MATH 2022
Linear Algebra II

Joe Tran

July 9, 2025

Preface

These are the first edition of these lecture notes for MATH 2022 (Linear Algebra II). Consequently, there may be several typographical errors, missing exposition on necessary background, and more advanced topics for which there will not be time in class to cover. Future iterations of these notes will hopefully be fairly self-contained provided one has the necessary background. If you come across any typos, errors, omissions, or unclear expositions, please feel free to contact me so that I may continually improve these notes.

Contents

Chapter 1

Vector Spaces

Linear algebra is fundamentally concerned with the study of vector spaces and linear transformations between them. A vector space is a mathematical structure formed by a collection of objects, called vectors, which can be added together and multiplied (”scaled”) by elements of a field, such as the real numbers $\mathbb{R}$ or complex numbers $\mathbb{C}$.

The abstraction of vector spaces allows us to unify and generalize many familiar mathematical objects—such as geometric vectors, sequences, matrices, polynomials, and functions—under a single framework. Despite their varied appearances, these objects all obey a shared set of algebraic rules.

Understanding vector spaces is crucial not only in pure mathematics but also in applied fields such as physics, engineering, computer science, and data science. Many problems in these domains can be reduced to questions about vectors, subspaces, and transformations.

In this chapter, we begin by defining vector spaces formally, exploring key examples and counterexamples, and laying the groundwork for deeper study of linear structure, spanning sets, linear independence, basis, and dimension.

1.1 Vector Spaces

A vector space consists of a nonempty set V and two operations, one called vector addition, and the other called scalar multiplication. In these notes, we will be more focused on vector spaces over $\mathbb{F}$, where $\mathbb{F} = \mathbb{Q}$, $\mathbb{F} = \mathbb{R}$ or $\mathbb{F} = \mathbb{C}$. Later on, we will be more focused on scalars of real or complex numbers, so we will denote $\mathbb{K}$ to be either $\mathbb{K} = \mathbb{R}$ or $\mathbb{K} = \mathbb{C}$.

Definition 1.1.1. A *vector space* (or *linear space*) over a field $\mathbb{F}$ is a set V equipped with:

- A binary operation $+ : V \times V \to V$ (called *vector addition*).
- An operation $\cdot : \mathbb{F} \times V \to V$ (called *scalar multiplication*)

such that the following axioms hold for all $\vec{u}, \vec{v}, \vec{w} \in V$ and $\alpha, \beta \in \mathbb{F}$:

1. (Associativity of Vector Addition) $(\vec{u} + \vec{v}) + \vec{w} = \vec{u} + (\vec{v} + \vec{w})$.
2. (Commutativity of Vector Addition) $\vec{u} + \vec{v} = \vec{v} + \vec{u}$.
3. (Additive Identity) There exists a $\vec{0}_V \in V$ such that for all $\vec{v} \in V$, $\vec{v} + \vec{0}_V = \vec{0}_V + \vec{v} = \vec{v}$.
4. (Additive Inverse) For every $\vec{v} \in V$, there exists a $-\vec{v} \in V$ such that $\vec{v} + (-\vec{v}) = -\vec{v} + \vec{v} = \vec{0}_V$.
5. (Multiplicative Identity) $\vec{v} \in V$, $1 \cdot \vec{v} = \vec{v}$.
6. (Distributivity over Vector Addition) $\alpha \cdot (\vec{u} + \vec{v}) = \alpha \cdot \vec{u} + \alpha \cdot \vec{v}$.
7. (Distributivity over Scalar Addition) $(\alpha + \beta) \cdot \vec{v} = \alpha \cdot \vec{v} + \beta \cdot \vec{v}$.
8. (Compatibility of Scalar Multiplication) $(\alpha\beta) \cdot \vec{v} = \alpha \cdot (\beta\vec{v})$.

Example 1.1.2. For $n \in \mathbb{N}$, we define the n-dimensional Euclidean space $\mathbb{R}^n$ by

$$\mathbb{R}^n = \left\{ \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} : x_1, x_2, ..., x_n \in \mathbb{R} \right\}$$

We equip $\mathbb{R}^n$ with the usual vector addition and scalar multiplication defined by

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_n + y_n \end{bmatrix} \quad \text{and} \quad \alpha \cdot \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} \alpha x_1 \\ \alpha x_2 \\ \vdots \\ \alpha x_n \end{bmatrix}$$

It is elementary to verify that $\mathbb{R}^n$ with the defined operations is a vector space. Note that instead of $\mathbb{R}^n$, we can also have $\mathbb{C}^n$ or $\mathbb{Q}^n$ with the same operations defined.

Example 1.1.3. For $m, n \in \mathbb{N}$, we define the set of all $m \times n$ matrices with entries in $\mathbb{F}$ by

$$\mathcal{M}_{mn}(\mathbb{F}) = \left\{ \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix} : x_{ij} \in \mathbb{F} \text{ for all } 1 \leq i \leq m,\ 1 \leq j \leq n \right\}$$

We equip $\mathcal{M}_{mn}(\mathbb{F})$ with the usual matrix addition defined by

$$\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix} + \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1n} \\ y_{21} & y_{22} & \cdots & y_{2n} \\ \vdots & \vdots & & \vdots \\ y_{m1} & y_{m2} & \cdots & y_{mn} \end{bmatrix} = \begin{bmatrix} x_{11} + y_{11} & x_{12} + y_{12} & \cdots & x_{1n} + y_{1n} \\ x_{21} + y_{21} & x_{22} + y_{22} & \cdots & x_{2n} + y_{2n} \\ \vdots & \vdots & & \vdots \\ x_{m1} + y_{m1} & x_{m2} + y_{m2} & \cdots & x_{mn} + y_{mn} \end{bmatrix}$$

and scalar multiplication defined by

$$\alpha \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix} = \begin{bmatrix} \alpha x_{11} & \alpha x_{12} & \cdots & \alpha x_{1n} \\ \alpha x_{21} & \alpha x_{22} & \cdots & \alpha x_{2n} \\ \vdots & \vdots & & \vdots \\ \alpha x_{m1} & \alpha x_{m2} & \cdots & \alpha x_{mn} \end{bmatrix}$$

It is elementary to verify that $\mathcal{M}_{mn}(\mathbb{F})$ with the defined operations is a vector space over $\mathbb{F}$. Note that in the case where we are working with $n \times n$ matrices over $\mathbb{F}$, we will simply denote it as $\mathcal{M}_n(\mathbb{F})$.

Example 1.1.4. For $n \in \mathbb{N}$, we define the set of all functions on the interval $[0, 1]$ over $\mathbb{F}$ by

$$\mathcal{F}[0,1] = \{f : [0,1] \to \mathbb{F} : f \text{ is a function}\}$$

We equip $\mathcal{F}[0, 1]$ with pointwise addition and scalar multiplication defined by

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

It is elementary to verify that $\mathcal{F}[0, 1]$ with the defined operations is a vector space over $\mathbb{F}$.

Example 1.1.5. For $n \in \mathbb{N}$, we define the set of all polynomials over $\mathbb{F}$ of degree at most n on the interval $[0,1]$ by

$$\mathcal{P}_n[0,1] = \left\{ p : [0,1] \to \mathbb{F} : n \in \mathbb{N}, a_0, a_1, ..., a_n \in \mathbb{F}, p(x) = \sum_{k=0}^{n} a_k x^k \right\}$$

We equip $\mathcal{P}_n[0,1]$ with pointwise addition and scalar multiplication as in Example 1.1.4. It is elementary to verify that $\mathcal{P}_n[0,1]$ with the defined operations is a vector space over $\mathbb{F}$.

Example 1.1.6. The set of all $\mathbb{F}$-valued sequences defined by

$$\mathbb{F}^{\mathbb{N}} = \{(x_n)_{n=1}^{\infty} : \text{for all } n \in \mathbb{N},\, x_n \in F\}$$

We equip $\mathbb{F}^{\mathbb{N}}$ with the usual vector addition and scalar multiplication defined by

$$\begin{aligned} (x_n)_{n=1}^{\infty} + (y_n)_{n=1}^{\infty} &= (x_n + y_n)_{n=1}^{\infty} \\ \alpha(x_n)_{n=1}^{\infty} &= (\alpha x_n)_{n=1}^{\infty} \end{aligned}$$

It is elementary to verify that $\mathbb{F}^{\mathbb{N}}$ with the defined operations is a vector space over $\mathbb{F}$.

We now present some examples of non-vector spaces.

Example 1.1.7. Let $V = \{x \in \mathbb{R} : x \geq 0\}$. We claim that V is not a vector space over $\mathbb{R}$. To see this, note that for any $x > 0$, there is no additive inverse $-x \in \mathbb{R}$ such that $x + (-x) = -x + x = 0$. For example, $2 \in V$, but $-2 \notin V$.

Example 1.1.8. Consider the 2-dimensional Euclidean space $\mathbb{R}^2$, but say we equip it with addition defined by

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \boxplus \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

and usual scalar multiplication. We claim that $\mathbb{R}^2$ with this form of addition and scalar multiplication, is not a vector space over $\mathbb{R}$. To see this, we show that $\mathbb{R}^2$ does not satisfy the associativity property. Observe that

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \boxplus \left(\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \boxplus \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} \right) = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \boxplus \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \end{bmatrix}$$

On the other hand, we have

$$\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \boxplus \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \right) \boxplus \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \boxplus \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \begin{bmatrix} x_1 + z_1 \\ x_2 + z_2 \end{bmatrix}$$

Clearly, the two resulting vectors are different, so $\mathbb{R}^n$ with such addition and scalar multiplication is not a vector space over $\mathbb{R}$.

1.2 Subspaces of a Vector Space

Subspaces form the foundational building blocks within a vector space. Just as subsets of a set inherit structure from the parent set, subspaces inherit the algebraic structure of a vector space. A subspace is a subset that is itself a vector space under the same operations of addition and scalar multiplication.

Understanding subspaces is central to linear algebra because many important constructs—such as the null space and column space of a matrix, the space of solutions to a homogeneous system, and spaces of polynomials or continuous functions—are all subspaces of some ambient vector space. Much of the power of linear algebra lies in analyzing the structure and relationships between these subspaces.

An essential aspect of studying subspaces involves understanding how they are generated. Given a subset of vectors, we are often interested in the smallest subspace that contains them. This leads naturally to the concept of the *span* of a set of vectors: the set of all linear combinations of those vectors. A spanning set provides a way to describe the entirety of a subspace in terms of simpler components.

In this section, we explore the definition and properties of subspaces, techniques to verify whether a given subset is a subspace, important examples and counterexamples, the concept of the span of a set of vectors, and the relationship between spanning sets and subspaces. These ideas form the theoretical foundation for understanding linear dependence, bases, and dimension in later sections.

Definition 1.2.1. Let V be a vector space over a field $\mathbb{F}$. A subset $W \subseteq V$ is called a *subspace* of V if W is itself a vector space under the same operations of vector addition and scalar multiplication defined on V.

To verify that a nonempty subset $W \subseteq V$ is a subspace, we require the following proposition.

Theorem 1.2.2 (Subspace Test). *Let V be a vector space over $\mathbb{F}$ and let $W \subseteq V$. Then W is a subspace of V if and only if*

1. *(Zero element) $\vec{0}_V \in W$*
2. *(Closed under vector addition) If $\vec{u}, \vec{v} \in W$, then $\vec{v} + \vec{w} \in W$.*
3. *(Closed under scalar multiplication) If $\vec{v} \in W$ and $\alpha \in \mathbb{F}$, then $\alpha\vec{v} \in W$.*

Proof. Follows directly from checking the vector space axioms restricted to the subset W. □

Example 1.2.3. If V is a vector space, then $\{\vec{0}_V\}$ and V are subspaces of V.

Example 1.2.4. Let $V = \mathbb{R}^3$ be a vector space over $\mathbb{R}$, and define

$$W = \left\{ \begin{bmatrix} x \\ y \\ z \end{bmatrix} \in V : x + y + z = 0 \right\}$$

We claim that W is a subspace of V. To see this, we need to verify the conditions of Theorem 1.2.2.

To see that (1) holds, note that the zero vector $\vec{0} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$ satisfies $x+y+z = 0 + 0 + 0 = 0$, so $\vec{0} \in W$.

To see that (2) holds, let $\vec{u} = \begin{bmatrix} x_1 \\ y_1 \\ z_1 \end{bmatrix}$ and $\vec{v} = \begin{bmatrix} x_2 \\ y_2 \\ z_2 \end{bmatrix}$ be vectors in W. Then we have $x_1 + y_1 + z_1 = 0$ and $x_2 + y_2 + z_2 = 0$. Then observe that

$$(x_1 + x_2) + (y_1 + y_2) + (z_1 + z_2) = (x_1 + y_1 + z_1) + (x_2 + y_2 + z_2) = 0$$

Therefore, the vector

$$\vec{u} + \vec{v} = \begin{bmatrix} x_1 \\ y_1 \\ z_1 \end{bmatrix} + \begin{bmatrix} x_2 \\ y_2 \\ z_2 \end{bmatrix} = \begin{bmatrix} x_1 + x_2 \\ y_1 + y_2 \\ z_1 + z_2 \end{bmatrix}$$

belongs to W.

Finally, to see that (3) holds, let $\vec{v} = \begin{bmatrix} x \\ y \\ z \end{bmatrix}$, and let $\alpha \in \mathbb{R}$. Then we have $x + y + z = 0$. Then observe that

$$\alpha x + \alpha y + \alpha z = \alpha(x + y + z) = 0$$

Therefore, the vector

$$\alpha\vec{v} = \begin{bmatrix} \alpha x \\ \alpha y \\ \alpha z \end{bmatrix}$$

belongs to W.

Therefore, by Theorem 1.2.2, we have shown that W is a subspace of V.

Example 1.2.5. Let $V = \mathcal{F}[0,1]$ be a vector space over $\mathbb{F}$, and let $W = \{f \in V : f(0) = 0\}$. We claim that W is a subspace of V. To see this, we need to verify the conditions of Theorem 1.2.2.

To see that (1) holds, it is easy to note that the zero function is a member of W.

To see that (2) holds, let $f, g \in W$. Then $f(0) = 0$ and $g(0) = 0$ and so

$$(f+g)(0) = f(0) + g(0) = 0 + 0 = 0$$

Therefore, $f + g \in W$.

Finally, to see that (3) holds, let $f \in W$ and $\alpha \in \mathbb{F}$. Then $f(0) = 0$ and so

$$(\alpha f)(0) = \alpha f(0) = \alpha \cdot 0 = 0$$

Therefore, $\alpha f \in W$.

Therefore, by Theorem 1.2.2, we have shown that W is a subspace of V.

Example 1.2.6. Let $V = \mathcal{M}_n(\mathbb{R})$, and for $A \in \mathcal{M}_n(\mathbb{R})$, define

$$Z_n(A) = \{B \in \mathcal{M}_n(\mathbb{R}) : AB = BA\}$$

Then $Z_n(A)$ is a subspace of $\mathcal{M}_n(\mathbb{R})$. To see this, we need to verify the conditions of Theorem 1.2.2.

To see that (1) holds, first note that the zero matrix $O_n \in \mathcal{M}_n(\mathbb{R})$ satisfies

$$AO_n = O_nA = O_n$$

so $O_n \in Z_n(A)$.

To see that (2) holds, let $B, C \in Z_n(\mathbb{R})$. Then we have $AB = BA$ and $AC = CA$. Consequently,

$$A(B+C) = AB + AC = BA + CA = (B+C)A$$

which shows that $B + C \in Z_n(\mathbb{R})$.

Finally, to see that (3) holds, let $B \in Z_n(\mathbb{R})$ and $\alpha \in \mathbb{R}$. Then we have $AB = BA$. Consequently,

$$A(\alpha B) = \alpha AB = \alpha BA = (\alpha B)A$$

which shows that $\alpha B \in Z_n(\mathbb{R})$.

Therefore, by Theorem 1.2.2, we have shown that $Z_n(\mathbb{R})$ is a subspace of $\mathcal{M}_n(\mathbb{R})$.

Example 1.2.7. Let $V = \mathcal{C}[0,1]$ be the space of all continuous functions from the unit interval $[0,1]$ to $\mathbb{F}$. It is elementary to verify that $\mathcal{C}[0,1]$ with pointwise addition and scalar multiplication is a vector space over $\mathbb{F}$. Define

$$\mathcal{D}[0,1] = \{f \in \mathcal{C}[0,1] : f \text{ is differentiable on } [0,1]\}$$

We claim that $\mathcal{D}[0,1]$ is a subspace of $\mathcal{C}[0,1]$. To see this, we need to verify the conditions of Theorem 1.2.2.

To see that (1) holds, note that clearly, the zero function $O(x) = 0$ is differentiable on $[0,1]$, so $O \in \mathcal{D}[0,1]$.

To see that (2) holds, let $f, g \in \mathcal{D}[0,1]$, so f and g are differentiable on $[0,1]$, so f', g' exists. Then from calculus, it follows that $f+g$ is differentiable on $[0,1]$ and

$$(f+g)' = f' + g'$$

so $f + g \in \mathcal{D}[0,1]$.

To see that (3) holds, let $f \in \mathcal{D}[0,1]$ and $\alpha \in \mathbb{F}$, so f is differentiable on $[0,1]$, so f' exists. From calculus, again, it follows that αf is differentiable on $[0,1]$, and

$$(\alpha f)' = \alpha f'$$

so $\alpha f \in \mathcal{D}[0,1]$.

Therefore, by Theorem 1.2.2, we have shown that $\mathcal{D}[0,1]$ is a subspace of $\mathcal{C}[0,1]$.

In linear algebra, one of the central goals is to describe vector spaces using simpler, more manageable building blocks. A key concept in this process is the idea of a *spanning set*.

Definition 1.2.8. Let V be a vector space over $\mathbb{F}$ and let $\{\vec{v}_1, \vec{v}_2, ..., \vec{v}_n\}$ be a collection of vectors in V.

1. A vector $\vec{v}$ is said to be a *linear combination* of the vectors $\{\vec{v}_1, \vec{v}_2, ..., \vec{v}_n\}$ if there are coefficients $\alpha_1, \alpha_2, ..., \alpha_n \in \mathbb{F}$ such that

$$\vec{v} = \alpha_1\vec{v}_1 + \alpha_2\vec{v}_2 + \cdots + \alpha_n\vec{v}_n$$

2. The set of all linear combinations of $\{\vec{v}_1, \vec{v}_2, ..., \vec{v}_n\}$ is the *span*, and is denoted by

$$\operatorname{span}\{\vec{v}_1, \vec{v}_2, ..., \vec{v}_n\} = \left\{ \sum_{k=1}^{n} \alpha_k \vec{v}_k : \alpha_1, \alpha_2, ..., \alpha_n \in \mathbb{F} \right\}$$

3. We say that a nonempty set $S \subseteq V$ is a *spanning set* for V if

$$\text{span}(S) = V$$

Example 1.2.9. Let $V = \mathbb{R}^2$, and consider the vectors $\vec{v}_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ and $\vec{v}_2 = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$. Then the vector $\vec{v} = \begin{bmatrix} 9 \\ 8 \end{bmatrix}$ is a linear combination of $\vec{v}_1$ and $\vec{v}2$ because

$$\vec{v} = 3\vec{v}_1 + 2\vec{v}_2 = 3\begin{bmatrix} 1 \\ 2 \end{bmatrix} + 2\begin{bmatrix} 3 \\ 1 \end{bmatrix}$$

Example 1.2.10. Let $V = \mathcal{P}_2(\mathbb{R})$ and let $p(x) = 1 + x + 4x^2$. We claim that $p(x)$ belongs to

$$\text{span}\{1 + 2x - x^2, 3 + 5x + 2x^2\}$$

To see this, let $a, b \in \mathbb{R}$ be such that

$$p(x) = a(1 + 2x - x^2) + b(3 + 5x + 2x^2)$$

Then

$$1 + x + 4x^2 = (a + 3b) + (2a + 5b)x + (-a + 2b)x^2$$

Equating coefficients of powers of x, we have

$$\begin{cases} a + 3b = 1 \\ 2a + 5b = 1 \\ -a + 2b = 4 \end{cases}$$

These equations have the solutions $a = -2$ and $b = 1$, so $p(x)$ is indeed in $\text{span}\{1 + 2x - x^2, 3 + 5x + 2x^2\}$.

Example 1.2.11. Let $V = \mathbb{R}^3$ and let

$$S = \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\} \subseteq \mathbb{R}^3$$

Then $\text{span}(S) = \mathbb{R}^3$ since every vector $\vec{v} = \begin{bmatrix} x \\ y \\ z \end{bmatrix}$ in $\mathbb{R}^3$ can be written as

$$\vec{v} = x\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + y\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + z\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

In this case, S is a spanning set for $\mathbb{R}^3$. In general, this can be extended to n-dimensional. That is, if

$$S = \left\{ \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, ..., \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix} \right\} \subseteq \mathbb{R}^n$$

Then $\text{span}(S) = \mathbb{R}^n$.

Example 1.2.12. Let $V = \mathcal{M}_2(\mathbb{R})$ and let

$$S = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\} \subseteq \mathcal{M}_2(\mathbb{R})$$

Then $\text{span}(S) = \mathcal{M}_2(\mathbb{R})$ since every vector $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ in $\mathcal{M}_2(\mathbb{R})$ can be written as

$$A = a \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + b \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} + c \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} + d \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

In this case, S is a spanning set for $\mathcal{M}_2(\mathbb{R})$. In general, this can be extended to $\mathcal{M}_n(\mathbb{R})$. In this case, if

$$S = \left\{ \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}, ..., \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}, \right\} \subseteq \mathcal{M}_n(\mathbb{R})$$

Then $\text{span}(S) = \mathcal{M}_n(\mathbb{R})$.

Example 1.2.13. Let $V = \mathcal{P}_2(\mathbb{R})$ and let

$$S = \{1, x, x^2\} \subseteq \mathcal{P}_2(\mathbb{R})$$

Then $\text{span}(S) = \mathcal{P}_2(\mathbb{R})$ since every vector $p(x) = a + bx + cx^2$ in $\mathcal{P}_2(\mathbb{R})$ can be written as

$$p(x) = a \cdot 1 + b \cdot x + c \cdot x^2$$

In this case, S is a spanning set for $\mathcal{P}_2(\mathbb{R})$. In general, this can be extended to $\mathcal{P}_n(\mathbb{R})$. In this case, if

$$S = \{1, x, x^2, ..., x^n\} \subseteq \mathcal{P}_n(\mathbb{R})$$

Then $\text{span}(S) = \mathcal{P}_n(\mathbb{R})$.

We now present a theorem that says that the span of any set of vectors is a subspace, and in fact, it is the smallest subspace containing such vectors.

Theorem 1.2.14. *Let V be a vector space over $\mathbb{F}$ and let $W = \text{span}\{\vec{v}_1, \vec{v}_2, ..., \vec{v}_n\}$. Then*

1. *W is a subspace of V containing each of $\vec{v}_1, \vec{v}_2, ..., \vec{v}_n$.*
2. *W is the smallest subspace containing these vectors in the sense that any subspace that contains each of $\vec{v}_1, \vec{v}_2, ..., \vec{v}_n$ must contain W.*

The following illustrates how (2) of Theorem 1.2.14 works.

Example 1.2.15. Let $V = \mathcal{P}_3(\mathbb{R})$ and let

$$S = \{x^2 + x^3, x, 2x^2 + 1, 3\}$$

We claim that $\text{span}(S) = \mathcal{P}_3(\mathbb{R})$. To see this, we need to show that $\text{span}(S) \subseteq \mathcal{P}_3(\mathbb{R})$ and $\mathcal{P}_3(\mathbb{R}) \subseteq \text{span}(S)$.

For the first, note that any linear combination of vectors in S belong to $\mathcal{P}_3(\mathbb{R})$ because if $a, b, c, d \in \mathbb{R}$ are such that

$$p(x) = a(x^2 + x^3) + bx + c(2x^2 + 1) + 3d$$

Then see that

$$\begin{aligned} p(x) &= ax^2 + ax^3 + bx + 2cx^2 + c + 3d \\ &= ax^3 + (a + c)x^2 + bx + (c + 3d) \end{aligned}$$

and we note from Example 1.2.13, $p(x)$ is a linear combination of $\{1, x, x^2, x^3\}$, so we have shown that $\text{span}(S) \subseteq \mathcal{P}_3(\mathbb{R})$.

For the second, let $p(x) = ax^3 + bx^2 + cx + d \in \mathcal{P}_3(\mathbb{R})$ be arbitrary. We need to show that $p(x)$ can be expressed as a linear combination of vectors in S. So we have for some $e, f, g, h \in \mathbb{R}$,

$$e(x^2 + x^3) + fx + g(2x^2 + 1) + 3h = ax^3 + bx^2 + cx + d$$

Using the above, we have on the left side

$$ex^3 + (e + g)x^2 + fx + (g + 3h) = ax^3 + bx^2 + cx + d$$

So comparing the coefficients, we have $a = e$, $b = e + g$, $c = f$, and $d = g + 3h$. This shows that it is possible to express any arbitrary polynomial $p(x) = ax^3 + bx^2 + cx + d$ as a linear combination of vectors in S, so we have $\mathcal{P}_3(\mathbb{R}) \subseteq \text{span}(S)$.

Therefore, since $\text{span}(S) \subseteq \mathcal{P}_3(\mathbb{R})$ and $\mathcal{P}_3(\mathbb{R}) \subseteq \text{span}(S)$, then $\text{span}(S) = \mathcal{P}_3(\mathbb{R})$, as desired.

Example 1.2.16. Let V be a vector space over $\mathbb{F}$, and let $\vec{u}, \vec{v} \in V$. We claim that

$$\operatorname{span}\{\vec{u}, \vec{v}\} = \operatorname{span}\{\vec{u} + 2\vec{v}, \vec{u} - \vec{v}\}$$

To see this we need to show "$\subseteq$" and "$\supseteq$".

To see the latter, it is obvious that $\operatorname{span}\{\vec{u} + 2\vec{v}, \vec{u} - \vec{v}\} \subseteq \operatorname{span}\{\vec{u}, \vec{v}\}$ by Theorem 1.2.14 since both $\vec{u} + 2\vec{v}$ and $\vec{u}, \vec{v}$ lie in $\operatorname{span}\{\vec{u}, \vec{v}\}$.

To see the former, note that

$$\vec{u} = \frac{1}{3}(\vec{u} + 2\vec{v}) + \frac{2}{3}(\vec{u} - \vec{v})$$

and

$$\vec{v} = \frac{1}{3}(\vec{u} + 2\vec{v}) - \frac{1}{3}(\vec{u} - \vec{v})$$

so $\operatorname{span}\{\vec{u}, \vec{v}\} \subseteq \operatorname{span}\{\vec{u} + 2\vec{v}, \vec{u} - \vec{v}\}$ by Theorem 1.2.14, as desired.

1.3 Linear Independence and Bases

In the previous sections, we studied how vector spaces can be constructed from smaller subsets using linear combinations and spanning sets. However, not all spanning sets are equally efficient or informative. Some may contain redundant vectors—vectors that can be written as combinations of others in the set. To identify the most efficient way to describe a vector space, we turn to the concept of *linear independence*.

Definition 1.3.1. Let V be a vector space over $\mathbb{F}$ and let $\{\vec{v}_1, \vec{v}_2, ..., \vec{v}_n\} \subseteq V$. We say that $\{\vec{v}_1, \vec{v}_2, ..., \vec{v}_n\}$ is *linearly independent* if for $\alpha_1, \alpha_2, ..., \alpha_n \in \mathbb{F}$ are such that

$$\alpha_1\vec{v}_1 + \alpha_2\vec{v}_2 + \cdots + \alpha_n\vec{v}_n = \vec{0}_V \implies \alpha_k = 0 \text{ for all } k = 1, 2, ..., n$$

If a set of vectors are not linearly independent, then we say that the set is *linearly dependent*.

Example 1.3.2. Let $V = \mathbb{R}^2$, and let $\vec{v}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\vec{v}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$. We claim that $\{\vec{v}_1, \vec{v}_2\}$ is linearly independent. To see this, assume that for $\alpha_1, \alpha_2 \in \mathbb{R}$ we have

$$\alpha_1\vec{v}_1 + \alpha_2\vec{v}_2 = \vec{0}$$

Then

$$\begin{bmatrix} \alpha_1 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ \alpha_2 \end{bmatrix} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Comparing entries, we see that $\alpha_1 = 0$ and $\alpha_2 = 0$. In general, the set

$$S = \left\{ \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \dots, \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix} \right\} \subseteq \mathbb{R}^n$$

is linearly independent.

Example 1.3.3. Let $V = \mathcal{M}_2(\mathbb{R})$ and let

$$S = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\} \subseteq \mathcal{M}_2(\mathbb{R})$$

We claim that S is a linearly independent subset of $\mathcal{M}_2(\mathbb{R})$. To see this, assume that for $\alpha_1, \alpha_2, \alpha_3, \alpha_4 \in \mathbb{R}$, we have that

$$\alpha_1 \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + \alpha_2 \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} + \alpha_3 \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} + \alpha_4 \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

Then

$$\begin{bmatrix} \alpha_1 & \alpha_2 \\ \alpha_3 & \alpha_4 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

So comparing entries, we see that $\alpha_1 = 0$, $\alpha_2 = 0$, $\alpha_3 = 0$, and $\alpha_4 = 0$. In general, the set

$$S = \left\{ \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}, \dots, \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}, \right\} \subseteq \mathcal{M}_n(\mathbb{R})$$

is linearly independent.

Example 1.3.4. Let $V = \mathcal{P}_2(\mathbb{R})$ and let

$$S = \{1, x, x^2\} \subseteq \mathcal{P}_2(\mathbb{R})$$

We claim that S is a linearly independent subset of $\mathcal{P}_2(\mathbb{R})$. To see this, assume that for $\alpha_1, \alpha_2, \alpha_3 \in \mathbb{R}$, we have

$$\alpha_1 \cdot 1 + \alpha_2 \cdot x + \alpha_3 \cdot x^2 = 0$$

Then observe that

$$\alpha_1 + \alpha_2 x + \alpha_3 x^2 = 0 + 0x + 0x^2$$

So comparing coefficients, we see that $\alpha_1 = 0$, $\alpha_2 = 0$, and $\alpha_3 = 0$. In general, the set

$$S = \{1, x, x^2, ..., x^n\} \subseteq \mathcal{P}_n(\mathbb{R})$$

is linearly independent.

Example 1.3.5. Let $V = \mathcal{P}_2(\mathbb{R})$, and let $S = \{1+x, 3x+x^2, 2+x-x^2\} \subseteq \mathcal{P}_2(\mathbb{R})$. We claim that S is a linearly independent subset of $\mathcal{P}_2(\mathbb{R})$. To see this, assume that for $\alpha_1, \alpha_2, \alpha_3 \in \mathbb{R}$, we have

$$\alpha_1(1+x) + \alpha_2(3x+x^2) + \alpha_3(2+x-x^2) = 0$$

Then rearranging the equation so that

$$(\alpha_2 - \alpha_3)x^2 + (\alpha_1 + 3\alpha_2 + \alpha_3)x + (\alpha_1 + 2\alpha_3) = 0x^2 + 0x + 0$$

Now comparing the coefficients, we have the system of equations given by

$$\begin{cases} \alpha_2 - \alpha_3 = 0 \\ \alpha_1 + 3\alpha_2 + \alpha_3 = 0 \\ \alpha_1 + 2\alpha_3 = 0 \end{cases}$$

One can check that the unique solution to this system is when $\alpha_1 = 0$, $\alpha_2 = 0$, and $\alpha_3 = 0$. So the set S is linearly independent.

Example 1.3.6. Let $V = \mathcal{C}[0, 2\pi]$, and let $S = \{\sin(x), \cos(x)\} \subseteq \mathcal{C}[0, 2\pi]$. We claim that S is linearly independent. To see this, assume that for $\alpha_1, \alpha_2 \in \mathbb{R}$,

$$\alpha_1 \sin(x) + \alpha_2 \cos(x) = 0$$

This must hold for all values of $x \in [0, 2\pi]$. Taking $x = 0$ yields $\alpha_2 = 0$ and taking $x = \frac{\pi}{2}$ yields $\alpha_1 = 0$.

Example 1.3.7. Let $V = \mathbb{R}^2$, and let $S = \left\{ \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 2 \\ 4 \end{bmatrix} \right\}$. We claim that S is linearly dependent. To see this, assume that for $\alpha_1, \alpha_2 \in \mathbb{R}$, we have

$$\alpha_1 \begin{bmatrix} 1 \\ 2 \end{bmatrix} + \alpha_2 \begin{bmatrix} 2 \\ 4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Then

$$\begin{bmatrix} \alpha_1 + 2\alpha_2 \\ 2\alpha_1 + 4\alpha_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

So comparing entries, we see that $\alpha_1 + 2\alpha_2 = 0$ and $2\alpha_1 + 4\alpha_2 = 0$, but then $\alpha_1 = -2\alpha_2$ and not always zero, so S has to be linearly dependent.

Example 1.3.8. We claim that the set of polynomials of distinct degrees is independent. To see this, let $p_1, p_2, ..., p_n$ be polynomials such that $\deg(p_k) = d_k$ for each k. By relabelling if necessary, we may assume that $d_1 > d_2 > \cdots > d_n$. Suppose that for $\alpha_1, \alpha_2, ..., \alpha_n \in \mathbb{R}$,

$$\alpha_1 p_1 + \alpha_2 p_2 + \cdots + \alpha_n p_n = 0$$

Since $\deg(p_1) = d_1$, let ax^{d_1} be the term in p_1 of highest degree with $a \neq 0$. Since $d_1 > d_2 > \cdots > d_n$, it follows that $\alpha_1 ax^{d_1}$ is the only term of degree d_1 in the linear combination $\alpha_1 p_1 + \alpha_2 p_2 + \cdots + \alpha_n p_n = 0$. This implies that $\alpha_1 ax^{d_1} = 0$, so $\alpha_1 = 0$. But then $\alpha_2 p_2 + \alpha_3 p_3 + \cdots + \alpha_n p_n = 0$, so we can repeat the argument to show that $\alpha_2 = 0$, and so on.

A set of vectors is linearly independent if $\vec{0}$ is a linear combination in a unique way. The following proposition shows that every linear combination of these vectors has uniquely determined coefficients.

Proposition 1.3.9. *Let V be a vector space over $\mathbb{F}$ and let $\{\vec{v}_1, \vec{v}_2, ..., \vec{v}_n\} \subseteq V$ be linearly independent. If $\vec{v}$ has two representations*

$$\begin{aligned} \vec{v} &= \alpha_1 \vec{v}_1 + \alpha_2 \vec{v}_2 + \cdots + \alpha_n \vec{v}_n \\ \vec{v} &= \beta_1 \vec{v}_1 + \beta_2 \vec{v}_2 + \cdots + \beta_n \vec{v}_n \end{aligned}$$

as a linear combination of these vectors, then for all $k = 1, 2, ..., n$, we have $\alpha_k = \beta_k$.

Proof. Note that by subtracting both equations we yield

$$\vec{0} = (\alpha_1 - \beta_1)\vec{v}_1 + (\alpha_2 - \beta_2)\vec{v}_2 + \cdots + (\alpha_n - \beta_n)\vec{v}_n$$

Then by linear independence, we have $\alpha_k - \beta_k = 0$ for each k, so $\alpha_k = \beta_k$ for each k, as desired. $\square$

The following theorem is one of the most useful results in linear algebra.

Theorem 1.3.10. *Let V be a a vector space over $\mathbb{F}$. Assume that V can be spanned by n vectors. If any set of m vectors in V is linearly independent, then $m \leq n$.*

Proof. Since V can be spanned by n vectors, let $V = \text{span}\{\vec{v}_1, \vec{v}_2, ..., \vec{v}_n\}$, and assume that $\{\vec{u}_1, \vec{u}_2, ..., \vec{u}_m\}$ is a linearly independent set in V. Then we can write

$$\vec{u}_1 = \alpha_1\vec{v}_1 + \alpha_2\vec{v}_2 + \cdots + \alpha_n\vec{v}_n$$

where each $\alpha_k \in \mathbb{R}$. As $\vec{u}_1 \neq \vec{0}_V$, not all of the α_k are zero. Without loss of generality, say that $\alpha_1 \neq 0$, after relabelling the $\vec{v}_k$. Then $V = \text{span}\{\vec{u}_1, \vec{v}_2, ..., \vec{v}_n\}$. Similarly, write

$$\vec{u}_2 = \beta_1\vec{v}_1 + \gamma_2\vec{v}_2 + \cdots + \gamma_n\vec{v}_n$$

Then some $\gamma_k \neq 0$ since $\{\vec{u}_1, \vec{u}_2\}$ is linearly independent, so $V = \text{span}\{\vec{u}_1, \vec{u}_2, ..., \vec{v}_n\}$. Proceeding inductively, if $m > n$, this procedure continues until all the vectors $\vec{v}_k$ are replaced by the vectors $\vec{u}_1, \vec{u}_2, ..., \vec{u}_n$. In particular, $V = \text{span}\{\vec{u}_1, \vec{u}_2, ..., \vec{u}_n\}$. But then $\vec{u}_{n+1}$ is a linear combination of $\vec{u}_1, \vec{u}_2, ..., \vec{u}_n$, contrary to the linear independence of the $\vec{u}_k$. Hence, we cannot have $m > n$, so $m \leq n$, as desired. □

We now introduce the concept of a basis of a vector space.

Definition 1.3.11. Let V be a vector space over $\mathbb{F}$. A subset $\mathcal{B} = \{\vec{e}_1, \vec{e}_2, ..., \vec{e}_n\} \subseteq V$ is called a *basis* for V if

1. $\mathcal{B}$ is linearly independent.
2. $V = \text{span}(\mathcal{B})$.

In other words, $\mathcal{B}$ is a basis of V if $\mathcal{B}$ is a linearly independent spanning set.

Example 1.3.12. Let $V = \mathbb{F}^n$. The set

$$\mathcal{B} = \left\{ \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, ..., \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix} \right\} \subseteq \mathbb{R}^n$$

is a basis for $\mathbb{F}^n$. From Example 1.2.11, we have shown that $\mathcal{B}$ is a spanning set for $\mathbb{R}^n$, and in Example 1.3.2, we have shown that $\mathcal{B}$ is linearly independent. We call $\mathcal{B}$ the *standard basis* of $\mathbb{F}^n$.

Example 1.3.13. Let $V = \mathcal{M}_n(\mathbb{F})$. The set

$$\mathcal{B} = \left\{ \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}, ..., \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}, \right\} \subseteq \mathcal{M}_n(\mathbb{R})$$

is a basis for $\mathcal{M}_n(\mathbb{F})$. From Example 1.2.12, we have shown that $\mathcal{B}$ is a spanning set for $\mathcal{M}_n(\mathbb{F})$, and in Example 1.3.3, we have shown that $\mathcal{B}$ is linearly independent. We call $\mathcal{B}$ the *standard basis* of $\mathcal{M}_n(\mathbb{F})$.

Example 1.3.14. Let $V = \mathcal{P}_n(\mathbb{F})$. The set

$$\mathcal{B} = \{1, x, x^2, ..., x^n\} \subseteq \mathcal{P}_n(\mathbb{R})$$

is a basis for $\mathcal{P}_n(\mathbb{F})$. From Example 1.2.13, we have shown that $\mathcal{B}$ is a spanning set for $\mathcal{P}_n(\mathbb{F})$, and in Example 1.3.4, we have shown that $\mathcal{B}$ is linearly independent. We call $\mathcal{B}$ the *standard basis* of $\mathcal{P}_n(\mathbb{F})$.

Thus, based on the examples above, if $\mathcal{B}$ is a basis for V, then every vector in V can be written as a linear combination of these vectors in a unique way. But even more, any two bases of V must contain the same number of vectors.

Proposition 1.3.15. *Let V be a vector space over $\mathbb{F}$, and suppose that $\mathcal{B}_1$ is a basis for V that has n elements and $\mathcal{B}_2$ is a basis for V that has m elements. Then $n = m$.*

Proof. Since $\mathcal{B}_1$ is a basis for V that contains n elements, and $\mathcal{B}_2$ is a basis for V that contains m elements, it follows from Theorem 1.3.10 that $m \leq n$. Similarly, $n \leq m$, so $n = m$, as asserted. □

Proposition 1.3.15 guarantees that no matter which basis V is chosen it contains the same number of vectors as any other basis. Hence, there is no ambiguity about the following definition.

Definition 1.3.16. Let V be a vector space over $\mathbb{F}$ and let $\mathcal{B}$ be a basis of V. The number of vectors in the basis $\mathcal{B}$ is called the *dimension* of V, and is denoted by

$$\dim(V) = |\mathcal{B}|$$

If $\mathcal{B}$ contains a finite number of elements, then we say that V is a *finite dimensional vector space*. Otherwise, if $\mathcal{B}$ contains an infinite number of elements, then we say that V is an *infinite dimensional vector space.*

Knowing more about infinite dimensional vector spaces would be nice, but it is not the focus of the course, but we mention it for our own understanding that if we have a vector space over $\mathbb{F}$, it is possible that we could have an infinite dimensional vector space. For example, the sequence space $\mathbb{F}^{\mathbb{N}}$ is an example of an infinite dimensional vector space.

Example 1.3.17. If $V = \mathbb{R}^n$, then the set $\mathcal{B}$ in Example 1.3.12 contains n elements, so $\dim(V) = n$.

Example 1.3.18. If $V = \mathcal{M}_n(\mathbb{R})$, then the set $\mathcal{B}$ in Example 1.3.13 contains n^2 elements, so $\dim(V) = n^2$.

Example 1.3.19. If $V = \mathcal{P}_n(\mathbb{R})$, then the set $\mathcal{B}$ in Example 1.3.14 contains $n+1$ elements, so $\dim(V) = n+1$.

The question that one may ask is, how do we construct a basis for a vector space? Proposition 1.3.15 tells us that we can have more than one basis for a vector space, on the condition that we make sure that this new basis contains the same number of elements. The following example illustrates how we can construct a basis for a given vector space.

Example 1.3.20. Let $A = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$ and consider the subspace of $\mathcal{M}_2(\mathbb{R})$ given by

$$U(A) = \{B \in \mathcal{M}_2(\mathbb{R}) : AB = BA\} \subseteq \mathcal{M}_2(\mathbb{R})$$

We first need to find a basis for $\mathcal{M}_2(\mathbb{R})$. Start with an arbitrary 2×2 matrix in $U(A)$, so say $B = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in U(A)$. Then by the given condition that $AB = BA$, we have that

$$\begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} a+c & b+d \\ 0 & 0 \end{bmatrix}$$

and

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} a & a \\ c & c \end{bmatrix}$$

Thus,

$$\begin{bmatrix} a+c & b+d \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} a & a \\ c & c \end{bmatrix}$$

Now, this would imply that $c = 0$, $a = b + d$, so

$$\begin{aligned} B &= \begin{bmatrix} b+d & b \\ 0 & d \end{bmatrix} \\ &= \begin{bmatrix} b & b \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} d & 0 \\ 0 & d \end{bmatrix} \\ &= b \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} + d \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \end{aligned}$$

Therefore, as $B \in U(A)$ was arbitrary and B is a linear combination of matrices $\begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$ and $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, we claim that

$$\mathcal{B} = \left\{ \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right\}$$

is a basis for $U(A)$. We have already showed that $\mathcal{B}$ is a spanning set for $U(A)$, so it suffices to show that $\mathcal{B}$ is linearly independent.

Assume that for $\alpha_1, \alpha_2 \in \mathbb{R}$, we have

$$\alpha_1 \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} + \alpha_2 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

Then

$$\begin{bmatrix} \alpha_1 + \alpha_2 & \alpha_1 \\ 0 & \alpha_2 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

Comparing entries, we easily see that $\alpha_1 = 0$ and $\alpha_2 = 0$, so $\mathcal{B}$ is indeed linearly independent. Therefore, $\mathcal{B}$ is indeed a basis of $U(A)$. Since $\mathcal{B}$ contains two elements, then $\dim(U(A)) = 2$.

Example 1.3.21. Let $V = \mathcal{M}_2(\mathbb{R})$, and let

$$W = \{A \in \mathcal{M}_2(\mathbb{R}) : A^T = A\}$$

i.e. W is the set of all 2×2 matrices such that A is symmetric. We want to find a basis for W and find its dimension, so let $A \in W$ be arbitrary, so $A^T = A$. Such matrices A takes the form

$$A = \begin{bmatrix} a & b \\ b & c \end{bmatrix}$$

for $a, b, c \in \mathbb{R}$. Then observe that

$$A = \begin{bmatrix} a & b \\ b & c \end{bmatrix} = a \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + b \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} + c \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

Therefore, A is a linear combination of the matrices $\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$, $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$, and $\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$. Thus, define

$$\mathcal{B} = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\}$$

We claim that $\mathcal{B}$ is a basis for W. Indeed, since any matrix in W is a linear combination of matrices in $\mathcal{B}$, $\mathcal{B}$ is a spanning set for W. To see that $\mathcal{B}$ is linearly independent, assume that for $\alpha_1, \alpha_2, \alpha_3 \in \mathbb{R}$, we have

$$\alpha_1 \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + \alpha_2 \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} + \alpha_3 \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

So then,

$$\begin{bmatrix} \alpha_1 & \alpha_2 \\ \alpha_2 & \alpha_3 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

and we easily see that by comparing the entries, $\alpha_1 = 0$, $\alpha_2 = 0$, and $\alpha_3 = 0$, so $\mathcal{B}$ is indeed linearly independent. Since $\mathcal{B}$ has three elements, $\dim(W) = 3$.

Up until this point, we only had a look at examples involving vector spaces over $\mathbb{R}$. Now let us have a look at some examples that involve vector spaces over $\mathbb{C}$.

Example 1.3.22. Let $V = \mathbb{C}^3$, and define a subspace

$$W = \left\{ \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} \in \mathbb{C}^3 : z_1 + iz_2 = 0 \right\}$$

We want to find a basis for W and the dimension of W, so let $\begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix}$ be arbitrary, so $z_1 + iz_2 = 0$, and thus, $z_1 = -iz_2$. Then observe that

$$\begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} = \begin{bmatrix} -iz_2 \\ z_2 \\ z_3 \end{bmatrix} = z_2 \begin{bmatrix} -i \\ 1 \\ 0 \end{bmatrix} + z_3 \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

Therefore, $\begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix}$ is a linear combination of the vectors $\begin{bmatrix} -i \\ 1 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$. Thus, define

$$\mathcal{B} = \left\{ \begin{bmatrix} -i \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}$$

We claim that $\mathcal{B}$ is a basis for W. Indeed, since any vector in W is a linear combination of the vectors in $\mathcal{B}$, $\mathcal{B}$ is a spanning set for W. To see that $\mathcal{B}$ is linearly independent, assume that for $\alpha_1, \alpha_2 \in \mathbb{C}$, we have

$$\alpha_1 \begin{bmatrix} -i \\ 1 \\ 0 \end{bmatrix} + \alpha_2 \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Then

$$\begin{bmatrix} -i\alpha_1 \\ \alpha_1 \\ \alpha_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Thus, comparing entries, we easily see that $\alpha_1 = 0$ and $\alpha_2 = 0$, and so $\mathcal{B}$ is linearly independent, as claimed. Since $\mathcal{B}$ has two elements, $\dim(W) = 2$.

Example 1.3.23. Let $V = \mathcal{P}_3(\mathbb{C})$ and consider the subspace

$$U = \{p(x) \in \mathcal{P}_3(\mathbb{C}) : p(1+i) = 0 \text{ and } p(1-i) = 0\}$$

We want to find a basis for U, and the dimension of U. Let $p(x) \in U$ be arbitrary, so $p(1+i) = 0$ and $p(1-i) = 0$. Then $1+i$ and $1-i$ are factors of the polynomial $p(x)$, so there is some polynomial $q(x)$ such that

$$p(x) = (x - (1+i))(x - (1-i))q(x)$$

where $\deg(q(x)) \leq 1$ as $\deg(p(x)) \leq 3$. In this case, we can write $q(x) = ax+b$ for some $a, b \in \mathbb{C}$, and so

$$\begin{aligned} p(x) &= (x^2 - 2x + 2)(ax + b) \\ &= ax^3 - 2ax^2 + 2ax + bx^2 - 2bx + 2b \\ &= a(x^3 - 2x^2 + 2x) + b(x^2 - 2x + 2) \end{aligned}$$

Therefore, $p(x)$ is a linear combination of the polynomials $x^3 - 2x^2 + 2x$ and $x^2 - 2x + 2$. Thus, define

$$\mathcal{B} = \{x^3 - 2x^2 + 2x, x^2 - 2x + 2\}$$

We claim that $\mathcal{B}$ is a basis for U. Indeed, since any polynomial in U is a linear combination of polynomials in $\mathcal{B}$, then $\mathcal{B}$ is indeed a spanning set for U. To see that $\mathcal{B}$ is linearly independent, assume that for $\alpha_1, \alpha_2 \in \mathbb{C}$, we have

$$\alpha_1(x^3 - 2x^2 + 2x) + \alpha_2(x^2 - 2x + 2) = 0$$

Then rearranging

$$\alpha_1 x^3 + (-2\alpha_1 + \alpha_2)x^2 + (2\alpha_1 - 2\alpha_2)x + 2\alpha_2 = 0$$

So comparing coefficients, we have

$$\begin{cases} \alpha_1 = 0 \\ -2\alpha_1 + \alpha_2 = 0 \\ 2\alpha_1 - 2\alpha_2 = 0 \\ 2\alpha_2 = 0 \end{cases}$$

Therefore, we easily see that then $\alpha_1 = 0$ and $\alpha_2 = 0$, so $\mathcal{B}$ is indeed linearly independent. Since $\mathcal{B}$ has two elements, $\dim(U) = 2$.

1.4 Finite Dimensional Bases

We have introduced the definition of finite dimensional and infinite dimensional vector spaces in the previous section, but up to this point, we had no guarantee that an arbitrary vector space has a basis, and hence, no guarantee that one can speak at all of the dimension of V. The following theorem shows that any space that is spanned by a finite set of vectors has a finite basis.

Lemma 1.4.1. *Let V be a vector space over $\mathbb{F}$ and let $\{\vec{v}_1, \vec{v}_2, ..., \vec{v}_n\}$ be an independent set of vectors in V. If $\vec{u} \in V \setminus \operatorname{span}\{\vec{v}_1, \vec{v}_2, ..., \vec{v}_n\}$, then $\{\vec{u}, \vec{v}_1, ..., \vec{v}_n\}$ is linearly independent.*

Proof. Assume that for $\alpha, \alpha_1, \alpha_2, ..., \alpha_n \in \mathbb{F}$, we have

$$\alpha\vec{u} + \alpha_1\vec{v}_1 + \cdots + \alpha_n\vec{v}_n = \vec{0}_V$$

First, note that $\alpha = 0$ since otherwise,

$$\vec{u} = -\frac{\alpha_1}{\alpha}\vec{v}_1 - \frac{\alpha_2}{\alpha}\vec{v}_2 - \cdots - \frac{\alpha_n}{\alpha}\vec{v}_n \in \operatorname{span}\{\vec{v}_1, \vec{v}_2, ..., \vec{v}_n\}$$

which is a contradiction. Therefore, we then have $\alpha_1\vec{v}_1 + \alpha_2\vec{v}_2 + \cdots + \alpha_n\vec{v}_n = \vec{0}_V$, and by assumption, $\alpha_k = 0$ for all $k = 1, 2, ..., n$. □

Remark 1.4.2. Note that the converse of Lemma 1.4.1 is also true, so if $\{\vec{u}, \vec{v}_1, ..., \vec{v}_n\}$ is linearly independent, then $\vec{u} \notin \operatorname{span}\{\vec{v}_1, \vec{v}_2, ..., \vec{v}_n\}$.

The question that we may ask ourselves is: if we have a linearly independent set that contains fewer elements than the dimension of a vector space, can we keep adding vectors to this set so that it eventually becomes a basis of the space?

The answer is yes! In finite-dimensional vector spaces, every linearly independent set can be extended to a basis. This powerful and foundational result guarantees that any "partial basis" (a linearly independent set) can be completed into a full basis.

To do so, we require the following lemma.

Lemma 1.4.3. *Let V be a finite-dimensional vector space over $\mathbb{F}$. If W is a subspace of V, then any independent subset of W can be enlarged to a finite basis of W.*

Proof. Suppose that A is a linearly independent subset of W. If $\operatorname{span}(A) = W$, then A is already a basis of W. Otherwise, if $\operatorname{span}(A) \neq W$, then choose $\vec{w}_1 \in W$ such that $\vec{w}_1 \notin \operatorname{span}(A)$. Then the set $A \cup \{\vec{w}_1\}$ is linearly independent by Lemma 1.4.1.

Now if $\operatorname{span}(A \cup \{\vec{w}_1\}) = W$, then we are done. Otherwise, if $\operatorname{span}(A \cup \{\vec{w}_1\}) \neq W$, then choose $\vec{w}_2 \in W$ such that $\vec{w}_2 \notin \operatorname{span}(A \cup \{\vec{w}_1\})$. Then the set $A \cup \{\vec{w}_1, \vec{w}_2\}$ is linearly independent by Lemma 1.4.1.

By proceeding inductively, we claim that a basis of W will be reached eventually. To see this, if no basis of W is ever reached, then the process would create arbitrary large independent sets in V. This is not possible by Theorem 1.3.10 since V is finite-dimensional and so is spanned by a finite set of vectors. $\square$

Theorem 1.4.4. *Let V be a finite-dimensional vector space over $\mathbb{F}$ spanned by $\{\vec{v}_1, \vec{v}_2, ..., \vec{v}_n\}$. The following hold.*

1. *V has a finite basis and $\dim(V) \leq n$.*
2. *Every independent set of vectors in V can be enlarged to a basis of V by adding vectors from any fixed basis of V.*
3. *If W is a subspace of V, then*
   (a) *W is a finite dimensional subspace and $\dim(W) \leq \dim(V)$.*
   (b) *If $\dim(W) = \dim(V)$, then $W = V$.*

Proof. To see that (1) holds, if $V = \{\vec{0}_V\}$, then V has an empty basis, so $\dim(V) = 0 \leq n$. Otherwise, if $\vec{v} \neq \vec{0}_V$, then $\{\vec{v}\}$ is linearly independent, so (1) follows from Lemma 1.4.3 with $W = V$.

To see that (2) holds, let $\mathcal{B}$ be a basis of V and let A be a linearly independent subset of V. If $\operatorname{span}(A) = V$, then A is a basis for V. Otherwise, $\mathcal{B}$ is not contained in A since $\mathcal{B}$ spans V, so choose $\vec{w}_1 \in \mathcal{B} \setminus \operatorname{span}(A)$ so that

$A \cup \{\vec{w}_1\}$ is linearly independent by Lemma 1.4.1. If $\operatorname{span}(A \cup \{\vec{w}_1\}) = V$, then we are done. Otherwise, a similar argument shows that $A \cup \{\vec{w}_1, \vec{w}_2\}$ is linearly independent for some $\vec{w}_2 \in \mathcal{B}$. Continuing this process, as in the proof of Lemma 1.4.3, a basis of V will be reached eventually.

To see that (3) part (a) holds, note that if $W = \{\vec{0}_V\}$, then this is easy. Otherwise, let $\vec{w} \neq \vec{0}_V \in W$. Then $\{\vec{w}\}$ can be enlarged to a finite basis $\mathcal{B}$ of W by Lemma 1.4.3, so W is finite-dimensional. But $\mathcal{B}$ is also linearly independent, so $\dim(W) \leq \dim(V)$ by Theorem 1.3.10. To see that part (b) holds, if $W = \{\vec{0}_V\}$, then this is trivial since V has a basis. Otherwise, it follows from (2). $\square$

Example 1.4.5. Let $V = \mathcal{M}_2(\mathbb{R})$ and consider the set

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

Right now, this is not a basis for $\mathcal{M}_2(\mathbb{R})$ just yet since $\dim(\mathcal{M}_2(\mathbb{R})) = 4$, we are missing one element. Recall that the standard basis for $\mathcal{M}_2(\mathbb{R})$ is given by

$$\mathcal{SB}_{\mathcal{M}_2(\mathbb{R})} = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\}$$

So including one of these in A will produce a basis by Theorem 1.4.4. In fact, *any* one of these matrices in A produces a linearly independent set, and hence, a basis, as we will mention later below.

Example 1.4.6. Let $V = \mathcal{M}_2(\mathbb{C})$ and consider the set

$$A = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & i \\ 0 & 0 \end{bmatrix} \right\}$$

Right now, this is not a basis for $\mathcal{M}_2(\mathbb{C})$ just yet, since $\dim(\mathcal{M}_2(\mathbb{C})) = 4$, we are missing two elements. To make it into a basis, we can simply add two standard basis elements to A to make it into a basis, for example,

$$\mathcal{B} = A \cup \left\{ \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\}$$

$\mathcal{B}$ is still a linearly independent set, and moreover, it is now a basis for $\mathcal{M}_2(\mathbb{C})$.

Example 1.4.7. Let $V = \mathcal{P}_3(\mathbb{R})$ and consider the linearly independent set $A = \{1+x, 1+x^2\}$. Suppose we want to find a basis containing A. Recall that the standard basis for $\mathcal{P}_3(\mathbb{R})$ is given by

$$\mathcal{SB}_{\mathcal{P}_3(\mathbb{R})} = \{1, x, x^2, x^3\}$$

so including two of these vectors will do. If we use 1 and x^3, then we have $\mathcal{B} = \{1, 1+x, 1+x^2, x^3\}$, and this is independent because the polynomials have distinct degrees, and so $\mathcal{B}$ is a basis by Theorem 1.4.4. However, note that if we add $\{1, x\}$ or $\{1, x^2\}$ instead, this would not work!

Example 1.4.8. Let $\mathcal{P}(\mathbb{F}) = \bigcup_{n=1}^{\infty} \mathcal{P}_n(\mathbb{F})$ be the set of all polynomials. It is elementary to verify that $\mathcal{P}(\mathbb{F})$ is a vector space over $\mathbb{F}$. We claim that $\mathcal{P}(\mathbb{F})$ is infinite-dimensional. To see this, note that for each $n \in \mathbb{N}$, $\mathcal{P}(\mathbb{F})$ has a subspace $\mathcal{P}_n(\mathbb{F})$ of dimension $n+1$.

Suppose for a contradiction that $\mathcal{P}(\mathbb{F})$ is finite dimensional, say $\dim(\mathcal{P}(\mathbb{F})) = m$. Then $\dim(\mathcal{P}_n(\mathbb{F})) \leq \dim(\mathcal{P}(\mathbb{F}))$ by Theorem 1.4.4, so $n+1 \leq m$. This is absurd since n is arbitrary, so $\mathcal{P}(\mathbb{F})$ must be infinite dimensional.

Proposition 1.4.9. *Let V be a finite-dimensional vector space over $\mathbb{F}$, and let U and W be subspaces of V.*

1. *If $U \subseteq W$, then $\dim(U) \leq \dim(W)$.*
2. *If $U \subseteq W$ and $\dim(U) = \dim(W)$, then $U = W$.*

Proof. Since W is finite-dimensional, (1) follows by taking $V = W$ in part (3) of Theorem 1.4.4. Now, assume that $\dim(U) = \dim(W) = n$, and let $\mathcal{B}$ be a basis for U. Then $\mathcal{B}$ is a linearly independent set in W. If $U \neq W$, then $\operatorname{span}(\mathcal{B}) \neq W$, so $\mathcal{B}$ can be extended to a linearly independent set of $n+1$ vectors in W, by Lemma 1.4.1. This contradicts Theorem 1.3.10, since W is spanned by $\dim(W) = n$ vectors. Hence, $U = W$, proving (2). □

Example 1.4.10. Let $V = \mathcal{P}_n(\mathbb{R})$, and let

$$W = \{p(x) \in \mathcal{P}_n(\mathbb{R}) : p(1) = 0\}$$

We claim that $\mathcal{B} = \{(x-1), (x-1)^2, ..., (x-1)^n\}$ is a basis for W. To see this, note that $(x-1), (x-1)^2, ..., (x-1)^n$ are all members of W, and that they are linearly independent because they have distinct degrees. Observe then that $\operatorname{span}(\mathcal{B}) \subseteq W \subseteq \mathcal{P}_n(\mathbb{R})$ and $\dim(\operatorname{span}(\mathcal{B})) = n$, and $\dim(\mathcal{P}_n(\mathbb{R})) = n+1$, so $n \leq \dim(W) \leq n+1$ by Proposition 1.4.9. Since $\dim(W)$ is an integer, we must have $\dim(W) = n$ or $\dim(W) = n+1$, but then $W = \operatorname{span}(\mathcal{B})$ or $W = \mathcal{P}_n(\mathbb{R})$ by Proposition 1.4.9. Since $W \neq \mathcal{P}_n(\mathbb{R})$, it follows that $W = \operatorname{span}(\mathcal{B})$, as required.

Lemma 1.4.11. *Let V be a vector space over $\mathbb{F}$ and let $A = \{\vec{v}_1, \vec{v}_2, ..., \vec{v}_n\} \subseteq V$. Then A is linearly dependent if and only if there exists some vector in A that is a linear combination of the others.*

Proof. Suppose $\vec{v}_2$ is a linear combination of $\vec{v}_1, \vec{v}_3, ..., \vec{v}_n$, so for some $\alpha_1, \alpha_3, ..., \alpha_n \in \mathbb{F}$,

$$\vec{v}_2 = \alpha_1\vec{v}_1 + \alpha_3\vec{v}_3 + \cdots + \alpha_n\vec{v}_N$$

Then

$$\alpha_1\vec{v}_1 + (-1)\vec{v}_2 + \cdots + \alpha_n\vec{v}_n = \vec{0}_V$$

is a nontrivial linear combination that vanishes, so A is linearly dependent.

Conversely, if A is linearly dependent, let

$$\alpha_1\vec{v}_1 + \alpha_2\vec{v}_2 + \cdots + \alpha_n\vec{v}_n = \vec{0}_V$$

where some coefficients is nonzero. If, say $\alpha_2 \neq 0$, then

$$\vec{v}_2 = -\frac{\alpha_1}{\alpha_2}\vec{v}_1 - \frac{\alpha_3}{\alpha_2}\vec{v}_3 - \cdots - \frac{\alpha_n}{\alpha_2}\vec{v}_n$$

is a linear combination of the others. □

Lemma 1.4.1 gives us a way to enlarge linearly independent sets to a basis, Theorem 1.4.12 shows that spanning sets can be cut down to a basis.

Theorem 1.4.12. *Let V be a finite-dimensional vector space over $\mathbb{F}$. Any spanning set for V can be cut down to a basis of V.*

Proof. Since V is finite-dimensional, it has a finite spanning set S. Among all spanning sets contained in S, choose S_0 containing the smallest number of vectors. It suffices to show that S_0 is linearly independent.

Assume for a contradiction that S_0 is linearly dependent. Then by Lemma 1.4.11, there exists a vector $\vec{u} \in S_0$ that is a linear combination of the set $S_1 = S_0 \setminus \{\vec{u}\}$ of vectors in S_0 other than $\vec{u}$. It follows that $\text{span}(S_0) = \text{span}(S_1)$, so $V = \text{span}(S_1)$, but S_1 has fewer elements than S_0, so this contradicts the choice of S_0, so S_0 is linearly independent, as claimed. □

Example 1.4.13. Let $V = \mathcal{P}_3(\mathbb{R})$, and consider

$$S = \{1, x + x^2, 2x - 3x^2, 1 + 3x - 2x^2, x^3\}$$

Right now, we have 5 elements in this set, and we know that $\dim(\mathcal{P}_3(\mathbb{R})) = 4$, so we need to eliminate an element from S. However, we note that we cannot remove x^3, since the span of the rest of S is contained in $\mathcal{P}_2(\mathbb{R})$, but by eliminating $1 + 3x - 2x^2$, we can get a basis, since $1 + 3x - 2x^2$ is the sum of the first three polynomials in S.

Theorem 1.4.4 and 1.4.12 have the following consequence.

Corollary 1.4.14. *Let V be a finite-dimensional vector space over $\mathbb{F}$ with $\dim(V) = n$ and let S be a set of exactly n vectors in V. Then S is linearly independent if and only if S spans V.*

Proof. Assume that S is linearly independent. Then by Theorem 1.4.4, S is contained in a basis $\mathcal{B}$ of V. Hence, $|S| = n = |\mathcal{B}|$ so, as $S \subseteq \mathcal{B}$, it follows that $S = B$, so S spans V.

Conversely, if S spans V, then S contains a basis $\mathcal{B}$ by Theorem 1.4.12. Thus, as $|S| = n = |\mathcal{B}|$, then $S \supseteq \mathcal{B}$, so $S = \mathcal{B}$. Therefore, S is linearly independent. $\square$

Example 1.4.15. Let $V = \mathcal{S}_2(\mathbb{R})$ denote the set of all 2×2 symmetric matrices with real entries and let $GL_2(\mathbb{R})$ denote the subspace of $\mathcal{S}_2(\mathbb{R})$ consisting of invertible matrices with real entries. Suppose we want to find a basis for $GL_2(\mathbb{R})$. From Example 1.3.21, we showed that $\dim(V) = 3$, so what is needed is a set of three invertible matrices that (using Corollary 1.4.14) is either independent or spans V. The set

$$\mathcal{B} = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \right\}$$

is indeed a linearly independent set, and spans $GL_2(\mathbb{R})$. We call $GL_n(\mathbb{R})$ the space of all general linear matrices with real entries, or all $n \times n$ matrices that are invertible.

Let V be a vector space over $\mathbb{F}$ and let U and W be subspaces of V. There are two subspaces that are of interest, their *sum* $U + W$, and their *intersection* $U \cap W$, defined by

$$U + W = \{\vec{u} + \vec{w} : \vec{u} \in U, \vec{w} \in W\}$$

and

$$U \cap W = \{\vec{v} \in V : \vec{v} \in U \text{ and } \vec{v} \in W\}$$

It is routine to verify that these are indeed subspaces of V, that $U \cap W$ is contained in both U and W, and that $U + W$ contains both U and W.

Theorem 1.4.16. *Let V be a vector space over $\mathbb{F}$, and let U and W be finite dimensional subspaces of V. Then $U+W$ is finite dimensional and*

$$\dim(U+W) = \dim(U) + \dim(W) - \dim(U \cap W)$$

Proof. Exercise. $\square$

Remark 1.4.17. Although we can have intersections of vector subspaces, it may not be the case that we have unions of vector subspaces. For example, suppose we have $V = \mathbb{R}^2$, and we let

$$U = \{(x,0) : x \in \mathbb{R}\} \quad \text{and} \quad W = \{(0,y) : y \in \mathbb{R}\}$$

i.e. U is the set of all points on the x-axis, and W is the set of all points on the y-axis. Their union $U \cup W$ is the set of all points on the x-axis or all points on the y-axis.

To see that $U \cup W$ is not a subspace of $\mathbb{R}^2$, we claim that it is not closed under vector addition. Consider the points $(3,0),(0,4) \in U \cup W$. But then the point

$$(3,4) = (3,0) + (0,4) \notin U \cup W$$

Therefore, as $(3,4) \notin U \cup W$, $U \cup W$ is not closed under vector addition, as claimed.

Chapter 2

Linear Transformations

In the previous chapter, we studied vector spaces, which provide the foundational setting for linear algebra. But vector spaces alone are only part of the story. Equally important is the study of functions between vector spaces that preserve the algebraic structure—these are called linear transformations.

A linear transformation is a rule that maps vectors from one vector space to another while preserving the two basic operations of vector addition and scalar multiplication. That is, the image of a sum is the sum of the images, and the image of a scalar multiple is the scalar multiple of the image.

Understanding linear transformations is essential because they allow us to translate abstract problems about vectors into concrete problems about matrices. Indeed, every linear transformation between finite-dimensional vector spaces can be represented by a matrix, and conversely, every matrix defines a linear transformation.

This chapter builds on the theory of vector spaces to develop a precise and rich understanding of linear transformations. We will define linear transformations and study their basic properties, explore the kernel (null space) and image (range) of a linear transformation, understand the conditions under which a transformation is injective (one-to-one), surjective (onto), or bijective, learn how linear transformations between finite-dimensional spaces correspond to matrix multiplication, introduce change of basis and the idea of similarity between matrices, and analyze the role of dimension through the powerful Dimension Theorem (or Rank-Nullity Theorem).

The study of linear transformations not only deepens our understanding of vector spaces, but it also sets the stage for key ideas in many advanced areas—such as eigenvalues and eigenvectors, diagonalization, and inner product spaces.

Throughout this chapter, we will move fluidly between the abstract world of transformations and the concrete world of matrices, building a bridge between structure and computation.

2.1 Linear Transformations

The most natural functions to consider between vector spaces are those that preserve the algebraic operations of vector addition and scalar multiplication. These functions are called *linear transformations*. They play a central role in linear algebra, as they encode structure-preserving maps between vector spaces.

Definition 2.1.1. Let V and W be vector spaces over $\mathbb{F}$. A *linear transformation* is a function $T : V \to W$ such that

1. (Preserves Vector Addition) For all $\vec{v}_1, \vec{v}_2 \in V$,
$$T(\vec{v}_1 + \vec{v}_2) = T(\vec{v}_1) + T(\vec{v}_2)$$
2. (Preserves Scalar Multiplication) For all $\vec{v} \in V$ and $\alpha \in \mathbb{F}$,
$$T(\alpha\vec{v}) = \alpha T(\vec{v})$$

A linear transformation $T : V \to V$ is called a *linear operator* on V.

Remark 2.1.2. The two properties of a linear transformation implies that $T(\vec{0}_V) = \vec{0}_W$ and that T preserves linear combinations, that is, if $\vec{v} = \alpha_1\vec{v}_1 + \alpha_2\vec{v}_2 + \cdots + \alpha_n\vec{v}_n$, where $\vec{v}_1, \vec{v}_2, ..., \vec{v}_n \in V$, then

$$T(\alpha_1\vec{v}_1 + \alpha_2\vec{v}_2 + \cdots + \alpha_n\vec{v}_n) = \alpha_1 T(\vec{v}_1) + \alpha_2 T(\vec{v}_2) + \cdots + \alpha_n T(\vec{v}_n)$$

Therefore, if one were to show that T is a linear transformation, it is sometimes quicker to check that for any $\vec{v}_1, \vec{v}_2 \in V$ and $\alpha_1, \alpha_2 \in \mathbb{F}$, that

$$T(\alpha_1\vec{v}_1 + \alpha_2\vec{v}_2) = \alpha_1 T(\vec{v}_1) + \alpha_2 T(\vec{v}_2)$$

Example 2.1.3. Let $V = W = \mathbb{R}^2$, and let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be the map defined by

$$T\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} 2x + y \\ x - y \end{bmatrix}$$

We claim that T is a linear transformation. To see this, we need to verify the properties of Definition 2.1.1.

To see that T preserves vector addition, let $\vec{v}_1 = \begin{bmatrix} x_1 \\ y_1 \end{bmatrix}$ and $\vec{v}_2 = \begin{bmatrix} x_2 \\ y_2 \end{bmatrix}$ are in $\mathbb{R}^2$, then

$$\begin{aligned} T(\vec{v}_1 + \vec{v}_2) &= T\left(\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} + \begin{bmatrix} x_2 \\ y_2 \end{bmatrix}\right) \\ &= T\left(\begin{bmatrix} x_1 + x_2 \\ y_1 + y_2 \end{bmatrix}\right) \\ &= \begin{bmatrix} 2(x_1 + x_2) + (y_1 + y_2) \\ (x_1 + x_2) - (y_1 + y_2) \end{bmatrix} \\ &= \begin{bmatrix} 2x_1 + 2x_2 + y_1 + y_2 \\ x_1 + x_2 - y_1 - y_2 \end{bmatrix} \\ &= \begin{bmatrix} 2x_1 + y_1 \\ x_1 - y_1 \end{bmatrix} + \begin{bmatrix} 2x_2 + y_2 \\ x_2 - y_2 \end{bmatrix} \\ &= T\left(\begin{bmatrix} x_1 \\ y_1 \end{bmatrix}\right) + T\left(\begin{bmatrix} x_2 \\ y_2 \end{bmatrix}\right) \\ &= T(\vec{v}_1) + T(\vec{v}_2) \end{aligned}$$

So T preserves vector addition.

To see that T preserves scalar multiplication, let $\vec{v} = \begin{bmatrix} x \\ y \end{bmatrix}$ and $\alpha \in \mathbb{R}$. Then

$$\begin{aligned} T(\alpha\vec{v}) &= T\left(\alpha \begin{bmatrix} x \\ y \end{bmatrix}\right) \\ &= T\left(\begin{bmatrix} \alpha x \\ \alpha y \end{bmatrix}\right) \\ &= \begin{bmatrix} 2(\alpha x) + \alpha y \\ \alpha x - \alpha y \end{bmatrix} \\ &= \begin{bmatrix} \alpha(2x + y) \\ \alpha(x - y) \end{bmatrix} \\ &= \alpha \begin{bmatrix} 2x + y \\ x - y \end{bmatrix} \\ &= \alpha T\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) \\ &= \alpha T(\vec{v}) \end{aligned}$$

So T preserves scalar multiplication.

Therefore, by Definition 2.1.1, we have shown that T is a linear transformation.

Example 2.1.4. Let $V = \mathbb{C}^2$, $W = \mathbb{C}$, and let $T : \mathbb{C}^2 \to \mathbb{C}$ be the map defined by

$$T\left(\begin{bmatrix} z_1 \\ z_2 \end{bmatrix}\right) = z_1 - iz_2$$

We claim that T is a linear transformation. To see this, we need to verify the properties of Definition 2.1.1.

To see that T preserves vector addition, let $\vec{v}_1 = \begin{bmatrix} z_1 \\ z_2 \end{bmatrix}, \vec{v}_2 = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} \in \mathbb{C}^2$. Then

$$\begin{aligned} T(\vec{v}_1 + \vec{v}_2) &= T\left(\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} + \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}\right) \\ &= T\left(\begin{bmatrix} z_1 + w_1 \\ z_2 + w_2 \end{bmatrix}\right) \\ &= (z_1 + w_1) - i(z_2 + w_2) \\ &= (z_1 - iz_2) + (w_1 - iw_2) \\ &= T\left(\begin{bmatrix} z_1 \\ z_2 \end{bmatrix}\right) + T\left(\begin{bmatrix} w_1 \\ w_2 \end{bmatrix}\right) \\ &= T(\vec{v}_1) + T(\vec{v}_2) \end{aligned}$$

So T preserves vector addition.

To see that T preserves scalar multiplication, let $\vec{v} = \begin{bmatrix} z_1 \\ z_2 \end{bmatrix}$ and let $\alpha \in \mathbb{C}$. Then

$$\begin{aligned} T(\alpha\vec{v}) &= T\left(\alpha \begin{bmatrix} z_1 \\ z_2 \end{bmatrix}\right) \\ &= T\left(\begin{bmatrix} \alpha z_1 \\ \alpha z_2 \end{bmatrix}\right) \\ &= (\alpha z_1) - i(\alpha z_2) \\ &= \alpha(z_1 - iz_2) \\ &= \alpha T\left(\begin{bmatrix} z_1 \\ z_2 \end{bmatrix}\right) \\ &= \alpha T(\vec{v}) \end{aligned}$$

So T preserves scalar multiplication.

Therefore, by Definition 2.1.1, we have shown that T is a linear transformation.

Example 2.1.5. Let $T : \mathcal{P}_2(\mathbb{R}) \to \mathbb{R}$ be defined by $T(p(x)) = p(1)$. We claim that T is a linear transformation. To see this, we need to verify the properties of Definition 2.1.1.

To see that T preserves vector addition, let $p(x), q(x) \in \mathcal{P}_2(\mathbb{R})$ be arbitrary. Then

$$\begin{aligned} T(p(x) + q(x)) &= T((p+q)(x)) \\ &= (p+q)(1) \\ &= p(1) + q(1) \\ &= T(p(x)) + T(q(x)) \end{aligned}$$

So T preserves vector addition.

To see that T preserves scalar multiplication, let $p(x) \in \mathcal{P}_2(\mathbb{R})$ and $\alpha \in \mathbb{R}$. Then

$$\begin{aligned} T(\alpha p(x)) &= T((\alpha p)(x)) \\ &= (\alpha p)(1) \\ &= \alpha p(1) \\ &= \alpha T(p(x)) \end{aligned}$$

So T preserves scalar multiplication.

Therefore, by Definition 2.1.1, we have shown that T is a linear transformation.

Linear transformations arise in countless mathematical settings, from the finite-dimensional spaces of vectors and matrices to the infinite-dimensional worlds of functions and operators. Beyond familiar examples like matrix multiplication and geometric transformations, we encounter linear operators in calculus (e.g., the derivative and integral), in probability (e.g., the expectation operator), and in analysis (e.g., the Laplace and Fourier transforms). These transformations all share the essential properties of linearity, preserving addition and scalar multiplication. The abundance and variety of linear transformations underscore their fundamental role in connecting different areas of mathematics, unifying diverse problems under a common framework. We leave the verification of the following examples to the reader.

Example 2.1.6. Let V be a vector space over $\mathbb{F}$. The *identity operator on* V $\mathrm{id}_V : V \to V$ defined by $\mathrm{id}_V(\vec{v}) = \vec{v}$ is a linear transformation.

Example 2.1.7. Let $V = \mathbb{F}^n$ and define the *left-* and *right-shift operators* by $\mathfrak{s}_l, \mathfrak{s}_r : \mathbb{F}^n \to \mathbb{F}^n$ by

$$\mathfrak{s}_l(x_1, x_2, ..., x_n) = (x_2, x_3, ..., x_n, 0)$$
$$\mathfrak{s}_r(x_1, x_2, ..., x_n) = (0, x_1, x_2, ..., x_{n-1})$$

for all $(x_1, x_2, ..., x_n) \in \mathbb{F}^n$. It is elementary to verify that both $\mathfrak{s}_l$ and $\mathfrak{s}_r$ are linear transformations.

Example 2.1.8. Let $\mathcal{D}[0,1]$ be the space of all differentiable functions on $[0,1]$ (see Example 1.2.7). Define the *differential operator* $D : \mathcal{D}[0,1] \to \mathcal{C}[0,1]$ by

$$D(f) = f'$$

for all $f \in \mathcal{D}[0,1]$. It is elementary to verify that D is a linear transformation.

Example 2.1.9. Let $V = \mathcal{C}[0,1]$. Define the *Volterra operator* $V : \mathcal{C}[0,1] \to \mathcal{C}[0,1]$ by

$$V(f(x)) = \int_0^x f(t)\ dt$$

for all $f \in \mathcal{C}[0,1]$. Indeed, this linear transformation is well-defined by the Fundamental Theorem of Calculus. It is elementary to verify that V is a linear transformation.

Example 2.1.10. Let $V = \mathcal{C}[0,1]$ and define the *expectation operator* $\mathbb{E} : \mathcal{C}[0,1] \to \mathbb{R}$ by

$$\mathbb{E}(f) = \int_0^1 xf(x)\ dx$$

for all $f \in \mathcal{C}[0,1]$. It is elementary to verify that $\mathbb{E}$ is a linear transformation.

Example 2.1.11. Let $A \in \mathcal{M}_{mn}(\mathbb{R})$ and define the map $T_A : \mathbb{R}^n \to \mathbb{R}^m$ by

$$T_A(\vec{x}) = A\vec{x}$$

for all $\vec{x} \in \mathbb{R}^n$. It is elementary to verify that T_A is a linear transformation.

Example 2.1.12. Let $V = \mathbb{R}^n$, and define the kth *coordinate projection* $\pi_k : \mathbb{R}^n \to \mathbb{R}$ by

$$\pi_k(x_1, x_2, ..., x_n) = x_k$$

for all $(x_1, x_2, ..., x_n) \in \mathbb{R}^n$. It elementary to verify that π_k is a linear transformation.

Example 2.1.13. Let $V = \mathcal{M}_n(\mathbb{R})$. Recall that for $A \in \mathcal{M}_n(\mathbb{R})$ given by

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}$$

The *trace of* A is defined to by

$$\operatorname{tr}(A) = a_{11} + a_{22} + \cdot + a_{nn}$$

That is, you take the sum of the diagonal entries of the matrix A. Define the *trace operator* $T : \mathcal{M}_n(\mathbb{R}) \to \mathbb{R}$ by

$$T(A) = \operatorname{tr}(A)$$

for all $A \in \mathcal{M}_n(\mathbb{R})$. It is elementary to verify that T is a linear transformation.

Example 2.1.14. Let $V = \mathcal{C}_p[0, \infty)$ be the space of all piecewise continuous functions on the interval $[0, \infty)$, and define the *Laplace transform* $\mathcal{L} : \mathcal{C}[0, \infty) \to \mathcal{F}[0, \infty)$ by

$$\mathcal{L}(f)(s) = \int_0^\infty e^{-st} f(t)\ dt$$

for all $f \in \mathcal{C}_p[0, \infty)$. It is elementary to verify that $\mathcal{L}$ is a linear transformation.

The following theorem collects three useful properties of all linear transformations. They can be described by saying that, in addition to preserving addition and scalar multiplication, linear transformations preserve the zero vector, negatives, and all linear combinations. Two of which are already mentioned in Remark 2.1.2, but we state as a general statement.

Proposition 2.1.15. *Let V and W be vector spaces over $\mathbb{F}$, and let $T : V \to W$ be a linear transformation. Then*

1. $T(\vec{0}_V) = \vec{0}_W$.
2. *For all $\vec{v} \in V$, $T(-\vec{v}) = -T(\vec{v})$.*
3. *For all $\vec{v}_1, \vec{v}_2, ..., \vec{v}_n \in V$ and $\alpha_1, \alpha_2, ..., \alpha_n \in \mathbb{F}$,*

$$T(\alpha_1\vec{v}_1 + \alpha_2\vec{v}_2 + \cdots + \alpha_n\vec{v}_n) = \alpha_1 T(\vec{v}_1) + \alpha_2 T(\vec{v}_2) + \cdots + \alpha_n T(\vec{v}_n)$$

As for functions in general, two linear transformations $T, S : V \to W$ are said to be equal, denoted by $T = S$, if they have the same action; that is, $T(\vec{v}) = S(\vec{v})$ for all $\vec{v} \in V$.

Theorem 2.1.16. *Let V, W be vector spaces over $\mathbb{F}$, and let $T, S : V \to W$ be linear transformations. If $V = \operatorname{span}\{\vec{v}_1, \vec{v}_2, ..., \vec{v}_n\}$ and $T(\vec{v}_k) = S(\vec{v}_k)$ for all k, then $T = S$.*

Proof. Let $\vec{v} \in V$. Then $\vec{v}$ can be written as a linear combination of the vectors $\vec{v}_1, \vec{v}_2, ..., \vec{v}_n$, say

$$\vec{v} = \alpha_1\vec{v}_1 + \alpha_2\vec{v}_2 + \cdots + \alpha_n\vec{v}_n$$

Then applying T to both sides, and using the fact that $T(\vec{v}_k) = S(\vec{v}_k)$ for all k, we obtain

$$\begin{aligned} T(\vec{v}) &= T(\alpha_1\vec{v}_1 + \alpha_2\vec{v}_2 + \cdots + \alpha_n\vec{v}_n) \\ &= \alpha_1 T(\vec{v}_1) + \alpha_2 T(\vec{v}_2) + \cdots + \alpha_n T(\vec{v}_n) \\ &= \alpha_1 S(\vec{v}_1) + \alpha_2 T(\vec{v}_2) + \cdots + \alpha_n S(\vec{v}_n) \\ &= S(\alpha_1\vec{v}_1 + \alpha_2\vec{v}_2 + \cdots + \alpha_n\vec{v}_n) \\ &= S(\vec{v}) \end{aligned}$$

Therefore, as $\vec{v} \in V$ was arbitrary, we have $T = S$, as desired. □

Theorem 2.1.16 can be expressed as follows: If we know what a linear transformation $T : V \to W$ does to each vector in a spanning set for V, then we know what T does to every vector in V. If the spanning set is a basis, we can say much more.

Theorem 2.1.17. *Let V, W be vector spaces over $\mathbb{F}$, and let $\{\vec{v}_1, \vec{v}_2, ..., \vec{v}_n\}$ be a basis for V. Given any vectors $\vec{w}_1, \vec{w}_2, ..., \vec{w}_n \in W$, there exists a unique linear transformation $T : V \to W$ such that $T(\vec{v}_k) = \vec{w}_k$ for each $k = 1, 2, ..., n$. In fact, the action of T is as follows: For $\vec{v} = \alpha_1\vec{v}_1 + \alpha_2\vec{v}_2 + \cdots + \alpha_n\vec{v}_n$ with $\alpha_k \in \mathbb{F}$, then*

$$T(\vec{v}) = T(\alpha_1\vec{v}_1 + \alpha_2\vec{v}_2 + \cdots + \alpha_n\vec{v}_n) = \alpha_1\vec{w}_1 + \alpha_2\vec{w}_2 + \cdots + \alpha_n\vec{w}_n$$

Proof. If a linear transformation T does exist with $T(\vec{v}_k) = \vec{w}_k$ for each k, and if S is any other such transformation, then $T(\vec{v}_k) = \vec{w}_k = S(\vec{v}_k)$ holds for each k, so $S = T$. Hence, T is linear if it exists.

We need to show that T is a linear transformation. For any $\vec{v} \in V$, we must specify $T(\vec{v}) \in W$. Since $\{\vec{v}_1, \vec{v}_2, ..., \vec{v}_n\}$ is a basis of V, we have

$$\vec{v} = \alpha_1\vec{v}_1 + \alpha_2\vec{v}_2 + \cdots + \alpha_n\vec{v}_n$$

where $\alpha_1, \alpha_2, ..., \alpha_n$ are uniquely determined by $\vec{v}$. Hence, we may define $T : V \to w$ by

$$T(\vec{v}) = T(\alpha_1\vec{v}_1 + \alpha_2\vec{v}_2 + \cdots + \alpha_n\vec{v}_n) = \alpha_1\vec{w}_1 + \alpha_2\vec{w}_2 + \cdots + \alpha_n\vec{w}_n$$

for all $\vec{v} = \alpha_1\vec{v}_1 + \alpha_2\vec{v}_2 + \cdots + \alpha_n\vec{v}_n$. This satisfies $T(\vec{v}_k) = \vec{w}_k$ for each k. The verification that T is linear is left to the reader. $\square$

This theorem shows that a linear transformation can be defined almost at will. Simply specifying where the basis vectors go, and the rest of the action is dictated by the linearity. Theorem 2.1.16 shows that deciding whether two linear transformations are equal comes down to determining whether they have the same effect on the basis vectors.

Example 2.1.18. Suppose we want to find a $T : \mathcal{P}_2(\mathbb{R}) \to \mathcal{M}_2(\mathbb{R})$ be defined such that

$$T(1+x) = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, T(x+x^2) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, T(1+x^2) = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

Note that $\{1+x, x+x^2, 1+x^2\}$ is a basis of $\mathcal{P}_2(\mathbb{R})$, so any polynomial $p(x) = a + bx + cx^2$ in $\mathcal{P}_2(\mathbb{R})$ is a linear combination of these vectors. Indeed,

$$p(x) = \frac{1}{2}(a+b-c)(1+x) + \frac{1}{2}(-a+b+c)(x+x^2) + \frac{1}{2}(a-b+c)(1+x^2)$$

Hence, Theorem 2.1.16 gives

$$\begin{aligned} T(p(x)) &= \frac{1}{2}(a+b-c)\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + \frac{1}{2}(-a+b+c)\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} + \frac{1}{2}(a-b+c)\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \\ &= \frac{1}{2}\begin{bmatrix} a+b-c & -a+b+c \\ -a+b+c & a-b+c \end{bmatrix} \end{aligned}$$

To conclude this section, we will introduce the concept of composition and composite linear operators.

Definition 2.1.19. Given linear transformations $V \xrightarrow{T} W \xrightarrow{S} U$, the *composite* $ST : V \to U$ of T and S is defined by

$$ST(\vec{v}) = S(T(\vec{v}))$$

for all $\vec{v} \in V$. The operation of forming the new function ST is called *composition*.

Sometimes we may see composition denoted as $S \circ T$, but ST will be used more frequently and for simplicity.

The action of ST can be described as follows: ST means T first, then S.

Remark 2.1.20. Not all pairs of linear transformations can be composed. For example, if $T : V \to W$ and $S : W \to U$ are linear transformations, then $ST : V \to U$ is defined, but TS cannot be formed, unless $U = V$ because $S : V \to U$ and $T : V \to W$ do not link in that order.

Moreover, if ST and TS can be formed, they may not be equal. In fact, if $S : \mathbb{R}^m \to \mathbb{R}^n$ and $T : \mathbb{R}^n \to \mathbb{R}^m$ are induced by matrices A and B, respectively, then ST and TS can be formed (they are induced by AB and BA respectively), but the matrix products AB and BA may not be equal (they may not even be the same size).

Example 2.1.21. Let $S : \mathcal{M}_2(\mathbb{R}) \to \mathcal{M}_2(\mathbb{R})$ and $T : \mathcal{M}_2(\mathbb{R}) \to \mathcal{M}_2(\mathbb{R})$ be defined by

$$S\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = \begin{bmatrix} c & d \\ a & b \end{bmatrix}, \quad T(A) = A^T$$

for $A \in \mathcal{M}_2(\mathbb{R})$. Here, we describe ST and TS as follows:

$$ST\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = S\left(\begin{bmatrix} a & c \\ b & d \end{bmatrix}\right) = \begin{bmatrix} b & d \\ a & c \end{bmatrix}$$

On the other hand,

$$TS\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = T\left(\begin{bmatrix} c & d \\ a & b \end{bmatrix}\right) = \begin{bmatrix} c & a \\ d & b \end{bmatrix}$$

It is easy to see that TS need not equal to ST, so $TS \neq ST$.

Example 2.1.22. Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be defined by $T(\vec{x}) = A\vec{x}$ where $A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$ (rotation by 90°), and let $S : \mathbb{R}^2 \to \mathbb{R}^2$ be defined by $S(\vec{x}) = B\vec{x}$ where $B = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$ (scaling in y-axis). Then ST and TS are both linear but yield different geometric effects and different matrices.

The following proposition collects some basic properties of the composition operation.

Proposition 2.1.23. *Let $V \xrightarrow{T} W \xrightarrow{S} U \xrightarrow{R} Z$ be linear transformations.*

1. *The composite $ST : V \to U$ is a linear transformation.*

2. $T\,\mathrm{id}_V = T$ *and* $\mathrm{id}_W\, T = T$.

3. $(RS)T = R(ST)$.

Proof. The proofs of (1) and (2) are left for the reader. To see that (3) holds, note that for every $\vec{v} \in V$,

$$[(RS)T](\vec{v}) = (RS)(T(\vec{v})) = R(S(T(\vec{v}))) = R((ST)(\vec{v})) = [R(ST)](\vec{v})$$

which shows that $(RS)T = R(ST)$. □

2.2 The Kernel and Range

When studying a linear transformation, it's not enough to simply understand how it maps inputs to outputs—we also want to analyze what gets sent to zero, and what the image of the transformation looks like. This leads to two fundamental subspaces associated with any linear transformation: the kernel and the range.

The kernel (also called the null space) of a linear transformation captures all vectors that are mapped to the zero vector. In many applications, the kernel tells us about the "degeneracy" or redundancy in the transformation—it tells us which vectors are "lost" or "collapsed" under the mapping.

On the other hand, the range (also called the image) consists of all possible outputs of the transformation. It tells us how much of the codomain is actually "reached" by the linear map. Together, the kernel and range give a precise way to understand the structure and behavior of a transformation.

In this section, we define the kernel and range formally, explore their properties, and compute them through concrete examples in both real and complex vector spaces. These concepts will be essential later on when we study the rank-nullity theorem, invertibility, and the classification of linear maps.

Definition 2.2.1. Let V and W be vector spaces over $\mathbb{F}$, and let $T : V \to W$ be a linear transformation.

1. The *kernel of* T is the set

$$\ker(T) = \{\vec{v} \in V : T(\vec{v}) = \vec{0}_W\}$$

   It consists of all vectors in V that are mapped to the zero vector in W. The kernel of T is a subspace of V.

2. The *range of* T is the set

$$\operatorname{ran}(T) = \{T(\vec{v}) : \vec{v} \in V\}$$

It consists of all vectors in W that are images of vectors in V under T. The range of T is a subspace of W.

Example 2.2.2. Let $T : \mathbb{R}^3 \to \mathbb{R}^2$ be defined by

$$T\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) = \begin{bmatrix} x+y \\ y+z \end{bmatrix}$$

To find the kernel of T, we need to solve

$$T\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

or

$$\begin{cases} x+y=0 \\ y+z=0 \end{cases}$$

Solving the system gives $x = -y$ and $z = -y$, so

$$\begin{aligned} \ker(T) &= \left\{ \begin{bmatrix} x \\ y \\ z \end{bmatrix} \in V : T\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \right\} \\ &= \left\{ \begin{bmatrix} -y \\ 1 \\ -y \end{bmatrix} : y \in \mathbb{R} \right\} \\ &= \operatorname{span}\left\{ \begin{bmatrix} -1 \\ 1 \\ -1 \end{bmatrix} \right\} \end{aligned}$$

To find the range of T, note that

$$\operatorname{ran}(T) = \operatorname{span}\left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}$$

So $\ker(T)$ is 1-dimensional, and $\operatorname{ran}(T) = \mathbb{R}^2$.

Example 2.2.3. Let $T : \mathbb{C}^2 \to \mathbb{C}^2$ be defined by

$$T\left(\begin{bmatrix} z_1 \\ z_2 \end{bmatrix}\right) = \begin{bmatrix} z_1 + iz_2 \\ z_2 \end{bmatrix}$$

To find the kernel, we need to solve

$$T\left(\begin{bmatrix} z_1 \\ z_2 \end{bmatrix}\right) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

or

$$\begin{cases} z_1 + iz_2 = 0 \\ z_2 = 0 \end{cases}$$

In this case, $z_2 = 0$, so $z_1 = 0$. Therefore, $\ker(T) = \{\vec{0}\}$.

For the range, since both components are independent linear expressions, $\operatorname{ran}(T) = \mathbb{C}^2$.

Example 2.2.4. Let $D : \mathcal{D}[0, 1] \to \mathcal{C}[0, 1]$ be the differential operator defined by $D(f) = f'$, where $\mathcal{D}[0, 1]$ is the space of differentiable functions on $[0, 1]$ and $\mathcal{C}[0, 1]$ is the space of continuous functions on $[0, 1]$.

To determine the kernel, we must find all functions $f \in \mathcal{D}[0, 1]$ such that $D(f) = f' = 0$. From calculus, if $f'(x) = 0$ for all $x \in [0, 1]$, then f must be a constant function. Thus, for some $c \in \mathbb{R}$, we have $f(x) = c$ for all $x \in [0, 1]$. Therefore,

$$\ker(D) = \{f \in \mathcal{D}[0, 1] : f(x) = c \text{ for some } c \in \mathbb{R}\} = \operatorname{span}\{1\}$$

To determine the range, observe that for any $g \in \mathcal{C}[0, 1]$, there exists a function $f \in \mathcal{D}[0, 1]$ such that $f'(x) = g(x)$ for all $x \in [0, 1]$. Indeed, by the Fundamental Theorem of Calculus, the function

$$f(x) = \int_0^x g(t)\, dt$$

satisfies $f' = g$. Therefore, every continuous function is the derivative of some differentiable function. Hence,

$$\operatorname{ran}(D) = \mathcal{C}[0, 1]$$

Definition 2.2.5. Let V and W be vector spaces over $\mathbb{F}$ and let $T : V \to W$ be a linear transformation.

1. The *nullity of* T is the dimension of the kernel of T, so $\dim(\ker(T))$.

2. The *rank of* T is the dimension of the range of T, so $\dim(\operatorname{ran}(T))$.

Often, a useful way to study a subspace of a vector space is to exhibit it as the kernel or range of a linear transformation. Here is an example.

Example 2.2.6. Let $T : \mathcal{M}_n(\mathbb{R}) \to \mathcal{M}_n(\mathbb{R})$ be the linear transformation defined by $T(A) = A - A^T$. Indeed, it is elementary to verify that T is linear. To find the kernel, note that a matrix A lies in $\ker(T)$ when

$$0 = T(A) = A - A^T$$

so $A = A^T$. In other words, the kernel of T is the set of all symmetric matrices, so

$$\ker(T) = \mathcal{S}_n(\mathbb{R})$$

To find the range, note that $\operatorname{ran}(T)$ consists of all matrices $T(A)$ for $A \in \mathcal{M}_n(\mathbb{R})$. Every such matrix is of the form of a skew-symmetric matrix because

$$T(A)^T = (A - A^T)^T = A^T - A = -T(A)$$

On the other hand, if S is skew-symmetric, so $S^T = -S$, then S lies in $\operatorname{ran}(T)$, so

$$T\left(\frac{1}{2}S\right) = \frac{1}{2}S - \left(\frac{1}{2}S\right)^T = \frac{1}{2}(S - S^T) = \frac{1}{2}(S + S) = S$$

In this case,

$$\operatorname{ran}(T) = \mathcal{S}_n^s(\mathbb{R})$$

where $\mathcal{S}_n^s(\mathbb{R})$ denotes the space of all skew-symmetric matrices.

In the context of linear transformations, it is natural to ask when a transformation behaves like a "function with an inverse." To answer this, we study two fundamental properties of linear transformations: being *one-to-one* (injective) and being *onto* (surjective).

Definition 2.2.7. Let V and W be vector spaces over $\mathbb{F}$ and let $T : V \to W$ be a linear transformation.

1. T is said to be *one-to-one* (*injective*) if $T(\vec{v}_1) = T(\vec{v}_2)$ implies $\vec{v}_1 = \vec{v}_2$.

2. T is said to be *onto* (*surjective*) if $\operatorname{ran}(T) = W$; that is, for any $\vec{w} \in W$, there exists a $\vec{v} \in V$ such that $T(\vec{v}) = \vec{w}$.

3. T is said to be a *bijection* (*bijective*) if T is both injective and surjective.

Clearly, the onto transformations T are those for which $\operatorname{ran}(T) = W$ is a large subspace of W as possible. By contrast, the following shows that one-to-one transformations are the ones with the kernel as small a subspace of V as possible.

Theorem 2.2.8. *Let V and W be vector spaces over $\mathbb{F}$, and let $T : V \to W$ be a linear transformation. Then T is one-to-one if and only if* $\ker(T) = \{\vec{0}_V\}$.

Proof. If T is one-to-one, let $\vec{v}$ be any vector in $\ker(T)$. Then $T(\vec{v}) = \vec{0}_W$, so $T(\vec{v}) = T(\vec{0}_V)$. Hence, $\vec{v} = \vec{0}_V$, because T is one-to-one, os $\ker(T) = \{\vec{0}_V\}$.

Conversely, if $\ker(T) = \{\vec{0}_V\}$ and $T(\vec{v}_1) = T(\vec{v}_2)$ where $\vec{v}_1, \vec{v}_2 \in V$, then

$$T(\vec{v}_1 - \vec{v}_2) = T(\vec{v}_1) - T(\vec{v}_2) = \vec{0}_W$$

$\vec{v}_1 - \vec{v}_2 \in \ker(T) = \{\vec{0}_V\}$, so $\vec{v}_1 - \vec{v}_2 = \vec{0}_V$, and thus, $\vec{v}_1 = \vec{v}_2$, showing that T is one-to-one, as desired. $\square$

Example 2.2.9. For any vector space V, the identity operator $\mathrm{id}_V : V \to V$ is a bijection.

Example 2.2.10. Let $T : \mathbb{R}^3 \to \mathbb{R}^2$ be the linear transformations given by

$$T\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) = \begin{bmatrix} x+y \\ x-y \end{bmatrix}$$

We claim that T is surjective, but not injective.

To see this, note that the vector $\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$ lies in $\ker(T)$, so T cannot be injective. On the other hand, for any $\begin{bmatrix} s \\ t \end{bmatrix} \in \mathbb{R}^2$ it lies in $\operatorname{ran}(T)$ because

$$\begin{bmatrix} s \\ t \end{bmatrix} = \begin{bmatrix} x+y \\ x-y \end{bmatrix}$$

whenever $x = \frac{1}{2}(s+t)$ and $y = \frac{1}{2}(s-t)$ and $z = 0$. Therefore, z is onto.

Example 2.2.11. Let $T : \mathbb{R}^2 \to \mathbb{R}^3$ be the linear transformation defined by

$$T\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} x+y \\ x-y \\ x \end{bmatrix}$$

We claim that T is injective, but not surjective.

To see this, note that T is injective because

$$\ker(T) = \left\{ \begin{bmatrix} x \\ y \end{bmatrix} \in \mathbb{R}^2 : x + y = x - y = x = 0 \right\} = \left\{ \begin{bmatrix} 0 \\ 0 \end{bmatrix} \right\}$$

However, note that T is not onto, because the vector $\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$ does not lie in $\operatorname{ran}(T)$ because if

$$\begin{bmatrix} x + y \\ x - y \\ x \end{bmatrix}$$

for some $x, y \in \mathbb{R}$, then $x + y = 0 = x - y$ and $x = 1$, which is absurd.

Example 2.2.12. Let $T : \mathbb{R}^3 \to \mathbb{R}^3$ be defined by $T(\vec{v}) = A\vec{v}$, where

$$A = \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix}$$

Since A is an invertible matrix (for example, check that $\det(A) \neq 0$), then T is both one-to-one and onto. Therefore, T is a bijection.

Example 2.2.13. Let $D : \mathcal{D}[0,1] \to \mathcal{C}[0,1]$ be the differential operator by $D(f) = f'$. As we have discussed earlier in Example 2.2.4, $\ker(D) = \operatorname{span}\{1\}$, so D cannot be injective. However since $\operatorname{ran}(D) = \mathcal{C}[0,1]$, it follows that D is onto.

Example 2.2.14. Let $V : \mathcal{C}[0,1] \to \mathcal{C}[0,1]$ be the Volterra operator

$$V(f(x)) = \int_0^x f(t)\ dt$$

Here, $\ker(T) = \{0\}$, since if $V(f(x)) = 0$ for all x, then $f = 0$, so V is injective. On the other hand, V is not nto, because not all continuous functions can be written in the above form.

2.3 The Dimension Theorem

One of the most elegant and powerful results in linear algebra is the *Dimension Theorem*, also known as the *Rank-Nullity Theorem*. This theorem describes a fundamental relationship between three key components of any linear transformation: the dimension of the domain, the dimension of the kernel (null space), and the dimension of the range (image).

Recall that the kernel of a linear transformation measures the extent to which the transformation *fails to be injective*, while the range tells us how much of the codomain is *actually reached* by the transformation. The Dimension Theorem tells us that if we know the dimension of the domain and the size of the kernel, then the size of the image is completely determined—and vice versa.

In essence, this result formalizes the intuitive idea that every vector in the domain contributes either to the kernel or to the image, but not both. This has profound implications across mathematics, including solutions of systems of equations, properties of matrices, and the study of invertibility and isomorphisms.

In this section, we present a formal statement and proof of the theorem, and work through illustrative examples to reinforce its importance.

Theorem 2.3.1 (The Dimension Theorem). *Let V and W be vector spaces over $\mathbb{F}$, and let $T : V \to W$ be a linear transformation. If* $\ker(T)$ *and* $\operatorname{ran}(T)$ *are finite-dimensional, then V is also finite-dimensional, and*

$$\dim(V) = \dim(\ker(T)) + \dim(\operatorname{ran}(T))$$

Proof. First note that every vector in $\operatorname{ran}(T)$ has the form $T(\vec{v})$ for some $\vec{v} \in V$. Hence, let $\{T(\vec{e}_1), T(\vec{e}_2), ..., T(\vec{e}_n)\}$ be a basis for $\operatorname{ran}(T)$, where the $\vec{e}_k$ lie in V. Let $\{\vec{f}_1, \vec{f}_2, ..., \vec{f}_m\}$ be any basis for $\ker(T)$. Then $\dim(\operatorname{ran}(T)) = n$ and $\dim(\ker(T)) = m$, so it suffices to show that

$$\mathcal{B} = \{\vec{e}_1, \vec{e}_2, ..., \vec{e}_n, \vec{f}_1, \vec{f}_2, ..., \vec{f}_m\}$$

is a basis of V.

To see that $\mathcal{B}$ spans V, first note that if $\vec{v} \in V$, then $T(\vec{v}) \in \operatorname{ran}(T)$, so

$$T(\vec{v}) = \alpha_1 T(\vec{e}_1) + \alpha_2 T(\vec{e}_2) + \cdots + \alpha_n T(\vec{e}_n)$$

where $\alpha_1, \alpha_2, ..., \alpha_n \in \mathbb{F}$. Then this implies that

$$\vec{v} - \alpha_1\vec{e}_1 - \alpha_2\vec{e}_2 - \cdots - \alpha_n\vec{e}_n$$

lies in $\ker(T)$ and so it is a linear combination of $\vec{f}_1, \vec{f}_2, ..., \vec{f}_m$. Therefore, $\vec{v}$ is a linear combination of the vectors in $\mathcal{B}$.

To see that $\mathcal{B}$ is linearly independent, assume that $\alpha_1, \alpha_2, ..., \alpha_n, \beta_1, \beta_2, ..., \beta_m \in \mathbb{F}$ are such that

$$\alpha_1\vec{e}_1 + \alpha_2\vec{e}_2 + \cdots + \alpha_n\vec{e}_n + \beta_1\vec{f}_1 + \beta_2\vec{f}_2 + \cdots + \beta_m\vec{f}_m = \vec{0}_V$$

Applying T gives

$$\alpha_1 T(\vec{e}_1) + \alpha_2 T(\vec{e}_2) + \cdots + \alpha_n T(\vec{e}_n) = \vec{0}_W$$

Hence, the linear independence of $\{T(\vec{e}_1), T(\vec{e}_2), ..., T(\vec{e}_n)\}$ gives $\alpha_1 = 0$, $\alpha_2 = 0$,..., $\alpha_n = 0$. But then

$$\beta_1\vec{f}_1 + \beta_2\vec{f}_2 + \cdots + \beta_m\vec{f}_m = \vec{0}_V$$

But by linear independence of $\{\vec{f}_1, \vec{f}_2, ..., \vec{f}_m\}$, we have $\beta_1 = 0$, $\beta_2 = 0$, ..., $\beta_m = 0$. Therefore, we have shown that $\mathcal{B}$ is linearly independent. $\square$

Remark 2.3.2. Note that the vector space V is not assumed to be finite-dimensional in Theorem 2.3.1. In fact, verifying that $\ker(T)$ and $\operatorname{ran}(T)$ are both finite-dimensional is often an important way to prove that V is finite dimensional.

Corollary 2.3.3. *Let V and W be vector spaces over $\mathbb{F}$, and let $T : V \to W$ be a linear transformation. Let $\{\vec{e}_1, \vec{e}_2, ..., \vec{e}_r, \vec{e}_{r+1}, ..., \vec{e}_n\}$ be a basis for V such that $\{\vec{e}_{r+1}, \vec{e}_{r+2}, ..., \vec{e}_n\}$ is a basis for* $\ker(T)$*. Then $\{T(\vec{e}_1), T(\vec{e}_2), ..., T(\vec{e}_r)\}$ is a basis for* $\operatorname{ran}(T)$*, and hence, $r = \operatorname{rank}(T)$.*

Example 2.3.4. Let $A \in \mathcal{M}_{mn}(\mathbb{R})$ with $\operatorname{rank}(A) = r$. We claim that the space $\operatorname{null}(A)$ of all solutions of the system $A\vec{v} = \vec{0}$ of m homogeneous equations in n variables has dimension $n - r$. Indeed, note that the space in question is simply $\ker(T_A)$, where $T_A(\vec{v}) = A\vec{v}$ for all columns $\vec{v} \in \mathbb{R}^n$. But $\dim(\operatorname{ran}(T_A)) = \operatorname{rank}(T_A) = \operatorname{rank}(A) = r$, so $\dim(\ker(T_A)) = n - r$ by the Dimension Theorem.

For example, let $T : \mathbb{R}^4 \to \mathbb{R}^3$ be defined by the matrix

$$A = \begin{bmatrix} 1 & 0 & 2 & 1 \\ 0 & 1 & -1 & 2 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Then $T(\vec{x}) = A\vec{x}$ defines a linear transformation. To apply the Dimension Theorem, we compute the rank and nullity of A.

From the matrix above, we see that the rows are already in row-echelon form, with two leading ones. Therefore, the rank of A is

$$\dim(\operatorname{ran}(T)) = \operatorname{rank}(A) = 2$$

Since the domain of T is $\mathbb{R}^4$, the Dimension Theorem tells us:

$$\dim(\ker(T)) = 4 - 2 = 2$$

So the nullity of T is 2. This means that the kernel of T is a 2-dimensional subspace of $\mathbb{R}^4$, and the image of T is a 2-dimensional subspace of $\mathbb{R}^3$.

Example 2.3.5. Consider the differential operator $D : \mathcal{P}_3(\mathbb{R}) \to \mathcal{P}_2(\mathbb{R})$ given by

$$D(p(x)) = p'(x)$$

We know that $\mathcal{P}_3(\mathbb{R})$ has dimension 4 and $\mathcal{P}_2(\mathbb{R})$ has dimension 3. The kernel of D consists of all polynomials $p(x)$ such that $p'(x) = 0$, which are precisely the constant polynomials, so $\ker(D) = \text{span}\{1\}$, and thus, $\dim(\ker(D)) = 1$. Then by the Dimension Theorem:

$$\dim(\text{ran}(D)) = \dim(\mathcal{P}_3(\mathbb{R})) - \dim(\ker(D)) = 4 - 1 = 3$$

So the image of D is all of $\mathcal{P}_2(\mathbb{R})$, i.e., D is onto.

Example 2.3.6. Let $T : \mathcal{M}_2(\mathbb{R}) \to \mathcal{M}_2(\mathbb{R})$ be defined by

$$T(A) = A - A^T$$

From Example 2.2.6, we know that $\ker(T) = \mathcal{S}_2(\mathbb{R})$ and $\text{ran}(T) = \mathcal{S}_2^s(\mathbb{R})$. Both $\ker(T)$ and $\text{ran}(T)$ are 3- and 1-dimensional, respectively. So, $\dim(\ker(T)) = 3$, $\dim(\text{ran}(T)) = 1$, and $\dim(\mathcal{M}_2(\mathbb{R})) = 4$. The Dimension Theorem tells us that:

$$\dim(\ker(T)) + \dim(\text{ran}(T)) = 3 + 1 = 4 = \dim(\mathcal{M}_2(\mathbb{R}))$$

Example 2.3.7. Let $V = \mathcal{P}_n(\mathbb{R})$ and $W = \mathbb{R}$, and consider the *evaluation map* $\varphi_a : \mathcal{P}_n(\mathbb{R}) \to \mathbb{R}$ by $\varphi_a(p(x)) = p(a)$. It is easy to note that φ_a is a linear operator and that φ_a is onto. In this case, $\dim(\text{ran}(\varphi_a)) = \dim(\mathbb{R}) = 1$, so $\dim(\ker(\varphi_a)) = (n+1) - 1 = n$ by the Dimension Theorem. Now, each of the n polynomials $(x-a)$, $(x-a)^2$,..., $(x-a)^n$ is in $\ker(\varphi_a)$, and they are linearly independent, so they are a basis because $\dim(\ker(\varphi_a)) = n$.

2.4 Isomorphisms of Linear Transformations

In linear algebra, an *isomorphism* is a special type of linear transformation that perfectly preserves the structure of vector spaces. Two vector spaces are said to be *isomorphic* if there exists a bijective linear transformation (i.e., one-to-one and onto) between them. In essence, isomorphic vector spaces are "the same" for all practical linear algebraic purposes—they may appear different, but their structure, dimension, and behavior under linear operations are identical.

Studying isomorphisms allows us to classify vector spaces according to their structure. For example, any n-dimensional vector space over a field $\mathbb{F}$ is isomorphic to $\mathbb{F}^n$. This means that every finite-dimensional vector space behaves, in some sense, like $\mathbb{R}^n$ or $\mathbb{C}^n$, depending on the field.

Understanding isomorphisms not only simplifies our perspective on vector spaces, but also plays a key role in applications across mathematics, physics, and computer science. In this section, we will define isomorphisms formally, discuss when two vector spaces are isomorphic, and explore how to construct and verify such transformations.

Definition 2.4.1. Let V and W be vector spaces over $\mathbb{F}$. A linear transformation $T : V \to W$ is said to be an *isomorphism* if T is both one-to-one and onto, that is, T is a bijection. If there exists an isomorphism $T : V \to W$, then we say that V and W are *isomorphic* vector spaces, and write $V \simeq W$.

Example 2.4.2. For any vector space V, the identity transformation $\mathrm{id}_V : V \to V$ given by $\mathrm{id}_V(\vec{v}) = \vec{v}$ is an isomorphism. In other words, $V \simeq V$.

Example 2.4.3. Let $T : \mathcal{M}_{nm}(\mathbb{R}) \to \mathcal{M}_{mn}(\mathbb{R})$ be the linear transformation defined by $T(A) = A^T$. Then T is an isomorphism, and thus, $\mathcal{M}_{nm} \simeq \mathcal{M}_{mn}$. To see this, we first show that T is injective. Indeed, for any $A \in \mathcal{M}_{nm}(\mathbb{R})$ such that $T(A) = \vec{0}$, we have $A^T = \vec{0}$, but then $A = (A^T)^T = \vec{0}^T = \vec{0}$, which shows that $\ker(T) = \{\vec{0}\}$.

Of course, it is also easy to determine that T is onto, since for any matrix $B \in \mathcal{M}_{mn}(\mathbb{R})$, there is another matrix $B^T \in \mathcal{M}_{nm}(\mathbb{R})$ such that $T(B^T) = B$, so T is also onto. Therefore, since T is both one-to-one and onto, T is an isomorphism, and thus, $\mathcal{M}_{nm}(\mathbb{R}) \simeq \mathcal{M}_{mn}(\mathbb{R})$.

Example 2.4.4. Let $V = \mathbb{R}^3$ and $W = \mathcal{P}_2(\mathbb{R})$, the vector space of real polynomials of degree at most 2. Define the map $T : V \to W$ by

$$T\left(\begin{bmatrix} a \\ b \\ c \end{bmatrix}\right) = a + bx + cx^2$$

It is easy to check that T is linear and that T is one-to-one and onto, so that T is an isomorphism, so $\mathbb{R}^3 \simeq \mathcal{P}_2(\mathbb{R})$.

The word *isomorphism* comes from two Greek roots: *iso* meaning "same", and *morphos*, meaning "form". An isomorphism $T : V \to W$ induces a pairing $\vec{v} \leftrightarrow T(\vec{v})$ between vectors $\vec{v} \in V$, and vectors in $T(\vec{v})$ that preserves vector addition and scalar multiplication. Hence, as far as their vector space properties are concerned, the spaces V and W are identical except for notation.

The following theorem gives a pretty useful characterization of isomorphisms.

Theorem 2.4.5. *Let V and W be finite-dimensional vector spaces over $\mathbb{F}$ and let $T : V \to W$ be a linear transformation. The following are equivalent.*

1. *T is an isomorphism.*
2. *If $\{\vec{e}_1, \vec{e}_2, ..., \vec{e}_n\}$ is any basis of V, then $\{T(\vec{e}_1), T(\vec{e}_2), ..., T(\vec{e}_n)\}$ is a basis of W.*
3. *There exists a basis $\{\vec{e}_1, \vec{e}_2, ..., \vec{e}_n\}$ of V such that $\{T(\vec{e}_1), T(\vec{e}_2), ..., T(\vec{e}_n)\}$ is a basis of W.*

Proof. (1) $\Rightarrow$ (2): Let $\{\vec{e}_1, \vec{e}_2, ..., \vec{e}_n\}$ be a basis of V. If

$$\alpha_1 T(\vec{e}_1) + \alpha_2 T(\vec{e}_2) + \cdots + \alpha_n T(\vec{e}_n) = \vec{0}_W$$

with $\alpha_1, \alpha_2, ..., \alpha_n \in \mathbb{F}$, then

$$T(\alpha_1 \vec{e}_1 + \alpha_2 \vec{e}_2 + \cdots + \alpha_n \vec{e}_n) = \vec{0}_V$$

(since $\ker(T) = \{\vec{0}_V\}$). But then each $\alpha_i = 0$ by the independence of $\vec{e}_i$, so $\{T(\vec{e}_1), T(\vec{e}_2), ..., T(\vec{e}_n)\}$ is linearly independent. To see that it spans W, let $\vec{w} \in W$. Since T is onto, $\vec{w} = T(\vec{v})$ for some $\vec{v} \in V$, so write

$$\vec{v} = \alpha_1 \vec{e}_1 + \alpha_2 \vec{e}_2 + \cdots + \alpha_n \vec{e}_n$$

Hence we obtain

$$\vec{w} = T(\vec{v}) = \alpha_1 T(\vec{e}_1) + \alpha_2 T(\vec{e}_2) + \cdots + \alpha_n T(\vec{e}_n)$$

proving that $\{T(\vec{e}_1), T(\vec{e}_2), ..., T(\vec{e}_n)\}$ spans W, and thus, $\{T(\vec{e}_1) T(\vec{e}_2), ..., T(\vec{e}_n)\}$ is a basis of W.

(2) $\Rightarrow$ (3): This holds because V is a basis.

(3) $\Rightarrow$ (1): If $T(\vec{v}) = \vec{0}_W$, then write

$$\vec{v} = \alpha_1 \vec{v}_1 + \alpha_2 \vec{v}_2 + \cdots + \alpha_n \vec{v}_n$$

where $\alpha_1, \alpha_2, ..., \alpha_n \in \mathbb{F}$. Then

$$\vec{0}_W = T(\vec{v}) = \alpha_1 T(\vec{e}_1) + \alpha_2 T(\vec{v}_2) + \cdots + \alpha_n T(\vec{e}_n)$$

so $\alpha_1 = 0$, $\alpha_2 = 0$, ..., $\alpha_n = 0$, so $\ker(T) = \{\vec{0}_V\}$, and T is one-to-one. To show that T is onto, let $\vec{w} \in W$. By assumption, there exists $\alpha_1, \alpha_2, ..., \alpha_n \in \mathbb{F}$ such that

$$\vec{w} = \alpha_1 T(\vec{e}_1) + \alpha_2 T(\vec{e}_2) + \cdots + \alpha_n T(\vec{e}_n) = T(\alpha_1 \vec{e}_1 + \alpha_2 \vec{e}_2 + \cdots + \alpha_n \vec{e}_n)$$

Therefore, T is onto. $\square$

Remark 2.4.6. Theorem 2.4.5 dovetails nicely with Theorem 2.1.17 as follows. Let V and W be vector spaces over $\mathbb{F}$ of dimension n, and suppose that $\{\vec{e}_1, \vec{e}_2, ..., \vec{e}_n\}$ and $\{\vec{f}_1, \vec{f}_2, ..., \vec{f}_n\}$ are bases of V and W, respectively. Theorem 2.1.17 asserts that there exists a linear transformation $T : V \to W$ such that

$$T(\vec{e}_i) = \vec{f}_i$$

for all $i = 1, 2, ..., n$. Then $\{T(\vec{e}_1), T(\vec{e}_2), ..., T(\vec{e}_n)\}$ is a basis of W, so T is an isomorphism by Theorem 2.4.5. Furthermore, the action of T is prescribed by

$$T(\alpha_1 \vec{e}_1 + \alpha_2 \vec{e}_2 + \cdots + \alpha_n \vec{e}_n) = \alpha_1 \vec{f}_1 + \alpha_2 \vec{f}_2 + \cdots + \alpha_n \vec{f}_n$$

so isomorphisms between spaces of equal dimension can easily be defined as soon as bases are known. In particular, this shows that if two vector spaces V and W have the same dimension, then they are isomorphic, that is, $V \simeq W$.

Theorem 2.4.7. *If V and W are finite-dimensional vector spaces over $\mathbb{F}$, then $V \simeq W$ if and only if $\dim(V) = \dim(W)$.*

Proof. It is easy to note that if $\dim(V) = \dim(W)$, then $V \simeq W$ by Remark 2.4.6, so it suffices to show that if $V \simeq W$, then $\dim(V) = \dim(W)$. In this case, there exists an isomorphism $T : V \to W$. Since V is finite-dimensional, let $\{\vec{e}_1, \vec{e}_2, ..., \vec{e}_n\}$ is a basis of V, then $\{T(\vec{e}_1), T(\vec{e}_2), ..., T(\vec{e}_n)\}$ is a basis of W by Theorem 2.4.5 so $\dim(W) = n = \dim(V)$. $\square$

As it turns out, the notion of composition of linear operators and isomorphisms have a well-established connection among each other.

Theorem 2.4.8. *Let V and W be finite-dimensional vector spaces over $\mathbb{F}$, and let $T : V \to W$ be a linear transformation. The following are equivalent.*

1. *T is an isomorphism.*
2. *There exists a linear transformation $S : W \to V$ such that $ST = \mathrm{id}_V$ and $TS = \mathrm{id}_W$.*

Moreover, in this case, S is also an isomorphism and is uniquely determined by T: If $\vec{w} \in W$ is written as $\vec{w} = T(\vec{v})$, then $S(\vec{w}) = \vec{v}$.

Proof. To see that (1) $\Rightarrow$ (2), let $\mathcal{B}_1 = \{\vec{e}_1, \vec{e}_2, ..., \vec{e}_n\}$ be a basis of V. Then $\mathcal{B}_2 = \{T(\vec{e}_1), T(\vec{e}_2), ..., T(\vec{e}_n)\}$ is a basis of W by Theorem **??**. Hence, by Theorem 2.1.17, define $S : W \to V$ by

$$S(T(\vec{e}_i)) = \vec{e}_i$$

for each i.

Since $\vec{e}_i = \mathrm{id}_V(\vec{e}_i)$, then $ST = \mathrm{id}_V$ by Theorem 2.1.16. But applying T gives

$$T(S(T(\vec{e}_i))) = T(\vec{e}_i)$$

for each i, so $TS = \mathrm{id}_W$.

To see that (2) $\Rightarrow$ (1), if $T(\vec{v}_1) = T(\vec{v}_2)$, then $S(T(\vec{v}_1)) = S(T(\vec{v}_2))$. Since $ST = \mathrm{id}_V$ by assumption, this implies $\vec{v}_1 = \vec{v}_2$, so T is one-to-one. Given $\vec{w} \in W$, the fact that $TS = \mathrm{id}_W$ means that $\vec{w} = T(S(\vec{w}))$, so T is onto.

Finally, S is uniquely determined by the condition $ST = \mathrm{id}_V$ since this condition implies $S(T(\vec{e}_i)) = \vec{e}_i$ for each i. S is an isomorphism since it carries the basis $\mathcal{B}_2$ to $\mathcal{B}_1$. As to the last assertion, for any $\vec{w} \in W$, write

$$\vec{w} = \alpha_1 T(\vec{e}_1) + \alpha_2 T(\vec{e}_2) + \cdots + \alpha_n T(\vec{e}_n)$$

Then $\vec{w} = T(\vec{v})$, where

$$\vec{v} = \alpha_1 \vec{e}_1 + \alpha_2 \vec{e}_2 + \cdots + \alpha_n \vec{e}_n$$

Then $S(\vec{w}) = \vec{v}$. □

Definition 2.4.9. Given an isomorphism $T : V \to W$, the unique isomorphism $S : W \to V$ satisfying (2) of Theorem 2.4.8 is called the *inverse* of T, denoted by T^{-1}. Hence, $T : V \to W$ and $T^{-1} : W \to V$ are related by fundamental identities

$$T^{-1}T(\vec{v}) = \vec{v} \text{ for all } \vec{v} \in V$$

and

$$TT^{-1}(\vec{w}) = \vec{w} \text{ for all } \vec{w} \in W$$

In other words, each of T and T^{-1} reverses the action of the other. In particular, the equation in the proof of Theorem 2.4.8 $S(T(\vec{e}_i)) = \vec{e}_i$ shows how to define T^{-1} using the mage of a basis under the isomorphism T.

Corollary 2.4.10. *"$\simeq$" forms an equivalence relation on the class of all isomorphic vector spaces That is, if U, V, and W are vector spaces over $\mathbb{F}$, then*

1. *$V \simeq V$ for every vector space V.*
2. *If $V \simeq W$, then $W \simeq V$.*
3. *If $U \simeq V$ and $V \simeq W$, then $U \simeq W$.*

Proof. To see that "$\simeq$" forms an equivalence relation, we need to verify reflexivity, symmetry, and transitivity. For (1), note that by Example 2.4.2, the identity operator is an isomorphism from V to V, so $V \simeq V$ for any vector space V. For (2), if $V \simeq W$ then there exists an isomorphism $T : V \to W$. But because T is a bijection, there exists an inverse $T^{-1} : W \to V$ such that T^{-1} is an isomorphism, so $W \simeq V$. Finally, if $U \simeq V$ and $V \simeq W$, then there exists isomorphisms $T : U \to V$ and $S : V \to W$. Then taking the composition $ST : U \to W$ is also an isomorphism, so $U \simeq W$. □

Corollary 2.4.11. *If V is a finite-dimensional vector space over $\mathbb{F}$ with $\dim(V) = n$, then $V \simeq \mathbb{R}^n$.*

If V is a vector space of dimension n, note that there are important explicit isomorphisms $V \to \mathbb{R}^n$. Fix a basis $\mathcal{B} = \{\vec{b}_1, \vec{b}_2, ..., \vec{b}_n\}$ of V, and write $\{\vec{e}_1, \vec{e}_2, ..., \vec{e}_n\}$ for the standard basis of $\mathbb{R}^n$. By Theorem 2.4.5, there is a unique linear combination $C_{\mathcal{B}} : V \to \mathbb{R}^n$ given by

$$C_{\mathcal{B}}(\alpha_1\vec{b}_1 + \alpha_2\vec{b}_2 + \cdots + \alpha_n\vec{b}_n) = \alpha_1\vec{e}_1 + \alpha_2\vec{e}_2 + \cdots + \alpha_n\vec{e}_n = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix}$$

where each $\alpha_i \in \mathbb{R}$. Moreover, $C_{\mathcal{B}}(\vec{b}_i) = \vec{e}_i$ for each i, so $C_{\mathcal{B}}$ is an isomorphism by Theorem 2.4.5, called the *coordinate isomorphism* corresponding to the basis $\mathcal{B}$. Such isomorphisms will be talked more about in the next chapter.

Example 2.4.12. Let $V = \mathcal{S}_2(\mathbb{R})$. Suppose we want to find an isomorphism $T : \mathcal{P}_2(\mathbb{R}) \to \mathcal{S}_2(\mathbb{R})$ such that $T(1) = I_2$, where I_2 is the 2×2 identity matrix.

Note that $\{1, x, x^2\}$ is a basis of $\mathcal{P}_2(\mathbb{R})$ and we want a basis of $\mathcal{S}_2(\mathbb{R})$ containing I_2. Indeed, the set

$$\mathcal{B} = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\}$$

is linearly independent of V, so it is a basis because $\dim(V) = 3$ (see Example 1.3.21). Hence, define $T : \mathcal{P}_2(\mathbb{R}) \to \mathcal{S}_2$ by taking $T(1) = I_2$, $T(x) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$, and $T(x^2) = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$, and extending linearly as in Theorem 2.1.17. T is an isomorphism by Theorem 2.1.17 and its action is given by

$$T(a + bx + cx^2) = aT(1) + bT(x) + cT(x^2) = \begin{bmatrix} a & b \\ b & a+c \end{bmatrix}$$

The Dimension Theorem (Theorem 2.3.1) gives the following useful fact about isomorphisms.

Theorem 2.4.13. *If V and W are finite-dimensional vector spaces over $\mathbb{F}$ with $\dim(V) = \dim(W) = n$, a linear transformation $T : V \to W$ is an isomorphism if it is either one-to-one or onto.*

Proof. The Dimension Theorem asserts that $\dim(\ker(T)) + \dim(\operatorname{ran}(T)) = n$, so $\dim(\ker(T)) = 0$ if and only if $\dim(\operatorname{ran}(T)) = n$, so T is one-to-one if and only if T is onto, and thus, the result follows. $\square$

Example 2.4.14. Let $T : \mathcal{P}_1(\mathbb{R}) \to \mathcal{P}_1(\mathbb{R})$ be defined by $T(a + bx) = (a - b) + ax$. Then T is a linear transformation. Note that since $T(1) = 1 + x$ and $T(x) = -1$, T carries the basis $\mathcal{B}_1 = \{1, X\}$ to the basis $\mathcal{B}_2 = \{1+x, -1\}$. Hence, T is an isomorphism, and T^{-1} carries $\mathcal{B}_2$ back to $\mathcal{B}_1$, that is,

$$T^{-1}(1 + x) = 1 \text{ and } T^{-1}(-1) = x$$

Since $a + bx = b(1 + x) + (b - a)(-1)$, we obtain

$$T^{-1}(a + bx) = bT^{-1}(1 + x) + (b - a)T^{-1}(-1) = b + (b - a)x$$

Example 2.4.15. If $\mathcal{B} = \{\vec{b}_1, \vec{b}_2, ..., \vec{b}_n\}$ is a basis of a vector space V, the coordinate transformation $C_{\mathcal{B}} : V \to \mathbb{R}^n$ is an isomorphism defined by

$$C_{\mathcal{B}}(\alpha_1\vec{b}_1 + \alpha_2\vec{b}_2 + \cdots + \alpha_n\vec{b}_n) = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix}$$

The way to reverse the action of $C_{\mathcal{B}}$ is clear: $C_{\mathcal{B}}^{-1} : \mathbb{R}^n \to V$ is given by

$$C_{\mathcal{B}}^{-1}\left(\begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix}\right) = \alpha_1\vec{b}_1 + \alpha_2\vec{b}_2 + \cdots + \alpha_n\vec{b}_n$$

for all $\alpha_i \in V$.

Example 2.4.16. Define $T : \mathbb{R}^3 \to \mathbb{R}^3$ to be the linear transformation defined by $T(x, y, z) = (z, x, y)$. Note that

$$T^2(x, y, z) = T(T(x, y, z)) = T(z, x, y) = (y, z, x)$$

and then

$$T^3(x, y, z) = T(T^2(x, y, z)) = T(y, z, x) = (x, y, z)$$

Since this holds for all $(x, y, z) \in \mathbb{R}^3$, $T^3 = \mathrm{id}_{\mathbb{R}^3}$, so $T(T^2) = (T^2)T$, so $T^{-1} = T^2$.

Chapter 3

Change of Basis

One of the key insights in linear algebra is that a vector space can be represented in many different ways, depending on the choice of basis. A basis provides a coordinate system for a vector space, and changing the basis corresponds to changing the perspective from which we view vectors and linear transformations.

In many applications—whether in computer graphics, differential equations, or quantum mechanics—it becomes useful or even necessary to switch between different bases. For example, a problem might be much simpler to analyze in a carefully chosen basis where the structure of a transformation is clearer, such as one that diagonalizes a matrix or aligns with the geometry of the problem.

The main goal of this chapter is to understand how vectors and linear transformations behave under a change of basis. We will explore the representation of vectors with respect to arbitrary bases, the computation of transition (or change-of-basis) matrices, and how matrix representations of linear transformations change when switching between different bases. This process of translating between coordinate systems lies at the heart of many areas of mathematics and physics and provides deeper insight into the intrinsic nature of linear operators.

3.1 The Matrix of a Linear Transformation

In this section, we focus on how to represent a linear transformation between finite-dimensional vector spaces as a matrix, once bases are chosen. This matrix allows us to compute the action of the transformation using matrix multiplication, and it serves as a concrete bridge between abstract linear maps and numerical computation.

Turning vectors into column vectors is straightforward, but there's one important adjustment to make. So far, the order of vectors in a basis hasn't mattered. However, in this section, we will work with an *ordered basis* $\vec{b}_1, \vec{b}_2, ..., \vec{b}_n$ a basis in which the sequence of the vectors is significant and must be preserved.

If $\mathcal{B} = \{\vec{b}_1, \vec{b}_2, ..., \vec{b}_n\}$ is an ordered basis in a vector space V, and if

$$\vec{v} = \alpha_1\vec{b}_1 + \alpha_2\vec{b}_2 + \cdots + \alpha_n\vec{b}_n$$

where $\alpha_i \in \mathbb{F}$, is a vector in V, then $\alpha_1, \alpha_2, ..., \alpha_n$ are called the *coordinates* of $\vec{v}$ with respect to the basis $\mathcal{B}$.

We introduced the notion of a coordinate transformation, so here we give the formal definition.

Definition 3.1.1. The *coordinate vector* of $\vec{v}$ with respect to $\mathcal{B}$ is

$$C_{\mathcal{B}}(\vec{v}) = [\vec{v}]_{\mathcal{B}} = [\alpha_1\vec{b}_1 + \alpha_2\vec{b}_2 + \cdots + \alpha_n\vec{b}_n]_{\mathcal{B}} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix}$$

Example 3.1.2. The coordinate vector for $\vec{v} = (2, 1, 3)$ with respect to the ordered basis

$$\mathcal{B} = \{(1, 1, 0), (1, 0, 1) + (0, 1, 1)\} \subseteq \mathbb{R}^3$$

is

$$[\vec{v}]_{\mathcal{B}} = \begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix}$$

because

$$\vec{v} = (2, 1, 3) + 0(1, 1, 0) + 2(1, 0, 1) + 1(0, 1, 1)$$

Example 3.1.3. Let $\vec{v} = (4, 2)$ and consider the ordered basis

$$\mathcal{B} = \{(1, 1), (2, 0)\} \subseteq \mathbb{R}^2$$

We want to find the coordinate vector of $\vec{v}$ with respect to $\mathcal{B}$. Suppose

$$\vec{v} = \alpha_1(1, 1) + \alpha_2(2, 0)$$

Then

$$(4, 2) = (\alpha_1 + 2\alpha_2, \alpha_1)$$

Solving this system, we get $\alpha_1 = 2$ and $\alpha_2 = 1$. Hence,

$$[\vec{v}]_{\mathcal{B}} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

Example 3.1.4. Let $\vec{v} = (3, 2, 1)$ and let the ordered basis be

$$\mathcal{B} = \{(1, 0, 0), (0, 2, 0), (0, 0, 3)\} \subseteq \mathbb{R}^3$$

To find the coordinate vector of $\vec{v}$ with respect to $\mathcal{B}$, write

$$\vec{v} = 3(1, 0, 0) + 1(0, 2, 0) + \frac{1}{3}(0, 0, 3)$$

That is,

$$\vec{v} = 3\vec{b}_1 + 1\vec{b}_2 + \tfrac{1}{3}\vec{b}_3$$

So the coordinate vector is

$$[\vec{v}]_{\mathcal{B}} = \begin{bmatrix} 3 \\ 1 \\ \frac{1}{3} \end{bmatrix}$$

Example 3.1.5. Let $\vec{v} = (1, 4, -1)$ and let the ordered basis be

$$\mathcal{B} = \{(1, 0, 1), (0, 1, 1), (1, 1, -1)\} \subseteq \mathbb{R}^3$$

We want to express $\vec{v}$ as a linear combination:

$$\vec{v} = \alpha_1(1, 0, 1) + \alpha_2(0, 1, 1) + \alpha_3(1, 1, -1)$$

Writing the equation:

$$(1, 4, -1) = (\alpha_1 + \alpha_3, \alpha_2 + \alpha_3, \alpha_1 + \alpha_2 - \alpha_3)$$

Solving the system gives $\alpha_1 = 1$, $\alpha_2 = 2$, $\alpha_3 = 0$, so

$$[\vec{v}]_{\mathcal{B}} = \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}$$

Example 3.1.6. Let $V = \mathcal{P}_2(\mathbb{R})$ and let

$$p(x) = 2 + 3x + 4x^2$$

Consider the ordered basis

$$\mathcal{B} = \{1, 1 + x, 1 + x + x^2\}$$

We want to find the coordinate vector of $p(x)$ relative to $\mathcal{B}$. Suppose

$$p(x) = \alpha_1(1) + \alpha_2(1+x) + \alpha_3(1+x+x^2)$$

Expanding,

$$p(x) = (\alpha_1 + \alpha_2 + \alpha_3) + (\alpha_2 + \alpha_3)x + \alpha_3 x^2$$

Matching coefficients:

$$\begin{aligned} \alpha_1 + \alpha_2 + \alpha_3 &= 2 \\ \alpha_2 + \alpha_3 &= 3 \\ \alpha_3 &= 4 \end{aligned}$$

Solving, we find $\alpha_3 = 4$, $\alpha_2 = -1$, and $\alpha_1 = -1$. Therefore,

$$[p(x)]_{\mathcal{B}} = \begin{bmatrix} -1 \\ -1 \\ 4 \end{bmatrix}$$

Example 3.1.7. Let $V = \mathcal{M}_2(\mathbb{R})$ and let

$$A = \begin{bmatrix} 2 & 1 \\ 0 & 3 \end{bmatrix}$$

Consider the ordered basis

$$\mathcal{B} = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\}$$

(i.e., the standard basis for $\mathcal{M}_2(\mathbb{R})$).

Then A is already expressed as:

$$A = 2\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + 1\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} + 0\begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} + 3\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

Hence,

$$[A]_{\mathcal{B}} = \begin{bmatrix} 2 \\ 1 \\ 0 \\ 3 \end{bmatrix}$$

Example 3.1.8. Let $V = \mathcal{P}_2(\mathbb{R})$ and consider

$$p(x) = 5x^2 - 3x + 2$$

and the ordered basis

$$\mathcal{B} = \{x^2, x, 1\}$$

Then $p(x)$ is already expressed in terms of the basis vectors:

$$p(x) = 5x^2 + (-3)x + 2 \cdot 1$$

Hence, the coordinate vector is

$$[p(x)]_{\mathcal{B}} = \begin{bmatrix} 5 \\ -3 \\ 2 \end{bmatrix}$$

Theorem 3.1.9. *If V is a finite-dimensional vector space over $\mathbb{F}$ with $\dim(V) = n$, and $\mathcal{B} = \{\vec{b}_1, \vec{b}_2, ..., \vec{b}_n\}$ is an ordered basis of V, the coordinate transformation $[\cdot]_{\mathcal{B}} : V \to \mathbb{F}^n$ is an isomorphism. In fact, $[\cdot]_{\mathcal{B}}^{-1} : \mathbb{F}^n \to V$ is given by*

$$\left[\begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix}\right]_{\mathcal{B}}^{-1} = \alpha_1 \vec{b}_1 + \alpha_2 \vec{b}_2 + \cdots + \alpha_n \vec{b}_n$$

for all $\begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix} \in \mathbb{F}^n$.

Proof. It is elementary to verify that $[\cdot]_{\mathcal{B}}$ is a linear transformation. If $T : \mathbb{F}^n \to V$ is the map denoted $[\cdot]_{\mathcal{B}}^{-1}$ in the theorem, then it is elementary to show that $T[\cdot]_{\mathcal{B}} = \mathrm{id}_V$ and $[\cdot]_{\mathcal{B}} T = \mathrm{id}_{\mathbb{F}^n}$. Note that $[\vec{b}_j]_{\mathcal{B}}$ is the column j of the identity matrix, so $[\cdot]_{\mathcal{B}}$ carries the basis $\mathcal{B}$ to the standard basis of $\mathbb{F}^n$, so $[\cdot]_{\mathcal{B}}$ is indeed an isomorphism. $\square$

Let $T : V \to W$ be a linear transformation where $\dim(V) = n$ and $\dim(W) = m$, and let $\mathcal{A} = \{\vec{v}_1, \vec{v}_2, ..., \vec{v}_n\}$ and $\mathcal{B} = \{\vec{w}_1, \vec{w}_2, ..., \vec{w}_m\}$ be ordered bases for V and W, respectively.

The coordinate mappings

$$[\cdot]_{\mathcal{A}} : V \to \mathbb{F}^n \quad \text{and} \quad [\cdot]_{\mathcal{B}} : W \to \mathbb{F}^m$$

are both vector space isomorphisms.

Thus, we obtain the following commutative diagram:

$$\begin{tikzcd} V \arrow[r, "T"] \arrow[d, dashed, "{[\cdot]_{\mathcal{A}}}"'] & W \arrow[d, dashed, "{[\cdot]_{\mathcal{B}}}"] \\ \mathbb{F}^n \arrow[r, dashed, "T_A"'] & \mathbb{F}^m \end{tikzcd}$$

where $T_A : \mathbb{F}^n \to \mathbb{F}^m$ is a linear transformation given by left-multiplication by an $m \times n$ matrix A.

In fact, the composite map

$$T_A = [\cdot]_{\mathcal{B}} \circ T \circ [\cdot]_{\mathcal{A}}^{-1} = [T([\cdot]_{\mathcal{A}}^{-1})]_{\mathcal{B}}$$

is linear, and there exists a unique matrix $A \in \mathcal{M}_{m\times n}(\mathbb{F})$ such that

$$T_A(\vec{x}) = A\vec{x} \quad \text{for all} \quad \vec{x} \in \mathbb{F}^n$$

which corresponds to the condition

$$[T(\vec{v})]_{\mathcal{B}} = A[\vec{v}]_{\mathcal{A}}$$

for all $\vec{v} \in V$.

Thus, the action of T on V is fully captured by the matrix A once bases $\mathcal{A}$ and $\mathcal{B}$ are fixed. This discussion leads to the following definition.

Definition 3.1.10. Let V and W be finite-dimensional vector spaces over $\mathbb{F}$, let $\mathcal{A} = \{\vec{v}_1, \vec{v}_2, ..., \vec{v}_n\}$ and $\mathcal{B} = \{\vec{w}_1, \vec{w}_2, ..., \vec{w}_m\}$ be ordered bases of V and W, respectively, and let $T : V \to W$ be a linear transformation. The *$\mathcal{A}$-to-$\mathcal{B}$ matrix of* T is the matrix $[T]_{\mathcal{B}}^{\mathcal{A}} \in \mathcal{M}_{mn}(\mathbb{F})$ defined by

$$[T]_{\mathcal{B}}^{\mathcal{A}} = \begin{bmatrix} \uparrow & \uparrow & & \uparrow \\ [T(\vec{v}_1)]_{\mathcal{B}} & [T(\vec{v}_2)]_{\mathcal{B}} & \cdots & [T(\vec{v}_n)]_{\mathcal{B}} \\ \downarrow & \downarrow & & \downarrow \end{bmatrix}$$

In other words, $[T]_{\mathcal{B}}^{\mathcal{A}}$ is the unique $m \times n$ matrix such that

$$[T(\vec{v})]_{\mathcal{B}} = [T]_{\mathcal{B}}^{\mathcal{A}}[\vec{v}]_{\mathcal{A}}$$

Example 3.1.11. Let $T : \mathbb{R}^3 \to \mathbb{R}^2$ be the linear transformation defined by

$$T\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) = \begin{bmatrix} x+y \\ y+z \end{bmatrix}$$

Let

$$\mathcal{A} = \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}, \quad \mathcal{B} = \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}$$

be the standard bases for $\mathbb{R}^3$ and $\mathbb{R}^2$, respectively. To find $[T]_{\mathcal{B}}^{\mathcal{A}}$, first we compute

$$T\left(\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad T\left(\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad T\left(\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

In this case, the coordinates of each vector is the same, so

$$\left[T\left(\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}\right)\right]_{\mathcal{B}} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad \left[T\left(\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}\right)\right]_{\mathcal{B}} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \left[T\left(\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}\right)\right]_{\mathcal{B}} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

Therefore, the $\mathcal{A}$-to-$\mathcal{B}$ matrix of T is

$$[T]_{\mathcal{B}}^{\mathcal{A}} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}$$

Example 3.1.12. Let $T : \mathcal{M}_2(\mathbb{R}) \to \mathcal{M}_2(\mathbb{R})$ be the linear transformation defined by

$$T(A) = A^T$$

and let

$$\mathcal{A} = \mathcal{B} = \left\{ E_{11} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, E_{12} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, E_{21} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, E_{22} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\}$$

i.e. the standard basis of $\mathcal{M}_2(\mathbb{R})$. To find $[T]_{\mathcal{B}}^{\mathcal{A}}$, we compute

$$T(E_{11}) = E_{11}, \quad T(E_{12}) = E_{21}, \quad T(E_{21}) = E_{12}, \quad T(E_{22}) = E_{22}$$

Then the coordinates of each vector are

$$[T(E_{11})]_{\mathcal{B}} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad [T(E_{12})]_{\mathcal{B}} = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \quad [T(E_{21})]_{\mathcal{B}} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \quad [T(E_{22})]_{\mathcal{B}} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

Therefore,

$$[T]_{\mathcal{B}}^{\mathcal{A}} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Now suppose that we want to determine the coordinate vector $[T(A)]_{\mathcal{B}}$, where

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$$

First, note that $[A]_{\mathcal{A}} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}$, and since

$$A = 1 \cdot E_{11} + 2 \cdot E_{12} + 3 \cdot E_{21} + 4 \cdot E_{22}$$

Then

$$[T(A)]_{\mathcal{B}} = [T]_{\mathcal{B}}^{\mathcal{A}} [A]_{\mathcal{A}} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \\ 2 \\ 4 \end{bmatrix}$$

Note that this coordinate corresponds to the matrix

$$A^T = \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix}$$

Example 3.1.13. Let $D : \mathcal{P}_3(\mathbb{R}) \to \mathcal{P}_2(\mathbb{R})$ be the derivative map defined by $D(p) = p'$. Let $\mathcal{A} = \{1, x, x^2, x^3\}$ and $\mathcal{B} = \{1, x, x^2\}$ be ordered bases for $\mathcal{P}_3(\mathbb{R})$ and $\mathcal{P}_2(\mathbb{R})$, respectively. Then to find $[D]_{\mathcal{B}}^{\mathcal{A}}$, here, we note that

$$D(1) = 0, \quad D(x) = 1, \quad D(x^2) = 2x, \quad D(x^3) = 3x^2$$

So then

$$[D(1)]_{\mathcal{B}} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \quad [D(x)]_{\mathcal{B}} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad [D(x^2)_{\mathcal{B}}] = \begin{bmatrix} 0 \\ 2 \\ 0 \end{bmatrix}, \quad [D(x^3)]_{\mathcal{B}} = \begin{bmatrix} 0 \\ 0 \\ 3 \end{bmatrix}$$

Therefore,

$$[D]_{\mathcal{B}}^{\mathcal{A}} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix}$$

Now say that we want to determine the coordinate $[D(2+3x+x^2+x^3)]_{\mathcal{B}}$.

Observe that

$$\begin{aligned}[D(2+3x+x^2+x^3)]_{\mathcal{B}} &= [D]_{\mathcal{B}}^{\mathcal{A}}[2+3x+x^2+x^3]_{\mathcal{A}} \\ &= \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \\ 1 \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} 3 \\ 2 \\ 3 \end{bmatrix}\end{aligned}$$

Example 3.1.14. Let $T : \mathcal{P}_2(\mathbb{R}) \to \mathcal{P}_2(\mathbb{R})$ be the linear transformation defined by

$$T(p(x)) = p(x) + p'(x)$$

and let $\mathcal{A} = \mathcal{B} = \{1, x, x^2\}$. To find $[T]_{\mathcal{B}}^{\mathcal{A}}$, we first compute

$$T(1) = 1, \quad T(x) = x+1, \quad T(x^2) = x^2 + 2x$$

Then the coordinates of each vector are

$$[T(1)]_{\mathcal{B}} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad [T(x)]_{\mathcal{B}} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \quad [T(x^2)]_{\mathcal{B}} = \begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix}$$

Therefore,

$$[T]_{\mathcal{B}}^{\mathcal{A}} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{bmatrix}$$

Example 3.1.15. Define the linear map $T : \mathbb{R}^2 \to \mathbb{R}^2$ by

$$T\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} \frac{3}{5}x + \frac{4}{5}y \\ \frac{4}{5}x - \frac{3}{5}y \end{bmatrix}$$

Let $\mathcal{A} = \left\{ \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 2 \end{bmatrix} \right\}$ be a basis of $\mathbb{R}^2$. Note that

$$T\left(\begin{bmatrix} 2 \\ 1 \end{bmatrix}\right) = 1 \cdot \begin{bmatrix} 2 \\ 1 \end{bmatrix} + 0 \cdot \begin{bmatrix} -1 \\ 2 \end{bmatrix}$$

and similarly,

$$T\left(\begin{bmatrix} -1 \\ 2 \end{bmatrix}\right) = 0 \cdot \begin{bmatrix} 2 \\ 1 \end{bmatrix} + 1 \cdot \begin{bmatrix} -1 \\ 2 \end{bmatrix}$$

Then
$$\left[T\left(\begin{bmatrix}2\\1\end{bmatrix}\right)\right]_{\mathcal{A}} = \begin{bmatrix}1\\0\end{bmatrix}, \quad \left[T\left(\begin{bmatrix}-1\\2\end{bmatrix}\right)\right]_{\mathcal{A}} = \begin{bmatrix}0\\1\end{bmatrix}$$
Therefore,
$$[T]_{\mathcal{A}}^{\mathcal{A}} = \begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}$$

In the following example, we will connect the kernel and the range of the linear map.

Example 3.1.16. Let $\mathcal{A} = \{e^x, e^{2x}, e^{3x}\}$ and $\mathcal{B} = \{1, \cos(x), \sin(x)\}$ which are linearly independent subsets of $\mathcal{C}(\mathbb{R})$. Let $V = \operatorname{span}(\mathcal{A})$ and $W = \operatorname{span}(\mathcal{B})$ and let $T : V \to W$ be a linear map such that
$$[T]_{\mathcal{B}}^{\mathcal{A}} = \begin{bmatrix}2 & 3 & 8\\3 & 0 & 3\\2 & 1 & 4\end{bmatrix}$$
First, let us find $T(\vec{v})$ for every $\vec{v} \in V$. Note that if $[\vec{v}]_{\mathcal{A}} = \begin{bmatrix}a\\b\\c\end{bmatrix}$, then
$$[T]_{\mathcal{B}}^{\mathcal{A}}[\vec{v}]_{\mathcal{A}} = \begin{bmatrix}2 & 3 & 8\\3 & 0 & 3\\2 & 1 & 4\end{bmatrix}\begin{bmatrix}a\\b\\c\end{bmatrix} = \begin{bmatrix}2a+3b+8c\\3a+3c\\2a+b+4c\end{bmatrix}$$
Then
$$T(\vec{v}) = T(ae^x + be^{2x} + ce^{3x}) = (2a+3b+c)\cdot 1 + (3a+3c)\cos(x) + (2a+b+4c)\sin(x)$$

Now let us determine $\ker(T)$. Indeed, note that $\vec{v} \in \ker(T)$ if and only if $T(\vec{v}) = \vec{0}$ if and only if $[T(\vec{v})]_{\mathcal{B}} = \begin{bmatrix}0\\0\\0\end{bmatrix}$ if and only if $[T]_{\mathcal{B}}^{\mathcal{A}}[\vec{v}]_{\mathcal{A}} = \begin{bmatrix}0\\0\\0\end{bmatrix}$, so $[\vec{v}]_{\mathcal{A}} \in \ker([T]_{\mathcal{B}}^{\mathcal{A}})$. Thus, we apply RREF
$$\left[\begin{array}{ccc|c}2 & 3 & 8 & 0\\3 & 0 & 3 & 0\\2 & 1 & 4 & 0\end{array}\right] \to \left[\begin{array}{ccc|c}1 & 0 & 1 & 0\\0 & 1 & 2 & 0\\0 & 0 & 0 & 0\end{array}\right]$$
and thus, $\ker(T) = \operatorname{span}\left\{\begin{bmatrix}-1\\2\\1\end{bmatrix}\right\}$. Thus, we then get
$$\ker(T) = \operatorname{span}\{-e^x - 2e^{2x} + e^{3x}\}$$

and $\{-e^x - 2x^{2x} + e^{3x}\}$ is a basis of $\ker(T)$.

Similarly, suppose we want to find $\operatorname{ran}(T)$. Note that

$$\operatorname{ran}(T) = \operatorname{span}\{T(e^x), T(e^{2x}), T(e^{3x})\}$$

So

$$T(e^x) = 2 + 3\cos(x) + 2\sin(x), T(e^{2x}) = 3 + \sin(x), T(e^{3x}) = 8 + 3\cos(x) + 4\sin(x)$$

So the coordinates of each vector is then given by

$$[T(e^x)]_{\mathcal{B}} = \begin{bmatrix} 2 \\ 3 \\ 2 \end{bmatrix}, \quad [T(e^{2x})]_{\mathcal{B}} = \begin{bmatrix} 3 \\ 0 \\ 1 \end{bmatrix}, \quad [T(e^{3x})]_{\mathcal{B}} = \begin{bmatrix} 8 \\ 3 \\ 4 \end{bmatrix}$$

Furthermore, observe that

$$\begin{bmatrix} 8 \\ 3 \\ 4 \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \\ 2 \end{bmatrix} + 2 \begin{bmatrix} 3 \\ 0 \\ 1 \end{bmatrix}$$

which implies that

$$[T(e^{3x})]_{\mathcal{B}} = [T(e^x)]_{\mathcal{B}} + 2[T(e^{2x})]_{\mathcal{B}} = [T(e^x) + 2T(e^{2x})]_{\mathcal{B}}$$

So

$$T(e^{3x}) = T(e^x) + 2T(e^{2x})$$

In other words, the vector e^{3x} is redundant, so we have

$$\operatorname{ran}(T) = \operatorname{span}\{T(e^x), T(e^{2x})\}$$

and since $T(e^x), T(e^{2x})$ are linearly independent, we have that $\{T(e^x), T(e^{2x})\}$ is a basis of $\operatorname{ran}(T)$.

Example 3.1.17. Let V and W be vector spaces over $\mathbb{F}$ with ordered bases $\mathcal{A}$ and $\mathcal{B}$. Let $\dim(V) = n$. Then

1. $\mathrm{id}_V : V \to V$ has the matrix $[\mathrm{id}_V]_{\mathcal{B}}^{\mathcal{B}} = I_n$.
2. $O_V : V \to W$ has matrix $[O_V]_{\mathcal{B}}^{\mathcal{A}} = O$

The following results show that composition of linear transformations is compatible with multiplication of the corresponding matrices.

Theorem 3.1.18. *Let $V \xrightarrow{T} W \xrightarrow{S} U$ be linear transformations and let $\mathcal{A}, \mathcal{B}, \mathcal{C}$ be finite ordered bases of V, W, and U, respectively. Then*

$$[ST]_{\mathcal{C}}^{\mathcal{A}} = [S]_{\mathcal{C}}^{\mathcal{B}}[T]_{\mathcal{B}}^{\mathcal{A}}$$

Proof. If $\vec{v} \in V$, then see that

$$\begin{aligned} [S]_{\mathcal{C}}^{\mathcal{B}}[T]_{\mathcal{B}}^{\mathcal{A}}[\vec{v}]_{\mathcal{B}} &= [T]_{\mathcal{C}}^{\mathcal{B}}[T(\vec{v})]_{\mathcal{B}} \\ &= [ST(\vec{v})]_{\mathcal{C}} \\ &= [ST]_{\mathcal{C}}^{\mathcal{A}}[\vec{v}]_{\mathcal{A}} \end{aligned}$$

If $\mathcal{A} = \{\vec{a}_1, \vec{a}_2, ..., \vec{a}_n\}$, then $[\vec{a}_j]_{\mathcal{A}}$ is the jth column of I_n. Hence, taking $\vec{v} = \vec{a}_j$ shows that $[S]_{\mathcal{C}}^{\mathcal{B}}[T]_{\mathcal{B}}^{\mathcal{A}}$ and $[ST]_{\mathcal{C}}^{\mathcal{A}}$ have equal jth columns, so the theorem follows. □

Theorem 3.1.19. *Let V and W be finite-dimensional vector spaces over $\mathbb{F}$ with $\dim(V) = \dim(W) = n$, and let $T : V \to W$ be a linear transformation. The following are equivalent.*

1. *T is an isomorphism.*
2. *$[T]_{\mathcal{B}}^{\mathcal{A}}$ is invertible for all ordered bases $\mathcal{A}$ and $\mathcal{B}$ of V and W.*
3. *$[T]_{\mathcal{B}}^{\mathcal{A}}$ is invertible for some pair of ordered bases $\mathcal{A}$ and $\mathcal{B}$ of V and W.*

Proof. The statement that (2) $\Rightarrow$ (3) is obvious, so we will show (1) $\Rightarrow$ (2) and (3) $\Rightarrow$ (1).

To see that (1) $\Rightarrow$ (2), note that we have $V \xrightarrow{T} W \xrightarrow{T^{-1}} V$, so Theorem 3.1.18 implies that

$$[T^{-1}]_{\mathcal{A}}^{\mathcal{B}}[T]_{\mathcal{B}}^{\mathcal{A}} = [T^{-1}T]_{\mathcal{B}}^{\mathcal{B}} = [\mathrm{id}_V]_{\mathcal{B}}^{\mathcal{B}} = I_n$$

Similarly, we have

$$[T]_{\mathcal{B}}^{\mathcal{A}}[T^{-1}]_{\mathcal{A}}^{\mathcal{B}} = I_n$$

proving (2).

To see that (3) $\Rightarrow$ (1), assume that $[T]_{\mathcal{B}}^{\mathcal{A}}$ is invertible for some bases $\mathcal{A}$ and $\mathcal{B}$, and write $A = [T]_{\mathcal{B}}^{\mathcal{A}}$. Then we have $[\cdot]_{\mathcal{B}}T = T_A[\cdot]_{\mathcal{A}}$, so

$$T = [\cdot]_{\mathcal{B}}^{-1}T_A[\cdot]_{\mathcal{A}}$$

by Theorem **??**, where $[\cdot]_{\mathcal{B}}^{-1}$ and $[\cdot]_{\mathcal{A}}$ are isomorphisms. Hence, (1) follows if we can demonstrate that $T_A : \mathbb{R}^n \to \mathbb{R}^n$ is an isomorphism. But A is invertible by (3), and one can show that $T_A T_{A^{-1}} = T_{A^{-1}}T_A = \mathrm{id}_{\mathbb{R}^n}$, so T_A is invertible. □

Theorem 3.1.20. *Let V and W be vector spaces over $\mathbb{F}$ with $\dim(V) = n$ and $\dim(W) = m$, and let $T : V \to W$ be a linear transformation. If $\mathcal{A}$ and $\mathcal{B}$ are any ordered bases of V and W, then $\operatorname{rank}(T) = \operatorname{rank}([T]_{\mathcal{B}}^{\mathcal{A}})$.*

Proof. Write $A = [T]_{\mathcal{B}}^{\mathcal{A}}$. The column space of A is $U = \{A\vec{x} : \vec{x} \in \mathbb{R}^n\}$, so $\operatorname{rank}(A) = \dim(U)$, so because $\operatorname{rank}(T) = \dim(\operatorname{ran}(T))$, it suffices to find an isomorphism $S : \operatorname{ran}(T) \to U$. Now every vector in $\operatorname{ran}(T)$ has the form $T(\vec{v})$ for $\vec{v} \in V$. Then $[T(\vec{v})]_{\mathcal{B}} = A[\vec{v}]_{\mathcal{A}}$ is in U, so define $S : \operatorname{ran}(T) \to U$ by

$$S(T(\vec{v})) = [T(\vec{v})]_{\mathcal{B}}$$

for all $T(\vec{v}) \in \operatorname{ran}(T)$.

The fact that $[\cdot]_{\mathcal{B}}$ is linear and one-to-one implies that S is linear and one-to-one. To see that S is onto, let $A\vec{x} \in U$, with $\vec{x} \in \mathbb{R}^n$. Then $\vec{x} = [\vec{v}]_{\mathcal{B}}$ for some $\vec{v} \in V$ since $[\cdot]_{\mathcal{A}}$ is onto. Hence,

$$A\vec{x} = A[\vec{v}]_{\mathcal{A}} = [T(\vec{v})]_{\mathcal{B}} = S(T(\vec{v}))$$

Therefore, S is an isomorphism. □

3.2 Change of Basis

In previous sections, we fixed a basis and computed the coordinate vector of a given vector or the matrix of a linear transformation relative to that basis. However, in many applications, we may wish to switch between different bases. The process of expressing vectors or linear transformations in a new basis is called a *change of basis*.

This section is devoted to understanding how coordinate vectors and linear transformations transform when we change the basis. The key tool is the *change of basis matrix*, which provides a bridge between different coordinate systems.

For example, let $V = \mathcal{P}_3(\mathbb{R})$ and consider its basis given by $\mathcal{A} = \{1, (x-2), (x-2)^2, (x-2)^3\}$, and consider $p(x) = 1 + 2x + 3x^2 + 4x^3 \in \mathcal{P}_3(\mathbb{R})$. Suppose we want to determine what $[p(x)]_{\mathcal{A}}$ is. What is means is that we need to first find $\alpha_1, \alpha_2, \alpha_3, \alpha_4$ such that

$$p(x) = \alpha_1 + \alpha_2(x-2) + \alpha_3(x-2)^2 + \alpha_4(x-2)^3$$

and then $[p(x)]_{\mathcal{A}} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \end{bmatrix}$. To find these scalars, however, is not an easy task. If we were given the standard basis $\mathcal{B} = \{1, x, x^2, x^3\}$ instead, then this would be very easy, as $[p(x)]_{\mathcal{B}} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}$. The question is, given a vector space V, a vector $\vec{v} \in V$, and two bases $\mathcal{A}$ and $\mathcal{B}$, can we find $[\vec{v}]_{\mathcal{A}}$ using $[\vec{v}]_{\mathcal{B}}$? The answer to this is yes! We require the concepts of change of basis.

Definition 3.2.1. Let V be a finite-dimensional vector space over $\mathbb{F}$, let $\mathcal{A} = \{\vec{v}_1, \vec{v}_2, ..., \vec{v}_n\}$ and $\mathcal{B} = \{\vec{w}_1, \vec{w}_2, ..., \vec{w}_n\}$ be ordered bases of V. The *$\mathcal{A}$-to-$\mathcal{B}$ change of basis matrix* is the matrix $A \in \mathcal{M}_n(\mathbb{F})$ such that $[\vec{v}]_{\mathcal{B}} = A[\vec{v}]_{\mathcal{A}}$ for all $\vec{v} \in V$. In this case, the matrix A is given as

$$A = [\mathrm{id}_V]_{\mathcal{B}}^{\mathcal{A}} = \begin{bmatrix} \uparrow & \uparrow & & \uparrow \\ [\vec{v}_1]_{\mathcal{B}} & [\vec{v}_2]_{\mathcal{B}} & \cdots & [\vec{v}_n]_{\mathcal{B}} \\ \downarrow & \downarrow & & \downarrow \end{bmatrix}$$

since $[\mathrm{id}_V]_{\mathcal{B}}^{\mathcal{A}}[\vec{v}]_{\mathcal{A}} = [\mathrm{id}_V(\vec{v})]_{\mathcal{B}}$ for all $\vec{v} \in V$.

Example 3.2.2. Going back to the discussion when $V = \mathcal{P}_3(\mathbb{R})$, $\mathcal{A} = \{1, (x-2), (x-2)^2, (x-2)^3\}$ and $\mathcal{B} = \{1, x, x^2, x^3\}$, we have the $\mathcal{A}$-to-$\mathcal{B}$ change of basis matrix given by

$$[\mathrm{id}_V]_{\mathcal{B}}^{\mathcal{A}} = \begin{bmatrix} 1 & -2 & 4 & -8 \\ 0 & 1 & -4 & 12 \\ 0 & 0 & 1 & -6 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

On the other hand, it is not hard to see that $[\mathrm{id}_V]_{\mathcal{B}}^{\mathcal{A}}$ is an invertible matrix, so we can find its inverse, which is then the $\mathcal{B}$-to$-\mathcal{A}$ change of basis matrix, given as

$$[\mathrm{id}_V]_{\mathcal{A}}^{\mathcal{B}} = ([\mathrm{id}_V]_{\mathcal{B}}^{\mathcal{A}})^{-1} = \begin{bmatrix} 1 & 2 & 4 & 8 \\ 0 & 1 & 4 & 12 \\ 0 & 0 & 1 & 6 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Now considering the polynomial $p(x) = 1 + 2x + 3x^2 + 4x^3 \in \mathcal{P}_3(\mathbb{R})$ from above, suppose we want to find $\alpha_1, \alpha_2, \alpha_3, \alpha_4 \in \mathbb{R}$ such that

$$p(x) = \alpha_1 + \alpha_2(x-2) + \alpha_3(x-2)^2 + \alpha_4(x-2)^3$$

Note that $[\vec{v}]_{\mathcal{B}} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}$ from above, so then

$$[\vec{v}]_{\mathcal{A}} = ([\mathrm{id}_V]_{\mathcal{B}}^{\mathcal{A}})^{-1}[\vec{v}]_{\mathcal{B}} = \begin{bmatrix} 1 & 2 & 4 & 8 \\ 0 & 1 & 4 & 12 \\ 0 & 0 & 1 & 6 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix} = \begin{bmatrix} 49 \\ 62 \\ 21 \\ 4 \end{bmatrix}$$

In other words, $p(x) = 49 + 62(x-2) + 21(x-2)^2 + 4(x-2)^3$.

Example 3.2.3. In $\mathbb{R}^2$, we have the standard basis given by $\mathcal{SB} = \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}$. Given $\mathcal{B} = \left\{ \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \end{bmatrix} \right\}$, say we want to determine the coordinates of $\begin{bmatrix} 4 \\ 16 \end{bmatrix}$ with respect to $\mathcal{B}$. Here, we know that

$$[\vec{v}]_{\mathcal{SB}} = \begin{bmatrix} 4 \\ 16 \end{bmatrix}$$

Also,

$$[\mathrm{id}_V]_{\mathcal{SB}}^{\mathcal{B}} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

which is invertible, so

$$[\mathrm{id}_V]_{\mathcal{B}}^{\mathcal{SB}} = \frac{1}{3} \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$$

Therefore,

$$[\vec{v}]_{\mathcal{B}} = [\mathrm{id}_V]_{\mathcal{B}}^{\mathcal{SB}}[\vec{v}]_{\mathcal{SB}} = \frac{1}{3} \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} 4 \\ 16 \end{bmatrix} = \frac{1}{3} \begin{bmatrix} 8 \\ 28 \end{bmatrix}$$

Chapter 4

Inner Product Spaces

In previous chapters, we studied vector spaces purely from an algebraic point of view, focusing on their structure in terms of addition, scalar multiplication, bases, and linear transformations. However, many problems in mathematics and applied disciplines such as physics, engineering, and data science require more than just algebraic structure—they also depend on geometric notions such as angles, lengths, and orthogonality.

To incorporate these ideas into linear algebra, we enrich our vector spaces with an additional structure known as an *inner product*. An inner product allows us to define and compute lengths of vectors (via norms), angles between vectors (via dot products), and the concept of orthogonality (perpendicularity). This geometric perspective leads to powerful techniques such as projection, orthogonal decomposition, and the Gram-Schmidt process.

The study of inner product spaces opens the door to many important results and applications. It lays the foundation for understanding orthonormal bases, diagonalization of symmetric matrices, Fourier analysis, and least-squares approximations. These tools are essential in numerical methods, signal processing, machine learning, and more.

In this chapter, we will formally define inner products, explore their properties, and investigate how they give rise to geometry in vector spaces. We will study orthogonality, normed spaces, and orthogonal projections, and we will learn how to construct orthonormal bases using the Gram-Schmidt procedure. These concepts will prepare us to work with advanced topics where geometry and algebra meet in a profound and elegant way. We will be focusing on vector spaces over $\mathbb{K}$, in which case $\mathbb{K} = \mathbb{R}$ or $\mathbb{K} = \mathbb{C}$.

4.1 Inner Products and Norms

In earlier chapters, we studied vector spaces from an algebraic perspective: focusing on linear combinations, bases, and linear transformations. Now we introduce a geometric structure on vector spaces through the concept of an *inner product*, which allows us to define angles, lengths, and orthogonality.

Definition 4.1.1. Let V be a vector space over $\mathbb{K}$. An *inner product* is a function $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{K}$ such that

1. $\langle \vec{v}, \vec{v} \rangle \geq 0$ for all $\vec{v} \in V$ with equality if and only if $\vec{v} = \vec{0}$.
2. $\langle \vec{v} + \vec{w}, \vec{u} \rangle = \langle \vec{v}, \vec{u} \rangle + \langle \vec{w}, \vec{u} \rangle$ for all $\vec{v}, \vec{w}, \vec{u} \in V$.
3. $\langle \alpha\vec{v}, \vec{w} \rangle = \alpha \langle \vec{v}, \vec{w} \rangle$ for all $\vec{v}, \vec{w} \in V$ and $\alpha \in \mathbb{K}$.
4. $\langle \vec{v}, \vec{w} \rangle = \overline{\langle \vec{w}, \vec{v} \rangle}$ if $\mathbb{K} = \mathbb{C}$ and $\langle \vec{v}, \vec{w} \rangle = \langle \vec{w}, \vec{v} \rangle$ if $\mathbb{K} = \mathbb{R}$ for all $\vec{v}, \vec{w} \in V$.

The pair $(V, \langle \cdot, \cdot \rangle)$ is called an *inner product space*.

Remark 4.1.2. It is easy to see that $\left\langle \vec{0}, \vec{v} \right\rangle = \left\langle \vec{v}, \vec{0} \right\rangle = 0$ for all $\vec{v} \in V$. Indeed, note that by using (3),

$$\left\langle \vec{0}, \vec{v} \right\rangle = \langle 0 \cdot \vec{v}, \vec{v} \rangle = 0 \langle \vec{v}, \vec{v} \rangle = 0$$

as claimed.

Remark 4.1.3. As with linear transformations, we can compress axioms (2) and (3) of Definition 4.1.1 into one line, that is, for all $\vec{v}, \vec{w}, \vec{u} \in V$ and $\alpha \in \mathbb{K}$,

$$\langle \alpha\vec{v} + \vec{w}, \vec{u} \rangle = \alpha \langle \vec{v}, \vec{u} \rangle + \langle \vec{w}, \vec{u} \rangle$$

We now present some examples of inner product spaces. Some of these examples may seem familiar.

Example 4.1.4. Let $V = \mathbb{R}^n$ be over $\mathbb{R}$. Then $\mathbb{R}^n$ is an inner product space with the *dot product* as an inner product. That is, if $\vec{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$ and

$\vec{w} = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix}$ are in $\mathbb{R}^n$, then

$$\langle \vec{v}, \vec{w} \rangle = \vec{v} \cdot \vec{w} = \sum_{k=1}^{n} v_k w_k = v_1 w_1 + v_2 w_2 + \cdots + v_n w_n$$

This is also called the *Euclidean inner product*, and $\mathbb{R}^n$, equipped with the dot product, is called the *Euclidean n-space.* It is elementary to verify that the dot product is an inner product.

Example 4.1.5. Similarly, $\mathbb{C}^n$ over $\mathbb{C}$ is an inner product with the *dot product* as an inner product. That is, if $\vec{z} = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{bmatrix}$ and $\vec{w} = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix}$ are in $\mathbb{C}^n$, then

$$\langle \vec{z}, \vec{w} \rangle = \vec{z} \cdot \vec{w} = \sum_{k=1}^{n} z_k \bar{w}_k = z_1 \bar{w}_1 + z_2 \bar{w}_2 + \cdots + z_n \bar{w}_n$$

It is elementary to verify that this is an inner product.

Example 4.1.6. Consider $\mathcal{C}[0,1]$ over $\mathbb{R}$. Then for all $f, g \in \mathcal{C}[0,1]$, define $\langle f, g \rangle$ as follows:

$$\langle f, g \rangle = \int_0^1 f(x)g(x)\ dx$$

It is elementary to verify that $\mathcal{C}[0,1]$ with the above operation is an inner product space.

Example 4.1.7. If $A, B \in \mathcal{M}_{mn}(\mathbb{K})$ we define $\langle A, B \rangle = \operatorname{tr}(AB^*)$, where tr is the trace of the square matrix X, and B^* denotes the conjugate transpose of B. For example, if

$$B = \begin{bmatrix} 1 & 2i \\ 3+4i & 5+6i \end{bmatrix}$$

Then

$$B^T = \begin{bmatrix} 1 & 3+4i \\ 2i & 5+6i \end{bmatrix} \implies B^* = \begin{bmatrix} \bar{1} & \overline{3+4i} \\ \overline{2i} & \overline{5+6i} \end{bmatrix} = \begin{bmatrix} 1 & 3-4i \\ -2i & 5-6i \end{bmatrix}$$

It is elementary to verify that $\mathcal{M}_{mn}(\mathbb{K})$ with the above is an inner product space.

Theorem 4.1.8. *Let V be an inner product space over $\mathbb{K}$. Then for all $\vec{v}, \vec{w}, \vec{u} \in V$ and $\alpha \in \mathbb{K}$,*

1. $\langle \vec{v}, \alpha\vec{w} + \vec{u} \rangle = \overline{\alpha} \langle \vec{v}, \vec{w} \rangle + \langle \vec{v}, \vec{u} \rangle$.
2. *If* $\langle \vec{v}, \vec{v} \rangle = 0$ *if and only if* $\vec{v} = \vec{0}$.

Proof. To see that (1) holds, note that by using the definition of inner products

$$\begin{aligned}
\langle \vec{v}, \alpha\vec{w} + \vec{u} \rangle &= \overline{\langle \alpha\vec{w} + \vec{u}, \vec{v} \rangle} \\
&= \overline{\alpha \langle \vec{w}, \vec{v} \rangle + \langle \vec{u}, \vec{v} \rangle} \\
&= \overline{\alpha}\overline{\langle \vec{w}, \vec{v} \rangle} + \overline{\langle \vec{u}, \vec{v} \rangle} \\
&= \overline{\alpha} \langle \vec{v}, \vec{w} \rangle + \langle \vec{v}, \vec{u} \rangle
\end{aligned}$$

To see that (2) holds, note that one direction is obvious. Conversely, assume for a contradiction that $\vec{v} \neq 0$. Then $\langle \vec{v}, \vec{v} \rangle > 0$, so this contradicts the assumption that $\langle \vec{v}, \vec{v} \rangle = 0$, so $\vec{v} = \vec{0}$, as desired. □

In a general vector space V, we want to somehow be able to define the length of a vector $\vec{v} \in V$ in the same way we did in calculus.

Definition 4.1.9. Let V be an inner product space over $\mathbb{K}$. The *length* or *norm* of $\vec{v} \in V$, denoted by $\|\vec{v}\|$, is defined by

$$\|\vec{v}\| = \sqrt{\langle \vec{v}, \vec{v} \rangle}$$

We define the *distance* between vectors $\vec{v}$ and $\vec{w}$ in V to be

$$d(\vec{v}, \vec{w}) = \|\vec{v} - \vec{w}\|$$

Example 4.1.10. The norm of a continuous function $f \in \mathcal{C}[0,1]$ with the inner product defined in Example 4.1.6 is given as

$$\|f\| = \sqrt{\int_0^1 f^2(x)\ dx}$$

Hence, $\|f\|^2$ is the area beneath the graph of $f^2(x)$ between $x = 0$ and $x = 1$.

For example, if $f(x) = x$, then

$$\|f\| = \sqrt{\int_0^1 x^2\ dx} = \frac{1}{\sqrt{3}}$$

We now present some important identities that are commonly discussed using inner products.

Proposition 4.1.11 (Paralleogram Law). *Let V be an inner product space over $\mathbb{R}$. For every $\vec{v}, \vec{w} \in V$,*

$$\|\vec{v}+\vec{w}\|^2 + \|\vec{v}-\vec{w}\|^2 = 2\|\vec{v}\|^2 + 2\|\vec{w}\|^2$$

Proof. First, observe that

$$\begin{aligned}
\|\vec{v}+\vec{w}\|^2 &= \langle \vec{v}+\vec{w}, \vec{v}+\vec{w} \rangle \\
&= \langle \vec{v}, \vec{v}+\vec{w} \rangle + \langle \vec{w}, \vec{v}+\vec{w} \rangle \\
&= \langle \vec{v}, \vec{v} \rangle + \langle \vec{v}, \vec{w} \rangle + \langle \vec{w}, \vec{v} \rangle + \langle \vec{w}, \vec{w} \rangle \\
&= \|\vec{v}\|^2 + \langle \vec{v}, \vec{w} \rangle + \overline{\langle \vec{v}, \vec{w} \rangle} + \|\vec{w}\|^2 \\
&= \|\vec{v}\|^2 + 2\mathrm{Re}(\langle \vec{v}, \vec{w} \rangle) + \|\vec{w}\|^2
\end{aligned}$$

and similarly, $\|\vec{v}-\vec{w}\|^2 = \|\vec{v}\|^2 - 2\mathrm{Re}(\langle \vec{v}, \vec{w} \rangle) + \|\vec{w}\|^2$. Therefore, by adding the two quantities, we obtain

$$\|\vec{v}+\vec{w}\|^2 + \|\vec{v}-\vec{w}\|^2 = 2\|\vec{v}\|^2 + 2\|\vec{w}\|^2$$

as desired. □

Proposition 4.1.12 (Polarization Identity, $\mathbb{K} = \mathbb{R}$). *Let V be a vector space over $\mathbb{R}$. Then for all $\vec{v}, \vec{w} \in V$,*

$$\langle \vec{v}, \vec{w} \rangle = \frac{1}{4}(\|\vec{v}+\vec{w}\|^2 - \|\vec{v}-\vec{w}\|^2)$$

Proof. As in the proof of Proposition 4.1.11, we have $\|\vec{v}+\vec{w}\| = \|\vec{v}\|^2 + 2\langle \vec{v}, \vec{w} \rangle + \|\vec{w}\|^2$ and $\|\vec{v}-\vec{w}\|^2 = \|\vec{v}\|^2 - 2\langle \vec{v}, \vec{w} \rangle + \|\vec{w}\|^2$. Therefore, by subtracting the two quantities, we obtain

$$\|\vec{v}+\vec{w}\|^2 - \|\vec{v}-\vec{w}\|^2 = 4\langle \vec{v}, \vec{w} \rangle$$

Therefore,

$$\langle \vec{v}, \vec{w} \rangle = \frac{1}{4}(\|\vec{v}+\vec{w}\|^2 - \|\vec{v}-\vec{w}\|^2)$$

as claimed. □

Proposition 4.1.13 (Polarization Identity, $\mathbb{K} = \mathbb{C}$). *Let V be an inner vector space over $\mathbb{C}$. Then for all $\vec{v}, \vec{w} \in V$,*

$$\langle \vec{v}, \vec{w} \rangle = \frac{1}{4}\left(\|\vec{v}+\vec{w}\|^2 - \|\vec{v}-\vec{w}\|^2 + i\|\vec{v}+i\vec{w}\|^2 - i\|\vec{v}-i\vec{w}\|^2\right)$$

Proof. Exercise. □

The norm of a vector in a vector space has some interesting properties in which one can relate to the Euclidean n-space. But to see the properties in an abstract setting, we require a preliminary result.

Theorem 4.1.14 (Cauchy-Schwartz Inequality). *Let V be an inner product space over $\mathbb{K}$. Then for all $\vec{v}, \vec{w} \in V$,*

$$|\langle \vec{v}, \vec{w} \rangle| \leq \|\vec{v}\|\|\vec{w}\|$$

Moreover, equality occurs if and only if one of $\vec{v}$ and $\vec{w}$ is a scalar multiple of the other.

Proof. We prove for the case when $\mathbb{K} = \mathbb{R}$. Consider the vector $\vec{v} + \lambda\vec{w}$ for $\lambda \in \mathbb{R}$. Then observe that

$$\begin{aligned} \|\vec{v} - \lambda\vec{w}\|^2 &\geq 0 \\ \|\vec{v}\|^2 - 2\lambda \langle \vec{v}, \vec{w} \rangle + \lambda^2\|\vec{w}\|^2 &\geq 0 \end{aligned}$$

which is a quadratic equation in λ, and we note that the equation is nonnegative if and only if the discriminant is nonpositive. Therefore,

$$\begin{aligned} (-2\langle \vec{v}, \vec{w} \rangle)^2 - 4\|\vec{w}\|^2\|\vec{v}\|^2 &\leq 0 \\ 4\langle \vec{v}, \vec{w} \rangle^2 &\leq 4\|\vec{v}\|^2\|\vec{w}\|^2 \\ |\langle \vec{v}, \vec{w} \rangle| &\leq \|\vec{v}\|\|\vec{w}\| \end{aligned}$$

as desired. □

Example 4.1.15. If $V = \mathbb{K}^n$, then for any $\vec{v}, \vec{w} \in \mathbb{K}^n$,

$$\left|\sum_{k=1}^{n} v_k \bar{w}_k\right| \leq \left(\sum_{k=1}^{n} v_k^2\right)^{\frac{1}{2}} \left(\sum_{k=1}^{n} \bar{w}_k^2\right)^{\frac{1}{2}}$$

Example 4.1.16. If $V = \mathcal{C}[0,1]$, then for any $f, g \in \mathcal{C}[0,1]$,

$$\left|\int_0^1 f(x)g(x)\ dx\right| \leq \left(\int_0^1 f^2(x)\ dx\right)^{\frac{1}{2}} \left(\int_0^1 g^2(x)\ dx\right)^{\frac{1}{2}}$$

As it turns out, any inner product space is defined to be a *normed linear space.*

Definition 4.1.17. Let V be a vector space over $\mathbb{K}$. A function $\|\cdot\| : V \to [0,\infty)$ is called a *norm* if

1. $\|\vec{v}\| = 0$ if and only if $\vec{v} = \vec{0}$.
2. For all $\alpha \in \mathbb{K}$ and $\vec{v} \in V$, $\|\alpha\vec{v}\| = |\alpha|\|\vec{v}\|$.
3. For all $\vec{v}, \vec{w} \in V$, $\|\vec{v} + \vec{w}\| \leq \|\vec{v}\| + \|\vec{w}\|$.

The pair $(V, \|\cdot\|)$ is called a *normed linear space.*

Theorem 4.1.18. *If V is an inner product space over $\mathbb{K}$, then the function defined in Definition 4.1.9 is a norm. We call $\|\cdot\|$ the* inner product norm.

Proof. To see that $\|\cdot\|$ is a norm, we need to verify Definition 4.1.17.

To see that (1) holds, note that $\|\vec{v}\| = 0$ if and only if $\|\vec{v}\|^2 = 0$ if and only if $\langle \vec{v}, \vec{v} \rangle = 0$ if and only if $\vec{v} = \vec{0}$ by Theorem 4.1.8.

To see that (2) holds, note that for any $\alpha \in \mathbb{K}$ and $\vec{v} \in V$, we have

$$\begin{aligned}
\|\alpha\vec{v}\|^2 &= \langle \alpha\vec{v}, \alpha\vec{v} \rangle \\
&= \alpha \langle \vec{v}, \alpha\vec{v} \rangle \\
&= \alpha\bar{\alpha} \langle \vec{v}, \vec{v} \rangle \\
&= |\alpha|^2 \|\vec{v}\|^2
\end{aligned}$$

The result then holds by taking square root on both sides.

Finally, to see that (3) holds, note that for any $\vec{v}, \vec{w} \in V$, we have by Cauchy-Schwartz Inequality

$$\begin{aligned}
\|\vec{v} + \vec{w}\|^2 &= \|\vec{v}\|^2 + 2\text{Re}(\langle \vec{v}, \vec{w} \rangle) + \|\vec{w}\|^2 \\
&\leq \|\vec{v}\|^2 + 2|\langle \vec{v}, \vec{w} \rangle| + \|\vec{w}\|^2 \\
&\leq \|\vec{v}\|^2 + 2\|\vec{v}\|\|\vec{w}\| + \|\vec{w}\|^2 \\
&= (\|\vec{v}\| + \|\vec{w}\|)^2
\end{aligned}$$

The result then holds by taking square roots on both sides. □

To conclude this section, we present an application of inner products and norms. In particular, recall that in $\mathbb{R}^n$, we computed angles between two vectors $\vec{v}, \vec{w}$ as

$$\cos(\theta) = \frac{\vec{v} \cdot \vec{w}}{\|\vec{v}\|\|\vec{w}\|}$$

If V is an inner product space over $\mathbb{R}$, then the angle of a nonzero vector $\vec{v} \in V$ to a nonzero vector in $\vec{w} \in V$ is defined to be the unique angle $\theta \in [0, \pi)$ such that

$$\cos(\theta) = \frac{\langle \vec{v}, \vec{w} \rangle}{\|\vec{v}\| \|\vec{w}\|}$$

Example 4.1.19. Let $V = \mathcal{C}[0, 1]$ be equipped with the inner product

$$\langle f, g \rangle = \int_0^1 f(x)g(x)\ dx$$

for all $f, g \in V$. If $f(x) = x$ and $g(x) = x^2$, then we have

$$\begin{aligned} \left\langle x, x^2 \right\rangle &= \int_0^1 x^3\ dx = \frac{1}{4} \\ \|x\| &= \sqrt{\int_0^1 x^2\ dx} = \frac{1}{\sqrt{3}} \\ \|x^2\| &= \sqrt{\int_0^1 x^4\ dx} = \frac{1}{\sqrt{5}} \end{aligned}$$

Therefore,

$$\theta = \cos^{-1}\left(\frac{\left\langle x, x^2 \right\rangle}{\|x\|\|x^2\|}\right) = \cos^{-1}\left(\frac{\frac{1}{4}}{\frac{1}{\sqrt{3}}\frac{1}{\sqrt{5}}}\right) \approx 0.25 \text{ rad}$$

4.2 Orthogonal Projections and Complements

In this section, we introduce one of the most important constructions in inner product spaces: the orthogonal projection. Orthogonal projections allow us to decompose vectors into components that lie in a subspace and its orthogonal complement. This decomposition is central in applications such as least squares approximation, Fourier analysis, and numerical linear algebra.

Moreover, in the next section, we will use orthogonal projections to construct orthonormal bases via the *Gram-Schmidt orthogonalization process.*

Recall from calculus that two vectors $\vec{v}, \vec{w} \in \mathbb{R}^n$ are said to be *orthogonal* if $\vec{v} \cdot \vec{w} = 0$. The same idea can be transferred to a general vector space.

Definition 4.2.1. Let V be an inner product space over $\mathbb{K}$. Two vectors $\vec{v}, \vec{w} \in V$ are said to be *orthogonal* if $\langle \vec{v}, \vec{w} \rangle = 0$.

Example 4.2.2. Let $V = \mathcal{C}[-1,1]$ equipped with the inner product

$$\langle f, g\rangle = \int_{-1}^{1} f(x)g(x)\ dx$$

for $f, g \in \mathcal{C}[-1,1]$. It can be easily checked that $f(x) = x$ and $g(x) = x^2$ are orthogonal. Indeed,

$$\langle f, g\rangle = \left\langle x, x^2\right\rangle = \int_{-1}^{1} x \cdot x^2\ dx = \int_{-1}^{1} x^3\ dx = 0$$

Theorem 4.2.3 (Pythagorean Theorem). *Let V be an inner product space over $\mathbb{K}$. Assume that $\vec{v}, \vec{w} \in V$ are orthogonal. Then*

$$\|\vec{v} + \vec{w}\|^2 = \|\vec{v}\|^2 + \|\vec{w}\|^2$$

Proof. Note that

$$\|\vec{v} + \vec{w}\|^2 = \|\vec{v}\|^2 + \langle \vec{v}, \vec{w}\rangle + \langle \vec{w}, \vec{v}\rangle + \|\vec{w}\|^2$$

But since $\vec{v}, \vec{w} \in V$ are orthogonal, then $\langle \vec{v}, \vec{w}\rangle = \langle \vec{w}, \vec{v}\rangle = 0$, so then

$$\|\vec{v} + \vec{w}\|^2 = \|\vec{v}\|^2 + \|\vec{w}\|^2$$

as claimed. □

The concept of orthogonality can be transferred from two vectors to a set of vectors.

Definition 4.2.4. Let V be an inner product space over $\mathbb{K}$. A subset $S \subseteq V$ is said to be an *orthogonal set* if $\vec{v}, \vec{w} \in S$ are distinct, then $\vec{v}$ and $\vec{w}$ are orthogonal. A vector $\vec{v} \in V$ is said to be a *unit vector* if $\|\vec{v}\| = 1$. A subset $S \subseteq V$ is said to be an *orthonormal set* if S is an orthogonal set and every element of S is a unit vector.

Example 4.2.5. The standard basis of $\mathbb{R}^n$ given by $\{\vec{e}_1, \vec{e}_2, ..., \vec{e}_n\}$, i.e. $\vec{e}_i$ is 1 if $i = j$, and 0 otherwise, is an orthonormal set. Indeed, note that each vector in the basis has norm one, and it is easy to check that for every $i \neq j$, $\langle \vec{e}_i, \vec{e}_j\rangle = 0$.

Remark 4.2.6. If S_0 is an orthogonal set that does not contain the zero vector, then it is very easy to turn it into an orthonormal set S, by dividing each element of S_0 by its length. That is,

$$S = \left\{\frac{\vec{v}}{\|\vec{v}\|} : \vec{v} \in S_0\right\}$$

Example 4.2.7. Consider the set

$$S = \left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 4 \\ -5 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ -3 \end{bmatrix} \right\}$$

We first show that S is an orthogonal set in $\mathbb{R}^3$ with respect to the standard inner product. Indeed, denoting $\vec{v}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$, $\vec{v}_2 = \begin{bmatrix} 4 \\ -5 \\ 1 \end{bmatrix}$, and $\vec{v}_3 = \begin{bmatrix} 2 \\ 1 \\ -3 \end{bmatrix}$, we have

$$\begin{aligned} \langle \vec{v}_1, \vec{v}_2 \rangle &= 4 - 5 + 1 = 0 \\ \langle \vec{v}_1, \vec{v}_3 \rangle &= 2 + 1 - 3 = 0 \\ \langle \vec{v}_2, \vec{v}_3 \rangle &= 8 - 5 - 3 = 0 \end{aligned}$$

Therefore, the set S is orthogonal. However, it is not orthonormal since $\|\vec{v}_1\| = \sqrt{3}$.

If we want to turn S into an orthonormal set, we simply divide by the norm, so $\|\vec{v}_1\| = \sqrt{3}$, $\|\vec{v}_2\| = \sqrt{42}$ and $\|\vec{v}_3\| = \sqrt{14}$. Therefore, the set

$$T = \left\{ \frac{1}{\sqrt{3}}\vec{v}_1, \frac{1}{\sqrt{42}}\vec{v}_2, \frac{1}{\sqrt{14}}\vec{v}_3 \right\}$$

is orthonormal.

The reason why we have discussions about orthogonal or orthonormal sets, is because that orthogonal or orthonormal sets are linearly independent as the theorem below states this fact.

Theorem 4.2.8. *Let V be an inner product space over $\mathbb{K}$ and let S be an orthonormal set. Then S is linearly independent.*

Proof. Assume that for $\alpha_1, \alpha_2, ..., \alpha_n \in \mathbb{K}$ are such that

$$\alpha_1\vec{v}_1 + \alpha_2\vec{v}_2 + \cdots + \alpha_n\vec{v}_n = \vec{0}$$

Note that for any $j = 1, 2, ..., n$

$$\left\langle \vec{0}, \vec{v}_j \right\rangle = \langle \alpha_1\vec{v}_1 + \alpha_2\vec{v}_2 + \cdots + \alpha_n\vec{v}_n, \vec{v}_j \rangle$$

So then

$$0 = \sum_{k=1}^{n} \alpha_k \langle \vec{v}_k, \vec{v}_j \rangle = \alpha_j \langle \vec{v}_j, \vec{v}_j \rangle$$

Therefore, $\alpha_j \|\vec{v}_j\|^2 = \alpha_j = 0$. Since this holds for any j, we are done. □

The other reason why we care about orthogonal or orthonormal sets is because it is very easy to find the scalars to write a vector $\vec{v}$ as a linear combination of elements of a basis $\mathcal{B}$ which is an orthogonal or orthonormal set.

Example 4.2.9. If $\mathcal{B} = \{\vec{e}_1, \vec{e}_2\}$ is the standard basis of $\mathbb{R}^2$ which is an orthonormal set, then if $\vec{v} = \begin{bmatrix} -1 \\ 2 \end{bmatrix}$, then

$$\vec{v} = (-1)\vec{e}_1 + 2\vec{e}_2$$

and note that $\alpha_1 = -1 = \langle \vec{v}, \vec{e}_1 \rangle$ and $\alpha_2 = 2 = \langle \vec{v}, \vec{e}_2 \rangle$.

Theorem 4.2.10. *Let V be an inner product space over $\mathbb{K}$, and let $S = \{\vec{x}_1, \vec{x}_2, ..., \vec{x}_n\}$ be an orthonormal set. If $\vec{v} \in \operatorname{span}(S)$, then*

$$\vec{v} = \langle \vec{v}, \vec{x}_1 \rangle \vec{x}_1 + \langle \vec{v}, \vec{x}_2 \rangle \vec{x}_2 + \cdots + \langle \vec{v}, \vec{x}_n \rangle \vec{x}_n$$

Proof. Since $\vec{v} \in \operatorname{span}(S)$ there exists $\alpha_1, \alpha_2, ..., \alpha_n \in \mathbb{K}$ such that

$$\vec{v} = \sum_{k=1}^{n} \alpha_k \vec{x}_k$$

We claim that $\alpha_j = \langle \vec{v}, \vec{x}_j \rangle$. For any $j \in \{1, 2, ..., n\}$, we have

$$\langle \vec{v}, \vec{x}_k \rangle = \left\langle \sum_{k=1}^{N} \alpha_k \vec{x}_k, \vec{x}_j \right\rangle = \sum_{k=1}^{n} \alpha_k \langle \vec{x}_k, \vec{x}_j \rangle = \alpha_j \langle \vec{x}_j, \vec{x}_j \rangle = \alpha_j$$

Therefore,

$$\alpha_j = \langle \vec{v}, \vec{x}_j \rangle$$

for any $j = 1, 2, ..., n$. $\square$

Definition 4.2.11. Let V be a finite-dimensional inner product space over $\mathbb{K}$. A set $\mathcal{B}$ is said to be an *orthonormal basis* of V if $\mathcal{B}$ is a basis of V and $\mathcal{B}$ is an orthonormal set.

Example 4.2.12. The standard basis of $\mathbb{K}^n$ is an orthonormal basis for $\mathbb{K}^n$ with respect to the standard inner product.

Example 4.2.13. The standard basis of $\mathcal{M}_{mn}(\mathbb{K})$ is an orthonormal basis of $\mathcal{M}_{mn}(\mathbb{K})$ with respect to the inner product $\langle A, B \rangle = \operatorname{tr}(AB^*)$.

According to Theorem 4.2.10, if $\mathcal{B} = \{\vec{x}_1, \vec{x}_2, ..., \vec{x}_n\}$ is an orthonormal basis of V, then for any $\vec{v} \in V$, we have

$$[\vec{v}]_{\mathcal{B}} = \begin{bmatrix} \langle \vec{v}, \vec{x}_1 \rangle \\ \langle \vec{v}, \vec{x}_2 \rangle \\ \vdots \\ \langle \vec{v}, \vec{x}_n \rangle \end{bmatrix}$$

Example 4.2.14. Let $\mathcal{B} = \left\{ \frac{1}{\sqrt{5}} \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \frac{1}{\sqrt{5}} \begin{bmatrix} -1 \\ 2 \end{bmatrix} \right\}$. Denoting $\vec{v}_1 = \frac{1}{\sqrt{5}} \begin{bmatrix} 2 \\ 1 \end{bmatrix}$ and $\vec{v}_2 = \frac{1}{\sqrt{5}} \begin{bmatrix} -1 \\ 2 \end{bmatrix}$, note that

$$\langle \vec{v}_1, \vec{v}_2 \rangle = \frac{2}{\sqrt{5}} \cdot \left(-\frac{1}{\sqrt{5}} \right) + \frac{1}{\sqrt{5}} \cdot \frac{2}{\sqrt{5}} = 0$$

Since $\vec{v}_1, \vec{v}_2$ are orthogonal and $\|\vec{v}_1\| = \|\vec{v}_2\| = 1$, then $\mathcal{B}$ is an orthonormal set. By Theorem 4.2.8, $\mathcal{B}$ is linearly independent and since $\mathcal{B}$ has two elements, it is an orthonormal basis of $\mathbb{R}^2$.

Now if $\vec{v} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$, we have

$$\langle \vec{v}, \vec{v}_1 \rangle = 2 \cdot \frac{2}{\sqrt{5}} + 3 \cdot \frac{1}{\sqrt{5}} = \frac{7}{\sqrt{5}}$$
$$\langle \vec{v}, \vec{v}_2 \rangle = 2 \cdot \left(-\frac{1}{\sqrt{5}} \right) + 3 \cdot \frac{2}{\sqrt{5}} = \frac{4}{\sqrt{5}}$$

Therefore,

$$[\vec{v}]_{\mathcal{B}} = \begin{bmatrix} \frac{7}{\sqrt{5}} \\ \frac{4}{\sqrt{5}} \end{bmatrix}$$

So far, we have seen why orthonormal bases are important. Now, we may ask if every vector space has an orthonormal basis? We will answer this question gradually. First, we need to see how we can find vectors orthogonal to a given set of vectors.

Definition 4.2.15. Let V be an inner product space over $\mathbb{K}$ and let $S \subseteq V$. The *orthogonal complement* of S, denoted $S^\perp$ is the set

$$S^\perp = \{\vec{v} \in V : \langle \vec{v}, \vec{w} \rangle = 0 \text{ for all } \vec{w} \in S\}$$

In other words, the orthogonal complement of S is the set of vectors orthogonal to every vector in S.

Example 4.2.16. Let $V = \mathbb{R}^3$ be equipped with the standard inner product and let $S = \left\{ \vec{x}_1 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \vec{x}_2 = \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix} \right\}$. Then note that $\vec{v} = \begin{bmatrix} a \\ b \\ c \end{bmatrix} \in S^\perp$ if and only if $\langle \vec{v}, \vec{x}_1 \rangle = \langle \vec{v}, \vec{x}_2 \rangle = 0$, so

$$\begin{cases} a + 2b + 3c = 0 \\ 3a + 2b + c = 0 \end{cases}$$

Solving this system gives $b = -2c$ and $a = c$, so then

$$S^\perp = \left\{ \begin{bmatrix} a \\ b \\ c \end{bmatrix} \in \mathbb{R}^3 : b = -2c \text{ and } a = c \right\} = \operatorname{span} \left\{ \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix} \right\}$$

As it turns out based on Example 4.2.16, $S^\perp$ is a subspace of V.

Theorem 4.2.17. *Let V be an inner product space over $\mathbb{K}$ and let $S \subseteq V$. Then $S^\perp$ is a subspace of V.*

Remark 4.2.18. Note that S is not necessarily a subspace. What the theorem mentions is that no matter what S is, $S^\perp$ is a subspace of V.

Proof. To see that $S^\perp$ is a subspace of V, we need to verify the following properties.

Note that $\left\langle \vec{0}_V, \vec{v} \right\rangle = 0$ for all $\vec{v} \in V$, so $\vec{0}_V \in S^\perp$.

Let $\vec{v}_1, \vec{v}_2 \in S^\perp$, so for all $\vec{v} \in S$, we have $\langle \vec{v}_1, \vec{v} \rangle = \langle \vec{v}_2, \vec{v} \rangle = 0$. Then

$$\langle \vec{v}_1 + \vec{v}_2, \vec{v} \rangle = \langle \vec{v}_1, \vec{v} \rangle + \langle \vec{v}_2, \vec{v} \rangle = 0$$

which implies that $\vec{v}_1 + \vec{v}_2 \in S^\perp$.

Finally, let $\vec{w} \in S^\perp$, so $\langle \vec{w}, \vec{v} \rangle = 0$ for all $\vec{v} \in S$. Let $\alpha \in \mathbb{K}$. Then

$$\langle \alpha \vec{w}, \vec{v} \rangle = \alpha \langle \vec{w}, \vec{v} \rangle = \alpha \cdot 0 = 0$$

which implies that $\alpha \vec{w} \in S^\perp$. $\square$

The following theorem tells us a slightly faster way to compute $W^\perp$ when W is a subspace of V.

Theorem 4.2.19. *Let V be an inner product space over $\mathbb{K}$, let W be a subspace of V, and let $\mathcal{B}$ be a basis of W. Then $W^\perp = \mathcal{B}^\perp$.*

Remark 4.2.20. Note that this theorem tells you that if you want to find $W^\perp$ when W is a subspace, then you need to find a basis of W, and then find $\mathcal{B}^\perp$.

Proof. To see that $W^\perp = \mathcal{B}^\perp$, we need to show that $W^\perp \subseteq \mathcal{B}^\perp$ and $\mathcal{B}^\perp \subseteq W^\perp$.

For the former, let $\mathcal{B} = \{\vec{b}_1, \vec{b}_2, ..., \vec{b}_n\}$ be a basis of W, and let $\vec{w} \in W^\perp$, so $\left\langle \vec{w}, \vec{0} \right\rangle = 0$ for all $\vec{v} \in W$. Then $\left\langle \vec{w}, \vec{b}_k \right\rangle = 0$ for all $k = 1, 2, ..., n$ since $\mathcal{B} \subseteq W$ and this implies that $\vec{w} \in \mathcal{B}^\perp$.

For the latter, let $\vec{v} \in \mathcal{B}^\perp$, so $\left\langle \vec{v}, \vec{b}_k \right\rangle = 0$ for all $k = 1, 2, ..., n$. Then let $\vec{w} \in W$. Since $W = \text{span}(\mathcal{B})$, then

$$\vec{w} = \alpha_1 \vec{b}_1 + \alpha_2 \vec{b}_2 + \cdots + \alpha_n \vec{b}_n$$

and so

$$\langle \vec{v}, \vec{w} \rangle = \left\langle \vec{v}, \sum_{k=1}^n \alpha_k \vec{b}_k \right\rangle = \sum_{k=1}^n \bar{\alpha}_k \left\langle \vec{v}, \vec{b}_k \right\rangle = 0$$

Therefore, $\vec{v} \in W^\perp$. $\square$

We have previously studied orthogonal projections in previous courses. The reason we need orthogonal projections is to answer questions like this: Given a point $P(a, b, c)$ and the plane, find the shortest distance from the point P to the plane. We would like to do the same in general abstract vector spaces.

Definition 4.2.21. Let V be an inner product space over $\mathbb{K}$, let W be a subspace of V, and let $\mathcal{B} = \{\vec{x}_1, \vec{x}_2, ..., \vec{x}_n\}$ be an orthonormal basis for W. The *orthogonal projection* of V onto W is the map $P_W : V \to V$ defined by

$$P_W(\vec{v}) = \sum_{k=1}^n \langle \vec{v}, \vec{x}_k \rangle \vec{x}_k$$

for all $\vec{v} \in V$.

Example 4.2.22. Let $V = \mathbb{R}^3$ and $W = \text{span}(\mathcal{B})$, where $\mathcal{B} = \left\{ \frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \frac{1}{\sqrt{2}} \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \right\}$.

Note that $\mathcal{B}$ is an orthonormal for W. If $\vec{v} = \begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix}$, and denoting $\vec{x}_1 =$

$\frac{1}{\sqrt{3}}\begin{bmatrix}1\\1\\1\end{bmatrix}$ and $\vec{x}_2 = \frac{1}{\sqrt{2}}\begin{bmatrix}-1\\0\\1\end{bmatrix}$, then one can check that

$$\langle \vec{v}, \vec{x}_1 \rangle = \frac{2}{\sqrt{3}}, \quad \langle \vec{v}, \vec{x}_2 \rangle = -\frac{1}{\sqrt{2}}$$

Then

$$P_W(\vec{v}) = \frac{2}{\sqrt{3}}\vec{x}_1 - \frac{1}{\sqrt{2}}\vec{x}_2$$

The following theorem provides the geometric interpretation of $P_W(\vec{v})$ familiar from our calculus courses.

Theorem 4.2.23. *Let V be an inner product space over $\mathbb{K}$, let W be a subspace of V, and let $\mathcal{B} = \{\vec{x}_1, \vec{x}_2, ..., \vec{x}_n\}$ be an orthonormal basis of W. Then*

1. *P_W is a linear map.*
2. *For all $\vec{v} \in V$, $P_W(\vec{v}) \in W$ and $\vec{v} - P_W(\vec{v}) \in W^{\perp}$.*
3. *$P_W(\vec{v})$ is the unique vector in W that is closest to $\vec{v}$.*

Proof. Exercise. □

Example 4.2.24. Let $V = \mathcal{C}[-1, 1]$ with inner product given by

$$\langle f, g \rangle = \int_{-1}^{1} f(x)g(x)\ dx$$

Also, consider $\mathcal{P}_W(\mathbb{R}) = \text{span}(\mathcal{B})$, where $\mathcal{B} = \{1, x, 3x^2 - 1\}$. Convince yourself that $\mathcal{B}$ is an orthogonal basis for $\mathcal{P}_2(\mathbb{R})$, but you do not need to show this. Given $\vec{v} = f(x) = |x|$, then we want to find the vector $\vec{w}$ in $\mathcal{P}_2(\mathbb{R})$ that is closest to $\vec{v}$.

Here, write $\vec{x}_1 = 1$, $\vec{x}_2 = x$, and $\vec{x}_3 = 3x^2 - 1$. Note that $\mathcal{B}$ in the problem is not an orthonormal basis. So we instead compute instead

$$\vec{w} = P_W(\vec{v}) = \frac{\langle \vec{v}, \vec{x}_1 \rangle}{\|\vec{x}_1\|^2} + \frac{\langle \vec{v}, \vec{x}_2 \rangle}{\|\vec{x}_2\|^2} + \frac{\langle \vec{v}, \vec{x}_3 \rangle}{\|\vec{x}_3\|^2}$$

Where in this case, one can check that

$$\langle \vec{v}, \vec{x}_1 \rangle = 1, \|\vec{x}_1\|^2 = 2, \langle \vec{v}, \vec{x}_2 \rangle = 0, \|\vec{x}_2\|^2 = \frac{2}{3}, \langle \vec{v}, \vec{x}_3 \rangle = \frac{1}{2}, \|\vec{x}_3\|^2 = \frac{8}{5}$$

Therefore,

$$\vec{w} = \frac{1}{2}\vec{x}_1 + \frac{5}{16}\vec{x}_3 = \frac{1}{2} + \frac{5}{16}(3x^2 - 1)$$

In the above example, we were given an orthogonal basis of $\mathcal{P}_2(\mathbb{R})$. But how did we find such a basis? Could we construct the basis $\mathcal{B}$ given in Example 4.2.24 from our familiar basis $\{1, x, x^2\}$ of $\mathcal{P}_2(\mathbb{R})$?

The answer is yes, and we need the Gram-Schmidt Orthogonalization Process.

4.3 The Gram-Schmidt Orthogonalization Process

The inner product structure of a vector space allows us to define orthogonality and construct orthonormal sets — sets of vectors that are both orthogonal and of unit length. Orthonormal sets are particularly desirable because they greatly simplify computations involving projections, norms, and linear transformations.

In this section, we study the *Gram-Schmidt orthogonalization process*, a method for transforming any linearly independent set into an orthonormal set that spans the same subspace. This procedure is essential in constructing orthonormal bases and plays a critical role in applications such as QR-factorization and numerical algorithms.

Example 4.3.1. Consider two vectors $\vec{v}_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ and $\vec{v}_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ in $\mathbb{R}^2$. Note that $\mathcal{B} = \{\vec{v}_1, \vec{v}_2\}$ is a basis of $\mathbb{R}^2$, but it is not orthogonal. But can we construct an orthogonal basis $\{\vec{f}_1, \vec{f}_2\}$ using $\mathcal{B} = \{\vec{v}_1, \vec{v}_2\}$?

The answer is yes, and we define these vectors as follows: Set $\vec{f}_1 = \vec{v}_1$, and $\vec{f}_2 = \vec{v}_2 - \text{proj}_{\vec{v}_1}(\vec{v}_2)$, where

$$\vec{f}_2 = \vec{v}_2 - \frac{\langle \vec{v}_2, \vec{v}_1 \rangle}{\|\vec{v}_1\|^2} = \begin{bmatrix} \frac{2}{5} \\ -\frac{1}{5} \end{bmatrix}$$

Thus, $\{\vec{f}_1, \vec{f}_2\}$ is an orthogonal basis.

Lemma 4.3.2. *Let V be an inner product space over $\mathbb{K}$, let $\{\vec{f}_1, \vec{f}_2, ..., \vec{f}_n\}$ be an orthogonal set of vectors in V, and let $\vec{v} \notin \text{span}\{\vec{f}_1, \vec{f}_2, ..., \vec{f}_n\}$. Define*

$$\vec{f}_{n+1} = \vec{v} - \frac{\left\langle \vec{v}, \vec{f}_1 \right\rangle}{\|\vec{f}_1\|^2} - \frac{\left\langle \vec{v}, \vec{f}_2 \right\rangle}{\|\vec{f}_2\|^2} - \cdots - \frac{\left\langle \vec{v}, \vec{f}_n \right\rangle}{\|\vec{f}_n\|^2}$$

Then $\{\vec{f}_1, \vec{f}_2, ..., \vec{f}_{n+1}\}$ is an orthogonal set of vectors.

Proof. For convenience, write $t_k = \frac{\langle \vec{x}, \vec{f}_k \rangle}{\|\vec{f}_k\|^2}$ for each k. For any $j = 1, 2, ..., n$,

$$\left\langle \vec{f}_{n+1}, \vec{f}_j \right\rangle = \left\langle \vec{v} - \sum_{k=1}^{n} t_j \vec{f}_j, \vec{f}_j \right\rangle = \left\langle \vec{v}, \vec{f}_j \right\rangle - t_j \|\vec{f}_j\|^2 = 0$$

This holds since $\vec{f}_{n+1} \neq \vec{0}_V$ and $\vec{v} \notin \operatorname{span}\{\vec{f}_1, \vec{f}_2, ..., \vec{f}_n\}$. □

Theorem 4.3.3 (**Gram-Schmidt Orthogonalization Process**). *Let V be an inner product space and let $\mathcal{B} = \{\vec{v}_1, \vec{v}_2, ..., \vec{v}_n\}$ be any basis of V. Define $\vec{f}_1, \vec{f}_2, ..., \vec{f}_n$ in V recursively successively as follows:*

$$\begin{aligned}
\vec{f}_1 &= \vec{v}_1 \\
\vec{f}_2 &= \vec{v}_2 - \frac{\left\langle \vec{v}_2, \vec{f}_1 \right\rangle}{\|\vec{f}_1\|^2} \vec{f}_1 \\
\vec{f}_3 &= \vec{v}_3 - \frac{\left\langle \vec{v}_3, \vec{f}_1 \right\rangle}{\|\vec{f}_1\|^2} \vec{f}_1 - \frac{\left\langle \vec{v}_3, \vec{f}_2 \right\rangle}{\|\vec{f}_2\|^2} \vec{f}_2 \\
&\vdots \\
\vec{f}_n &= \vec{v}_n - \sum_{k=1}^{n-1} \frac{\left\langle \vec{v}_n, \vec{f}_k \right\rangle}{\|\vec{f}_k\|^2} \vec{f}_k
\end{aligned}$$

Then $\{\vec{f}_1, \vec{f}_2, ..., \vec{f}_n\}$ is an orthogonal basis of V.

Remark 4.3.4. If you want to have an orthonormal basis, then divide each $\vec{f}_k$ by its length $\|\vec{f}_k\|$. In other words, $\left\{ \frac{\vec{f}_1}{\|\vec{f}_1\|}, \frac{\vec{f}_2}{\|\vec{f}_2\|}, ..., \frac{\vec{f}_n}{\|\vec{f}_n\|} \right\}$ would be an *orthonormal basis* of V.

Example 4.3.5. Let $V = \mathbb{R}^4$ be an inner product space over $\mathbb{R}$ with the standard inner product, and let

$$\vec{v}_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \vec{v}_2 = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \vec{v}_3 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

and let $W = \operatorname{span}\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$. Then using the Gram-Schmidt Orthogonalization Process, one can check that

$$\vec{f_1} = \vec{v}_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}$$

$$\vec{f_2} = \vec{v}_2 - \frac{\left\langle \vec{v}_2, \vec{f_1} \right\rangle}{\|\vec{f_1}\|^2}\vec{f_1} = \begin{bmatrix} \frac{1}{2} \\ -\frac{1}{2} \\ 1 \\ 0 \end{bmatrix}$$

$$\vec{f_3} = \vec{v}_3 - \frac{\left\langle \vec{v}_3, \vec{f_1} \right\rangle}{\|\vec{f_1}\|^2}\vec{f_1} - \frac{\left\langle \vec{v}_3, \vec{f_2} \right\rangle}{\|\vec{f_2}\|^2}\vec{f_2} = \begin{bmatrix} \frac{1}{3} \\ -\frac{1}{3} \\ -\frac{1}{3} \\ 1 \end{bmatrix}$$

Then $\{\vec{f_1}, \vec{f_2}, \vec{f_3}\}$ is an orthogonal basis of W. If we want it to be an orthonormal basis of W instead, then divide each vector by its norm, where

$$\|\vec{f_1}\| = \sqrt{2}, \|\vec{f_2}\| = \sqrt{\frac{3}{2}}, \|\vec{f_3}\| = \frac{2}{\sqrt{3}}$$

Then

$$\left\{ \frac{1}{\sqrt{2}}\vec{f_1}, \sqrt{\frac{2}{3}}\vec{f_2}, \frac{\sqrt{3}}{2}\vec{f_3} \right\}$$

is an orthonormal basis of W.

Example 4.3.6. Let $V = \mathcal{P}_2(\mathbb{R})$ over $\mathbb{R}$ be equipped with the inner product

$$\langle f, g \rangle = \int_{-1}^{1} f(x)g(x)\ dx$$

We apply the Gram-Schmidt Orthogonalization Process to $\{1, x, x^2\}$ to find an orthogonal basis, as follows:

Here, $\vec{v}_1 = 1$, $\vec{v}_2 = x$, and $\vec{v}_3 = x^2$, and we want to find $\vec{f_1}, \vec{f_2}, \vec{f_3}$. Indeed, set $\vec{f_1} = 1$.

For $\vec{f_2}$, we have

$$\vec{f_2} = \vec{v}_2 - \frac{\left\langle \vec{v}_2, \vec{f_1} \right\rangle}{\|\vec{f_1}\|^2}\vec{f_1}$$

where in this case,

$$\left\langle \vec{v}_2, \vec{f}_1 \right\rangle = \int_{-1}^{1} x \cdot 1 \; dx = 0$$

and

$$\|\vec{f}_1\|^2 = \left\langle \vec{f}_1, \vec{f}_1 \right\rangle = \int_{-1}^{1} \; dx = 2$$

Therefore,

$$\vec{f}_2 = x - \frac{0}{2} \cdot 1 = x$$

For $\vec{f}_3$, we have

$$\vec{f}_3 = \vec{v}_3 - \frac{\left\langle \vec{v}_3, \vec{f}_1 \right\rangle}{\|\vec{f}_1\|^2}\vec{f}_1 - \frac{\left\langle \vec{v}_3, \vec{f}_2 \right\rangle}{\|\vec{f}_2\|^2}\vec{f}_2$$

where in this case,

$$\begin{aligned} \left\langle \vec{v}_3, \vec{f}_1 \right\rangle &= \int_{-1}^{1} x^2 \; dx = \frac{2}{3} \\ \left\langle \vec{v}_3, \vec{f}_2 \right\rangle &= \int_{-1}^{1} x^3 \; dx = 0 \\ \|\vec{f}_2\|^2 &= \int_{-1}^{1} x^2 \; dx = \frac{2}{3} \end{aligned}$$

Therefore,

$$\vec{f}_3 = x^2 - \frac{1}{3}$$

So $\{\vec{f}_1, \vec{f}_2, \vec{f}_3\}$ is an orthogonal basis of $\{1, x, x^2\}$.

For an orthonormal basis, divide each vector by its length, to obtain that

$$\left\{ \frac{\vec{f}_1}{\|\vec{f}_1\|}, \frac{\vec{f}_2}{\|\vec{f}_2\|}, \frac{\vec{f}_3}{\|\vec{f}_3\|} \right\} = \left\{ \frac{1}{\sqrt{2}}, \sqrt{\frac{3}{2}}x, \sqrt{\frac{5}{8}}(3x^2 - 1) \right\}$$

Such a basis is called the *Legendre polynomials.*

4.4 QR-Factorization

In the previous section, we studied the Gram-Schmidt orthogonalization process, which transforms a linearly independent set of vectors into an orthonormal set that spans the same subspace. This construction allows us to write a matrix as a product of an orthonormal matrix and an upper triangular matrix — a decomposition known as the *QR-factorization*.

QR-factorization plays a central role in numerical linear algebra, especially for solving systems of linear equations, computing eigenvalues, and performing least squares approximations. Before we see the QR-factorization of matrices, let us review some concepts about matrices.

Definition 4.4.1. Let $R = [a_{ij}]_{i,j=1}^{n} \in \mathcal{M}_n(\mathbb{K})$. It is said that R is *upper triangular* if $a_{ij} = 0$ for all $i > j$.

For example, the matrix

$$A = \begin{bmatrix} 2 & 1 & 5 \\ 0 & 1 & 0 \\ 0 & 0 & 4 \end{bmatrix}$$

is an upper triangular matrix.

Definition 4.4.2. Let $A \in \mathcal{M}_{mn}(\mathbb{K})$ be a matrix such that $\operatorname{rank}(A) = n$. Write

$$A = \begin{bmatrix} \uparrow & \uparrow & & \uparrow \\ \vec{v}_1 & \vec{v}_2 & \cdots & \vec{v}_n \\ \downarrow & \downarrow & & \downarrow \end{bmatrix}$$

Apply the Gram-Schmidt Orthogonalization process to $\{\vec{v}_1, \vec{v}_2, ..., \vec{v}_n\}$ with respect to the standard inner product on $\mathbb{K}^n$ to get

$$\vec{f}_1 = \vec{v}_1, \quad \vec{f}_k = \vec{v}_k - \sum_{i=1}^{k-1} \frac{\left\langle \vec{v}_k, \vec{f}_i \right\rangle\rangle}{\|\vec{f}_i\|^2} \vec{f}_i, \quad \text{and} \quad \vec{u}_k = \frac{\vec{f}_k}{\|\vec{f}_k\|}$$

Then the matrix A can be decomposed into two matrices Q and R, where

$$A = QR = \begin{bmatrix} \uparrow & \uparrow & & \uparrow \\ \vec{u}_1 & \vec{u}_2 & \cdots & \vec{u}_n \\ \downarrow & \downarrow & & \downarrow \end{bmatrix} \begin{bmatrix} \|\vec{f}_1\| & \langle \vec{v}_2, \vec{u}_1 \rangle & \langle \vec{v}_3, \vec{u}_1 \rangle & \langle \vec{v}_4, \vec{u}_1 \rangle & \cdots \\ 0 & \|\vec{f}_2\| & \langle \vec{v}_3, \vec{u}_2 \rangle & \langle \vec{v}_4, \vec{u}_2 \rangle & \cdots \\ 0 & 0 & \|\vec{f}_3\| & \langle \vec{v}_4, \vec{u}_3 \rangle & \cdots \\ \vdots & \vdots & \ddots & \ddots & \ddots \end{bmatrix}$$

If A can be decomposed into Q and R, then A is the *QR-decomposition* of matrices Q and R.

Remark 4.4.3. If Q is the left matrix in the matrix multiplication, and R is the right matrix in the matrix multiplication, then $Q \in \mathcal{M}_{mn}(\mathbb{K})$ and $R \in \mathcal{M}_{nm}(\mathbb{K})$. R is an upper triangular matrix with each entry along the diagonal nonzero, and the columns Q form an orthonormal set.

Example 4.4.4. Consider the matrix $A \in \mathcal{M}_{43}$ given by

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

Here, $n = 3$, and the column vectors of A are given by

$$\vec{v}_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \vec{v}_2 = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \vec{v}_3 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

By the Gram-Schmidt Orthogonalization Process (or see Example 4.3.5), we have

$$\vec{u}_1 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \\ 0 \end{bmatrix}, \vec{u}_2 = \begin{bmatrix} \frac{1}{\sqrt{6}} \\ -\frac{1}{\sqrt{6}} \\ \frac{2}{\sqrt{6}} \\ 0 \end{bmatrix}, \vec{u}_3 = \begin{bmatrix} \frac{1}{2\sqrt{3}} \\ -\frac{1}{2\sqrt{3}} \\ -\frac{1}{2\sqrt{3}} \\ \frac{3}{2\sqrt{3}} \end{bmatrix}$$

The Q matrix is then given as

$$Q = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & \frac{1}{2\sqrt{3}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} & -\frac{1}{2\sqrt{3}} \\ 0 & \frac{2}{\sqrt{6}} & -\frac{1}{2\sqrt{3}} \\ 0 & 0 & \frac{3}{2\sqrt{3}} \end{bmatrix}$$

and the R matrix is given as

$$R = \begin{bmatrix} \sqrt{2} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ 0 & \frac{\sqrt{6}}{2} & \frac{1}{\sqrt{6}} \\ 0 & 0 & \frac{2\sqrt{3}}{3} \end{bmatrix}$$

Therefore, the matrix A can be decomposed into matrices Q and R, i.e.

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & \frac{1}{2\sqrt{3}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} & -\frac{1}{2\sqrt{3}} \\ 0 & \frac{2}{\sqrt{6}} & -\frac{1}{2\sqrt{3}} \\ 0 & 0 & \frac{3}{2\sqrt{3}} \end{bmatrix} \begin{bmatrix} \sqrt{2} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ 0 & \frac{\sqrt{6}}{2} & \frac{1}{\sqrt{6}} \\ 0 & 0 & \frac{2\sqrt{3}}{3} \end{bmatrix}$$

4.5 Adjoints and Unitaries of Linear Maps

In inner product spaces, we can define a natural analogue of the transpose of a matrix — called the *adjoint* of a linear operator. The adjoint provides a way to "move" a linear transformation from one side of an inner product to the other. This concept is fundamental in the study of self-adjoint, unitary, and normal operators, all of which generalize key properties of symmetric and orthogonal matrices.

Definition 4.5.1. Let V and W be inner product spaces over $\mathbb{K}$ and $T : V \to W$ be a linear map. The *adjoint* of T is a linear map $T^* : W \to V$ with the following property that

$$\langle T(\vec{x}), \vec{y} \rangle_W = \langle \vec{x}, T^*(\vec{y}) \rangle_V$$

where $\langle \cdot, \cdot \rangle_V$ and $\langle \cdot, \cdot \rangle_W$ are inner products with respect to V and W, respectively.

Example 4.5.2. Let $V = \{f \in \mathcal{P}_2(\mathbb{R}) : f(0) = f(1) = 0\}$ be the vector space equipped with the inner product

$$\langle f, g \rangle = \int_0^1 f(x)g(x)\ dx$$

and let $D : V \to V$ be the derivative map defined by

$$D(f) = f'$$

Let $f, g \in V$ be arbitrary. Observe that

$$\begin{aligned}
\langle T(f), g \rangle &= \left\langle f', g \right\rangle \\
&= \int_0^1 f'(x)g(x)\ dx \\
&= [f(x)g(x)]_0^1 - \int_0^1 f(x)g'(x)\ dx \quad \text{integration by parts} \\
&= f(1)g(1) - f(0)g(0) + \int_0^1 f(x)(-g'(x))\ dx \\
&= \int_0^1 f(x)(-T(g))\ dx \\
&= \langle f, -T(g) \rangle
\end{aligned}$$

Therefore, $T^* = -T$, that is, $T^*(g) = -T(g) = -g'$.

Remark 4.5.3. You would realize that it takes some effort to find the adjoint of T. We would like to have a straightforward way to find T^*. This can be achieved via connecting the matrices that correspond to T and T^*.

Theorem 4.5.4. *Let V and W be finite-dimensional inner product spaces over $\mathbb{K}$, and let $T : V \to W$ be a linear map. If $\mathcal{A}$ is an orthonormal basis of V and $\mathcal{B}$ is an orthonormal basis of W, then $[T^*]_{\mathcal{A}}^{\mathcal{B}}$ is the conjugate transpose of $[T]_{\mathcal{B}}^{\mathcal{A}}$.*

Proof. Exercise. □

Example 4.5.5. Let $V = \mathbb{C}^2$ be a vector space over $\mathbb{C}$ equipped with the standard inner product, nad having the for a basis a standard basis $\mathcal{A} = \{\vec{e}_1, \vec{e}_2\}$. Let $T : V \to V$ be a linear map defined by

$$T(a_1, a_2) = (2ia_1 + 2a_2, a_1 - a_2)$$

Then note that

$$T(\vec{e}_1) = T(1, 0) = (2i, 1), T(\vec{e}_2) = T(0, 1) = (2, -1)$$

and so,

$$[T(\vec{e}_1)]_{\mathcal{A}} = \begin{bmatrix} 2i \\ 1 \end{bmatrix}, [T(\vec{e}_2)]_{\mathcal{A}} = \begin{bmatrix} 2 \\ -1 \end{bmatrix}$$

and thus,

$$[T]_{\mathcal{A}}^{\mathcal{A}} = \begin{bmatrix} 2i & 2 \\ 1 & -1 \end{bmatrix}$$

By Theorem 4.5.4, we have

$$[T^*]_{\mathcal{A}}^{\mathcal{A}} = ([T]_{\mathcal{A}}^{\mathcal{A}})^* = \begin{bmatrix} -2i & 1 \\ 2 & -1 \end{bmatrix}$$

Furthermore, for any $(a_1, a_2) \in \mathbb{C}^2$, recall that

$$[T^*(a_1, a_2)]_{\mathcal{A}} = [T^*]_{\mathcal{A}}^{\mathcal{A}}[(a_1, a_2)]_{\mathcal{A}} = \begin{bmatrix} -2i & 1 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} -2ia_1 + a_2 \\ 2a_1 - a_2 \end{bmatrix}$$

Therefore,

$$T^*(a_1, a_2) = (-2ia_1 + a_2)\vec{e}_1 + (2a_1 - a_2)\vec{e}_2 = \begin{bmatrix} -2ia_1 + a_2 \\ 2a_1 - a_2 \end{bmatrix}$$

Remark 4.5.6. We can generalize Example 4.5.5 for a linear map $T_A : \mathbb{K}^n \to \mathbb{K}^n$ defined by $T_A(\vec{x}) = A\vec{x}$, where $A \in \mathcal{M}_{mn}(\mathbb{K})$. If $V = \mathbb{K}^n$ and $W = \mathbb{K}^n$ equipped with the standard bases $\mathcal{SB}$, then $[T_A]_{\mathcal{SB}}^{\mathcal{SB}} = A$, by Theorem 4.5.4,

$$[T_A^*]_{\mathcal{SB}}^{\mathcal{SB}} = ([T_A]_{\mathcal{SB}}^{\mathcal{SB}})^* = A^*$$

and as a result,

$$T_A^*(\vec{x}) = A^*\vec{x} = T_{A^*}(\vec{x})$$

With the help of the adjoint of a linear map T, we can introduce special cases of linear maps.

Definition 4.5.7. Let V be an inner product space over $\mathbb{K}$.

1. A linear map $T : V \to V$ such that $T^* = T$ is said to be *self-adjoint* in the case when $\mathbb{K} = \mathbb{C}$, and is said to be *symmetric* in the case when $\mathbb{K} = \mathbb{R}$.

2. A linear map $T : V \to V$ such that $T^*T = TT^* = \mathrm{id}_V$ is said to be *unitary* in the case when $\mathbb{K} = \mathbb{C}$, and is said to be *orthogonal* in the case $\mathbb{K} = \mathbb{R}$.

The definition of self-adjoint (or symmetric) and unitary (or orthogonal) linear maps is very similar to that of an $n \times n$ matrix.

Definition 4.5.8. Let $A \in \mathcal{M}_n(\mathbb{K})$.

1. A is said to be *self-adjoint* (or *symmetric*) if $A^* = A$.

2. A is said to be *unitary* (or *orthogonal*) if $A^*A = AA^* = I_n$.

Remark 4.5.9. It is easier to check whether T is self-adjoint or unitary via their corresponding matrices. More precisely, if V is an inner product space over $\mathbb{K}$ with an orthonormal basis $\mathcal{A}$. Then

1. T is self-adjoint if and only if $T^* = T$ if and only if $[T^*]_{\mathcal{A}}^{\mathcal{A}} = [T]_{\mathcal{A}}^{\mathcal{A}}$ if and only if (by Theorem 4.5.4) $([T]_{\mathcal{A}}^{\mathcal{A}})^* = [T]_{\mathcal{A}}^{\mathcal{A}}$ if and only if $[T]_{\mathcal{A}}^{\mathcal{A}}$ is self-adjoint.

2. T is unitary if and only if $T^*T = TT^* = \mathrm{id}_V$ if and only if $[T^*]_{\mathcal{A}}^{\mathcal{A}}[T]_{\mathcal{A}}^{\mathcal{A}} = [T]_{\mathcal{A}}^{\mathcal{A}}[T^*]_{\mathcal{A}}^{\mathcal{A}} = [\mathrm{id}_V]_{\mathcal{A}}^{\mathcal{A}}$ if and only if $([T]_{\mathcal{A}}^{\mathcal{A}})^*[T]_{\mathcal{A}}^{\mathcal{A}} = [T]_{\mathcal{A}}^{\mathcal{A}}([T]_{\mathcal{A}}^{\mathcal{A}})^* = [\mathrm{id}_V]_{\mathcal{A}}^{\mathcal{A}}$ if and only if $[T]_{\mathcal{A}}^{\mathcal{A}}$ is unitary.

Example 4.5.10. Let $\mathbb{K} = \mathbb{R}$ and consider the map $T_A : \mathbb{R}^2 \to \mathbb{R}^2$ given by $T_A(\vec{x}) = A\vec{x}$, where

$$A = \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix}$$

(rotation matrix by angle θ). Note that

$$A^* = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \neq A$$

and so T_A cannot be self-adjoint. On the other hand, note that

$$A^*A = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I_2$$

and similarly, one can show that $AA^* = I_2$. Therefore, A is unitary, and thus, T_A is unitary.

4.6 Isometries

In the study of inner product spaces, it is often useful to consider linear maps that preserve geometric structure — in particular, lengths and angles. Such transformations are called *isometries*.

An isometry is a linear transformation that preserves the norm (and hence the inner product) between vectors. These maps are the algebraic analogue of rigid motions in geometry: they do not distort shapes, but simply rotate or reflect them. Isometries play a central role in linear algebra, geometry, and functional analysis, and are especially important in the classification of orthogonal and unitary operators.

In this section, we formally define isometries, examine their properties, and explore examples in both real and complex vector spaces. We also connect the concept of isometries to unitary and orthogonal matrices, and show that isometries are precisely those linear maps whose adjoints are inverses.

Definition 4.6.1. Let V be an inner product space over $\mathbb{K}$ and let $T : V \to V$ be a linear map. Then T is called an *isometry* if for all $\vec{v}, \vec{w} \in V$,

$$\|T(\vec{v}) - T(\vec{w})\| = \|\vec{v} - \vec{w}\|$$

The following theorem tells us that an isometry is not only distance preserving, but also length preserving and inner product preserving.

Theorem 4.6.2. *Let V be a finite-dimensional inner product space with $\dim(V) = n$, and let $T : V \to V$ be a linear map. The following are equivalent.*

1. *T is an isometry.*
2. *$\|T(\vec{v})\| = \|\vec{v}\|$ for all $\vec{v} \in V$.*
3. *$\langle T(\vec{v}), T(\vec{w})\rangle = \langle \vec{v}, \vec{w}\rangle$ for all $\vec{v}, \vec{w} \in V$.*
4. *If $\{\vec{f_1}, \vec{f_2}, ..., \vec{f_n}\}$ is an orthonormal basis of V, then $\{T(\vec{f_1}), T(\vec{f_2}), ..., T(\vec{f_n})\}$ is an orthonormal basis of V.*

Proof. Assume that (1) holds. To see that (2) holds, note that for all $\vec{v} \in V$,

$$\|T(\vec{v}) - T(\vec{0}_V)\| = \|T(\vec{v})\| = \|\vec{v}\|$$

Next, assume that (2) holds. To see that (3) holds, note that for all $\vec{v}, \vec{w} \in V$, we have $\|T(\vec{v} - \vec{w})\| = \|\vec{v} - \vec{w}\|$, and so because T is linear, then $\|T(\vec{v}) - T(\vec{w})\| = \|\vec{v} - \vec{w}\|$. Therefore,

$$\begin{aligned}\|T(\vec{v}) - T(\vec{w})\|^2 &= \|\vec{v} - \vec{w}\|^2 \\ &= \langle T(\vec{v}) - T(\vec{w}), T(\vec{v}) - T(\vec{w})\rangle \\ &= \langle \vec{v} - \vec{w}, \vec{v} - \vec{w}\rangle\end{aligned}$$

Now observe that

$$\begin{aligned}\|T(\vec{v})\|^2 - 2\langle T(\vec{v}), T(\vec{w})\rangle + \|T(\vec{w})\|^2 &= \|\vec{v}\|^2 - 2\langle \vec{v}, \vec{w}\rangle + \|\vec{w}\|^2 \\ -2\langle T(\vec{v}), T(\vec{w})\rangle &= -2\langle \vec{v}, \vec{w}\rangle \\ \langle T(\vec{v}), T(\vec{w})\rangle &= \langle \vec{v}, \vec{w}\rangle\end{aligned}$$

proving (3).

Next, assume that (3) holds. To see that (4) holds, note that for any $1 \leq i, j \leq n$,

$$\left\langle T(\vec{f_i}), T(\vec{f_j})\right\rangle = \left\langle \vec{f_i}, \vec{f_j}\right\rangle = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j\end{cases}$$

Therefore, $\{T(\vec{f_1}), T(\vec{f_2}), ..., T(\vec{f_n})\}$ is orthonormal, since $\dim(V) = n$.

Finally, assume that (4) holds. To see that (1) holds, let

$$\begin{aligned}\vec{v} &= \alpha_1\vec{f_1} + \alpha_2\vec{f_2} + \cdots + \alpha_n\vec{f_n} \\ \vec{w} &= \beta_1\vec{f_1} + \beta_2\vec{f_2} + \cdots + \beta_n\vec{f_n}\end{aligned}$$

Then by the Pythagorean Theorem (Theorem 4.2.3),

$$\begin{aligned}\|T(\vec{v}) - T(\vec{w})\|^2 &= \|T(\vec{v} - \vec{w})\|^2 \\ &= \left\|\sum_{i=1}^{n}(\alpha_i - \beta_i)T(\vec{f}_i)\right\|^2 \\ &= \sum_{i=1}^{n}|\alpha_i - \beta_i|^2\end{aligned}$$

and also,

$$\|\vec{v} - \vec{w}\|^2 = \left\|\sum_{i=1}^{n}(\alpha_i - \beta_i)\vec{f}_i\right\|^2 = \sum_{i=1}^{n}|\alpha_i - \beta_i|^2$$

as desired. □

An isometry T can be characterized via the adjoint of T.

Theorem 4.6.3. *Let V be an inner product space over $\mathbb{K}$ and let $T : V \to V$ be a linear map. Then T is an isometry if and only if $T^*T = \mathrm{id}_V$.*

Proof. Assume that T is an isometry. Then we have for all $\vec{v}, \vec{w} \in V$,

$$\begin{aligned}\langle T(\vec{v}), T(\vec{w})\rangle &= \langle \vec{v}, \vec{w}\rangle \\ \langle \vec{v}, T^*T(\vec{w})\rangle &= \langle \vec{v}, \vec{w}\rangle = \langle \vec{v}, \mathrm{id}_V(\vec{w})\rangle \\ \langle \vec{v}, T^*T(\vec{w}) - \mathrm{id}_V(\vec{w})\rangle &= 0\end{aligned}$$

Now apply it for $\vec{v} = T^*T(\vec{w}) - \mathrm{id}_V(\vec{w})$ to get

$$\langle T^*T(\vec{w}) - \mathrm{id}_V(\vec{v}), T^*T(\vec{w}) - \mathrm{id}_V(\vec{w})\rangle = 0$$

and so,

$$\|T^*T - \mathrm{id}_V(\vec{w})\|^2 = 0$$

and so by properties of norm, $T^*T - \mathrm{id}_V(\vec{w}) = \vec{0}_V$, and since $T^*T(\vec{w}) = \mathrm{id}_V(\vec{w})$ for all $\vec{w} \in V$, then $T^*T = \mathrm{id}_V$.

Conversely, if $T^*T = \mathrm{id}_V$, to see that T is an isometry, note that for all $\vec{v}, \vec{w} \in V$,

$$\langle T(\vec{v}), T(\vec{w})\rangle = \langle \vec{v}, T^*T(\vec{w})\rangle = \langle \vec{v}, \mathrm{id}_V(\vec{w})\rangle = \langle \vec{v}, \vec{w}\rangle$$

Therefore, T is an isometry, as desired. □

Following the same steps, it can be shown that T is an isometry if and only if $TT^* = \mathrm{id}_V$, i.e. T is a unitary linear map. If $\mathcal{B}$ is an orthonormal basis of V, then T is an isometry if and only if $[T]_{\mathcal{B}}^{\mathcal{B}}$ is a unitary matrix.

Theorems 4.6.2 and 4.6.3 can be stated in terms of matrices. We just state the theorem for $T = T_A$.

Theorem 4.6.4. *Let $A \in \mathcal{M}_n(\mathbb{K})$. The following are equivalent.*

1. $\|A\vec{v} - A\vec{w}\| = \|\vec{v} - \vec{w}\|$ *for all* $\vec{v}, \vec{w} \in \mathbb{K}^n$.
2. $\|A\vec{v}\| = \|\vec{v}\|$ *for all* $\vec{v} \in \mathbb{K}^n$.
3. $\langle A\vec{v}, A\vec{w}\rangle = \langle \vec{v}, \vec{w}\rangle$ *for all* $\vec{v}, \vec{w} \in \mathbb{K}^n$.
4. *The columns of A are an orthonormal set in $\mathbb{K}^n$.*
5. $A^*A = AA^* = I_n$.

Proof. The proof follows from the same steps as in Theorem 4.6.2 and 4.6.3. Exercise. □

The above theorem is even true for non-square matrices $A \in \mathcal{M}_{mn}(\mathbb{K})$ with only difference is that in (5), we have only $A^*A = I_n$. The proof is the same as in the case of square matrices. In QR factorization of A, the columns of Q form an orthonormal set. By Theorem 4.6.4, the matrix Q has the property $Q^*Q = I_n$.

4.7 Method of Least Squares

In many practical situations, we are faced with systems of linear equations that have no exact solution. This typically occurs when there are more equations than unknowns — such systems are called *overdetermined*. In such cases, we seek an approximate solution that "best fits" the data. The standard approach is known as the *method of least squares*.

The core idea is to find the vector $\vec{x} \in \mathbb{R}^n$ that minimizes the distance between $A\vec{x}$ and the given vector $\vec{b} \in \mathbb{R}^m$, where $A \in \mathcal{M}_{mn}(\mathbb{R})$ and $m > n$. Geometrically, this means projecting $\vec{b}$ orthogonally onto the column space of A, yielding the closest possible approximation within the image of the transformation A.

In this section, we develop the least squares method using inner product space theory. We derive the *normal equations*, explain the geometric intuition behind the solution, and demonstrate how the QR-factorization can be used to efficiently compute least squares approximations. This method is essential in statistics, numerical analysis, machine learning, and many areas of applied mathematics.

Imagine you want to predict future prices of houses in your local area based on some features that houses have. For this purpose, you need to have data. Say you collected a sample of n houses in your local area, and you got the following data: For each $1 \leq i \leq n$, let b_i be the price of the house, a_{i1} be the number of bedrooms, and let a_{i2} be the square feet of the house. The goal is to find scalars x_1 and x_2 such that for each $1 \leq i \leq n$

$$b_i = a_{i1}x_1 + a_{i2}x_2$$

The expression is called a *linear model*. Equivalently, we can write the above as

$$\begin{aligned} b_1 &= a_{11}x_1 + a_{12}x_2 \\ b_2 &= a_{21}x_1 + a_{22}x_2 \\ &\vdots \\ b_n &= a_{n1}x_1 + a_{n2}x_2 \end{aligned}$$

Or in matrix notation,

$$\vec{b} = A\vec{x}$$

The above system can be inconsistent. Then, we can only ask for $\vec{x}$ such that $A\vec{x}$ is as close as possible to $\vec{b}$, i.e. to make

$$\|A\vec{x} - \vec{b}\|$$

as small as possible. This is the *least squares problem.*

Note that $\vec{b} - A\vec{x} \in \operatorname{ran}(A)^\perp$, so then $\vec{b} - A\vec{x} \in \ker(A^*)$, that is, $A^*(\vec{b} - A\vec{x}) = \vec{0}$, and so $A^*\vec{b} - A^*A\vec{x} = \vec{0}$, and thus, $A^*A\vec{x} = A^*\vec{b}$. We need to solve for $\vec{x}$.

Theorem 4.7.1. *Assuming that A^*A is an invertible matrix, then the solution to the least squares problem is given by*

$$\vec{x} = (A^*A)^{-1}A^*\vec{b}$$

Example 4.7.2. Consider a system of linear equations given by

$$\begin{aligned} 3x_1 - x_2 &= 4 \\ x_1 + 2x_2 &= 0 \\ 2x_1 + x_2 &= 1 \end{aligned}$$

we have

$$A = \begin{bmatrix} 3 & -1 \\ 1 & 2 \\ 2 & 1 \end{bmatrix}, \vec{b} = \begin{bmatrix} 4 \\ 0 \\ 1 \end{bmatrix}$$

Also, we have

$$A^* = \begin{bmatrix} 3 & 1 & 2 \\ -1 & 2 & 1 \end{bmatrix}, A^*A = \begin{bmatrix} 14 & 1 \\ 1 & 6 \end{bmatrix}$$

and thus,

$$(A^*A)^{-1} = \frac{1}{83} \begin{bmatrix} 6 & -1 \\ -1 & 14 \end{bmatrix}$$

Finally, the solution to the least squares problem is given by

$$\begin{aligned} \vec{x} &= (A^*A)^{-1}A^*\vec{b} \\ &= \frac{1}{83} \begin{bmatrix} 6 & -1 \\ -1 & 14 \end{bmatrix} \begin{bmatrix} 3 & 1 & 2 \\ -1 & 2 & 1 \end{bmatrix} \begin{bmatrix} 4 \\ 0 \\ 1 \end{bmatrix} \\ &= \frac{1}{83} \begin{bmatrix} 87 \\ -56 \end{bmatrix} \end{aligned}$$

Note that $\vec{x}$ is not a solution to the system.

Now, say that the model is described as follows: Let P be the price defined by x_1 number of bedrooms and x_2 square feet, that is,

$$P(x_1, x_2) = x_1 \times \text{number of bedrooms} + x_2 \times \text{square feet}$$

Then say if $x_1 = 5$ and $x_2 = 100$,

$$P(5, 100) = \frac{87}{56} \cdot 5 - \frac{56}{83} \cdot 100 \approx 8250$$

Chapter 5

Diagonalization

A central theme in linear algebra is the study of linear transformations and the structure they impose on vector spaces. One of the most powerful tools for analyzing linear transformations is *diagonalization*. The idea behind diagonalization is to represent a linear transformation with respect to a suitable basis so that its matrix becomes diagonal — a form that is simple to work with, both algebraically and computationally.

Diagonal matrices are especially desirable because their action on vectors is straightforward: applying a diagonal matrix simply scales each coordinate. When a transformation can be diagonalized, many problems — such as computing powers of a matrix, solving differential equations, or analyzing stability — become vastly simpler.

This chapter begins with the foundational concepts of *eigenvalues* and *eigenvectors*, which are the key to identifying diagonalizable transformations. We then study the necessary and sufficient conditions for diagonalizability and explore the relationship between eigenstructure and invariant subspaces.

The chapter also introduces the *Spectral Theorem*, one of the cornerstones of linear algebra, which guarantees that certain operators — specifically, self-adjoint and normal operators — are diagonalizable with respect to an orthonormal basis. We then examine *positive definite matrices*, which play an important role in optimization and numerical analysis.

Finally, we conclude the chapter with the *Singular Value Decomposition* (SVD), a far-reaching generalization of diagonalization that applies to all matrices, including non-square ones. The SVD is not only theoretically elegant but also a foundational tool in modern applications such as machine learning, image compression, and principal component analysis.

5.1 Eigenvalues and Eigenvectors

In previous courses, you may have encountered eigenvalues and eigenvectors in the context of matrices — finding scalars λ and nonzero vectors $\vec{v}$ such that

$$A\vec{v} = \lambda\vec{v}$$

for some square matrix A. This section builds upon that idea, but now in the context of abstract linear transformations between vector spaces.

Given a linear map $T : V \to V$, we are interested in finding vectors $\vec{v} \in V$ (with $\vec{v} \neq \vec{0}$) and scalars $\lambda \in \mathbb{F}$ such that

$$T(\vec{v}) = \lambda\vec{v}$$

We call such a scalar λ an *eigenvalue* of T, and the corresponding vector $\vec{v}$ an *eigenvector*.

Although eigenvectors and eigenvalues are defined abstractly in terms of a linear transformation T, we often compute them by working with a matrix representation of T with respect to a chosen basis. This section focuses on defining eigenvalues and eigenvectors formally, exploring their properties, and connecting the abstract theory to practical computations using the matrix of T.

The study of eigenstructure is fundamental to understanding the behavior of linear operators, and it lays the foundation for diagonalization, invariant subspaces, the spectral theorem, and many applications throughout mathematics and the applied sciences.

In this chapter, we will allow for our field to be $\mathbb{F} = \mathbb{C}, \mathbb{R}$, or $\mathbb{Q}$.

Definition 5.1.1. Let $A \in \mathcal{M}_n(\mathbb{F})$. A scalar $\lambda \in \mathbb{F}$ is said to be an *eigenvalue* of A if there exists a nonzero vector $\vec{v} \in \mathbb{F}^n$ such that $A\vec{v} = \lambda\vec{v}$. The nonzero vector $\vec{v}$ is called an *eigenvector* that corresponds to the eigenvalue λ.

The above definition can be restated in terms of $T = T_A$. That is, $A\vec{v} = \lambda\vec{v}$ if and only if $T_A(\vec{v}) = \lambda\vec{v}$ and this leads to the following definition.

Definition 5.1.2. Let V be a vector space over $\mathbb{F}$ and let $T : V \to V$ be a linear transformation. A scalar $\lambda \in \mathbb{F}$ is said to be an *eigenvalue* of T if there exists a nonzero vector $\vec{v}$ such that $T(\vec{v}) = \lambda\vec{v}$. The nonzero vector $\vec{v}$ is called an *eigenvector* that corresponds to the eigenvalue λ.

Example 5.1.3. Let $V = \mathcal{C}^2(\mathbb{R})$, i.e. V is the set of all twice differentiable functions defined on $\mathbb{R}$. It is elementary to verify that $\mathcal{C}^2(\mathbb{R})$ is a subspace of $\mathcal{C}(\mathbb{R})$. Let $D : V \to V$ be the derivative map defined by $D(f) = f'$.

If $f(x) = e^{\lambda x}$, where λ is some constant, we claim that $f(x)$ is an eigenvector of D that corresponds to the eigenvalue λ. Indeed, note that for any $x \in \mathbb{R}$

$$D(f)(x) = D(e^{\lambda x}) = (e^{\lambda x})' = \lambda e^{\lambda x} = \lambda f(x)$$

Since this holds for any $x \in \mathbb{R}$, then $D(f) = \lambda f$, as required.

Example 5.1.4. Let $V = \mathcal{C}^2(\mathbb{R})$. let $D : V \to V$ be the linear map defined by $D(f) = f''$. If $f(x) = \sin(\lambda x)$ and $g(x) = \cos(\lambda x)$, where λ is some constant, we claim that $f(x)$ and $g(x)$ are eigenvectors that correspond to the eigenvalue $-\lambda^2$. To see this, note that for any $x \in \mathbb{R}$, we have

$$\begin{aligned} D(f)(x) &= D(\sin(\lambda x)) = \sin''(\lambda x) = -\lambda^2 \sin(\lambda x) = -\lambda^2 f(x) \\ D(g)(x) &= D(\cos(\lambda x)) = \cos''(\lambda x) = -\lambda^2 \cos(\lambda x) = -\lambda^2 g(x) \end{aligned}$$

Since this holds for any $x \in \mathbb{R}$, then $D(f) = -\lambda^2 f$ and $D(g) = -\lambda^2 g$, as required.

We saw that two different eigenvectors correspond to the same eigenvalue $-\lambda^2$. We can easily check that any linear combination of $f(x)$ and $g(x)$ is an eigenvector that corresponds to the same eigenvalue $-\lambda^2$. This is in fact, a general result.

Theorem 5.1.5. *Let V be a vector space over $\mathbb{F}$, and $T : V \to V$ be a linear transformation. If $\vec{v}_1$ and $\vec{v}_2$ are eigenvectors that correspond to the same eigenvalue λ, then*

$$\vec{w} = \alpha_1 \vec{v}_1 + \alpha_2 \vec{v}_2$$

where $\alpha_1, \alpha_2 \in \mathbb{F}$, is an eigenvector that corresponds to the eigenvalue λ, as well.

Proof. We have $T(\vec{v}_1) = \lambda \vec{v}_1$ and $T(\vec{v}_2) = \lambda \vec{v}_2$, then note that for $\alpha_1, \alpha_2 \in \mathbb{F}$, if $\vec{w} = \alpha_1 \vec{v}_1 + \alpha_2 \vec{v}_2$,

$$\begin{aligned} T(\vec{w}) &= T(\alpha_1 \vec{v}_1 + \alpha_2 \vec{v}_2) \\ &= \alpha_1 T(\vec{v}_1) + \alpha_2 T(\vec{v}_2) \\ &= \alpha_1 \lambda \vec{v}_1 + \alpha_2 \lambda \vec{v}_2 \\ &= \lambda(\alpha_1 \vec{v}_1 + \alpha_2 \vec{v}_2) \\ &= \lambda \vec{w} \end{aligned}$$

Therefore, $\vec{w}$ is an eigenvector that corresponds to the eigenvalue λ, as claimed. $\square$

In Example 5.1.3, we saw that for the eigenvalue $\lambda = 1$, the corresponding eigenvector was e^x, and for the eigenvalue $\lambda = 2$, the corresponding eigenvector was e^{2x}. We can actually verify that these two eigenvectors e^x and e^{2x} are linearly independent. This is in fact a general result, which says in simple words: the eigenvectors that correspond to distinct eigenvalues are linearly independent.

Theorem 5.1.6. *Let V be a vector space over $\mathbb{F}$, let $T : V \to V$ be a linear transformation, and let $\lambda_1, \lambda_2, ..., \lambda_n \in \mathbb{F}$ be distinct eigenvalues of T. For each $1 \leq k \leq n$, if $\vec{v}_k$ is an eigenvector for T with eigenvalue λ_k, then $\vec{v}_1, \vec{v}_2, ..., \vec{v}_n$ is linearly independent.*

Proof. Assume for a contradiction that $\vec{v}_1, \vec{v}_2, ..., \vec{v}_n$ are linearly dependent. Then there exists $\alpha_1, \alpha_2, ..., \alpha_n \in \mathbb{F}$ not all zero such that

$$\alpha_1\vec{v}_1 + \alpha_2\vec{v}_2 + \cdots + \alpha_n\vec{v}_n = \vec{0}_V$$

Without loss of generality, say that $\alpha_n \neq 0$. Then applying T to both sides yields

$$T(\alpha_1\vec{v}_1 + \alpha_2\vec{v}_2 + \cdots + \alpha_n\vec{v}_n) = \vec{0}_V$$

and thus, by linearity,

$$\alpha_1T(\vec{v}_1) + \alpha_2T(\vec{v}_2) + \cdots + \alpha_nT(\vec{v}_n) = \vec{0}_V$$

But then for each $1 \leq k \leq n$, since $\vec{v}_k$ is an eigenvector of T corresponding to λ_k, then

$$\alpha_1\lambda_1\vec{v}_1 + \alpha_2\lambda_2\vec{v}_2 + \cdots + \alpha_n\lambda_n\vec{v}_n = \vec{0}_V$$

Now, multiply both sides of the above equation by λ_n, we have

$$\lambda_n(\alpha_1\lambda_1\vec{v}_1 + \alpha_2\lambda_2\vec{v}_2 + \cdots + \alpha_n\lambda_n\vec{v}_n) = \vec{0}_V$$

Then subtracting both equations, we have

$$\alpha_1(\lambda_1 - \lambda_n)\vec{v}_1 + \alpha_2(\lambda_2 - \lambda_n)\vec{v}_2 + \cdots + \alpha_{n-1}(\lambda_{n-1} - \lambda_n)\vec{v}_{n-1} = \vec{0}_V$$

Notice that $\vec{v}_n$ disappears so we have a linear dependence that is based on $\vec{v}_1, \vec{v}_2, ..., \vec{v}_{n-1}$. By repeating this argument, we can continue to eliminate all of the terms, ultimately concluding that $\alpha_k = 0$ for all $1 \leq k \leq n$, contradicting the assumption that not all α_k are zero. $\square$

Example 5.1.7. Let $\lambda_1, \lambda_2, ..., \lambda_n \in \mathbb{R}$ be distinct real numbers. We claim that the functions $\{e^{\lambda_1 x}, e^{\lambda_2 x}, ..., e^{\lambda_n x}\}$ are linearly independent.

To see this, note that by Example 5.1.3, $e^{\lambda_k x}$ is the eigenvector corresponding to the eigenvalue λ_k for each $1 \leq k \leq n$, and since $\lambda_1, \lambda_2, ..., \lambda_n$ are distinct, by Theorem 5.1.6, $\{e^{\lambda_1 x}, e^{\lambda_2 x}, ..., e^{\lambda_n x}\}$ are linearly independent.

The following theorem generalizes the previous theorem.

Corollary 5.1.8. *Let V be a vector space over $\mathbb{F}$, let $T : V \to V$ be a linear map, and let $\lambda_1, \lambda_2, ..., \lambda_n \in \mathbb{F}$ be distinct eigenvalues of T. For each $1 \leq k \leq n$, if S_k is a set of linearly independent eigenvectors of T with eigenvalue λ_k, then $\bigcup_{k=1}^n S_k$ is linearly independent.*

Proof. Exercise. □

Example 5.1.9. Let $\lambda_1, \lambda_2 \in (0, \infty)$. We claim that the functions

$$\{\cos(\lambda_1 x), \sin(\lambda_1 x), \cos(\lambda_2 x)\}$$

are linearly independent. To see this, note that $\cos(\lambda_1 x)$ and $\sin(\lambda_1 x)$ are eigenvectors corresponding to the eigenvalue $-\lambda_1^2$ of linear map $D(f) = f''$, and $\cos(\lambda_2 x)$ is the eigenvector corresponding to the eigenvalue $-\lambda_2^2$ of $D(f) = f''$.

Then define $S_1 = \{\cos(\lambda_1 x), \sin(\lambda_1 x)\}$ and $S_2 = \{\cos(\lambda_2 x)\}$. It is easy to note that S_2 is linearly independent. To see that S_1 is linearly independent, let $\alpha_1, \alpha_2 \in \mathbb{R}$ be such that for all $x \in \mathbb{R}$,

$$\alpha_1 \cos(r_1 x) + \alpha_2 \sin(\lambda_1 x) = 0$$

Note when $x = 0$,

$$\alpha_1 \cdot 1 + \alpha_2 \cdot 0 = 0$$

and so $\alpha_1 = 0$, and similarly, when $x = \frac{\pi}{2}$,

$$\alpha_1 \cdot 0 + \alpha_2 \cdot 1 = 0$$

and so $\alpha_2 = 0$. Therefore, S_1 is linearly independent.

Since S_1 and S_2 are linearly independent, then by Theorem 1.4.4 and Corollary 5.1.8, $S_1 \cup S_2$ is linearly independent. The same problem can be extended to the set of functions

$$\{\cos(\lambda_1 x), \sin(\lambda_1 x), \cos(\lambda_2 x), \sin(\lambda_2 x), ..., \cos(\lambda_n x), \sin(\lambda_n x)\}$$

Let V be a vector space over $\mathbb{F}$, let $T : V \to V$ be a linear map, and let $\lambda \in \mathbb{F}$. Then note that λ is an eigenvalue of T if and only if for some $\vec{v} \in V$, $T(\vec{v}) = \lambda\vec{v}$. However, note then that we can rearrange this equation to obtain

$$T(\vec{v}) - \lambda\vec{v} = \vec{0}_V$$

But then observe $\mathrm{id}_V(\vec{v}) = \vec{v}$, so

$$T(\vec{v}) - \lambda \, \mathrm{id}_V(\vec{v}) = \vec{0}_V$$

and thus,

$$(T - \lambda \, \mathrm{id}_V)(\vec{v}) = \vec{0}_V$$

Thus, $\vec{v}$ is an eigenvector if and only if $\vec{v} \in \ker(T - \lambda \, \mathrm{id}_V)$. This leads us to the following definition.

Definition 5.1.10. Let V be a vector space over $\mathbb{F}$, let $T : V \to V$ be a linear map, and let $\lambda \in \mathbb{F}$. If λ is an eigenvalue of T, then the λ-*eigenspace* of T, denoted by $\mathcal{E}_T(\lambda)$ is defined to be

$$\mathcal{E}_T(\lambda) = \ker(T - \lambda \, \mathrm{id}_V) = \{\text{all } \lambda\text{-eigenvectors of } T \text{ and } \vec{0}_V\}$$

Remark 5.1.11. $\mathcal{E}_T(\lambda)$ is a subspace of V. As a result, all nonzero linear combinations of eigenvectors with eigenvalue λ is an eigenvector with the same eigenvalue λ, as we have shown in Theorem 5.1.6.

In previous linear algebra courses, we learned how to compute eigenvalues and eigenvectors of matrices. This skill is very useful, since we will be computing eigenvalues and eigenvectors of linear maps via their corresponding matrices.

Let V be a finite-dimensional vector space over $\mathbb{F}$, say $\dim(V) = n$, let $T : V \to V$ be a linear map, let $\lambda \in \mathbb{F}$, and let $\mathcal{B}$ be a basis of V. Notice that if $T(\vec{v}) = \lambda\vec{v}$, then

$$[T]_{\mathcal{B}}^{\mathcal{B}}[\vec{v}]_{\mathcal{B}} = \lambda[\vec{v}]_{\mathcal{B}}$$

and so,

$$[T]_{\mathcal{B}}^{\mathcal{B}}[\vec{v}]_{\mathcal{B}} - \lambda[\vec{v}]_{\mathcal{B}} = \vec{0}$$

and thus,

$$([T]_{\mathcal{B}}^{\mathcal{B}} - \lambda I_n)[\vec{v}_B] = \vec{0}$$

which shows that $[\vec{v}]_{\mathcal{B}} \in \ker([T]_{\mathcal{B}}^{\mathcal{B}} - \lambda I_n)$. This would also mean that the matrix $[T]_{\mathcal{B}}^{\mathcal{B}} - \lambda I_n$ is not invertible, and thus,

$$\det([T]_{\mathcal{B}}^{\mathcal{B}} - \lambda I_n) = 0$$

This means that we can find eigenvalues and eigenvalues of T by using $[T]_{\mathcal{B}}^{\mathcal{B}}$. In this case, $\det([T]_{\mathcal{B}}^{\mathcal{B}} - \lambda I_n)$ is the characteristic polynomial of $[T]_{\mathcal{B}}^{\mathcal{B}}$.

Definition 5.1.12. Let $A \in \mathcal{M}_n(\mathbb{F})$. The *characteristic polynomial* of A, denoted by χ_A is the polynomial

$$\chi_A(\lambda) = \det(A - \lambda I_n)$$

The λ-*eigenspace* of a matrix $A \in \mathcal{M}_n(\mathbb{F})$, denoted by $\mathcal{E}_A(\lambda)$ is defined to be

$$\mathcal{E}_A(\lambda) = \ker(A - \lambda I_n) = \{\text{all } \lambda\text{-eigenvectors of } A \text{ and the vector } \vec{0}\}$$

In order to find the eigenvalues of the matrix, we need to solve the equation $\chi_A(\lambda) = 0$.

Example 5.1.13. Let $A = \begin{bmatrix} 1 & -1 & 1 \\ -1 & 1 & -1 \\ 1 & -1 & 1 \end{bmatrix}$. To find the eigenvalues of A, we need to solve the equation $\chi_A(\lambda) = 0$. First note that

$$\begin{aligned}
\det(A - \lambda I_3) &= \begin{vmatrix} 1-\lambda & -1 & 1 \\ -1 & 1-\lambda & -1 \\ 1 & -1 & 1-\lambda \end{vmatrix} \\
&= (1-\lambda)\begin{vmatrix} 1-\lambda & -1 \\ -1 & 1-\lambda \end{vmatrix} - (-1)\begin{vmatrix} -1 & -1 \\ 1 & 1-\lambda \end{vmatrix} + \begin{vmatrix} -1 & 1-\lambda \\ 1 & -1 \end{vmatrix} \\
&= (1-\lambda)((1-\lambda)^2 - 1) - ((1-\lambda) - 1) + (1 - (1-\lambda)) \\
&= (1-\lambda)^3 - (1-\lambda) + 1 - (1-\lambda) - (1-\lambda) + 1 \\
&= (1-\lambda)^3 - 3 + 2 + 3\lambda \\
&= (1-\lambda)^3 + 3\lambda - 1 \\
&= (1 - 3\lambda + 3\lambda^2 - \lambda^3) + 3\lambda - 1 \\
&= 3\lambda^2 - \lambda^3 \\
&= \lambda^2(3-\lambda)
\end{aligned}$$

Therefore, $\det(A - \lambda I_3) = 0$ gives us $\lambda^2(3-\lambda) = 0$, so the eigenvalues of A are $\lambda = 0$ and $\lambda = 3$.

Next, to find the eigenspaces corresponding to each λ, if $\vec{v} = (x, y, z) \in \ker(A - \lambda I_3)$, then $(A - \lambda I_3)(\vec{v}) = \vec{0}$. If $\lambda = 0$, then $A - 0I_3 = \vec{0}$ becomes

$$\left[\begin{array}{ccc|c} 1 & -1 & 1 & 0 \\ -1 & 1 & -1 & 0 \\ 1 & -1 & 1 & 0 \end{array}\right] \xrightarrow{RR} \left[\begin{array}{ccc|c} 1 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

In this case, $x = y - z$, and y and z are arbitrary real numbers. Thus,

$$\vec{v} = \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} y - z \\ y \\ z \end{bmatrix} = \begin{bmatrix} y \\ y \\ 0 \end{bmatrix} + \begin{bmatrix} -z \\ 0 \\ z \end{bmatrix} = y \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + z \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$$

Thus, as $\vec{v} \in \ker(A - \lambda I_3)$ was arbitrary, then

$$\mathcal{E}_A(0) = \operatorname{span}\left\{ \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \right\}$$

Similarly, if $\lambda = 3$, then $A - 3\lambda I_3 = \vec{0}$ becomes

$$\left[\begin{array}{ccc|c} -2 & -1 & 1 & 0 \\ -1 & -2 & -1 & 0 \\ 1 & -1 & -2 & 0 \end{array}\right] \xrightarrow{RR} \left[\begin{array}{ccc|c} 1 & 0 & -1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

In this case, $x = z$ and $y = -z$, and z is an arbitrary real number. Thus,

$$\vec{v} = \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} z \\ -z \\ z \end{bmatrix} = z \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$$

Thus, as $\vec{v} \in \ker(A - 3\lambda I_3)$ was arbitrary, then

$$\mathcal{E}_A(3) = \operatorname{span}\left\{ \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} \right\}$$

Example 5.1.14. Let $V = \mathcal{P}_2(\mathbb{R})$ be equipped with the basis $\mathcal{B} = \{1, x, x^2\}$. Let $T : \mathcal{P}_2(\mathbb{R}) \to \mathcal{P}_2(\mathbb{R})$ be defined by

$$T(ax^2 + bx + c) = (a - b + c)x^2 + (-a + b - c)x + (a - b + c)$$

Denoting $A = [T]_{\mathcal{B}}^{\mathcal{B}}$, observe that

$$T(1) = x^2 - x + 1, T(x) = -x^2 + x - 1, T(x^2) = x^2 - x + 1$$

so then

$$[T(1)]_{\mathcal{B}} = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}, [T(x)]_{\mathcal{B}} = \begin{bmatrix} -1 \\ 1 \\ -1 \end{bmatrix}, [T(x^2)]_{\mathcal{B}} = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$$

Thus,

$$A = [T]_{\mathcal{B}}^{\mathcal{B}} = \begin{bmatrix} 1 & -1 & 1 \\ -1 & 1 & -1 \\ 1 & -1 & 1 \end{bmatrix}$$

which is the same matrix as in Example 5.1.13. In this case, the eigenvalues of T are the eigenvalues of $[T]_{\mathcal{B}}^{\mathcal{B}}$, so $\lambda = 0$ and $\lambda = 3$ from Example 5.1.13. In this case, because the 0-eigenspace of A is given by

$$\mathcal{E}_A(0) = \operatorname{span}\left\{ \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \right\}$$

Then the 0-eigenspace of T is given by

$$\mathcal{E}_T(0) = \operatorname{span}\{1 + x, -1 + x^2\}$$

Similarly, since the 3-eigenspace of A is given by

$$\mathcal{E}_A(3) = \operatorname{span}\left\{ \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} \right\}$$

Then the 3-eigenspace of T is given by

$$\mathcal{E}_T(3) = \operatorname{span}\{1 - x + x^2\}$$

5.2 Diagonalization

In the previous section, we introduced the concepts of eigenvalues and eigenvectors, which describe how certain vectors are scaled by a linear transformation without changing direction. We now turn to an important structural question: when can a linear operator or a matrix be represented in its simplest possible form — as a diagonal matrix?

This leads us to the notion of *diagonalization*, which is the process of finding a basis of eigenvectors for a linear operator. If such a basis exists, the operator can be represented by a diagonal matrix with respect to that basis, making computations such as powers of the transformation, exponentials, and function applications extremely straightforward.

Assume for a moment, that we have a linear map $T : V \to V$ and a basis $\mathcal{B} = \{\vec{v}_1, \vec{v}_2, ..., \vec{v}_n\}$ for V consisting of the eigenvectors of T, that is,

$$\begin{aligned} T(\vec{v}_1) &= \lambda_1 \vec{v}_1 = \lambda_1 \vec{v}_1 + 0\vec{v}_2 + \cdots + 0\vec{v}_n \\ T(\vec{v}_2) &= \lambda_2 \vec{v}_2 = 0\vec{v}_1 + \lambda_2 \vec{v}_2 + \cdots + 0\vec{v}_n \\ &\vdots \\ T(\vec{v}_n) &= \lambda_n \vec{v}_n = 0\vec{v}_1 + 0\vec{v}_2 + \cdots + \lambda_n \vec{v}_n \end{aligned}$$

then the matrix $[T]_{\mathcal{B}}^{\mathcal{B}}$ that corresponds to the linear map T is given by

$$[T]_{\mathcal{B}}^{\mathcal{B}} = \begin{bmatrix} \lambda_1 & 0 & 0 & \cdots & 0 \\ 0 & \lambda_2 & 0 & \cdots & 0 \\ 0 & 0 & \lambda_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & \lambda_n \end{bmatrix} = \operatorname{diag}(\lambda_1, \lambda_2, ..., \lambda_n)$$

Such a $[T]_{\mathcal{B}}^{\mathcal{B}}$ is called a *diagonal matrix*.

One question that we may ask is, can all linear maps be represented by diagonal matrices? The answer to this is not always. Only those linear maps whose eigenvectors form a basis for V. For example, symmetric, self-adjoint, and unitary maps possesses this property.

Definition 5.2.1. Let V be a vector space over $\mathbb{F}$. A linear map $T : V \to V$ is said to be *diagonalizable* over $\mathbb{F}$ if there exists a basis of V consisting of eigenvectors.

Example 5.2.2. Let V be an inner product space over $\mathbb{K}$ and $\mathcal{B} = \{\vec{v}_1, \vec{v}_2, ..., \vec{v}_n\}$ be an orthonormal basis. Let $W = \operatorname{span}(\{\vec{v}_1, \vec{v}_2, ..., \vec{v}_m\})$, where $m < n$ be a subspace of V, and let $P : V \to V$ be the orthogonal projection onto W defined by

$$P(\vec{v}) = \langle \vec{v}, \vec{v}_1 \rangle \vec{v}_1 + \langle \vec{v}, \vec{v}_2 \rangle \vec{v}_2 + \cdots + \langle \vec{v}, \vec{v}_m \rangle \vec{v}_m$$

We consider the following cases.

<u>Case 1: If $1 \leq k \leq n$.</u> Then $\vec{v}_k \in W$ is an eigenvector with eigenvalue 1. To see this, observe that

$$P(\vec{v}_k) = \langle \vec{v}_k, \vec{v}_1 \rangle \vec{v}_1 + \langle \vec{v}_k, \vec{v}_2 \rangle \vec{v}_2 + \cdots + \langle \vec{v}_k, \vec{v}_k \rangle \vec{v}_k + \cdots + \langle \vec{v}_k, \vec{v}_m \rangle \vec{v}_m = \vec{v}_k$$

so each $\vec{v}_k$ is the eigenvalue of P with $\lambda = 1$.

Case 2: If $m+1 \le k \le n$. Then $\vec{v}_k$ is an eigenvector with eigenvalue 0. To see this, note that

$$P(\vec{v}_k) = \langle \vec{v}_k, \vec{v}_1 \rangle \vec{v}_1 + \cdots + \langle \vec{v}_k, \vec{v}_m \rangle \vec{v}_m = 0 = 0\vec{v}_k$$

so each $\vec{v}_k$ is an eigenvector with $\lambda = 0$.

From Case 1 and Case 2, the matrix of P with respect to basis $\mathcal{B}$ is diagonal, with

$$[P]_{\mathcal{B}}^{\mathcal{B}} = \operatorname{diag}(\underbrace{1, 1, ..., 1}_{m \text{ times}}, \underbrace{0, 0, ..., 0}_{n-m \text{ times}})$$

Moreover, $[P]_{\mathcal{B}}^{\mathcal{B}}$ is self-adjoint, so the map P is also self-adjoint, that is, $P^* = P$.

We have already seen the concept of diagonalization of a matrix in previous linear algebra courses. We define the concept here.

Definition 5.2.3. A matrix $A \in \mathcal{M}_n(\mathbb{F})$ is said to be *diagonalizable* over $\mathbb{F}$ if $T_A : \mathbb{F}^n \to \mathbb{F}^n$ defined by

$$T_A(\vec{x}) = A\vec{x}$$

is diagonalizable over $\mathbb{F}$.

According to the above definition, a matrix A is diagonalizable if and only if there exists a basis $\mathcal{B}$ for $\mathbb{F}^n$ such that $[T_A]_{\mathcal{B}}^{\mathcal{B}}$ is diagonal. Of course, the basis $\mathcal{B}$ must consist of eigenvectors of T_A. Recall that if $\mathcal{SB}$ is the standard basis of $\mathbb{F}^n$, then $[T_A]_{\mathcal{SB}}^{\mathcal{SB}} = A$, and

$$A = [T_A]_{\mathcal{SB}}^{\mathcal{SB}} = [\mathrm{id}_V]_{\mathcal{SB}}^{\mathcal{B}} [T_A]_{\mathcal{B}}^{\mathcal{B}} ([\mathrm{id}_V]_{\mathcal{SB}}^{\mathcal{B}})^{-1}$$

If we set $P = [\mathrm{id}_V]_{\mathcal{SB}}^{\mathcal{B}}$, then

$$A = P[T_A]_{\mathcal{B}}^{\mathcal{B}} P^{-1}$$

This discussion leads us to the following definition.

Definition 5.2.4. Two matrices $A, B \in \mathcal{M}_n(\mathbb{F})$ are said to be *similar* if there exists an invertible matrix $P \in \mathcal{M}_n(\mathbb{F})$ such that

$$A = PBP^{-1}$$

Two similar matrices have the same characteristic polynomial, which means they have the same eigenvalues.

Proposition 5.2.5. *If $A, B \in \mathcal{M}_n(\mathbb{F})$ are similar matrices, the eigenvalues of A coincide with the eigenvalues of B.*

Proof. To show that the eigenvalues of A and B are the same, we show that $\chi_A(\lambda) = \chi_B(\lambda)$. Indeed, observe that

$$\begin{aligned}
\chi_A(\lambda) &= \det(A - \lambda I_n) \\
&= \det(PBP^{-1} - \lambda I_n) \\
&= \det(PBP^{-1} - \lambda PP^{-1}) \\
&= \det(P(B - \lambda I_n)P^{-1}) \\
&= \det(P)\det(B - \lambda I_n)\det(P^{-1}) \\
&= \det(P)\det(B - \lambda I_n)\frac{1}{\det(P)} \\
&= \det(B - \lambda I_n) \\
&= \chi_B(\lambda)
\end{aligned}$$

as desired. $\square$

Example 5.2.6. Consider the matrix defined by

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

We claim that A is diagonalizable over $\mathbb{R}$. To see this, note that from Example 5.1.13, we have found eigenvalues $\lambda = 3$ and $\lambda = 0$, and we also found that $\begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$ is an eigenvalue for A corresponding to $\lambda = 3$, and $\begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$ are eigenvectors for A with $\lambda = 0$.

By Corollary 5.1.8,

$$\mathcal{B} = \left\{ \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \right\}$$

is a basis of $\mathbb{R}^3$ consisting of eigenvectors of A, so A is diagonalizable.

Note that

$$D = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \operatorname{diag}(3, 0, 0)$$

and

$$P = [I_3]^{\mathcal{B}}_{\mathcal{SB}} = \begin{bmatrix} 1 & -1 & 1 \\ -1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}$$

One can show that P is an invertible matrix with

$$P^{-1} = ([I_3]^{\mathcal{B}}_{\mathcal{SB}})^{-1} = \begin{bmatrix} 1 & -1 & 1 \\ -1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}^{-1}$$

and one can also check that $A = PDP^{-1}$.

Remark 5.2.7. The order of the eigenvectors matters when defining the invertible matrix P. In Example 5.2.6, the first column of P corresponds to the eigenvector that corresponds to 3, second column of P corresponds to the eigenvector that corresponds to 0, and third column of P corresponds to the eigenvector that corresponds to 0.

Example 5.2.8. Let A be the matrix defined by

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

This matrix is diagonalizable over $\mathbb{C}$, but it cannot be diagonalizable over $\mathbb{R}$. To see the latter, note that $\chi_A(\lambda) = \lambda^2 + 1$ has no roots, so we do not have eigenvalues, and thus, no eigenvectors.

On the other hand, over $\mathbb{C}$, the eigenvalues we obtain are $\lambda = i$ and $\lambda = -i$. Then one can check that $\begin{bmatrix} -i \\ 1 \end{bmatrix}$ is an eigenvector for A corresponding to the eigenvalue i and $\begin{bmatrix} i \\ 1 \end{bmatrix}$ is an eigenvector for A corresponding to the eigenvalue $-i$. Thus, $\left\{ \begin{bmatrix} -i \\ 1 \end{bmatrix}, \begin{bmatrix} i \\ 1 \end{bmatrix} \right\}$ is a basis of eigenvectors for $\mathbb{C}^2$. Therefore, we can define $D = \operatorname{diag}(i, -i)$, and an invertible matrix P by

$$P = \begin{bmatrix} -i & i \\ 1 & 1 \end{bmatrix}$$

and one can check that $A = PDP^{-1}$.

Example 5.2.9. Not all matrices are diagonalizable over $\mathbb{C}$. For example, $\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$ is not diagonalizable over $\mathbb{C}$. Note that the eigenvalues are $\lambda = 0$, but then if A were to be diagonalizable, then we can write $A = PDP^{-1}$ for some invertible matrix P. But then $D = \operatorname{diag}(0,0)$, which gives us $A = O_2$, which is a contradiction.

Remark 5.2.10. Keep in mind that checking whether a linear map is diagonalizable or not can be done through checking whether its corresponding matrix $[T]_{\mathcal{B}}^{\mathcal{B}}$ is diagonalizable or not. For example, the matrix in Example 5.2.6 comes from Example 5.1.13. In Example 5.2.6, we showed that the matrix is diagonalizable, which implies that the corresponding linear map is diagonalizable as well. As a result, it is very important to discover conditions under which a given matrix is diagonalizable or not.

Definition 5.2.11. Let $A \in \mathcal{M}_n(\mathbb{F})$ and let λ be an eigenvalue of A. The *geometric multiplicity* of λ, denoted by $\operatorname{gm}_A(\lambda)$, is defined to be

$$\operatorname{gm}_A(\lambda) = \dim(\ker(A - \lambda I_n))$$

and the *algebraic multiplicity* of λ, denoted by $\operatorname{am}_A(\lambda)$ is defined to be the number which is the exponent appearing in the term $(x - \lambda)$, when the characteristic polynomial is factored.

Example 5.2.12. If $\chi_A(x) = x^3(x-1)^7(x-2)^{11}(x+2)^8$, then $\operatorname{am}_A(0) = 3$, $\operatorname{am}_A(1) = 7$, $\operatorname{am}_A(2) = 11$, and $\operatorname{am}_A(-2) = 8$. Note that

$$\sum_{\lambda \text{ is an eigenvalue}} \operatorname{am}_A(\lambda) = \deg(\chi_A) = n$$

Theorem 5.2.13. *Let $A \in \mathcal{M}_n(\mathbb{F})$. If $\lambda \in \mathbb{F}$ is an eigenvalue of A, then*

$$1 \leq \operatorname{gm}_A(\lambda) \leq \operatorname{am}_A(\lambda)$$

Proof. Let $\{\vec{w}_1, \vec{w}_2, ..., \vec{w}_m\}$ be a basis for $\mathcal{E}_A(\lambda)$. Then $\operatorname{gm}_A(\lambda) = m$ and now, we extend it to a basis

$$\mathcal{B} = \{\vec{w}_1, \vec{w}_2, ..., \vec{w}_m, \vec{v}_1, \vec{v}_2, ..., \vec{v}_k\}$$

of V. Then

$$E = [L_A]_{\mathcal{B}}^{\mathcal{B}} = \begin{bmatrix} \begin{matrix} \lambda & 0 & \cdots & 0 \\ 0 & \lambda & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \lambda \end{matrix} & C \\ \begin{matrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 \end{matrix} & D \end{bmatrix}$$

where the upper left block is $m \times m$, lower left block is $k \times m$, C block is $m \times k$, and D block is $k \times k$. Also note that $A = [L_A]_{\mathcal{SB}}^{\mathcal{SB}}$, so A and E are similar.

Hence,

$$\chi_A(x) = \chi_E(x) = (\lambda - x)^m \det(D - xI_n)$$

so $\operatorname{am}_A(\lambda) \geq m = \operatorname{gm}_A(\lambda)$. □

The following states the conditions for a matrix A to be diagonalizable.

Theorem 5.2.14. *Let $A \in \mathcal{M}_n(\mathbb{F})$. Then A is diagonalizable if and only if $\operatorname{gm}_A(\lambda) = \operatorname{am}_A(\lambda)$ for every eigenvalue λ of the matrix A.*

Proof. Assume that A is diagonalizable. Then there exists a basis of eigenvectors, n in total. Note that $\sum_{\lambda \text{ is an eigenvalue of } A} \operatorname{am}_A(\lambda) = n$. Assume for a contradiction that there exists λ_0 such that $\operatorname{gm}_A(\lambda_0) \neq \operatorname{am}_A(\lambda_0)$, i.e. $\operatorname{gm}_A(\lambda_0) < \operatorname{am}_A(\lambda)$. Then by Theorem 5.2.13,

$$\sum_\lambda \operatorname{gm}_A(\lambda) < \sum_\lambda \operatorname{am}_A(\lambda) = n$$

So the number of eigenvectors is less than n, which is absurd.

Conversely, since $\operatorname{gm}_A(\lambda) = \operatorname{am}_A(\lambda)$, so for any eigenvalue λ,

$$\sum_\lambda \operatorname{gm}_A(\lambda) = \sum_\lambda \operatorname{am}_A(\lambda) = n$$

so A is diagonalizable. Note that the first term on the left is the number of linearly independent eigenvectors, since we can combine linearly independent sets of $\operatorname{gm}_A(\lambda)$ vectors from each eigenspace to set n linearly independent eigenvector. □

In Example 5.2.6, the characteristic polynomial of A was

$$\chi_A(x) = x^2(x - 3)$$

with $\operatorname{am}_A(0) = 2$ and $\operatorname{am}_A(3) = 1$. We have also found that $\operatorname{gm}_A(0) = \dim(A - 0I_3) = 1$, and $\operatorname{gm}_A(2) = \dim(A - 2I_3) = 2$, which match the algebraic multiplicities. According to Theorem 5.2.14, A must be diagonalizable and that is what we have shown in Example 5.2.6.

5.3 Invariant Subspaces

Theorem 5.2.14 gives the conditions under which the matrix A is diagonalizable. However, sometimes it is not easy to check the conditions stated in the previous theorem. Recall that if $T : V \to V$ is a linear map and if V is has a basis of eigenvectors of T, then $[T]_{\mathcal{B}}^{\mathcal{B}}$ is diagonal, i.e. T is diagonalizable. If we do not know in advance whether V has a basis of eigenvectors of T, then we do not know if T can be represented by a diagonal matrix. However, the concept of invariant subspaces can guarantee that T can be represented to some extent by a simple matrix.

Before we define the concept of invariant subspaces, let us introduce some useful notations.

Definition 5.3.1. Let V be a vector space over $\mathbb{F}$ and let $T : V \to V$ be a linear map, and let $X \subseteq V$. Then the *image of* X is the set

$$T(\vec{x}) = \{T(\vec{x}) : \vec{x} \in X\} \subseteq V$$

Example 5.3.2. If $T : \mathcal{P}_2(\mathbb{R}) \to \mathcal{P}_2(\mathbb{R})$ is defined by

$$T(ax^2 + bx + c) = ax^2 + (a + b)x + (c + a)$$

and $X = \{x, x^2\}$, then $T(x) = x$ and $T(x^2) = x^2 + x + 1$, so

$$T(X) = \{x, x^2 + x + 1\}$$

Definition 5.3.3. Let V be a vector space over $\mathbb{F}$ and let $T : V \to V$ be a linear map. A subspace W of V is said to be T-*invariant* if $T(W) \subseteq W$.

Example 5.3.4. Let V be a vector space over $\mathbb{F}$ and let $T : V \to V$ be a linear map. If $W = \{\vec{0}\}$, then

$$T(W) = \{T(\vec{0})\} = \{\vec{0}\} \subseteq W$$

Example 5.3.5. Let V be a vector space over $\mathbb{F}$ and let $T : V \to V$ be a linear map and let $\vec{v}$ be an eigenvector of T with eigenvalue λ. Consider $W = \operatorname{span}\{\vec{v}\}$. Then we have $T(W) \subseteq W$.

Indeed, if $\vec{w} \in T(W)$, then there exists $\alpha \in \mathbb{F}$ such that $\vec{w} = T(\alpha\vec{v})$. But then

$$T(\alpha\vec{v}) = \alpha T(\vec{v}) = \alpha\lambda\vec{v} \in W$$

so $\vec{w} \in W$.

Example 5.3.6. Let V be a vector space over $\mathbb{F}$, $T : V \to V$ be a linear map, and let $\lambda \in \mathbb{F}$ be an eigenvalue of T. The λ-eigenspace $\mathcal{E}_T(\lambda)$ is also a T-invariant subspace.

Theorem 5.3.7. *Let V be a vector space over $\mathbb{F}$, let $T : V \to V$ be a linear map, and let W be a T-invariant subspace of V. Let $\mathcal{A} = \{\vec{w}_1, \vec{w}_2, ..., \vec{w}_m\}$ be a basis for W and let $\mathcal{B} = \{\vec{w}_1, \vec{w}_2, ..., \vec{w}_m, \vec{v}_1, \vec{v}_2, ..., \vec{v}_k\}$ be a basis for V. Then*

$$[T]_{\mathcal{B}}^{\mathcal{B}} = \begin{bmatrix} A & B \\ O_{k,m} & D \end{bmatrix}$$

where $A \in \mathcal{M}_m(\mathbb{F})$, $B \in \mathcal{M}_{mk}(\mathbb{F})$, and $O_{k,m}$ is the $k \times m$ zero matrix.

Proof. Since $T(W) \subseteq W = \operatorname{span}(\mathcal{A})$, then for each $\vec{w}_j$, $T(\vec{w}_j) \in \operatorname{span}(\mathcal{A})$. So then for each $1 \leq j \leq m$

$$T(\vec{w}_j) = \alpha_{1,j}\vec{w}_1 + \alpha_{2,j}\vec{w}_2 + \cdots + \alpha_{m,j}\vec{w}_m + 0\vec{v}_1 + 0\vec{v}_2 + \cdots + 0\vec{v}_k$$

Then

$$[T]_{\mathcal{B}}^{\mathcal{B}} = \begin{bmatrix} \begin{matrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mm} \end{matrix} & B \\ \begin{matrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 \end{matrix} & D \end{bmatrix}$$

□

Example 5.3.8. Let $V = \operatorname{span}\{\sin(kx), \cos(kx), 1, x, x^2\}$ and $T : V \to V$ be defined as $T(f) = f''$ for all $f \in V$. Let $W = \operatorname{span}\{\sin(kx), \cos(kx)\}$ with $\mathcal{A} = \{\sin(kx), \cos(kx)\}$ as a basis of W and $\mathcal{B} = \{\sin(kx), \cos(kx), 1, x, x^2\}$ as a basis for V.

We first show that W is T-invariant. Indeed, let $\vec{w} = \alpha_1 \sin(kx) + \alpha_2 \cos(kx)$. Then

$$\begin{aligned} T(\vec{w}) &= \alpha_1 T(\cos(kx)) + \alpha_2 T(\sin(kx)) \\ &= -\alpha_1 k^2 \sin(kx) - \alpha_2 k^2 \cos(kx) \\ &= -k^2 \vec{w} \in W \end{aligned}$$

Now observe that

$$\begin{aligned}
T(\sin(kx)) &= -k^2\sin(kx) + 0\cos(kx) + 0\cdot 1 + 0x + 0x^2\\
T(\cos(kx)) &= 0\sin(kx) - k^2\cos(kx) + 0\cdot 1 + 0x + 0x^2\\
T(1) &= 0\sin(kx) + 0\cos(kx) + 0\cdot 1 + 0x + 0x^2\\
T(x) &= 0\sin(kx) + 0\cos(kx) + 0\cdot 1 + 0x + 0x^2\\
T(x^2) &= 0\sin(kx) + 0\cos(kx) + 2\cdot 1 + 0x + 0x^2
\end{aligned}$$

Therefore,

$$[T]_{\mathcal{B}}^{\mathcal{B}} = \begin{bmatrix} -k^2 & 0 & 0 & 0 & 0\\ 0 & -k^2 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 2\\ 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

With the previous theorem, we made some progress regarding a representation of a linear map $T : V \to V$ by a simple matrix. However, we still have the same question. *What classes of linear maps are diagonalizable?* The answer to this question will be given by the spectral theorem, which we will discuss in the next section.

Definition 5.3.9. Let V be an inner product space over $\mathbb{K}$. A linear map $T : V \to V$ is called *normal* if $T^*T = TT^*$. A matrix $A \in \mathcal{M}_n(\mathbb{K})$ is called *normal* if $AA^* = A^*A$.

Self-adjoint and unitary linear maps (matrices) are always normal maps. For example, if T is self-adjoint, i.e. $T^* = T$, then $TT^* = TT = T^2$ and $T^*T = TT = T^2$, then $TT^* = T^*T$.

Example 5.3.10. Let $V = \mathcal{M}_2(\mathbb{C})$ be equipped with the following inner product $\langle A, B\rangle = \mathrm{tr}(AB^*)$ for all $A, B \in \mathcal{M}_2(\mathbb{C})$. Consider the linear map $T : V \to V$ defined by

$$T\left(\begin{bmatrix} a & b\\ c & d\end{bmatrix}\right) = \begin{bmatrix} ia & b\\ c & id\end{bmatrix}$$

Consider the standard basis $\mathcal{B}$ of $\mathcal{M}_2(\mathbb{C})$. It is elementary to verify that $\mathcal{B}$ is an orthonormal basis with respect to the given inner product. The matrices

$[T]_{\mathcal{B}}^{\mathcal{B}}$ and $([T]_{\mathcal{B}}^{\mathcal{B}})^*$ are given by

$$[T]_{\mathcal{B}}^{\mathcal{B}} = \begin{bmatrix} i & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & i \end{bmatrix}, \quad ([T]_{\mathcal{B}}^{\mathcal{B}})^* = \begin{bmatrix} -i & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -i \end{bmatrix}$$

Clearly, T is not self-adjoint. Observe that

$$[T]_{\mathcal{B}}^{\mathcal{B}}([T]_{\mathcal{B}}^{\mathcal{B}})^* = \begin{bmatrix} i & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & i \end{bmatrix} \begin{bmatrix} -i & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -i \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = I_4$$

and in a similar way, $([T]_{\mathcal{B}}^{\mathcal{B}})^*[T]_{\mathcal{B}}^{\mathcal{B}} = I_4$. Therefore, T is unitary, and furthermore, is normal.

Recall that T is self-adjoint if $T = T^*$ or equivalently the corresponding matrix is self-adjoint. Recall also that we have showed that eigenvectors correspond to distinct eigenvalues are linearly independent. In the case of self-adjoint maps, they are in fact orthogonal.

Theorem 5.3.11. *Let V be an inner product space over $\mathbb{K}$ and let $T : V \to V$ be a self-adjoint linear map. Then*

1. *The eigenvalues of T are real.*

2. *The eigenvectors of T corresponding to distinct eigenvalues are orthogonal.*

Proof. To see that (1) is true, let $\vec{v}$ be an eigenvector corresponding to its eigenvalue λ. We need to show that $\lambda = \overline{\lambda}$. Indeed, observe that

$$\lambda\langle \vec{v}, \vec{v}\rangle = \langle \lambda\vec{v}, \vec{v}\rangle = \langle A\vec{v}, \vec{v}\rangle = \langle \vec{v}, A^*\vec{v}\rangle = \langle \vec{v}, \lambda\vec{v}\rangle = \overline{\lambda}\langle \vec{v}, \vec{v}\rangle$$

Therefore, $(\lambda - \overline{\lambda})\langle \vec{v}, \vec{v}\rangle = 0$, but since $\langle \vec{v}, \vec{v}\rangle \neq 0$, it follows that $\lambda = \overline{\lambda}$, as desired.

To see that (2) is true, let $\vec{v}_1$ be an eigenvector corresponding to the eigenvalue λ_1, and let $\vec{v}_2$ be an eigenvector corresponding to the eigenvalue

λ_2. Then note that $T(\vec{v}_1) = \lambda_1\vec{v}_1$ and $T(\vec{v}_2) = \lambda_2\vec{v}_2$. Then

$$\begin{aligned}\lambda_1\langle\vec{v}_1, \vec{v}_2\rangle &= \langle\lambda_1\vec{v}_1, \vec{v}_2\rangle \\ &= \langle A\vec{v}_1, \vec{v}_2\rangle \\ &= \langle\vec{v}_1, A^*\vec{v}_2\rangle \\ &= \langle\vec{v}_1, A\vec{v}_2\rangle \\ &= \langle\vec{v}_1, \lambda_2\vec{v}_2\rangle \\ &= \lambda_2\langle\vec{v}_1, \vec{v}_2\rangle\end{aligned}$$

and so $\lambda_1 = \lambda_2$. Implying that $\vec{v}_1$ is orthogonal to $\vec{v}_2$, as desired. □

Similar to properties of transpose matrices, adjoints also have the same properties: For all $A, B \in \mathcal{M}_n(\mathbb{K})$

- $(A^*)^* = A$
- $(AB)^* = B^*A^*$
- $(A + B)^* = A^* + B^*$

Theorem 5.3.12. *Let $A, B \in \mathcal{M}_n(\mathbb{K})$ be unitary matrices. Then*

1. *AB is a unitary matrix.*
2. *The columns of A are orthonormal with respect to the standard inner product.*
3. *If λ is an eigenvalue of A, then $|\lambda| = 1$.*
4. *The eigenvectors from distinct eigenvalues are orthogonal.*

Proof. Exercise. □

5.4 Spectral Theorem

We posed a question earlier in this chapter. *What classes of linear maps are diagonalizable?* Let $T : V \to V$ be a linear map and let $\mathcal{B}$ be a basis of V. Recall that T is diagonalizable if and only if its corresponding matrix $[T]_{\mathcal{B}}^{\mathcal{B}}$ is diagonalizable.

As a result, our question can be equivalently stated: *What classes of matrices are diagonalizable?* or *For what classes of matrices A can we find an invertible matrix P such that $A = PDP^{-1}$?* The answer to this question is given by the Spectral Theorem.

Theorem 5.4.1 (**Spectral Theorem, Complex Case**). *Let $A \in \mathcal{M}_n(\mathbb{C})$. Then A is normal if and only if there exists a unitary matrix U such that $A = UDU^*$.*

Theorem 5.4.2 (**Spectral Theorem, Real Case**). *Let $A \in \mathcal{M}_n(\mathbb{R})$. Then A is symmetric if and only if there exists an orthogonal matrix U such that $A = UDU^T$.*

When $\mathbb{F} = \mathbb{R}$, according to the Spectral Theorem, only symmetric matrices are diagonalizable.

The proof of the Spectral Theorem relies on Schur's Theorem.

Theorem 5.4.3 (**Schur's Theorem**). *Let $A \in \mathcal{M}_n(\mathbb{F})$. Then there exists a unitary matrix U such that*

$$U^*AU = B$$

where $B \in \mathcal{M}_n(\mathbb{F})$ is an upper triangular matrix having the eigenvalues of A on the main diagonal.

Proof. We prove Schur's Theorem using induction.

Base Case: If $n = 1$, then $A = [a_{1,1}]$ is upper triangular, so in this case, $U = [1]$, which is 1×1 and is unitary.

Inductive Step: Suppose the theorem holds for any $(n-1)\times(n-1)$ matrix. We want to show that the theorem is true for any $n \times n$ matrix. Let A be an $n \times n$ matrix. Then there exists an eigenvalue λ_1 (which is guaranteed by the Fundamental Theorem of Algebra), and eigenvector $\vec{v}_1$ such that $A\vec{v}_1 = \lambda_1\vec{v}_1$ and $\|\vec{v}_1\| = 1$. Extend $\{\vec{v}_1\}$ to a basis $\{\vec{v}_1, \vec{v}_2, ..., \vec{v}_n\}$ of $\mathbb{F}^n$. We can assume it is orthonormal, otherwise, use Gram-Schmidt Orthogonalization Process to make it orthonormal. Let $U_0 = [\vec{v}_1, \vec{v}_2, ..., \vec{v}_n]$. Note that U_0 is unitary since the columns of U_0 are orthonormal. Then

$$U_0^*AU_0 = \begin{bmatrix} \lambda_1 & * & * & * \\ 0 & & & \\ \vdots & & A_1 & \\ 0 & & & \end{bmatrix}$$

Note that A_1 is $(n-1) \times (n-1)$ matrix so by the inductive step, there exists $(n-1) \times (n-1)$ unitary matrix $\tilde{U}_1$ such that $\tilde{U}_1^* A_1 \tilde{U}_1 = B_{n-1}$. Consider

$$U_1 = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & \tilde{U}_1 & \\ 0 & & & \end{bmatrix}$$

Then U_1 is unitary since its columns are orthonormal and

$$U_1^* = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & \tilde{U}_1^* & \\ 0 & & & \end{bmatrix}$$

Now we note that

$$\begin{aligned} U_1^* U_0^* A U_0 U_1 &= \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & \tilde{U}_1^* & \\ 0 & & & \end{bmatrix} \begin{bmatrix} \lambda_1 & * & * & * \\ 0 & & & \\ \vdots & & A_1 & \\ 0 & & & \end{bmatrix} \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & \tilde{U}_1 & \\ 0 & & & \end{bmatrix} \\ &= \begin{bmatrix} \lambda_1 & * & * & * \\ 0 & & & \\ \vdots & & \tilde{U}_1^* A & \\ 0 & & & \end{bmatrix} \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & \tilde{U}_1 & \\ 0 & & & \end{bmatrix} \\ &= \begin{bmatrix} \lambda_1 & * & * & * \\ 0 & & & \\ \vdots & & \tilde{U}_1^* A \tilde{U}_1 & \\ 0 & & & \end{bmatrix} \\ &= \begin{bmatrix} \lambda_1 & * & * & * \\ 0 & & & \\ \vdots & & B_{n-1} & \\ 0 & & & \end{bmatrix} \end{aligned}$$

If we set $U = U_0 U_1$, we just showed that $U^* A U = B$, i.e. A and B are similar; they have the same eigenvalues and as a result, the diagonal elements of B are the eigenvalues of A. □

Proof of Spectral Theorem, Complex Case (Theorem 5.4.1). From Schur's Theorem (Theorem 5.4.3), we have a unitary matrix U such that

$$U^* A U = B$$

We want to show that B is diagonal. Note that A has real eigenvalues since A is self-adjoint. Observe that

$$B^* = (U^* A U)^* = (AU)^* (U^*)^* = U^* A^* U = U^* A U = B$$

So $B^* = B$ and since B is an upper triangular matrix, then B must be diagonal. □

Proof of Spectral Theorem, Real Case (Theorem 5.4.2). Note that A is a symmetric matrix, so from Theorem 5.4.1, there exists a unitary matrix U such that

$$U^* A U = D$$

Since $\mathbb{F} = \mathbb{R}$, we have that $U^* = U^T$, and $U^T A U = D$. $\square$

Example 5.4.4. Consider the matrix

$$A = \begin{bmatrix} 2 & -1 & 2 & -2 \\ -1 & 2 & -2 & 2 \\ -2 & 2 & 2 & -1 \\ 2 & -2 & -1 & 2 \end{bmatrix}$$

It is elementary to verify that A is a normal matrix.

The characteristic polynomial is given by $\chi_A(\lambda) = (\lambda - 1)^2(\lambda^2 - 6\lambda + 25)$. For the case when $\mathbb{F} = \mathbb{R}$, note that A is not diagonalizable since $\lambda^2 - 6\lambda + 25 = 0$ has complex roots. On the other hand, if $\mathbb{F} = \mathbb{C}$, A is diagonalizable, since A is normal.

The eigenvalues of A are easily found, and they are $\lambda = 1$ with $\mathrm{am}_A(1) = 2$, and the roots of $\lambda^2 - 6\lambda + 25 = 0$ are $\lambda = 3 \pm 4i$, each with $\mathrm{am}_A(3 \pm 4i) = 1$. The corresponding eigenspaces for each eigenvalue are given as follows

$$\mathcal{E}_A(1) = \operatorname{span}\left\{ \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix} \right\},$$

$$\mathcal{E}_A(3+4i) = \operatorname{span}\left\{ \begin{bmatrix} i \\ -i \\ -1 \\ 1 \end{bmatrix} \right\},$$

$$\mathcal{E}_A(3-4i) = \operatorname{span}\left\{ \begin{bmatrix} -i \\ i \\ -1 \\ 1 \end{bmatrix} \right\}$$

Denote the diagonal matrix as

$$D = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 3+4i & 0 \\ 0 & 0 & 0 & 3-4i \end{bmatrix}$$

Taking the vectors $\vec{v}_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}$, $\vec{v}_2 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}$, $\vec{v}_3 = \begin{bmatrix} i \\ -i \\ -1 \\ 1 \end{bmatrix}$, and $\vec{v}_4 = \begin{bmatrix} -i \\ i \\ -1 \\ 1 \end{bmatrix}$, we find an invertible matrix P given by

$$P = \begin{bmatrix} \vec{v}_1 & \vec{v}_2 & \vec{v}_3 & \vec{v}_4 \end{bmatrix} = \begin{bmatrix} 1 & 0 & i & -i \\ 1 & 0 & -i & i \\ 0 & 1 & -1 & -1 \\ 0 & 1 & 1 & 1 \end{bmatrix}$$

such that $A = PDP^{-1}$.

Note that $\vec{v}_1 \perp \vec{v}_3$ and $\vec{v}_1 \perp \vec{v}_4$ since they correspond to distinct eigenvalues and A is normal. Also, for the same reason $\vec{v}_3 \perp \vec{v}_4$ since they correspond to distinct eigenvalues. Also, it can be checked that $\vec{v}_1 \perp \vec{v}_2$ since they correspond to the same eigenvalue. We need orthonormal vectors so divide each vector by its norm to get

$$\vec{u}_1 = \frac{\vec{v}_1}{\sqrt{2}}, \quad \vec{u}_2 = \frac{\vec{v}_2}{\sqrt{2}}, \quad \vec{u}_3 = \frac{\vec{v}_3}{2}, \quad \vec{u}_4 = \frac{\vec{v}_4}{2}$$

so that we can find a unitary matrix U given by

$$U = \begin{bmatrix} \vec{u}_1 & \vec{u}_2 & \vec{u}_3 & \vec{u}_4 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} & 0 & \frac{1}{2}i & -\frac{1}{2}i \\ \frac{1}{\sqrt{2}} & 0 & -\frac{1}{2}i & \frac{1}{2}i \\ 0 & \frac{1}{\sqrt{2}} & -\frac{1}{2} & -\frac{1}{2} \\ 0 & \frac{1}{\sqrt{2}} & \frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

such that $A = UDU^*$.

According to our process, we want to have an orthonormal set of vectors since the eigenvectors corresponding to distinct eigenvalues are already orthogonal and with an application of the Gram-Schmidt Orthogonalization Process, the eigenvectors corresponding to the same eigenvalue are orthogonal too. Also, this set of orthonormal eigenvectors is a basis of V as we know A is diagonalizable so the number of vectors must be the dimension of V.

5.5 Positive Matrices

The concept of positive matrices is ubiquitous in applied and pure sciences such as statistics, physics, computer science, etc. This concept is also used in MATH 2310 (Calculus of Several Variables with Applications) in optimization problems.

Definition 5.5.1. Let $A \in \mathcal{M}_n(\mathbb{F})$. Then A is said to be

1. *positive semidefinite* if A is self-adjoint and all eigenvalues of A are non-negative.

2. *positive definite* if A is self-adjoint and all eigenvalues of A are strictly positive.

Example 5.5.2. Consider the matrix $A = \begin{bmatrix} 1 & i \\ -i & 1 \end{bmatrix}$. It is elementary to verify that A is a self-adjoint matrix. The eigenvalues of A are given by $\lambda_1 = 0$ and $\lambda_2 = 2$. Therefore, A is a positive semidefinite matrix.

Example 5.5.3. Consider $D = \operatorname{diag}(\lambda_1, \lambda_2)$. Then D is positive semidefinite if $\lambda_1, \lambda_2 \geq 0$ and D is positive definite if $\lambda_1, \lambda_2 > 0$. For the case when $\mathbb{F} = \mathbb{R}$, and considering the standard inner product $\langle \vec{x}, \vec{y} \rangle = x_1 y_1 + x_2 + y_2$, note that

$$\langle \vec{x}, \vec{y} \rangle = \vec{x}^T \vec{y}$$

and so computing $Q(\vec{x}) = \langle D\vec{x}, \vec{x} \rangle$, we have

$$\begin{aligned} Q(\vec{x}) &= (D\vec{x})^T \vec{x} = \vec{x}^T D \vec{x} = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \\ &= \lambda x_1^2 + \lambda_2 x_2^2 \end{aligned}$$

Note that if $\lambda_1, \lambda_2 \geq 0$, then $Q(x) \geq 0$. This is true in general.

Theorem 5.5.4. *Let $A \in \mathcal{M}_n(\mathbb{K})$ be a self-adjoint matrix. Then A is positive semidefinite (positive definite) if and only if $\langle A\vec{v}, \vec{v} \rangle \geq 0$ (> 0) for all $\vec{v} \in \mathbb{K}^n$ (for all $\vec{v} \in \mathbb{K}^n \setminus \{\vec{0}\}$).*

Proof. We will prove for the case when A is a positive semidefinite matrix. Assume that A is positive semidefinite, and let $\lambda_1, \lambda_2, ..., \lambda_n$ be the non-negative eigenvalues of A. Them we have

$$\langle A\vec{v}, \vec{v} \rangle = (A\vec{v})^* \vec{v} = \vec{v}^* A^* \vec{v} = \vec{v}^* A \vec{v}$$

so by the Spectral Theorem (Theorem 5.4.1), there exists a unitary matrix U such that $A = U^* D U$. Then

$$\langle A\vec{v}, \vec{v} \rangle = \vec{v}^* U^* D U \vec{v}$$

Denote $U\vec{v} = \vec{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$ and note that

$$\langle A\vec{v}, \vec{v}\rangle = \vec{y}^* D\vec{y} = \lambda_1|y_1|^2 + \lambda_2|y_2|^2 + \cdots + \lambda_n|y_n|^2$$

Since all eigenvalues are nonnegative, it follows that $\langle A\vec{v}, \vec{v}\rangle \geq 0$.

Conversely, by the Spectral Theorem (Theorem 5.4.1), we have

$$\langle A\vec{v}, \vec{v}\rangle = \vec{y}^* D\vec{y} = \lambda_1|y_1|^2 + \lambda_2|y_2|^2 + \cdots + \lambda_n|y_n|^2$$

Using $\vec{y} = U\vec{v}$, note that for $\vec{v} = U^*\vec{e}_j$, we have

$$\vec{y} = UU^8\vec{e}_j = \vec{e}_j$$

and thus,

$$0 \leq \langle A\vec{v}, \vec{v}\rangle = \lambda_j|y_j|^2$$

so $\lambda_j \geq 0$, and thus, A is positive semidefinite as desired. □

We can determine if A is positive semidefinite (definite) just by checking the determinants of submatrices of A, but we need the following definition first.

Definition 5.5.5. Let $A \in \mathcal{M}_n(\mathbb{K})$ and let $1 \leq m \leq n$. The *mth principle submatrix* of A, denoted by $A_{(m)}$ is the upper left $m \times m$ submatrix of A, that is,

$$A = \begin{bmatrix} A_{(m)} & B \\ C & D \end{bmatrix}$$

where $B \in \mathcal{M}_{m,n-m}(\mathbb{K})$, $C \in \mathcal{M}_{n-m,m}(\mathbb{K})$, and $D \in \mathcal{M}_{n-m}(\mathbb{K})$.

Theorem 5.5.6. *Let $A \in \mathcal{M}_n(\mathbb{K})$ be a self-adjoint. Then A is positive semidefinite (definite) if and only if*

$$\det(A_{(m)} \geq 0 \quad (> 0)$$

for all $1 \leq m \leq n$.

Proof. Exercise. □

Example 5.5.7. Let $A = \begin{bmatrix} 10 & 0 & 0 \\ 5 & 3 & 2 \\ 2 & 2 & 3 \end{bmatrix}$. Note that

$$\begin{aligned} A_{(1)} = [10] &\implies \det(A_{(1)}) > 0 \\ A_{(2)} = \begin{bmatrix} 10 & 0 \\ 5 & 3 \end{bmatrix} &\implies \det(A_{(2)}) > 0 \\ A_{(3)} = A &\implies \det(A_{(3)}) > 0 \end{aligned}$$

In this case, however, we cannot conclude that A is positive definite because A is not even self-adjoint.

5.6 Singular Value Decomposition

Given a square matrix $A \in \mathcal{M}_n(\mathbb{K})$, according to the Spectral Theorem (Theorem 5.4.1), under some conditions on A, the matrix A can be factored as

$$A = UDU^*$$

where U is a unitary matrix.

In the case of a rectangular matrix $A \in \mathcal{M}_{n,m}(\mathbb{K})$, can we find a similar factorization in this case? The answer lies within the Singular Value Decomposition. But first, we need to introduce some definitions and some theory.

Theorem 5.6.1. *Let $A \in \mathcal{M}_{m,n}(\mathbb{K})$. Then $A^*A \in \mathcal{M}_{n,n}(\mathbb{K})$ is self-adjoint and positive semidefinite.*

Proof. Note that $(A^*A)^* = A^*(A^*)^* = A^*A$, so A is self-adjoint. Furthermore,

$$\langle A^*A\vec{v}, \vec{v}\rangle = \langle A\vec{v}, A\vec{v}\rangle \geq 0$$

Therefore, A^*A is positive semidefinite by Theorem 5.5.6. □

Since A^*A is positive semidefinite, its eigenvalues are nonnegative.

Definition 5.6.2. Let $A \in \mathcal{M}_{m,n}(\mathbb{K})$ and let $\lambda_1, \lambda_2, , .., \lambda_n$ be the eigenvalues of A^*A. The *singular values* of A are the square roots of the eigenvalues of A^*A, denoted by $\sigma_1 = \sqrt{\lambda_1}$, $\sigma_2 = \sqrt{\lambda_2}$,..., $\sigma_n = \sqrt{\lambda_n}$.

Note that if the rank of A is r, we place the singular values of A in non-increasing order, that is, $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n$, and then $\sigma_{r+1} = \sigma_{r+2} = \cdots = \sigma_n = 0$.

Theorem 5.6.3 (Singular Value Decomposition). *Let $A \in \mathcal{M}_{m,n}(\mathbb{K})$ with rank r and $m \times n$ diagonal matrix*

$$\Sigma = \begin{bmatrix} D_{r\times r} & O_{r\times n-r} \\ O_{m-r\times r} & O_{m-r\times n-r} \end{bmatrix}$$

where the diagonal entries of $D_{r\times r}$ are the first singular values of A, $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$. Then there exists a $m \times m$ unitary matrix U and a $n \times n$ unitary matrix V such that

$$A = U\Sigma V^*$$

Example 5.6.4. Consider the matrix $A = \begin{bmatrix} 1 & -1 \\ 1 & -1 \\ \sqrt{3} & 0 \end{bmatrix}$. Note that $A^* = \begin{bmatrix} 1 & 1 & \sqrt{3} \\ -1 & 1 & 0 \end{bmatrix}$. Multiplying A^* and A together, we obtain

$$A^*A = \begin{bmatrix} 1 & 1 & \sqrt{3} \\ -1 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 1 & -1 \\ \sqrt{3} & 0 \end{bmatrix} = \begin{bmatrix} 5 & -2 \\ -2 & 2 \end{bmatrix}$$

and its eigenvalues are $\lambda_1 = 6$ and $\lambda_2 = 1$. Thus, its singular values are $\sigma_1 = \sqrt{6}$ and $\sigma_2 = 1$.

The eigenspaces of each eigenvalue are given as

$$\mathcal{E}_{A^*A}(6) = \text{span}\left\{ \begin{bmatrix} 2 \\ -1 \end{bmatrix} \right\}, \quad \mathcal{E}_{A^*A}(1) = \text{span}\left\{ \begin{bmatrix} 1 \\ 2 \end{bmatrix} \right\}$$

Taking $\vec{x}_1 = \begin{bmatrix} 2 \\ -1 \end{bmatrix}$ and $\vec{x}_2 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$, we note that $\vec{x}_1 \perp \vec{x}_2$ since they correspond to distinct eigenvalues and A is self-adjoint. We do not need to apply the Gram-Schmidt Orthogonalization Process, however, we need to normalize them. Indeed, setting

$$\vec{v}_1 = \frac{\vec{x}_1}{\sqrt{5}}, \quad \vec{v}_2 = \frac{\vec{x}_2}{\sqrt{5}}$$

Then the unitary matrix V is given as

$$V = \begin{bmatrix} \vec{v}_1 & \vec{v}_2 \end{bmatrix} = \begin{bmatrix} \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \\ -\frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \end{bmatrix}$$

Since we have $\lambda_1 > \lambda_2$, we have $r = 2$, so

$$\vec{u}_1 = \frac{1}{\sigma_1} A\vec{v}_1 = \frac{1}{\sqrt{30}} \begin{bmatrix} 3 \\ 3 \\ 2\sqrt{3} \end{bmatrix}$$

$$\vec{u}_2 = \frac{1}{\sigma_2} A\vec{v}_2 = \frac{1}{\sqrt{5}} \begin{bmatrix} -1 \\ -1 \\ \sqrt{3} \end{bmatrix}$$

Note that if $\{\vec{u}_1, \vec{u}_2, ..., \vec{u}_r\}$ is already a basis of $\mathbb{K}^m$, then skip this step. Otherwise, extend $\{\vec{u}_1, \vec{u}_2, ..., \vec{u}_r\}$ to an orthonormal basis of $\{\vec{u}_1, \vec{u}_2, ..., \vec{u}_m\}$ of $\mathbb{K}^m$.

In this case, we have $\{\vec{u}_1, \vec{u}_2\}$, but we need to extend it to an orthonormal basis of $\mathbb{K}^3$. Need to find $\vec{w} = [a, b, c]^T$ such that $\langle \vec{u}_1, \vec{w} \rangle = \langle \vec{u}_2, \vec{w} \rangle = 0$. We have a system given as

$$\begin{cases} 3a + 3b + 2\sqrt{3}c = 0 \\ -a - b - \sqrt{3}c = 0 \end{cases}$$

and so $\vec{w} \in \operatorname{span} \left\{ \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} \right\}$. Now we normalize $\vec{w}$ to get

$$\vec{u}_3 = \frac{\vec{w}}{\|\vec{w}\|} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}$$

So we have an orthonormal basis $\{\vec{u}_1, \vec{u}_2, \vec{u}_3\}$ of $\mathbb{K}^3$. Therefore, we find a unitary matrix U given by

$$U = \begin{bmatrix} \vec{u}_1 & \vec{u}_2 & \vec{u}_3 \end{bmatrix} = \begin{bmatrix} \frac{3}{\sqrt{30}} & -\frac{1}{\sqrt{5}} & \frac{1}{\sqrt{2}} \\ \frac{3}{\sqrt{30}} & -\frac{1}{\sqrt{5}} & -\frac{1}{\sqrt{2}} \\ \frac{2\sqrt{3}}{\sqrt{30}} & \frac{\sqrt{3}}{\sqrt{5}} & 0 \end{bmatrix}$$

Furthermore, we have

$$\Sigma = \begin{bmatrix} 6 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$$

So we have

$$A = U\Sigma V^* = \begin{bmatrix} \frac{3}{\sqrt{30}} & -\frac{1}{\sqrt{5}} & \frac{1}{\sqrt{2}} \\ \frac{3}{\sqrt{30}} & -\frac{1}{\sqrt{5}} & -\frac{1}{\sqrt{2}} \\ \frac{2\sqrt{3}}{\sqrt{30}} & \frac{\sqrt{3}}{\sqrt{5}} & 0 \end{bmatrix} \begin{bmatrix} 6 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \\ -\frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \end{bmatrix}$$

Index