Table of contents

Preface

Welcome to **Advanced Linear Algebra Notes**, a comprehensive guide to spectral theory and optimization techniques in linear algebra.

About This Book

This book provides a rigorous yet accessible introduction to advanced topics in linear algebra, with a focus on:

- **Spectral Theory**: Eigenvalues, eigenvectors, and matrix decompositions
- **Optimization**: Convex optimization and its applications
- **Computational Methods**: Algorithms for numerical linear algebra

How to Use This Book

Each chapter contains:

- **Definitions** with precise mathematical statements
- **Theorems** with complete proofs
- **Examples** illustrating key concepts
- **Exercises** for practice

Prerequisites

Readers should be familiar with:

- Basic linear algebra (vectors, matrices, linear transformations)
- Calculus (derivatives, integrals)
- Basic proof techniques

Acknowledgments

This book was created using Quarto with a custom LaTeX template for beautiful mathematical typesetting.

Chapter 0

Preliminary

This chapter covers the essential background needed for the study of linear algebra. Feel free to skip sections that you are already familiar with. In addition to the following knowledge, basic knowledge of logic and proof techniques is recommended, including:

- Direct proof
- Proof by contradiction
- Proof by induction
- Proof by contrapositive
- Proof by cases (exhaustion)
- Constructive proof (constructive existence)
- Non-constructive existence proof
- Uniqueness proof
- Proof of equivalence (iff proof)
- Disproof by counterexample
- Pigeonhole principle

0.1 Sets and Notation

We begin by defining the standard sets of numbers used throughout this book.

Definition 0.1.1 (Common Number Sets)**.** We use the following notation for common sets of numbers:

- $\mathbb{N}$: The set of natural numbers $\{1, 2, 3, ...\}$.
- $\mathbb{Z}$: The set of integers $\{..., -2, -1, 0, 1, 2, ...\}$.
- $\mathbb{Q}$: The set of rational numbers.
- $\mathbb{R}$: The set of real numbers.
- $\mathbb{C}$: The set of complex numbers.

Definition 0.1.2 (Cartesian Product)**.** Let A and B be sets. The **Cartesian product** of A and B, denoted $A \times B$, is the set of all ordered pairs:

$$A \times B = \{(a, b) : a \in A, b \in B\}$$

More generally, for sets $A_1, A_2, \dots, A_n$, the Cartesian product is:

$$A_1 \times A_2 \times \cdots \times A_n = \{(a_1, a_2, \dots, a_n) : a_i \in A_i \text{ for } i = 1, 2, \dots, n\}$$

Definition 0.1.3 (n-Tuple)**.** An ***n*-tuple** is an ordered list of n elements. For a set A, the set of all n-tuples with entries from A is denoted:

$$A^n = \underbrace{A \times A \times \cdots \times A}_{n \text{ times}} = \{(a_1, a_2, \dots, a_n) : a_i \in A \text{ for } i = 1, 2, \dots, n\}$$

Example (n-Tuples). Here are examples of n-tuples from various sets:

1. $\mathbb{R}^2$: The Cartesian plane. Elements include $(0, 0)$, $(1, 2)$, $(-3, \pi)$, $(\sqrt{2}, -5.7)$.
2. $\mathbb{R}^3$: 3-dimensional space. Elements include $(1, 0, 0)$, $(1, 2, 3)$, $(-1, \pi, e)$.
3. $\mathbb{Z}^2$: Pairs of integers. Elements include $(0, 0)$, $(1, -2)$, $(5, 7)$. Note that $(\frac{1}{2}, 3) \notin \mathbb{Z}^2$.
4. $\mathbb{C}^2$: Pairs of complex numbers. Elements include $(1 + i, 2 - 3i)$, $(i, 0)$, $(3, 4)$.
5. $\{0, 1\}^3$: Binary 3-tuples. This set has exactly 8 elements:

$$\{0, 1\}^3 = \{(0, 0, 0), (0, 0, 1), (0, 1, 0), (0, 1, 1), (1, 0, 0), (1, 0, 1), (1, 1, 0), (1, 1, 1)\}$$

In general, if A is a finite set with $|A| = k$ elements, then A^n has k^n elements.

0.2 Functions

Definition 0.2.1 (Function)**.** A **function** f from a set A to a set B, denoted $f : A \to B$, is a rule that assigns to each element $a \in A$ exactly one element $f(a) \in B$. The set A is called the **domain** and B is called the **codomain**.

Definition 0.2.2 (Injective, Surjective, Bijective)**.** Let $f : A \to B$ be a function.

- f is **injective** (one-to-one) if $f(a_1) = f(a_2)$ implies $a_1 = a_2$.
- f is **surjective** (onto) if for every $b \in B$, there exists $a \in A$ such that $f(a) = b$.
- f is **bijective** if it is both injective and surjective.

$$\int_{-\infty}^{\infty} e^{-x^2}\, dx = \sqrt{\pi}$$

Example (Injective, Surjective, and Bijective Functions). Consider the following functions:

1. $f : \mathbb{R} \to \mathbb{R}$, $f(x) = 2x + 1$ is **bijective**.
   - Injective: If $2x_1 + 1 = 2x_2 + 1$, then $x_1 = x_2$.

- Surjective: For any $y \in \mathbb{R}$, we have $f\left(\frac{y-1}{2}\right) = y$.

2. $g : \mathbb{R} \to \mathbb{R}$, $g(x) = x^2$ is **neither injective nor surjective**.
   - Not injective: $g(1) = g(-1) = 1$, but $1 \neq -1$.
   - Not surjective: There is no $x \in \mathbb{R}$ such that $g(x) = -1$.
3. $h : \mathbb{R} \to [0, \infty)$, $h(x) = x^2$ is **surjective but not injective**.
   - Not injective: $h(1) = h(-1) = 1$.
   - Surjective: For any $y \geq 0$, we have $h(\sqrt{y}) = y$.
4. $k : [0, \infty) \to \mathbb{R}$, $k(x) = x^2$ is **injective but not surjective**.
   - Injective: If $x_1^2 = x_2^2$ with $x_1, x_2 \geq 0$, then $x_1 = x_2$.
   - Not surjective: There is no $x \geq 0$ such that $k(x) = -1$.

0.3 Polynomials

Definition 0.3.1 (Polynomial)**.** A **polynomial** over a field F in the indeterminate x is an expression of the form:

$$p(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0 = \sum_{k=0}^{n} a_k x^k$$

where $a_0, a_1, \ldots, a_n \in F$ are called the **coefficients** of $p(x)$.

Definition 0.3.2 (Degree)**.** Let $p(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0$ be a nonzero polynomial with $a_n \neq 0$. The **degree** of $p(x)$, denoted $\deg p(x)$, is n. The coefficient a_n is called the **leading coefficient**.

By convention, the zero polynomial $p(x) = 0$ has degree $-\infty$.

Definition 0.3.3 (Polynomial Ring)**.** The set of all polynomials over a field F is denoted $F[x]$:

$$F[x] = \left\{ \sum_{k=0}^{n} a_k x^k : n \in \mathbb{N},\ a_k \in F \right\}$$

This set forms a **ring** under polynomial addition and multiplication. (You can omit what ring means here.)

Example (Elements of the Polynomial Ring). The following are elements of $\mathbb{R}[x]$:

$$3x^2 - 2x + 1, \quad x^5 + \pi x^3 - \sqrt{2}, \quad 7, \quad 0$$

Note that constant polynomials (including 0) are also polynomials.

0.3.1 Polynomial Spaces

In linear algebra, we often work with polynomials of bounded degree.

Definition 0.3.4 (Polynomials of Degree at Most n)**.** Let F be a field and $n \in \mathbb{N}$. The set of all polynomials over F of degree at most n is denoted:

$$F[x]_{\leq n} = \{p(x) \in F[x] : \deg p(x) \leq n\}$$

Equivalently:

$$F[x]_{\leq n} = \{a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0 : a_k \in F\}$$

Example (Polynomials of Degree at Most 2). The set $\mathbb{R}[x]_{\leq 2}$ consists of all polynomials of degree at most 2:

$$\mathbb{R}[x]_{\leq 2} = \{ax^2 + bx + c : a, b, c \in \mathbb{R}\}$$

Examples include $3x^2 - 2x + 1$, $5x + 7$, and -4.

Definition 0.3.5 (Polynomials of Degree Exactly n)**.** We use $F[x]_{=n}$ to denote polynomials of degree **exactly** n:

$$F[x]_{=n} = \{p(x) \in F[x] : \deg p(x) = n\}$$

Equivalently:

$$F[x]_{=n} = \{a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0 : a_n \neq 0,\ a_k \in F\}$$

Remark. Note the distinction:

- $F[x]_{\leq n}$ contains polynomials of degree $\leq n$ (including the zero polynomial)
- $F[x]_{=n}$ contains polynomials of degree **exactly** n (so $a_n \neq 0$)

Thus $F[x]_{\leq n} \supsetneq F[x]_{=n}$, and in fact $F[x]_{\leq n} = F[x]_{=n} \cup F[x]_{\leq n-1}$.

Example (Comparing Notations).

- $\mathbb{R}[x]_{\leq 2} = \{ax^2 + bx + c : a, b, c \in \mathbb{R}\}$ includes $x^2 + 1$, $3x - 2$, and 5
- $\mathbb{R}[x]_{=2} = \{ax^2 + bx + c : a \neq 0,\ a, b, c \in \mathbb{R}\}$ includes $x^2 + 1$ but **not** $3x - 2$ or 5

0.4 Fields

Definition 0.4.1 (Field)**.** A set F is called a **field** if we can define two binary operations addition $+$, and multiplication $\cdot$

$$+ : F \times F \to F, \quad \cdot : F \times F \to F$$

such that:

(F1) $a + b = b + a$

(F2) $(a + b) + c = a + (b + c)$

(F3) There exists $0 \in F$ such that $a + 0 = a$.

(F4) For every $a \in F$, there exists $-a \in F$ such that $a + (-a) = 0$.

(F5) $(a \cdot b) \cdot c = a \cdot (b \cdot c)$

(F6) $a \cdot b = b \cdot a$

(F7) There exists $1 \in F$ such that $1 \neq 0$ and $a \cdot 1 = a$.

(F8) For every $a \in F \setminus \{0\}$, there exists $a^{-1} \in F$ such that $a \cdot a^{-1} = 1$.

(F9) $a \cdot (b + c) = (a \cdot b) + (a \cdot c)$

Remark. We can shorten "F is a field with operations $+$ and $\cdot$" to "$(F, +, \cdot)$ is a field". Moreover when the context of $+$ and $\cdot$ is clear, we simply say "F is a field", which we will do most of the time in this book.

Example (Examples of Fields).

1. $\mathbb{R}$ is a field.
2. $\mathbb{C}$ is a field.
3. $\mathbb{Q} = \left\{ \frac{p}{q} : p, q \in \mathbb{Z},\ \gcd(p, q) = 1 \right\}$ is a field.
4. $\mathbb{Z}$ is **not** a field since **(F8)** fails.
5. $M_n(\mathbb{R})$ is **not** a field since **(F6)** fails.

Example (Finite Fields). Let $\mathbb{F}_n$ be the finite field with n elements. Then $(\mathbb{F}_2, +, \cdot)$ is defined to be:

$$F_2 = \{0, 1\}, \quad \begin{array}{c|cc} + & 0 & 1 \\ \hline 0 & 0 & 1 \\ 1 & 1 & 0 \end{array}, \quad \begin{array}{c|cc} \cdot & 0 & 1 \\ \hline 0 & 0 & 0 \\ 1 & 0 & 1 \end{array}.$$

Then one can verify that $(\mathbb{F}_2, +, \cdot)$ is a field.

0.5 Matrices

Definition 0.5.1 (Matrix)**.** An $m \times n$ **matrix** over a field F is a rectangular array of elements from F arranged in m rows and n columns:

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}.$$

Remark. $M_{m\times n}(F)$ is the set of all matrices of size $m \times n$ and have entries in F. In the above definition, we can write $\mathbf{A} \in M_{m\times n}(F)$.

Moreover, when $m = n$, we abbreviate $M_{m\times n}(F)$ to $M_n(F)$.

Example (Matrices). Here are some examples of matrices:

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \in M_2(\mathbb{R}), \quad \begin{bmatrix} 1 & 0 & -1 \\ 2 & 5 & 3 \end{bmatrix} \in M_{2\times 3}(\mathbb{R}), \quad \begin{bmatrix} i & 1+i \\ 0 & 2-i \end{bmatrix} \in M_2(\mathbb{C}).$$

0.5.1 Special Matrices

Definition 0.5.2 (Zero Matrix)**.** The **zero matrix** $\in M_{m\times n}(F)$ is the matrix with all entries equal to zero:

$$= \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}.$$

Definition 0.5.3 (Identity Matrix)**.** The **identity matrix** $\mathbf{I}_n \in M_n(F)$ is the $n \times n$ square matrix with 1s on the diagonal and 0s elsewhere:

$$\mathbf{I}_n = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}.$$

When the size is clear from context, we simply write $\mathbf{I}$.

Definition 0.5.4 (Transpose)**.** Let $\mathbf{A} = (a_{ij}) \in M_{m\times n}(F)$. The **transpose** of $\mathbf{A}$, denoted $\mathbf{A}^\top$, is the $n \times m$ matrix obtained by swapping rows and columns:

$$(\mathbf{A}^\top)_{ij} = a_{ji}.$$

Visually, the rows of $\mathbf{A}$ become the columns of $\mathbf{A}^\top$:

$$\begin{bmatrix} a & b & c \\ d & e & f \end{bmatrix}^\top = \begin{bmatrix} a & d \\ b & e \\ c & f \end{bmatrix}.$$

Definition 0.5.5 (Symmetric Matrix)**.** A square matrix $\mathbf{A} \in M_n(F)$ is **symmetric** if $\mathbf{A} = \mathbf{A}^\top$. This means it is symmetric across its main diagonal:

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{12} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{nn} \end{bmatrix}.$$

Definition 0.5.6 (Diagonal Matrix)**.** A square matrix $\mathbf{A} \in M_n(F)$ is **diagonal** if all entries off the main diagonal are zero:

$$\mathbf{A} = \begin{bmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_n \end{bmatrix}.$$

We often write $\mathbf{A} = \operatorname{diag}(d_1, d_2, \dots, d_n)$.

Definition 0.5.7 (Upper Triangular Matrix)**.** A square matrix $\mathbf{A} \in M_n(F)$ is **upper triangular** if all entries below the main diagonal are zero:

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ 0 & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{nn} \end{bmatrix}.$$

Example (Zero and Identity Matrices).

$${}_{2\times 3} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad \mathbf{I}_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \mathbf{I}_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Example (Examples of Special Matrices).

- **Transpose:** $\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}^{\top} = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}$.
- **Symmetric:** $\begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 5 \\ 3 & 5 & 6 \end{bmatrix}$ is symmetric since $a_{ij} = a_{ji}$.
- **Diagonal:** $\operatorname{diag}(3, -1, 0) = \begin{bmatrix} 3 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$.
- **Upper Triangular:** $\begin{bmatrix} 1 & 2 & 3 \\ 0 & 4 & 5 \\ 0 & 0 & 6 \end{bmatrix}$.

0.5.2 Matrix Addition

Definition 0.5.8 (Matrix Addition)**.** Let $\mathbf{A} = (a_{ij})$ and $\mathbf{B} = (b_{ij})$ be matrices in $M_{m\times n}(F)$. Their **sum** $\mathbf{A} + \mathbf{B}$ is the matrix $\mathbb{C} = (c_{ij}) \in M_{m\times n}(F)$ where:

$$c_{ij} = a_{ij} + b_{ij}.$$

Definition 0.5.9 (Additive Inverse)**.** The **additive inverse** (or **negation**) of a matrix $\mathbf{A} = (a_{ij}) \in M_{m\times n}(F)$ is the matrix $-\mathbf{A} = (-a_{ij}) \in M_{m\times n}(F)$.

Example (Matrix Addition). Let $\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$ and $\mathbf{B} = \begin{bmatrix} 5 & 6 \\ 7 & 8 \end{bmatrix}$. Then:

$$\mathbf{A} + \mathbf{B} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} + \begin{bmatrix} 5 & 6 \\ 7 & 8 \end{bmatrix} = \begin{bmatrix} 1+5 & 2+6 \\ 3+7 & 4+8 \end{bmatrix} = \begin{bmatrix} 6 & 8 \\ 10 & 12 \end{bmatrix}.$$

Theorem 0.5.1 (Properties of Matrix Addition). Let $\mathbf{A}, \mathbf{B}, \mathbb{C} \in M_{m\times n}(F)$. Then:

1. $\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$ (Commutativity)
2. $(\mathbf{A} + \mathbf{B}) + \mathbb{C} = \mathbf{A} + (\mathbf{B} + \mathbb{C})$ (Associativity)
3. $\mathbf{A} + \ = \mathbf{A}$ (Additive identity)
4. $\mathbf{A} + (-\mathbf{A}) = \ $ (Additive inverse)

Proof. We prove commutativity. Let $\mathbf{A} = (a_{ij})$ and $\mathbf{B} = (b_{ij})$. Then:

$$(\mathbf{A} + \mathbf{B})_{ij} = a_{ij} + b_{ij} = b_{ij} + a_{ij} = (\mathbf{B} + \mathbf{A})_{ij}$$

where the second equality uses commutativity of addition in the underlying field. Since

> this holds for all i, j, we have $\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$.
>
> The remaining properties follow similarly. ■

0.5.3 Matrix Multiplication

Definition 0.5.10 (Matrix Multiplication)**.** Let $\mathbf{A} = (a_{ij}) \in M_{m\times n}(F)$ and $\mathbf{B} = (b_{jk}) \in M_{n\times p}(F)$. Their **product AB** is the matrix $\mathbb{C} = (c_{ik}) \in M_{m\times p}(F)$ where:

$$c_{ik} = \sum_{j=1}^{n} a_{ij} b_{jk}$$

In other words, the (i, k)-entry of **AB** is the dot product of the i-th row of **A** with the k-th column of **B**.

Example (Matrix Multiplication). Let $\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \in M_2(\mathbb{R})$ and $\mathbf{B} = \begin{bmatrix} 5 & 6 \\ 7 & 8 \end{bmatrix} \in M_2(\mathbb{R})$. We compute **AB** entry by entry:

$$\begin{aligned}
(\mathbf{AB})_{11} &= (1)(5) + (2)(7) = 5 + 14 = 19 \\
(\mathbf{AB})_{12} &= (1)(6) + (2)(8) = 6 + 16 = 22 \\
(\mathbf{AB})_{21} &= (3)(5) + (4)(7) = 15 + 28 = 43 \\
(\mathbf{AB})_{22} &= (3)(6) + (4)(8) = 18 + 32 = 50
\end{aligned}$$

Therefore:

$$\mathbf{AB} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 5 & 6 \\ 7 & 8 \end{bmatrix} = \begin{bmatrix} 19 & 22 \\ 43 & 50 \end{bmatrix}$$

Example (Matrix Multiplication with Different Dimensions). Let $\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \in M_{2\times 3}(\mathbb{R})$ and $\mathbf{B} = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} \in M_{3\times 1}(\mathbb{R})$. The product $\mathbf{AB} \in M_{2\times 1}(\mathbb{R})$:

$$\mathbf{AB} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} = \begin{bmatrix} (1)(1) + (2)(0) + (3)(-1) \\ (4)(1) + (5)(0) + (6)(-1) \end{bmatrix} = \begin{bmatrix} 1 + 0 - 3 \\ 4 + 0 - 6 \end{bmatrix} = \begin{bmatrix} -2 \\ -2 \end{bmatrix}$$

Note that **BA** is **not defined** since the number of columns of **B** (which is 1) does not equal the number of rows of **A** (which is 2).

Theorem 0.5.2 (Properties of Matrix Multiplication). Let $\mathbf{A}, \mathbf{B}, \mathbb{C}$ be matrices of compatible dimensions. Then:

1. $(\mathbf{AB})\mathbb{C} = \mathbf{A}(\mathbf{B}\mathbb{C})$ (Associativity)

2. $\mathbf{A}(\mathbf{B}+\mathbb{C}) = \mathbf{AB} + \mathbf{A}\mathbb{C}$ (Left distributivity)
3. $(\mathbf{A}+\mathbf{B})\mathbb{C} = \mathbf{A}\mathbb{C} + \mathbf{B}\mathbb{C}$ (Right distributivity)
4. $\mathbf{AI} = \mathbf{IA} = \mathbf{A}$ (Multiplicative identity)

Proof. We prove associativity. Let $\mathbf{A}$ be $m \times n$, $\mathbf{B}$ be $n \times p$, and $\mathbb{C}$ be $p \times q$. For any entry (i, ℓ):

$$\begin{aligned} ((\mathbf{AB})\mathbb{C})_{i\ell} &= \sum_{k=1}^{p} (\mathbf{AB})_{ik} c_{k\ell} = \sum_{k=1}^{p} \left(\sum_{j=1}^{n} a_{ij} b_{jk} \right) c_{k\ell} \\ &= \sum_{k=1}^{p} \sum_{j=1}^{n} a_{ij} b_{jk} c_{k\ell} = \sum_{j=1}^{n} \sum_{k=1}^{p} a_{ij} b_{jk} c_{k\ell} \\ &= \sum_{j=1}^{n} a_{ij} \left(\sum_{k=1}^{p} b_{jk} c_{k\ell} \right) = \sum_{j=1}^{n} a_{ij} (\mathbf{B}\mathbb{C})_{j\ell} \\ &= (\mathbf{A}(\mathbf{B}\mathbb{C}))_{i\ell} \end{aligned}$$

■

Remark. Matrix multiplication is **not** commutative in general. That is, $\mathbf{AB} \neq \mathbf{BA}$ even when both products are defined.

0.5.4 Scalar Multiplication

Definition 0.5.11 (Scalar Multiplication)**.** Let $\mathbf{A} = (a_{ij}) \in M_{m\times n}(F)$ and $c \in F$ be a scalar. The **scalar multiple** $c\mathbf{A}$ is the matrix in $M_{m\times n}(F)$ with entries (ca_{ij}).

0.5.5 Trace of a Matrix

Definition 0.5.12 (Trace)**.** Let $\mathbf{A} = (a_{ij}) \in M_n(F)$ be a square matrix. The **trace** of $\mathbf{A}$, denoted $\operatorname{tr}(\mathbf{A})$, is the sum of the entries on its main diagonal:

$$\operatorname{tr}(\mathbf{A}) = \sum_{i=1}^{n} a_{ii} = a_{11} + a_{22} + \cdots + a_{nn}.$$

Example (Trace). Let $\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$. Then $\operatorname{tr}(\mathbf{A}) = 1 + 5 + 9 = 15$.

0.5.6 Properties of Transpose and Trace

Theorem 0.5.3 (Properties of Transpose). Let $\mathbf{A}, \mathbf{B}$ be matrices of compatible sizes and $c \in F$. Then:

1. $(\mathbf{A}^\top)^\top = \mathbf{A}$
2. $(\mathbf{A}+\mathbf{B})^\top = \mathbf{A}^\top + \mathbf{B}^\top$
3. $(c\mathbf{A})^\top = c\mathbf{A}^\top$
4. $(\mathbf{A}\mathbf{B})^\top = \mathbf{B}^\top\mathbf{A}^\top$ (Reversal rule)

Proof. We prove the reversal rule $(\mathbf{A}\mathbf{B})^\top = \mathbf{B}^\top\mathbf{A}^\top$. Let $\mathbf{A} \in M_{m\times n}(F)$ and $\mathbf{B} \in M_{n\times p}(F)$.

$$((\mathbf{A}\mathbf{B})^\top)_{ij} = (\mathbf{A}\mathbf{B})_{ji} = \sum_{k=1}^{n} a_{jk}b_{ki}$$

Now consider $\mathbf{B}^\top\mathbf{A}^\top$. Its (i,j)-entry is:

$$(\mathbf{B}^\top\mathbf{A}^\top)_{ij} = \sum_{k=1}^{n}(\mathbf{B}^\top)_{ik}(\mathbf{A}^\top)_{kj} = \sum_{k=1}^{n} b_{ki}a_{jk} = \sum_{k=1}^{n} a_{jk}b_{ki}$$

Since the entries are equal for all i, j, we have $(\mathbf{A}\mathbf{B})^\top = \mathbf{B}^\top\mathbf{A}^\top$. ■

Theorem 0.5.4 (Properties of Trace). Let $\mathbf{A}, \mathbf{B} \in M_n(F)$ and $c \in F$. Then:

1. $\mathrm{tr}(\mathbf{A}+\mathbf{B}) = \mathrm{tr}(\mathbf{A}) + \mathrm{tr}(\mathbf{B})$
2. $\mathrm{tr}(c\mathbf{A}) = c\,\mathrm{tr}(\mathbf{A})$
3. $\mathrm{tr}(\mathbf{A}^\top) = \mathrm{tr}(\mathbf{A})$
4. $\mathrm{tr}(\mathbf{A}\mathbf{B}) = \mathrm{tr}(\mathbf{B}\mathbf{A})$ (Cyclic property)

Proof. We prove $\mathrm{tr}(\mathbf{A}\mathbf{B}) = \mathrm{tr}(\mathbf{B}\mathbf{A})$.

$$\mathrm{tr}(\mathbf{A}\mathbf{B}) = \sum_{i=1}^{n}(\mathbf{A}\mathbf{B})_{ii} = \sum_{i=1}^{n}\left(\sum_{j=1}^{n} a_{ij}b_{ji}\right)$$

Rearranging the summation:

$$\mathrm{tr}(\mathbf{A}\mathbf{B}) = \sum_{j=1}^{n}\sum_{i=1}^{n} b_{ji}a_{ij} = \sum_{j=1}^{n}(\mathbf{B}\mathbf{A})_{jj} = \mathrm{tr}(\mathbf{B}\mathbf{A}).$$

■

Chapter 1
Vector Spaces and Dimensions

In this chapter, we introduce the fundamental concept of a vector space and discuss its properties, including basis and dimension.

1.1 Vector Spaces

Why do we bother with a list of eight axioms? To a student first encountering Linear Algebra, this can feel like a bureaucratic exercise in "checking boxes." However, there is a profound beauty hidden in this formality.

In our preliminary chapter, we saw three very different worlds: the world of geometric arrows ($\mathbb{R}^n$), the world of functions ($F[x]$), and the world of matrices ($M_{m\times n}$). On the surface, an arrow is not a polynomial, and a polynomial is not a matrix. But if you squint, they all behave the same: you can add them, and you can scale them.

By defining a **Vector Space** through these axioms, we are choosing to ignore what the objects *are* and focus entirely on how they *act*. If we prove a theorem using only these eight axioms, that theorem becomes a "universal law" that applies to arrows, functions, and matrices all at once. This is the power of abstraction: solve the problem once, and you solve it for every universe that obeys these rules.

Definition 1.1.1 (Vector Space)**.** A set V over a field F with two operations: addition $+$ and scalar multiplication $\cdot$

$$+ : V \times V \to V, \quad \cdot : F \times V \to V$$

such that satisfy the following axioms for all $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3 \in V$ and $\alpha, \beta \in F$:

(VS1) $\mathbf{v}_1 + \mathbf{v}_2 = \mathbf{v}_2 + \mathbf{v}_1$

(VS2) $(\mathbf{v}_1 + \mathbf{v}_2) + \mathbf{v}_3 = \mathbf{v}_1 + (\mathbf{v}_2 + \mathbf{v}_3)$

(VS3) There exists $\mathbf{0} \in V$ such that $\mathbf{v}_1 + \mathbf{0} = \mathbf{v}_1$

(VS4) There exists $\mathbf{v}_1' \in V$ such that $\mathbf{v}_1 + \mathbf{v}_1' = \mathbf{0}$

(VS5) $\alpha(\mathbf{v}_1 + \mathbf{v}_2) = \alpha\mathbf{v}_1 + \alpha\mathbf{v}_2$

(VS6) $(\alpha + \beta)\mathbf{v}_1 = \alpha\mathbf{v}_1 + \beta\mathbf{v}_1$

(VS7) $\alpha(\beta\mathbf{v}_1) = (\alpha\beta)\mathbf{v}_1$

(VS8) $1\mathbf{v}_1 = \mathbf{v}_1$

Remark. Instead of writing $(V, +, \cdot)$ is a vector space over F, we usually simplify it to V is a vector space over F, or even just V is a vector space if the context is clear.

Remark. We sometimes will abuse the name and refer to $\mathbf{0} = \mathbf{0}_V \in V$ as the "zero vector" of the vector space V. Readers should not confuse $\mathbf{0}_V$ with the zero scalar $0 = 0_F \in F$ from the field F.

Example (Examples of Vector Spaces).

1. Most defaultly, F^n with the usual operations of $+$ and $\cdot$ is a vector space over F.
2. Working with polynomials is common too: $F[x]$ with the usual operations of $+$ and $\cdot$ is a vector space over F.
3. Let $\mathcal{D}$ be any open interval. Let

$$C(\mathcal{D}) := \{f : \mathcal{D} \to \mathbb{R} : f \text{ is continuous}\}.$$

   Then $C(\mathcal{D})$ is a vector space over $\mathbb{R}$. The zero vector of this vector space is given by the zero polynomial $(x \mapsto 0)$.
4. Let F be any field and fix $n \in \mathbb{N}$. Then $M_n(F)$ is a vector space over F.
5. The set $V = \mathbb{R}_{>0}$ of positive real numbers forms a vector space over $F = \mathbb{R}$ under the following operations: for $x, y \in V$ and $\alpha \in F$, define

$$x \oplus y := xy, \quad \alpha \odot x := x^\alpha = e^{\alpha \log x}.$$

   Under these operations, the zero vector is $\mathbf{0} = 1$, and the additive inverse of x is x^{-1}.

You can check if addition and scalar multiplication make sense and follow all axioms of vector spaces.

Example ($\mathbb{Q}$ Adjoin $\sqrt{2}$). We verify that $\mathbb{Q}(\sqrt{2}) := \left\{a + b\sqrt{2} : a, b \in \mathbb{Q}\right\}$ is a vector space over $\mathbb{Q}$ by checking all eight axioms. Let $\mathbf{v}_1 = a_1 + b_1\sqrt{2}$, $\mathbf{v}_2 = a_2 + b_2\sqrt{2}$, $\mathbf{v}_3 = a_3 + b_3\sqrt{2} \in \mathbb{Q}(\sqrt{2})$ and $\alpha, \beta \in \mathbb{Q}$.

(VS1) $\mathbf{v}_1 + \mathbf{v}_2 = (a_1 + a_2) + (b_1 + b_2)\sqrt{2} = (a_2 + a_1) + (b_2 + b_1)\sqrt{2} = \mathbf{v}_2 + \mathbf{v}_1$.

(VS2) $(\mathbf{v}_1 + \mathbf{v}_2) + \mathbf{v}_3 = (a_1 + a_2 + a_3) + (b_1 + b_2 + b_3)\sqrt{2} = \mathbf{v}_1 + (\mathbf{v}_2 + \mathbf{v}_3)$.

(VS3) The element $\mathbf{0} = 0 + 0\sqrt{2} = 0 \in \mathbb{Q}(\sqrt{2})$ satisfies $\mathbf{v}_1 + \mathbf{0} = (a_1 + 0) + (b_1 + 0)\sqrt{2} = \mathbf{v}_1$.

(VS4) For $\mathbf{v}_1 = a_1 + b_1\sqrt{2}$, define $-\mathbf{v}_1 := (-a_1) + (-b_1)\sqrt{2} \in \mathbb{Q}(\sqrt{2})$. Then $\mathbf{v}_1 + (-\mathbf{v}_1) = (a_1 - a_1) + (b_1 - b_1)\sqrt{2} = \mathbf{0}$.

(VS5) $\alpha(\mathbf{v}_1+\mathbf{v}_2) = \alpha(a_1+a_2)+\alpha(b_1+b_2)\sqrt{2} = (\alpha a_1+\alpha a_2)+(\alpha b_1+\alpha b_2)\sqrt{2} = \alpha\mathbf{v}_1+\alpha\mathbf{v}_2$.

(VS6) $(\alpha+\beta)\mathbf{v}_1 = (\alpha+\beta)a_1+(\alpha+\beta)b_1\sqrt{2} = (\alpha a_1+\beta a_1)+(\alpha b_1+\beta b_1)\sqrt{2} = \alpha\mathbf{v}_1+\beta\mathbf{v}_1$.

(VS7) $\alpha(\beta\mathbf{v}_1) = \alpha(\beta a_1+\beta b_1\sqrt{2}) = \alpha\beta a_1+\alpha\beta b_1\sqrt{2} = (\alpha\beta)\mathbf{v}_1$.

(VS8) $1\cdot\mathbf{v}_1 = 1\cdot a_1+1\cdot b_1\sqrt{2} = a_1+b_1\sqrt{2} = \mathbf{v}_1$.

Thus $\mathbb{Q}(\sqrt{2})$ is a vector space over $\mathbb{Q}$.

Definition 1.1.2 (Vectors and Scalars)**.** Let V be a vector space over F. Then the elements of V are called **vectors**, and the elements of F are called **scalars**.

Additionally, $\mathbf{0}\in V$ is called the **zero vector**, and $(\mathbf{v}_1')$ in **(VS4)** is called the **inverse element** of $\mathbf{v}_1$.

Theorem 1.1.1 (Left Cancellation Law). Let V be a vector space over F, let $\mathbf{u},\mathbf{v}_1,\mathbf{v}_2\in V$. If $\mathbf{u}+\mathbf{v}_1=\mathbf{u}+\mathbf{v}_2$, then $\mathbf{v}_1=\mathbf{v}_2$.

Proof. Let $\mathbf{u}'$ be an inverse of $\mathbf{u}$.

$$\begin{aligned} \mathbf{u}+\mathbf{v}_1 &= \mathbf{u}+\mathbf{v}_2 && \text{(given)}\\ \mathbf{u}'+(\mathbf{u}+\mathbf{v}_1) &= \mathbf{u}'+(\mathbf{u}+\mathbf{v}_2) && \text{(add } \mathbf{u}' \text{ to both sides)}\\ (\mathbf{u}'+\mathbf{u})+\mathbf{v}_1 &= (\mathbf{u}'+\mathbf{u})+\mathbf{v}_2 && \text{(by VS2)}\\ \mathbf{0}+\mathbf{v}_1 &= \mathbf{0}+\mathbf{v}_2 && \text{(by VS4)}\\ \mathbf{v}_1 &= \mathbf{v}_2 && \text{(by VS3)} \end{aligned}$$

■

Theorem 1.1.2 (Right Cancellation Law). Let V be a vector space over F, let $\mathbf{u},\mathbf{v}_1,\mathbf{v}_2\in V$. If $\mathbf{v}_1+\mathbf{u}=\mathbf{v}_2+\mathbf{u}$, then $\mathbf{v}_1=\mathbf{v}_2$.

Proof. Since we have $\mathbf{u}+\mathbf{v}_1=\mathbf{u}+\mathbf{v}_2 \implies \mathbf{v}_1=\mathbf{v}_2$, applying **(VS1)** to it gives $\mathbf{v}_1+\mathbf{u}=\mathbf{v}_2+\mathbf{u} \implies \mathbf{v}_1=\mathbf{v}_2$. ■

Theorem 1.1.3 (Zero Vector is Unique). Let V be a vector space over F. The zero vector $\mathbf{0}\in V$ is unique.

Proof. Suppose that there are two vectors $\mathbf{0}_1,\mathbf{0}_2$.

$$\begin{aligned} \mathbf{0}_1+\mathbf{0}_2 &= \mathbf{0}_2+\mathbf{0}_1 && \text{(by VS1)}\\ \mathbf{0}_1 &= \mathbf{0}_2 && \text{(by VS3)} \end{aligned}$$

■

Theorem 1.1.4 (Additive Inverse is Unique). Let V be a vector space over F. Then for every $\mathbf{v} \in V$, its additive inverse described in **(VS4)**, which is $\mathbf{v}'$, is unique.

Proof. Let $\mathbf{v}_1', \mathbf{v}_2'$ both be inverses of $\mathbf{v}$. Then

$$\begin{aligned}
\mathbf{v}_1' &= \mathbf{v}_1' + \mathbf{0} && \text{(by VS3)} \\
&= \mathbf{v}_1' + (\mathbf{v} + \mathbf{v}_2') && \text{(by VS4)} \\
&= (\mathbf{v}_1' + \mathbf{v}) + \mathbf{v}_2' && \text{(by VS2)} \\
&= \mathbf{0} + \mathbf{v}_2' && \text{(by VS4)} \\
&= \mathbf{v}_2' && \text{(by VS3)}
\end{aligned}$$

■

Definition 1.1.3 (Notation of Additive Inverse)**.** The unique inverse of a vector $\mathbf{v} \in V$, for a vector space V over F, will be denoted as $-\mathbf{v}$.

Theorem 1.1.5 (Zero Scalar Annihilates). Let $\mathbf{v} \in V$ for a vector space V over F. Then $0 \cdot \mathbf{v} = \mathbf{0}$.

Proof. Let $\mathbf{w} = 0 \cdot \mathbf{v}$. We show that $\mathbf{w} = \mathbf{0}$.

$$\begin{aligned}
\mathbf{w} + \mathbf{w} &= 0 \cdot \mathbf{v} + 0 \cdot \mathbf{v} \\
&= (0 + 0) \cdot \mathbf{v} && \text{(by VS6)} \\
&= 0 \cdot \mathbf{v} && \text{(arithmetic in } F\text{)} \\
&= \mathbf{w} \\
(\mathbf{w} + \mathbf{w}) + (-\mathbf{w}) &= \mathbf{w} + (-\mathbf{w}) && \text{(add } -\mathbf{w} \text{ to both sides)} \\
\mathbf{w} + (\mathbf{w} + (-\mathbf{w})) &= \mathbf{0} && \text{(by VS2, VS4)} \\
\mathbf{w} + \mathbf{0} &= \mathbf{0} && \text{(by VS4)} \\
\mathbf{w} &= \mathbf{0} && \text{(by VS3)}
\end{aligned}$$

■

Theorem 1.1.6 (Negation as Scalar Multiplication). Let V be a vector space over F and $\mathbf{v} \in V$ be any vector. Then $-\mathbf{v} = (-1) \cdot \mathbf{v}$.

Proof. We show that $(-1) \cdot \mathbf{v}$ is an additive inverse of $\mathbf{v}$:

$$\begin{aligned}
\mathbf{v} + (-1) \cdot \mathbf{v} &= 1 \cdot \mathbf{v} + (-1) \cdot \mathbf{v} && \text{(by VS8)} \\
&= (1 + (-1)) \cdot \mathbf{v} && \text{(by VS6)} \\
&= 0 \cdot \mathbf{v} && \text{(arithmetic in } F\text{)} \\
&= \mathbf{0} && \text{(by @thm-zero-scalar-mult)}
\end{aligned}$$

Since $(-1) \cdot \mathbf{v}$ is an additive inverse of $\mathbf{v}$, by uniqueness we have $-\mathbf{v} = (-1) \cdot \mathbf{v}$. ■

Theorem 1.1.7 (Negative Scalar Distribution). Let V be a vector space over F. For any $\alpha \in F$ and $\mathbf{v} \in V$, we have $(-\alpha)\mathbf{v} = -(\alpha\mathbf{v})$.

Proof. We show that $(-\alpha)\mathbf{v}$ is an additive inverse of $\alpha\mathbf{v}$:

$$\begin{aligned}
\alpha\mathbf{v} + (-\alpha)\mathbf{v} &= (\alpha + (-\alpha))\mathbf{v} && \text{(by VS6)} \\
&= 0 \cdot \mathbf{v} && \text{(arithmetic in } F\text{)} \\
&= \mathbf{0} && \text{(by @thm-zero-scalar-mult)}
\end{aligned}$$

Since $(-\alpha)\mathbf{v}$ is an additive inverse of $\alpha\mathbf{v}$, by uniqueness we have $(-\alpha)\mathbf{v} = -(\alpha\mathbf{v})$. ■

Theorem 1.1.8 (Scalar Multiplication by Zero Vector). Let V be a vector space over F. For any scalar $\alpha \in F$, we have $\alpha\mathbf{0} = \mathbf{0}$.

Proof. Let $\mathbf{w} = \alpha\mathbf{0}$. We show that $\mathbf{w} = \mathbf{0}$.

$$\begin{aligned}
\mathbf{w} &= \alpha\mathbf{0} \\
&= \alpha(\mathbf{0} + \mathbf{0}) && \text{(by VS3)} \\
&= \alpha\mathbf{0} + \alpha\mathbf{0} && \text{(by VS5)} \\
&= \mathbf{w} + \mathbf{w} \\
\mathbf{w} + (-\mathbf{w}) &= (\mathbf{w} + \mathbf{w}) + (-\mathbf{w}) && \text{(add } -\mathbf{w} \text{ to both sides)} \\
\mathbf{0} &= \mathbf{w} + (\mathbf{w} + (-\mathbf{w})) && \text{(by VS4, VS2)} \\
\mathbf{0} &= \mathbf{w} + \mathbf{0} && \text{(by VS4)} \\
\mathbf{0} &= \mathbf{w} && \text{(by VS3)}
\end{aligned}$$

■

Theorem 1.1.9 (Zero Product Law). Let V be a vector space over F, $\mathbf{v} \in V$, and $\alpha \in F$. If $\alpha\mathbf{v} = \mathbf{0}$, then either $\alpha = 0$ or $\mathbf{v} = \mathbf{0}$.

Proof. Suppose $\alpha\mathbf{v} = \mathbf{0}$. We consider two cases:

- **Case $\alpha = 0$:** Then the statement holds.
- **Case $\alpha \neq 0$:** Since F is a field, α has a multiplicative inverse α^{-1}.

$$\begin{aligned} \alpha\mathbf{v} &= \mathbf{0} && \text{(given)} \\ \alpha^{-1}(\alpha\mathbf{v}) &= \alpha^{-1}\mathbf{0} && \text{(multiply by } \alpha^{-1}\text{)} \\ (\alpha^{-1}\alpha)\mathbf{v} &= \mathbf{0} && \text{(by VS7 and previous Theorem)} \\ 1 \cdot \mathbf{v} &= \mathbf{0} && \text{(field inverse property)} \\ \mathbf{v} &= \mathbf{0} && \text{(by VS8)} \end{aligned}$$

Thus, if $\alpha \neq 0$, we must have $\mathbf{v} = \mathbf{0}$.

■

1.2 Subspaces

Some vector spaces might be "too big" and be unwieldy to deal with. Therefore, it is often convenient to look at "smaller vector spaces" inside a given vector space. These are called **subspaces**.

Definition 1.2.1 (Subspace)**.** Let V be a vector space over F. A subset W of V is called a **subspace** of V, if W is also a vector space over F with addition and scalar multiplication inherited from V.

Remark. Let $(V, +, \cdot)$ be a vector space over F, and $W \subseteq V$. If $(W, +, \cdot)$ is also a vector space over F then W is a subspace of V.

Example (Trivial Subspaces). Given a vector space V, then

- V is a subspace of V itself;
- $\{\mathbf{0}\}$ is a subspace of V (also known as the "zero subspace");
- $\emptyset$ is **not** a subspace of V.

Theorem 1.2.1 (Subspace Test). Let V be a vector space over F and $W \subseteq V$. Then W is a subspace of V if and only if the following are satisfied:

- W is non-empty;
- W is closed under addition: $\mathbf{w}_1 + \mathbf{w}_2 \in W$ for all $\mathbf{w}_1, \mathbf{w}_2 \in W$;
- W is closed under scalar multiplication: $\alpha\mathbf{w} \in W$ for all $\alpha \in F$ and $\mathbf{w} \in W$.

Proof. ($\Rightarrow$) If W is a subspace, then W is a vector space, so it contains the zero vector (hence non-empty) and is closed under addition and scalar multiplication by definition.

($\Leftarrow$) Assume the three conditions hold. Since $W \subseteq V$, the operations on W are inherited from V, so associativity, commutativity, and distributivity are automatically satisfied. It remains to verify the existence of identity elements and inverses.

- **Zero vector:** Since $W \neq \emptyset$, there exists $\mathbf{w} \in W$. By closure under scalar multiplication, $0 \cdot \mathbf{w} = \mathbf{0} \in W$.
- **Additive inverse:** For any $\mathbf{w} \in W$, closure under scalar multiplication gives $(-1) \cdot \mathbf{w} = -\mathbf{w} \in W$.

Thus W is a vector space under the inherited operations, i.e., a subspace of V. ■

Remark. Note that from the proof, the first condition of the Subspace Test can be replaced by the requirement that W contains the zero vector of V.

Remark. In practice, to test whether a given subset of a vector space is a subspace, we use Theorem 1.2.1 instead of checking all eight vector space axioms from scratch, which is much less tedious. Moreover, some introductory textbooks use Theorem 1.2.1 as the definition of a subspace. While this is logically equivalent, it can be somewhat less intuitive than defining a subspace as a subset that is itself a vector space.

Example (Line as Subspace). $W = \{(x, y) \in \mathbb{R}^2 : x + 2y = 0\}$ is a subspace of $\mathbb{R}^2$. Indeed:

- **Zero vector:** $(0, 0) \in W$ since $0 + 2(0) = 0$.
- **Addition:** If $(x_1, y_1), (x_2, y_2) \in W$, then $(x_1 + x_2) + 2(y_1 + y_2) = (x_1 + 2y_1) + (x_2 + 2y_2) = 0 + 0 = 0$, so their sum is in W.
- **Scalar multiplication:** If $(x, y) \in W$ and $c \in \mathbb{R}$, then $(cx) + 2(cy) = c(x + 2y) = c(0) = 0$, so $c(x, y) \in W$.

Example (Translated Line is Not a Subspace). $W = \{(x, y) \in \mathbb{R}^2 : x + 2y = 3\}$ is **not** a subspace of $\mathbb{R}^2$. This is because it does not contain the zero vector: $0 + 2(0) = 0 \neq 3$, so $(0, 0) \notin W$.

Example (Polynomials of Bounded Degree). Let $F[x]$ be the vector space of all polynomials over a field F. For a fixed $n \in \mathbb{N}$, the set $F[x]_{\leq n}$ of polynomials of degree at most n is a subspace of $F[x]$.

- **Zero vector:** The zero polynomial has degree $-\infty$, so $0 \in F[x]_{\leq n}$.
- **Addition:** If $\deg p \leq n$ and $\deg q \leq n$, then $\deg(p + q) \leq \max(\deg p, \deg q) \leq n$.
- **Scalar multiplication:** If $\deg p \leq n$ and $c \in F$, then $\deg(cp) \leq \deg p \leq n$.

Example (Polynomials of Exact Degree). The set $F[x]_{=n}$ of polynomials of degree **exactly** n is **not** a subspace of $F[x]$.

- It does not contain the zero vector (since $\deg 0 = -\infty \neq n$).
- It is not closed under addition. For example, if $p(x) = x^n + x$ and $q(x) = -x^n$, both have degree n, but $p(x) + q(x) = x$, which has degree $1 \neq n$ (assuming $n > 1$).

Example (Differentiable Functions). Let $C(\mathbb{R}) = \{f : \mathbb{R} \to \mathbb{R} \mid f \text{ is continuous}\}$ be the set of all continuous functions on $\mathbb{R}$. We consider the set of differentiable functions:

$$D(\mathbb{R}) = \{f : \mathbb{R} \to \mathbb{R} \mid f \text{ is differentiable}\}$$

To verify that $D(\mathbb{R})$ is a subspace of $C(\mathbb{R})$:

- **Subset:** Since differentiability implies continuity, we have $D(\mathbb{R}) \subseteq C(\mathbb{R})$.
- **Zero vector:** The zero function $f(x) = 0$ is differentiable, so $\mathbf{0} \in D(\mathbb{R})$.
- **Closure under addition:** If $f, g \in D(\mathbb{R})$, then $f+g$ is differentiable with $(f+g)' = f' + g'$, so $f + g \in D(\mathbb{R})$.
- **Closure under scalar multiplication:** If $f \in D(\mathbb{R})$ and $c \in \mathbb{R}$, then cf is differentiable with $(cf)' = cf'$, so $cf \in D(\mathbb{R})$.

Thus, $D(\mathbb{R})$ is a subspace of $C(\mathbb{R})$.

Example (Symmetric Matrices). The set of all symmetric matrices in $M_n(F)$, denoted by $S_n(F) = \{\mathbf{A} \in M_n(F) : \mathbf{A} = \mathbf{A}^\top\}$, is a subspace of $M_n(F)$.

- **Zero vector:** The zero matrix $\mathbf{O}$ is symmetric since $\mathbf{O}^\top = \mathbf{O}$, so $\mathbf{O} \in S_n(F)$.
- **Addition:** If $\mathbf{A}, \mathbf{B} \in S_n(F)$, then $(\mathbf{A}+\mathbf{B})^\top = \mathbf{A}^\top + \mathbf{B}^\top = \mathbf{A}+\mathbf{B}$, so $\mathbf{A}+\mathbf{B} \in S_n(F)$.
- **Scalar multiplication:** If $\mathbf{A} \in S_n(F)$ and $c \in F$, then $(c\mathbf{A})^\top = c\mathbf{A}^\top = c\mathbf{A}$, so $c\mathbf{A} \in S_n(F)$.

Example (Trace-free Matrices). The set of all trace-free matrices in $M_n(F)$, denoted by $W = \{\mathbf{A} \in M_n(F) : \operatorname{tr}(\mathbf{A}) = 0\}$, is a subspace of $M_n(F)$.

- **Zero vector:** $\operatorname{tr}(\mathbf{O}) = 0 + 0 + \cdots + 0 = 0$, so $\mathbf{O} \in W$.
- **Addition:** If $\mathbf{A}, \mathbf{B} \in W$, then $\operatorname{tr}(\mathbf{A}+\mathbf{B}) = \operatorname{tr}(\mathbf{A}) + \operatorname{tr}(\mathbf{B}) = 0+0 = 0$, so $\mathbf{A}+\mathbf{B} \in W$.
- **Scalar multiplication:** If $\mathbf{A} \in W$ and $c \in F$, then $\operatorname{tr}(c\mathbf{A}) = c\operatorname{tr}(\mathbf{A}) = c(0) = 0$, so $c\mathbf{A} \in W$.

Theorem 1.2.2 (Intersection of Subspaces). Let W_1 and W_2 be subspaces of a vector space V. Then the intersection $W_1 \cap W_2$ is also a subspace of V.

Proof. We use the Subspace Test on $W = W_1 \cap W_2$:

- **Zero vector:** Since W_1 and W_2 are subspaces, $\mathbf{0} \in W_1$ and $\mathbf{0} \in W_2$. Thus $\mathbf{0} \in W_1 \cap W_2$.
- **Closure under addition:** Let $\mathbf{w}, \mathbf{z} \in W_1 \cap W_2$. Then $\mathbf{w}, \mathbf{z} \in W_1$ and $\mathbf{w}, \mathbf{z} \in W_2$. Since W_1 and W_2 are subspaces, they are closed under addition, so $\mathbf{w} + \mathbf{z} \in W_1$ and $\mathbf{w} + \mathbf{z} \in W_2$. Hence $\mathbf{w} + \mathbf{z} \in W_1 \cap W_2$.
- **Closure under scalar multiplication:** Let $\mathbf{w} \in W_1 \cap W_2$ and $c \in F$. Then $\mathbf{w} \in W_1$ and $\mathbf{w} \in W_2$. Since W_1 and W_2 are closed under scalar multiplication, $c\mathbf{w} \in W_1$ and $c\mathbf{w} \in W_2$. Hence $c\mathbf{w} \in W_1 \cap W_2$.

Since all three conditions of the Subspace Test are satisfied, $W_1 \cap W_2$ is a subspace of V. ■

Remark. Unlike the intersection, the **union** of two subspaces is not necessarily a subspace. For $W_1 \cup W_2$ to be a subspace, one subspace must be contained within the other (i.e., $W_1 \subseteq W_2$ or $W_2 \subseteq W_1$).

Example (Union of Axes). Consider the vector space $V = \mathbb{R}^2$. Let W_1 be the x-axis and W_2 be the y-axis:

$$W_1 = \{(x, 0) : x \in \mathbb{R}\}, \quad W_2 = \{(0, y) : y \in \mathbb{R}\}.$$

Both W_1 and W_2 are subspaces of $\mathbb{R}^2$. However, their union $W_1 \cup W_2$ is **not** a subspace because it is not closed under addition.

Indeed, $(1, 0) \in W_1 \subseteq W_1 \cup W_2$ and $(0, 1) \in W_2 \subseteq W_1 \cup W_2$, but their sum:

$$(1, 0) + (0, 1) = (1, 1)$$

is not in $W_1 \cup W_2$ since $(1, 1)$ is neither on the x-axis nor the y-axis.

Proposition 1.2.1 (Union of Subspaces). Let W_1, W_2 be subspaces of V. Then $W_1 \cup W_2$ is a subspace of V if and only if $W_1 \subseteq W_2$ or $W_2 \subseteq W_1$.

Proof. ($\Leftarrow$) If $W_1 \subseteq W_2$, then $W_1 \cup W_2 = W_2$, which is a subspace. Similarly, if $W_2 \subseteq W_1$, the union is W_1, which is also a subspace.

($\Rightarrow$) We prove the contrapositive: if neither is contained in the other, then the union is not a subspace. Assume $W_1 \nsubseteq W_2$ and $W_2 \nsubseteq W_1$.

- Since $W_1 \nsubseteq W_2$, there exists a vector $\mathbf{u} \in W_1$ such that $\mathbf{u} \notin W_2$.
- Since $W_2 \nsubseteq W_1$, there exists a vector $\mathbf{v} \in W_2$ such that $\mathbf{v} \notin W_1$.

Consider the sum $\mathbf{w} = \mathbf{u} + \mathbf{v}$. We show that $\mathbf{w} \notin W_1 \cup W_2$:

- If $\mathbf{w} \in W_1$, then $\mathbf{v} = \mathbf{w} - \mathbf{u}$. Since $\mathbf{w} \in W_1$ and $\mathbf{u} \in W_1$, their difference $\mathbf{v}$ must be in W_1 (by closure). But we chose $\mathbf{v} \notin W_1$, which is a contradiction.

- If $\mathbf{w} \in W_2$, then $\mathbf{u} = \mathbf{w} - \mathbf{v}$. Since $\mathbf{w} \in W_2$ and $\mathbf{v} \in W_2$, their difference $\mathbf{u}$ must be in W_2. But we chose $\mathbf{u} \notin W_2$, which is a contradiction.

Therefore, $\mathbf{u} + \mathbf{v}$ is in neither W_1 nor W_2, so it is not in $W_1 \cup W_2$. Thus, the union is not closed under addition and is therefore not a subspace. ■

1.3 Linear Combinations and Span

In the previous section, we studied subspaces. A natural question arises: given a set of vectors, how can we describe the "smallest" subspace that contains them? This leads us to the concept of linear combinations and the span.

Definition 1.3.1 (Linear Combination)**.** Let V be a vector space over F, $S = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m\}$ be a subset of V. Then the expression

$$a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \dots + a_m\mathbf{v}_m$$

is called a **linear combination** of S. The scalars a_i are called the **coefficients** of the linear combination.

Example (Examples of Linear Combinations).

1. The vector $(2, 1, 0)$ in $\mathbb{R}^3$ is a linear combination of $S = \{(1, 2, 3), (4, 5, 6)\}$ because

$$(2, 1, 0) = -2 \cdot (1, 2, 3) + 1 \cdot (4, 5, 6).$$

2. The polynomial $f(x) = x^3+2x+1$ in $\mathbb{R}[x]$ is a linear combination of $S = \{1, x, x^2, x^3\}$ because

$$f(x) = 1 \cdot 1 + 2 \cdot x + 0 \cdot x^2 + 1 \cdot x^3.$$

Definition 1.3.2 (Span)**.** Let S be a subset of a vector space V over F. The **span** of S, denoted by $\operatorname{span}_F(S)$, is the set of all linear combinations of finite subsets of elements of S.

If $S = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$, then $\operatorname{span}_F(S) = \{a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \dots + a_n\mathbf{v}_n : a_i \in F\}$. By convention, $\operatorname{span}_F(\emptyset) = \{\mathbf{0}\}$.

Remark. When the underlying field is clear, we abuse the notation of span_F and write it as span.

Theorem 1.3.1 (Spanning Set as Subspace). The span of any subset $S \subseteq V$ is a subspace of V. Moreover, it is the smallest subspace of V containing S.

Proof. Let $W = \text{span}(S)$.

- **Zero vector:** Since $\text{span}(\emptyset) = \{\mathbf{0}\}$, and for non-empty S, taking all coefficients as zero gives $\mathbf{0}$, we have $\mathbf{0} \in W$.
- **Addition:** Let $\mathbf{x}, \mathbf{y} \in W$. Then $\mathbf{x} = \sum a_i \mathbf{v}_i$ and $\mathbf{y} = \sum b_i \mathbf{v}_i$. Their sum $\mathbf{x} + \mathbf{y} = \sum (a_i + b_i)\mathbf{v}_i$ is also a linear combination of elements in S, so $\mathbf{x} + \mathbf{y} \in W$.
- **Scalar multiplication:** For any $c \in F$, $c\mathbf{x} = \sum (ca_i)\mathbf{v}_i \in W$.

To see it is the smallest subspace, note that any subspace containing S must be closed under addition and scalar multiplication, and thus must contain all linear combinations of elements in S. ■

Definition 1.3.3 (Spanning Set)**.** If $\text{span}(S) = V$, we say that S **spans** V, or that S is a **spanning set** for V.

Example (Spanning Set for Polynomials). The set $\{1, x, x^2\}$ spans $\mathbb{R}[x]_{\leq 2}$, the space of polynomials of degree at most 2.

1.4 Linear Independence

A spanning set is like a toolbox that is guaranteed to have every tool you need. But a messy toolbox might have three different hammers that all do the same thing. In mathematics, as in engineering, we value **efficiency**.

If one of the vectors in your spanning set can already be built using the others, then that vector is "dead weight"—it isn't helping you reach any new territory. We call such a set *linearly dependent*.

In this section, we search for the "cleanest" possible sets. We want to know: "Is every vector in this collection actually contributing something unique, or is someone just coasting on the work of the others?"

Definition 1.4.1 (Linear Independence)**.** A subset S of a vector space V is said to be **linearly independent** if for any distinct vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n \in S$, the only solution to the equation:

$$a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_n\mathbf{v}_n = \mathbf{0}$$

is the trivial solution $a_1 = a_2 = \cdots = a_n = 0$.

If there exists a non-trivial solution (where at least one $a_i \neq 0$), then S is said to be **linearly dependent**.

Remark. Geometrically, in $\mathbb{R}^2$, two vectors are linearly dependent if and only if they lie on the same line through the origin (one is a multiple of the other). In $\mathbb{R}^3$, three vectors are linearly dependent if and only if they lie on the same plane through the origin.

Example (Examples of Linear Independence).

1. Let $S_1 = \{(1,2,3),(4,5,6)\} \subseteq \mathbb{R}^3$. To check linear independence, suppose

$$a(1,2,3)+b(4,5,6)=(0,0,0) \implies \begin{cases} a+4b=0 \\ 2a+5b=0 \\ 3a+6b=0 \end{cases}$$

The first two equations give $a=-4b$ and $2(-4b)+5b=0 \implies -3b=0$. Thus $a=b=0$, so S_1 is **linearly independent**.

2. Let $S_2 = \{(1,2,3),(4,5,6),(7,8,9)\} \subseteq \mathbb{R}^3$. Note that

$$(1,2,3)-2(4,5,6)+(7,8,9)=(1-8+7,2-10+8,3-12+9)=(0,0,0).$$

Since there exists a non-trivial solution, S_2 is **linearly dependent**.

3. Consider $\{2x+1, x^2+3\}$ in $\mathbb{R}[x]_{\leq 2}$. Suppose

$$a(2x+1)+b(x^2+3)=0 \implies bx^2+2ax+(a+3b)=0.$$

Comparing coefficients of $x^2, x, 1$, we have $b=0$, $2a=0$, and $a+3b=0$. This implies $a=b=0$, so the set is **linearly independent**.

4. Consider $\{\sin x, \cos x, e^x\}$ in $D(\mathbb{R})$. Suppose for all $x \in \mathbb{R}$:

$$a\sin x + b\cos x + ce^x = 0.$$

Setting $x=0$ gives $b+c=0$. Setting $x=\pi$ gives $-b+ce^\pi=0$. Adding these gives $c(1+e^\pi)=0$, so $c=0$ and hence $b=0$. Finally, setting $x=\pi/2$ gives $a+ce^{\pi/2}=0 \implies a=0$. Thus, the set is **linearly independent**.

Theorem 1.4.1 (Linear Dependence Lemma). $\{\mathbf{v}_1,\mathbf{v}_2,\dots,\mathbf{v}_m\}$ is linearly dependent if and only if either $\mathbf{v}_1=\mathbf{0}$ or for some r, $\mathbf{v}_r$ is a linear combination of $\{\mathbf{v}_1,\mathbf{v}_2,\dots,\mathbf{v}_{r-1}\}$.

Proof. ($\Rightarrow$) Suppose $S=\{\mathbf{v}_1,\dots,\mathbf{v}_m\}$ is linearly dependent. Then there exist $a_1,\dots,a_m \in F$, not all zero, such that

$$a_1\mathbf{v}_1+\cdots+a_m\mathbf{v}_m=\mathbf{0}.$$

Let r be the largest index such that $a_r \neq 0$. Then

$$a_1\mathbf{v}_1+\cdots+a_{r-1}\mathbf{v}_{r-1}+a_r\mathbf{v}_r=\mathbf{0}.$$

So

$$a_r\mathbf{v}_r=-\left(a_1\mathbf{v}_1+\cdots+a_{r-1}\mathbf{v}_{r-1}\right),$$

and dividing by a_r gives

$$\mathbf{v}_r=-\frac{a_1}{a_r}\mathbf{v}_1-\cdots-\frac{a_{r-1}}{a_r}\mathbf{v}_{r-1}.$$

Hence $\mathbf{v}_r$ is a linear combination of $\{\mathbf{v}_1,\dots,\mathbf{v}_{r-1}\}$. (If $r=1$, this says $a_1\mathbf{v}_1=\mathbf{0}$ with $a_1\neq 0$, so $\mathbf{v}_1=\mathbf{0}$.)

($\Leftarrow$) We prove by cases:

- If $\mathbf{v}_1 = \mathbf{0}$, then
$$1 \cdot \mathbf{v}_1 + 0 \cdot \mathbf{v}_2 + \cdots + 0 \cdot \mathbf{v}_m = \mathbf{0},$$
with coefficients not all zero, so S is linearly dependent.
- If for some r we have $\mathbf{v}_r$ is a linear combination of $\{\mathbf{v}_1, \dots, \mathbf{v}_{r-1}\}$, then there exist scalars $c_1, \dots, c_{r-1} \in F$ such that
$$\mathbf{v}_r = c_1\mathbf{v}_1 + \cdots + c_{r-1}\mathbf{v}_{r-1}.$$
Move everything to one side:
$$c_1\mathbf{v}_1 + \cdots + c_{r-1}\mathbf{v}_{r-1} - 1 \cdot \mathbf{v}_r = \mathbf{0}.$$
Extending coefficients by zeros for $\mathbf{v}_{r+1}, \dots, \mathbf{v}_m$, we get a nontrivial linear combination of $\mathbf{v}_1, \dots, \mathbf{v}_m$ equal to $\mathbf{0}$, so S is linearly dependent.

■

Remark. The above theorem can be paraphrased to: A set $S = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ with $n \geq 2$ is linearly dependent if and only if at least one vector in S can be written as a linear combination of the others.

1.4.1 The Redundant Members

Having the last remark in mind, we can develop an intuition that says linear independence of a set is equivalent to requiring the set containing no "redundant members." Next we want to develop the intuition saying that throwing away these "redundant members" will not change the linear span.

Theorem 1.4.2 (Span Preservation). Let $S = \{\mathbf{v}_1, \dots, \mathbf{v}_m, \mathbf{w}\} \subseteq V$. If $\mathbf{w} \in \operatorname{span}\{\mathbf{v}_1, \dots, \mathbf{v}_m\}$, then
$$\operatorname{span}(S) = \operatorname{span}\{\mathbf{v}_1, \dots, \mathbf{v}_m\}.$$

Proof. Let $W = \operatorname{span}\{\mathbf{v}_1, \dots, \mathbf{v}_m\}$ and $W' = \operatorname{span}(S)$. Since $\{\mathbf{v}_1, \dots, \mathbf{v}_m\} \subseteq S$, it is clear that $W \subseteq W'$.

For the reverse inclusion, let $\mathbf{v} \in W'$. Then $\mathbf{v}$ is a linear combination of vectors in S:
$$\mathbf{v} = a_1\mathbf{v}_1 + \cdots + a_m\mathbf{v}_m + b\mathbf{w}$$
for some scalars a_i, b. Since $\mathbf{w} \in W$, there exist scalars $c_1, \dots, c_m$ such that $\mathbf{w} = c_1\mathbf{v}_1 + \cdots + c_m\mathbf{v}_m$. Substituting this into the expression for $\mathbf{v}$:
$$\mathbf{v} = a_1\mathbf{v}_1 + \cdots + a_m\mathbf{v}_m + b(c_1\mathbf{v}_1 + \cdots + c_m\mathbf{v}_m) = (a_1 + bc_1)\mathbf{v}_1 + \cdots + (a_m + bc_m)\mathbf{v}_m.$$
Thus $\mathbf{v}$ is a linear combination of $\mathbf{v}_1, \dots, \mathbf{v}_m$, so $\mathbf{v} \in W$. This shows $W' \subseteq W$. Hence $W' = W$. ■

1.5 Bases

We have reached what many consider the “Goldilocks Zone” of Linear Algebra.

If a set is too small, it won’t span the whole space—you’ll be left with points you can’t reach. If a set is too large, it will be linearly dependent—you’ll have redundant vectors cluttering your workspace. A **basis** is a set that is “just right.” It is large enough to build everything, but small enough that every piece is essential.

Because a basis has no redundancy, it gives us something incredible: a **unique address system**. In an abstract vector space, it’s hard to tell someone where a vector is. But once we pick a basis, every vector can be described by a unique list of numbers—its coordinates. A basis is the bridge that allows us to turn abstract geometry into concrete arithmetic.

Definition 1.5.1 (Basis)**.** A subset B of a vector space V is called a **basis** for V if:

1. B is linearly independent;
2. B spans V.

Example (Standard Bases).

- The **standard basis** for F^n is $\{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n\}$, where $\mathbf{e}_i$ has a 1 in the i-th position and 0 elsewhere.
- The standard basis for $F[x]_{\leq n}$ is $\{1, x, x^2, \dots, x^n\}$.
- The standard basis for $M_{m\times n}(F)$ is the set of matrices $\mathbf{E}_{ij}$ having a 1 at entry (i, j) and 0 elsewhere.

1.5.1 Unique Representation and Coordinates

The most important property of a basis is that every vector in the space has a unique “address” relative to it.

Theorem 1.5.1 (Unique Representation). Let $B = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ be a basis for V. Then every vector $\mathbf{v} \in V$ can be written as a linear combination of elements in B in **exactly one way**.

Proof. Since B spans V, there exist scalars a_i such that $\mathbf{v} = \sum a_i\mathbf{v}_i$. Suppose there is another representation $\mathbf{v} = \sum b_i\mathbf{v}_i$. Then:

$$\mathbf{0} = \mathbf{v} - \mathbf{v} = \sum a_i\mathbf{v}_i - \sum b_i\mathbf{v}_i = \sum (a_i - b_i)\mathbf{v}_i.$$

Since B is linearly independent, we must have $a_i - b_i = 0$ for all i, meaning $a_i = b_i$. ■

Definition 1.5.2 (Coordinates)**.** Let $B = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ be an **ordered** basis for V. For any $\mathbf{v} \in V$, the unique scalars $c_1, \dots, c_n$ such that $\mathbf{v} = \sum c_i\mathbf{v}_i$ are called the **coordinates**

of $\mathbf{v}$ with respect to B. We write this as a column vector:

$$[\mathbf{v}]_B = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix}.$$

Remark. The map $\mathbf{v} \mapsto [\mathbf{v}]_B$ provides a way to treat any abstract n-dimensional vector space as if it were simply F^n. This is the power of a basis.

1.6 Sifting and the Replacement Theorem

How "big" is a vector space? We intuitively know that a plane is "bigger" than a line, and 3D space is "bigger" than a plane. But in the abstract world of vector spaces, we need a rigorous way to measure this.

The answer lies in the number of vectors in a basis. But there's a potential problem: what if one person finds a basis with 3 vectors, and another person finds a basis for the same space with 4 vectors? If that were possible, our entire concept of "size" would collapse.

In this section, we prove the most important technical result of the chapter: the **Steinitz Replacement Theorem**. This theorem is the "engine" that guarantees every basis of a space has the exact same number of vectors. Once we have this, we can finally define the **Dimension** of a space—a single number that captures its fundamental complexity.

Theorem 1.6.1 (Sifting Method). Let $S = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m\}$ be a subset of V. Then there exists a subset $S' \subseteq S$ such that:

1. S' is linearly independent, and
2. $\text{span}(S') = \text{span}(S)$.

Proof. For each $r = 1, 2, \dots, m$, we check whether $\mathbf{v}_r$ is a linear combination of the previous terms $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_{r-1}$. We:

1. Keep $\mathbf{v}_r$ in S' if $\mathbf{v}_r$ is not a linear combination of $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_{r-1}$ (not redundant);
2. Remove $\mathbf{v}_r$ from S' if $\mathbf{v}_r$ is a linear combination of $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_{r-1}$ (redundant).

After this process, we obtain a subset S' in which no element is a linear combination of the preceding members anymore. Therefore, by Theorem 1.4.1, S' is linearly independent. Moreover, according to Theorem 1.4.2, removing "redundant members" from S does not change its linear span. Therefore, we have $\text{span}(S') = \text{span}(S)$. ■

Remark. The sifting method is a powerful algorithmic tool that allows us to "clean up" any set of vectors. Given any collection of vectors, we can systematically remove the redundant ones while preserving the span. This gives us a "cleaned" version that is linearly independent and spans the same space.

> This method becomes our "new weapon" in linear algebra proofs. Whenever we encounter a set that might have redundant vectors, we can apply the sifting method to obtain a linearly independent subset with the same span. This technique will be particularly useful in proving results about dimension and in constructing bases from spanning sets.

Example (Sifting Example). Consider $S = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4, \mathbf{v}_5, \mathbf{v}_6\}$ in $\mathbb{R}^3$ where:

$$\begin{aligned} &\mathbf{v}_1 = (1,1,1), && \mathbf{v}_2 = (2,2,2), && \mathbf{v}_3 = (1,0,0), \\ &\mathbf{v}_4 = (3,2,2), && \mathbf{v}_5 = (1,1,0), && \mathbf{v}_6 = (0,0,1). \end{aligned}$$

We apply the sifting method:

- $\mathbf{v}_1 \neq \mathbf{0}$, so $\mathbf{v}_1$ is kept.
- $\mathbf{v}_2 = 2\mathbf{v}_1$, so $\mathbf{v}_2$ is removed.
- $\mathbf{v}_3$ is not a linear combination of $\mathbf{v}_1$, so $\mathbf{v}_3$ is kept.
- $\mathbf{v}_4 = 2\mathbf{v}_1 + 1\mathbf{v}_3$, so $\mathbf{v}_4$ is removed.
- Suppose $\mathbf{v}_5 = a\mathbf{v}_1 + b\mathbf{v}_3$. Then $(1,1,0) = (a+b, a, a)$, which gives $a = 0$ and $a + b = 1$, so $b = 1$. But then the second coordinate gives $1 = a = 0$, which is a contradiction. So $\mathbf{v}_5$ is not a linear combination of $\mathbf{v}_1, \mathbf{v}_3$. Hence, $\mathbf{v}_5$ is kept.
- Since $\dim(\mathbb{R}^3) = 3$, any four vectors cannot be linearly independent. Therefore, $\mathbf{v}_6$ must be a linear combination of $\mathbf{v}_1, \mathbf{v}_3, \mathbf{v}_5$. So $\mathbf{v}_6$ is removed.

The sifted set is $S' = \{\mathbf{v}_1, \mathbf{v}_3, \mathbf{v}_5\} = \{(1,1,1), (1,0,0), (1,1,0)\}$, which is linearly independent and spans the same space as S.

Theorem 1.6.2 (Steinitz Exchange/Replacement Theorem). Let V be a vector space over F. Suppose:

- $G = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$ spans V;
- $L = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_m\}$ is linearly independent.

Then $m \leq n$.

Proof. The strategy is to insert the vectors from L one-by-one into G and "sift" the resulting set to maintain the spanning property. We rely on the fact that if we add a vector $\mathbf{w}$ to a spanning set S, the set $\{\mathbf{w}\} \cup S$ becomes linearly dependent, allowing us to remove a redundant vector (by Theorem 1.4.2).

Consider the first step. We form the set B_1 by inserting $\mathbf{u}_1$ into G:

$$B_1 = \{\mathbf{u}_1, \mathbf{v}_1, \dots, \mathbf{v}_n\}.$$

Since G spans V, B_1 is linearly dependent. By applying the sifting process (Theorem 1.6.1), we can remove a vector to obtain a new spanning set G_1. Since L is linearly independent, $\mathbf{u}_1 \neq \mathbf{0}$, so the removed vector must be some $\mathbf{v}_i$. Thus:

$$G_1 = \{\mathbf{u}_1\} \cup S_1,$$

where S_1 is a subset of G with size $n-1$.

Now, consider the general step. Suppose we have successfully exchanged k vectors such that the set

$$G_k = \{\mathbf{u}_1, \dots, \mathbf{u}_k\} \cup S_k$$

spans V, where $S_k \subset G$ contains $n-k$ vectors. We insert the next vector $\mathbf{u}_{k+1}$ to form:

$$B_{k+1} = \{\mathbf{u}_{k+1}\} \cup G_k.$$

Since G_k is a spanning set, $\mathbf{u}_{k+1}$ depends on the vectors in G_k, making B_{k+1} dependent. We must remove a vector to restore the spanning property. We cannot remove any of the $\mathbf{u}_i$'s, as this would imply a linear dependence relation exclusively among elements of L, contradicting the hypothesis that L is linearly independent. Therefore, we must remove some $\mathbf{v}_j$ from S_k.

This process generates a sequence of subsets of remaining $\mathbf{v}$'s:

$$n = |G| > |S_1| > |S_2| > \cdots > |S_m| \geq 0.$$

Since we are able to perform this exchange m times (once for each $\mathbf{u} \in L$), we must have started with enough $\mathbf{v}$'s to accommodate every removal. Therefore, $n \geq m$. ■

1.6.1 Consequences of the Replacement Theorem

Having proved the Steinitz Replacement Theorem, the major conclusion we can draw from it is that any (finite) basis of the same vector space will have the same size.

Theorem 1.6.3 (Basis Theorem). If a vector space V over F has a finite basis, then every basis of V has the same number of vectors.

Proof. Let B_1 and B_2 be two bases with n and m vectors respectively.

- Since B_1 is linearly independent and B_2 spans, by Replacement Theorem, $n \leq m$.
- Since B_2 is linearly independent and B_1 spans, by Replacement Theorem, $m \leq n$.

Thus $n = m$. ■

Definition 1.6.1 (Dimension)**.** A vector space V over F is called **finite-dimensional** if it has a finite basis. The **dimension** of V, denoted $\dim_F(V)$, is the number of vectors in any basis for V. By convention, $\dim_F(\{\mathbf{0}\}) = 0$.

Remark. When the underlying field is clear, we abuse the notation of $\dim_F$ and write it as dim.

Example (Examples of Dimensions). $\dim(F^n) = n$, $\dim(F[x]_{\leq n}) = n+1$, and $\dim(M_{m\times n}(F)) = mn$.

Theorem 1.6.4 (Size Bounds for Independent and Spanning Sets). Let V be a vector space of dimension n over F, and $S \subseteq V$ be a subset.

1. If $|S| > n$, then S cannot be linearly independent.
2. If $|S| < n$, then S cannot span V.

Proof.

1. Given $|S| > n$, and suppose for the sake of contradiction that S is linearly independent. Let B be a basis of V, then we have $|B| = n$ and B is a spanning set of V. By Replacement Theorem, we have

$$\underbrace{|S|}_{\text{size of lin. indep. set}} \leq \underbrace{|B|}_{\text{size of spanning set}} = n.$$

   Giving a contradiction, hence S cannot be linearly independent.

2. Given $|S| < n$, and suppose for the sake of contradiction that S spans V. Let B be a basis of V, then we have $|B| = n$ and B is linearly independent. By Replacement Theorem, we have

$$n = \underbrace{|B|}_{\text{size of lin. indep. set}} \leq \underbrace{|S|}_{\text{size of spanning set}}.$$

   This implies $n \leq |S|$, which contradicts the assumption that $|S| < n$. Hence S cannot span V.

■

Remark.

Size of lin. indep. set $\leq$ Size of basis $\leq$ Size of spanning set.

Theorem 1.6.5 (Basis Extension Theorem). Let V be a finite-dimensional vector space and S be a linearly independent subset of V. Then S can be extended into a basis of V.

Proof. Since V is finite-dimensional, it has a finite basis $B = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$. Consider the set $S \cup B$. Since B spans V, $S \cup B$ also spans V. We apply the Sifting Method (Theorem 1.6.1) to the ordered set formed by listing the elements of S first, followed by the elements of B.

The Sifting Method will keep all elements of S because S is linearly independent. It then proceeds to examine each element of B, keeping those that are not in the span of the vectors already kept. The resulting subset S' spans $\operatorname{span}(S \cup B) = V$ and is linearly independent by construction. Thus, S' is a basis for V that contains S. ■

Corollary 1.6.1 (Existence of Basis). Every finite spanning set of a vector space V contains a basis.

Proof. Apply the Sifting Method to the spanning set. The resulting subset S' is linearly independent by construction and still spans V. Therefore, S' is a basis for V. ■

Theorem 1.6.6 (Dimension Implies Equality). Let V and W be finite-dimensional vector spaces over F. If

1. $W \subseteq V$;
2. $\dim_F(W) = \dim_F(V)$,

then $W = V$.

Proof. If we show $V \subseteq W$ then we are done. Suppose for contradiction that $V \nsubseteq W$. That means there exists some $\mathbf{v} \in V$ that $\mathbf{v} \notin W$.

Let $S = \{\mathbf{w}_1, \dots, \mathbf{w}_n\}$ be a basis of W, (consequently, $\dim_F(W) = n$). Since $\mathbf{v} \notin W$, we have $\mathbf{v} \notin \operatorname{span}(S)$.

That would mean that $S' = \{\mathbf{w}_1, \dots, \mathbf{w}_n, \mathbf{v}\}$ is linearly independent in V. By Replacement Theorem, we have

$$\underbrace{|S'|}_{\text{size of lin. indep. set}} \le \dim_F(V),$$

which implies $n + 1 \le n$, which is a contradiction, therefore $V \subseteq W$. ■

Remark. This theorem makes proving $W = V$ somewhat easier. As if we wanted to prove equivalence before, we can only do it by proving $W \subseteq V$ and $V \subseteq W$. Now we can just prove one side and then state that their dimension is the same.

An immediate consequence of the above theorem is the following:

Theorem 1.6.7 (Maximal Independent Subset is a Basis). Let V be a vector space of dimension n over F. Then any n independent vectors form a basis of V. (We say that "any maximal linearly independent subset forms a basis").

Proof. Let S be a linearly independent set of size n. We want to show that $\operatorname{span}(S) = V$.

Let $W = \operatorname{span}(S)$, this implies that W is a subspace of V, which implies $W \subseteq V$.