}}$, from $\langle ix, iy\rangle = \langle x, y\rangle$, we have:

$$g(J(x), J(y)) + i\omega(J(x), J(y)) = g(x, y) + i\omega(x, y)$$

which implies that:

$$g(J(x), J(y)) = g(x, y), \tag{7}$$

$$\omega(J(x), J(y)) = \omega(x, y) \tag{8}$$

Shortly, we can write $J^*g = g$ and $J^*\omega = \omega$ where J^* is the pullback of J.

Note that the Hermitian product is positive-definite, i.e., $\langle x, x\rangle > 0$ for all $x \in V \setminus \{0\}$. Therefore, for any $x \in V_{\mathbb{R}} \setminus \{0\}$, we have:

$$g(x, x) > 0$$
$$\omega(x, x) = 0$$

If $x = 0$, then we have $\langle 0, 0\rangle = 0$, $g(0, 0) = 0$ and $\omega(0, 0) = 0$. Also, for any $x, y \in V_{\mathbb{R}}$, from the conjugate symmetry of Hermitian product, i.e., $\langle y, x\rangle = \overline{\langle x, y\rangle}$, we have:

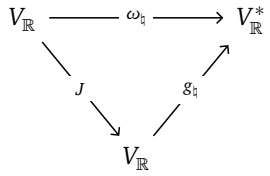
$$g(y, x) + i\omega(y, x) = g(x, y) - i\omega(x, y)$$

which implies that:

$$g(y,x) = g(x,y)$$
$$\omega(y,x) = -\omega(x,y)$$

This shows that g is symmetric and ω is skew-symmetric.

As $\omega(x,y) = g(x,J(y))$ for all $x,y \in V_{\mathbb{R}}$, i.e., $\omega_\natural = g_\natural \circ J$, so ω is non-degenerate if g is non-degenerate. Then we have the following commutative diagram:



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,J) where $iv = Jv$ for all $v \in V$.

The following three propositions show the meaning of compatibility between these three structures and how to recover the Hermitian product from any two compatible structures.

Proposition 8.2.1 — Compatibility of Complex and Riemannian Structures. If (g,J) are compatible, then there is a unique symplectic structure ω such that (g,ω,J) are compatible.

Proof. The pair (g,J) is compatible if $J^*g = g$, i.e., $J \in \mathrm{Aut}(W,g) = \mathrm{O}(W,g)$; Then we can define $\omega(x,y) = g(x,J(y))$ and $\langle -,-\rangle = g + i\omega$. We can check that ω is skew-symmetric and non-degenerate, and $\langle -,-\rangle$ is a Hermitian product:

$$\omega(y,x) = g(y,J(y)) = g(J(y),J^2(x))) = g(J(y),-x) = -g(J(y),x) = -g(x,J(y)) = -\omega(x,y)$$

Also, if $\omega(x,y) = 0$ for all $y \in V$, then we have $g(x,J(y)) = 0$ for all $y \in V$, which implies that $Jx = 0$ 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:

$$\begin{aligned}\langle ix,y\rangle &= g(J(x),y) + i\omega(J(x),y) = g(x,-J(y)) + ig(J(x),J(y))\\ &= -\omega(x,y) + ig(x,y) = i(g(x,y) + i\omega(x,y))\\ &= i\langle x,y\rangle\\ \langle x,iy\rangle &= g(x,J(y)) + i\omega(x,J(y)) = \omega(x,y) + ig(x,J^2(y))\\ &= \omega(x,y) - ig(x,y) = -i(g(x,y) + i\omega(x,y))\\ &= -i\langle x,y\rangle\end{aligned}$$

The conjugate symmetry is shown as follows for any $x,y \in V$:

$$\langle y,x\rangle = g(y,x) + i\omega(y,x) = g(x,y) - i\omega(x,y) = \overline{\langle x,y\rangle}$$

The positive-definiteness is shown as follows for any $x \in V \setminus \{0\}$:

$$\langle x,x\rangle = g(x,x) + i\omega(x,x) = g(x,x) > 0$$

□

Proposition 8.2.2 — Compatibility of Complex and Symplectic Structures. If (ω,J) are compatible, then there is a unique Riemannian structure g such that (g,ω,J) are compatible.

Proof. The pair (ω,J) is compatible if $J^*\omega = \omega$, i.e., $J \in \mathrm{Aut}(W,\omega) = \mathrm{Sp}(W,\omega)$ and $\omega(x,J(x)) \geq 0$ and equality holds if and only if $x = 0$. Then we can define $g(x,y) = \omega(x,J(y))$ and $\langle -,-\rangle = g + i\omega$. We can check that g is symmetric and positive-definite, and $\langle -,-\rangle$ is a Hermitian product:

$$g(y,x) = \omega(y,J(x)) = \omega(J(y),J^2(x))) = \omega(J(y),-x) = -\omega(J(y),x) = \omega(x,J(y)) = g(x,y)$$

Also, as $\omega(x,J(x)) \geq 0$ 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. □

Proposition 8.2.3 — Compatibility of Riemannian and Symplectic Structures. If (g,ω) are compatible, then there is a unique complex structure J such that (g,ω,J) are compatible.

Proof. The pair (g,ω) is compatible if $\omega(x,y) = g(A(x),y)$ for some $A \in \operatorname{End} V$. If $A^2 = -1$, then $J = A$. In general, A is skew-symmetric relative to g, i.e., $g(A(x),y) = g(x,-A(y)) = -g(x,A(y))$, as ω is skew-symmetric. Since $AA^* = -A^2$ is symmetric and positive-definite, we define $P = \sqrt{AA^*}$ and the complex structure $J = -P^{-1}A$, which satisfies that $J^2 = -1$, as J commutes with A and P. Then we have $A = -PJ$. Therefore, we have:

$$\omega(J(x),J(y)) = g(AJ(x),J(y)) = g(-PJ^2(x),J(y)) = g(P(x),J(y)) = g(-PJ(x),y) = \omega(x,y).$$

Also, we have:

$$\omega(x,J(x)) = g(A(x),J(x)) = g(-PJ(x),J(x)) = g(P(x),x) > 0$$

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. □

Consider a F-linear space V where the characteristic of F is not 2. Recall the definition of double $D(V)$ in Definition 3.16. Then we have a natural symplectic structure on $D(V)$ defined as:

$$\omega((u,\alpha),(v,\beta)) = \alpha(v) - \beta(u)$$

Then $D(V)$ is also 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 $D(V)$.

8.3. Matrix Representation and Canonical Form of Symplectic Structures

Consider a bilinear form $\omega : V \times V \to F$ on a vector space V of dimension n. Then we have:

$$\begin{array}{ccc} V \times V & \xrightarrow{\quad\omega\quad} & F \\ \uparrow & & \nearrow \\ {\scriptstyle [-]_{\mathcal{B}}} & & \\ \downarrow & & \\ F^n \times F^n & & \end{array}$$

Then $[\omega]_{\mathcal{B}} = [\omega(v_i,v_j)]$ is the matrix representation of ω with respect to the basis $\mathcal{B} = \{v_1,v_2,\cdots,v_n\}$ of V. If we change the basis of V to $\mathcal{B}' = \{u_1,u_2,\cdots,u_n\}$, then there is a unique invertible matrix, P in $\mathrm{GL}_n(F)$, such that $u_j = \sum_i v_i P^i_j$ for all j. Then we have:

$$[\omega]_{\mathcal{B}'} = [\omega(u_i,u_j)] = \left[\omega\left(\sum_k v_k P^k_i, \sum_l v_l P^l_j\right)\right] = \left[\sum_{k,l} P^k_i \omega(v_k,v_l) P^l_j\right]$$

$$= \left[\sum_{k,l} (P^T)^i_k \omega(v_k,v_l) P^l_j\right] = P^T[\omega(v_k,v_l)]P$$

So we have the right group action of $\mathrm{GL}_n(F)$ on the set of $n \times n$ matrices $\mathrm{Mat}_{n\times n}(F)$ defined as:

$$A \cdot P = P^T A P$$

We may check that $(A \cdot P_1) \cdot P_2 = A \cdot (P_1 P_2)$ for all $A \in \mathrm{Mat}_{n\times n}(F)$ and $P_1,P_2 \in \mathrm{GL}_n(F)$.

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 $(P^TAP)^T = P^TAP$ (or $(P^TAP)^T = -P^TAP$) for all $P \in \mathrm{GL}_n(F)$. For symmetric 2-forms, as $(P^TAP)^T = P^TA^T(P^T)^T = P^TA^TP$, where $A^T = A$, so we have $(P^TAP)^T = P^TAP$. For skew-symmetric 2-forms, as $(P^TAP)^T = P^TA^T(P^T)^T = P^T(-A)P$, where $A^T = -A$, so we have $(P^TAP)^T = -P^TAP$.

When $F = \mathbb{R}$, then the skew-symmetric ω corresponds to a real skew-symmetric matrix. The matrix iA is a Hermitian matrix, and we have $iA = UDU^*$ 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 $J_2^2 = -I_2$.

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.

8.4. Exercises

Problem 8.1. The goal of this exercise is to help you understand the relation between complex linear spaces and real linear spaces with complex structures.

(a) Let V be a complex linear space and J be the endomorphism $i \cdot \mathrm{id}_V$ on $V_{\mathbb{R}}$. Show that $J^2 = -\mathrm{id}_{V_{\mathbb{R}}}$.

(b) Show that if $v = (v_1, \ldots, v_n)$ is a basis of the complex linear space V, then $v_{\mathbb{R}} := (v_1, Jv_1, \ldots, v_n, Jv_n)$ is a basis of the real linear space $V_{\mathbb{R}}$. Thus $\dim_{\mathbb{R}} V_{\mathbb{R}} = 2 \dim_{\mathbb{C}} V$.

(c) Show that if $w = (w_1, \ldots, w_m)$ is a basis of the real linear space W, then $w \equiv w \otimes 1$ is a basis of the complex linear space $W_{\mathbb{C}}$. Thus $\dim_{\mathbb{R}} W = \dim_{\mathbb{C}} W_{\mathbb{C}}$.

(d) Show that a complex linear space with underlying real linear space W is isomorphic to the real linear space W with a complex structure J on W. Also, if $V := (W, J)$ is a complex linear space, then its complex conjugate $\overline{V}$ is the complex linear space $(W, -J)$.

(e) Let V be a complex linear space and $J \in \mathrm{End}\, V_{\mathbb{R}}$ be the scalar multiplication by i on V. We prefer to write $V_{\mathbb{R}} \otimes_{\mathbb{R}} \mathbb{C}$ as $V \otimes_{\mathbb{R}} \mathbb{C}$ or simply $V_{\mathbb{C}}$. Note that J extends to $J_{\mathbb{C}} := J \otimes_{\mathbb{R}} \mathrm{id}_{\mathbb{C}}$. Show that $J_{\mathbb{C}}$ is an endomorphism of the complex linear space $V_{\mathbb{C}}$ such that $J_{\mathbb{C}}^2 = -\mathrm{id}_{V_{\mathbb{C}}}$. Note: the complex structure on $V_{\mathbb{C}}$ comes from the complex structure on $\mathbb{C}$, not from that of V.

(f) Show that the real linear map $V \to V_{\mathbb{C}}$ that sends v to $v \otimes_{\mathbb{R}} 1 - Jv \otimes_{\mathbb{R}} i$ is an *embedding* of complex linear spaces, i.e., an injective complex linear map. Let us denote the image of this embedding as V'. Then $V \equiv V'$.

(g) Show that the real linear map $\overline{V} \to V_{\mathbb{C}}$ that sends v to $v \otimes_{\mathbb{R}} 1 + Jv \otimes_{\mathbb{R}} i$ is an embedding of complex linear spaces. Let us denote the image of this embedding as V''. Then $\overline{V} \equiv V''$.

(h) Show that $V_{\mathbb{C}} = V' \oplus V'' \equiv V \oplus \overline{V}$. People usually write $V_{\mathbb{C}} = V \oplus \overline{V}$.

Let V be a finite-dimensional complex linear space. An endomorphism σ on $V_{\mathbb{R}}$ is called a *complex conjugation* on the complex linear space V if σ satisfies the following two conditions:

– $\sigma(cu) = \overline{c}\sigma(u)$ for any $c \in \mathbb{C}$ and $u \in V$;

– $\sigma^2 = \mathrm{id}_{V_{\mathbb{R}}}$.

If V is a finite-dimensional complex linear space with complex conjugation σ, we let V^σ be the set of σ-invariant vectors, i.e.,

$$V^\sigma = \{v \in V : \sigma(v) = v\}.$$

Problem 8.2. We have seen that a complex linear space is nothing but a real lineaer space together with a complex structure J. The goal of this exercise is to enable you to see that a real linear space is nothing but a complex linear space together with a complex conjugation.

(a) Let W be a finite-dimensional real linear space. Let σ_{std} be the complex conjugation on $W_{\mathbb{C}}$ that sends $w \otimes c$ to $w \otimes \overline{c}$. Show that $W \equiv W_{\mathbb{C}}^{\sigma_{\mathrm{std}}}$ under which w in W becomes $w \otimes 1$ in $W_{\mathbb{C}}^{\sigma_{\mathrm{std}}}$.

(b) Let V be a finite-dimensional complex linear space with complex conjugation σ. Show that V^σ is a real linear space and $V^\sigma \otimes \mathbb{C} \equiv V$ under which σ_{std} on $V^\sigma \otimes \mathbb{C}$ becomes σ on V.

(c) Let T be an endomorphism on the real linear space W. Then $T_{\mathbb{C}} := T \otimes \mathrm{id}_{\mathbb{C}}$ is an endomorphism on the complex linear space $W_{\mathbb{C}}$. Show that $T_{\mathbb{C}}\sigma_{\mathrm{std}} = \sigma_{\mathrm{std}} T_{\mathbb{C}}$.

(d) Show that an eigenvalue of $T_{\mathbb{C}}$ is either a real number or a complex number with its complex conjugation being also an eigenvalue.

(e) Assume that $T_{\mathbb{C}}$ is "diagonalisable", then its eigenspace decomposition of $W_{\mathbb{C}}$ must be of the form

$$W_{\mathbb{C}} = E_{\lambda_1} \oplus \cdots \oplus E_{\lambda_t} \oplus E_{\mu_1} \oplus \overline{E_{\mu_1}} \oplus \cdots \oplus E_{\mu_s} \oplus \overline{E_{\mu_s}}$$

where λ_i are real numbers, μ_i are complex numbers, E_{λ_i} are the eigenspaces associated to λ_i, E_{μ_i} are the eigenspaces associated to μ_i and $\overline{E_{\mu_i}}$ are the complex conjugate of E_{μ_i}, thus must be the eigenspace with eigenvalue $\overline{\mu_i}$. By convention, $t = 0$ means no real eigenvalues and $s = 0$ means no complex numbers.

(f) Continuing the previous part, show that W has a decomposition of the form

$$W = E_{\lambda_1}(T) \oplus \cdots \oplus E_{\lambda_t}(T) \oplus E_{\mu_1}(T) \oplus \cdots \oplus E_{\mu_s}(T)$$

with respect to which we have $T = T_{\lambda_1} \oplus \cdots \oplus T_{\lambda_t} \oplus T_{\mu_1} \oplus \cdots \oplus T_{\mu_s}$; Moreover, $E_{\lambda_i}(T)$ is the eigenspace of T associated to the real eigenvalue λ_i and $E_{\mu_i}(T)$ is a real linear space with a complex structure J_i such that the characteristic polynomial of T_{μ_i} is a power of the real irreducible quadratic polynomial $x^2 - (\mu_i + \overline{\mu_i})x + |\mu_i|^2$.

The conclusions here are useful and are corollaries of spectral theorem for Hermitian matrices and unitary matrices. By now I hope you feel comfortable with switching between linear maps or forms and their matrix representations. You need some help from Problem 1. In principle, you don't need Problem 2 for help because we only deal with matrices here. This shows again that the "cheaty method", i.e., the method of matrix representation, is powerful.

Problem 8.3. Let V be an Euclidean vector space. Then its complexification $V_{\mathbb{C}}$ becomes a Hermitian vector space. One way to see it is this: the Hermitian inner product is the one such that $\langle u \otimes \alpha, v \otimes \beta \rangle = \overline{\alpha}\beta \langle u, v \rangle$ for any $u, v \in V$ and complex numbers α, β.

(a) Show that the canonical form of a Hermitian 2-form on a complex linear space is a matrix of the form $I_p \oplus -I_q \oplus [0] \oplus \cdots \oplus [0]$. Then conclude that the canonical form of a pseudo-Hermitian inner product is a matrix of the form $I_{p,q} := I_p \oplus -I_q$.

(b) Let T be a self-adjoint operator on V. Show that $T_{\mathbb{C}} := T \otimes \mathrm{id}_{\mathbb{C}}$ is a self-adjoint operator on $V_{\mathbb{C}}$. In terms of matrix representation with respect to orthonormal bases, it says that a real symmetric matrix is a Hermitian matrix.

(c) Show that any real symmetric matrix A can be diagonalised by an orthogonal matrix, i.e., $A = O^T D O$ for some orthogonal matrix and D is a real diagonal matrix.

(d) Show that the canonical form of a symmetric 2-form on a real linear space is a matrix of the form $I_p \oplus -I_q \oplus [0] \oplus \cdots \oplus [0]$. Then conclude that the canonical form of a pseudo-Euclidean inner product is a matrix of the form $I_{p,q} := I_p \oplus -I_q$.

(e) Let T be an orthogonal transformation on V. Show that $T_{\mathbb{C}}$ is a unitary transformation on $V_{\mathbb{C}}$. In terms of matrix representation with respect to orthonormal bases, it says that an orthogonal matrix is a unitary matrix.

(f) Denote by $R(\theta)$ the 2×2 rotation matrix with rotation angle θ. By definition, a special orthogonal matrix is an orthogonal matrix whose determinant is 1. Show that any special orthogonal matrix A can be factorised this way: $A = O^T \overline{A} O$ where O is an orthogonal matrix and $\overline{A}$ is a canonical special orthogonal matrix, i.e., a matrix of the form $R(\theta_1) \oplus \cdots \oplus R(\theta_k)$ for some angles θ_i if $n = 2k$ or a matrix of the form $R(\theta_1) \oplus \cdots \oplus R(\theta_k) \oplus [1]$ for some angles θ_i if $n = 2k + 1$. In fact O can be chosen to be a special orthogonal matrix.

(g) Denote by J_2 the real skew-symmetric matrix with its $(2,1)$-entry being 1. Let A be a real skew-symmetric matrix. Show that $A = P^T \overline{A} P$ where P is an invertible matrix and $\overline{A}$ is a real skew-symmetric matrix of the form $J_2 \oplus J_2 \oplus \cdots \oplus J_2 \oplus [0] \oplus \cdots \oplus [0]$. Then conclude that the canonical form of a symplectic form ω on a real linear space V is a matrix of the form $J_2 \oplus J_2 \oplus \cdots \oplus J_2$. In this case, the basis made of the columns of P^T is called a symplectic basis of the symplectic space (V, ω).

CHAPTER 9

Further Topics

9.1. Polar Decomposition and Singular Value Decomposition

9.1.1. 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:

$$\mathrm{GL}_1(\mathbb{C}) = \mathrm{U}(1) \cdot \mathrm{H}_1^{>0}(\mathbb{C}), \quad [z] \mapsto [e^{i\theta}] \cdot [\rho]$$

where $\mathrm{H}_1^{>0}(\mathbb{C})$ is the set of positive Hermitian 1×1 matrices, i.e., positive real numbers, and $\mathrm{U}(1)$ is the set of complex numbers of modulus 1.

Then we may generalise this idea to matrices.

Theorem 9.1 — Polar Decomposition.

For any invertible complex matrix A of order n, there exists a unique decomposition:

$$A = UP \tag{9}$$

where $P \in \mathrm{H}_n^{>0}(\mathbb{C})$ and $U \in \mathrm{U}(n)$. This shows that there is an isomorphism:

$$\mathrm{GL}_n(\mathbb{C}) \cong \mathrm{U}(n) \cdot \mathrm{H}_n^{>0}(\mathbb{C}).$$

Proof. Assume the existance, if $A = UP$ then $A^* = PU^*$. Then we have:

$$A^*A = PU^*UP = P^2$$

As A is invertible, so is A^*A. Therefore, $P = \sqrt{A^*A}$ is a positive Hermitian matrix. Then we have $U = AP^{-1}$. Also, we have:

$$(A^*A)^* = A^*A \implies P^*P^* = P^2 \implies P^* = P$$

and

$$\vec{z}^*A^*A\vec{z} = (A\vec{z})^*(A\vec{z}) > 0 \implies \|P\vec{z}\| \geq 0$$

for all $\vec{z}$ and equal to 0 if and only if $\vec{z} = 0$ as $A \in \mathrm{GL}_n(\mathbb{C})$. Therefore, P and A^*A are positive Hermitian. Then we know that $A^*A = U'DU'^*$ where $U' \in \mathrm{U}(n)$ and D is a diagonal matrix with positive real numbers on the diagonal. Then we have $P = U'\sqrt{D}U'^*$. Also, we have:

$$P^2 = U'\sqrt{D}U'^*U'\sqrt{D}U'^* = U'DU'^* = A^*A$$

Then we have:

$$U^*U = P^{-1}A^*AP^{-1} = P^{-1}P^2P^{-1} = I_n$$

Therefore, $U \in \mathrm{U}(n)$. □

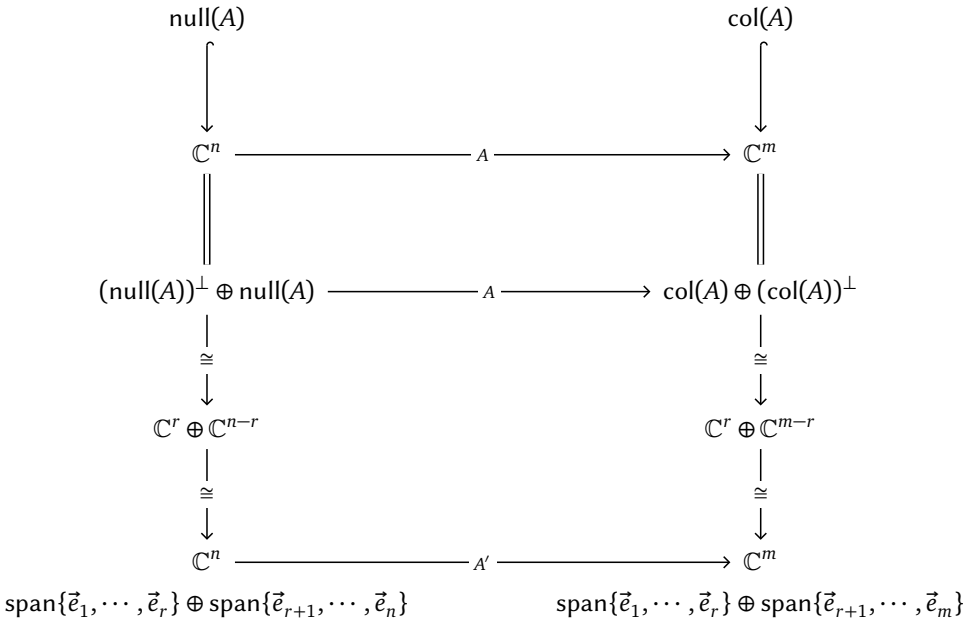
If it is real number, then we have the similar polar decomposition:

$$\mathrm{GL}_n(\mathbb{R}) = \mathrm{O}(n) \cdot \mathrm{S}_n^{>0}(\mathbb{R}), \quad [A] \mapsto [O] \cdot [S]$$

where $\mathrm{S}_n^{>0}(\mathbb{R})$ is the set of positive symmetric $n \times n$ matrices and $\mathrm{O}(n)$ is the orthogonal group of order n.

9.1.2. 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 $(\text{null}A)^\perp \oplus \text{null}A$ and $\text{col}A \oplus (\text{col}A)^\perp$ 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 $\text{span}\{\vec{e}_1, \cdots, \vec{e}_r\} \oplus \text{span}\{\vec{e}_{r+1}, \cdots, \vec{e}_n\}$ and $\text{span}\{\vec{e}_1, \cdots, \vec{e}_r\} \oplus \text{span}\{\vec{e}_{r+1}, \cdots, \vec{e}_m\}$ are internal direct sums. Note that all the isomorphisms in the diagram are of Hermitian spaces. Then we may simplify the diagram as follows:

$$\begin{array}{ccc} \mathbb{C}^n & \xrightarrow{A} & \mathbb{C}^m \\ \downarrow{\scriptstyle U_1} & & \downarrow{\scriptstyle U_2} \\ \mathbb{C}^n & \xrightarrow{A'} & \mathbb{C}^m \end{array}$$

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 $P = U_4 D_\lambda U_4^*$ 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:

$$\begin{aligned} U_2 A &= \begin{bmatrix} \overline{A} & 0 \\ 0 & 0 \end{bmatrix} U_1 \\ A &= U_2^* \begin{bmatrix} U_3 U_4 D_\lambda U_4^* & 0 \\ 0 & 0 \end{bmatrix} U_1 \\ &= \left(U_2^* \begin{bmatrix} U_3 U_4 & 0 \\ 0 & I_{m-r} \end{bmatrix} \right) \begin{bmatrix} D_\lambda & 0 \\ 0 & 0 \end{bmatrix} \left(\begin{bmatrix} U_4^* & 0 \\ 0 & I_{n-r} \end{bmatrix} U_1 \right) \end{aligned}$$

Then we have the singular value decomposition of A:

$$A = U \Sigma V^*$$

where $U = U_2^* \begin{bmatrix} U_3 U_4 & 0 \\ 0 & I_{m-r} \end{bmatrix}$, $\Sigma = \begin{bmatrix} D_\lambda & 0 \\ 0 & 0 \end{bmatrix}$, $V = U_1^* \begin{bmatrix} U_4 & 0 \\ 0 & I_{n-r} \end{bmatrix}$.

Theorem 9.2 — Singular Value Decomposition.
For any $A \in \mathrm{Mat}_{m\times n}(\mathbb{C})$, there exist unitary matrices $U \in \mathrm{U}(m)$, $V \in \mathrm{U}(n)$ and a set of positive numbers $\{\lambda_1, \cdots, \lambda_r\}$ such that:

$$A = U\Sigma V^*, \quad \Sigma = \begin{bmatrix} \lambda_1 & & & & \\ & \lambda_2 & & & \\ & & \ddots & & \\ & & & \lambda_r & \\ & & & & 0 \end{bmatrix}$$

9.2. Simultaneous Diagonalisation Theorem

Theorem 9.3 — Simultaneous Diagonalisation Theorem.
Suppose that $A_1, \cdots, A_k$ are mutually commuting Hermitian matrices of order n, i.e., $A_i \in \mathrm{H}_n(\mathbb{C})$ and $[A_i, A_j] := A_iA_j - A_jA_i = 0$ for all $1 \le i, j \le k$, 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 $1 \le i \le k$. In particular, there is a unitary matrix $U \in \mathrm{U}(n)$ such that:

$$A_i = UD_iU^*, \quad D_i = \begin{bmatrix} d_1(i) & & \\ & \ddots & \\ & & d_n(i) \end{bmatrix} \in \mathrm{Mat}_n(\mathbb{R})$$

for all $1 \le i \le k$ 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 9.4 — Spectral Theorem for Normal Operators.
A complex square matrix can be diagonalised by a unitary matrix if and only if it is normal.

Proof.
(⇒): Assume that A can be diagonalised by a unitary matrix, i.e., there is a unitary matrix U such that $A = UDU^*$ where D is a diagonal matrix. Then we have:

$$A^* = UD^*U^*$$

where D^* is also a diagonal matrix. Then we have:

$$AA^* = UDU^*UD^*U^* = UDD^*U^* = UD^*DU^* = A^*A$$

$DD^* = D^*D$ 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.

(⇐): Assume that A is normal, i.e., $AA^* = A^*A$. Then we write $A = B + iC$ where $B = \frac{A+A^*}{2}$ and $C = \frac{A-A^*}{2i}$. Then we claim that $[B, C] = 0$ if and only if A is normal. We have:

$$AA^* = (B + iC)(B - iC) = B^2 + C^2 - i[B, C]$$
$$A^*A = (B - iC)(B + iC) = B^2 + C^2 + i[B, C]$$

Therefore, $AA^* = A^*A$ if and only if $[B, C] = 0$. Also, we may check that B and C are Hermitian:

$$B^* = \left(\frac{A+A^*}{2}\right)^* = \frac{A^*+A}{2} = B, \quad C^* = \left(\frac{A-A^*}{2i}\right)^* = \frac{A^*-A}{-2i} = C.$$

Then, by the simultaneous diagonalisation theorem, there is a unitary matrix U such that:

$$B = UD_BU^*, \quad C = UD_CU^*$$

where D_B and D_C are diagonal matrices. Therefore, we have:

$$A = B + iC = UD_BU^* + iUD_CU^* = U(D_B + iD_C)U^* = UD_AU^*$$

where $D_A = D_B + iD_C$ is a diagonal matrix. Hence, A can be diagonalised by a unitary matrix. □

9.3. 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 F be a field. An affine space of dimension n over F, $\mathbb{A}$, is a principal $(F^n, +)$-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 9.3.1. F^n is an affine space of dimension n over F with the usual addition action of $(F^n, +)$ on itself.

$$(F^n, +) \times F^n \to F^n$$
$$(\vec{v}, \vec{x}) \mapsto \vec{v} + \vec{x}$$

For any $\vec{x}, \vec{y} \in F^n$, there is a unique $\vec{v} = \vec{y} - \vec{x} \in F^n$ such that $\vec{v} + \vec{x} = \vec{y}$. Therefore, the action is transitive and free.

In fact, any F-linear space is a F-affine space.

Problem 9.1. 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 $\mathbb{A}_F^n := \{(x_1, \cdots, x_n) \mid x_i \in F\}$. Then the group action is:

$$\begin{aligned}(F^n, +) \times \mathbb{A}_F^n &\to \mathbb{A}_F^n \\ (\vec{v}, \vec{x}) &\mapsto \vec{v} + \vec{x} := (v_1 + x_1, \cdots, v_n + x_n)\end{aligned}$$

for all $\vec{v} = (v_1, \cdots, v_n) \in F^n$ and $\vec{x} = (x_1, \cdots, x_n) \in \mathbb{A}_F^n$. Moreover, up to isomorphism, there is only one affine space of dimension n over 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:

$$p_0, p_1, \cdots, p_k \in \mathbb{A}, \quad c^0, c^1, \cdots, c^k \in F$$

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 Figure 9: The red line is the smallest

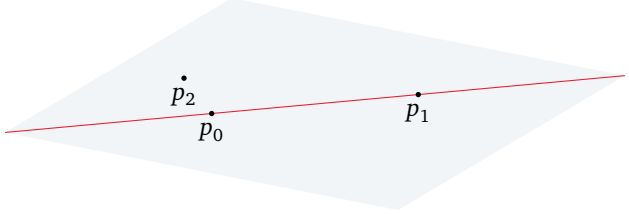


Figure 9. Affine Span and Affine Frame

affine subspace containing p_0 and p_12, i.e., the affine span of p_0 and p_1. We may write $\text{span}\{p_0, p_1\} := \{c^0p_0+c^1p_1 \mid c^0+c^1 = 1, c^i \in \mathbb{R}\} = \{tp_0+(1-t)p_1 \mid t \in \mathbb{R}\}$. Note that $\overline{p_0p_1} = \{tp_0+(1-t)p_1 \mid t \in [0,1]\}$ 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:

$$\begin{array}{ccc} \mathbb{A}_1 & \xrightarrow{\quad\phi\quad} & \mathbb{A}_2 \\ \downarrow & & \downarrow \\ \mathbb{A}_F^n & \longrightarrow & \mathbb{A}_F^m \\ \downarrow & & \downarrow \\ F^n & \xrightarrow{\quad A\quad} & F^m \\ \vec{x} = x - 0 & \longmapsto & A\vec{x} + \vec{b} \end{array}$$

where $A \in \text{Mat}_{m\times n}(F)$ and $\vec{b} \in F^m$. 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:

$$\alpha_1q_1 + \alpha_2q_2 = p + \alpha_1(q_1 - p) + \alpha_2(q_2 - p)$$

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., $\langle \vec{pq}, \vec{pr}\rangle = \langle \vec{p'q'}, \vec{p'r'}\rangle$. Consider the following Figure 10: Then $q = q' + \vec{v}$ and $r = r' + \vec{v}$. Note that $\mathcal{T}_p\,\mathbb{A}$ is different from $\mathcal{T}_{p'}\,\mathbb{A}$

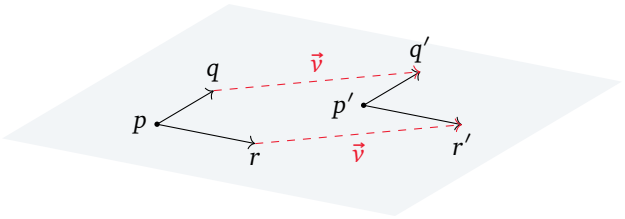


Figure 10. Translation Invariance

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 $\mathbb{E}^n := (\mathbb{A}^n_{\mathbb{R}}, \langle -, -\rangle)$ where $\langle -, -\rangle$ 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.

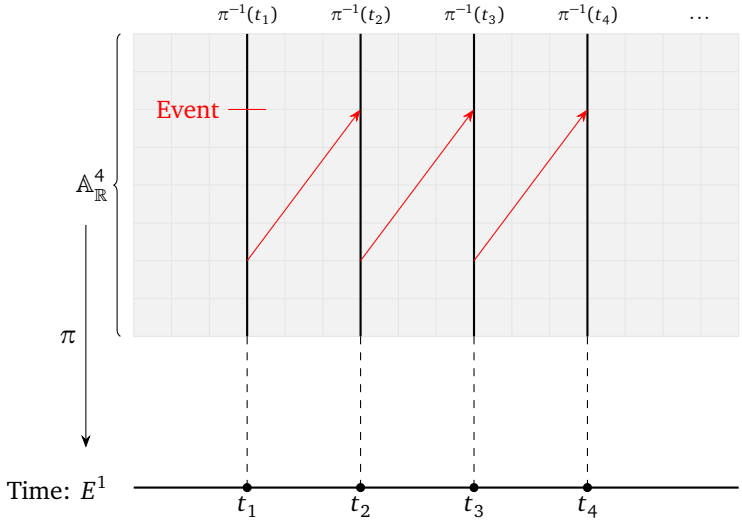
For an affine map $\phi : \mathbb{A}_1 \to \mathbb{A}_2$ between two affine spaces, we say that ϕ is injective implies that $\dim \mathbb{A}_1 \leq \dim \mathbb{A}_2$. The proof is by picking a point $p_1 \in \mathbb{A}_1$ and take $p_2 = \phi(p_1)$. Then we have the following commutative diagram:

$$\begin{array}{ccc} (\mathbb{A}_1, p_1) & \xrightarrow{\mathcal{T}_{p_1}\phi} & (\mathbb{A}_2, p_2) \\ \downarrow \cong & & \downarrow \cong \\ \mathbb{A}_1 & \xrightarrow{\phi} & \mathbb{A}_2 \end{array}$$

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:



9.4. Quadratic Form and Clifford Algebra

Let V be a vector space over a field F. A quadratic form on V is a map $q : V \to F$ such that:

– $q(\alpha v) = \alpha^2 q(v)$ for all $\alpha \in F$ and $v \in V$;

– The map $B : V \times V \to F$ defined by $B(u, v) = q(u + v) - q(u) - q(v)$ is bilinear.

In case the characteristic of F is not equal to 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 F of order $\dim V = n$. Moreover, (V, q) forms a quadratic space.

> **Remark.** When the characteristic of F is 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 $\mathrm{Cl}(V, q) := \mathcal{T}^\bullet V / I_q$ is an associative algebra over 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 $\mathrm{Cl}(V, q)$ is $\mathbb{Z}/2$ graded algebra.

We have the following isomorphisms:

– $\mathrm{Cl}(\mathbb{R}^{0,1}) \cong \mathbb{C}$ as $\mathbb{R}$-algebras, where elements in $\mathrm{Cl}(\mathbb{R}^{0,1})$ are of the form $a + be_1$ with $e_1^2 = -1$;

– $\mathrm{Cl}(\mathbb{R}^{1,0}) \cong \mathbb{R} \oplus \mathbb{R}$, the split-complex number, where elements in $\mathrm{Cl}(\mathbb{R}^{1,0})$ are of the form $a + be_1$ with $e_1^2 = 1$;

– $\mathrm{Cl}(\mathbb{R}^{0,2}) \cong \mathbb{H}$, the quaternion, as $\mathbb{R}$-algebras, where elements in $\mathrm{Cl}(\mathbb{R}^{0,2})$ are of the form $a + be_1 + ce_2 + de_1e_2$ with $e_1^2 = e_2^2 = -1$ and $e_1e_2 = -e_2e_1$;

– $\mathrm{Cl}(\mathbb{R}^{1,1}) \cong \mathrm{Mat}_2(\mathbb{R})$, the split-quaternion, as $\mathbb{R}$-algebras, where elements in $\mathrm{Cl}(\mathbb{R}^{1,1})$ are of the form $a + be_1 + ce_2 + de_1e_2$ with $e_1^2 = 1$, $e_2^2 = -1$ and $e_1e_2 = -e_2e_1$;

– $\mathrm{Cl}(\mathbb{R}^{2,0}) \cong \mathrm{Mat}_2(\mathbb{R})$, the split-quaternion, as $\mathbb{R}$-algebras.

The $\mathbb{R}$, $\mathbb{C}$ and $\mathbb{H}$ are called the associative real division algebras.

Appendix: Fudan University Problems

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

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"

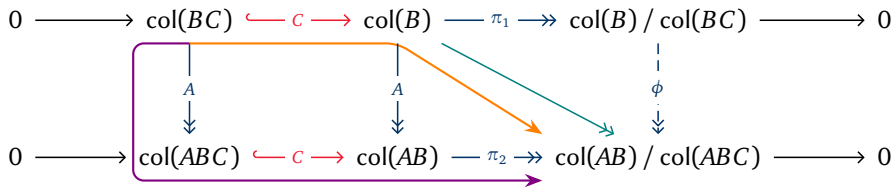
— Guowu Meng on the lecture of September 19, 2025.

Problem 1.

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

$$\mathrm{rank}(B) + \mathrm{rank}(ABC) \geq \mathrm{rank}(AB) + \mathrm{rank}(BC)$$

Solution. 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 $\mathrm{col}(B)$ to $\mathrm{col}(AB) / \mathrm{col}(ABC)$ 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 $\mathrm{col}(BC)$ to $\mathrm{col}(AB) / \mathrm{col}(ABC)$ is zero. If the map is zero, then we can construct a unique surjective map ϕ from $\mathrm{col}(B) / \mathrm{col}(BC)$ to $\mathrm{col}(AB) / \mathrm{col}(ABC)$ due to the universal property of quotient space.

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

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

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

$$\begin{aligned}
\dim(\mathrm{col}(B) \,/\, \mathrm{col}(BC)) &\geq \dim(\mathrm{col}(AB) \,/\, \mathrm{col}(ABC)) \\
\dim(\mathrm{col}(B)) - \dim(\mathrm{col}(BC)) &\geq \dim(\mathrm{col}(AB)) - \dim(\mathrm{col}(ABC)) \\
\dim(\mathrm{col}(B)) + \dim(\mathrm{col}(ABC)) &\geq \dim(\mathrm{col}(AB)) + \dim(\mathrm{col}(BC)) \\
\mathrm{rank}(B) + \mathrm{rank}(ABC) &\geq \mathrm{rank}(AB) + \mathrm{rank}(BC)
\end{aligned}$$

Problem 2.

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

$$\mathrm{rank}(A^n) = \mathrm{rank}(A^{n+1})$$

Solution. We consider the following diagram:

$$I_n \xrightarrow{A} \mathrm{im}(A) \xrightarrow{A} \mathrm{im}(A^2) \xrightarrow{A} \cdots \xrightarrow{A} \mathrm{im}(A^n) \xrightarrow{A} \cdots$$

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

$$n = \dim(I_n) \geq \mathrm{rank}(A) \geq \mathrm{rank}(A^2) \geq \cdots$$

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 $\mathrm{rank}(A^j) = \mathrm{rank}(A^{j+1})$.

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 $\mathrm{rank}(A^j) = \mathrm{rank}(A^{j+1})$ implies $\mathrm{rank}(A^n) = \mathrm{rank}(A^{n+1})$.

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 $k > n$.

Appendix: Zariski Topology

Before studying Zariski topology, we first introduce *affine spaces*.

Definition — Affine Space.

A set $\mathbb{A}$ is called an *affine space* over a field F if it is a principal $(F^n, +)$-set, i.e., there is a free and transitive action of the additive group $(F^n, +)$ on $\mathbb{A}$:

$$+ : \mathbb{A} \times F^n \to \mathbb{A}, \quad (P, \vec{v}) \mapsto P + \vec{v}$$

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 $\mathrm{SO}(2)$ 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 — Topology.

Let X be a set. A *topology* on X is a collection τ of subsets of X such that:

(a) the empty set $\emptyset$ and the whole set X are in τ;

(b) the union of any collection of sets in τ is also in τ;

(c) 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 — Topology (Closed Set Version).

Let X be a set. A *topology* on X is a collection τ of subsets of X such that:

(a) the empty set $\emptyset$ and the whole set X are in τ;

(b) the intersection of any collection of sets in τ is also in τ;

(c) 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 — Zariski Topology.

Let $\mathbb{A}$ be an affine space over a field F. The *Zariski topology* on $\mathbb{A}$ is defined by taking the closed sets to be the zero loci of sets of polynomials in $F[x_1, \cdots, x_n]$. More precisely, for any set of polynomials $S \subseteq F[x_1, \cdots, x_n]$, 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 $A \in F^{n^2} \simeq \mathbb{A}_F^{n^2}$ can be viewed as a point in the affine space $\mathbb{A}_F^{n^2}$ over F. Then the set of all diagonalisable endomorphisms with distinct eigenvalues forms a dense open subset of $\mathrm{End}(V) = F^{n^2}$ 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 $F[x_1, \cdots, x_{n^2}]$.

References

[1] А.Г. *Why in the proof of* $A \cdot \operatorname{adj}(A) = \det(A) \cdot I_n$ *entries not on the diagonal are zero?* 2015. URL: https://math.stackexchange.com/q/1404250 (cited on page 74).