Honors in Linear and Abstract Algebra I

Lecture Notes for
MATH 2131

Department of Mathematics
Hong Kong University of Science and Technology

January 4, 2026

Copyright Notice

Contents

Preface

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

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

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

CHAPTER 1

Linear Spaces

1.1. Introduction

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

1.2. Operations and Structures

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

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

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

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

Example 1.2.1. Common examples of unary operations include:

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

and

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

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

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

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

and

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

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

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

Some common algebraic structures include:

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

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

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

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

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

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

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

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

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

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

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

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

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

1.3. Homomorphisms

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

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

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

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

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

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

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

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

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

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

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

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

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

Definition 1.16 — Ring Homomorphism.

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

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

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

On top of homomorphisms, we have 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 $\mathrm{End}(G)$, is the set of all endomorphisms from G to itself, equipped with the pointwise addition and composition of functions as the two binary operations. The two operations are defined as follows:

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

The identity element for the addition operation is the zero map $0 : G \to G$ 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 $\mathrm{id}_G : G \to G$ defined by $\mathrm{id}_G(a) = a$ for all $a \in G$.

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

1.4. Linear Spaces

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

Definition 1.20 — 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

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

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

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

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

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

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

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

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

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

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

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

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

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 $\mathsf{Map}_{\mathsf{fin}}(X, F)$, 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.

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

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 $v_1, v_2, \ldots, v_n$ in a linear space V over F is any vector of the form

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

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

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

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

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

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

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

Then it is natural to ask: the union of any collection of linear subspaces of a linear space V over F is also a linear subspace of V? The answer is no in general. However, if we perform "completion", or

technically taking the *linear span*, we can get a linear subspace again and it is called the *sum* of those linear subspaces.

Definition 1.23 — 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 — Weak Linear Independence.
The subspaces $W_1, W_2, \ldots, W_n$ of a linear space V over F are *weakly linearly independent* if the only way to express the zero vector 0_V as a linear combination of vectors from these subspaces is the trivial way, i.e., if

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

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

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

Definition 1.28 — 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 $i = 1, 2, \ldots, n$, then we must have $\alpha^1 = \alpha^2 = \cdots = \alpha^n = 0_F$.

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

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

Definition 1.29 — 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

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

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

1.7. Exercise

Problem 1.1. On the logic set $X = \{\mathrm{true}, \mathrm{false}\}$, 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 $\mathrm{Map}_{\mathrm{fin}}(X, F)$ is the set of finitely-supported F-valued functions on X. Both $\mathrm{Map}(X, F)$ and $\mathrm{Map}_{\mathrm{fin}}(X, F)$ are linear spaces over the field F.
Let $T : X \to Y$ be a set map and $T_* : \mathrm{Map}_{\mathrm{fin}}(X, F) \to \mathrm{Map}_{\mathrm{fin}}(Y, F)$ 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 $(1_X)_* = 1_{\mathrm{Map}_{\mathrm{fin}}(X,F)}$ 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 $\mathrm{Map}_{\mathrm{fin}}(Y, F)$ to $\mathrm{Map}_{\mathrm{fin}}(X, F)$ for any set map $T : X \to Y$ between two infinite sets X and Y?

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 $V \cong W$.

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.

(⇒) : 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 $T(v) = T(T^{-1}(w)) = w$.

(⇐) : 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 $\dim(V)$.

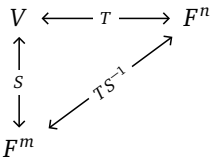
If a linear space is not finite-dimensional, it is called *infinite-dimensional*. Then we have to show that the dimension is well-defined.

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

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

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



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

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:

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

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

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

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

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

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 $\text{Hom}(F^n, F^m)$ and $\text{Mat}_{m\times n}(F)$. 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 $\text{Mat}_{m\times n}(F)$, the corresponding linear map $T_A : F^n \to F^m$ is given by

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

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

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

Then, using the linearity of T, we have

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

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 $T_A : F^n \to F^m$ 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 $D : F[x] \to F[x]$ 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:
$$\text{null}(A) = \{\vec{x} \in F^n \mid A\vec{x} = 0\}.$$

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

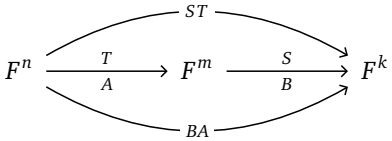
Definition 2.13 — Rank.
The *rank* of an $m \times n$ matrix A over F is the dimension of its column space:
$$\text{rank}(A) = \dim(\text{col}(A)).$$

Definition 2.14 — Nullity.
The *nullity* of an $m \times n$ matrix A over F is the dimension of its null space:
$$\text{nullity}(A) = \dim(\text{null}(A)).$$

Definition 2.15 — 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 $ST : F^n \to F^k$.



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 $x^1, x^2, \ldots, x^n$, we have
$$\vec{x} = x^1\vec{e}_1 + x^2\vec{e}_2 + \cdots + x^n\vec{e}_n.$$
The j-th column of BA is given by
$$(ST)(\vec{e}_j) = S(T(\vec{e}_j)) = S(\vec{a}_j) = B\vec{a}_j = B(A\vec{e}_j) = (BA)(\vec{e}_j).$$
Therefore, the standard matrix of ST is BA. □

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

The matrix multiplication BA can be computed as follows.

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

2.7. 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:

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

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

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

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

(4) the normal form of A is I_n.

Proof.

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

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

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

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

Exercise 2.8.1. 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) $\mathrm{rank}(A) = n$;

(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 $AC = I_m$;

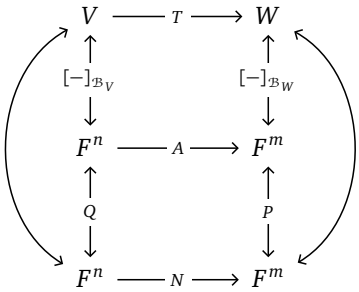
(2) A is surjective;

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

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

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

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



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

2.9. 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

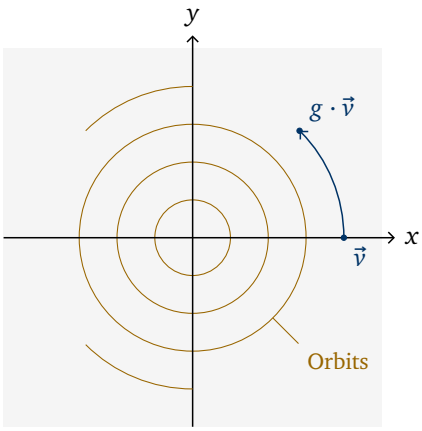
- Compatibility: $g \cdot (h \cdot x) = (gh) \cdot x$;
- Unital: $e \cdot x = x$, where e is the identity element of G.

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 follows:



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 $\pi : X \to X \,/\, G$ 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:

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

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

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

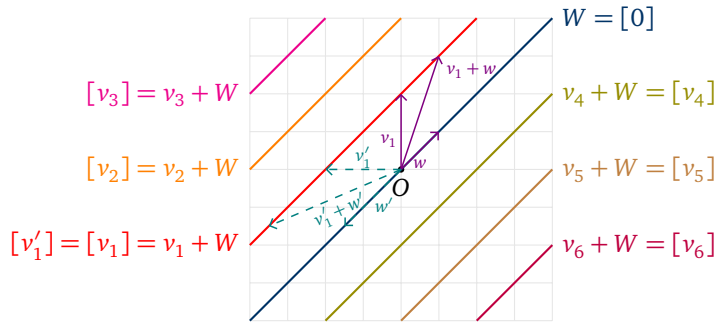
The orbits of this group action are the sets of the form $v + W = \{v + w \mid w \in W\}$ for each $v \in V$. These orbits partition the linear space V into disjoint subsets. Algebraically, such subsets are called *cosets* of W in V. The cosets can be written as $[v]$ or $\overline{v}$ for simplicity.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

Definition 2.20 — 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 with the following diagram:



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.

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* $\pi : V \to V / W$ 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 $\pi : V \to V / W$ is a linear map.

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

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

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

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

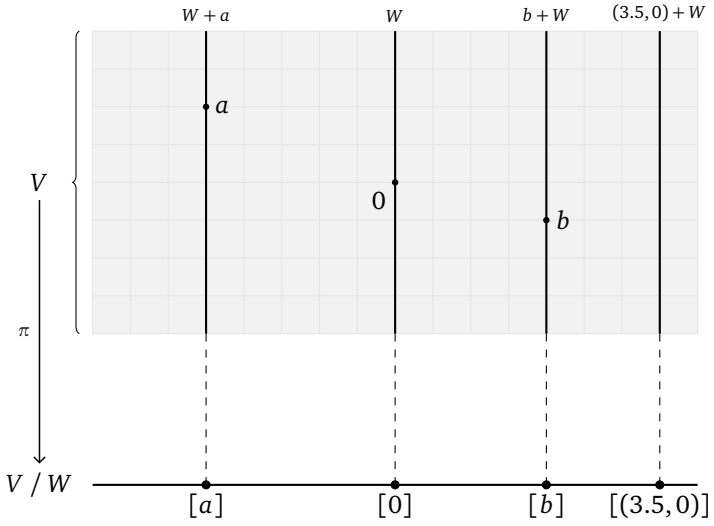
We need to verify that this definition is well-defined. Suppose $[v_1] = [v_1']$ and $[v_2] = [v_2']$, i.e., $v_1' - v_1 \in W$ and $v_2' - v_2 \in W$. Then,

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

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

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

If we want to visualise the graphical representation of the quotient space V / W and the quotient map $\pi : V \to V / W$, the following diagram may help:



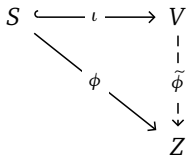
2.11. Universal Property

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:



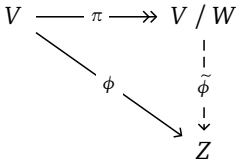
Proof. If such a linear map $\widetilde{\phi}$ exists, then for any $s \in S$, we must have $\widetilde{\phi} \circ \iota(s) = \phi(s)$, which suggests that $\widetilde{\phi}$ is defined by extending ϕ linearly to the whole space V. Specifically, for any $v \in V$, we can express v as a linear combination 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:



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*.

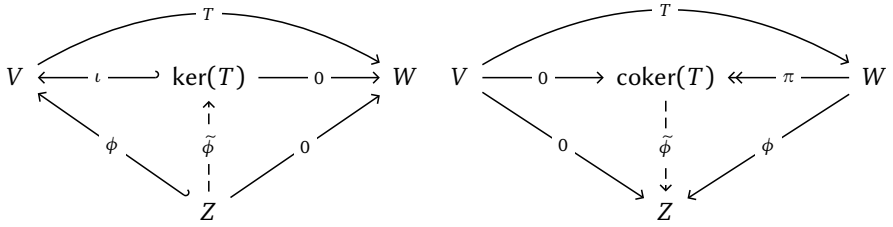
Definition 2.22 — Cokernel.
The *cokernel* of a linear map $T : V \to W$ is the quotient space of W by the image of T:

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

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 the following universal property for kernel and cokernel with the following commutative diagrams:



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

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

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

(c) 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 $[-]_S$.