Honors in Linear and Abstract Algebra I

Lecture Notes for
MATH 2131

Department of Mathematics
Hong Kong University of Science and Technology

January 9, 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 $\mathbb{F}_p=\mathbb{Z}/p\mathbb{Z}=\{0,1,2,\dots,p-1\}$ 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 a special type of homomorphisms called endomorphism.

Definition 1.17 — Endomorphism.

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

Several maps can form a set as below.

Definition 1.18 — 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 1.19 — Endomorphism Ring.

Given an abelian group $(G,+)$, the *endomorphism ring* of G, denoted by $\operatorname{End}(G)$, 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:

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

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

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 1.20 — 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

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

> **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 $\mathrm{End}(V)$, while the last three axioms corresponds to the properties of ring homomorphism from F to $\mathrm{End}(V)$. 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 $\mathrm{Map}(X, F)$ or F^X, forms a linear space over F with the following operations:

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

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 $\mathrm{Map}_{\mathrm{fin}}(X, F)$ 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 1.21 — 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 1.22 — Linear Combination.
A *linear combination* of vectors $v_1,v_2,\dots,v_n$ in a linear space V over F is any vector of the form

$$\alpha^1v_1+\alpha^2v_2+\cdots+\alpha^nv_n,$$

where $\alpha^1,\alpha^2,\dots,\alpha^n$ 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 1.23 — Linear Span.

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

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

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

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

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

Definition 1.24 — 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., $\operatorname{span} S = V$.

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 $\operatorname{span}(S)$ is the same as $\operatorname{span}\{\vec{e}_1, \vec{e}_2\}$. 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 $\operatorname{span}(T) = F^3$.

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., $\operatorname{span}(S) = F[x]$. 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 1.25 — 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 $\operatorname{span}(S') \subset \operatorname{span}(S) = V$.

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 1.26 — 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 1.27 — Weakly Linear Independence.
The subspaces $W_1, W_2, \dots, W_n$ 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 $i = 1, 2, \dots, n$, 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 1.28 — 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 $i = 1, 2, \dots, n$, 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 1.29 — 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 1.30 — Sum of Linear Subspaces.

The *sum* of the linear subspaces $W_1, W_2, \dots, W_n$ 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.,

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

Equivalently, the sum can be expressed as

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

Definition 1.31 — Internal Direct Sum of Linear Subspaces.

The *internal direct sum* of the linear subspaces $W_1, W_2, \dots, W_n$ 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 $i = 1, 2, \dots, n$ 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 $i = 1, 2, \dots, n$,

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

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 $V \cong W_1 \oplus W_2 \oplus \cdots \oplus W_n$.

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 $X = \{\mathsf{true}, \mathsf{false}\}$, 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 $\mathrm{Map}(X, F)$ is the set of F-valued functions on X and $F[X]$ is the set of finitely-supported F-valued functions on X. Both $\mathrm{Map}(X, F)$ 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 $T^* : \mathrm{Map}(Y, F) \to \mathrm{Map}(X, F)$ via the formula $T^* f = f T$. Show that $1_X^* = 1_{\mathrm{Map}(X,F)}$ and $(TS)^* = S^* T^*$.

(d) Can we get a natural map from $\mathrm{Map}(X, F)$ to $\mathrm{Map}(Y, F)$ 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 and Linear Combinations

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. □

2.2. 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 2.2.1. 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 2.2.2. 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. □

2.3. 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 2.3.1. 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 2.3.2. 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 2.3.1. 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.

- **Inverse property**: 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 2.3.1.** 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.

2.4. 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 2.4.1. 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 $S^{-1}\colon F^m \to V$ that is also a linear map. Then the composition $TS^{-1}\colon F^m \to F^n$ is also a linear isomorphism. Mutatis mutandis for the opposite direction. Therefore, it suffices to show that if there exists a linear isomorphism $L\colon F^m \to F^n$, 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:

$$\begin{tikzcd}
V \arrow[r, leftrightarrow, "T"] \arrow[d, leftrightarrow, "S"'] & F^n \\
F^m \arrow[ur, leftrightarrow, "TS^{-1}"'] &
\end{tikzcd}$$

> **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 $V \longleftrightarrow W$ to denote an isomorphism.

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

Exercise 2.4.1. 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 $\phi_S\colon \mathbb{F}^k \to V$ 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 2.4.2. 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 $T\colon F^n \to V$ 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 2.4.1, the linear map $\phi_S\colon F^k \to V$ 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 2.4.3. $\mathrm{dim}(V_1 + V_2) \leq \mathrm{dim}(V_1) + \mathrm{dim}(V_2)$ 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 $\dim(V_1 + V_2) \le |S_1| + |S_2| = \dim(V_1) + \dim(V_2)$, 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:

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

2.5. Matrices

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

$$\begin{aligned} F^n &\to F^m \\ \vec{x} &\mapsto A\vec{x} \end{aligned}$$

is a linear map over F.

Proposition 2.5.1. 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 $\vec{e}_1, \vec{e}_2, \dots, \vec{e}_n$:

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

Then, using the linearity of T, we have

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

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

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

Example 2.5.1. 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 2.11 — 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:

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

Definition 2.12 — 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:

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

Definition 2.13 — Rank.
The *rank* of an $m \times n$ matrix A over F is the dimension of its column space:

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

Definition 2.14 — Nullity.
The *nullity* of an $m \times n$ matrix A over F is the dimension of its null space:

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

Definition 2.15 — 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.

2.6. 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 $ST : F^n \to F^k$.

$$\begin{tikzcd} F^n \arrow[r, "A"] \arrow[rr, bend left, "ST"] \arrow[rr, bend right, "BA"'] & F^m \arrow[r, "B"] & F^k \end{tikzcd}$$

Proposition 2.6.1. 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 $x^1, x^2, \dots, x^n$, 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.

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

2.7. 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 2.7.1. 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 2.7.1. 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 2.7.2. 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.

2.8. 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 2.8.1. 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:

$$\begin{tikzcd}
F^n \arrow[r, "A" description] \arrow[d, leftrightarrow, "Q" description] & F^m \arrow[d, leftrightarrow, "P" description] \\
F^n \arrow[r, "N" description] & F^m
\end{tikzcd}$$

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 2.8.2. 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) $\operatorname{rank}(A) = n = m$;

(4) the normal form of A is I_n.

Proof.

(1) $\implies$ (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) $\implies$ (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, $\operatorname{rank}(PAQ^{-1}) = n$. Moreover, as rank is invariant under multiplication by invertible matrices, we have $\operatorname{rank}(A) = \operatorname{rank}(PAQ^{-1}) = n$. Since PAQ^{-1} is an $m \times n$ invertible matrix, we must have $m = n$.

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

(4) $\implies$ (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 2.8.1. 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) $\operatorname{rank}(A) = n$;

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

Exercise 2.8.2. 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:

$$\begin{tikzcd}
V \arrow[r, "T" description] \arrow[d, leftrightarrow, "{[-]_{\mathcal{B}_V}}" description] \arrow[dd, leftrightarrow, bend right=60] & W \arrow[d, leftrightarrow, "{[-]_{\mathcal{B}_W}}" description] \arrow[dd, leftrightarrow, bend left=60] \\
F^n \arrow[r, "A" description] \arrow[d, leftrightarrow, "Q" description] & F^m \arrow[d, leftrightarrow, "P" description] \\
F^n \arrow[r, "N" description] & F^m
\end{tikzcd}$$

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.

2.9. Group Actions

Before studying quotient spaces, we introduce the concept of (left) group actions.

Definition 2.16 — 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 SO(2) on the set of points in the plane. Each element of 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.

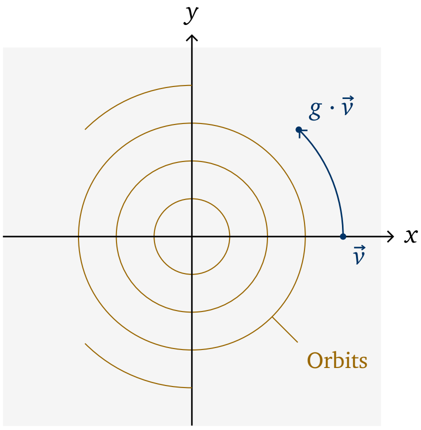


FIGURE 1. A group action of SO(2) on the plane.

Definition 2.17 — 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 2.18 — 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 2.19 — 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 $\pi \colon X \to X \,/\, G$ 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*.

2.10. 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:

$$\begin{aligned} W \times V &\to V \\ (w, v) &\mapsto v + w. \end{aligned}$$

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:

$$\begin{tikzcd} W \times V \arrow[r, hook, "\iota\times\mathrm{id}_V"] & V \times V \arrow[r, "+"] & V \end{tikzcd}$$

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 2.20 — 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, 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.

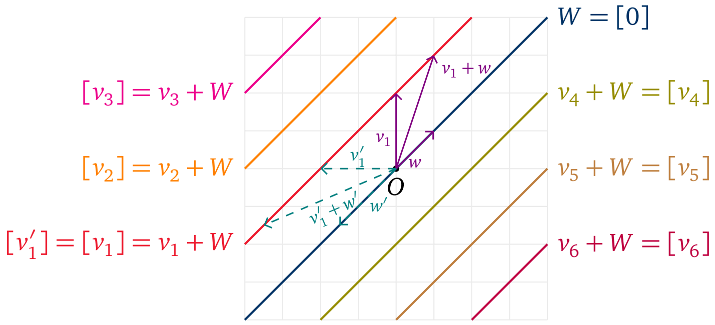


FIGURE 2. A graphical representation of the quotient space V / W.

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 2.21 — Linear Quotient Map.
The *linear quotient map* $\pi\colon V \to V / W$ 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 2.10.1. There is a unique linear structure on the quotient space V/W such that the quotient map $\pi\colon V \to V / W$ 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 $\pi\colon V\to V\,/\,W$, we can refer to Figure 3.

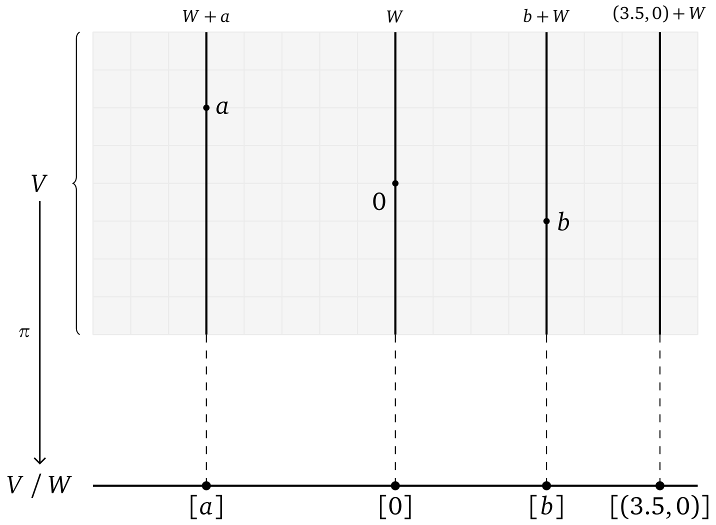


FIGURE 3. A graphical representation of the quotient map $\pi\colon V\to V\,/\,W$.

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

Proposition 2.10.2. 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:

$$\begin{tikzcd}
V \arrow[r, two heads] \arrow[d, leftrightarrow, "{[-]_{\mathcal{B}_V}}"'] & W \\
F^n \arrow[ur, dashed, two heads, "\phi"] &
\end{tikzcd}$$

Here, the map $\phi\colon F^n\to W$ 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:

$$\begin{tikzcd}
W \arrow[r, hook, "\iota"] & V \arrow[r, two heads, "\pi"] \arrow[d, leftrightarrow, "{[-]_{\mathcal{B}_V}}"'] & V\,/\,W \\
& F^n \arrow[ur, dashed, two heads, "\phi"] &
\end{tikzcd}$$

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. □

2.11. Universal Properties of 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 2.11.1 — 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:

$$\begin{tikzcd}
S \arrow[r, hook, "\iota"] \arrow[dr, "\phi"'] & V \arrow[d, dashed, "\widetilde{\phi}"] \\
& Z
\end{tikzcd}$$

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

Proposition 2.11.2 — 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:

$$\begin{tikzcd}
V \arrow[r, two heads, "\pi"] \arrow[dr, "\phi"'] & V\,/\,W \arrow[d, dashed, "\widetilde{\phi}"] \\
& Z
\end{tikzcd}$$

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 2.22 — 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 2.23 — 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:

$$\begin{tikzcd}
V \arrow[rr, bend left, "T"] & \ker(T) \arrow[l, hook', "\iota"] \arrow[r, "0"] & W \\
& Z \arrow[ul, "\phi"] \arrow[u, dashed, "\widetilde{\phi}"] \arrow[ur, "0"'] &
\end{tikzcd}$$

$$\begin{tikzcd}
V \arrow[rr, bend left, "T"] \arrow[r, "0"] \arrow[dr, "0"'] & \mathrm{coker}(T) \arrow[d, dashed, "\widetilde{\phi}"] & W \arrow[l, two heads, "\pi"'] \arrow[dl, "\phi"] \\
& Z &
\end{tikzcd}$$

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

2.12. Exact Sequences

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

Definition 2.24 — 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:

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

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

Example 2.12.1. Consider the following *short exact sequence* of linear spaces:

$$\begin{tikzcd} 0 \arrow[r] & V_1 \arrow[r, "\iota_1"] & V \arrow[r, "\pi_2"] & V_2 \arrow[r] & 0 \end{tikzcd}$$

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

- the exactness at V_1 implies that $\{0_{V_1}\} = \mathsf{im}(0) = \mathsf{ker}(\iota_1)$, thus ι_1 is injective.
- the exactness at V implies that $\mathsf{im}(\iota_1) = \mathsf{ker}(\pi_2)$, thus $V_1 \cong \mathsf{im}(\iota_1) \subseteq V$.
- the exactness at V_2 implies that $\mathsf{im}(\pi_2) = \mathsf{ker}(0) = V_2$, 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 $\iota_2 \colon V_2 \to V$ such that $\pi_2 \circ \iota_2 = \mathsf{id}_{V_2}$. This is because V_2 has a minimal spanning set. Thus, for each element in the minimal spanning set of V_2, we can choose one representative in V and define the map on the minimal spanning set. Then we can extend it to the whole space.
- ι_1 has a left inverse, or a retraction, i.e., there exists a linear map $\pi_1 \colon V \to V_1$ such that $\pi_1 \circ \iota_1 = \mathsf{id}_{V_1}$. 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:

$$\begin{tikzcd} 0 \arrow[r] & V_1 \arrow[r, hook, "\iota_1"] & V \arrow[l, bend right, "\pi_1"'] \arrow[r, two heads, "\pi_2"] & V_2 \arrow[l, bend left, "\iota_2"] \arrow[r] & 0 \end{tikzcd}$$

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

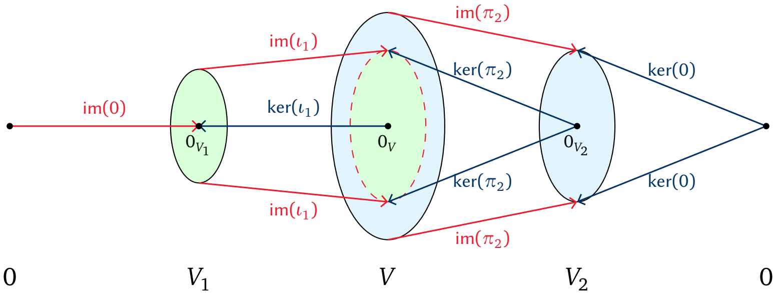


FIGURE 4. An Euler diagram illustrating a short exact sequence of linear spaces.

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

- $\pi_1 \circ \iota_1 = \mathsf{id}_{V_1}$ because π_1 is a left inverse of ι_1.
- $\pi_2 \circ \iota_2 = \mathsf{id}_{V_2}$ because π_2 is a right inverse of ι_2.
- $\pi_2 \circ \iota_1 = 0$ because $\mathsf{im}(\iota_1) = \mathsf{ker}(\pi_2)$.
- $\pi_1 \circ \iota_2 = 0$ because $\mathsf{im}(\iota_2) = \mathsf{ker}(\pi_1)$.
- $\iota_1 \circ \pi_1 + \iota_2 \circ \pi_2 = \mathsf{id}_V$ because for all $v \in V$, we have $v = (v - \iota_2(\pi_2(v))) + \iota_2(\pi_2(v))$ where $v - \iota_2(\pi_2(v)) \in \mathsf{im}(\iota_1)$ and $\iota_2(\pi_2(v)) \in \mathsf{im}(\iota_2)$. Also, $\mathsf{im}(\iota_1) \cap \mathsf{im}(\iota_2) = \{0_V\}$.

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

Proposition 2.12.1. 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 $\iota_2 \colon s \mapsto \iota_2(s)$ 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:

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

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

$$\begin{tikzcd} 0 \arrow[r] & V_1 \cap V_2 \arrow[r, hook, "\iota"] & V_1 \arrow[r, two heads, "\pi"] & (V_1 + V_2)\,/\,V_2 \arrow[r] & 0 \end{tikzcd}$$

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 4.3 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 $T\colon V \to W$ between finite-dimensional linear spaces, we have:

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

Proof. Consider the following short exact sequence:

$$\begin{tikzcd} 0 \arrow[r] & \ker(T) \arrow[r, hook, "\iota"] & V \arrow[r, two heads, "T"] & \operatorname{im}(T) \arrow[r] & 0 \end{tikzcd}$$

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

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

Proof. Consider the following short exact sequence:

$$\begin{tikzcd} 0 \arrow[r] & \operatorname{im}(T) \arrow[r, hook, "\iota"] & W \arrow[r, two heads, "\pi"] & \operatorname{coker}(T) \arrow[r] & 0 \end{tikzcd}$$

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

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

Proof. Consider the following short exact sequence:

$$\begin{tikzcd} 0 \arrow[r] & \ker(T) \arrow[r, hook, "\iota"] & V \arrow[r, two heads, "\pi"] & \operatorname{coim}(T) \arrow[r] & 0 \end{tikzcd}$$

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

2.13. Canonical Form of Linear Maps

We have already known that any linear map $T\colon V \to W$ 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:

$$\begin{tikzcd}
0 \arrow[d] & 0 \\
\ker(T) \arrow[d, hook] & \operatorname{coker}(T) \arrow[u] \arrow[d, bend left, "s_2"] \\
V \arrow[r, "T"] \arrow[d, two heads] \arrow[dr, two heads, "\overline{T}"] & W \arrow[u, two heads] \\
\operatorname{coim}(T) \arrow[u, bend left, "s_1"] \arrow[r, leftrightarrow, "T'"] \arrow[d] & \operatorname{im}(T) \arrow[u, hook] \\
0 & 0 \arrow[u]
\end{tikzcd}$$

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 $\operatorname{coim}(T)$ and from W to $\operatorname{coker}(T)$ respectively. Thus, we can decompose V and W into $V = \operatorname{im}(s_1) \oplus \ker(T)$ and $W = \operatorname{im}(s_2) \oplus \operatorname{im}(T)$ respectively. Then, with respect to these decompositions, the linear map $T\colon V \to W$ can be represented as:

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

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

$$\begin{tikzcd}
\operatorname{im}(s_1) \arrow[r, leftrightarrow, dashed, "\widetilde{T}"] \arrow[d, leftrightarrow, "s_1"] & \operatorname{im}(T) \\
\operatorname{coim}(T) \arrow[ur, leftrightarrow, "T'"] &
\end{tikzcd}$$

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:

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

2.14. 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 $\bar{f} : V \to \mathrm{im}(f)$ 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

$$\begin{tikzcd} 0 \arrow[r] & V_1 \arrow[r, "\iota_1"] & V \arrow[r, "\pi_2"] & V_2 \arrow[r] & 0 \end{tikzcd}$$

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

$$\begin{tikzcd} 0 & V_1 \arrow[l] & V \arrow[l, "\pi_1"'] & V_2 \arrow[l, "\iota_2"'] & 0 \arrow[l] \end{tikzcd}$$

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.

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 Spaces. 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:

$$\begin{tikzcd}
X \arrow[r, hook, "\iota_X"] \arrow[dr, "\phi"'] & F[X] \arrow[d, dashed, "\widetilde{\phi}"] \\
 & Z
\end{tikzcd}$$

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:

$$\begin{tikzcd}
 & X \arrow[dl, hook', "\iota_X"'] \arrow[dr, hook, "\iota'_X"] & \\
F[X] \arrow[rr, dashed, "\psi"] & & F'[X]
\end{tikzcd}$$

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

$$\begin{tikzcd}
X \arrow[r, "T"] \arrow[d, hook, "\iota_X"] & Y \arrow[d, hook, "\iota_Y"] \\
F[X] \arrow[r, dashed, "T_*"] & F[Y]
\end{tikzcd}$$

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 $T_*\colon F[S] \to F[T]$:

$$\begin{tikzcd}
\mathbf{Set} \arrow[r, "F{[-]}"] & \mathbf{Vec}_F
\end{tikzcd}$$

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:

$$\begin{tikzcd}
* \arrow[r, "a"] \arrow[rr, bend left, "ab"] & * \arrow[r, "b"] & *
\end{tikzcd}$$

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: $A\colon n \to m$. If we have a $k \times m$ matrix B that sends m to k, then we have the composition $BA\colon n \to k$. Note that $I_n\colon n \to n$ is the identity, which is not unique, there can be I_m and I_k. We have

$$\begin{tikzcd}
n \arrow["I_n"', loop left] \arrow[r, "A"] & m \arrow["I_m", loop right] \arrow[d, "B"] \\
& k
\end{tikzcd}$$

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

$$\begin{tikzcd}
n \arrow[r, "C"] \arrow[rrr, bend left=40, "(AB)C"] \arrow[rr, bend right=30, "BC"'] \arrow[rrr, bend right=40, "A(BC)"'] & m \arrow[r, "B"] \arrow[rr, bend left=30, "AB"] & k \arrow[r, "A"] & l
\end{tikzcd}$$

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

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

$$\begin{tikzcd}
& \text{Categories} \\
\text{Monoids} \arrow[r] & \text{Small Categories} \arrow[u] \\
\text{Groups} \arrow[u] \arrow[r] & \text{Groupoids} \arrow[u]
\end{tikzcd}$$

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 Categories.

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:

$$\begin{tikzcd}
\mathcal{C} \times_{s,t} \mathcal{C} \arrow[r, "\pi_1"] \arrow[d, "\pi_2"] \arrow[dr, phantom, "\ulcorner", very near start] & \mathcal{C} \arrow[d, two heads, "t"] \\
\mathcal{C} \arrow[r, two heads, "s"] & \mathcal{C}_0
\end{tikzcd}$$

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:

$$\begin{tikzcd}
t(f) & s(f) \arrow[l, "f"'] & t(g) & s(g) \arrow[l, "g"'] & t(f) & s(g) \arrow[l, dashed, "f \circ g"']
\end{tikzcd}$$

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:

$$\begin{tikzcd}
t(f) \arrow[loop left, "\mathrm{id}_{t(f)}"] & s(f) \arrow[l, "f"'] & t(f) & s(f) \arrow[l, "f"'] & t(f) & s(f) \arrow[l, "f"'] \arrow[loop right, "\mathrm{id}_{s(f)}"]
\end{tikzcd}$$

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:

$$\begin{tikzcd}
t(f) & s(f)=t(g) \arrow[l, "f"'] & s(g)=t(h) \arrow[l, "g"'] \arrow[ll, bend left, "fg"] & s(h) \arrow[l, "h"'] \arrow[ll, bend right, "gh"'] \arrow[lll, bend right, "f(gh)"'] \arrow[lll, bend left, "(fg)h"]
\end{tikzcd}$$

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 $f : X \to Y$.

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 $\mathrm{Hom}(X,Y)$ 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:

$$\begin{tikzcd}
\mathrm{Hom}(Y,Z)\times\mathrm{Hom}(X,Y) \arrow[r] & \mathrm{Hom}(X,Z) \\
(Z \xleftarrow{f} Y, Y \xleftarrow{g} X) \arrow[r, mapsto] & (Z \xleftarrow{fg} X)
\end{tikzcd}$$

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

Definition 3.2 — 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.

If we allow the collection of objects $\mathcal{C}_0$ and the collection of morphisms $\mathcal{C}_1$ to be *proper classes*, a larger collection than set, instead of sets, then we obtain the definition of a *category*.

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:

$$\begin{tikzcd}
& X \arrow[dl, "f"'] \arrow[dr, "f'"] & \\
Y \arrow[rr, "g"] & & Y'
\end{tikzcd}$$

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:

$$\begin{tikzcd}
& V \arrow[dl, "f"'] \arrow[dr, "f'"] & \\
Z \arrow[rr, "g"] & & Z'
\end{tikzcd}$$

Definition 3.3 — 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 3.4 — 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 $\pi\colon V \to V \,/\, W$ is an initial object and the zero map $0\colon V \to 0$ 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 vector space $\{0\}$ is both an initial object and a terminal object, hence it is a *zero object*.

3.4. Products and Coproducts

Definition 3.5 — 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:

$$\begin{tikzcd}
& Z \arrow[ld, "f_X"'] \arrow[d, dashed, "f"] \arrow[rd, "f_Y"] & \\
X & X \prod Y \arrow[l, "\pi_X"] \arrow[r, "\pi_Y"'] & Y
\end{tikzcd}$$

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:

$$\begin{tikzcd}
& Z \arrow[ld, "f_X"'] \arrow[rd, "f_Y"] \arrow[dd, "f"] & \\
X & & Y \\
& Z' \arrow[lu, "f'_X"] \arrow[ru, "f'_Y"'] &
\end{tikzcd}$$

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 3.6 — 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:

$$\begin{tikzcd}
X \arrow[r, "\iota_X"] \arrow[dr, "f_X"'] & X \coprod Y \arrow[d, dashed, "f"] & Y \arrow[l, "\iota_Y"'] \arrow[dl, "f_Y"] \\
& Z &
\end{tikzcd}$$

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:

$$\begin{tikzcd}
& Z \arrow[dd, "f"] & \\
X \arrow[ur, "f_X"] \arrow[dr, "f'_X"'] & & Y \arrow[ul, "f_Y"'] \arrow[dl, "f'_Y"] \\
& Z' &
\end{tikzcd}$$

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: $V \prod W \cong V \coprod W \cong V \oplus W$. In this case, we will say the *biproduct* of V and W is $V \oplus W$.

Definition 3.7 — 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.

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:

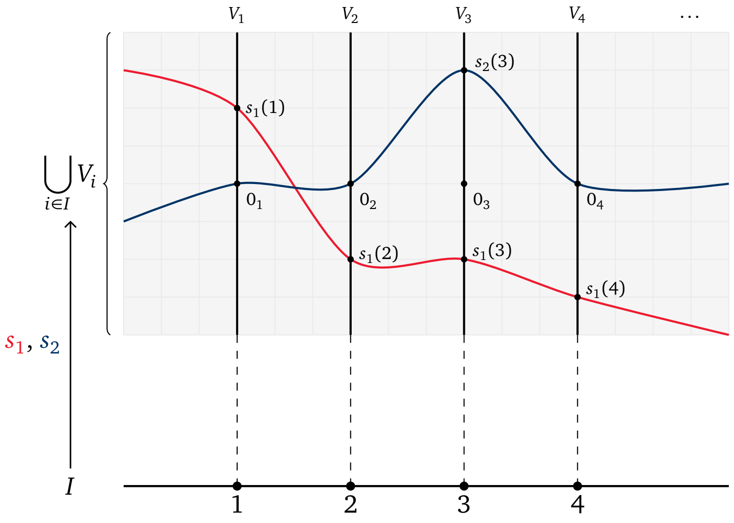
$$\begin{tikzcd}
X_i & \prod X_i \arrow[l, "f_i"'] \\
 & Z \arrow[ul, "\pi_i"] \arrow[u, dashed, "f"']
\end{tikzcd}
\qquad
\begin{tikzcd}
X_i \arrow[r, "\iota_i"] \arrow[dr, "f_i"'] & \coprod X_i \arrow[d, dashed, "f"] \\
 & Z
\end{tikzcd}$$

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 diagram:

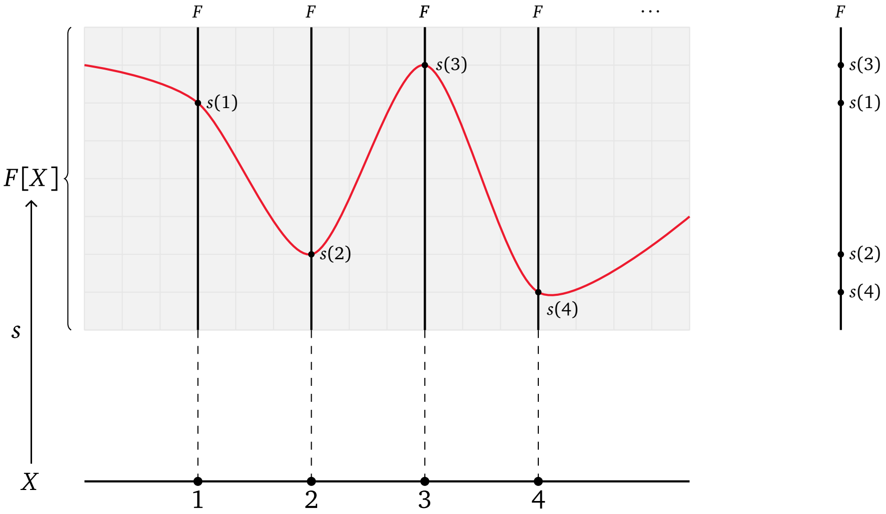


Remark. The right sections s_1 and s_2 are two elements in the product $\prod V_i$. Note that s_2 is likely to be "finitely supported" since it is zero in almost all components shown in the diagram. However, if I is an infinite set, then s_2 may not be finitely supported since there may be infinitely 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 $s\colon I \to \bigcup V_i$, 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 diagrams:



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

CHAPTER 4

Determinants

4.1. 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 4.1** — 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 $\mathbf{Vec}_F^1$:

$$\begin{tikzcd}
\mathbf{Vec}_F^n \arrow[r, "\det"] & \mathbf{Vec}_F^1 \\
V_1 \arrow[d, dash] & \det(V_1) \arrow[d, dash] \\
f \arrow[d] \arrow[r, mapsto] & \det(f) \arrow[d] \\
V_2 & \det(V_2)
\end{tikzcd}$$

Here, for a linear map $f : V_1 \to V_2$, the induced map $\det(f)\colon \det(V_1) \to \det(V_2)$ 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 $g\colon V_2 \to V_3$, 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:

$$\begin{tikzcd}
F^n \arrow[r, "A'"] \arrow[dd, bend right=60, "P"'] & F^n \arrow[dd, bend left=60, "P"] \\
V \arrow[u, "{[-]_{\mathcal{B}'}}"] \arrow[d, "{[-]_{\mathcal{B}}}"'] \arrow[r, "f"] & V \arrow[u, "{[-]_{\mathcal{B}'}}"'] \arrow[d, "{[-]_{\mathcal{B}}}"] \\
F^n \arrow[r, "A"] & F^n
\end{tikzcd}$$

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.

4.2. 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 4.2 — 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 4.2.1. 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 4.2.2. 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 4.3 — 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.

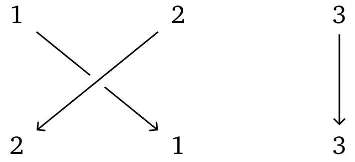
Example 4.2.3. The permutation group S_2 has two elements: the identity permutation 1 and the transposition σ_1 defined by $\sigma_1(1) = 2$ and $\sigma_1(2) = 1$.

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:

$$\sigma_i(j) = \begin{cases} j+1, & j = i \\ j-1, & j = i+1 \\ j, & \text{otherwise} \end{cases} = (i, i+1)$$

Example 4.2.4. The generator σ_1 of S_3 can be represented by the following diagram:



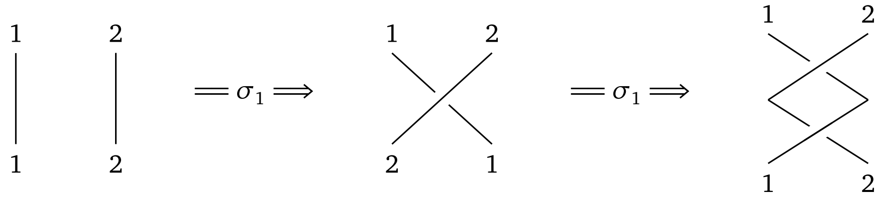
It can also be written as $\sigma_1 = (12)$ or $(12)(3)$ or $\begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix}$. Moreover, we have a cycle with 3 elements denoted as (123) defined by the $\begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}$. Then the presentation of S_3 is:

$$S_3 = \langle \sigma_1, \sigma_2 \mid \sigma_1^2 = 1, \sigma_2^2 = 1, \sigma_1\sigma_2\sigma_1 = \sigma_2\sigma_1\sigma_2 \rangle$$

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:

$$\begin{tikzcd} 1 \arrow[r] & A_n \arrow[r, hook] & S_n \arrow[r, two heads, "\mathrm{sgn}"] & \mathbb{Z}/2\mathbb{Z} \arrow[r] & 1 \end{tikzcd}$$

where A_n is the *alternating group* on n elements, i.e., the subgroup of S_n consisting of all even permutations, and $\mathrm{sgn} : S_n \to \mathbb{Z}/2\mathbb{Z} = \{\pm 1\}$, the *sign homomorphism*, is the unique group homomorphism such that $\mathrm{sgn}(\sigma_i) = -1$ for all $1 \leq i \leq n-1$. Note that $\ker(\mathrm{sgn}) = A_n$ and $\mathrm{im}(\mathrm{sgn}) = \mathbb{Z}/2\mathbb{Z}$.

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:

- $\mathrm{sgn}(1) = 1$;
- $\mathrm{sgn}(\sigma\tau) = \mathrm{sgn}(\sigma) \cdot \mathrm{sgn}(\tau)$ for all $\sigma, \tau \in S_n$.

4.3. 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, we have the following commutative diagram:

$$\begin{tikzcd} V^n \arrow[r, "\sigma"] \arrow[d, "\phi"] & V^n \arrow[d, "\phi"] \\ \Lambda^n V \arrow[r, dashed, "\widetilde{\sigma}"] & \Lambda^n V \end{tikzcd}$$

The induced map $\widetilde{\sigma} \colon \Lambda^n V \to \Lambda^n V$ is defined by

$$\widetilde{\sigma}(v_1 \wedge v_2 \wedge \cdots \wedge v_n) := v_{\sigma(1)} \wedge v_{\sigma(2)} \wedge \cdots \wedge v_{\sigma(n)}.$$

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

$$\operatorname{rank}(B) + \operatorname{rank}(ABC) \geq \operatorname{rank}(AB) + \operatorname{rank}(BC)$$

Solution. We consider the following diagram:

$$\begin{tikzcd}
0 \arrow[r] & \operatorname{col}(BC) \arrow[r, hook, red, "C"] \arrow[d, two heads, "A"] \arrow[drr, bend left, orange] \arrow[drr, bend right=60, purple] & \operatorname{col}(B) \arrow[r, two heads, "\pi_1"] \arrow[d, two heads, "A"] \arrow[dr, two heads, teal] & \operatorname{col}(B)\,/\,\operatorname{col}(BC) \arrow[r] \arrow[d, dashed, two heads, "\phi"] & 0 \\
0 \arrow[r] & \operatorname{col}(ABC) \arrow[r, hook, red, "C"] & \operatorname{col}(AB) \arrow[r, two heads, "\pi_2"] & \operatorname{col}(AB)\,/\,\operatorname{col}(ABC) \arrow[r] & 0
\end{tikzcd}$$

We denote the injective map with red color and the surjective map with blue color. Notice that there is a surjective map from $\operatorname{col}(B)$ to $\operatorname{col}(AB)\,/\,\operatorname{col}(ABC)$ 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 $\operatorname{col}(BC)$ to $\operatorname{col}(AB)\,/\,\operatorname{col}(ABC)$ is zero. If the map is zero, then we can construct a unique surjective map ϕ from $\operatorname{col}(B)\,/\,\operatorname{col}(BC)$ to $\operatorname{col}(AB)\,/\,\operatorname{col}(ABC)$ due to the universal property of quotient space.

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

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

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

$$\begin{gathered}
\mathrm{dim}(\mathrm{col}(B)\,/\,\mathrm{col}(BC)) \geq \mathrm{dim}(\mathrm{col}(AB)\,/\,\mathrm{col}(ABC)) \\
\mathrm{dim}(\mathrm{col}(B)) - \mathrm{dim}(\mathrm{col}(BC)) \geq \mathrm{dim}(\mathrm{col}(AB)) - \mathrm{dim}(\mathrm{col}(ABC)) \\
\mathrm{dim}(\mathrm{col}(B)) + \mathrm{dim}(\mathrm{col}(ABC)) \geq \mathrm{dim}(\mathrm{col}(AB)) + \mathrm{dim}(\mathrm{col}(BC)) \\
\mathrm{rank}(B) + \mathrm{rank}(ABC) \geq \mathrm{rank}(AB) + \mathrm{rank}(BC)
\end{gathered}$$

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

$$n = \mathrm{dim}(I_n) \geq \mathrm{rank}(A) \geq \mathrm{rank}(A^2) \geq \cdots$$

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 $k > n$.