MAT 3040 - Advanced Linear Algebra

Foreword

This book is taken notes from the MAT 3040 in spring semester, 2019. These lecture notes were taken and compiled in LaTeX by Jie Wang, an undergraduate student in spring 2019. The tex writter would like to thank Prof. Daniel Wong and some students for their detailed and valuable comments and suggestions, which significantly improved the quality of this notebook. Students taking this course may use the notes as part of their reading and reference materials. This version of the lecture notes were revised and extended for many times, but may still contain many mistakes and typos, including English grammatical and spelling errors. It would be greatly appreciated if those students, who will use the notes as their reading or reference material, tell any mistakes and typos to Jie Wang for improving this notebook.

Contents

Notations and Conventions

$\mathbb{F}^n$	n-dimensional $\mathbb{F}$-valued space
$M_{m \times n}(\mathbb{F})$	set of all $(m \times n)$ $\mathbb{F}$-valued matrices
$\oplus$	Direct Sum
$\ker(T)$	The null space of T
$V \cong W$	vector spaces V and W are isomorphic
$(T)_{\mathcal{B},\mathcal{A}}$	Matrix representation of T w.r.t. $\mathcal{A}$ and $\mathcal{B}$
$\mathbf{v} + W$	coset of $\mathbf{v}$ in $W \leq V$, i.e., $\{\mathbf{v} + \mathbf{w} \mid \mathbf{w} \in W\}$
A_i^{T}	i^{th} row of matrix A
V/W	Quotient space of V by the subspace $W \leq V$
V^*	Dual space of V, i.e. the set of linear transformations from V to $\mathbb{F}$
$\mathrm{Ann}(S)$	The annihilator of $S \subseteq V$, i.e. $\{f \in V^* \mid f(s) = 0, \forall s \in S\}$
$T^* : W^* \to V^*$	Adjoint map for the mapping $T : V \to W$
A^t	transpose of A, i.e, $B = A^t$ means $b_{ji} = a_{ij}$ for all i, j
A^{H}	Hermitian transpose of A, i.e, $B = A^{\mathrm{H}}$ means $b_{ji} = \bar{a}_{ij}$ for all i, j
$\mathcal{X}_T(x)$	characteristic polynomial of T
$m_T(x)$	Minimal polynomial of the linear operator T
$T' : V \to V$	Hermitian Adjoint for $T : V \to V$
$\langle \mathbf{v}, \mathbf{w} \rangle$	Inner product between vectors $\mathbf{v}$ and $\mathbf{w}$
$V \otimes W$	Tensor product between vector spaces V and W
$V \wedge V$	Wedge product for vector space V

Chapter 1

Review of Vector Space and Basis

1.1 Introduction

Advanced Linear Algebra (MAT 3040) is one of the most important courses in MATH major, with prerequisite MAT 2040 and MAT 1001/1011. This course will offer **true** linear algebra knowledge.

What will be covered?

- In MAT 2040 we have studied the space $\mathbb{R}^n$; while in MAT 3040 we will study the general vector space V.

- In MAT 2040 we have studied the linear transformation between Euclidean spaces, i.e., $T : \mathbb{R}^n \to \mathbb{R}^m$; while in MAT 3040 we will study the linear transformation from vector spaces to vector spaces: $T : V \to W$

- In MAT 2040 we have studied the eigenvalues of $n \times n$ matrix A ; while in MAT 3040 we will study the eigenvalues of a linear operator $T : V \to V$.

- In MAT 2040 we have studied the dot product $\mathbf{x} \cdot \mathbf{y} = \sum\limits_{i=1}^{n} x_i y_i$; while in MAT 3040 we will study the inner product $\langle \mathbf{v}_1, \mathbf{v}_2 \rangle$.

Why generalize? In mathematics (and beyond), we come across many other spaces than $\mathbb{R}^n$, for example:

- $\mathcal{C}(\mathbb{R})$, the space of all functions on $\mathbb{R}$,

- $\mathcal{C}^\infty(\mathbb{R})$, the space of all infinitely differentiable functions on $\mathbb{R}$,

- $\mathbb{R}[x]$, the space of polynomials of one real variable.

Here are some applications in partial differential equations and physics:

Example 1.1. 1. Consider the Laplace equation $\Delta f = 0$ with Laplace operator:

$$\Delta : C^\infty\left(\mathbb{R}^3\right) \to C^\infty\left(\mathbb{R}^3\right), \qquad f \mapsto \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\right) f$$

The solution to the partial differential equation $\Delta f = 0$ is the 0-eigenspace of Δ.

2. Consider the Schrödinger equation $\widehat{H} f = E f$ with the Schrödinger operator

$$\widehat{H} : C^\infty \left(\mathbb{R}^3 \right) \to C^\infty \left(\mathbb{R}^3 \right), \qquad f \mapsto \left(\frac{-\hbar^2}{2\mu} \nabla^2 + V \left(x, y, z \right) \right) f$$

Solving the equation $\widehat{H} f = \lambda f$ is equivalent to finding the λ-eigenspace of $\widehat{H}$. In fact, a major observation in quantum mechanics is that the eigenvalues of $\widehat{H}$ are discrete.

1.2 Vector Spaces

Definition 1.2 (Vector Space). A **vector space** over a field $\mathbb{F}$ (in particular, $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$) is a set of objects V equipped with vector addiction $+$ and scalar multiplication $\cdot$ such that the following rules are satisfied:

- Commutativity: $\mathbf{v}_1 + \mathbf{v}_2 = \mathbf{v}_2 + \mathbf{v}_1$.
- Associativity: $\mathbf{v}_1 + (\mathbf{v}_2 + \mathbf{v}_3) = (\mathbf{v}_1 + \mathbf{v}_2) + \mathbf{v}_3$.
- Zero vector: there exists $\mathbf{0} \in V$ such that $\mathbf{0} + \mathbf{v} = \mathbf{v}$ for all $\mathbf{v} \in V$.
- Additive inverse: for all $\mathbf{v} \in V$, there exists $-\mathbf{v} \in V$ such that $\mathbf{v} + (-\mathbf{v}) = \mathbf{0}$.
- Distributivity of vectors: $\alpha \left(\mathbf{v}_1 + \mathbf{v}_2 \right) = \alpha \mathbf{v}_1 + \alpha \mathbf{v}_2$.
- Distributivity of scalars: $\left(\alpha_1 + \alpha_2 \right) \mathbf{v} = \alpha_1 \mathbf{v} + \alpha_2 \mathbf{v}$.
- Compatibility of scalar multiplication: $\alpha_1 \left(\alpha_2 \mathbf{v} \right) = \left(\alpha_1 \alpha_2 \right) \mathbf{v}$.
- Multiplicative identity: $1\mathbf{v} = \mathbf{v}$.

Here we study several examples of vector spaces:

Example 1.3. For $V = \mathbb{F}^n$, we can define scalar multiplication:

$$\alpha \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} \alpha x_1 \\ \vdots \\ \alpha x_n \end{pmatrix}$$

and vector addition:

$$\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} + \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} x_1 + y_1 \\ \vdots \\ x_n + y_n \end{pmatrix}.$$

In such a case, $\mathbf{0} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}$ is the zero vector.

Example 1.4. It is clear that the set $V = M_{m \times n} \left(\mathbb{F} \right)$ (the set of all $m \times n$ real matrices) is a vector space as well. In such a case, the zero 'vector' is the zero matrix $\mathbf{0}_{m \times n}$.

Example 1.5. The set $V = \mathcal{C}(\mathbb{R})$ is a vector space with scalar multiplication:

$$(\alpha f)(x) = \alpha f(x), \forall \alpha \in \mathbb{R}, f \in V$$

and vector addition:

$$(f + g)(x) = f(x) + g(x), \forall f, g \in V$$

Here the zero vector is the zero function, i.e., $\mathbf{0}(x) = 0$ for all $x \in \mathbb{R}$.

Definition 1.6. A sub-collection $W \subseteq V$ of a vector space V is called a **vector subspace** of V if W itself forms a vector space, denoted by $W \leq V$.

Example 1.7. For $V = \mathbb{R}^3$, $W := \{(x, y, 0) \mid x, y \in \mathbb{R}\} \leq V$, but $W' := \{(x, y, 1) \mid x, y \in \mathbb{R}\}$ is not a vector subspace of V.

Proposition 1.8. *A subset $W \subseteq V$ is a vector subspace of V iff for all $\mathbf{w}_1, \mathbf{w}_2 \in W$, we have*

$$\alpha \mathbf{w}_1 + \beta \mathbf{w}_2 \in W$$

for all $\alpha, \beta \in \mathbb{F}$.

The proof is identical to that of MAT 2040.

Example 1.9. We look at a few more examples and non-examples of vector subspaces:

- Let $V = M_{n \times n}(\mathbb{F})$, then the subset of all symmetric matrices $W = \{A \in V \mid A^t = A\} \leq V$ is a subspace of V.

- Let $V = C^\infty(\mathbb{R})$. Define $W = \left\{f \in V \mid \frac{\mathrm{d}^2}{\mathrm{d}x^2}f + f = 0\right\} \leq V$. For $f, g \in W$, we have

$$(\alpha f + \beta g)'' = \alpha f'' + \beta g'' = \alpha(-f) + \beta(-g) = -(\alpha f + \beta g),$$

  which implies $(\alpha f + \beta g)'' + (\alpha f + \beta g) = 0$. Therefore, $W \leq V$.

- Let $V = \mathbb{R}^2$. Then the set $\mathbb{R}_+^2 := \{(x, y) \in \mathbb{R}^2 \mid x, y > 0\}$ is not a vector subspace, since W is not closed under scalar multiplication.

- Moreover, the set $\mathbb{R}_+^2 \cup \mathbb{R}_-^2 := \{(x, y) \in \mathbb{R}^2 \mid xy > 0\}$ is not a vector subspace of $V = \mathbb{R}^2$ since it is not closed under addition.

- For $V = \mathbb{M}_{3 \times 3}(\mathbb{R})$, the set of invertible 3×3 matrices is not a vector subspace, since the zero 'vector' $\mathbf{0}_{3 \times 3}$ is not invertible (Exercise: How about the set of all singular matrices?).

1.2.1 Spanning Set and Linear Independence

Definition 1.10. Let V be a vector space over $\mathbb{F}$:

1. A **linear combination** of a subset S in V is of the form

$$\sum_{i=1}^{n} \alpha_i \mathbf{s}_i, \quad \alpha_i \in \mathbb{F}, \mathbf{s}_i \in S$$

for $n \in \mathbb{N}$. Note that the summation should be finite.

2. The **span** of a subset $S \subseteq V$ is

$$\mathrm{span}\,(S) = \left\{ \sum_{i=1}^{n} \alpha_i \mathbf{s}_i \,\middle|\, \alpha_i \in \mathbb{F}, \mathbf{s}_i \in S \right\}$$

3. S is a **spanning set** of V, or say S **spans** V, if

$$\mathrm{span}\,(S) = V \ .$$

Example 1.11.

1. For $V = \mathbb{R}\,[x]$, define the set $S := \{1, x^2, x^4, \ldots, x^6\} \subseteq V$. Then $2 + x^4 + \pi x^{106} \in \mathrm{span}\,(S)$, while the series $1 + x^2 + x^4 + \cdots \notin \mathrm{span}\,(S)$.

   It is clear that $\mathrm{span}\,(S) \neq V$, but S is the spanning set of $W = \{p \in V \mid p\,(x) = p\,(-x)\}$.

2. For $V = M_{3\times3}\,(\mathbb{R})$, let $\mathbf{w}_1 = \left\{ A \in V \mid A^{\mathrm{T}} = A \right\}$ and $\mathbf{w}_2 = \left\{ B \in V \mid B^{\mathrm{T}} = -B \right\}$ (the set of skew-symmetric matrices) be two vector subspaces. **Exercise:** show that the set $S := \mathbf{w}_1 \cup \mathbf{w}_2$ spans V.

Proposition 1.12. *Let S be a subset in a vector space V.*

1. *$S \subseteq \mathrm{span}\,(S)$*

2. *$\mathrm{span}\,(S) = \mathrm{span}\,(\mathrm{span}\,(S))$*

3. *(Exchange Lemma) If $\mathbf{w} \in \mathrm{span}\,\{\mathbf{v}_1, \ldots, \mathbf{v}_n\} \smallsetminus \mathrm{span}\,\{\mathbf{v}_2, \ldots, \mathbf{v}_n\}$, then*

$$\mathbf{v}_1 \in \mathrm{span}\,\{\mathbf{w}, \mathbf{v}_2, \ldots, \mathbf{v}_n\} \smallsetminus \mathrm{span}\,\{\mathbf{v}_2, \ldots, \mathbf{v}_n\}$$

Proof. 1. For each $s \in S$, we have

$$\mathbf{s} = 1 \cdot \mathbf{s} \in \mathrm{span}\,(S)$$

2. From (1), it's clear that $\mathrm{span}\,(S) \subseteq \mathrm{span}\,(\mathrm{span}\,(S))$, and therefore suffices to show $\mathrm{span}\,(\mathrm{span}\,(S)) \subseteq \mathrm{span}\,(S)$:

   Pick $\mathbf{v} = \sum\limits_{i=1}^{n} \alpha_i \mathbf{v}_i \in \mathrm{span}\,(\mathrm{span}\,(S))$, where $\mathbf{v}_i \in \mathrm{span}\,(S)$. Rewrite

$$\mathbf{v}_i = \sum_{j=1}^{n_i} \beta_{ij} \mathbf{s}_j, \ \mathbf{s}_j \in S,$$

   which implies

$$\mathbf{v} = \sum_{i=1}^{n} \alpha_i \sum_{j=1}^{n_i} \beta_{ij} \mathbf{s}_j = \sum_{i=1}^{n} \sum_{j=1}^{n_i} (\alpha_i \beta_{ij})\,\mathbf{s}_j,$$

i.e., $\mathbf{v}$ is a finite linear combination of elements in S, which implies $\mathbf{v} \in \text{span}\,(S)$.

3. By hypothesis, $\mathbf{w} = \alpha_1 \mathbf{v}_1 + \cdots + \alpha_n \mathbf{v}_n$ with $\alpha_1 \neq 0$, which implies

$$\mathbf{v}_1 = -\frac{\alpha_2}{\alpha_1}\mathbf{v}_2 + \cdots + \left(-\frac{1}{\alpha_1}\mathbf{w}\right)$$

which implies $\mathbf{v}_1 \in \text{span}\,\{\mathbf{w}, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$. It suffices to show $\mathbf{v}_1 \notin \text{span}\,\{\mathbf{v}_2, \ldots, \mathbf{v}_n\}$. Suppose on the contrary that $\mathbf{v}_1 \in \text{span}\,\{\mathbf{v}_2, \ldots, \mathbf{v}_n\}$. It's clear that $\text{span}\,\{\mathbf{v}_1, \ldots, \mathbf{v}_n\} = \text{span}\,\{\mathbf{v}_2, \ldots, \mathbf{v}_n\}$. (left as exercise). Therefore,

$$\varnothing = \text{span}\,\{\mathbf{v}_1, \ldots, \mathbf{v}_n\} \smallsetminus \text{span}\,\{\mathbf{v}_2, \ldots, \mathbf{v}_n\}\,,$$

which is a contradiction.

$\square$

Definition 1.13 (Linear Independence)**.** Let S be a (not necessarily finite) subset of V. Then S is **linearly independent** (l.i.) on V if for any finite subset $\{\mathbf{s}_1, \ldots, \mathbf{s}_k\}$ in S,

$$\sum_{i=1}^{k} \alpha_i \mathbf{s}_i = 0 \Leftrightarrow \alpha_i = 0 \; \forall i.$$

Otherwise, we say S is **linearly dependent**.

Example 1.14.

- For $V = C\,(\mathbb{R})$, $S_1 = \{\sin x, \cos x\}$ is linearly independent, since

$$\alpha \sin x + \beta \cos x = \mathbf{0} \text{ (means zero function)}$$

Taking $x = 0$ both sides leads to $\beta = 0$; taking $x = \frac{\pi}{2}$ both sides leads to $\alpha = 0$.

- For $V = C\,(\mathbb{R})$, $S_2 = \{\sin^2 x, \cos^2 x, 1\}$ is linearly dependent, since

$$1 \cdot \sin^2 x + 1 \cdot \cos^2 x + (-1) \cdot 1 = 0, \forall x$$

- For $V = \mathbb{R}\,[x]$, $S = \{1, x, x^2, x^3, \ldots,\}$ is linearly independent: Pick $x^{k_1}, \ldots, x^{k_n} \in S$ with $k_1 < \cdots < k_n$. Consider that the equation

$$\alpha_1 x^{k_1} + \cdots + \alpha_n x^{k_n} = \mathbf{0}$$

holds for all x, and try to solve for $\alpha_1, \ldots, \alpha_n$ (one way is differentation.)

Definition 1.15 (Basis)**.** A subset S is a **basis** of V if

(a) S spans V; and

(b) S is linearly independent.

Example 1.16. 1. For $V = \mathbb{R}^n, S = \{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ is a basis of V

    2. For $V = \mathbb{R}[x], S = \{1, x, x^2, \ldots\}$ is a basis of V

    3. For $V = M_{2\times 2}(\mathbb{R})$,

$$S = \left\{ \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \right\}$$

is a basis of V.

Proposition 1.17. *Let $V = \mathrm{span}\{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$. Then there exists a subset of $\{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$ which is a basis of V.*

Proof. If $\{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$ is linearly independent, the proof is complete. Otherwise, $\alpha_1 \mathbf{v}_1 + \cdots + \alpha_m \mathbf{v}_m = \mathbf{0}$ has a non-trivial solution. w.l.o.g., $\alpha_1 \neq 0$, which implies

$$\mathbf{v}_1 = \left(-\frac{\alpha_2}{\alpha_1}\right) \mathbf{v}_2 + \cdots + \left(-\frac{\alpha_m}{\alpha_1}\right) \mathbf{v}_m \Rightarrow \mathbf{v}_1 \in \mathrm{span}\{\mathbf{v}_2, \ldots, \mathbf{v}_m\}$$

By the proof in Proposition 1.12 (3),

$$V = \mathrm{span}\{\mathbf{v}_1, \ldots, \mathbf{v}_m\} = \mathrm{span}\{\mathbf{v}_2, \ldots, \mathbf{v}_m\},$$

which implies $V = \mathrm{span}\{\mathbf{v}_2, \ldots, \mathbf{v}_m\}$. So one can continue this argument finitely many times until one has $\{\mathbf{v}_i, \mathbf{v}_{i+1}, \ldots, \mathbf{v}_m\}$ is linearly independent, and spans V. The proof is complete. $\square$

Corollary 1.18. *Suppose V is* **finitely-generated***, i.e. there exists a finite set $\{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$ such that $V = \mathrm{span}\{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$. Then V has a basis.*

The same conclusion holds for non-finitely generated V if one assumes Zorn's lemma.

Proposition 1.19. *If $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is a basis of V, then every $v \in V$ can be expressed uniquely as*

$$\mathbf{v} = \alpha_1 \mathbf{v}_1 + \cdots + \alpha_n \mathbf{v}_n$$

Proof. Since $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ spans V, so $\mathbf{v} \in V$ can be written as in the form given in the Proposition. Suppose further that

$$\mathbf{v} = \beta_1 \mathbf{v}_1 + \cdots + \beta_n \mathbf{v}_n,$$

for some $\beta_i \in \mathbb{F}$, it suffices to show that $\alpha_i = \beta_i$ for all i: Indeed, by subtracting the above equations, one has

$$(\alpha_1 - \beta_1) \mathbf{v}_1 + \cdots + (\alpha_n - \beta_n) \mathbf{v}_n = 0.$$

By the hypothesis of linear independence, we have $\alpha_i - \beta_i = 0$ for all i, i.e., $\alpha_i = \beta_i$. $\square$

1.2.2 Basis and Dimension

We begin by showing that the bases of any finitely generated vector space have the same number of vectors:

Theorem 1.20. *Let V be a finitely generated vector space (so that V has a basis by Corollary 1.18. Suppose $\{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$ and $\{\mathbf{w}_1, \ldots, \mathbf{w}_n\}$ are two bases of V. Then $m = n$.*

Proof. Suppose on the contrary that $m \neq n$. Without loss of generality, assume that $m < n$. Let

$$\mathbf{v}_1 = \alpha_1 \mathbf{w}_1 + \cdots + \alpha_n \mathbf{w}_n$$

with some $\alpha_i \neq 0$. By rearranging the terms if necessary, assume $\alpha_1 \neq 0$. Then

$$\mathbf{v}_1 \in \operatorname{span}\{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_n\} \setminus \operatorname{span}\{\mathbf{w}_2, \ldots, \mathbf{w}_n\}. \tag{1.1}$$

This implies that

$$\mathbf{w}_1 \in \operatorname{span}\{\mathbf{v}_1, \mathbf{w}_2, \ldots, \mathbf{w}_n\} \setminus \operatorname{span}\{\mathbf{w}_2, \ldots, \mathbf{w}_n\}.$$

We now claim that $\{\mathbf{v}_1, \mathbf{w}_2, \ldots, \mathbf{w}_n\}$ is a basis of V.

1. **Spanning Set:** Firstly, note that $\mathbf{w}_1 \in \operatorname{span}\{\mathbf{v}_1, \mathbf{w}_2, \ldots, \mathbf{w}_n\}$, so

$$\operatorname{span}\{\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_n\} \subseteq \operatorname{span}\{\mathbf{v}_1, \mathbf{w}_2, \ldots, \mathbf{w}_n\}.$$

Since $\operatorname{span}\{\mathbf{w}_1, \ldots, \mathbf{w}_n\} = V$, we have $\operatorname{span}\{\mathbf{v}_1, \mathbf{w}_2, \ldots, \mathbf{w}_n\} = V$.

2. **Linear independence:** Suppose

$$\beta_1 \mathbf{v}_1 + \beta_2 \mathbf{w}_2 + \cdots + \beta_n \mathbf{w}_n = \mathbf{0}.$$

- If $\beta_1 \neq 0$, then
$$\mathbf{v}_1 = -\frac{\beta_2}{\beta_1} \mathbf{w}_2 - \cdots - \frac{\beta_n}{\beta_1} \mathbf{w}_n \in \operatorname{span}\{\mathbf{w}_2, \ldots, \mathbf{w}_n\},$$
contradicting (1.1).
- If $\beta_1 = 0$, then independence of $\{\mathbf{w}_2, \ldots, \mathbf{w}_n\}$ implies $\beta_2 = \cdots = \beta_n = 0$.

Thus $\{\mathbf{v}_1, \mathbf{w}_2, \ldots, \mathbf{w}_n\}$ is a basis. By repeating this replacement process, we ultimately construct

$$\mathcal{B} := \{\mathbf{v}_1, \ldots, \mathbf{v}_m, \mathbf{w}_{m+1}, \ldots, \mathbf{w}_n\}$$

as a basis of V. But then $\mathbf{w}_{m+1}$ would be a linear combination of $\mathbf{v}_1, \ldots, \mathbf{v}_m$, contradicting the linear independence the basis $\mathcal{B}$. Hence, $m = n$. $\qquad\square$

Definition 1.21 (Dimension)**.** Let V be a vector space over $\mathbb{F}$.

- If V is finitely-generated, then the **dimension** of V is defined as $\dim(V) = n$, where n the size of *any* basis of V (c.f. Theorem 1.20).

- If V is not finitely-generated, then the dimension of V is infinite, i.e. $\dim(V) = \infty$.

Example 1.22. A vector space may have more than one basis. For instance, suppose $V = \mathbb{F}^n$. Then clearly $\dim(V) = n$, and

$$\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$$

is a basis of V, where $\mathbf{e}_i$ denotes the standard unit vector. Also, another basis of V can be taken as:

$$\left\{ \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \ldots, \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \right\}.$$

In fact, the columns of any invertible $n \times n$ matrix form a basis of V.

Example 1.23. Let $V = M_{m \times n}(\mathbb{R})$. Then $\dim(V) = mn$. More explicitly, a basis of V is given by the set

$$\{E_{ij} \mid 1 \leq i \leq m, \ 1 \leq j \leq n\},$$

where E_{ij} is the $m \times n$ matrix with a 1 in the (i, j)-th entry and 0s elsewhere.

Example 1.24. Let V be the space of all real polynomials of degree $\leq n$. Then

$$\dim(V) = n + 1.$$

A basis of V is given by $\{1, x, x^2, \ldots, x^n\}$.

Example 1.25. Let $V = \left\{ A \in M_{n \times n}(\mathbb{R}) \mid A^{\mathrm{T}} = A \right\}$, the space of real symmetric matrices. Then

$$\dim(V) = \frac{n(n + 1)}{2}.$$

Example 1.26. Let $W = \left\{ B \in M_{n \times n}(\mathbb{R}) \mid B^{\mathrm{T}} = -B \right\}$, the space of real skew-symmetric matrices. Then

$$\dim(W) = \frac{n(n - 1)}{2}.$$

Example 1.27. Sometimes we need to specify the field $\mathbb{F}$ for the scalar multiplication when we define a vector space. Consider the example below:

1. Let $V = \mathbb{C}$, then $\dim_{\mathbb{C}}(\mathbb{C}) = 1$ for as a vector space over the field $\mathbb{F} = \mathbb{C}$.

2. Let $V = \mathrm{span}_{\mathbb{R}}\{1, i\} = \mathbb{C}$, then $\dim_{\mathbb{R}}(\mathbb{C}) = 2$ as a vector space over $\mathbb{F} = \mathbb{R}$, since all $z \in V$ can be uniquely written as $z = a + bi$, for some $a, b \in \mathbb{R}$. In cases where there may be confusion, it is safer to write

$$\dim_{\mathbb{C}}(\mathbb{C}) = 1, \ \dim_{\mathbb{R}}(\mathbb{C}) = 2.$$

Note that a basis for a vector space is characterized as the maximal linearly independent set.

Theorem 1.28 (Basis Extension)**.** *Let V be a finite-dimensional vector space, and let $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ be a linearly independent subset of V. Then we can extend it to a basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_k, \mathbf{v}_{k+1}, \ldots, \mathbf{v}_n\}$ of V.*

Proof. Suppose $\dim(V) = n > k$, and let $\{\mathbf{w}_1, \ldots, \mathbf{w}_n\}$ be a basis of V. Consider the set

$$\{\mathbf{w}_1, \ldots, \mathbf{w}_n\} \cup \{\mathbf{v}_1, \ldots, \mathbf{v}_k\},$$

which spans V but is linearly dependent, i.e., there exists non-trivial solutions of the linear equation

$$\alpha_1 \mathbf{w}_1 + \cdots + \alpha_n \mathbf{w}_n + \beta_1 \mathbf{v}_1 + \cdots + \beta_k \mathbf{v}_k = \mathbf{0}.$$

Note that there must be some $\alpha_i \neq 0$, otherwise the above equation will only involve the linearly independent set $\mathbf{v}_1, \ldots, \mathbf{v}_k$, which forces $\beta_1 = \cdots = \beta_k = 0$.

Without loss of generality, assume $\alpha_1 \neq 0$, so that $\mathbf{w}_1 \in \text{span}(\{\mathbf{w}_2, \ldots, \mathbf{w}_n\} \cup \{\mathbf{v}_1 \ldots, \mathbf{v}_k\})$. Consider the set

$$\{\mathbf{w}_2, \ldots, \mathbf{w}_n\} \cup \{\mathbf{v}_1, \ldots, \mathbf{v}_k\}.$$

with $\mathbf{w}_1$ removed. As in the proof of spanning set in Theorem 1.20, one has

$$\text{span}(\{\mathbf{w}_2, \ldots, \mathbf{w}_n\} \cup \{\mathbf{v}_1, \ldots, \mathbf{v}_k\}) = \text{span}(\{\mathbf{w}_1, \ldots, \mathbf{w}_n\} \cup \{\mathbf{v}_1, \ldots, \mathbf{v}_k\}) = V.$$

In other words, it is still a spanning set. If it is linearly dependent, we repeat the argument by removing $\mathbf{w}_2$ and get a smaller spanning set

$$\{\mathbf{w}_3, \ldots, \mathbf{w}_n\} \cup \{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$$

Continue this process by removing more vectors $\mathbf{w}_3, \ldots$, so that the remaining set still spans V, until we hit the first occasion when the set

$$\{\mathbf{w}_\ell, \ldots, \mathbf{w}_m\} \cup \{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$$

is linearly independent and spans V. Then this becomes a basis of V. Moreover, one has $\ell = k + 1$ by Theorem 1.20. $\qquad\square$

1.2.3 Intersection and Sum of Vector Subspaces

Definition 1.29. Let W_1, W_2 be two vector subspaces of V. Then:

1. $W_1 \cap W_2 := \{\mathbf{w} \in V \mid \mathbf{w} \in W_1 \text{ and } \mathbf{w} \in W_2\}$

2. $W_1 + W_2 := \{\mathbf{w}_1 + \mathbf{w}_2 \mid \mathbf{w}_i \in W_i \text{ for } i = 1, 2\}$

3. Furthermore, if $W_1 \cap W_2 = \{\mathbf{0}\}$, then $W_1 + W_2 = W_1 \oplus W_2$ is called the *direct sum* of W_1 and W_2.

Proposition 1.30. $W_1 \cap W_2$ and $W_1 + W_2$ are vector subspaces of V.

Proof. Exercise (use Proposition 1.8). $\qquad\square$

Proposition 1.31. *Let* W_1, W_2 *be subspaces of a vector space* V. *The following are equivalent (TFAE):*

1. $W_1 + W_2 = W_1 \oplus W_2$ *is a direct sum*

2. *Every* $\mathbf{w} \in W_1 + W_2$ *can be uniquely written as* $\mathbf{w} = \mathbf{w}_1 + \mathbf{w}_2$ *for* $\mathbf{w}_1 \in W_1$ *and* $\mathbf{w}_2 \in W_2$.

3. *For* $\mathbf{w}_1 \in W_1$ *and* $\mathbf{w}_2 \in W_2$,

$$\mathbf{w}_1 + \mathbf{w}_2 = \mathbf{0} \quad \Leftrightarrow \quad \mathbf{w}_1 = \mathbf{w}_2 = \mathbf{0}.$$

Proof. XXX $\qquad\square$

Obviously, one can generalize Definition 1.29 to any number of (even for uncountably many) vector subspaces:

Definition 1.32. Let $\{W_i \mid i \in I\}$ be a collection of vector subspaces of V. Then:

1. $\bigcap_{i \in I} W_i := \{\mathbf{w} \in V \mid \mathbf{w} \in W_i \text{ for all } i \in I\}$

2. $\sum_{i \in I} W_i := \{\mathbf{w}_{i_1} + \cdots + \mathbf{w}_{i_n} \mid i_k \in I \text{ for all } k, \ n \in \mathbb{N}\}$ (Note that the sum in the definition above must be finite).

Here is a generalization of direct sum of more than two subspaces:

Definition 1.33. Let $\{W_i \mid i \in I\}$ be a collection of subspaces of a vector space V. We write

$$\bigoplus_{i \in I} W_i = \sum_{i \in I} W_i$$

as the **direct sum** of all W_i's if for **any** finite subset $\{i_1, \ldots, i_n\} \subseteq I$ and $\mathbf{w}_{i_k} \in W_{i_k}$,

$$\mathbf{w}_{i_1} + \cdots + \mathbf{w}_{i_n} = \mathbf{0} \quad \Leftrightarrow \quad \mathbf{w}_{i_1} = \cdots = \mathbf{w}_{i_n} = \mathbf{0},$$

As in the two vector subspaces case, every element $\mathbf{w} \in \bigoplus_{i \in I} W_i$ in the direct sum has a **unique** expression as a finite sum $\mathbf{w} = \mathbf{w}_{i_1} + \cdots + \mathbf{w}_{i_n}$ of vectors $\mathbf{w}_{i_k} \in W_{i_k}$.

Proposition 1.34 (Complementation)**.** *Let* $W \leq V$ *be a subspace of a finite-dimensional vector space* V. *Then there exists a subspace* $W' \leq V$ *such that*

$$V = W \oplus W',$$

i.e., every vector in V *can be uniquely written as a sum of vectors from* W *and* W'.

Proof. Let $\dim(W) = k \leq n = \dim(V)$, and let $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ be a basis of W. By the basis extension theorem, we can extend this to a basis of V:

$$\{\mathbf{v}_1, \ldots, \mathbf{v}_k, \mathbf{v}_{k+1}, \ldots, \mathbf{v}_n\}.$$

Define

$$W' := \mathrm{span}\{\mathbf{v}_{k+1}, \ldots, \mathbf{v}_n\}.$$

We now verify the two conditions for a direct sum:

1. **Sum:** For every $\mathbf{v} \in V$, we can write

$$\mathbf{v} = \sum_{i=1}^{k} \alpha_i \mathbf{v}_i + \sum_{j=k+1}^{n} \alpha_j \mathbf{v}_j,$$

where the first sum lies in W and the second in W', so $V = W + W'$.

2. **Intersection:** Suppose $\mathbf{v} \in W \cap W'$. Then

$$\mathbf{v} = \sum_{i=1}^{k} \beta_i \mathbf{v}_i = \sum_{j=k+1}^{n} \gamma_j \mathbf{v}_j.$$

But since the $\mathbf{v}_i$ form a basis of V, their representation is unique. Hence all coefficients must vanish:

$$\beta_1 = \cdots = \beta_k = \gamma_{k+1} = \cdots = \gamma_n = 0,$$

so $\mathbf{v} = \mathbf{0}$, and $W \cap W' = \{\mathbf{0}\}$.

Thus, $V = W \oplus W'$. $\qquad\square$

Chapter 2

Linear Transformations

In linear algebra, we will study a special kind of functions between vector spaces:

Definition 2.1 (Linear Transformation)**.** Let V and W be vector spaces over a field $\mathbb{F}$. A map $T : V \to W$ is called a *linear transformation* if for all $\alpha, \beta \in \mathbb{F}$ and all $\mathbf{v}_1, \mathbf{v}_2 \in V$, we have

$$T(\alpha \mathbf{v}_1 + \beta \mathbf{v}_2) = \alpha T(\mathbf{v}_1) + \beta T(\mathbf{v}_2).$$

Proposition 2.2. *Let V, W, U be vector spaces and let T, S be linear transformations. Then:*

1. *If $S : V \to W$ and $T : W \to U$, then the composition $T \circ S : V \to U$ is also a linear transformation.*

2. *For any linear transformation $T : V \to W$, we have*

$$T(\mathbf{0}_V) = \mathbf{0}_W.$$

Proof.   1. Let $\mathbf{v}_1, \mathbf{v}_2 \in V$ and $\alpha, \beta \in \mathbb{F}$. Then,

$$(T \circ S)(\alpha \mathbf{v}_1 + \beta \mathbf{v}_2) = T(S(\alpha \mathbf{v}_1 + \beta \mathbf{v}_2)) = T(\alpha S(\mathbf{v}_1) + \beta S(\mathbf{v}_2)) = \alpha T(S(\mathbf{v}_1)) + \beta T(S(\mathbf{v}_2)).$$

Hence $T \circ S$ is linear.

2. Since T is linear,
$$T(\mathbf{0}_V) = T(0 \cdot \mathbf{v}) = 0 \cdot T(\mathbf{v}) = \mathbf{0}_W$$

for any $\mathbf{v} \in V$.

$\square$

Example 2.3.

1. Let $A \in \mathbb{R}^{m \times n}$. The map $T : \mathbb{R}^n \to \mathbb{R}^m$ defined by

$$T(\mathbf{x}) = A\mathbf{x}$$

is a linear transformation.

2. Let $T : \mathbb{R}[x] \to \mathbb{R}[x]$ be defined by either of the following:

$$T(p(x)) = \frac{d}{dx} p(x), \quad \text{or} \quad T(p(x)) = \int_0^x p(t)\, dt.$$

Both are linear transformations.

3. The map $T : M_{n \times n}(\mathbb{R}) \to \mathbb{R}$ defined by

$$T(A) = \text{trace}(A) := \sum_{i=1}^{n} a_{ii}$$

is a linear transformation.

4. However, the map $A \mapsto \det(A)$ is *not* a linear transformation.

As we saw above, many of the functions we studied in other branches of mathematics are linear transformations. One of the many advantages of linear transformation over a general function is the following:

Remark 2.4. Let $T : V \to W$ be a linear transformation, and $\mathcal{B} = \{\mathbf{v}_i \mid i \in I\}$ be a basis of V, then

> **the image of T is uniquely determined by the values of $\{T(\mathbf{v}_i) \mid \mathbf{v}_i \in \mathcal{B}\}$**.

Namely, since every element $\mathbf{v} \in V$ can be expressed as $\mathbf{v} = \sum_{\ell=1}^{k} \alpha_\ell \mathbf{v}_{i_\ell}$, then

$$T(\mathbf{v}) = T \left(\sum_{\ell=1}^{k} \alpha_\ell \mathbf{v}_{i_\ell} \right) = \sum_{\ell=1}^{k} \alpha_\ell T\left(\mathbf{v}_{i_\ell}\right). \tag{2.1}$$

This is certainly not true for a general function - for instance, if I tell you $f : \mathbb{R} \to \mathbb{R}$ satisfies $f(1) = 1$ and $f(2) = 4$, one cannot determine $f(x)$ for all $x \in \mathbb{R}$.

Conversely, one can

> **Define linear transformation $T : V \to W$ by *just* specifying $\{T(\mathbf{v}_i) = \mathbf{w}_i \mid \mathbf{v}_i \in \mathcal{B}\}$.**

In such a case, the image $T(\mathbf{v})$ for all $\mathbf{v} \in V$ can be obtained by extending T linearly using Equation 2.1.

2.1   Kernel and Image

Definition 2.5 (Kernel and Image)**.** Let $T : V \to W$ be a linear transformation.

1. The **kernel** of T is

$$\ker(T) = T^{-1}(\mathbf{0}) = \{\mathbf{v} \in V \mid T(\mathbf{v}) = \mathbf{0}\}.$$

2. The **image** (or range) of T is

$$\text{Im}(T) = T(V) = \{T(\mathbf{v}) \in W \mid \mathbf{v} \in V\}.$$

Example 2.6.

1. Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be a linear transformation defined by $T(\mathbf{x}) = A\mathbf{x}$, where $A \in \mathbb{R}^{n \times n}$. Then

$$\ker(T) = \{\mathbf{x} \in \mathbb{R}^n \mid A\mathbf{x} = \mathbf{0}\} = \text{Null}(A)$$

and

$$\text{Im}(T) = \{A\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^n\} = \text{Col}(A) = \text{span}\{\text{columns of } A\} \quad \text{(Column Space)}.$$

2. For $T(p(x)) = \frac{d}{dx}p(x)$, we have

$$\ker(T) = \{\text{constant polynomials}\}, \quad \text{Im}(T) = \mathbb{R}[x].$$

Proposition 2.7. *The kernel and image of a linear transformation $T : V \to W$ are vector subspaces:*

$$\ker(T) \leq V, \quad \text{Im}(T) \leq W.$$

Proof. Let $\mathbf{v}_1, \mathbf{v}_2 \in \ker(T)$. Then for any $\alpha, \beta \in \mathbb{F}$, we have

$$T(\alpha\mathbf{v}_1 + \beta\mathbf{v}_2) = \alpha T(\mathbf{v}_1) + \beta T(\mathbf{v}_2) = \alpha \cdot \mathbf{0} + \beta \cdot \mathbf{0} = \mathbf{0},$$

so $\alpha\mathbf{v}_1 + \beta\mathbf{v}_2 \in \ker(T)$. The case for $\text{Im}(T)$ is similar. $\square$

Definition 2.8 (Rank/Nullity)**.** Let V, W be finite dimensional vector spaces over a field $\mathbb{F}$ and $T : V \to W$ a linear transformation. Then we define

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

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

Proposition 2.9. *There are alternative characterizations for the injectivity and surjectivity of a linear transformation T:*

1. *The linear transformation T is injective if and only if*

$$\ker(T) = \{\mathbf{0}\} \quad \Leftrightarrow \quad \text{nullity}(\mathbf{T}) = \mathbf{0}.$$

2. *The linear transformation T is surjective if and only if*

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

3. *If T is bijective, then T^{-1} is a linear transformation.*

Proof. (1) (a) For the forward direction:

$$\mathbf{x} \in \ker(T) \Rightarrow T(\mathbf{x}) = 0 = T(\mathbf{0}) \Rightarrow \mathbf{x} = \mathbf{0}.$$

(b) For the reverse direction:

$$T(\mathbf{x}) = T(\mathbf{y}) \Rightarrow T(\mathbf{x} - \mathbf{y}) = 0 \Rightarrow \mathbf{x} - \mathbf{y} \in \ker(T) = \{0\} \Rightarrow \mathbf{x} = \mathbf{y}.$$

(2) The proof follows a similar idea as in (1).

(3) Let $T^{-1} : W \to V$. For all $\mathbf{w}_1, \mathbf{w}_2 \in W$, there exist $\mathbf{v}_1, \mathbf{v}_2 \in V$ such that $T(\mathbf{v}_i) = \mathbf{w}_i$, i.e.,

$$T^{-1}(\mathbf{w}_i) = \mathbf{v}_i, \quad i = 1, 2.$$

Consider the mapping:

$$T(\alpha \mathbf{v}_1 + \beta \mathbf{v}_2) = \alpha T(\mathbf{v}_1) + \beta T(\mathbf{v}_2) = \alpha \mathbf{w}_1 + \beta \mathbf{w}_2,$$

which implies

$$\alpha \mathbf{v}_1 + \beta \mathbf{v}_2 = T^{-1}(\alpha \mathbf{w}_1 + \beta \mathbf{w}_2),$$

i.e.,

$$\alpha T^{-1}(\mathbf{w}_1) + \beta T^{-1}(\mathbf{w}_2) = T^{-1}(\alpha \mathbf{w}_1 + \beta \mathbf{w}_2).$$

$\square$

2.2   Space of Linear Transformations

Definition 2.10 (Hom space)**.** Let

$$\mathrm{Hom}_{\mathbb{F}}(V, W) = \{\text{all linear transformations } T : V \to W\},$$

and we can define the addition and scalar multiplication to make it a vector space:

1. For $T, S \in \mathrm{Hom}_{\mathbb{F}}(V, W)$, define

$$(T + S)(\mathbf{v}) := T(\mathbf{v}) + S(\mathbf{v}),$$

one can check easily that $(T + S)(\alpha \mathbf{v} + \beta \mathbf{w}) = \alpha(T + S)(\mathbf{v}) + \beta(T + S)(\mathbf{w})$. Therefore, $T + S \in \mathrm{Hom}_{\mathbb{F}}(V, W)$.

2. Also, define

$$(\gamma T)(\mathbf{v}) := \gamma T(\mathbf{v}), \quad \text{for all } \gamma \in \mathbb{F},$$

as before, one can easily check that $\gamma T \in \mathrm{Hom}_{\mathbb{F}}(V, W)$.

Therefore, $T + S$, γT are the addition and scalar multiplication operations on the vector space $\mathrm{Hom}_{\mathbb{F}}(V, W)$.

In particular, if $V = \mathbb{F}^n, W = \mathbb{F}^m$, then by MAT2040, all linear transformations between V and

W are just matrix multiplications, i.e.:

$$\text{Hom}_{\mathbb{F}}(V, W) = \{T : V \to W \mid T(\mathbf{v}) = A\mathbf{v} \text{ for some } A \in M_{m \times n}(\mathbb{F})\}.$$

So

$$\text{Hom}_{\mathbb{F}}(V, W) \overset{1:1}{\longleftrightarrow} M_{m \times n}(\mathbb{F}).$$

In fact, the relationship between them is more than just a bijection - as we will see later, it is an **isomorphism** (Definition 2.12) between vector spaces!

Proposition 2.11. *If* $\dim(V) = n, \dim(W) = m$, *then* $\dim\left(\text{Hom}_{\mathbb{F}}(V, W)\right) = mn.$

2.3   Isomorphism of Vector Spaces

Definition 2.12. [Isomorphism] We say that the vector subspaces

$$V \cong W$$

are **isomorphic** if there exists a bijective linear transformation $T : V \to W$. This mapping T is called an *isomorphism* from V to W.

Example 2.13. Let $\dim(V) = n$ and $\dim(W) = m$ with $n, m < \infty$. Then $n = m$ implies $V \cong W$. More precisely, suppose $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}, \{\mathbf{w}_1, \ldots, \mathbf{w}_n\}$ are bases of V and W respectively, define $T : V \to W$ by:

$$T(\alpha_1 \mathbf{v}_1 + \cdots + \alpha_n \mathbf{v}_n) = \alpha_1 \mathbf{w}_1 + \cdots + \alpha_n \mathbf{w}_n, \quad \forall \alpha_i \in \mathbb{F}.$$

In particular, $T(\mathbf{v}_i) = \mathbf{w}_i$ for all i, and it is clear that our constructed T is a bijective linear transformation.

Note that the converse also holds, i.e. if $V \cong W$ then $n = m$. This is given by the proposition below.

Proposition 2.14. *If* $T : V \to W$ *is an isomorphism, then:*

1. *The set* $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ *is linearly independent in* V *if and only if* $\{T\mathbf{v}_1, \ldots, T\mathbf{v}_k\}$ *is linearly independent in* W. *Note the same holds if we replace "linearly independent" with "spans."*

2. *If* $\dim(V) = n$, *then* $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ *forms a basis of* V *if and only if* $\{T\mathbf{v}_1, \ldots, T\mathbf{v}_n\}$ *forms a basis of* W. *In particular,* $\dim(V) = \dim(W)$.

3. *Two vector spaces with finite dimensions are isomorphic if and only if they have the same dimension.*

Proof. Exercise. □

Remark 2.15. Note that $V \cong W$ does not imply that any linear transformation $T : V \to W$ is an isomorphism. For example, $T(\mathbf{v}) = \mathbf{0}$ is not an isomorphism if $W \neq \{\mathbf{0}\}$.

Theorem 2.16 (Rank–Nullity Theorem)**.** *Let $T : V \to W$ be a linear transformation with* $\dim(V) < \infty$. *Then*

$$\text{rank}(T) + \text{nullity}(T) = \dim(V).$$

Proof. Since $\ker(T) \leq V$, by Proposition 1.34, there exists a subspace $U \leq V$ such that

$$V = \ker(T) \oplus U.$$

1. Consider the restricted map $T|_U : U \to T(U)$. This is an isomorphism because:

   - It is surjective by definition of the codomain $T(U)$.

   - If $\mathbf{v} \in \ker(T|_U)$, then $T(\mathbf{v}) = 0$, so $\mathbf{v} \in \ker(T)$. But $\mathbf{v} \in U$ as well, and since $U \cap \ker(T) = \{\mathbf{0}\}$, we conclude $\mathbf{v} = \mathbf{0}$.

   Therefore, $\dim(U) = \dim(T(U))$ by the end of Example 2.13.

2. Note that $\text{im}(T) = T(V) = T(U)$, because for any $v \in V$, we can write $\mathbf{v} = \mathbf{v}_0 + \mathbf{u}$ with $\mathbf{v}_0 \in \ker(T), \mathbf{u} \in U$, and then

$$T(\mathbf{v}) = T(\mathbf{v}_0 + \mathbf{u}) = T(\mathbf{v}_0) + T(\mathbf{u}) = \mathbf{0} + T(\mathbf{u}) = T(\mathbf{u}).$$

   Hence $T(V) \subseteq T(U)$. But obviously one has $U \subseteq V \Rightarrow T(U) \subseteq T(V)$ as well. So $T(V) = T(U)$.

3. Since $V = \ker(T) \oplus U$, we have:

$$\dim(V) = \dim(\ker(T)) + \dim(U).$$

   by the proof of Proposition 1.34.

Therefore,

$$\dim(V) = \text{nullity}(T) + \dim(T(U)) = \text{nullity}(T) + \dim(T(V)) = \text{nullity}(T) + \text{rank}(T). \qquad \square$$

Chapter 3

Change of Basis

3.1 Change of Basis and Matrix Representation

3.1.1 Coordinate Vector

Definition 3.1 (Coordinate Vector). Let V be a finite-dimensional vector space and $\mathcal{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ an ordered basis of V. Any vector $\mathbf{v} \in V$ can be uniquely written as

$$\mathbf{v} = \alpha_1 \mathbf{v}_1 + \cdots + \alpha_n \mathbf{v}_n.$$

We define the map $[\cdot]_{\mathcal{B}} : V \to \mathbb{F}^n$, which maps any vector $\mathbf{v}$ to its coordinate vector:

$$[\mathbf{v}]_{\mathcal{B}} = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix}.$$

Note that $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ and $\{\mathbf{v}_2, \mathbf{v}_1, \ldots, \mathbf{v}_n\}$ are distinct ordered bases.

Example 3.2. Given $V = M_{2\times 2}(\mathbb{F})$ and the ordered basis

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

any matrix has a coordinate vector with respect to $\mathcal{B}$, e.g.,

$$\left[\begin{pmatrix} 1 & 4 \\ 2 & 3 \end{pmatrix} \right]_{\mathcal{B}} = \begin{pmatrix} 1 \\ 4 \\ 2 \\ 3 \end{pmatrix}.$$

However, if given another ordered basis

$$\mathcal{B}_1 = \left\{ \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \right\},$$

the matrix may have a different coordinate vector:

$$\left[\begin{pmatrix} 1 & 4 \\ 2 & 3 \end{pmatrix} \right]_{\mathcal{B}_1} = \begin{pmatrix} 4 \\ 1 \\ 2 \\ 3 \end{pmatrix}.$$

Theorem 3.3. *The mapping* $[\cdot]_{\mathcal{B}} : V \to \mathbb{F}^n$ *is an isomorphism.*

Proof. We prove this in several steps:

1. **Well-definedness.** Let $\mathbf{v} = \alpha_1 \mathbf{v}_1 + \cdots + \alpha_n \mathbf{v}_n$. and $\mathbf{v}' = \alpha'_1 \mathbf{v}_1 + \cdots + \alpha'_n \mathbf{v}_n$. Suppose $\mathbf{v} = \mathbf{v}'$, then by uniqueness of coordinates $\alpha_i = \alpha'_i$ for all $i = 1, \ldots, n$ and hence

$$[\mathbf{v}]_{\mathcal{B}} = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix} = \begin{pmatrix} \alpha'_1 \\ \vdots \\ \alpha'_n \end{pmatrix} = [\mathbf{v}']_{\mathcal{B}}.$$

2. **Linearity.** It is clear that the operator is a linear transformation:

$$[p\mathbf{v} + q\mathbf{w}]_{\mathcal{B}} = p[\mathbf{v}]_{\mathcal{B}} + q[\mathbf{w}]_{\mathcal{B}} \quad \forall p, q \in \mathbb{F}.$$

3. **Injectivity.** If

$$[\mathbf{v}]_{\mathcal{B}} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix},$$

   then $\mathbf{v} = 0\mathbf{v}_1 + \cdots + 0\mathbf{v}_n = \mathbf{0}$.

4. **Surjectivity.** Given any $\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{F}^n$, let

$$\mathbf{v} := x_1 \mathbf{v}_1 + \cdots + x_n \mathbf{v}_n.$$

   Then $[\mathbf{v}]_{\mathcal{B}} = \mathbf{x}$, so every element in $\mathbb{F}^n$ is hit.

   Therefore, $[\cdot]_{\mathcal{B}}$ is an isomorphism. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Example 3.4. Given a vector space $V = P_3[x]$ and its basis $B = \{1, x, x^2, x^3\}$.

To check if the set $\{1+x^2, 3-x^3, x-x^3\}$ is linearly independent, by part (1) in Proposition 2.14 and Theorem 3.3, it suffices to check whether the corresponding coordinate vectors

$$\left\{ \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 3 \\ 0 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ -1 \end{bmatrix} \right\}$$

are linearly independent, i.e., do Gaussian Elimination and check the number of pivots.

Remark 3.5. Here gives rise to the question: if $\mathcal{B}_1, \mathcal{B}_2$ form two bases of V, then how are $[\mathbf{v}]_{\mathcal{B}_1}, [\mathbf{v}]_{\mathcal{B}_2}$ related to each other?

Here we consider an easy example first:

Example 3.6. Consider $V = \mathbb{R}^n$ and its basis $\mathcal{B}_1 = \{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$. For any $\mathbf{v} \in V$,

$$\mathbf{v} = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix} = \alpha_1 \mathbf{e}_1 + \cdots + \alpha_n \mathbf{e}_n \Rightarrow [\mathbf{v}]_{\mathcal{B}_1} = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix}$$

Also, we can construct a different basis $\mathcal{B}_2$ of V as:

$$\mathcal{B}_2 = \left\{ \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \ldots, \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \right\}$$

which gives a different coordinate vector of $\mathbf{v}$:

$$[\mathbf{v}]_{B_2} = \begin{bmatrix} \alpha_1 - \alpha_2 \\ \alpha_2 - \alpha_3 \\ \vdots \\ \alpha_{n-1} - \alpha_n \\ \alpha_n \end{bmatrix}$$

Proposition 3.7 (Change of Basis). *Let $\mathcal{A} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ and $\mathcal{A}' = \{\mathbf{w}_1, \ldots, \mathbf{w}_n\}$ be two ordered bases of a vector space V. Define the change of basis matrix from $\mathcal{A}$ to $\mathcal{A}'$, say $C_{\mathcal{A}',\mathcal{A}} := [\alpha_{ij}]$, where*

$$\mathbf{v}_j = \sum_{i=1}^n \alpha_{ij} \mathbf{w}_i.$$

Then for any vector $\mathbf{v} \in V$, the change of basis amounts to left-multiplying the change of basis matrix:

$$C_{\mathcal{A}',\mathcal{A}}[\mathbf{v}]_{\mathcal{A}} = [\mathbf{v}]_{\mathcal{A}'}. \tag{3.1}$$

Define matrix $C_{\mathcal{A},\mathcal{A}'} := [\beta_{ij}]$, where

$$\mathbf{w}_j = \sum_{i=1}^{n} \beta_{ij}\mathbf{v}_i.$$

Then $(C_{\mathcal{A},\mathcal{A}'})^{-1} = C_{\mathcal{A}',\mathcal{A}}$.

Proof. 1. **Validity check.** Show (3.1) holds for $\mathbf{v} = \mathbf{v}_j$, $j = 1, \ldots, n$:

$$\text{LHS of } (3.1) = C_{\mathcal{A}'\mathcal{A}}[\mathbf{v}_j]_{\mathcal{A}} = [\alpha_{ij}]\mathbf{e}_j = \begin{pmatrix} \alpha_{1j} \\ \vdots \\ \alpha_{nj} \end{pmatrix},$$

$$\text{RHS of } (3.1) = [\mathbf{v}_j]_{\mathcal{A}'} = \left[\sum_{i=1}^{n} \alpha_{ij}\mathbf{w}_i \right]_{\mathcal{A}'} = \begin{pmatrix} \alpha_{1j} \\ \vdots \\ \alpha_{nj} \end{pmatrix}.$$

So (3.1) holds for all basis vectors $\mathbf{v}_j$.

2. **Linearity extension.** Let $\mathbf{v} = \sum_{j=1}^{n} r_j\mathbf{v}_j$. Then:

$$\begin{aligned}
C_{\mathcal{A}',\mathcal{A}}[\mathbf{v}]_{\mathcal{A}} &= C_{\mathcal{A}',\mathcal{A}} \left[\sum_{j=1}^{n} r_j\mathbf{v}_j \right]_{\mathcal{A}} \\
&= C_{\mathcal{A}',\mathcal{A}} \sum_{j=1}^{n} r_j[\mathbf{v}_j]_{\mathcal{A}} \\
&= \sum_{j=1}^{n} r_j C_{\mathcal{A}',\mathcal{A}}[\mathbf{v}_j]_{\mathcal{A}} \\
&= \sum_{j=1}^{n} r_j[\mathbf{v}_j]_{\mathcal{A}'} \\
&= \left[\sum_{j=1}^{n} r_j\mathbf{v}_j \right]_{\mathcal{A}'} = [\mathbf{v}]_{\mathcal{A}'}
\end{aligned}$$

where the first and last line follow from the linearity of the coordinate map $[\cdot]_{\mathcal{A}}$ and $[\cdot]_{\mathcal{A}'}$, and the fourth equality follows from (1) above. Therefore, (3.1) is shown for all $\mathbf{v} \in V$.

3. **Matrix inverse.** We verify $C_{\mathcal{A}',\mathcal{A}}C_{\mathcal{A},\mathcal{A}'} = I_n$:

$$\begin{aligned}
\mathbf{v}_j &= \sum_{i=1}^{n} \alpha_{ij}\mathbf{w}_i = \sum_{i=1}^{n} \alpha_{ij} \left(\sum_{k=1}^{n} \beta_{ki}\mathbf{v}_k \right) = \sum_{i=1}^{n}\sum_{k=1}^{n} \alpha_{ij}\beta_{ki}\mathbf{v}_k \\
&= \sum_{k=1}^{n} \left(\sum_{i=1}^{n} \beta_{ki}\alpha_{ij} \right) \mathbf{v}_k.
\end{aligned}$$

By the uniqueness of coordinates, we imply

$$\left(\sum_{i=1}^{n} \beta_{ki}\alpha_{ij} \right) = \delta_{jk} := \begin{cases} 1, & j = k \\ 0, & j \neq k \end{cases}$$

By matrix multiplication, the (k, j)-th entry of $C_{\mathcal{A}',\mathcal{A}}C_{\mathcal{A},\mathcal{A}'}$ is

$$[C_{\mathcal{A}',\mathcal{A}}C_{\mathcal{A},\mathcal{A}'}]_{kj} = \left(\sum_{i=1}^{n} \beta_{ki}\alpha_{ij} \right) = \delta_{jk} \Rightarrow C_{\mathcal{A}',\mathcal{A}}C_{\mathcal{A},\mathcal{A}'} = I_n.$$

4. **Symmetry.** Similarly,

$$\mathbf{w}_j = \sum_{i=1}^{n} \beta_{ij}\mathbf{v}_i = \sum_{i=1}^{n} \beta_{ij} \left(\sum_{k=1}^{n} \alpha_{ki}\mathbf{w}_k \right) = \sum_{k=1}^{n} \left(\sum_{i=1}^{n} \alpha_{ki}\beta_{ij} \right) \mathbf{w}_k,$$

giving $\sum_{i=1}^{n} \alpha_{ki}\beta_{ij} = \delta_{kj}$, so

$$C_{\mathcal{A},\mathcal{A}'}C_{\mathcal{A}',\mathcal{A}} = I_n.$$

$\square$

Example 3.8. Back to Example 3.6, write $\mathcal{B}_1, \mathcal{B}_2$ as

$$\mathcal{B}_1 = \{\mathbf{e}_1, \ldots, \mathbf{e}_n\}, \quad \mathcal{B}_2 = \{\mathbf{w}_1, \ldots, \mathbf{w}_n\}$$

and therefore $\mathbf{w}_i = \mathbf{e}_1 + \cdots + \mathbf{e}_i$. The change of basis matrix is given by

$$C_{\mathcal{B}_1,\mathcal{B}_2} = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 0 & 1 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}.$$

For instance, take $\mathbf{v}$ in the example, then one has:

$$C_{\mathcal{B}_1,\mathcal{B}_2}[\mathbf{v}]_{\mathcal{B}_2} = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 0 & 1 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} \begin{pmatrix} \alpha_1 - \alpha_2 \\ \vdots \\ \alpha_{n-1} - \alpha_n \\ \alpha_n \end{pmatrix} = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix} = [\mathbf{v}]_{\mathcal{B}_1}$$

3.1.2 Matrix Representations

Definition 3.9. Let $T : V \to W$ be a linear transformation, and

$$\mathcal{A} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}, \quad \mathcal{B} = \{\mathbf{w}_1, \ldots, \mathbf{w}_m\}$$

be bases of V and W, respectively. The matrix representation of T with respect to (w.r.t.) $\mathcal{A}$ and $\mathcal{B}$ is defined as $(T)_{\mathcal{BA}} := (\alpha_{ij}) \in M_{m \times n}(\mathbb{F})$, where

$$T(\mathbf{v}_j) = \sum_{i=1}^{m} \alpha_{ij} \mathbf{w}_i$$

for $j = 1, \ldots, n$.

Example 3.10. In MAT2040, we studied *matrix transformations* $T : \mathbb{F}^n \to \mathbb{F}^m$ given by $\mathbf{x} \mapsto A\mathbf{x}$ for some $A \in M_{m \times n}(\mathbb{F})$. In this section, we want to put it in the perspective of linear transformations:

Let $A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$, and $T : \mathbb{R}^2 \to \mathbb{R}^2$ given by $T\mathbf{x} = A\mathbf{x}$, and $\mathcal{B} = \left\{ \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\}$, then the matrix representation of $T : \mathbb{R}^2_{\mathcal{B}} \to \mathbb{R}^2_{\mathcal{B}}$ can be calculated by:

$$T \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} -1 \\ -1 \end{pmatrix} = 0 \begin{pmatrix} 1 \\ -1 \end{pmatrix} + -1 \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

$$T \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 3 \\ 7 \end{pmatrix} = -2 \begin{pmatrix} 1 \\ -1 \end{pmatrix} + 5 \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

Therefore,

$$T_{\mathcal{B},\mathcal{B}} = \begin{pmatrix} 0 & -2 \\ -1 & 5 \end{pmatrix}$$

So we end up having a different matrix than A!

However, if we use the usual basis $\mathcal{E} = \{\mathbf{e}_1, \mathbf{e}_2\}$, then the matrix representation of $T : \mathbb{R}^2_{\mathcal{E}} \to \mathbb{R}^2_{\mathcal{E}}$ becomes:

$$T(\mathbf{e}_1) = \begin{pmatrix} 1 \\ 3 \end{pmatrix} = 1\{\mathbf{e}_1 + 3\mathbf{e}_2$$

$$T(\mathbf{e}_2) = \begin{pmatrix} 2 \\ 4 \end{pmatrix} = 2\mathbf{e}_1 + 4\mathbf{e}_2,$$

so that

$$T_{\mathcal{E},\mathcal{E}} = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} = A$$

and one can get back our original matrix A.

This generalized perspective is extremely useful in our latter studies - it allows us to understand matrix transformations using different bases (coordinate systems). In the special case when one chooses the usual basis, one can retrieve the original matrix (see section 3.2 below.

Here is an example of matrix representation of a non-matrix transformation:

Example 3.11. Let $V = \mathbb{P}_3[x]$ and $\mathcal{A} = \{1, x, x^2, x^3\}$.

Let $T : V \to V$ be defined as $p(x) \mapsto p'(x)$:

$$\begin{cases} T(1) = 0 \cdot 1 + 0 \cdot x + 0 \cdot x^2 + 0 \cdot x^3 \\ T(x) = 1 \cdot 1 + 0 \cdot x + 0 \cdot x^2 + 0 \cdot x^3 \\ T(x^2) = 0 \cdot 1 + 2 \cdot x + 0 \cdot x^2 + 0 \cdot x^3 \\ T(x^3) = 0 \cdot 1 + 0 \cdot x + 3 \cdot x^2 + 0 \cdot x^3 \end{cases}$$

We can define the change of basis matrix for the linear transformation $T : V_{\mathcal{A}} \to V_{\mathcal{A}}$ with respect to the bases $\mathcal{A}$ and $\mathcal{A}$ as:

$$T_{\mathcal{A},\mathcal{A}} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Suppose a different basis $\mathcal{A}' = \{x^3, x^2, x, 1\}$ for the output space of T, i.e., $T : V_{\mathcal{A}} \to V_{\mathcal{A}'}$. Then,

$$T_{\mathcal{A}',\mathcal{A}} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 2 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$

for instance, since $T(x^2) = 0 \cdot x^3 + 0 \cdot x^2 + 2 \cdot x + 0 \cdot 1$ (written in the order of $\mathcal{A}'$, the third column

of $T_{\mathcal{A}',\mathcal{A}}$ reads $\begin{pmatrix} 0 \\ 0 \\ 2 \\ 0 \end{pmatrix}$.

Observe that the coordinate vectors before and after applying T has the same matrix multiplication. For instance, if we express $T(2x^2 + 4x^3) = (4x + 12x^2)$ in terms of matrices and vectors:

$$[2x^2 + 4x^3]_{\mathcal{A}} = \begin{pmatrix} 0 \\ 0 \\ 2 \\ 4 \end{pmatrix}, \quad [4x + 12x^2]_{\mathcal{A}} = \begin{pmatrix} 0 \\ 4 \\ 12 \\ 0 \end{pmatrix},$$

Then one can easily check that the matrix representation $T_{\mathcal{A},\mathcal{A}}$ satisfies

$$T_{\mathcal{A},\mathcal{A}} \cdot [2x^2 + 4x^3]_{\mathcal{A}} = [4x + 12x^2]_{\mathcal{A}} = [T(2x^2 + 4x^3)]_{\mathcal{A}}.$$

In other words, one can compute $T : V \to V$ by using $T_{\mathcal{A},\mathcal{A}}$. More generally:

Theorem 3.12 (Matrix Representation)**.** *Let $T : V \to W$ be a linear transformation of finite-dimensional vector spaces. Let $\mathcal{A}, \mathcal{B}$ be the ordered bases of V and W, respectively. Then the*

following diagram holds:

$$
\begin{array}{ccc}
V & \xrightarrow{\;\;T\;\;} & W \\
{\scriptstyle [\cdot]_{\mathcal{A}}}\big\downarrow & & \big\downarrow{\scriptstyle [\cdot]_{\mathcal{B}}} \\
\mathbb{F}^n & \xrightarrow[\;T_{\mathcal{B}\mathcal{A}}\;]{} & \mathbb{F}^m
\end{array}
$$

Namely, for any $\mathbf{v} \in V$,

$$T_{\mathcal{B},\mathcal{A}}[\mathbf{v}]_{\mathcal{A}} = [T(\mathbf{v})]_{\mathcal{B}}.$$

(Informally, the linear transformation T 'corresponds' to multiplication of coordinate matrix $T_{\mathcal{B},\mathcal{A}}$).

Proof. Suppose $\mathcal{A} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ and $\mathcal{B} = \{\mathbf{w}_1, \ldots, \mathbf{w}_m\}$. The proof of this theorem follows the same procedure as in .

1. **Basis vector** We compare both sides:

$$\text{LHS} = T_{\mathcal{B}\mathcal{A}}[\mathbf{v}_j]_{\mathcal{A}} = [\alpha_{ij}] \cdot \mathbf{e}_j = \begin{pmatrix} \alpha_{1j} \\ \vdots \\ \alpha_{nj} \end{pmatrix} = \left(\sum_{i=1}^{m} \alpha_{ij} \mathbf{w}_i \right)_{\mathcal{B}} = [T(\mathbf{v}_j)]_{\mathcal{B}} = \text{RHS}$$

So the identity $T_{\mathcal{B}\mathcal{A}}[\mathbf{v}_j]_{\mathcal{A}} = [T\mathbf{v}_j]_{\mathcal{B}}$ holds for all basis vectors $\mathbf{v}_j$.

2. **Extension to arbitrary vector** Let $\mathbf{v} = \sum_{j=1}^{n} r_j \mathbf{v}_j$ with coefficients $r_j \in \mathbb{F}$. Then:

$$
\begin{aligned}
T_{\mathcal{B}\mathcal{A}}([\mathbf{v}]_{\mathcal{A}}) = T_{\mathcal{B}\mathcal{A}} \left(\sum_{j=1}^{n} r_j [\mathbf{v}_j]_{\mathcal{A}} \right) &= \sum_{j=1}^{n} r_j T_{\mathcal{B}\mathcal{A}}[\mathbf{v}_j]_{\mathcal{A}} \\
&= \sum_{j=1}^{n} r_j [T(\mathbf{v}_j)]_{\mathcal{B}} \\
&= \left[\sum_{j=1}^{n} r_j T(\mathbf{v}_j) \right]_{\mathcal{B}} \\
&= [T(\sum_{j=1}^{n} r_j \mathbf{v}_j)]_{\mathcal{B}} = [T(\mathbf{v})]_{\mathcal{B}}.
\end{aligned}
$$

Thus, for all $\mathbf{v} \in V$, we have

$$T_{\mathcal{B}\mathcal{A}}[\mathbf{v}]_{\mathcal{A}} = [T(\mathbf{v})]_{\mathcal{B}},$$

completing the proof. $\qquad\square$

Remark 3.13. Suppose

$$\text{id} : V_{\mathcal{A}} \to V_{\mathcal{A}'}$$

is the identity transformation $\text{id}(\mathbf{v}) := \mathbf{v}$ for all $\mathbf{v} \in V$, and $\mathcal{A}, \mathcal{A}'$ are two ordered bases of V. We would like to study the matrix representation

$$\text{id}_{\mathcal{A}',\mathcal{A}}$$

(from now on, we write $T : V_{\mathcal{A}} \to W_{\mathcal{B}}$ instead of $T : V \to W$ whenever we study the matrix representation $T_{\mathcal{B},\mathcal{A}}$).

Firstly, Theorem 3.12 implies that

$$\text{id}_{\mathcal{A}',\mathcal{A}}([\mathbf{v}]_{\mathcal{A}}) = [\text{id}(\mathbf{v})]_{\mathcal{A}'} = [\mathbf{v}]_{\mathcal{A}'}.$$

On the other hand, Proposition 3.7 implies that the change of basis matrix $C_{\mathcal{A}',\mathcal{A}}$ satisfies:

$$C_{\mathcal{A}',\mathcal{A}}([\mathbf{v}]_{\mathcal{A}}) = [\mathbf{v}]_{\mathcal{A}'}.$$

Therefore,

$$\text{id}_{\mathcal{A}',\mathcal{A}} = C_{\mathcal{A}',\mathcal{A}}$$

This shows that Theorem 3.12 generalizes the change-of-basis theorem Proposition 3.7.

Another advantage of using matrix representation to understand linear transformation is that:

Composing linear transformations $S \circ T$ $\longleftrightarrow$ Multiplying matrix representations $S_{\mathcal{C},\mathcal{B}} \cdot T_{\mathcal{B},\mathcal{A}}$

Proposition 3.14 (Functoriality)**.** *Let $T : V \to W$, $S : W \to U$ be linear transformations, and $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$ be ordered bases of V, W and U respectively. Then the corresponding matrix representations satisfy*

$$(S \circ T)_{\mathcal{C},\mathcal{A}} = S_{\mathcal{C},\mathcal{B}} \cdot T_{\mathcal{B},\mathcal{A}}. \tag{3.2}$$

Proof. Let $\mathcal{A} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$, $\mathcal{B} = \{\mathbf{w}_1, \ldots, \mathbf{w}_m\}$, $\mathcal{C} = \{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$. As before, it suffices to check (3.2) holds on each column:

$$(S \circ T)_{\mathcal{C},\mathcal{A}}\mathbf{e}_j = (S_{\mathcal{C},\mathcal{B}} \cdot T_{\mathcal{B},\mathcal{A}})\mathbf{e}_j.$$

Recall that

$$T(\mathbf{v}_j) = \sum_i (T_{\mathcal{B},\mathcal{A}})_{ij}\mathbf{w}_i, \qquad S(\mathbf{w}_i) = \sum_k (S_{C,\mathcal{B}})_{ki}\mathbf{u}_k$$

Therefore,

$$(S \circ T)_{\mathcal{C},\mathcal{A}} \mathbf{e}_j = (S \circ T)_{\mathcal{C},\mathcal{A}} (\mathbf{v}_j)_{\mathcal{A}} = (S \circ T(\mathbf{v}_j))_{\mathcal{C}}$$

$$= \left[S \circ \left(\sum_i (T_{\mathcal{B},\mathcal{A}})_{ij} \mathbf{w}_i \right) \right]_{\mathcal{C}}$$

$$= \sum_i (T_{\mathcal{B},\mathcal{A}})_{ij} (S(\mathbf{w}_i))_{\mathcal{C}}$$

$$= \sum_i (T_{\mathcal{B},\mathcal{A}})_{ij} \left(\sum_k (S_{C,\mathcal{B}})_{ki} \mathbf{u}_k \right)_{\mathcal{C}}$$

$$= \sum_k \sum_i (S_{C,\mathcal{B}})_{ki} (T_{\mathcal{B},\mathcal{A}})_{ij} (\mathbf{u}_k)_{\mathcal{C}}$$

$$= \sum_k (S_{C,\mathcal{B}} \cdot T_{\mathcal{B},\mathcal{A}})_{kj} (\mathbf{u}_k)_C$$

$$= \sum_k (S_{C,\mathcal{B}} \cdot T_{\mathcal{B},\mathcal{A}})_{kj} \mathbf{e}_k$$

$$= (S_{C\mathcal{B}} \cdot T_{\mathcal{B},\mathcal{A}}) \cdot \mathbf{e}_j$$

Consequently, the result follows. □

Example 3.15. Let id : $V \to V$ be the identity map, and $\mathcal{A}, \mathcal{A}'$ are two ordered bases of V. By Remark 3.13, $\mathrm{id}_{\mathcal{A}',\mathcal{A}} = C_{\mathcal{A}',\mathcal{A}}$ and $\mathrm{id}_{\mathcal{A},\mathcal{A}'} = C_{\mathcal{A},\mathcal{A}'}$. Now Proposition 3.14 implies that

$$(\mathrm{id} \circ \mathrm{id})_{\mathcal{A},\mathcal{A}} = \mathrm{id}_{\mathcal{A},\mathcal{A}'} \cdot \mathrm{id}_{\mathcal{A}',\mathcal{A}} = C_{\mathcal{A},\mathcal{A}'} \cdot C_{\mathcal{A}',\mathcal{A}}$$

But obviously $(\mathrm{id} \circ \mathrm{id})_{\mathcal{A},\mathcal{A}} = \mathrm{id}_{\mathcal{A},\mathcal{A}} = I_{\dim(V) \times \dim(V)}$ is the identity matrix. Therefore, we have reproved that

$$I_{\dim(V) \times \dim(V)} = C_{\mathcal{A},\mathcal{A}'} \cdot C_{\mathcal{A}',\mathcal{A}}.$$

3.2 Relation to Similar Matrices

We end this section with a very important observations in linear algebra: Let $T : V \to V$ be a linear operator with $\mathcal{A}, \mathcal{A}'$ being two ordered bases of V. We wish to relate

$$T_{\mathcal{A}',\mathcal{A}'} \text{ and } T_{\mathcal{A},\mathcal{A}}.$$

More explicitly, rewrite T by $T = \mathrm{id} \circ T \circ \mathrm{id}$. By specifying which bases we are working on, $T : V_{\mathcal{A}'} \to V_{\mathcal{A}'}$ can be understood as:

$$V_{\mathcal{A}'} \xrightarrow{\mathrm{id}} V_{\mathcal{A}} \xrightarrow{T} V_{\mathcal{A}} \xrightarrow{\mathrm{id}} V_{\mathcal{A}'}$$

By Proposition 3.14,

$$T_{\mathcal{A}',\mathcal{A}'} = \mathrm{id}_{\mathcal{A}',\mathcal{A}} \cdot T_{\mathcal{A},\mathcal{A}} \cdot \mathrm{id}_{\mathcal{A},\mathcal{A}'} = C_{\mathcal{A}',\mathcal{A}} \cdot T_{\mathcal{A},\mathcal{A}} \cdot C_{\mathcal{A},\mathcal{A}'} = (C_{\mathcal{A},\mathcal{A}'})^{-1} T_{\mathcal{A},\mathcal{A}} C_{\mathcal{A},\mathcal{A}'}.$$

Therefore, the two matrix representations $T_{\mathcal{A}',\mathcal{A}'}$ and $T_{\mathcal{A},\mathcal{A}}$ are **similar matrices**. In particular, they share the same characteristic polynomials, eigenvalues, determinant, etc.

Putting this into the perspective of matrix representations, recall that we do the following in Example 3.10:

- Let $T : \mathbb{F}^n \to \mathbb{F}^n$ be defined by $T(\mathbf{x}) := A\mathbf{x}$.

- Let $\mathcal{B} = \{\mathbf{m}_1, \ldots, \mathbf{m}_n\}$ be an ordered basis of $\mathbb{F}^n$, then

$$T_{\mathcal{E},\mathcal{E}} = B$$

- Let $\mathcal{E} = \{\mathbf{e}_1, \ldots \mathbf{e}_n\}$ be the usual ordered basis of $\mathbb{F}^n$. Then one gets back the original matrix

$$T_{\mathcal{E},\mathcal{E}} = A$$

Let M be the matrix whose columns are given by $\mathcal{B}$, then Proposition 3.7 implies that $C_{\mathcal{E},\mathcal{B}} = M$, and our discussions above gives
$$B = M^{-1}AM.$$
In particular, A and B are similar.

To showcase the above arguments in the setting of Example 3.10, note that

$$T_{\mathcal{E},\mathcal{E}} = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \qquad T_{\mathcal{B},\mathcal{B}} = \begin{pmatrix} 0 & -2 \\ -1 & 5 \end{pmatrix}$$

and the change of basis matrix is $C_{\mathcal{E},\mathcal{B}} = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}$. Then one can check that

$$\begin{pmatrix} 0 & -2 \\ -1 & 5 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}^{-1} \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}$$

Now one can understand diagonalizable matrices better: Recall in MAT2040, the following are equivalent:

- A is diagonalizable;

- There exists a basis $\mathcal{B}$ of eigenvectors;

- There exists an invertible M such that

$$\begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix} = M^{-1}AM,$$

i.e. A is similar to a diagonal matrix.

Under our new perspective, we have a natural account (rather than direct computation) on why the columns of M are given by the eigen-basis $\mathcal{B}$, and diagonalization is simply a change-of-basis of the **same** linear transformation $T(\mathbf{x}) = A\mathbf{x}$ from the usual basis $\mathcal{E}$ to the (eigen)-basis $\mathcal{B}$.

To conclude:

A, B are **similar** $\longleftrightarrow$	A, B are matrix representations of the **same** linear transformation under **different** coordinate systems.

Chapter 4

Quotient Spaces & Its Universal Properties

4.1 Cosets and Quotient Space

One important aspect in mathematics is to construct *new* objects from *known* ones. In this chapter, we will study the quotient space of a (known) vector space V by a (known) vector subspace W. Informally speaking, we will 'divide' the big vector space V into many 'slices' of W.

Definition 4.1 (Coset). Let V be a vector space and $W \leq V$. For any element $\mathbf{v} \in V$, the (right) coset determined by $\mathbf{v}$ is the set

$$\mathbf{v} + W := \{\mathbf{v} + \mathbf{w} \mid \mathbf{w} \in W\}.$$

Example 4.2. Consider $V = \mathbb{R}^3$ and $W = \operatorname{span}\{(1, 2, 0)\}$. Then the coset determined by $\mathbf{v} = (5, 6, -3)$ can be written as

$$\mathbf{v} + W = \{(5 + t, 6 + 2t, -3) \mid t \in \mathbb{R}\}.$$

It's interesting that the coset determined by $\mathbf{v}' = (4, 4, -3)$ is exactly the same as the coset shown above:

$$\mathbf{v}' + W = \{(4 + t, 4 + 2t, -3) \mid t \in \mathbb{R}\} = \mathbf{v} + W.$$

Therefore, writing the exact expression of $\mathbf{v} + W$ may sometimes become tedious and hard to check the equivalence. We say $\mathbf{v}$ is a *representative* of a coset $\mathbf{v} + W$.

Example 4.3. One motivation to understand cosets is to describe the solutions of the linear system $A\mathbf{x} = \mathbf{b}$ with $A \in \mathbb{R}^{m \times n}$. Recall in MAT2040 that the general step for solving this linear system is as follows:

1. Find the solution set for $A\mathbf{x} = 0$, i.e., the set $\ker(A) \leq \mathbb{R}^n$.

2. Find a particular solution $\mathbf{x}_0$ such that $A\mathbf{x}_0 = \mathbf{b}$.

Then the general solution set to this linear system is the coset $\mathbf{x}_0 + \ker(A)$ of $\mathbb{R}^n$.

Recall that the particular solution $\mathbf{x}_0$ is **not** unique. Indeed, if $\mathbf{y}_0$ is another particular solution of $A\mathbf{x} = \mathbf{b}$, then

$$\mathbf{x}_0 + \ker(A) = \mathbf{y}_0 + \ker(A)$$

describes the same set of solutions of the system of equations.

As directed by the previous examples, we have:

Proposition 4.4. *Two cosets are the same if and only if the subtraction for the corresponding representatives is in W, i.e.,*

$$\mathbf{v}_1 + W = \mathbf{v}_2 + W \iff \mathbf{v}_1 - \mathbf{v}_2 \in W.$$

Proof. **Necessity.** Suppose that $\mathbf{v}_1 + W = \mathbf{v}_2 + W$, then $\mathbf{v}_1 + \mathbf{w}_1 = \mathbf{v}_2 + \mathbf{w}_2$ for some $\mathbf{w}_1, \mathbf{w}_2 \in W$, which implies

$$\mathbf{v}_1 - \mathbf{v}_2 = \mathbf{w}_2 - \mathbf{w}_1 \in W.$$

Sufficiency. Suppose that $\mathbf{v}_1 - \mathbf{v}_2 = \mathbf{w} \in W$. It suffices to show $\mathbf{v}_1 + W \subseteq \mathbf{v}_2 + W$. For any $\mathbf{v}_1 + \mathbf{w}' \in \mathbf{v}_1 + W$, this element can be expressed as

$$\mathbf{v}_1 + \mathbf{w}' = (\mathbf{v}_2 + \mathbf{w}) + \mathbf{w}' = \mathbf{v}_2 + \underbrace{(\mathbf{w} + \mathbf{w}')}_{\in W} \in \mathbf{v}_2 + W.$$

Therefore, $\mathbf{v}_1 + W \subseteq \mathbf{v}_2 + W$. Similarly, we can show that $\mathbf{v}_2 + W \subseteq \mathbf{v}_1 + W$, and thus $\mathbf{v}_1 + W = \mathbf{v}_2 + W$. $\qquad\square$

(Exercise: If two cosets $\mathbf{v}_1 + W \neq \mathbf{v}_2 + W$ are not equal, then they have no intersection, i.e. $(\mathbf{v}_1 + W) \cap (\mathbf{v}_2 + W) = \phi$).

Remark 4.5. To many of us, this may be the first occasion to come across some mathematical objects that has **different expressions of the same element**, i.e. one may have

$$\mathbf{v}_1 \neq \mathbf{v}_2 \text{ but } \mathbf{v}_1 + W = \mathbf{v}_2 + W.$$

Therefore, extra care is needed whenever we study such objects (see **WARNING** above Proposition 4.7).

Definition 4.6 (Quotient Space)**.** The quotient space V/W of V by the subspace W is the collection of all cosets $\mathbf{v} + W$, i.e.

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

To make the quotient space a vector space structure, we define the addition and scalar multiplication on V/W by:

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

$$\alpha \cdot (\mathbf{v} + W) := (\alpha \cdot \mathbf{v}) + W$$

For example, consider $V = \mathbb{R}^2$ and $W = \text{span}\{\begin{pmatrix} 1 \\ 1 \end{pmatrix}\}$. Then note that:

$$\left(\begin{pmatrix} 1 \\ 0 \end{pmatrix} + W\right) + \left(\begin{pmatrix} 2 \\ 0 \end{pmatrix} + W\right) = \begin{pmatrix} 3 \\ 0 \end{pmatrix} + W$$

$$\pi \cdot \left(\begin{pmatrix} 1 \\ 0 \end{pmatrix} + W\right) = \left(\begin{pmatrix} \pi \\ 0 \end{pmatrix} + W\right)$$

WARNING: As mentioned in Remark 4.5, one has different expression of the same element in V/W. For instance,

$$\begin{pmatrix} 1 \\ 0 \end{pmatrix} + W = \begin{pmatrix} 11 \\ 10 \end{pmatrix} + W, \qquad \begin{pmatrix} 2 \\ 0 \end{pmatrix} + W = \begin{pmatrix} -98 \\ -100 \end{pmatrix} + W$$

by Proposition 4.4. It is unclear from our definition of addition that

$$\left(\begin{pmatrix} 1 \\ 0 \end{pmatrix} + W\right) + \left(\begin{pmatrix} 2 \\ 0 \end{pmatrix} + W\right) = \left(\begin{pmatrix} 11 \\ 10 \end{pmatrix} + W\right) + \left(\begin{pmatrix} -98 \\ -100 \end{pmatrix} + W\right)$$

are both equal to $\begin{pmatrix} 3 \\ 0 \end{pmatrix} + W$ (similar problem also occurs in scalar multiplication). In other words, we do not know if the functions $+ : V/W \times V/W \to V/W$ given by

$$(\mathbf{v}_1 + W, \mathbf{v}_2 + W) \mapsto (\mathbf{v}_1 + \mathbf{v}_2) + W$$

and $\cdot : \mathbb{F} \times V/W \to V/W$ given by

$$(\alpha, \mathbf{v} + W) \mapsto (\alpha \cdot \mathbf{v}) + W$$

are **well-defined** [1].

Proposition 4.7. *Addition and scalar multiplication on V/W is well-defined.*

Proof. **Addition.** Suppose that

$$\begin{cases} \mathbf{v}_1 + W = \mathbf{v}_1' + W \\ \mathbf{v}_2 + W = \mathbf{v}_2' + W \end{cases} \tag{4.1}$$

We need to show that

$$(\mathbf{v}_1 + \mathbf{v}_2) + W = (\mathbf{v}_1' + \mathbf{v}_2') + W.$$

From (4.1) and Proposition 4.4, we have:

$$\mathbf{v}_1 - \mathbf{v}_1' \in W, \quad \mathbf{v}_2 - \mathbf{v}_2' \in W,$$

---

[1] A function $f : A \to B$ is well-defined if for $a_1 = a_2 \in A$, $f(a_1) = f(a_2) \in B$.

which implies
$$(\mathbf{v}_1 - \mathbf{v}_1') + (\mathbf{v}_2 - \mathbf{v}_2') = (\mathbf{v}_1 + \mathbf{v}_2) - (\mathbf{v}_1' + \mathbf{v}_2') \in W.$$

By Proposition 4.4 again, we conclude:

$$(\mathbf{v}_1 + \mathbf{v}_2) + W = (\mathbf{v}_1' + \mathbf{v}_2') + W.$$

Multiplication.   For scalar multiplication, similarly, we can show that $\mathbf{v}_1 + W = \mathbf{v}_1' + W$ implies

$$\alpha\mathbf{v}_1 + W = \alpha\mathbf{v}_1' + W \quad \text{for all } \alpha \in \mathbb{F}.$$

$\square$

Consequently
$$(V/W, +, \ \cdot)$$

is a vector space, and is called the **quotient space**. In particular, its zero vector is

$$\mathbf{0}_{V/W} = \mathbf{0}_V + W.$$

Proposition 4.8. *The* **canonical projection** *mapping defined by*

$$\pi_W : V \to V/W, \quad \mathbf{v} \mapsto \mathbf{v} + W$$

is a surjective linear transformation, with $\ker(\pi_W) = W$.

Proof. We follow a few steps:

1. To show the mapping π_W is a linear transformation, note that

$$\begin{aligned}
\pi_W(\alpha\mathbf{v}_1 + \beta\mathbf{v}_2) &= (\alpha\mathbf{v}_1 + \beta\mathbf{v}_2) + W \\
&= (\alpha\mathbf{v}_1 + W) + (\beta\mathbf{v}_2 + W) \\
&= \alpha(\mathbf{v}_1 + W) + \beta(\mathbf{v}_2 + W) \\
&= \alpha\pi_W(\mathbf{v}_1) + \beta\pi_W(\mathbf{v}_2).
\end{aligned}$$

2. Then we show that $\ker(\pi_W) = W$:

$$\pi_W(\mathbf{v}) = \mathbf{0}_{V/W} \ \Leftrightarrow \ \mathbf{v} + W = 0 + W \ \Leftrightarrow \ \mathbf{v} = (\mathbf{v} - 0) \in W.$$

3. Finally, for any $\mathbf{v}_0 + W \in V/W$, we can construct $\mathbf{v}_0 \in V$ such that $\pi_W(\mathbf{v}_0) = \mathbf{v}_0 + W$. Therefore the mapping π_W is surjective.

So the proposition is proved. $\square$

4.2   First Isomorphism Theorem

Proposition 4.9 (Universal Property of the Quotient Map I)**.** *Suppose that $T : V \to U$ is a linear transformation, and that $W \leq \ker(T)$. Then the mapping*

$$\widetilde{T} : V/W \to U, \quad \mathbf{v} + W \mapsto T(\mathbf{v})$$

is a well-defined linear transformation. As a result, the following diagram commutes:

$$
\begin{array}{ccc}
V & \xrightarrow{\ \pi_W\ } & V/W \\
 & T \searrow & \Big\downarrow \widetilde{T} \\
 & & U
\end{array}
$$

In other words, we have $T = \widetilde{T} \circ \pi_W$.

Proof. First we show the well-definedness. Suppose that $\mathbf{v}_1 + W = \mathbf{v}_2 + W$, and we suffice to show

$$\widetilde{T}(\mathbf{v}_1 + W) = \widetilde{T}(\mathbf{v}_2 + W),$$

i.e.,

$$T(\mathbf{v}_1) = T(\mathbf{v}_2).$$

By Proposition 4.4, we imply

$$\mathbf{v}_1 - \mathbf{v}_2 \in W \leq \ker(T) \Rightarrow T(\mathbf{v}_1 - \mathbf{v}_2) = 0 \Rightarrow T(\mathbf{v}_1) - T(\mathbf{v}_2) = \mathbf{0}.$$

Then we show $\widetilde{T}$ is a linear transformation:

$$
\begin{aligned}
\widetilde{T}(\alpha(\mathbf{v}_1 + W) + \beta(\mathbf{v}_2 + W)) &= \widetilde{T}((\alpha\mathbf{v}_1 + \beta\mathbf{v}_2) + W) \\
&= T(\alpha\mathbf{v}_1 + \beta\mathbf{v}_2) \\
&= \alpha T(\mathbf{v}_1) + \beta T(\mathbf{v}_2) \\
&= \alpha\widetilde{T}(\mathbf{v}_1 + W) + \beta\widetilde{T}(\mathbf{v}_2 + W)
\end{aligned}
$$

$\square$

We now study a special case of Proposition 4.9: under the same setting, we let $W = \ker(T)$ and $U = T(V) = \operatorname{im}(T)$. Then one has a linear transformation mapping

$$\widetilde{T} : V/\ker(T) \to \operatorname{im}(T).$$

In fact, more is true:

Theorem 4.10 (First Isomorphism Theorem)**.** *Let $T : V \to U$ be a linear transformation. Then the mapping*

$$\widetilde{T} : V/\ker(T) \to \mathrm{im}(T), \quad \mathbf{v} + \ker(T) \mapsto T(\mathbf{v})$$

is an isomorphism.

Proof. **Injectivity.** Suppose that

$$\widetilde{T}(\mathbf{v}_1 + \ker(T)) = \widetilde{T}(\mathbf{v}_2 + \ker(T)),$$

then we imply

$$T(\mathbf{v}_1) = T(\mathbf{v}_2) \Rightarrow T(\mathbf{v}_1 - \mathbf{v}_2) = 0_U \Rightarrow \mathbf{v}_1 - \mathbf{v}_2 \in \ker(T),$$

i.e.,

$$\mathbf{v}_1 + \ker(T) = \mathbf{v}_2 + \ker(T).$$

Surjectivity. For $\mathbf{u} \in U$, due to the surjectivity of T, we can find a $\mathbf{v}_0 \in V$ such that $T(\mathbf{v}_0) = \mathbf{u}$. Therefore, we can construct a set $\mathbf{v}_0 + \ker(T)$ such that

$$\widetilde{T}(\mathbf{v}_0 + \ker(T)) = \mathbf{u}.$$

$\square$

Example 4.11. Suppose that $U, W \leq V$ with $U \cap W = \{\mathbf{0}\}$, the mapping

$$\phi : U \oplus W \to U, \quad \phi(\mathbf{u} + \mathbf{w}) = \mathbf{u}.$$

is a well-defined surjective linear transformation with $\ker(\phi) = W$. (Exercise: If $U \cap W \neq \{\mathbf{0}\}$, then $\phi : U + W \to U$ given by $\phi(\mathbf{u} + \mathbf{w}) := \mathbf{u}$ is **not** well-defined.)

By Theorem 4.10, the new linear transformation

$$\widetilde{\phi} : (U \oplus W)/W \to U, \quad \widetilde{\phi}((\mathbf{u} + \mathbf{w}) + W) = \mathbf{u}.$$

is an isomorphism of vector spaces.

4.3   Universal Property of Quotient Spaces
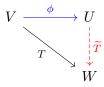
Definition 4.12 (Universal Property of Quotient Spaces)**.** Let V be a vector space and let $V' \leq V$ be a subspace. Consider the collection of linear transformations:

$$\mathrm{Obj} = \left\{T : V \to W \mid T \text{ is linear}, \ V' \leq \ker(T)\right\}.$$

(*Important example:* the canonical projection $\pi_{V'} : V \to V/V'$ in Proposition 4.8 belongs to Obj.)

An element $(\phi : V \to U) \in \mathrm{Obj}$ is said to satisfy the **universal property** if the following holds: Given any $T : V \to W$ in Obj, there exists a unique linear transformation $\widetilde{T} : U \to W$ such that the following diagram commutes:

$$
\begin{array}{ccc}
V & \xrightarrow{\ \phi\ } & U \\
 & {\scriptstyle T}\searrow & \big\downarrow {\scriptstyle \widetilde{T}} \\
 & & W
\end{array}
$$

Equivalently, for any $T : V \to W$ in Obj, there exists a unique map $\widetilde{T} : U \to W$ such that

$$T = \widetilde{T} \circ \phi.$$

Theorem 4.13 (Universal Property of Quotient Space II). *Let V be a vector space and $V' \le V$. Then:*

1. *The canonical projection $\pi_{V'} : V \to V/V'$ is a universal object in the category* Obj, *i.e., it satisfies the universal property of Definition 4.12.*

2. *If $\phi : V \to U$ is a universal object in* Obj, *then $U \cong V/V'$. In other words, there is intrinsically "one" element (up to isomorphism) in the set of universal objects.*

Proof.   1. Consider any linear transformation $T : V \to W$ such that $V' \le \ker(T)$. By Proposition 4.9, there is a linear transformation $\widetilde{T} : V/V' \to W$ such that

$$T = \widetilde{T} \circ \pi_{V'},$$

i.e., $\pi_{V'}$ satisfies the commuting diagram in the theorem.

To show uniqueness of $\widetilde{T}$, suppose there exists another map $\widetilde{S} : V/V' \to W$ such that

$$\widetilde{T} \circ \pi_{V'} = T = \widetilde{S} \circ \pi_{V'}$$

Then for any $\mathbf{v} + V' \in V/V'$, we compute:

$$\widetilde{S}(\mathbf{v} + V') := \widetilde{S}(\pi_{V'}(\mathbf{v})) = T(\mathbf{v}),$$

and by definition of $\widetilde{T}$, we also have $T(\mathbf{v}) = \widetilde{T}(\mathbf{v} + V')$. Hence,

$$\widetilde{S}(\mathbf{v} + V') = \widetilde{T}(\mathbf{v} + V')$$

for all $\mathbf{v} + V' \in V/V'$, proving $\widetilde{S} = \widetilde{T}$ is unique.

2. Suppose there exists another $(\phi : V \to U) \in \mathrm{Obj}$ also satisfies the universal property, i.e. for all $(T : V \to W) \in \mathrm{Obj}$, one has:

So one can take $T = \pi_{V'} : V \to V/V'$, i.e.

$$V \xrightarrow{\ \phi\ } U$$

with T from V to W and $\widetilde{T}$ from U to W, reaching W.

$$V \xrightarrow{\ \phi\ } U$$

with $\pi_{V'}$ from V and α from U, reaching V/V'.

for $\alpha = \widetilde{\pi}_{V'}$. Similarly, by the fact that $(\pi_{V'} : V \to V/V') \in \mathrm{Obj}$ is universal, one can take $T = \phi : V \to U$ and get

$$V \xrightarrow{\ \pi_{V'}\ } V/V'$$

with ϕ from V and β from V/V', reaching U.

for $\beta = \widetilde{\phi}$. Combining the two, we obtain the following diagram:

$$V \xrightarrow{\ \pi_{V'}\ } V/V'$$

with ϕ, $\pi_{V'}$ from V, β from V/V' to U, and α from U to V/V'.

Therefore,

$$\alpha \circ \beta \circ \pi_{V'} = \pi_{V'} = \mathrm{id}_{V/V'} \circ \pi_{V'}.$$

By (1), this implies $\alpha \circ \beta = \mathrm{id}_{V/V'}$. Similarly, one can show that $\beta \circ \alpha = \mathrm{id}_U$ and hence

$$\alpha : U \to V/V' \qquad \beta : V/V' \to U$$

are inverses of each other, i.e.

$$V/V' \cong U.$$

$\square$

Chapter 5

Dual Spaces, Annihilators, and the Adjoint Map

5.1 Dual Spaces

Definition 5.1 (Dual Space). Let V be a vector space over a field $\mathbb{F}$. The *dual vector space V^** is defined as

$$V^* = \operatorname{Hom}_{\mathbb{F}}(V, \mathbb{F}) = \{f : V \to \mathbb{F} \mid f \text{ is a linear transformation}\}.$$

Example 5.2.

1. Let $V = \mathbb{R}^n$, and define $\phi_i : V \to \mathbb{R}$ to be the i-th coordinate function:

$$\phi_i \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = x_i.$$

   Then clearly $\phi_i \in V^*$. On the contrary, the function

$$\phi_i^2 \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = x_i^2$$

   is not linear, hence $\phi_i^2 \notin V^*$.

2. Let $V = \mathbb{F}[x]$, and define $\phi : V \to \mathbb{F}$ by $\phi(p(x)) = p(1)$. Then $\phi \in V^*$ because:

$$\phi(ap(x) + bq(x)) = ap(1) + bq(1) = a\phi(p(x)) + b\phi(q(x)).$$

3. Let $\psi : V \to \mathbb{F}$ be defined by $\psi(p(x)) = \int_0^1 p(x)\, dx$. Then $\psi \in V^*$.

4. Let $V = M_{n \times n}(\mathbb{F})$, and define $\operatorname{tr} : V \to \mathbb{F}$ by $\operatorname{tr}(M) = \sum_{i=1}^{n} M_{ii}$. Then $\operatorname{tr} \in V^*$. However, the

determinant function $\det : V \to \mathbb{F}$ is not linear, hence $\det \notin V^*$.

5.1.1 Dual Basis

Definition 5.3 (Dual Basis)**.** Let V be a vector space with basis $\mathcal{B} = \{\mathbf{v}_i \mid i \in I\}$ (where I may be finite, countable, or uncountable). Define

$$\mathcal{B}^* = \{f_i : V \to \mathbb{F} \mid i \in I\},$$

where the functionals $f_i \in V^*$ satisfies:

$$f_i(\mathbf{v}_j) = \delta_{ij} = \begin{cases} 1, & \text{if } i = j, \\ 0, & \text{if } i \neq j. \end{cases}$$

for all $\mathbf{v}_j \in \mathcal{B}$; and for all $\mathbf{v} = \sum_{\ell=1}^{k} \alpha_\ell v_{i_\ell} \in V$:

$$f_i \left(\sum_{\ell=1}^{k} \alpha_\ell \mathbf{v}_{i_\ell} \right) = \sum_{\ell=1}^{k} \alpha_\ell f_i(\mathbf{v}_{i_\ell}) = \alpha_i.$$

(c.f. Remark 2.4). The natural question is: **Does the set $\mathcal{B}^*$ form a basis for V^*?**

Proposition 5.4. *Let $B = \{\mathbf{v}_i \mid i \in I\}$ be a basis of a vector space V, and define the dual set*

$$\mathcal{B}^* = \{f_i : V \to \mathbb{F} \mid i \in I\}$$

as in Definition 5.3. Then:

1. *$\mathcal{B}^*$ is always linearly independent. That is, any finite subset of $\mathcal{B}^*$ is linearly independent.*

2. *If V is finite-dimensional, then $\mathcal{B}^*$ forms a basis of V^*.*

Proof.   1. Let $f_1, \ldots, f_k \in \mathcal{B}^*$, and suppose there exists a linear dependence:

$$\alpha_1 f_1 + \alpha_2 f_2 + \cdots + \alpha_k f_k = 0_{V^*}.$$

We will show all $\alpha_i = 0$. Since each $f_i \in \mathcal{B}^*$ satisfies $f_i(\mathbf{v}_j) = \delta_{ij}$, evaluate the left-hand side on $\mathbf{v}_1 \in B$:

$$\alpha_1 f_1(\mathbf{v}_1) + \alpha_2 f_2(\mathbf{v}_1) + \cdots + \alpha_k f_k(\mathbf{v}_1) = \alpha_1 = 0.$$

Repeating this for $\mathbf{v}_2, \ldots, \mathbf{v}_k$ gives $\alpha_2 = \cdots = \alpha_k = 0$. Thus, $\{f_1, \ldots, f_k\}$ is linearly independent.

2. Suppose V is finite-dimensional with basis $B = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$, and let $\mathcal{B}^* = \{f_1, \ldots, f_n\}$. For any $f \in V^*$, define

$$g := \sum_{i=1}^{n} f(\mathbf{v}_i) f_i \in \text{span}(\mathcal{B}^*).$$

For any $j = 1, \ldots, n$, we compute:

$$g(\mathbf{v}_j) = \sum_{i=1}^{n} f(\mathbf{v}_i) f_i(\mathbf{v}_j) = \sum_{i=1}^{n} f(\mathbf{v}_i) \delta_{ij} = f(\mathbf{v}_j).$$

Hence $g(\mathbf{v}) = f(\mathbf{v})$ for all $\mathbf{v} \in V$, by linearity and since the $\mathbf{v}_j$'s form a basis. Therefore, $f = g \in \mathrm{span}(\mathcal{B}^*)$, and we conclude:

$$V^* = \mathrm{span}(\mathcal{B}^*).$$

Since $\mathcal{B}^*$ is linearly independent and spans V^*, it forms a basis.

$\square$

Corollary 5.5. *If* $\dim(V) = n$, *then* $\dim(V^*) = n$.

Proof. Define the map

$$\Phi : V \longrightarrow V^*, \qquad \Phi(\mathbf{v}_i) := f_i,$$

where $f_i \in \mathcal{B}^*$ is the dual basis corresponding to the basis $\mathcal{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ of V (once again, we use Remark 2.4 to define Φ). This is a linear isomorphism.

Alternatively, recall that $V^* = \mathrm{Hom}(V, \mathbb{F})$ and hence

$$\dim(V^*) = \dim(V) \dim(\mathbb{F}) = \dim(V) \cdot 1 = \dim(V).$$

$\square$

Remark 5.6. The above definition of the isomorphism Φ requires a specified choice of basis in V, so it is not a *natural isomorphism.*

Remark 5.7. Part (2) of Proposition 5.4 does not hold when V is infinite-dimensional. That is, the set $\mathcal{B}^*$ may not span all of V^*, since we only allow finite linear combinations in the definition of span. The following example illustrates a counterexample.

Example 5.8. Let $V = \mathbb{F}[x]$, the space of polynomials, and define a basis

$$B = \{1, x, x^2, x^3, \ldots\}, \quad \text{so} \quad \mathcal{B}^* = \{\phi_0, \phi_1, \phi_2, \ldots\},$$

where $\phi_i : V \to \mathbb{F}$ is defined by evaluation on monomials:

$$\phi_i(x^j) = \begin{cases} 1, & \text{if } i = j, \\ 0, & \text{otherwise.} \end{cases}$$

Consider a functional $\phi \in V^*$ given by:

$$\phi(p(x)) = p(1),$$

i.e., ϕ evaluates the polynomial at $x = 1$. Then we have:

$$\phi(1) = 1, \quad \phi(x) = 1, \quad \phi(x^2) = 1, \quad \cdots, \quad \phi(x^n) = 1 \quad \text{for all } n \in \mathbb{N}.$$

If ϕ were in the span of $\{\phi_0, \phi_1, \phi_2, \dots\}$, then it could be written as a finite linear combination:

$$g := \sum_{n=0}^{\infty} \phi(x^n)\phi_n = \sum_{n=0}^{\infty} \phi_n.$$

This is a contradiction, since $\text{span}(\mathcal{B}^*)$ only consists of finite linear combinations of the ϕ_i's.

Conclusion: If V is infinite-dimensional, then V^* strictly contains more elements than $\text{span}(\mathcal{B}^*)$. In fact, $\dim(V^*)$ is uncountable, while $\mathcal{B}^*$ is countable.

Any subspace of a given vector space has some gap with the whole space when considered from the perspective of the dual space. We now turn to describing this more precisely.

5.1.2 Annihilators

Definition 5.9 (Annihilator). Let V be a vector space and $S \subseteq V$ a subset. The *annihilator* of S, denoted $\text{Ann}(S)$, is defined as:

$$\text{Ann}(S) = \{f \in V^* \mid f(s) = 0, \ \forall s \in S\}.$$

Example 5.10. Let $V = \mathbb{R}^4$, with basis $B = \{\mathbf{e}_1, \dots, \mathbf{e}_4\}$, and let the dual basis be $\mathcal{B}^* = \{f_1, \dots, f_4\}$. Consider the subset $S = \{\mathbf{e}_3, \mathbf{e}_4\}$.

Then $f_1 \in \text{Ann}(S)$, since

$$f_1(\mathbf{e}_3) = 0, \quad f_1(\mathbf{e}_4) = 0.$$

Indeed, any linear combination $af_1 + bf_2 \in V^*$ such that the coefficients only involve functionals vanishing on S, will be in $\text{Ann}(S)$.

Proposition 5.11 (Properties of Annihilators). *Let V be a vector space. Then:*

1. *The set $\text{Ann}(S) \subseteq V^*$ is a vector subspace.*

2. *The annihilator is **inclusion-reversing**, i.e., if $W_1 \subseteq W_2 \subseteq V$, then*

$$\text{Ann}(W_1) \supseteq \text{Ann}(W_2).$$

3. *The annihilator is **idempotent**, i.e.,*

$$\text{Ann}(S) = \text{Ann}(\text{span}(S)).$$

4. *If V is finite-dimensional and $W \leq V$, then the annihilator of W "fills in the gap":*

$$\dim(W) + \dim(\text{Ann}(W)) = \dim(V).$$

Proof.   1. Suppose that $f, g \in \text{Ann}(S)$, i.e., $f(s) = g(s) = 0, \forall s \in S$. It's clear that $(af + bg) \in$ Ann(S).

2. Suppose that $f \in \text{Ann}(W_2)$, we imply $f(\mathbf{w}) = 0$ for any $\mathbf{w} \in W_2$. Therefore, $f(\mathbf{w}_1) = 0$ for any $\mathbf{w}_1 \in W_1 \subseteq W_2$, i.e., $f \in \text{Ann}(W_1)$.

3. Note that $S \subseteq \text{span}(S)$. Therefore we imply $\text{Ann}(S) \supseteq \text{Ann}(\text{span}(S))$. It suffices to show $\text{Ann}(S) \subseteq \text{Ann}(\text{span}(S))$:

   For any $f \in \text{Ann}(S)$ and any $\sum_{i=1}^n k_i \mathbf{s}_i \in \text{span}(S)$, we imply

$$f\left(\sum_{i=1}^n k_i \mathbf{s}_i\right) = \sum_{i=1}^n k_i f(\mathbf{s}_i) = \sum_{i=1}^n k_i \cdot 0 = 0,$$

   i.e., $f \in \text{Ann}(\text{span}(S))$.

4. Let $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ be a basis of W. By basis extension, we construct a basis of V:

$$\mathcal{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_k, \mathbf{v}_{k+1}, \ldots, \mathbf{v}_n\}.$$

   Let $\mathcal{B}^* = \{f_1, \ldots, f_k, f_{k+1}, \ldots, f_n\}$ be a basis of V^*. We claim that $\{f_{k+1}, \ldots, f_n\}$ is a basis of Ann(W):

   (a) f_j are elements in Ann(W) for $j = k+1, \ldots, n$, since for any $\mathbf{w} = \sum_{i=1}^k \alpha_i \mathbf{v}_i \in W$, we have

$$f_j(\mathbf{w}) = \sum_{i=1}^k \alpha_i f_j(\mathbf{v}_i) = 0.$$

   (b) The set $\{f_{k+1}, \ldots, f_n\}$ is linearly independent.

   (c) $\{f_{k+1}, \ldots, f_n\}$ spans Ann(W): for any $g \in \text{Ann}(W) \leq V^*$, it can be expressed as a linear combination of $\mathcal{B}^*$, i.e. $g = \sum_{i=1}^n \beta_i f_i$. Since $g \in \text{Ann}(W)$

$$0 = g(\mathbf{v}_j) = \sum_{i=1}^n \beta_i f_i(\mathbf{v}_j) = \beta_j,$$

   for all $1 \leq j \leq k$. We conclude

$$g = \beta_{k+1} f_{k+1} + \cdots + \beta_n f_n \in \text{span}\{f_{k+1}, \ldots, f_n\}.$$

   Therefore, $\{f_{k+1}, \ldots, f_n\}$ forms a basis for Ann(W), i.e., $\dim \text{Ann}(W) = n - k$.
   Let $W \leq V$, with $\dim V < \infty$, then

$$\dim(\text{Ann}(W)) = \dim(V) - \dim(W), \quad \dim((V/W)^*) = \dim(V/W) = \dim(V) - \dim(W).$$

So the result follows.   □

5.2   Adjoint Map

In the previous section, if $\dim(V) < \infty$, then both V/W and $\mathrm{Ann}(W)$ has dimensional equal to $\dim(V) - \dim(W)$. One wishes to define a *natural* isomorphism between them without specifying any bases.

In fact, the more natural isomorphism is instead:

$$(V/W)^* \cong \mathrm{Ann}(W),$$

and this isomorphism holds even for infinite dimensional vector spaces!

We now construct the natural isomorphisms $\Phi : \mathrm{Ann}(W) \to (V/W)^*$ and $\Psi : (V/W)^* \to \mathrm{Ann}(W)$ explicitly, so that

$$\Phi \circ \Psi = \mathrm{id}_{(V/W)^*} \qquad \Psi \circ \Phi = \mathrm{id}_{\mathrm{Ann}(W)}.$$

Firstly, for $\Phi : \mathrm{Ann}(W) \to (V/W)^*$, we need to define $\Phi(f)$ for $f \in \mathrm{Ann}(W) \leq V^*$. In other words,

$$f : V \to \mathbb{F} \text{ with } W \leq \ker(f).$$

By Proposition 4.9, one can define a linear transformation $\widetilde{f} : V/W \to \mathbb{F}$ such that the diagram commutes:

$$
\begin{array}{ccc}
V & \xrightarrow{\ \pi\ } & V/W \\
& {\scriptstyle f} \searrow & \Big\downarrow {\scriptstyle \widetilde{f}} \\
& & \mathbb{F}
\end{array}
$$

i.e., $\widetilde{f}(\mathbf{v} + W) = f(\mathbf{v})$. Then $\Phi : \mathrm{Ann}(W) \to (V/W)^*$ is defined by

$$\Phi(f) := \widetilde{f}.$$

Proposition 5.12. *The map Φ is a linear transformation, i.e.,*

$$\Phi(af + bg) = a \cdot \Phi(f) + b \cdot \Phi(g).$$

Proof. It suffices to show that

$$\widetilde{(af + bg)} = a\widetilde{f} + b\widetilde{g}.$$

$\square$

Now we have constructed the natural linear transformation Φ without specifying any bases. To see whether Φ is a bijection, we will construct its inverse Ψ. In order to do so, we need the following:

Definition 5.13 (Adjoint Map)**.** Let $T : V \to W$ be a linear transformation. Define the *adjoint* of T by

$$T^* : W^* \to V^*$$

such that for any $f \in W^*$,

$$[T^*(f)](v) := f(T(v)), \quad \forall v \in V.$$

1. In other words, $T^*(f) = f \circ T$, i.e., a linear transformation from V to $\mathbb{F}$, so it belongs to V^*.

2. Moreover, the mapping T^* itself is a linear transformation. For $f, g \in W^*$, and for all $v \in V$,

$$
\begin{aligned}
[T^*(af + bg)](v) &= (af + bg)[T(v)] \\
&= af(T(v)) + bg(T(v)) \\
&= a[T^*(f)](v) + b[T^*(g)](v) \\
&= [aT^*(f) + bT^*(g)](v).
\end{aligned}
$$

Proposition 5.14. *Let $T : V \to W$ be a linear transformation.*

1. *If T is injective, then T^* is surjective.*

2. *If T is surjective, then T^* is injective.*

This statement is intuitive, since T^* reverses the dual of the output into the dual of the input:

$$T : V \to W, \quad T^* : W^* \to V^*.$$

Proof. We prove only part (2): If T is surjective, then T^* is injective.

Suppose $g \in W^*$ and $T^*(g) = 0_{V^*}$. Then, by definition of the adjoint map:

$$[T^*(g)](v) = g(T(v)) = 0, \quad \forall v \in V. \tag{5.1}$$

We want to show $g = 0_{W^*}$, i.e., $g(w) = 0$ for all $w \in W$.

Since T is surjective, for every $w \in W$, there exists $v' \in V$ such that

$$w = T(v').$$

Substituting into (5.1), we get

$$g(w) = g(T(v')) = 0.$$

Hence, $g = 0_{W^*} \in Ker(T*)$ and T^* is thus injective. $\square$

The following proposition shows the effect of adjoint map on matrix representations:

Proposition 5.15. *Let $T : V \to W$ be a linear transformation, and let $\mathcal{A} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$, $\mathcal{B} = \{\mathbf{w}_1, \ldots, \mathbf{w}_m\}$ be bases of V and W, respectively. Let $\mathcal{A}^* = \{f_1, \ldots, f_n\}$, $\mathcal{B}^* = \{g_1, \ldots, g_m\}$ be the*

dual bases of V^* and W^*, respectively. Then the adjoint $T^* : W^* \to V^*$ has a matrix representation

$$[T^*]_{\mathcal{A}^*, \mathcal{B}^*} = ([T]_{\mathcal{B}, \mathcal{A}})^\mathsf{T},$$

where $[T^*]_{\mathcal{A}^*, \mathcal{B}^*} \in \mathbb{R}^{n \times m}$ and $[T]_{\mathcal{B}, \mathcal{A}} \in \mathbb{R}^{m \times n}$.

Proof. Let $[T]_{\mathcal{B}, \mathcal{A}} = (\alpha_{ij})$ and $[T^*]_{\mathcal{A}^*, \mathcal{B}^*} = (\beta_{ij})$. By definition of the matrix representation:

$$T(\mathbf{v}_j) = \sum_{i=1}^{m} \alpha_{ij} \mathbf{w}_i, \quad T^*(g_j) = \sum_{k=1}^{n} \beta_{kj} f_k \in V^*.$$

Then for any i, j, we compute:

$$[T^*(g_j)](\mathbf{v}_j) = g_j(T(\mathbf{v}_j)) = g_j \left(\sum_{\ell=1}^{m} \alpha_{\ell j} \mathbf{w}_\ell \right) = \sum_{\ell=1}^{m} \alpha_{\ell j} g_j(\mathbf{w}_\ell) = \alpha_{ij}.$$

On the other hand:

$$[T^*(g_j)](\mathbf{v}_j) = \left(\sum_{k=1}^{n} \beta_{kj} f_k \right)(\mathbf{v}_j) = \sum_{k=1}^{n} \beta_{kj} f_k(\mathbf{v}_j) = \beta_{ji}.$$

Therefore, we conclude that $\beta_{ij} = \alpha_{ij}$, so $[T^*] = [T]^\mathsf{T}$. $\qquad \square$

5.2.1 Relationship between Annihilator and Dual of Quotient Spaces

Example 5.16. Consider the canonical projection mapping $\pi_W : V \to V/W$ with its adjoint mapping:

$$(\pi_W)^* : (V/W)^* \to V^*.$$

To understand $(\pi_W)^*$:

1. Take $h \in (V/W)^*$ and study $(\pi_W)^*(h) \in V^*$.

2. Take $\mathbf{v} \in V$ and compute:

$$[(\pi_W)^*(h)](\mathbf{v}) = h(\pi_W(\mathbf{v})) = h(\mathbf{v} + W).$$

In particular, for all $\mathbf{w} \in W \leq V$, we have:

$$[(\pi_W)^*(h)](\mathbf{w}) = h(\mathbf{w} + W) = h(\mathbf{0}_{V/W}) = 0_\mathbb{F}.$$

Therefore,

$$(\pi_W)^*(h) \in \mathrm{Ann}(W).$$

i.e.,

$$(\pi_W)^* : (V/W)^* \to \mathrm{Ann}(W).$$

So $\Psi = (\pi_W)^*$ is our candidate of the inverse of Φ.

Theorem 5.17.

$$\Phi \circ \Psi = \mathrm{id}_{(V/W)^*}, \quad \Psi \circ \Phi = \mathrm{id}_{\mathrm{Ann}(W)}.$$

Consequently, one has a natural isomorphism $(V/W)^* \cong \mathrm{Ann}(W)$ regardless of whether $\dim(V) < \infty$ or not.

Proof. Recall that for $f \in \mathrm{Ann}(W)$, $\Phi(f) \in (V/W)^*$ satisfies:

$$\Phi(f)(\mathbf{v} + W) = f(\mathbf{v}),$$

while for $h \in (V/W)^*$, $\Psi(h) \in \mathrm{Ann}(W)$ satisfies

$$\Psi(h)(\mathbf{v}) = h(\mathbf{v} + W).$$

Therefore, for all $\mathbf{v} + W \in V/W$,

$$(\Phi \circ \Psi(h))(\mathbf{v} + W) = \Psi(h)(\mathbf{v}) = h(\mathbf{v} + W)$$

and hence $\Phi \circ \Psi(h) = h$ for all $h \in (V/W)^*$, i.e. $\Phi \circ \Psi = \mathrm{id}_{(V/W)^*}$.

On the other hand, for all $\mathbf{v} \in V$,

$$(\Psi \circ \Phi(f))(\mathbf{v}) = \Phi(f)(\mathbf{v} + W) = f(\mathbf{v}),$$

which implies $\Psi \circ \Phi(f) = f$ for all $f \in \mathrm{Ann}(W)$. Consequently, $\Psi \circ \Phi = \mathrm{id}_{\mathrm{Ann}(W)}$ and the result follows. $\square$

Chapter 6

Polynomials, Eigenvalues and Eigenvectors

6.1 Basics of Polynomials

We recall some useful properties of polynomials before studying eigenvalues and eigenvectors.

Definition 6.1 (Polynomial)**.**

1. A polynomial over a field $\mathbb{F}$ has the form

$$p(z) = a_m z^m + \cdots + a_1 z + a_0, \quad \text{with } a_m \neq 0.$$

   Here $a_m z^m$ is called the *leading term* of $p(z)$; $\deg(p(z)) = m$ the *degree* of $p(z)$; a_m the *leading coefficient* of $p(z)$; and $a_0, \ldots, a_m$ the *coefficients* of $p(z)$.

2. A polynomial over $\mathbb{F}$ is said to be *monic* if its leading coefficient is $1_{\mathbb{F}}$.

3. A polynomial $p(z) \in \mathbb{F}[z]$ is *irreducible* if whenever

$$p(z) = a(z)b(z)$$

   for $a(z), b(z) \in \mathbb{F}[z]$, then either $\deg(a(z)) = 0$ or $\deg(b(z)) = 0$ (i.e. $a(z)$ or $b(z)$ is a constant polynomial). Otherwise, $p(z)$ is *reducible*.

Example 6.2. The polynomial $p(x) = x^2 + 1$ is irreducible over $\mathbb{R}$, but reducible over $\mathbb{C}$ since

$$p(x) = (x - i)(x + i).$$

6.1.1 Division Algorithm

Theorem 6.3 (Division Theorem)**.** *Let $p, q \in \mathbb{F}[z]$ with $q \neq 0$. Then there exist unique $s, r \in \mathbb{F}[z]$ such that*

$$p(z) = s(z) \cdot q(z) + r(z), \quad \deg(r) < \deg(q).$$

Here $r(z)$ is called the remainder.

Example 6.4. Let $p(x) = x^4 + 1$ and $q(x) = x^2 + 1$. Then, using basic algebra,

$$x^4 + 1 = (x^2 - 1)(x^2 + 1) + 2.$$

Theorem 6.5 (Root Theorem). *Let $p(x) \in \mathbb{F}[x]$, and $\lambda \in \mathbb{F}$. Then $x - \lambda$ divides $p(x)$ if and only if $p(\lambda) = 0$.*

Proof.

($\Rightarrow$) If $x - \lambda$ divides p, then $p = (x - \lambda)q$ for some $q \in \mathbb{F}[x]$. Substituting λ yields

$$p(\lambda) = (\lambda - \lambda)q(\lambda) = 0.$$

($\Leftarrow$) Suppose $p(\lambda) = 0$. By the Division Theorem, there exist $s, r \in \mathbb{F}[x]$ such that

$$p(x) = (x - \lambda)s(x) + r(x), \quad \deg(r) < \deg(x - \lambda) = 1. \tag{6.1}$$

Therefore, $r(x) = r$ is constant. Plugging in $x = \lambda$,

$$0 = p(\lambda) = 0 \cdot s(\lambda) + r \quad \Rightarrow \quad r = 0.$$

Hence $p(x) = (x - \lambda)s(x)$, and $x - \lambda$ divides $p(x)$.

$\square$

Corollary 6.6 (Root Bound). *A polynomial with degree n has at most n roots, counting multiplicity.*

Example 6.7. The polynomial $(x - 3)^2$ has one root $x = 3$ with multiplicity 2. When counting multiplicity, we say it has two roots.

6.1.2   Unique Factorization

Definition 6.8 (Algebraically Closed Field). A field $\mathbb{F}$ is called *algebraically closed* if every non-constant polynomial $p(x) \in \mathbb{F}[x]$ has a root $\lambda \in \mathbb{F}$.

Theorem 6.9 (Fundamental Theorem of Algebra). *The field $\mathbb{C}$ of complex numbers is algebraically closed.*

Sketch. One way to prove this is via complex analysis; another way is using the topology of $\mathbb{C} \setminus \{0\}$.

$\square$

By induction, we can show that every polynomial of degree n over an algebraically closed field $\mathbb{F}$ has exactly n roots, counting multiplicity. Therefore, any polynomial $p(x)$ over such a field can

be factorized as

$$p(x) = c(x - \lambda_1) \cdots (x - \lambda_n) \tag{6.1}$$

for some $c, \lambda_1, \ldots, \lambda_n \in \mathbb{F}$.

Polynomials over general fields $\mathbb{F}$ may not factor as in (6.1), but they still satisfy a unique factorization property. To begin with:

Definition 6.10 (Divisibility)**.** Let $p(x), q(x), s(x) \in \mathbb{F}[x]$ with $p(x) = q(x)s(x)$. Then:

- $p(x)$ is divisible by $s(x)$,

- $s(x)$ is a factor of $p(x)$,

- $s(x) \mid p(x)$,

- $s(x)$ divides $p(x)$,

- $p(x)$ is a multiple of $s(x)$.

Definition 6.11 (Irreducible Polynomials)**.** A polynomial $f(x) \in \mathbb{F}[x]$ is *irreducible* if whenever

$$s(x) | f(x),$$

then either $s(x) = c$ is a constant or $s(x) = cf(x)$ is a scalar multiple of $f(x)$.

Theorem 6.12 (Unique Factorization)**.** *Every* $f(x) = a_n x^n + \cdots + a_0 \in \mathbb{F}[x]$ *can be uniquely written as*

$$f(x) = a_n [p_1(x)]^{e_1} \cdots [p_k(x)]^{e_k}$$

where each p_i is monic, irreducible, and $p_i \neq p_j$ for $i \neq j$. This factorization is unique up to the ordering of the factors.

6.1.3   Greatest Common Divisor and Bézout's Theorem

Definition 6.13 (Common Factor and GCD)**.** Let $f_1, \ldots, f_k \in \mathbb{F}[x]$. A polynomial $g(x)$ is a *common factor* of these if $g \mid f_i$ for all i.

The *greatest common divisor GCD* $g = \gcd(f_1, \ldots, f_k)$ of $f_1, \ldots, f_k$ satisfies:

- g is monic,

- g is a common factor of $f_1, \ldots, f_k$,

- g has maximal degree among all common factors.

If $\gcd(f_1, \ldots, f_k) = 1$, we say the polynomials are *relatively prime.*

With the definition above, it is easy to check that $\gcd(f_1, \ldots, f_k)$ is unique, and

$$\gcd(f_1, \ldots, f_k) = \gcd(\gcd(f_1, f_2), f_3, \ldots, f_k) = \gcd(\gcd(\gcd(f_1, f_2), f_3), \ldots, f_k) = \cdots.$$

Remark 6.14. Polynomials $f_1, \ldots, f_k$ being relatively prime does not imply $\gcd(f_i, f_j) = 1$ for all $i \neq j$.

Example 6.15 (Counter-example). Let $a_1, \ldots, a_n$ be distinct irreducible polynomials, and define

$$f_i(x) := a_1(x) \cdots \widehat{a_i(x)} \cdots a_n(x),$$

where the hat means $a_i(x)$ is omitted. Then $\gcd(f_1, \ldots, f_n) = 1$, but

$$\gcd(f_i, f_j) = a_1 \cdots \widehat{a_i} \cdots \widehat{a_j} \cdots a_n$$

which may not be 1.

Example 6.16. Let

$$f_1(x) = (x^2 + x + 1)^3 (x - 3)^2 x^4, \quad f_2(x) = (x^2 + 1)(x - 3)^4 x^2 \in \mathbb{R}[x],$$

then

$$\gcd(f_1, f_2) = (x - 3)^2 x^2.$$

As mentioned in the paragraph after Definition 6.13, $\gcd(f_1, f_2, f_3, \cdots, f_k)$ can be computed inductively by understanding

$$\gcd(f_1, f_2), \gcd((\gcd(f_1, f_2), f_3)), \gcd(\gcd((\gcd(f_1, f_2), f_3)), f_4), \cdots$$

So it suffices to study the greatest common divisor of two polynomials. In such a case $\gcd(f_1, f_2)$ can be computed using **Euclidean Algorithm** (see Homework set).

One important theorem relating f_1, f_2 and $\gcd(f_1, f_2)$ is the following:

Theorem 6.17 (Bézout's Identity). *Let $g = \gcd(f_1, f_2)$. Then there exist polynomials $r_1, r_2 \in \mathbb{F}[x]$ such that*

$$g(x) = r_1(x)f_1(x) + r_2(x)f_2(x).$$

More generally, for $g = \gcd(f_1, \ldots, f_k)$, there exist $r_1, \ldots, r_k \in \mathbb{F}[x]$ such that

$$g(x) = r_1(x)f_1(x) + \cdots + r_k(x)f_k(x).$$

Example 6.18. Consider the polynomials $f_1(x) = x^3 + 6x + 7$ and $f_2(x) = x^2 + 3x + 2$. Applying the Euclidean algorithm:

$$\begin{aligned}
f_1(x) - (x - 3)f_2(x) &= x^3 + 6x + 7 - (x - 3)(x^2 + 3x + 2) \\
&= x^3 + 6x + 7 - (x^3 + 3x^2 + 2x - 3x^2 - 9x - 6) \\
&= 13x + 13.
\end{aligned}$$

Next,

$$f_2(x) - \frac{x+2}{13}(13x + 13) = x^2 + 3x + 2 - (x+2)(x+1)$$
$$= 0.$$

Therefore,

$$\gcd(f_1, f_2) = \gcd(f_2, 13x + 13) = x + 2.$$

By applying Theorem 6.17 repeatedly, one has the following:

Corollary 6.19. *Let* $g = \gcd(f_1, \ldots, f_k)$. *Then there exists polynomials* $r_1, \ldots, r_k \in \mathbb{F}[x]$ *such that*

$$g(x) = r_1(x)f_1(x) + \cdots + r_k(x)f_k(x).$$

6.2 Eigenvalues & Eigenvectors

Definition 6.20 (Eigenvalue and Eigenspace). Let $T : V \to V$ be a linear operator.

1. A nonzero vector $\mathbf{v} \in V \setminus \{\mathbf{0}\}$ is called an *eigenvector* of T with eigenvalue λ if

$$T(\mathbf{v}) = \lambda \mathbf{v}.$$

2. Equivalently, $\mathbf{v} \in \ker(T - \lambda I)$, where $I = \mathrm{id} : V \to V$ is the identity map, i.e. $I(\mathbf{v}) = \mathbf{v}$ for all $\mathbf{v} \in V$. The subspace $\ker(T - \lambda I)$ is called the λ-*eigenspace* of T.

Definition 6.21 (Generalized Eigenvector). A nonzero vector $\mathbf{v} \in V$ is a *generalized eigenvector* of T with generalized eigenvalue λ if

$$\mathbf{v} \in \ker\left((T - \lambda I)^k\right) \quad \text{for some } k \in \mathbb{N}_{>0}.$$

Remark 6.22. Every eigenvector is a generalized eigenvector, but the converse does not necessarily hold.

Example 6.23. Let $A : \mathbb{R}^2 \to \mathbb{R}^2$ be the linear transformation defined by matrix $A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$..

1. The vector $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ is an eigenvector with eigenvalue 1, since $A\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$..

2. The vector $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$ is not an eigenvector of eigenvalue 1, since $A\begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$. However, we compute:

$$(A - I)^2 = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}^2 = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix},$$

so

$$\begin{pmatrix} 0 \\ 1 \end{pmatrix} \in \ker((A - I)^2),$$

i.e., it is a generalized eigenvector with generalized eigenvalue 1.

Example 6.24. Let $V = C^\infty(\mathbb{R})$ be the space of infinitely differentiable functions. Define the linear operator

$$T(f) = f''.$$

Then $f \in V$ is a (-1)-eigenvector of T if and only if $f'' = -f$. From ODE theory, the general solution is

$$f(x) = a \sin x + b \cos x,$$

so the (-1)-eigenspace is spanned by $\{\sin x, \cos x\}$.

From now on, we assume that V is finite-dimensional unless otherwise stated.

Definition 6.25 (Determinant of a Linear Operator). Let $T : V \to V$ be a linear operator. The determinant of T is defined as

$$\det(T) = \det(T_{\mathcal{A},\mathcal{A}}),$$

where $T_{\mathcal{A},\mathcal{A}}$ is the matrix representation of T with respect to **any** choice of basis $\mathcal{A}$ of V.

Remark 6.26. The determinant is independent of the choice of basis. For any other basis $\mathcal{B}$, we have the following from section 3.2:

$$\det(T_{\mathcal{B},\mathcal{B}}) = \det(C_{\mathcal{B},\mathcal{A}} T_{\mathcal{A},\mathcal{A}} C_{\mathcal{A},\mathcal{B}}) = \det(C_{\mathcal{A},\mathcal{B}}^{-1}) \cdot \det(T_{\mathcal{A},\mathcal{A}}) \cdot \det(C_{\mathcal{A},\mathcal{B}}) = \det(T_{\mathcal{A},\mathcal{A}}).$$

Definition 6.27 (Characteristic Polynomial). Let $T : V \to V$ be a linear operator. The *characteristic polynomial* $\mathcal{X}_T(x) \in \mathbb{F}[x]$ of T is defined by

$$\mathcal{X}_T(x) := \det\left(T_{\mathcal{A},\mathcal{A}} - xI\right),$$

where $T_{\mathcal{A},\mathcal{A}}$ is the matrix representation of T with respect to any basis $\mathcal{A}$ of V, and I is the identity matrix of the same size.

Remark 6.28. The above definition is independent of the choice of basis of V. Indeed, let $\mathcal{B}$ be another basis of V, then

$$T_{\mathcal{B},\mathcal{B}} = M^{-1} T_{\mathcal{A},\mathcal{A}} M$$

for $M := C_{\mathcal{A},\mathcal{B}}$. As a result, one has:

$$\det(T_{\mathcal{B},\mathcal{B}} - xI) = \det(M^{-1} T_{\mathcal{A},\mathcal{A}} M - xM^{-1}IM) = \det(M)^{-1} \det(T_{\mathcal{A},\mathcal{A}} - xI) \det(M) = \det(T_{\mathcal{A},\mathcal{A}} - xI).$$

In the next couple of chapters, we will study two fundamental theorems that reveal the internal structure of linear operators:

- Cayley–Hamilton Theorem

- Jordan Canonical Form

These results are usually stated in terms of matrices, but they are truly theorems about linear operators—independent of the choice of basis. Our goal is to understand them in the more general setting of finite-dimensional vector spaces, beyond just $\mathbb{R}^n$ or $\mathbb{C}^n$.

Remark 6.29. Why do we care about these theorems?

1. The **Cayley–Hamilton Theorem** tells us that every linear operator satisfies its own characteristic polynomial. This bridges algebraic information (the determinant and characteristic polynomial) with operator action.

2. The **Jordan Canonical Form** provides a refined classification of linear operators up to similarity. It allows us to break down a complicated operator into blocks that reveal both eigenvalues and the structure of generalized eigenspaces.

3. Understanding these theorems gives insight into the "shape" of an operator—how it stretches, rotates, or shears a space—and how we can simplify its action by choosing the right basis.

Chapter 7

Minimal Polynomial & Cayley-Hamilton Theorem

7.1 Minimal Polynomial

Definition 7.1 (Linear Operator Induced From Polynomial)**.** Let $f(x) := a_m x^m + \cdots + a_0 \in \mathbb{F}[x]$, and let $T : V \to V$ be a linear operator. Then the operator

$$f(T) := a_m T^m + \cdots + a_1 T + a_0 I : V \to V$$

is called the linear operator induced by the polynomial $f(x)$.

The composition of linear operators is not commutative, i.e. $S \circ T \neq T \circ S$ in general. This follows similarly to the fact that matrix multiplication is not commutative. However, we always have

$$f(T)T = Tf(T),$$

for any polynomial $f \in \mathbb{F}[x]$ - Indeed, we can show that $T^n T = TT^n$ for all n by induction. Now for $f(x) = \sum_i a_i x^i$, then

$$f(T)T = \sum_i a_i T^i T = \sum_i a_i TT^i = T \sum_i a_i T^i = Tf(T).$$

More generally, one can apply induction argument again to show that the composition of polynomial-induced operators is commutative:

$$f(T)g(T) = g(T)f(T),$$

for all $f(x), g(x) \in \mathbb{F}[x]$.

Definition 7.2 (Minimal Polynomial)**.** Let $T : V \to V$ be a linear operator. The *minimal polyno-*

mial $m_T(x)$ is the unique monic polynomial of least degree such that

$$m_T(T) = \mathbf{0}_{V \to V},$$

where $\mathbf{0}_{V \to V}$ denotes the zero operator in $\mathrm{Hom}(V, V)$.

Example 7.3 (Minimal Polynomials of Matrices)**.**

1. Let $A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, then

$$\mathcal{X}_A(x) = (x-1)^2, \quad m_A(x) = x - 1,$$

   since $A - I = 0$.

2. Let $B = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$, then

$$\mathcal{X}_B(x) = (x-1)^2.$$

   Could the minimal polynomial be $x - 1$? No, because

$$B - I = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \neq 0.$$

   However,

$$(B - I)^2 = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \Rightarrow m_B(x) = (x-1)^2.$$

Two natural questions arise:

1. Does $m_T(x)$ always exist? Is it unique?

2. What is the relationship between $m_T(x)$ and $\mathcal{X}_T(x)$?

Regarding the existence of minimal polynomial, note that $m_T(x)$ may not exist if V is infinite-dimensional:

Example 7.4. Let $V = \mathbb{R}[x]$, and define

$$T : V \to V, \quad p(x) \mapsto \int_0^x p(t)\, dt.$$

Then, for monomials: $T(x^n) = \frac{1}{n+1}x^{n+1}$. Suppose $m_T(x) = x^n + \cdots + a_0$, then

$$m_T(T) = T^n + \cdots + a_0 I = 0.$$

Applying this to $x \in V$, we would have:

$$m_T(T)(x) = \frac{1}{n!}x^n + a_{n-1}\frac{1}{(n-1)!}x^{n-1} + \cdots + a_0 x = 0.$$

But this is a contradiction since the LHS is a nonzero polynomial while the RHS is zero. Hence, no such minimal polynomial exists.

7.1.1 Properties of the Minimal Polynomial

Proposition 7.5 (Existence). *The minimal polynomial $m_T(x)$ always exists if $\dim(V) = n < \infty$.*

Proof. We consider the set

$$\{I, T, T^2, \ldots, T^{n^2}\} \subseteq \mathrm{Hom}(V, V),$$

which lies in a vector space $\mathrm{Hom}(V, V)$ of dimension n^2. Therefore, this set is linearly dependent, and there exist scalars $a_0, \ldots, a_{n^2}$, not all zero, such that

$$a_0 I + a_1 T + \cdots + a_{n^2} T^{n^2} = 0.$$

That is, there exists a nonzero polynomial $g(x) \in \mathbb{F}[x]$ with $\deg(g) < n^2$ such that $g(T) = 0$. Thus a minimal-degree such polynomial exists. $\qquad\square$

Proposition 7.6 (Uniqueness). *The minimal polynomial $m_T(x)$, if it exists, is unique.*

Proof. Suppose $f_1(x) \neq f_2(x)$ are two minimal polynomials with the same minimal degree. Then their difference

$$g(x) = f_1(x) - f_2(x) \neq 0,$$

satisfies $\deg(g) < \deg(f_1)$ and

$$g(T) = f_1(T) - f_2(T) = 0.$$

After scaling $g(x)$ to be monic, we obtain a polynomial of strictly lower degree annihilating T, contradicting the minimality of f_1. Hence $m_T(x)$ is unique. $\qquad\square$

Proposition 7.7 (Divisibility). *Let $f(x) \in \mathbb{F}[x]$ satisfy $f(T) = 0$. Then*

$$m_T(x) \mid f(x).$$

Proof. By the division algorithm, write

$$f(x) = q(x) m_T(x) + r(x), \quad \deg(r) < \deg(m_T).$$

Then

$$f(T) = q(T) m_T(T) + r(T) = 0 + r(T) = r(T).$$

Since $f(T) = 0$, we get $r(T) = 0$. But then $r(x)$ is a polynomial of smaller degree than m_T that annihilates T, contradicting the minimality of m_T. Thus $r(x) \equiv 0$, and $m_T(x) \mid f(x)$. $\qquad\square$

Proposition 7.8 (Similarity Invariance)**.** *Let $A, B \in \mathbb{F}^{n \times n}$ be similar, i.e., $B = P^{-1}AP$ for some invertible matrix P. Then*

$$m_A(x) = m_B(x).$$

Proof. Let $m_A(x) = x^k + a_{k-1}x^{k-1} + \cdots + a_0$. Then

$$m_A(B) = m_A(P^{-1}AP) = P^{-1}m_A(A)P = P^{-1}(0)P = 0.$$

Hence $m_A(x)$ annihilates B, so by Proposition 7.7, $m_B(x) \mid m_A(x)$. A symmetric argument shows $m_A(x) \mid m_B(x)$. Since both are monic, we conclude

$$m_A(x) = m_B(x).$$

$\square$

Remark 7.9. Proposition 7.8 states that the minimal polynomial is invariant under similarity. In fact, the characteristic polynomial is similarity-invariant as well.

We now assume $\dim(V) < \infty$. The minimal polynomial $m_T(x)$ denotes the unique monic polynomial of least degree such that

$$m_T(T) = 0 \in \operatorname{Hom}(V, V).$$

7.2   Cayley–Hamilton Theorem

7.2.1   Statement of the Theorem

We begin with an example: consider the matrix $A = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$, and its associated linear operator on $\mathbb{F}^2$. The characteristic polynomial of A is

$$\chi_A(x) = (x - 1)(x - 2).$$

Note that the minimal polynomial $m_A(x)$ cannot have degree one. Otherwise, $m_A(x) = x - k$ for some scalar k, and we would have

$$m_A(A) = A - kI = \begin{pmatrix} 1 - k & 0 \\ 0 & 2 - k \end{pmatrix} \neq 0,$$

for any $k \in \mathbb{F}$, contradicting the definition of minimal polynomial. Therefore, $m_A(x)$ must be

$$m_A(x) = (x - 1)(x - 2),$$

which coincides with the characteristic polynomial. In fact, one has:

Theorem 7.10 (Cayley–Hamilton). *Let $T : V \to V$ be a linear operator on a finite-dimensional vector space over a field $\mathbb{F}$. Then the characteristic polynomial $\mathcal{X}_T(x)$ annihilates T, i.e.,*

$$\mathcal{X}_T(T) = 0.$$

Equivalently, the minimal polynomial $m_T(x)$ divides the characteristic polynomial:

$$m_T(x) \mid \mathcal{X}_T(x).$$

In the course of the proof of Theorem 7.10, we would assume $\mathcal{X}_T(x) = (x - \lambda_1)^{e_1} \cdots (x - \lambda_k)^{e_k}$ can be factorized into linear terms. However, this is not always possible over arbitrary $\mathbb{F}$. For instance, consider

$$A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \in M_{2\times 2}(\mathbb{R}).$$

Its characteristic polynomial is $\mathcal{X}_A(x) = x^2 + 1$, which cannot be factorized into linear terms in $\mathbb{R}[x]$. However, we may always extend the field to its **algebraic closure**

$$\overline{\mathbb{F}} \supseteq \mathbb{F}$$

(e.g. $\mathbb{C} \supseteq \mathbb{R}$) so that $\mathcal{X}_A(x) \in \mathbb{F}[x] \subseteq \overline{\mathbb{F}}[x]$ always admits the factorization (e.g. $\mathcal{X}_A(x) = (x-i)(x+1)$ in $\mathbb{C}[x]$).

Our strategy to prove Theorem 7.10 is:

- Consider the case where $m_T(x), \mathcal{X}_T(x)$ are in $\overline{\mathbb{F}}[x]$

- Show that $m_T(x) \mid \mathcal{X}_T(x)$ under $\overline{\mathbb{F}}[x]$.

- Since both $m_T(x)$ and $\mathcal{X}_T(x)$ are in $\mathbb{F}[x]$, one must also have $m_T(x) \mid \mathcal{X}_T(x)$ under $\mathbb{F}[x]$.

7.2.2 Invariant Subspace

Before we give the full proof, we develop tools that allow us to decompose the vector space using invariant subspaces.

Definition 7.11 (Invariant Subspace). Let $T : V \to V$ be a linear operator. A subspace $W \leq V$ is called T-invariant if

$$T(W) \leq W.$$

Remark 7.12. If $W \leq V$ is T-invariant, then the restriction $T|_W : W \to W$ defines a well-defined linear operator.

Example 7.13. The following are examples of T-invariant subspaces:

1. The whole space V is trivially T-invariant.

2. For each eigenvalue λ, the eigenspace $\ker(T - \lambda I)$ is T-invariant.

3. More generally, for any polynomial $g \in \mathbb{F}[x]$, the space

$$U = \ker(g(T)) \leq V$$

is T-invariant.

Proof: If $\mathbf{v} \in \ker(g(T))$, then $g(T)(\mathbf{v}) = 0$. We want to show $T(\mathbf{v}) \in \ker(g(T))$, i.e., $g(T)(T(\mathbf{v})) = 0$. Observe:

$$g(T)(T(\mathbf{v})) = (a_m T^m + \cdots + a_1 T + a_0 I)(T(\mathbf{v})) = T(g(T)(\mathbf{v})) = T(\mathbf{0}) = \mathbf{0}.$$

4. If $\mathbf{v} \in \ker(T - \lambda I)$, then $\text{span}\{\mathbf{v}\}$ is a T-invariant subspace.

Proposition 7.14. *Let $T : V \to V$ be a linear operator, and suppose $W \leq V$ is a T-invariant subspace. Then we have the induced linear operators:*

$$T|_W : W \to W, \quad \mathbf{w} \mapsto T(\mathbf{w}) \tag{7.1}$$

$$\widetilde{T} : V/W \to V/W, \quad \mathbf{v} + W \mapsto T(\mathbf{v}) + W \tag{7.2}$$

(The well-definedness of $\widetilde{T}$ is proved in Homework 2, Exercise 4.)

With respect to a suitable basis, the matrix representation of T has the block form:

$$[T]_{\mathcal{B},\mathcal{B}} = \begin{pmatrix} [T|_W]_{\mathcal{C},\mathcal{C}} & * \\ 0 & [\widetilde{T}]_{\overline{\mathcal{B}},\overline{\mathcal{B}}} \end{pmatrix}$$

where $\mathcal{C} = \{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ is a basis of W, and $\overline{\mathcal{B}} = \{\mathbf{v}_{k+1} + W, \ldots, \mathbf{v}_n + W\}$ is a basis of V/W.

It follows that the characteristic polynomial of T satisfies:

$$\mathcal{X}_T(x) = \mathcal{X}_{T|_W}(x) \cdot \mathcal{X}_{\widetilde{T}}(x).$$

7.2.3   Triangularizable Transformation

As mentioned before, we will now focus on the case when

$$\mathcal{X}_T(x) = (x - \lambda_1) \cdots (x - \lambda_n) \in \mathbb{F}[x]$$

can be factorized into linear terms (for instance, when $\mathbb{F} = \overline{\mathbb{F}}$ is algebraically closed.

Proposition 7.15. *Suppose*
$$\mathcal{X}_T(x) = (x - \lambda_1) \cdots (x - \lambda_n)$$

where the $\lambda_i \in \mathbb{F}$ are not necessarily distinct. Then there exists a basis $\mathcal{A}$ of V such that

$$T_{\mathcal{A},\mathcal{A}} = \begin{pmatrix} \lambda_1 & * & \cdots & * \\ 0 & \lambda_2 & \cdots & * \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}.$$

*In such a case, we call T a **triangularizable** linear transformation.*

Obviously, the converse of the above proposition also holds, i.e. if $T : V \to V$ is triangularizable, then $\mathcal{X}_T(x) = (x - \lambda_1) \cdots (x - \lambda_n)$ can be factorized into linear terms. We will not make use of this fact in the proof of Theorem 7.10, though.

Proof. We proceed by induction on $n = \dim V$.

Base Case: For $n = 1$, the matrix of T is scalar (λ_1), and the statement holds.

Induction Hypothesis: Assume the result holds for all dimensions $< n$. We show it holds for $\dim V = n$.

Step 1: Existence of eigenvector. Let λ_1 be an eigenvalue of T. By the fundamental theorem of algebra (applied over $\mathbb{C}$), there exists an eigenvector $\mathbf{v} \in V \setminus \{\mathbf{0}\}$ such that

$$T(\mathbf{v}) = \lambda_1 \mathbf{v}.$$

Step 2: Invariant subspace and quotient. Define $W = \mathrm{span}\{\mathbf{v}\} \leq V$, which is T-invariant. By Proposition 7.14, T induces a well-defined operator $\widetilde{T} : V/W \to V/W$ with characteristic polynomial

$$\mathcal{X}_{\widetilde{T}}(x) = (x - \lambda_2) \cdots (x - \lambda_n).$$

Step 3: Apply induction. By the induction hypothesis, there exists a basis $\overline{\mathcal{C}} = \{\mathbf{w}_2 + W, \ldots, \mathbf{w}_n + W\}$ of V/W such that

$$\widetilde{T}_{\overline{\mathcal{C}},\overline{\mathcal{C}}} = \begin{pmatrix} \lambda_2 & * & \cdots & * \\ 0 & \lambda_3 & \cdots & * \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}.$$

Step 4: Lift basis to V. Define $\mathcal{A} := \{\mathbf{v}, \mathbf{w}_2, \ldots, \mathbf{w}_n\}$. Then $\mathcal{A}$ is a basis of V (see Homework 2, Exercise 2), and in this basis we have:

$$T_{\mathcal{A},\mathcal{A}} = \begin{pmatrix} \lambda_1 & * \\ 0 & \widetilde{T}_{\overline{\mathcal{C}},\overline{\mathcal{C}}} \end{pmatrix} = \begin{pmatrix} \lambda_1 & * & * & * \\ 0 & \lambda_2 & \cdots & * \\ 0 & \cdots & \ddots & * \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}$$

as given in Proposition 7.14. So the result follows.   □

Proposition 7.16. *Suppose that the characteristic polynomial of a linear operator T is*

$$\mathcal{X}_T(x) = (x - \lambda_1) \cdots (x - \lambda_n) \in \mathbb{F}[x].$$

Then

$$\mathcal{X}_T(T) = \mathbf{0}_{V \to V}.$$

In other words, Theorem 7.10 holds for T.

Proof. Since $\mathcal{X}_T(x) = (x - \lambda_1) \cdots (x - \lambda_n)$, we imply T is triangularizable under some basis $\mathcal{A}$, i.e.

$$T_{\mathcal{A},\mathcal{A}} = \begin{pmatrix} \lambda_1 & \times & \cdots & \times \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \times \\ 0 & \cdots & 0 & \lambda_n \end{pmatrix}.$$

Note that

- $* \mapsto (*)_{\mathcal{A},\mathcal{A}}$ is an isomorphism between $\mathrm{Hom}(V, V)$ and $M_{n \times n}(\mathbb{F})$ (c.f. Theorem 3.3 with $S = T$ and $\mathcal{A} = \mathcal{B} = \mathcal{C}$).
- $(T^m)_{\mathcal{A},\mathcal{A}} = [T_{\mathcal{A},\mathcal{A}}]^m$ for any m (c.f. Proposition 3.14).

Therefore, if $\mathcal{X}_T(x) = x^n + \alpha_{n-1}x^{n-1} + \cdots + \alpha_1 x + \alpha_0$, then

$$\begin{aligned} [\mathcal{X}_T(T)]_{\mathcal{A},\mathcal{A}} &= [T^n + \alpha_{n-1}T^{n-1} + \cdots + \alpha_1 T + \alpha_0 I]_{\mathcal{A},\mathcal{A}} \\ &= (T_{\mathcal{A},\mathcal{A}})^n + \alpha_{n-1}(T_{\mathcal{A},\mathcal{A}})^{n-1} + \cdots + \alpha_1 T_{\mathcal{A},\mathcal{A}} + \alpha_0 I_{\mathcal{A},\mathcal{A}} \\ &= \mathcal{X}_T(T_{\mathcal{A},\mathcal{A}}) \end{aligned}$$

Therefore,

$$\mathcal{X}_T(T) = \mathbf{0}_{V \to V} \quad \Leftrightarrow \quad [\mathcal{X}_T(T)]_{\mathcal{A},\mathcal{A}} = \mathbf{0}_{n \times n} \quad \Leftrightarrow \quad \mathcal{X}_T(T_{\mathcal{A},\mathcal{A}}) = \mathbf{0}_{n \times n},$$

and it suffices to show $\mathcal{X}_T(T_{\mathcal{A},\mathcal{A}})$ is the zero matrix. To see so, note that

$$\mathcal{X}_T(T_{\mathcal{A},\mathcal{A}}) = (T_{\mathcal{A},\mathcal{A}} - \lambda_1 \mathbf{I}) \cdots (T_{\mathcal{A},\mathcal{A}} - \lambda_n \mathbf{I}),$$

where $T_{\mathcal{A},\mathcal{A}}$ is as given above. Write

$$M_i := (T_{\mathcal{A},\mathcal{A}} - \lambda_i \mathbf{I}) = \begin{pmatrix} \lambda_1 - \lambda_i & \times & \cdots & \cdots & \times \\ 0 & \ddots & \ddots & & \vdots \\ \vdots & \ddots & \mathbf{0} & \ddots & \vdots \\ \vdots & & \ddots & \ddots & \times \\ 0 & \cdots & \cdots & 0 & \lambda_n - \lambda_i \end{pmatrix}$$

so that $\mathcal{X}_T (T_{\mathcal{A},\mathcal{A}}) = M_1 \cdots M_i \cdots M_n$. Then

$$\mathcal{X}_T (T_{\mathcal{A},\mathcal{A}}) \begin{pmatrix} x_1 \\ \vdots \\ x_{n-1} \\ x_n \end{pmatrix} = M_1 \cdots M_{n-1} M_n \begin{pmatrix} x_1 \\ \vdots \\ x_{n-1} \\ x_n \end{pmatrix}$$

$$= M_1 \cdots M_{n-1} \begin{pmatrix} \lambda_1 - \lambda_n & \times & \cdots & \times \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \lambda_{n-1} - \lambda_n & \times \\ 0 & \cdots & 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_{n-1} \\ x_n \end{pmatrix}$$

$$= M_1 \cdots M_{n-1} \begin{pmatrix} x_1' \\ \vdots \\ x_{n-1}' \\ 0 \end{pmatrix}$$

So we 'kill off' the last coordinate of $\mathbf{v} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$. Continuing the process, one will kill of the $n-1$, $n-2$, $\ldots$ until the top entry of $\mathbf{v}$, i.e.

$$\mathcal{X}_T (T_{\mathcal{A},\mathcal{A}}) \mathbf{v} = M_n \cdots M_1 \mathbf{v} = \mathbf{0}.$$

Consequently, $\mathcal{X}_T (T_{\mathcal{A},\mathcal{A}})$ is a zero matrix. $\qquad \square$

7.3   Proof of Cayley–Hamilton Theorem

Now we are ready to give a proof for the Cayley-Hamilton Theorem:

Theorem 7.17 (Cayley–Hamilton). *Let $T : V \to V$ be a linear operator on a finite-dimensional vector space over a field $\mathbb{F}$. Then the characteristic polynomial $\mathcal{X}_T(x)$ annihilates T, i.e.,*

$$\mathcal{X}_T(T) = \mathbf{0}.$$

Proof. Suppose that $\mathcal{X}_T(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_0 \in \mathbb{F}[x]$. By considering an algebraically closed field $\overline{\mathbb{F}} \supseteq \mathbb{F}$, we write

$$\mathcal{X}_T(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_0, \quad a_i \in \mathbb{F} \tag{7.3}$$

$$= (x - \lambda_1)\cdots(x - \lambda_n), \quad \lambda_i \in \overline{\mathbb{F}}. \tag{7.4}$$

By applying Proposition 7.16, we conclude $\mathcal{X}_T(T) = \mathbf{0}$, where the coefficients in the formula $\mathcal{X}_T(T) = \mathbf{0}$ with respect to T lie in $\overline{\mathbb{F}}$.

Then we argue that these coefficients are actually in $\mathbb{F}$. Expand the full expression for $\mathcal{X}_T(T)$:

$$\mathcal{X}_T(T) = (T - \lambda_1 I)\cdots(T - \lambda_n I) \tag{7.5}$$

$$= T^n - (\lambda_1 + \cdots + \lambda_n)T^{n-1} + \cdots + (-1)^n \lambda_1 \cdots \lambda_n I \tag{7.6}$$

$$= T^n + a_{n-1}T^{n-1} + \cdots + a_0 I. \tag{7.7}$$

The equality (7.7) holds because the coefficients of the polynomials in (7.3) and (7.4) are equal.

Therefore, we conclude that $\mathcal{X}_T(T) = 0$ over the field $\mathbb{F}$. $\qquad\square$

Corollary 7.18. *$m_T(x) \mid \mathcal{X}_T(x)$. More precisely, if*

$$\mathcal{X}_T(x) = [p_1(x)]^{e_1} \cdots [p_k(x)]^{e_k}, \quad e_i > 0 \text{ for all } i,$$

where p_i are distinct, monic, and irreducible polynomials, then

$$m_T(x) = [p_1(x)]^{f_1} \cdots [p_k(x)]^{f_k}, \quad \text{for some } 0 < f_i \le e_i \text{ for all } i.$$

Proof. The statement $m_T(x) \mid \mathcal{X}_T(x)$ follows from Theorem 7.17 and Proposition 7.7. Therefore, $0 < f_i \le e_i$ for all i.

Suppose on the contrary that $f_i = 0$ for some i. Without loss of generality, let $i = 1$.

It's clear that $\gcd(p_1, p_j) = 1$ for all $j \ne 1$, which implies

$$a(x)p_1(x) + b(x)p_j(x) = 1, \quad \text{for some } a(x), b(x) \in \mathbb{F}[x].$$

Considering the field extension $\overline{\mathbb{F}} \supseteq \mathbb{F}$, we have

$$p_1(x) = (x - \mu_1)\cdots(x - \mu_\ell).$$

For any root μ_m of p_1, $m = 1, \ldots, \ell$, we have

$$a(\mu_m)p_1(\mu_m) + b(\mu_m)p_j(\mu_m) = 1 \Rightarrow b(\mu_m)p_j(\mu_m) = 1 \Rightarrow p_j(\mu_m) \ne 0,$$

i.e., μ_m is not a root of p_j for all $j \ne 1$.

Therefore, μ_m is a root of $\mathcal{X}_T(x)$ but not a root of $m_T(x)$. Then μ_m is an eigenvalue of T, i.e.,

$T\mathbf{v} = \mu_m \mathbf{v}$ for some $\mathbf{v} \neq 0$.

Recall that $m_{T,\mathbf{v}} = x - \mu_m$, so $m_{T,\mathbf{v}} = x - \mu_m \mid m_T(x)$, which contradicts the assumption $f_1 = 0$. $\qquad\square$

Example 7.19. We can use Corollary 7.18, a stronger version of the Cayley–Hamilton Theorem, to determine minimal polynomials:

1. For the matrix $A = \begin{pmatrix} 0 & -1 \\ 1 & 1 \end{pmatrix}$, we compute

$$\mathcal{X}_A(x) = \left(x^2 + x + 1\right)^1.$$

   Since $x^2 + x + 1$ is irreducible over $\mathbb{R}$, we conclude

$$m_A(x) = x^2 + x + 1.$$

2. For the matrix

$$A = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \end{pmatrix},$$

   we compute

$$\mathcal{X}_A(x) = (x-1)^2(x-2)^2.$$

   By Corollary 7.18, both $(x-1)$ and $(x-2)$ must divide $m_A(x)$, so the minimal polynomial must be one of:

$$(x-1)^2(x-2)^2, \quad (x-1)(x-2)^2, \quad (x-1)^2(x-2), \quad (x-1)(x-2).$$

   By trial and error, we find that

$$m_A(x) = (x-1)^2(x-2).$$

Chapter 8

Primary Decomposition and Jordan Normal Form

We begin this chapter by re-defining the well-known notion of diagonalizability for linear operators:

Definition 8.1 (Diagonalizable). The linear operator $T : V \to V$ is *diagonalizable* over $\mathbb{F}$ if and only if there exists a basis $\mathcal{A}$ of V such that

$$T_{\mathcal{A},\mathcal{A}} = \mathrm{diag}(\lambda_1, \ldots, \lambda_n),$$

where the λ_i's are not necessarily distinct.

Proposition 8.2. *If the linear operator $T : V \to V$ is diagonalizable, then its minimal polynomial is*

$$m_T(x) = (x - \mu_1) \cdots (x - \mu_k),$$

where the μ_i's are distinct.

Proof. Suppose T is diagonalizable. Then there exists a basis $\mathcal{A}$ of V such that

$$T_{\mathcal{A},\mathcal{A}} = \mathrm{diag}(\underbrace{\mu_1, \ldots, \mu_1}_{r_1}, \underbrace{\mu_2, \ldots, \mu_2}_{r_2}, \ldots, \underbrace{\mu_k, \ldots, \mu_k}_{r_k}),$$

for some multiplicities $r_i > 0$ and distinct scalars $\mu_i \in \mathbb{F}$. Then

$$(T_{\mathcal{A},\mathcal{A}} - \mu_1 I) \cdots (T_{\mathcal{A},\mathcal{A}} - \mu_k I) = 0,$$

i.e., $m_T(x) \mid (x - \mu_1) \cdots (x - \mu_k)$.

To show minimality, suppose we omit any factor $(x - \mu_i)$ from the product. Then the matrix

$$(T_{\mathcal{A},\mathcal{A}} - \mu_1 I) \cdots (\widehat{T_{\mathcal{A},\mathcal{A}} - \mu_i I}) \cdots (T_{\mathcal{A},\mathcal{A}} - \mu_k I)$$

is nonzero, since all μ_i's are distinct and appear on the diagonal. Hence, no proper sub-product

annihilates T, and we conclude

$$m_T(x) = (x - \mu_1) \cdots (x - \mu_k).$$

□

8.1 Primary Decomposition Theorem

The converse of Proposition 8.2 is also true, which is a special case of the Primary Decomposition Theorem.

Theorem 8.3 (Primary Decomposition Theorem). *Let $T : V \to V$ be a linear operator with minimal polynomial*

$$m_T(x) = [p_1(x)]^{e_1} \cdots [p_k(x)]^{e_k},$$

where $p_i(x)$ are distinct, monic, and irreducible polynomials over $\mathbb{F}$. For each $i = 1, \ldots, k$, define the subspace

$$V_i = \ker\left([p_i(T)]^{e_i}\right) \leq V.$$

Then the following hold:

1. *Each V_i is T-invariant, i.e.,*

$$T(V_i) \leq V_i.$$

2. *The space V decomposes as a direct sum:*

$$V = V_1 \oplus V_2 \oplus \cdots \oplus V_k.$$

3. *The minimal polynomial of the restriction $T|_{V_i} : V_i \to V_i$ is*

$$m_{T|_{V_i}}(x) = [p_i(x)]^{e_i}.$$

Proof. 1. This follows from Example 7.13(3).

2. Define

$$q_i(x) := \frac{m_T(x)}{p_i(x)^{e_i}} = \prod_{\substack{j=1 \\ j \neq i}}^{k} p_j(x)^{e_j}.$$

Then:

- $\gcd(q_1, \ldots, q_k) = 1$,
- $\gcd(q_i, p_i^{e_i}) = 1$,
- $q_i \cdot p_i^{e_i} = m_T$,
- $m_T(x) \mid q_i(x)q_j(x)$ for $i \neq j$.

By Bézout's Identity, there exist polynomials $a_1(x), \ldots, a_k(x) \in \mathbb{F}[x]$ such that

$$a_1(x)q_1(x) + \cdots + a_k(x)q_k(x) = 1. \tag{8.1}$$

Evaluating at T, we obtain

$$a_1(T)q_1(T) + \cdots + a_k(T)q_k(T) = I.$$

Thus, for any $\mathbf{v} \in V$,

$$\mathbf{v} = \sum_{i=1}^{k} a_i(T)q_i(T)\mathbf{v}, \tag{8.2}$$

and each term $a_i(T)q_i(T)\mathbf{v} \in V_i$. Therefore,

$$V = V_1 + \cdots + V_k. \tag{8.3}$$

To show the sum is direct, suppose

$$\mathbf{0} = \mathbf{v}_1' + \cdots + \mathbf{v}_k', \quad \text{where } \mathbf{v}_i' \in V_i. \tag{8.4}$$

Fix any index i. Since $\gcd(q_i, p_i^{e_i}) = 1$, Bézout's identity gives polynomials $b_i(x), c_i(x)$ such that

$$b_i(x)q_i(x) + c_i(x)p_i(x)^{e_i} = 1 \quad \Rightarrow \quad b_i(T)q_i(T) + c_i(T)p_i(T)^{e_i} = I.$$

Applying both sides to $\mathbf{v}_i'$, and using $p_i(T)^{e_i}\mathbf{v}_i' = 0$, we get

$$b_i(T)q_i(T)\mathbf{v}_i' = \mathbf{v}_i'.$$

Apply $b_i(T)q_i(T)$ to both sides of (8.4):

$$\mathbf{0} = b_i(T)q_i(T)(\mathbf{v}_1' + \cdots + \mathbf{v}_k') = \sum_{j=1}^{k} b_i(T)q_i(T)\mathbf{v}_j'.$$

But for $j \neq i$, $q_i(T)\mathbf{v}_j' = 0$, since $p_j(x)^{e_j} \mid q_i(x)$. So:

$$\mathbf{0} = b_i(T)q_i(T)\mathbf{v}_i' = \mathbf{v}_i' \quad \Rightarrow \quad \mathbf{v}_i' = 0.$$

Since this holds for each i, the sum is direct:

$$V = V_1 \oplus \cdots \oplus V_k. \tag{8.5}$$

3. Since $p_i(T)^{e_i}\mathbf{v} = 0$ for any $\mathbf{v} \in V_i$, it follows that $m_{T|_{V_i}}(x) \mid p_i(x)^{e_i}$. By Corollary 7.18, we

have

$$m_{T|_{V_i}}(x) = p_i(x)^{f_i}, \quad \text{for some } 1 \le f_i \le e_i.$$

Suppose $f_i < e_i$ for some i. Let $\mathbf{v} := \sum_{j=1}^{k} \mathbf{v}_j$, with $\mathbf{v}_j \in V_j$, and define

$$f(x) := \prod_{j=1}^{k} p_j(x)^{f_j}.$$

Then $f(T)\mathbf{v} = 0$, so $f(T) = 0$ on all of V, and hence $m_T(x) \mid f(x)$, contradicting the minimality of $m_T(x)$. Therefore,

$$m_{T|_{V_i}}(x) = p_i(x)^{e_i}. \qquad \square$$

Now we can prove the converse of Proposition 8.2:

Corollary 8.4. *Let the minimal polynomial of T be*

$$m_T(x) = (x - \mu_1) \cdots (x - \mu_k),$$

with $\mu_i \in \mathbb{F}$ distinct. Then T is diagonalizable over $\mathbb{F}$.

Proof. By the Primary Decomposition Theorem , we have

$$V = \bigoplus_{i=1}^{k} V_i, \quad \text{where } V_i := \ker(T - \mu_i I).$$

In other words, each V_i is the μ_i-eigenspace of T. Let $\mathcal{B}_i$ be a basis of V_i, and define $\mathcal{B} := \bigcup_{i=1}^{k} \mathcal{B}_i$. Then $\mathcal{B}$ is a basis of V consisting entirely of eigenvectors of T.

With respect to $\mathcal{B}$, the matrix representation of T is block-diagonal:

$$T_{\mathcal{A},\mathcal{A}} = \mathrm{diag}(\underbrace{\mu_1, \ldots, \mu_1}_{\dim V_1}, \ldots, \underbrace{\mu_k, \ldots, \mu_k}_{\dim V_k}),$$

which is diagonal. Hence T is diagonalizable. $\qquad \square$

Corollary 8.5 (Spectral Decomposition)**.** *Suppose $T : V \to V$ is diagonalizable. Then there exist linear operators $P_i : V \to V$, for $1 \le i \le k$, such that:*

1. *$P_i P_j = 0$ for all $i \ne j$,*
2. *$\sum_{i=1}^{k} P_i = I$,*
3. *$P_i^2 = P_i$,*
4. *$P_i T = T P_i$ for all i,*
5. *$T = \mu_1 P_1 + \cdots + \mu_k P_k$, for distinct scalars $\mu_1, \ldots, \mu_k \in \mathbb{F}$.*

Proof. Since $T : V \to V$ is diagonalizable, the minimal polynomial of T splits into distinct linear factors:

$$m_T(x) = (x - \mu_1)(x - \mu_2) \cdots (x - \mu_k),$$

with $\mu_1, \ldots, \mu_k \in \mathbb{F}$ distinct. Define:

$$q_i(x) := (x - \mu_1) \cdots \widehat{(x - \mu_i)} \cdot (c - \mu_k)$$

As in the proof of Theorem 8.3 such that $m_T(x) = (x - \mu_i)q_i(x)$, and there exist polynomials $a_1(x), \ldots, a_k(x) \in \mathbb{F}[x]$ such that

$$\sum_{i=1}^{k} a_i(x)q_i(x) = 1.$$

Define operators $P_i : V \to V$ by

$$P_i := a_i(T)q_i(T).$$

We now verify the required properties:

(1) **Orthogonality:** For $i \neq j$, we compute:

$$P_i P_j = a_i(T)q_i(T) \cdot a_j(T)q_j(T) = a_i(T)a_j(T)q_i(T)q_j(T).$$

Since $q_i(x)q_j(x)$ is divisible by $m_T(x)$, and $m_T(T) = 0$, it follows that $P_i P_j = 0$.

(2) **Completeness:** By construction:

$$\sum_{i=1}^{k} P_i = \sum_{i=1}^{k} a_i(T)q_i(T) = I.$$

(3) **Idempotency:** Since $I = \sum_{j=1}^{k} P_j = \sum_{j=1}^{k} a_j(T)q_j(T)$, and $P_i P_j = 0$ if $i \neq j$, we compute:

$$P_i^2 = P_i \cdot (P_1 + \cdots + P_i + \cdots + P_k) = P_i \cdot I = P_i.$$

(4) **Commutativity:** Each $P_i = a_i(T)q_i(T)$ is a polynomial in T, so it commutes with T. Thus:

$$P_i T = T P_i.$$

(5) **Spectral decomposition:** For each i, define the subspace

$$V_i := \ker(T - \mu_i I) = \ker(m_i(T)),$$

where $m_i(x) := (x - \mu_i)$. Let $\mathbf{v} \in V$. Then using $\sum P_i = I$, write:

$$\mathbf{v} = \sum_{i=1}^{k} P_i \mathbf{v}.$$

Apply T:

$$T\mathbf{v} = \sum_{i=1}^{k} T(P_i\mathbf{v}) = \sum_{i=1}^{k} P_i T(\mathbf{v}) \quad \text{(since } TP_i = P_iT\text{)}.$$

But $P_i\mathbf{v} \in V_i$, and $T|_{V_i} = \mu_i \cdot \mathrm{id}_{V_i}$, so:

$$T(P_i\mathbf{v}) = \mu_i P_i\mathbf{v}.$$

Hence,

$$T\mathbf{v} = \sum_{i=1}^{k} \mu_i P_i\mathbf{v} \quad \Rightarrow \quad T = \sum_{i=1}^{k} \mu_i P_i.$$

$\square$

8.2   Jordan Normal Form

In the previous section, we have proved that a linear operator T is diagonalizable if and only if its minimal polynomial $m_T(x)$ consists of distinct linear terms. One may ask What happens if the minimal polynomial

$$m_T(x) = (x - \lambda_1)^{e_1} \cdots (x - \lambda_k)^{e_k}$$

contains repeated linear factors?

Theorem 8.6 (Jordan Normal Form). *Let $\mathbb{F}$ be an algebraically closed field. Suppose that $T : V \to V$ is a linear operator whose minimal polynomial has the form*

$$m_T(x) = \prod_{i=1}^{k} (x - \lambda_i)^{e_i},$$

where the $\lambda_i \in \mathbb{F}$ are distinct eigenvalues.

Then there exists a basis $\mathcal{A}$ of V such that the matrix representation of T with respect to $\mathcal{A}$ is block diagonal:

$$T_{\mathcal{A},\mathcal{A}} = \mathrm{diag}(J_1, \ldots, J_k),$$

where each block J_i is a Jordan block *of the form*

$$J_i = \begin{pmatrix} \mu & 1 & 0 & \cdots & 0 \\ 0 & \mu & 1 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & \mu & 1 \\ 0 & \cdots & \cdots & 0 & \mu \end{pmatrix},$$

for some $\mu \in \{\lambda_1, \ldots, \lambda_k\}$.

8.2.1 First Step of Proof:

By the Primary Decomposition Theorem , we have a direct sum decomposition:

$$V = V_1 \oplus \cdots \oplus V_k, \quad \text{where } V_i = \ker\left((T - \lambda_i I)^{e_i}\right),$$

and each subspace $\mathbf{v}_i$ is T-invariant. Moreover, choose a basis $\mathcal{B}_i$ of $\mathbf{v}_i$ for each V_i. Then the union $\mathcal{B} := \bigcup_{i=1}^{k} \mathcal{B}_i$ forms a basis of V, and the matrix representation of T in this basis is block diagonal:

$$T_{\mathcal{B},\mathcal{B}} = \begin{bmatrix} [T|_{\mathbf{v}_1}]_{\mathcal{B}_1,\mathcal{B}_1} & 0 & \cdots & 0 \\ 0 & [T|_{\mathbf{v}_2}]_{\mathcal{B}_2,\mathcal{B}_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & [T|_{\mathbf{v}_k}]_{\mathcal{B}_k,\mathcal{B}_k} \end{bmatrix},$$

with $m_{T|_{\mathbf{v}_i}}(x) = (x - \lambda_i)^{e_i}$.

Therefore, to prove the full Jordan normal form, it suffices to treat the case where $T : V \to V$ has minimal polynomial:

$$m_T(x) = (x - \lambda)^e$$

is a single primary factor. In such a case, all vectors $\mathbf{v} \in V$ satisfies

$$(T - \lambda I)^e \mathbf{v} = \mathbf{0},$$

that is, all $\mathbf{v} \in V$ are generalized eigenvectors (Definition 6.21) of T with generalized eigenvalue λ.

8.2.2 Second Step of Proof

Firstly, we consider the case when the generalized eigenvalue is $\lambda = 0$, and the minimal polynomial has the form $m_T(x) = x^m$:

Proposition 8.7 (Nilpotent Case of Jordan Normal Form). *Suppose $T : V \to V$ is a linear operator with minimal polynomial*

$$m_T(x) = x^m.$$

Then there exists a basis $\mathcal{A}$ of V such that

$$[T]_{\mathcal{A},\mathcal{A}} = \text{diag}(J_1, \ldots, J_\ell),$$

where each block J_i is a Jordan block of the form

$$J_i = \begin{bmatrix} 0 & 1 & & \\ & 0 & \ddots & \\ & & \ddots & 1 \\ & & & 0 \end{bmatrix}.$$

Proof. Assume $m_T(x) = x^m$. Then we have the ascending chain of kernels:

$$\{0\} = \ker(T^0) \subsetneq \ker(T^1) \subsetneq \cdots \subsetneq \ker(T^m) = V,$$

we have $\ker\left(T^{i-1}\right) \subsetneq \ker\left(T^i\right)$ for $i = 1, \ldots, m$: Note that $\ker\left(T^{m-1}\right) \subsetneq \ker\left(T^m\right) = V$ due to the minimality of $m_T(x)$; and $\ker\left(T^{m-2}\right) \subsetneq \ker\left(T^{m-1}\right)$ since otherwise for any $\mathbf{x} \in \ker\left(T^m\right)$,

$$T^{m-1}(T\mathbf{x}) = \mathbf{0} \Rightarrow T\mathbf{x} \in \ker\left(T^{m-1}\right) = \ker\left(T^{m-2}\right) \Rightarrow T^{m-2}(T\mathbf{x}) = T^{m-1}(\mathbf{x}) = \mathbf{0},$$

i.e., $\mathbf{x} \in \ker\left(T^{m-1}\right)$, which contradicts to the fact that $\ker\left(T^{m-1}\right) \subsetneq \ker\left(T^m\right)$. Proceeding this trick sequentially for $i = m, m-1, \ldots, 1$, we proved the desired result.

For each $i = 1, \ldots, m$, define the quotient space:

$$W_i = \ker(T^i)/\ker(T^{i-1}),$$

Pictorially,

$$\{0\} = \underbrace{\ker(T^0) \subsetneq \ker(T^1)}_{W_1} \subsetneq \cdots \subsetneq \underbrace{\ker(T^{m-1}) \subsetneq \ker(T^m)}_{W_m} = V,$$

Choose a basis of W_m:

$$\mathcal{B}'_m = \{\mathbf{u}_1 + \ker(T^{m-1}), \ldots, \mathbf{u}_\ell + \ker(T^{m-1})\} \subseteq W_m.$$

And consider a **set** $\mathcal{C}'_m$ in W_{m-1}:

$$\mathcal{C}'_m = \{T(\mathbf{u}_1) + \ker(T^{m-2}), \ldots, T(\mathbf{u}_\ell) + \ker(T^{m-2})\} \subseteq W_{m-1}.$$

(Note that each $T^{m-1}(T(\mathbf{u}_j)) = T(\mathbf{u}_j) = \mathbf{0}$ since $\mathbf{u}_j \in \ker(T)$, therefore $T(\mathbf{u}_j) \in \ker(T^{m-1})$)

Claim: $\mathcal{C}'_m$ is linearly independent in W_{m-1}: consider the equation

$$\sum_j k_j \left(T(\mathbf{u}_j) + \ker(T^{m-2})\right) = \mathbf{0}_{W_i} \Leftrightarrow T\left(\sum_j k_j \mathbf{u}_j\right) + \ker\left(T^{m-2}\right) = \mathbf{0}_{W_i}$$

i.e.,

$$T\left(\sum_j k_j \mathbf{u}_j\right) \in \ker\left(T^{m-2}\right) \Leftrightarrow T^{m-2}\left(T\left(\sum_j k_j \mathbf{u}_j\right)\right) = \mathbf{0}_V.$$

In other words, $\sum_j k_j \mathbf{u}_j \in \ker(T^{m-1})$, and hence

$$\sum_j k_j \mathbf{u}_j + \ker\left(T^{m-1}\right) = \mathbf{0}_{W_m} \Leftrightarrow \sum_j k_j \left(\mathbf{u}_j + \ker\left(T^{m-1}\right)\right) = \mathbf{0}_{W_m}.$$

Since $\mathcal{B}_m = \{\mathbf{u}_j + \ker(T^{m-1}) \mid 1 \le j \le \ell_m\}$ forms a basis of W_m, we have $k_j = 0$ for all j, and $\mathcal{C}'_m$ is linearly independent in W_{m-1}.

By Basis Extension Theorem, one can get a basis of W_{m-1} of the form

$$\mathcal{B}'_{m-1} = \{T(\mathbf{u}_1)+\ker(T^{m-2}),\dots,T(\mathbf{u}_\ell)+\ker(T^{m-2}), \quad \mathbf{v}_1+\ker(T^{m-2}),\dots,\mathbf{v}_p+\ker(T^{m-2})\} \subseteq W_{m-1}.$$

By similar arguments, one can get

$$\mathcal{C}'_{m-1} = \{T^2(\mathbf{u}_1)+\ker(T^{m-3}),\dots,T^2(\mathbf{u}_\ell)+\ker(T^{m-3}), \quad T(\mathbf{v}_1)+\ker(T^{m-3}),\dots,T(\mathbf{v}_p)+\ker(T^{m-3})\}$$

is linearly independent in W_{m-2}, and one can construct a basis

$$\mathcal{B}'_{m-2} = \{T^2(\mathbf{u}_1) + \ker(T^{m-3}),\dots,T^2(\mathbf{u}_\ell) + \ker(T^{m-3}), \quad T(\mathbf{v}_1) + \ker(T^{m-3}),\dots,T(\mathbf{v}_p) + $$
$$\ker(T^{m-3}), \mathbf{w}_1 + \ker(T^{m-3}),\dots,\mathbf{w}_q + \ker(T^{m-3})\}$$

of W_{m-2}.

Consequently, we have the following collection of bases of W_i:

$$W_1 \qquad\qquad \cdots \qquad\qquad W_{m-2} \qquad\qquad W_{m-1} \qquad\qquad W_m$$

$$\left\{\begin{array}{c} T^{m-1}(\mathbf{u}_1) + \ker(T^0) \\ \vdots \\ T^{m-1}(\mathbf{u}_\ell) + \ker(T^0) \end{array}\right\} \cdots \left\{\begin{array}{c} T^2(\mathbf{u}_1) + \ker(T^{m-3}) \\ \vdots \\ T^2(\mathbf{u}_\ell) + \ker(T^{m-3}) \end{array}\right\} \left\{\begin{array}{c} T(\mathbf{u}_1) + \ker(T^{m-2}) \\ \vdots \\ T(\mathbf{u}_\ell) + \ker(T^{m-2}) \end{array}\right\} \left\{\begin{array}{c} \mathbf{u}_1 + \ker(T^{m-1}) \\ \vdots \\ \mathbf{u}_\ell + \ker(T^{m-1}) \end{array}\right\}$$

$$\left\{\begin{array}{c} T^{m-2}(\mathbf{v}_1) + \ker(T^0) \\ \vdots \\ T^{m-2}(\mathbf{v}_p) + \ker(T^0) \end{array}\right\} \cdots \left\{\begin{array}{c} T(\mathbf{v}_1) + \ker(T^{m-3}) \\ \vdots \\ T(\mathbf{v}_p) + \ker(T^{m-3}) \end{array}\right\} \left\{\begin{array}{c} \mathbf{v}_1 + \ker(T^{m-2}) \\ \vdots \\ \mathbf{v}_p + \ker(T^{m-2}) \end{array}\right\}$$

$$\left\{\begin{array}{c} T^{m-3}(\mathbf{w}_1) + \ker(T^0) \\ \vdots \\ T^{m-3}(\mathbf{w}_q) + \ker(T^0) \end{array}\right\} \cdots \left\{\begin{array}{c} \mathbf{w}_1 + \ker(T^{m-3}) \\ \vdots \\ \mathbf{w}_q + \ker(T^{m-3}) \end{array}\right\}$$

$$\vdots$$

Indeed, by removing the $\ker(T^i)$'s in the above list, the vectors will give a basis of V (Exercise).

Now choose the ordered basis $\mathcal{A}$ of V by arranging the vectors:

$$
\mathcal{A} = \left\{
\begin{array}{ccccc}
T^{m-1}(\mathbf{u}_1), & T^{m-2}(\mathbf{u}_1), & \cdots & T^2(\mathbf{u}_1), & T(\mathbf{u}_1), & \mathbf{u}_1, \\
\vdots & \vdots & & \vdots & \vdots & \vdots \\
T^{m-1}(\mathbf{u}_\ell), & T^{m-2}(\mathbf{u}_\ell), & \cdots & T^2(\mathbf{u}_1), & T(\mathbf{u}_\ell), & \mathbf{u}_\ell, \\
T^{m-2}(\mathbf{v}_1), & T^{m-3}(\mathbf{v}_1), & \cdots & T(\mathbf{v}_1), & \mathbf{v}_1, \\
\vdots & \vdots & & \vdots & \vdots \\
T^{m-2}(\mathbf{v}_p), & T^{m-3}(\mathbf{v}_p), & \cdots & T(\mathbf{v}_p), & \mathbf{v}_p, \\
T^{m-3}(\mathbf{w}_1), & T^{m-4}(\mathbf{w}_1), & \cdots & \mathbf{w}_1, \\
\vdots & \vdots & & \vdots \\
T^{m-3}(\mathbf{w}_q), & T^{m-4}(\mathbf{w}_q), & \cdots & \mathbf{w}_q, \\
\vdots & \vdots
\end{array}
\right\}
$$

Under this basis (read from left to right, top to bottom), $T_{\mathcal{A},\mathcal{A}}$ is block diagonal with nilpotent Jordan blocks:

- All diagonal entries are zero since $T(T^{i-1}(\mathbf{x})) = T^i(\mathbf{x})$, and both $\{T^{i-1}(\mathbf{x}), T^i(\mathbf{x})\}$ are basis vectors.

- All superdiagonal entries are 1, corresponding to $T(T^k(\mathbf{x})) = T^{k+1}(\mathbf{x})$.

In particular, the matrix representation is of the form:

$$
T_{\mathcal{A},\mathcal{A}} = \begin{pmatrix}
N_m \\
& \ddots \\
& & N_m \\
& & & N_{m-1} \\
& & & & \ddots \\
& & & & & N_{m-1} \\
& & & & & & N_{m-2} \\
& & & & & & & \ddots \\
& & & & & & & & N_{m-2} \\
& & & & & & & & & \ddots
\end{pmatrix}
$$

where $N_i = \begin{pmatrix} 0 & 1 \\ & 0 & \ddots \\ & & \ddots & 1 \\ & & & 0 \end{pmatrix}$ is an $i \times i$-matrix, and J_m appears ℓ times, J_{m-1} appears p times, J_{m-2} appears q times in the above matrix. $\qquad \square$

8.2.3   Third Step of Proof

Then we consider the case where $m_T(x) = (x - \lambda)^e$:

Corollary 8.8. *Suppose $T : V \to V$ is such that the minimal polynomial is*

$$m_T(x) = (x - \lambda)^e,$$

for some $\lambda \in \mathbb{F}$. Then there exists a basis $\mathcal{A}$ of V such that

$$T_{\mathcal{A},\mathcal{A}} = \mathrm{diag}(J_1, \ldots, J_\ell),$$

where each block J_i is of the form

$$J_i = \begin{pmatrix} \lambda & 1 & & \\ & \lambda & \ddots & \\ & & \ddots & 1 \\ & & & \lambda \end{pmatrix}.$$

Proof. Let $U := T - \lambda I$. Then the minimal polynomial of U is $m_U(x) = x^e$. By Proposition 8.7, there exists a basis $\mathcal{A}$ such that

$$U_{\mathcal{A},\mathcal{A}} = \mathrm{diag}(M_1, \ldots, M_\ell),$$

where $M_i = N_{\alpha_i}$ as given in Proposition 8.7 for some $\alpha_i \in \mathbb{N}$. Therefore,

$$T_{\mathcal{A},\mathcal{A}} = [U + \lambda I]_{\mathcal{A},\mathcal{A}} = \mathrm{diag}(M_1 + \lambda I, \ldots, M_\ell + \lambda I),$$

which gives the desired Jordan blocks centered at λ. $\qquad\square$

Corollary 8.9. *Let $A \in M_{n \times n}(\mathbb{C})$. Then there exists an invertible matrix P such that*

$$P^{-1}AP = \mathrm{diag}(J_1, \ldots, J_\ell),$$

where each J_i is a Jordan block of the form described above. That is, any complex square matrix is similar to a matrix in Jordan normal form.

Proof. By the Fundamental Theorem of Algebra, the minimal polynomial of T can always be factorized into linear terms, i.e.

$$m_T(x) = (x - \lambda_1)^{e_1} \cdots (x - \lambda_k)^{e_k}.$$

Therefore, the result follows from Theorem 8.6 and the discussions in section 3.2. $\qquad\square$

Chapter 9

Inner Product Space

9.1   Introduction to Inner Product Space

Definition 9.1 (Bilinear Form)**.** Let V be a vector space over $\mathbb{R}$. A *bilinear form* on V is a map

$$F : V \times V \to \mathbb{R}$$

such that for all $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$ and all scalars $\lambda \in \mathbb{R}$,

1. $F(\mathbf{u} + \mathbf{v}, \mathbf{w}) = F(\mathbf{u}, \mathbf{w}) + F(\mathbf{v}, \mathbf{w})$

2. $F(\mathbf{u}, \mathbf{v} + \mathbf{w}) = F(\mathbf{u}, \mathbf{v}) + F(\mathbf{u}, \mathbf{w})$

3. $F(\lambda \mathbf{u}, \mathbf{v}) = \lambda F(\mathbf{u}, \mathbf{v}) = F(\mathbf{u}, \lambda \mathbf{v})$

   We say:

   - F is *symmetric* if $F(\mathbf{u}, \mathbf{v}) = F(\mathbf{v}, \mathbf{u})$ for all $\mathbf{u}, \mathbf{v} \in V$.

   - F is *non-degenerate* if $F(\mathbf{u}, \mathbf{w}) = 0$ for all $\mathbf{u} \in V$ implies $\mathbf{w} = 0$.

   - F is *positive definite* if $F(\mathbf{v}, \mathbf{v}) > 0$ for all $\mathbf{v} \in V \setminus \{0\}$.

Remark 9.2. If F is positive definite, then F is non-degenerate. Indeed, suppose $F(\mathbf{v}, \mathbf{v}) > 0$ for all $\mathbf{v} \neq 0$, and that $F(\mathbf{u}, \mathbf{w}) = 0$ for all $\mathbf{u} \in V$. Then in particular, $F(\mathbf{w}, \mathbf{w}) = 0$. But by positive definiteness, this implies $\mathbf{w} = 0$. Hence F is non-degenerate.

When we say that V is a vector space over a field $\mathbb{F}$, we treat $\alpha \in \mathbb{F}$ as a scalar.

Definition 9.3 (Sesquilinear Form)**.** Let V be a vector space over $\mathbb{C}$. A *sesquilinear form* on V is a function

$$F : V \times V \to \mathbb{C}$$

satisfying, for all $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$ and all $\lambda \in \mathbb{C}$,

1. $F(\mathbf{u} + \mathbf{v}, \mathbf{w}) = F(\mathbf{u}, \mathbf{w}) + F(\mathbf{v}, \mathbf{w})$

2. $F(\mathbf{u}, \mathbf{v} + \mathbf{w}) = F(\mathbf{u}, \mathbf{v}) + F(\mathbf{u}, \mathbf{w})$

3. $F(\bar{\lambda}\mathbf{v}, \mathbf{w}) = F(\mathbf{v}, \lambda\mathbf{w}) = \lambda F(\mathbf{v}, \mathbf{w})$

We say that F is *conjugate symmetric* if

$$F(\mathbf{v}, \mathbf{w}) = \overline{F(\mathbf{w}, \mathbf{v})}, \quad \forall \mathbf{v}, \mathbf{w} \in V.$$

The definitions of *non-degenerate* and *positive definite* are the same as those for bilinear forms:

- F is non-degenerate if $F(\mathbf{u}, \mathbf{w}) = 0$ for all $\mathbf{u} \in V$ implies $\mathbf{w} = 0$.

- F is positive definite if $F(\mathbf{v}, \mathbf{v}) > 0$ for all $\mathbf{v} \neq 0$.

Remark 9.4. Why is the complex conjugate $\bar{\lambda}$ necessary in the definition?

Suppose we want F to be positive definite. If we define sesquilinearity without the conjugation, then for any $\mathbf{v} \in V$, we would have

$$F(i\mathbf{v}, i\mathbf{v}) = i^2 F(\mathbf{v}, \mathbf{v}) = -F(\mathbf{v}, \mathbf{v}) < 0,$$

contradicting positive definiteness.

But with the conjugation, we instead compute:

$$F(i\mathbf{v}, i\mathbf{v}) = \bar{i} \cdot i \cdot F(\mathbf{v}, \mathbf{v}) = |i|^2 F(\mathbf{v}, \mathbf{v}) = F(\mathbf{v}, \mathbf{v}),$$

which preserves positivity. Hence, the use of $\bar{\lambda}$ is essential to ensure the existence of positive definite sesquilinear forms on complex vector spaces.

Example 9.5 (Hermitian Inner Product). Let $V = \mathbb{C}^n$. A fundamental sesquilinear form on V is the Hermitian inner product, defined by

$$F(\mathbf{v}, \mathbf{w}) = \mathbf{v}^H \mathbf{w} = \begin{bmatrix} \overline{v_1} & \cdots & \overline{v_n} \end{bmatrix} \begin{bmatrix} w_1 \\ \vdots \\ w_n \end{bmatrix} = \sum_{i=1}^{n} \overline{v_i} w_i.$$

In this case, F is not symmetric in the usual sense, i.e., $F(\mathbf{v}, \mathbf{w}) \neq F(\mathbf{w}, \mathbf{v})$, but it is *conjugate symmetric*:

$$F(\mathbf{v}, \mathbf{w}) = \overline{F(\mathbf{w}, \mathbf{v})}.$$

Definition 9.6 (Inner Product). Let V be a real (respectively, complex) vector space. A bilinear (respectively, sesquilinear) form

$$\langle \cdot, \cdot \rangle : V \times V \to \mathbb{R} \text{ (respectively, } \mathbb{C})$$

is called an *inner product* if it satisfies:

- Symmetry (or conjugate symmetry): $\langle \mathbf{v}, \mathbf{w} \rangle = \overline{\langle \mathbf{w}, \mathbf{v} \rangle}$

- Positive definiteness: $\langle \mathbf{v}, \mathbf{v} \rangle > 0$ for all $\mathbf{v} \neq 0$

A vector space equipped with an inner product is called an *inner product space.*

Remark 9.7. We denote the inner product using the bracket notation $\langle \cdot, \cdot \rangle$ instead of a general form symbol like $F(\cdot, \cdot)$.

Definition 9.8 (Norm). The *norm* induced by the inner product is defined as

$$\|\mathbf{v}\| := \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}.$$

In particular, for any scalar $\alpha \in \mathbb{F}$,

$$\|\alpha \mathbf{v}\| = \sqrt{\langle \alpha \mathbf{v}, \alpha \mathbf{v} \rangle} = \sqrt{\bar{\alpha}\alpha \langle \mathbf{v}, \mathbf{v} \rangle} = |\alpha| \cdot \|\mathbf{v}\|.$$

The norm is well-defined by the positive definiteness of the inner product.

Definition 9.9 (Orthogonality and Orthonormality). Let $S = \{\mathbf{v}_i \mid i \in I\} \subset V$. We say S is an *orthogonal set* if

$$\langle \mathbf{v}_i, \mathbf{v}_j \rangle = 0 \quad \text{for all } i \neq j.$$

If in addition each vector has unit norm, i.e.,

$$\langle \mathbf{v}_i, \mathbf{v}_i \rangle = 1 \quad \text{for all } i,$$

then S is called an *orthonormal set.*

9.2   Cauchy–Schwarz Inequality

Proposition 9.10 (Cauchy–Schwarz Inequality). *Let V be an inner product space. Then for all $\mathbf{u}, \mathbf{v} \in V$,*

$$|\langle \mathbf{u}, \mathbf{v} \rangle| \leq \|\mathbf{u}\| \cdot \|\mathbf{v}\|.$$

Proof. We divide the proof into two cases depending on whether the inner product is real-valued or complex-valued.

   Case 1: $\langle \mathbf{u}, \mathbf{v} \rangle \in \mathbb{R}$: Define the function $f(t) := \|\mathbf{u} + t\mathbf{v}\|^2$ for real $t \in \mathbb{R}$. Then

$$f(t) = \langle \mathbf{u} + t\mathbf{v}, \mathbf{u} + t\mathbf{v} \rangle = \|\mathbf{u}\|^2 + 2t\langle \mathbf{u}, \mathbf{v} \rangle + t^2\|\mathbf{v}\|^2.$$

Since $f(t) \geq 0$ for all $t \in \mathbb{R}$, the discriminant of this quadratic must be non-positive:

$$(2\langle \mathbf{u}, \mathbf{v} \rangle)^2 - 4\|\mathbf{u}\|^2\|\mathbf{v}\|^2 \leq 0,$$

which simplifies to

$$|\langle \mathbf{u}, \mathbf{v} \rangle|^2 \leq \|\mathbf{u}\|^2\|\mathbf{v}\|^2,$$

and hence

$$|\langle \mathbf{u}, \mathbf{v} \rangle| \leq \|\mathbf{u}\| \cdot \|\mathbf{v}\|.$$

Case 2: $\langle \mathbf{u}, \mathbf{v} \rangle \in \mathbb{C} \setminus \mathbb{R}$: We apply a rescaling argument to reduce to the real case. Define

$$\mathbf{w} := \frac{1}{\langle \mathbf{u}, \mathbf{v} \rangle} \mathbf{u}.$$

Then

$$\langle \mathbf{w}, \mathbf{v} \rangle = \left\langle \frac{1}{\langle \mathbf{u}, \mathbf{v} \rangle} \mathbf{u}, \mathbf{v} \right\rangle = \overline{\frac{1}{\langle \mathbf{u}, \mathbf{v} \rangle}} \cdot \langle \mathbf{u}, \mathbf{v} \rangle = 1 \in \mathbb{R}.$$

Now applying the Cauchy–Schwarz inequality in the real case to vectors $\mathbf{w}$ and $\mathbf{v}$, we have

$$|\langle \mathbf{w}, \mathbf{v} \rangle| \leq \|\mathbf{w}\| \cdot \|\mathbf{v}\|,$$

i.e.,

$$1 \leq \left\| \frac{1}{\langle \mathbf{u}, \mathbf{v} \rangle} \mathbf{u} \right\| \cdot \|\mathbf{v}\| = \frac{\|\mathbf{u}\|}{|\langle \mathbf{u}, \mathbf{v} \rangle|} \cdot \|\mathbf{v}\|.$$

Rearranging gives

$$|\langle \mathbf{u}, \mathbf{v} \rangle| \leq \|\mathbf{u}\| \cdot \|\mathbf{v}\|,$$

as desired. $\qquad \square$

Proposition 9.11 (Triangle Inequality)**.** *Let V be an inner product space. Then for all $\mathbf{u}, \mathbf{v} \in V$,*

$$\|\mathbf{u} + \mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|.$$

Proof. We expand:

$$\|\mathbf{u} + \mathbf{v}\|^2 = \langle \mathbf{u} + \mathbf{v}, \mathbf{u} + \mathbf{v} \rangle = \|\mathbf{u}\|^2 + 2 \operatorname{Re}(\langle \mathbf{u}, \mathbf{v} \rangle) + \|\mathbf{v}\|^2.$$

Now use $\operatorname{Re}(\langle \mathbf{u}, \mathbf{v} \rangle) \leq |\langle \mathbf{u}, \mathbf{v} \rangle|$ and apply Cauchy–Schwarz:

$$\|\mathbf{u} + \mathbf{v}\|^2 \leq \|\mathbf{u}\|^2 + 2\|\mathbf{u}\|\|\mathbf{v}\| + \|\mathbf{v}\|^2 = (\|\mathbf{u}\| + \|\mathbf{v}\|)^2.$$

Taking square roots on both sides completes the proof. $\qquad \square$

Theorem 9.12 (Gram–Schmidt Process)**.** *Let $S = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\} \subset V$ be a finite linearly independent set in an inner product space V. Then there exists an orthonormal set $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ such that* $\operatorname{span}\{\mathbf{v}_1, \ldots, \mathbf{v}_k\} = \operatorname{span}\{\mathbf{e}_1, \ldots, \mathbf{e}_k\}$ *for all k.*

The construction is given inductively:

$$\mathbf{w}_1 := \mathbf{v}_1, \quad \mathbf{w}_{i+1} := \mathbf{v}_{i+1} - \sum_{j=1}^{i} \frac{\langle \mathbf{v}_{i+1}, \mathbf{w}_j \rangle}{\|\mathbf{w}_j\|^2} \mathbf{w}_j, \quad \text{for } i = 1, \ldots, n-1,$$

and the orthonormal set is obtained by normalizing:

$$\mathbf{e}_i := \frac{\mathbf{w}_i}{\|\mathbf{w}_i\|}.$$

Corollary 9.13. *Every finite-dimensional inner product space admits an orthonormal basis.*

9.3   Orthogonal Complement

Definition 9.14 (Orthogonal Complement)**.** Let $U \leq V$ be a subspace of an inner product space. Then the orthogonal complement of U is

$$U^\perp = \{\mathbf{v} \in V \mid \langle \mathbf{v}, \mathbf{u} \rangle = 0, \forall \mathbf{u} \in U\}$$

The analysis for orthogonal complement for vector spaces over $\mathbb{C}$ is quite similar as what we have studied in MAT2040.

Proposition 9.15. *Let V be an inner product space, $U \leq V$ a subspace of V, and $U_1, U_2 \subseteq V$ are subsets of V.*

1. *$U^\perp \leq V$ is a subspace of V.*

2. *$U \cap U^\perp = \{0\}$.*

3. *$U_1 \subseteq U_2$ implies $U_2^\perp \leq U_1^\perp$.*

Proof. 1. Suppose that $\mathbf{v}_1, \mathbf{v}_2 \in U^\perp$, where $a, b \in K$ ($K = \mathbb{C}$ or $\mathbb{R}$) , then for all $\mathbf{u} \in U$,

$$\langle a\mathbf{v}_1 + b\mathbf{v}_2, \mathbf{u} \rangle = \bar{a}\langle \mathbf{v}_1, \mathbf{u} \rangle + \bar{b}\langle \mathbf{v}_2, \mathbf{u} \rangle = \bar{a} \cdot 0 + \bar{b} \cdot 0 = 0$$

Therefore, $a\mathbf{v}_1 + b\mathbf{v}_2 \in U^\perp$.

2. Suppose that $\mathbf{u} \in U \cap U^\perp$, then we imply $\langle \mathbf{u}, \mathbf{u} \rangle = 0$. By the positive-definiteness of inner product, $\mathbf{u} = \mathbf{0}$.

3. Exercise. $\qquad\qquad\square$

Proposition 9.16. *Let $U, W \leq V$ be subspaces of an inner product space V.*

1. *If $\dim(V) < \infty$ and $U \leq V$, then $V = U \oplus U^\perp$.*

2. *One has*

    (a) $(U + W)^\perp = U^\perp \cap W^\perp$;

    (b) $(U \cap W)^\perp \supseteq U^\perp + W^\perp$;

    (c) $\left(U^\perp\right)^\perp \supseteq U$.

Moreover, if $\dim(V) < \infty$, then these are equalities.

Proof. 1. Suppose that $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ forms a basis for U, and by basis extension, we obtain $\{\mathbf{v}_1, \ldots, \mathbf{v}_k, \mathbf{v}_{k+1}, \ldots, \mathbf{v}_n\}$ is a basis for V.

By Gram-Schmidt Process, any finite basis induces an orthonormal basis. Therefore, suppose that $\{\mathbf{e}_1, \ldots, \mathbf{e}_k\}$ forms an orthonormal basis for U, and $\{\mathbf{e}_{k+1}, \ldots, \mathbf{e}_n\}$ forms an orthonormal basis for $U^\perp$.

It's easy to show $V = U + U^\perp$ using orthonormal basis.

2. (a) The inclusion $(U + W)^\perp \supseteq U^\perp \cap W^\perp$ is trivial; for the other inclusion, suppose $\mathbf{v} \in (U + W)^\perp$. Then

$$\langle \mathbf{v}, \mathbf{u} + \mathbf{w} \rangle = 0, \forall \mathbf{u} \in U, \mathbf{w} \in W$$

Taking $\mathbf{u} \equiv \mathbf{0}$ in the equality above gives $\langle \mathbf{v}, \mathbf{w} \rangle = 0$, i.e., $\mathbf{v} \in W^\perp$. Similarly, $\mathbf{v} \in U^\perp$.

(b) Follow the similar argument as in (2a). If $\dim(V) < \infty$, then write down the orthonormal basis for $U^\perp + W^\perp$ and $(U \cap W)^\perp$.

(c) Follow the similar argument as in (2a). If $\dim(V) < \infty$, then

$$V = U^\perp \oplus \left(U^\perp\right)^\perp = U \oplus U^\perp.$$

Therefore, $\dim\left(U^\perp\right)^\perp = \dim(U)$ and hence we have $\left(U^\perp\right)^\perp = (U)$ □

9.4   The Riesz Representation Theorem

Theorem 9.17 (Riesz Representation). *Let V be an inner product space over $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$. Define the mapping*

$$\phi : V \to V^*, \quad \mathbf{v} \mapsto \phi_{\mathbf{v}},$$

where $\phi_{\mathbf{v}}(\mathbf{w}) := \langle \mathbf{v}, \mathbf{w} \rangle$ for all $\mathbf{w} \in V$.

Then:

1. *The mapping ϕ is well-defined and $\mathbb{R}$-linear.*

2. *If V is finite-dimensional, then ϕ is an isomorphism of real vector spaces.*

Remark 9.18. Note that:

- If V is over $\mathbb{R}$, then ϕ is a linear map in the usual sense.

- If V is over $\mathbb{C}$, then ϕ is only $\mathbb{R}$-linear. In particular, $\phi(i\mathbf{v}) \neq i\phi(\mathbf{v})$, but $\phi(2\mathbf{v}) = 2\phi(\mathbf{v})$.

Example 9.19. Let V be a complex vector space with basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$. Any vector $\mathbf{v} \in V$ can be written as

$$\mathbf{v} = \sum_{j=1}^n \alpha_j \mathbf{v}_j, \quad \alpha_j = p_j + iq_j \in \mathbb{C}.$$

Then

$$\mathbf{v} = \sum_j p_j \mathbf{v}_j + \sum_j q_j (i\mathbf{v}_j), \quad p_j, q_j \in \mathbb{R},$$

so the set $\{\mathbf{v}_1, \ldots, \mathbf{v}_n, i\mathbf{v}_1, \ldots, i\mathbf{v}_n\}$ forms a basis for V over $\mathbb{R}$.

Proof. **(1) Well-definedness.** We need to show $\phi_{\mathbf{v}} \in V^*$, i.e., that it defines a linear functional. Let $\mathbf{w}_1, \mathbf{w}_2 \in V$ and $a, b \in \mathbb{F}$. Then:

$$\phi_{\mathbf{v}}(a\mathbf{w}_1 + b\mathbf{w}_2) = \langle \mathbf{v}, a\mathbf{w}_1 + b\mathbf{w}_2 \rangle = a\langle \mathbf{v}, \mathbf{w}_1 \rangle + b\langle \mathbf{v}, \mathbf{w}_2 \rangle = a\phi_{\mathbf{v}}(\mathbf{w}_1) + b\phi_{\mathbf{v}}(\mathbf{w}_2),$$

so $\phi_{\mathbf{v}}$ is linear (over $\mathbb{R}$ or $\mathbb{C}$, depending on V).

(2) $\mathbb{R}$-linearity of ϕ. Let $\mathbf{v}_1, \mathbf{v}_2 \in V$ and $c, d \in \mathbb{R}$. For any $\mathbf{w} \in V$,

$$\phi_{c\mathbf{v}_1 + d\mathbf{v}_2}(\mathbf{w}) = \langle c\mathbf{v}_1 + d\mathbf{v}_2, \mathbf{w} \rangle = c\langle \mathbf{v}_1, \mathbf{w} \rangle + d\langle \mathbf{v}_2, \mathbf{w} \rangle = c\phi_{\mathbf{v}_1}(\mathbf{w}) + d\phi_{\mathbf{v}_2}(\mathbf{w}),$$

so $\phi(c\mathbf{v}_1 + d\mathbf{v}_2) = c\phi(\mathbf{v}_1) + d\phi(\mathbf{v}_2)$, i.e., ϕ is $\mathbb{R}$-linear.

(3) Isomorphism when $\dim V < \infty$. Since V and V^* have the same dimension, and ϕ is injective (because $\phi_{\mathbf{v}} = 0$ implies $\langle \mathbf{v}, \mathbf{w} \rangle = 0$ for all $\mathbf{w}$, which by positive-definiteness implies $\mathbf{v} = 0$), ϕ is a linear isomorphism. $\square$

We now relate orthogonal complement with the annihilator space:

Proposition 9.20. *The mapping* $\phi : V \to V^*$ *defined by*

$$\phi(\mathbf{v}) := \phi_{\mathbf{v}}$$

in Theorem 9.17 maps $U^\perp \le V$ *injectively to* $\mathrm{Ann}\,(U) \le V^*$.

If $\dim(V) < \infty$ *, then* $U^\perp \cong \mathrm{Ann}\,(U)$ *as* $\mathbb{R}$*-vector spaces.*

Proof. The injectivity of ϕ is given in Theorem 9.17. For any $\mathbf{v} \in U^\perp$, we imply $\phi_{\mathbf{v}}(\mathbf{u}) = 0$ for all $\mathbf{u} \in U$, i.e. $\phi_{\mathbf{v}} \in \mathrm{Ann}\,(U)$. Therefore, $\phi\left(U^\perp\right) \le \mathrm{Ann}\,(U)$.

Suppose $\dim(V) < \infty$, by Proposition 9.16(1),

$$\dim(U) + \dim\left(U^\perp\right) = \dim(V).$$

Since $\dim(U) + \dim(\mathrm{Ann}\,(U)) = \dim(V)$, we imply $\dim\left(U^\perp\right) = \dim(\mathrm{Ann}\,(U))$. Consequently, the injection

$$\phi|_{U^\perp} : U^\perp \to \mathrm{Ann}\,(U)$$

is an isomorphism between $\mathbb{R}$-vector spaces $U^\perp$ and $\mathrm{Ann}\,(U)$. $\square$

Chapter 10

Adjoint Operators on Inner Product Spaces

In this chapter, we generalize the notion of taking transpose (or Hermitian transpose) of a matrix for linear operators.

Definition 10.1. Let $T : V \to V$ be a linear operator between inner product spaces. The *adjoint* of T is defined as $T' : V \to V$ satisfying

$$\langle T'(\mathbf{v}), \mathbf{w} \rangle = \langle \mathbf{v}, T(\mathbf{w}) \rangle, \forall \mathbf{w} \in V \tag{10.1}$$

Remark 10.2. Previously we have another definition of the adjoint of $T : V \to W$, denoted as $T^* : W^* \to V^*$. It is an unfortunate abuse of the same terminology but with different meaning. To distinguish the two adjoints, we will use

$$T' : V \to V$$

for the adjoint of operators of inner product spaces, and

$$T : W^* \to V^*$$

for the adjoint of linear transformations.

It is not clear that whether such $T' : V \to V$ exists from the definition, let alone whether it is a linear operator or not. These issues are dealt with by the following:

Proposition 10.3. *If* $\dim(V) < \infty$, *then* T' *exists, and it is unique. Moreover,* T' *is a linear map.*

Proof. Fix any $\mathbf{v} \in V$. Consider the mapping

$$\alpha_{\mathbf{v}} : \mathbf{w} \xrightarrow{T} T(\mathbf{w}) \xrightarrow{\phi_{\mathbf{v}}} \langle \mathbf{v}, T(\mathbf{w}) \rangle$$

This is a linear transformation from V to $\mathbb{F}$, i.e., $\alpha_{\mathbf{v}} \in V^*$. By Theorem 9.17, ϕ is an isomorphism from V to V^*. Therefore, for any $\alpha_{\mathbf{v}} \in V^*$, there exists a vector $T'(\mathbf{v}) \in V$ such that

$$\phi\left(T'(\mathbf{v})\right) = \alpha_{\mathbf{v}} \in V^*$$

Or equivalently, $\phi_{T'(\mathbf{v})}(\mathbf{w}) = \alpha_{\mathbf{v}}(\mathbf{w}), \forall \mathbf{w} \in V$, i.e.,

$$\langle T'(\mathbf{v}), \mathbf{w} \rangle = \langle \mathbf{v}, T(\mathbf{w}) \rangle. \tag{10.1}$$

Therefore, from $\mathbf{v}$ we have constructed $T'(\mathbf{v})$ satisfying Equation 10.1. Now define

$$T' : V \to V \qquad \mathbf{v} \mapsto T'(\mathbf{v}).$$

Since the choice of $T'(v)$ is unique by the injectivity of ϕ, T' is well-defined.

Now we show T' is a linear transformation: Let $\mathbf{v}_1, \mathbf{v}_2 \in V, a, b \in \mathbb{F}$. For all $\mathbf{w} \in V$, we have

$$\begin{aligned}
\langle T'(a\mathbf{v}_1 + b\mathbf{v}_2), \mathbf{w} \rangle = \langle a\mathbf{v}_1 + b\mathbf{v}_2, T(\mathbf{w}) \rangle &= \bar{a}\langle \mathbf{v}_1, T(\mathbf{w}) \rangle + \bar{b}\langle \mathbf{v}_2, T(\mathbf{w}) \rangle \\
&= \bar{a}\langle T'(\mathbf{v}_1), \mathbf{w} \rangle + \bar{b}\langle T'(\mathbf{v}_2), \mathbf{w} \rangle \\
&= \langle aT'(\mathbf{v}_1) + bT'(\mathbf{v}_2), \mathbf{w} \rangle
\end{aligned}$$

Therefore,
$$\langle T'(a\mathbf{v}_1 + b\mathbf{v}_2) - \left[aT'(\mathbf{v}_1) + bT'(\mathbf{v}_2)\right], \mathbf{w} \rangle = 0, \forall \mathbf{w} \in V.$$

By the non-degeneracy of inner product,

$$T'(a\mathbf{v}_1 + b\mathbf{v}_2) - \left[aT'(\mathbf{v}_1) + bT'(\mathbf{v}_2)\right] = \mathbf{0},$$

i.e., $T'(a\mathbf{v}_1 + b\mathbf{v}_2) = aT'(\mathbf{v}_1) + bT'(\mathbf{v}_2)$. $\qquad\qquad\square$

Example 10.4. Let $V = (\mathbb{R}^n, \langle \cdot, \cdot \rangle)$ be the usual inner product. Consider the matrix-multiplication mapping $T : V \to V$ given by:

$$T(\mathbf{v}) = A\mathbf{v}.$$

Then $\langle T'(\mathbf{v}), \mathbf{w} \rangle = \langle \mathbf{v}, T(\mathbf{w}) \rangle$ implies

$$(T'(\mathbf{v}))^{\mathrm{T}}\mathbf{w} = \langle \mathbf{v}, A\mathbf{w} \rangle = \mathbf{v}^{\mathrm{T}}A\mathbf{w} = \left(A^{\mathrm{T}}\mathbf{v}\right)^{\mathrm{T}}\mathbf{w}$$

Therefore,
$$T'(\mathbf{v}) = A^{\mathrm{T}}\mathbf{v}.$$

Meanwhile, for the usual inner product $(\mathbb{C}^n, \langle \cdot, \cdot \rangle)$, the adjoint is equal to

$$T'(\mathbf{v}) = A^{\mathrm{H}}\mathbf{v} = \bar{A}^{\mathrm{T}}\mathbf{v},$$

where the extra complex conjugate comes from the conjugate symmetry of $\langle \cdot, \cdot \rangle$.

Proposition 10.5. *Let $T : V \to V$ be a linear transformation, V a inner product space. Suppose that $\mathcal{B} = \{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ is an orthonormal basis of V, then*

$$(T')_{\mathcal{B},\mathcal{B}} = \overline{(T_{\mathcal{B},\mathcal{B}})}^{\mathrm{T}}.$$

Proof. Suppose that $T_{\mathcal{B},\mathcal{B}} = (a_{ij})$, where $T(\mathbf{e}_j) = \sum\limits_{k=1}^{n} a_{kj}\mathbf{e}_k$, then

$$\langle \mathbf{e}_i, T(\mathbf{e}_j) \rangle = \left\langle \mathbf{e}_i, \sum_{k=1}^{n} a_{kj}\mathbf{e}_k \right\rangle = \sum_{k=1}^{n} a_{kj} \langle \mathbf{e}_i, \mathbf{e}_k \rangle = a_{ij}$$

Also, suppose $(T')_{\mathcal{B},\mathcal{B}} = (b_{ij})$, we imply $T'(\mathbf{e}_j) = \sum\limits_{k=1}^{n} b_{ij}\mathbf{e}_k$, which follows that $\langle \mathbf{e}_i, T'(\mathbf{e}_j) \rangle = b_{ij}$ i.e., $\overline{a_{ji}} = b_{ij}$. $\qquad\square$

Note that Proposition 10.5 does not hold if $\mathcal{B}$ is not an orthonormal basis.

10.1   Self-Adjoint Operators

Definition 10.6 (Self-Adjoint)**.** Let V be an inner product space and $T : V \to V$ be a linear operator. Then T is self-adjoint if $T' = T$.

Example 10.7. Let $V = \mathbb{C}^n$, and $\mathcal{B} = \{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ be a orthonormal basis. Let $T : V \to V$ be given by

$$T(\mathbf{v}) = A\mathbf{v}, \quad \text{where } A \in M_{n \times n}(\mathbb{C}).$$

Or equivalently, there exists basis $\mathcal{B}$ such that $T_{\mathcal{B},\mathcal{B}} = A$.

In such case, T is self-adjoint if and only if $(T')_{\mathcal{B},\mathcal{B}} = T_{\mathcal{B},\mathcal{B}}$, i.e., $\overline{T_{\mathcal{B},\mathcal{B}}^{\mathrm{T}}} = T_{\mathcal{B},\mathcal{B}}$, i.e., $A^{\mathrm{H}} = A$. Therefore, $T(\mathbf{v}) = \mathbf{A}\mathbf{v}$ is self-adjoint if and only if $A^{\mathrm{H}} = A$, i.e. A is a Hermitian matrix.

Similarly, if $\mathbb{C}$ is replaced by $\mathbb{R}$, then T is self-adjoint if and only if A is symmetric.

The notion of self-adjoint for linear operator is essentially the generalized notion of Hermitian for matrix that we have studied in MAT 2040. Therefore, we expect to have some nice properties for self-adjoint operators, and the proof for which are essentially the same for the proof in the case of Hermitian matrices in MAT 2040.

Proposition 10.8. *If λ is an eigenvalue of a self-adjoint operator T, then $\lambda \in \mathbb{R}$.*

Proof. Suppose there is an eigen-pair $(\lambda, \mathbf{w})$ for $\mathbf{w} \neq \mathbf{0}$, then

$$\lambda\langle \mathbf{w}, \mathbf{w} \rangle = \langle \mathbf{w}, \lambda\mathbf{w} \rangle = \langle \mathbf{w}, T(\mathbf{w}) \rangle = \langle T'(\mathbf{w}), \mathbf{w} \rangle = \langle T(\mathbf{w}), \mathbf{w} \rangle = \langle \lambda\mathbf{w}, \mathbf{w} \rangle = \bar{\lambda}\langle \mathbf{w}, \mathbf{w} \rangle$$

Since $\langle \mathbf{w}, \mathbf{w} \rangle \neq 0$ by non-degeneracy property, we have $\lambda = \bar{\lambda}$, i.e., $\lambda \in \mathbb{R}$. $\qquad\square$

Proposition 10.9. *If $U \leq V$ is T-invariant over the self-adjoint operator T, then so is $U^{\perp}$.*

Proof. It suffices to show $T(\mathbf{v}) \in U^{\perp}, \forall \mathbf{v} \in U^{\perp}$, i.e., for any $\mathbf{u} \in U$, check that

$$\langle \mathbf{u}, T(\mathbf{v}) \rangle = \langle T'(\mathbf{u}), \mathbf{v} \rangle = \langle T(\mathbf{u}), \mathbf{v} \rangle = 0,$$

where the last equality is because that $T(\mathbf{u}) \in U$ and $\mathbf{v} \in U^{\perp}$. Therefore, $T(\mathbf{v}) \in U^{\perp}$.   □

Theorem 10.10. *If $T : V \to V$ is self-adjoint, and $\dim(V) < \infty$, then there exists an orthonormal basis of eigenvectors of T, i.e., an orthonormal basis of V such that any element from this basis is an eigenvector of T.*

Proof. We use the induction on $\dim(V)$: The result is trivial for $\dim(V) = 1$. By induction hypothesis, suppose the theorem holds for all vector spaces V with $\dim(V) \leq k$. We now want to show the theorem holds when $\dim(V) = k + 1$:

Let $T : V \to V$ is self-adjoint with $\dim(V) = k + 1$.

<div style="border:1px solid black; display:inline-block; padding:4px;">

Claim: There is an eigen-pair $(\lambda_1, \mathbf{v})$ for T.

</div>

To see why the claim holds, we separate the case when V is a vector space over $\mathbb{C}$ or $\mathbb{R}$:

- If $\mathbb{F} = \mathbb{C}$, then $\mathcal{X}_T(x)$ can be decomposed as

$$\mathcal{X}_T(x) = (x - \lambda_1) \cdots (x - \lambda_{k+1})$$

  by the fundamental theorem of algebra (c.f. Theorem 6.9). Since there is a linear factor $(x - \lambda_1)$, we obtain an eigen-pair $(\lambda_1, \mathbf{v})$ for T.

- If $\mathbb{F} = \mathbb{R}$, let $\mathcal{A}$ be an orthogonal basis of V and

$$A := T_{\mathcal{A},\mathcal{A}} \in M_{(k+1) \times (k+1)}(\mathbb{R}).$$

By Example 10.7, A is a symmetric matrix.

Consider $\tau : \mathbb{R}^{k+1} \to \mathbb{R}^{k+1}$ the matrix transformation $\tau(\mathbf{x}) = A\mathbf{x}$ ($\mathbf{x} \in \mathbb{R}^{k+1}$), and $\tau_{\mathbb{C}} : \mathbb{C}^{k+1} \to \mathbb{C}^{k+1}$ be the **complexified** matrix transformation $T(\mathbf{z}) = A\mathbf{z}$ ($\mathbf{z} \in \mathbb{C}^{k+1}$). Obviously, one has

$$\mathcal{X}_T(x) = \mathcal{X}_\tau(x) = \mathcal{X}_{\tau_{\mathbb{C}}}(x) = \det(A - xI). \tag{10.2}$$

On the other hand, since A is symmetric, Example 10.7 implies that $\tau_{\mathbb{C}}$ is self-adjoint. By the fundamental theorem of algebra again,

$$\mathcal{X}_{\tau_{\mathbb{C}}}(x) = (x - \lambda_1) \cdots (x - \lambda_{k+1}),$$

where all $\lambda_i \in \mathbb{R}$ are real by Proposition 10.8. So Equation 10.2 implies that $\mathcal{X}_T(x)$ can be factorized into real linear factors, and we obtain the eigen-pair $(\lambda_1, \mathbf{v})$ for T.

Consider $U = \text{span}\{\mathbf{v}\}$, then

- U is T-invariant by $T(\mathbf{v}) = \lambda\mathbf{v}$;

- $U^\perp$ is T-invariant by Proposition 10.8;

- $V = U \oplus U^\perp$, since V is finite dimensional inner product space.

Consider $T|_{U^\perp}$, which is a self-adjoint operator on $U^\perp$, with $\dim\left(U^\perp\right) = k$. By induction hypothesis, there exists an orthonormal basis $\{\mathbf{e}_2, \ldots, \mathbf{e}_{k+1}\}$ of eigenvectors of $T|_{U^\perp}$.

Consider the basis $\mathcal{B} = \left\{\mathbf{v}' = \frac{\mathbf{v}}{\|\mathbf{v}\|}, \mathbf{e}_2, \ldots, \mathbf{e}_{k+1}\right\}$. Then one can check the following:

1. $\mathcal{B}$ forms a basis of V.

2. All $\mathbf{v}', \mathbf{e}_i$ are of norm 1 eigenvectors of T.

3. $\mathcal{B}$ is an orthonormal set, e.g., $\langle \mathbf{v}', \mathbf{e}_i \rangle = 0$, where $\mathbf{v}' \in U$ and $\mathbf{e}_i \in U^\perp$.

Therefore, $\mathcal{B}$ is a basis of orthonormal eigenvectors of V. $\qquad\square$

Corollary 10.11. *If* $\dim(V) < \infty$, *and* $T : V \to V$ *is self-adjoint, then there exists an orthonormal basis* $\mathcal{B}$ *such that*

$$T_{\mathcal{B},\mathcal{B}} = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}.$$

In particular, for all real symmetric matrix $A \in M_{n \times n}(\mathbb{R})$, *there exists orthogonal matrix* P *i.e.* $\left(P^{\mathrm{T}}P = I_n\right)$ *such that*

$$P^{-1}AP = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}.$$

Proof. 1. By applying Theorem 10.10, there exists orthonormal basis of V, say $\mathcal{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ such that $T(\mathbf{v}_i) = \lambda_i \mathbf{v}_i$. Directly writing the basis representation gives

$$T_{\mathcal{B},\mathcal{B}} = \text{diag}\left(\lambda_1, \ldots, \lambda_n\right).$$

2. For the second part, consider $T : \mathbb{R}^n \to \mathbb{R}^n$ by $T(\mathbf{v}) = \mathbf{A}\mathbf{v}$. Since $A^{\mathrm{T}} = A$, we imply T is self-adjoint. There exists orthonormal basis $\mathcal{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ such that

$$T_{\mathcal{B},\mathcal{B}} = \text{diag}\left(\lambda_1, \ldots, \lambda_n\right).$$

In particular, if $\mathcal{A} = \{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$, then $T_{\mathcal{A},\mathcal{A}} = A$. We construct $P := \mathcal{C}_{\mathcal{A},\mathcal{B}}$, which is the change of basis matrix from $\mathcal{B}$ to $\mathcal{A}$, then

$$P = \left(\begin{array}{ccc} \mathbf{v}_1 & \cdots & \mathbf{v}_n \end{array} \right)$$

and

$$P^{-1}T_{\mathcal{A},\mathcal{A}}P = T_{\mathcal{B},\mathcal{B}}$$

Or equivalently, $P^{-1}AP = \mathrm{diag}\,(\lambda_1, \ldots, \lambda_n)$, with

$$P^{\mathrm{T}}P = \begin{pmatrix} \mathbf{v}_1^{\mathrm{T}} \\ \vdots \\ \mathbf{v}_n^{\mathrm{T}} \end{pmatrix} \begin{pmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_n \end{pmatrix} = I$$

$\square$

10.2   Orthogonal and Unitary Operators

Definition 10.12 (Orthogonal/Unitary Operator)**.** A linear operator $T : V \to V$ over $\mathbb{F}$ with $\langle T(\mathbf{w}), T(\mathbf{v}) \rangle = \langle \mathbf{w}, \mathbf{v} \rangle, \forall \mathbf{v}, \mathbf{w} \in V$, is called:

- Orthogonal if $\mathbb{F} = \mathbb{R}$

- Unitary if $\mathbb{F} = \mathbb{C}$

Proposition 10.13. *T is orthogonal (real case) or unitary (complex case) if and only if $T' \circ T = I$.*

Proof. The reverse direction is by directly checking that

$$\langle T(\mathbf{w}), T(\mathbf{v}) \rangle = \langle T' \circ T(\mathbf{w}), \mathbf{v} \rangle = \langle \mathbf{w}, \mathbf{v} \rangle.$$

The forward direction is by checking that $T' \circ T(\mathbf{w}) = \mathbf{w}$ for all $\mathbf{w} \in V$:

$$\langle T' \circ T(\mathbf{w}), \mathbf{v} \rangle = \langle T(\mathbf{w}), T(\mathbf{v}) \rangle = \langle \mathbf{w}, \mathbf{v} \rangle \Rightarrow \langle T' \circ T(\mathbf{w}) - \mathbf{w}, \mathbf{v} \rangle = 0, \quad \forall \mathbf{v} \in V.$$

By non-degeneracy of the inner product, $T' \circ T(\mathbf{w}) - \mathbf{w} = 0$, i.e.,

$$T' \circ T(\mathbf{w}) = \mathbf{w}, \quad \forall \mathbf{w} \in V.$$

and hence $T' \circ T = I$. $\square$

Example 10.14. Let $T : \mathbb{F}^n \to \mathbb{F}^n$ be given by $T(v) = Av$. Then:

- If $\mathbb{F} = \mathbb{R}$: T is orthogonal $\Leftrightarrow A^T A = I$

- If $\mathbb{F} = \mathbb{C}$: T is unitary $\Leftrightarrow A^{\mathrm{H}} A = I$

Definition 10.15 (Orthogonal/Unitary Group)**.**

$$\text{Orthogonal Group}: O\,(n) = \left\{ A \in M_{n \times n}\,(\mathbb{R}) \mid A^{\mathrm{T}} A = I \right\}$$

$$\text{Unitary Group}: U\,(n) = \left\{ A \in M_{n \times n}\,(\mathbb{C}) \mid A^{\mathrm{H}} A = I \right\}$$

Proposition 10.16. *Let* $T : V \to V$ *be a linear operator on a vector space over* $\mathbb{F}$ *satisfying* $T'T = I$. *Then for all eigenvalues* λ *of* T, *we have* $|\lambda| = 1$.

Proof. Suppose we have the eigen-pair $(\lambda, \mathbf{v})$, then

$$\langle T\mathbf{v}, T\mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{v} \rangle \Leftrightarrow \langle \lambda\mathbf{v}, \lambda\mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{v} \rangle \Leftrightarrow \overline{\lambda}\lambda\langle \mathbf{v}, \mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{v} \rangle$$

Since $\langle \mathbf{v}, \mathbf{v} \rangle \neq 0$ (i.e., $\mathbf{v} \neq 0$), we imply $|\lambda|^2 = 1$, i.e., $|\lambda| = 1$. $\qquad\square$

Proposition 10.17. *Let* $T : V \to V$ *be an operator on a finite dimension* V *over* $\mathbb{F}$ *satisfying* $T'T = I$. *If* $U \leq V$ *is* T-*invariant, then* U *is also* T^{-1}-*invariant.*

Proof. Since $T'T = I$, i.e., T is invertible, we imply 0 is not a root of $X_T(x)$, i.e., 0 is not a root of $m_T(x)$. Since $m_T(0) \neq 0$, $m_T(x)$ has the form

$$m_T(x) = x^m + \cdots + a_1 x + a_0, \quad a_0 \neq 0,$$

which follows that

$$m_T(T) = T^m + \cdots + a_0 I = 0 \Rightarrow T\left(T^{m-1} + \cdots + a_1 I\right) = -a_0 I$$

Or equivalently,

$$T\left(-\frac{1}{a_0}(T^{m-1} + \cdots + a_1 I)\right) = I$$

Therefore,

$$T^{-1} = -\frac{1}{a_0}T^{m-1} - \cdots - \frac{a_2}{a_0}T - \frac{a_1}{a_0}I,$$

i.e., the inverse T^{-1} can be expressed as a polynomial involving T only.

Since U is T-invariant, we imply U is T^m-invariant for $m \in \mathbb{N}$, and therefore U is T^{-1}-invariant since T^{-1} is a polynomial of T. $\qquad\square$

Proposition 10.18. *Let* $T : V \to V$ *satisfy* $T'T = I$ ($\dim(V) < \infty$), *then if* $U \leq V$ *is* T-*invariant, it implies* $U^{\perp}$ *is* T-*invariant.*

Proof. Let $\mathbf{v} \in U^{\perp}$, it suffices to show $T(\mathbf{v}) \in U^{\perp}$. Indeed, for all $\mathbf{u} \in U$, we have

$$\langle \mathbf{u}, T(\mathbf{v}) \rangle = \langle T'(\mathbf{u}), \mathbf{v} \rangle = \langle T^{-1}(\mathbf{u}), \mathbf{v} \rangle$$

Since U is T^{-1}-invariant, we imply $T^{-1}(\mathbf{u}) \in U$, and therefore

$$\langle \mathbf{u}, T(\mathbf{v}) \rangle = \langle T^{-1}(\mathbf{u}), \mathbf{v} \rangle = 0 \Rightarrow T(\mathbf{v}) \in U^{\perp}.$$

$\qquad\square$

Theorem 10.19. *Let* $T : V \to V$ *be a unitary operator on finite dimension* V *(over* $\mathbb{C}$*),*

then there exists an orthonormal basis $\mathcal{A}$ such that

$$T_{\mathcal{A},\mathcal{A}} = \operatorname{diag}(\lambda_1, \ldots, \lambda_n), |\lambda_i| = 1, \forall i.$$

Proof. Note that $\mathcal{X}_T(x)$ always admits a root in $\mathbb{C}$, so we can always find an eigenvector $v \in V$ of T.

Then the theorem follows by the same argument before on self-adjoint operators.

- Consider $U = \operatorname{span}\{\mathbf{v}\}$

- $V = U \oplus U^\perp$ and $U^\perp$ is T -invariant

- Use induction on the unitary operator $T|_{U^\perp} : U^\perp \to U^\perp$

$\square$

The argument above fails for orthogonal operators. For instance, let $T : \mathbb{R} \to \mathbb{R}^2$ be given by

$$T(\mathbf{v}) = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \mathbf{v}$$

The matrix A is not diagonalizable over $\mathbb{R}$. It has no real eigenvalues. However, if we treat A as $T : \mathbb{C}^2 \to \mathbb{C}^2$ with $T(\mathbf{v}) = \mathbf{A}\mathbf{v}$, then $A^H A = I$, and therefore T is unitary. Then A is diagonalizable over $\mathbb{C}$ with eigenvalues $e^{i\theta}, e^{-i\theta}$

- As a corollary of the theorem, for all $A \in M_{n \times n}(\mathbb{C})$ satisfying $A^H A = I$, there exists $P \in M_{n \times n}(\mathbb{C})$ such that

$$P^{-1} A P = \operatorname{diag}(\lambda_1, \ldots, \lambda_n), \ |\lambda_i| = 1,$$

where $P = (\mathbf{u}_1, \ldots, \mathbf{u}_n)$, with $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ forming orthonormal basis of $\mathbb{C}^n$. In fact,

$$P^H P = \begin{pmatrix} \mathbf{u}_1^H \\ \vdots \\ \mathbf{u}_n^H \end{pmatrix} (\mathbf{u}_1 \cdots \mathbf{u}_n) = \begin{pmatrix} \langle \mathbf{u}_1, \mathbf{u}_1 \rangle & \cdots & \langle \mathbf{u}_1, \mathbf{u}_n \rangle \\ \vdots & \ddots & \vdots \\ \langle \mathbf{u}_n, \mathbf{u}_1 \rangle & \cdots & \langle \mathbf{u}_n, \mathbf{u}_n \rangle \end{pmatrix}$$

Conclusion: all matrices $A \in M_{n \times n}(\mathbb{C})$ with $A^H A = I$ can be written as

$$A = P^{-1} \operatorname{diag}(\lambda_1, \ldots, \lambda_n) P,$$

with some P satisfying $P^H P = I$.

Notation. Let $U(n) = \left\{ A \in M_{n \times n}(\mathbb{C}) \mid A^H A = I \right\}$ be the unitary group, then all $A \in U(n)$ can be diagonalized by

$$A = P^{-1} \operatorname{diag}(\lambda_1, \ldots, \lambda_n) P, \ P \in U(n).$$

10.3 Normal Operators

Definition 10.20 (Normal Operator)**.** Let $T : V \to V$ be a linear operator over a $\mathbb{C}$-inner product vector space V. We say that T is **normal** if

$$T'T = TT',$$

where T' denotes the adjoint of T.

Example 10.21 (Examples of Normal Operators)**.**

- All self-adjoint operators are normal:

$$T = T' \ \Rightarrow \ TT' = T'T = T^2.$$

- All (finite-dimensional) unitary operators are normal:

$$T'T = TT' = I.$$

Proposition 10.22. *Let T be a normal operator on V. Then:*

1. *$\|T(\mathbf{v})\| = \|T'(\mathbf{v})\|$ for all $\mathbf{v} \in V$.*
   In particular, $T(\mathbf{v}) = 0$ if and only if $T'(\mathbf{v}) = 0$.

2. *$(T - \lambda I)$ is also a normal operator for any $\lambda \in \mathbb{C}$.*

3. *$T(\mathbf{v}) = \lambda \mathbf{v}$ if and only if $T'(\mathbf{v}) = \bar{\lambda}\mathbf{v}$.*

Proof.   1. One has $\langle \mathbf{v}, T\mathbf{v} \rangle = \langle T'\mathbf{v}, \mathbf{v} \rangle = \langle T'T\mathbf{v}, \mathbf{v} \rangle = \langle TT'\mathbf{v}, \mathbf{v} \rangle = \langle T'\mathbf{v}, T'\mathbf{v} \rangle = \|T'(\mathbf{v})\|^2$. There-
   fore, $\|T(\mathbf{v})\|^2 = \|T'(\mathbf{v})\|^2 \Rightarrow \|T(\mathbf{v})\| = \|T'(\mathbf{v})\|$.

2. By the hypothesis that T is normal, we have $T'T = TT'$. We check:

$$(T - \lambda I)'(T - \lambda I) = (T - \lambda I)(T - \lambda I)'.$$

Expanding both sides:

$$(T - \lambda I)'(T - \lambda I) = (T' - \bar{\lambda}I)(T - \lambda I) = T'T - \bar{\lambda}T - \lambda T' + |\lambda|^2 I,$$
$$(T - \lambda I)(T - \lambda I)' = TT' - \lambda T' - \bar{\lambda}T + |\lambda|^2 I.$$

These expressions are equal since $T'T = TT'$, hence $T - \lambda I$ is normal.

3. **Forward direction:** If $(T - \lambda I)\mathbf{v} = 0$, then $T\mathbf{v} = \lambda \mathbf{v}$. By (2), $T - \lambda I$ is normal, so:

$$\|(T - \lambda I)'(\mathbf{v})\| = 0 \Rightarrow (T - \lambda I)'(\mathbf{v}) = 0 \Rightarrow T'\mathbf{v} = \bar{\lambda}\mathbf{v}.$$

Reverse direction: Suppose $(T' - \bar{\lambda}I)\mathbf{v} = 0$, i.e., $T'\mathbf{v} = \bar{\lambda}\mathbf{v}$. Then:

$$(T' - \bar{\lambda}I)\mathbf{v} = 0 \Rightarrow ((T')' - \lambda I)\mathbf{v} = 0,$$

but $(T')' = T$ by the self-adjoint property in Hilbert spaces. Thus, $(T - \lambda I)\mathbf{v} = 0$.

4. Suppose $T\mathbf{v} = \lambda\mathbf{v}$, $T\boldsymbol{w} = \mu\boldsymbol{w}$, and $\lambda \neq \mu$. Then:

$$\lambda\langle\mathbf{v}, \boldsymbol{w}\rangle = \langle T\mathbf{v}, \boldsymbol{w}\rangle = \langle\mathbf{v}, T'\boldsymbol{w}\rangle = \langle\mathbf{v}, \bar{\mu}\boldsymbol{w}\rangle = \bar{\mu}\langle\mathbf{v}, \boldsymbol{w}\rangle.$$

Since $\lambda \neq \bar{\mu}$, we must have $\langle\mathbf{v}, \boldsymbol{w}\rangle = 0$. The proof is complete.

$\square$

Theorem 10.23. *Let T be an operator on a finite dimensional $(\dim(V) = n)\,\mathbb{C}$ -inner product vector space V satisfying $T'T = TT'$. Then there is an orthonormal basis of eigenvectors of V, i.e., an orthonormal basis of V such that any element from this basis is an eigenvector of T.*

Proof. Since $\mathcal{X}_T(x)$ must have a root in $\mathbb{C}$, there must exist an eigen-pair $(\mathbf{v}, \lambda)$ of T.

- Construct $U = \text{span}\{\mathbf{v}\}$, and it follows that

$$T\mathbf{v} = \lambda\mathbf{v} \Rightarrow U \text{ is } T \text{ -invariant.}$$

$$T'\mathbf{v} = \bar{\lambda}\mathbf{v} \Rightarrow U \text{ is } T' \text{ -invariant.}$$

- Moreover, we claim that $U^\perp$ is T and T' invariant: let $\mathbf{w} \in U^\perp$, and for all $\mathbf{u} \in U$, we have

$$\langle\mathbf{u}, T(\mathbf{w})\rangle = \langle T'(\mathbf{u}), \mathbf{w}\rangle = \langle\bar{\lambda}\mathbf{u}, \mathbf{w}\rangle = \lambda\langle\mathbf{u}, \mathbf{w}\rangle = 0,$$

i.e., $U^\perp$ is T invariant.

$$\langle\mathbf{u}, T'(\mathbf{w})\rangle = \langle T(\mathbf{u}), \mathbf{w}\rangle = \langle\lambda\mathbf{u}, \mathbf{w}\rangle = \bar{\lambda}\langle\mathbf{u}, \mathbf{w}\rangle = 0,$$

which implies $U^\perp$ is T' invariant.

- Therefore, we construct the operator $T|_{U^\perp} : U^\perp \to U^\perp$, and

$$TT' = T'T \Rightarrow \left(T|_{U^\perp}\right)\left(T'|_{U^\perp}\right) = \left(T'|_{U^\perp}\right)\left(T|_{U^\perp}\right),$$

i.e., $\left(T|_{U^\perp}\right)$ is normal on $U^\perp$. Moreover, $\dim\left(U^\perp\right) = n - 1$.

Applying the same trick as in Theorem 10.10, we imply there exists an orthonormal basis

$\{\mathbf{e}_2, \ldots, \mathbf{e}_n\}$ of eigenvectors of $(T|_{U^\perp})$. Then we can argue that

$$\mathcal{B} = \left\{ \mathbf{v}' = \frac{\mathbf{v}}{\| \mathbf{v} \|}, \mathbf{e}_2, \ldots, \mathbf{e}_{k+1} \right\}$$

is a basis of orthonormal eigenvectors of V. □

Corollary 10.24 (Spectral Theorem for Normal Operators). *Let $T : V \to V$ be a normal operator on a $\mathbb{C}$ -inner product space with $\dim(V) < \infty$. Then there exists self-adjoint operators $P_1, \ldots, P_k$ such that*

$$P_i^2 = P_i, \ P_i P_j = 0, i \neq j, \ \sum_{i=1}^{k} P_i = I,$$

and $T = \sum_{i=1}^{k} \lambda_i P_i$, where λ_i 's are the eigenvalues of T.

These P_i 's are the orthogonal projections from V to the λ_i -eigenspace $\ker(T - \lambda_i I)$ of T, i.e., we have

$$\mathbf{v} = P_i(\mathbf{v}) + (\mathbf{v} - P_i(\mathbf{v}))$$

where $P_i(\mathbf{v}) \in \ker(T - \lambda_i I)$, and $\mathbf{v} - P_i(\mathbf{v}) \in (\ker(T - \lambda_i I))^\perp$.

You should know how to compute the projections P_i when $T(\mathbf{v}) = A\mathbf{v}$ in MAT2040.

Proof. Since T has a basis of eigenvectors, it is diagonalizable. By Proposition (8.2),

$$m_T(x) = (x - \lambda_1) \cdots (x - \lambda_k),$$

with distinct eigenvalues λ_i. By Corollary (9.2), it suffices to show that each projection P_i is self-adjoint.

Recall that $P_i = a_i(T)q_i(T) := b_m T^m + \cdots + b_1 T + b_0 I$, i.e., a polynomial in T. Then,

$$P_i' = \bar{b}_m(T')^m + \cdots + \bar{b}_1 T' + \bar{b}_0 I.$$

We claim P_i is normal. Since $T'T = TT'$, we have:

$$(T')^p T^q = T^q (T')^p \quad \forall p, q \in \mathbb{N}.$$

Thus:

$$P_i P_i' = (b_m T^m + \cdots + b_0 I)\left(\bar{b}_m(T')^m + \cdots + \bar{b}_0 I\right) = \sum_{x,y=0}^{m} b_x \bar{b}_y T^x (T')^y$$

$$= \sum_{x,y=0}^{m} \bar{b}_y b_x (T')^y T^x = P_i' P_i.$$

In general, if an operator S is self-adjoint, then it is normal. The converse is not true unless all eigenvalues of S are real. But in this case:

By Theorem 10.23, a normal operator S is orthonormally diagonalizable, so its matrix representation in some orthonormal basis $\mathcal{B}$ is

$$S_{\mathcal{B},\mathcal{B}} = \mathrm{diag}(\lambda_1, \ldots, \lambda_k).$$

This same basis diagonalizes S' as well, and by part (3) of Proposition (12.1),

$$(S')_{\mathcal{B},\mathcal{B}} = \mathrm{diag}(\lambda_1, \ldots, \lambda_k).$$

Therefore, $S = S'$, so S is self-adjoint.

In particular, for $S = P_i$, all eigenvalues are 0 or 1, which are real. Hence, each P_i is self-adjoint. $\qquad\square$

Corollary 10.25. *Let $T : V \to V$ be a linear operator on a $\mathbb{C}$-inner product space with $\dim(V) < \infty$. Then T is normal if and only if $T' = f(T)$ for some polynomial $f(x) \in \mathbb{C}[x]$.*

Proof. For the reverse direction, if $T' = f(T)$, then

$$T'T = f(T)T = Tf(T) = TT'.$$

For the forward direction, suppose that T is normal. Then by Corollary 10.24, we have:

$$T = \sum_{i=1}^{k} \lambda_i P_i, \quad P_i = f_i(T),$$

where the P_i are self-adjoint. It follows that

$$T' = \left(\sum_{i=1}^{k} \lambda_i P_i \right)' = \sum_{i=1}^{k} \bar{\lambda}_i P_i' = \sum_{i=1}^{k} \bar{\lambda}_i P_i = \sum_{i=1}^{k} \bar{\lambda}_i f_i(T).$$

$\qquad\square$

Chapter 11

Introduction to Tensor Products

11.1 Motivation

Let U, V, W be vector spaces. We want to study bilinear maps $f : V \times W \to U$, i.e., for all $\mathbf{v}, \mathbf{v}_1, \mathbf{v}_2 \in V$, $\mathbf{w}, \mathbf{w}_1, \mathbf{w}_2 \in W$, $a, b, c, d \in \mathbb{F}$, one has

$$f(a\mathbf{v}_1 + b\mathbf{v}_2, \mathbf{w}) = af(\mathbf{v}_1, \mathbf{w}) + bf(\mathbf{v}_2, \mathbf{w})$$

$$f(\mathbf{v}, c\mathbf{w}_1 + d\mathbf{w}_2) = cf(v, \mathbf{w_1}) + df(\mathbf{v}, \mathbf{w}_2)$$

Example 11.1.   • $f : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ defined by the usual inner product $(\mathbf{u}, \mathbf{v}) \mapsto \langle \mathbf{u}, \mathbf{v} \rangle$.

- $f : M_{n \times n}(\mathbb{F}) \times M_{n \times n}(\mathbb{F}) \to M_{n \times n}(\mathbb{F})$ defined by matrix multiplication $f(A, B) = AB$.

- $f : \mathbb{F}[x] \times \mathbb{F}[x] \to \mathbb{F}[x]$ defined by multiplication $f(p(x), q(x)) = p(x)q(x)$.

- $f : \mathbb{F}[x] \times \mathbb{F}[x] \to \mathbb{F}$ defined by $f(p(x), q(x)) = p(1)q(2)$.

- $f : \mathbb{R}^3 \times \mathbb{R}^3$ defined by the *cross product* $f(\mathbf{u}, \mathbf{v}) = \mathbf{u} \times \mathbf{v}$.

Unfortunately, bilinear maps are almost always **not a linear transformation**. For instance, in the last example, one has:

$$f(3\mathbf{v}, 3\mathbf{w}) = (3\mathbf{v}) \times (3\mathbf{w}) = 9\mathbf{v} \times \mathbf{w} \neq 3f(\mathbf{v}, \mathbf{w}).$$

Since f is not a linear transformation, one cannot apply any of the tools (e.g., matrix representations, rank-nullity theorem, etc.) we developed in this course so far to study f.

Indeed, the fundamental issue is that the vector space structure of $V \times W$ is not suited to studying bilinear maps.
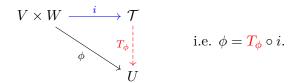
As a consequence, we begin by giving an abstract, category-theoretic definition of tensor product $V \otimes W$.

11.1.1 Universal Property of Tensor Products

Definition 11.2 (Universal Property of Tensor Product)**.** Let V, W be vector spaces. Consider the set

$$\text{Obj} := \{\phi : V \times W \to U \mid \phi \text{ is a bilinear map }\}.$$

We say the **tensor product space** $\mathcal{T}$, or the bilinear map $(i : V \times W \to \mathcal{T}) \in Obj$ satisfies the **universal property of tensor product** if for any $(\phi : V \times W \to U) \in \text{Obj}$, there exists an unique linear transformation $T_\phi : \mathcal{T} \to U$ such that the diagram below commutes:

$$
\begin{array}{ccc}
V \times W & \xrightarrow{\ \ i \ \ } & \mathcal{T} \\
& {\scriptstyle \phi} \searrow & \big\downarrow {\scriptstyle T_\phi} \\
& & U
\end{array}
\qquad \text{i.e. } \phi = T_\phi \circ i.
$$

In other words, rather than studying the **bilinear map** ϕ, it is better to study the **linear transformation** T_ϕ. Since $\phi = T_\phi \circ i$, T_ϕ contains all the information about ϕ, and one can apply all the theorems we know about linear algebra to study T_ϕ!

The question is: Does the tensor product space $\mathcal{T}$ exist?

In the next section, we will construct $\mathcal{T}$ explicitly, and show that it satisfies the properties we mentioned above.

11.2 Tensor Product Space

We begin with defining the tensor product $\mathcal{T} = V \otimes W$ of two vector spaces V and W. This can be generalized into any (countable) number of vector spaces.

Definition 11.3. Let V, W be vector spaces. Let $S = \{(\mathbf{v}, \mathbf{w}) \mid \mathbf{v} \in V, \mathbf{w} \in W\}$, we define

$$\mathfrak{X} = \text{span}(S).$$

Remark 11.4. Note that we assume no relations on the elements $(\mathbf{v}, \mathbf{w}) \in \mathcal{S}$. In other words, $1 \cdot (\mathbf{v}, \mathbf{w})$ and $1 \cdot (\mathbf{v}', \mathbf{w}') \in \mathfrak{X}$ are linearly independent unless $\mathbf{v} = \mathbf{v}'$ and $\mathbf{w} = \mathbf{w}'$

For example, if $\mathbf{w} \neq \mathbf{0}$, then $\mathbf{w} \neq 2\mathbf{w}$. Therefore, $(0, \mathbf{w})$ and $(0, 2\mathbf{w})$ are linearly independent and hence:

$$2 \cdot (0, \mathbf{w}) \neq (0, 2\mathbf{w})$$
$$3 \cdot (0, \mathbf{w}) \neq 1 \cdot (0, \mathbf{w}) + 1 \cdot (0, 2\mathbf{w})$$

Similarly, if $\mathbf{v}, \mathbf{w} \neq \mathbf{0}$:

$$(\mathbf{v}, \mathbf{w}) \neq (\mathbf{v}, 0) + (0, \mathbf{w}).$$

The only legitimate relationship in $\mathfrak{X}$ is

$$2 \cdot (\mathbf{v}, \mathbf{w}) + 3 \cdot (\mathbf{v}, \mathbf{w}) = 5 \, (\mathbf{v}, \mathbf{w}),$$

yet it is not equal to $(5\mathbf{v}, 5\mathbf{w})$.

In other words, $\mathcal{S}$ is a basis of $\mathfrak{X}$, and consequently $\mathfrak{X}$ is of uncountable dimension.

Definition 11.5 (Tensor Product of V and W)**.** Let $\mathfrak{Y} \leq \mathfrak{X}$ be a vector subspace spanned by vectors of the form

$$\{1 \, (\mathbf{v}_1 + \mathbf{v}_2, \mathbf{w}) - 1 \, (\mathbf{v}_1, \mathbf{w}) - 1 \, (\mathbf{v}_2, \mathbf{w})\}, \quad \{1 \, (\mathbf{v}, \mathbf{w}_1 + \mathbf{w}_2) - 1 \, (\mathbf{v}, \mathbf{w}_1) - 1 \, (\mathbf{v}, \mathbf{w}_2)\}$$

and

$$\{1 \, (k\mathbf{v}, \mathbf{w}) - k \, (\mathbf{v}, \mathbf{w}) \mid k \in \mathbb{F}\}, \quad \{1 \, (\mathbf{v}, k\mathbf{w}) - k \, (\mathbf{v}, \mathbf{w}) \mid k \in \mathbb{F}\}$$

for all $\mathbf{v}, \mathbf{v}_1, \mathbf{v}_2 \in V$, $\mathbf{w}, \mathbf{w}_1, \mathbf{w}_2 \in W$. Then the **tensor product** $V \otimes W$ is defined by

$$V \otimes W := \mathfrak{X}/\mathfrak{Y}$$

Also, for $\mathbf{v} \in V$ and $\mathbf{w} \in W$, we define

$$\mathbf{v} \otimes \mathbf{w} := (\mathbf{v}, \mathbf{w}) + \mathfrak{Y} \quad \in \mathfrak{X}/\mathfrak{Y} = V \otimes W.$$

Using our definition of $V \otimes W$, the expression $\mathbf{v} \otimes \mathbf{w} \in V \otimes W$ is 'bilinear', for instance:

$$\begin{aligned}
(\mathbf{v}_1 + \mathbf{v}_2) \otimes \mathbf{w} &= (\mathbf{v}_1 + \mathbf{v}_2, \mathbf{w}) + \mathfrak{Y} \\
&= (\mathbf{v}_1 + \mathbf{v}_2, \mathbf{w}) - [(\mathbf{v}_1 + \mathbf{v}_2, \mathbf{w}) - (\mathbf{v}_1, \mathbf{w}) - (\mathbf{v}_2, \mathbf{w})] + \mathfrak{Y} \\
&= 0 \, (\mathbf{v}_1 + \mathbf{v}_2, \mathbf{w}) + (\mathbf{v}_1, \mathbf{w}) + (\mathbf{v}_2, \mathbf{w}) + \mathfrak{Y} \\
&= [(\mathbf{v}_1, \mathbf{w}) + \mathfrak{Y}] + [(\mathbf{v}_2, \mathbf{w}) + \mathfrak{Y}] \\
&= \mathbf{v}_1 \otimes \mathbf{w} + \mathbf{v}_2 \otimes \mathbf{w}
\end{aligned} \tag{11.1}$$

Similarly, one can check that

$$\begin{aligned}
\mathbf{v} \otimes (\mathbf{w}_1 + \mathbf{w}_2) &= (\mathbf{v} \otimes \mathbf{w}_1) + (\mathbf{v} \otimes \mathbf{w}_2) \\
(k\mathbf{v}) \otimes \mathbf{w} &= k \, (\mathbf{v} \otimes \mathbf{w}) \\
\mathbf{v} \otimes (k\mathbf{w}) &= k \, (\mathbf{v} \otimes \mathbf{w})
\end{aligned} \tag{11.2}$$

Making use of the rules above, we present an example of arithmetic on tensor product spaces:

Example 11.6. Let $V = W = \mathbb{R}^2$, with $\mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, $\mathbf{e}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$. Then

$$\begin{pmatrix} 3 \\ 1 \end{pmatrix} \otimes \begin{pmatrix} -4 \\ 2 \end{pmatrix} = (3\mathbf{e}_1 + 2\mathbf{e}_2) \otimes (-4\mathbf{e}_1 + 2\mathbf{e}_2)$$

$$= (3\mathbf{e}_1) \otimes (-4\mathbf{e}_1 + 2\mathbf{e}_2) + (\mathbf{e}_2) \otimes (-4\mathbf{e}_1 + 2\mathbf{e}_2)$$
$$= (3\mathbf{e}_1) \otimes (-4\mathbf{e}_1) + (3\mathbf{e}_1) \otimes (2\mathbf{e}_2) + (\mathbf{e}_2) \otimes (-4\mathbf{e}_1) + \mathbf{e}_2 \otimes (2\mathbf{e}_2)$$
$$= -12 (\mathbf{e}_1 \otimes \mathbf{e}_1) + 6 (\mathbf{e}_1 \otimes \mathbf{e}_2) - 4 (\mathbf{e}_2 \otimes \mathbf{e}_1) + 2 (\mathbf{e}_2 \otimes \mathbf{e}_2)$$

Exercise: Check that $\mathbf{e}_1 \otimes \mathbf{e}_2 + \mathbf{e}_2 \otimes \mathbf{e}_1$ cannot be re-written as

$$(a\mathbf{e}_1 + b\mathbf{e}_2) \otimes (c\mathbf{e}_1 + d\mathbf{e}_2).$$

for any $a, b, c, d \in \mathbb{R}$.

Remark 11.7. The product space $V \times W$ is different from the tensor product space $V \otimes W$ in the following sense:

(a) $(\mathbf{v}, \mathbf{0}) \neq \mathbf{0}_{V \times W}$ in $V \times W$; but $\mathbf{v} \otimes \mathbf{0} \in \mathbf{0}_{V \otimes W}$, since

$$\mathbf{v} \otimes 0 = \mathbf{v} \otimes (0\mathbf{w}) = 0 (\mathbf{v} \otimes \mathbf{w}) = \mathbf{0}_{V \otimes W}$$

(b) $(\mathbf{v}_1, \mathbf{w}_1) + (\mathbf{v}_2, \mathbf{w}_2) = (\mathbf{v}_1 + \mathbf{v}_2, \mathbf{w}_1 + \mathbf{w}_2)$; but $\mathbf{v}_1 \otimes \mathbf{w}_1 + \mathbf{v}_2 \otimes \mathbf{w}_2$ cannot be simplified further in general, unless (for instance) $\mathbf{v}_1 = \mathbf{v}_2$, so that:

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

As we saw in the exercise above, a general element in $V \otimes W$ **cannot** is not necessarily of the form $v \otimes w$. How does a general element in $V \otimes W$ look like?

We begin with a general element in $\mathfrak{X}$:

$$a_1 (\mathbf{v}_1, \mathbf{w}_1) + \cdots + a_k (\mathbf{v}_k, \mathbf{w}_k),$$

where $(\mathbf{v}_i, \mathbf{w}_i)$ are distinct. Then a general element in $\mathfrak{X}/\mathfrak{Y} := V \otimes W$ looks like:

$$a_1 (\mathbf{v}_1, \mathbf{w}_1) + \cdots + a_k (\mathbf{v}_k, \mathbf{w}_k) + \mathfrak{Y} = a_1 ((\mathbf{v}_1, \mathbf{w}_1) + \mathfrak{Y}) + \cdots + a_k ((\mathbf{v}_k, \mathbf{w}_k) + \mathfrak{Y})$$
$$= a_1 (\mathbf{v}_1 \otimes \mathbf{w}_1) + \cdots + a_k (\mathbf{v}_k \otimes \mathbf{w}_k)$$
$$= (a_1\mathbf{v}_1) \otimes \mathbf{w}_1 + \cdots + (a_k\mathbf{v}_k) \otimes \mathbf{w}_k$$

Therefore, a general element in $V \otimes W$ is of the form

$$\mathbf{v}^{(1)} \otimes \mathbf{w}^{(1)} + \cdots + \mathbf{v}^{(k)} \otimes \mathbf{w}^{(k)} \qquad (\mathbf{v}^{(i)} \in V, \quad \mathbf{w}^{(i)} \in W). \qquad (11.3)$$

Theorem 11.8. *The bilinear map*

$$i : V \times W \to V \otimes W$$

defined by

$$i(\mathbf{v}, \mathbf{w}) := \mathbf{v} \otimes \mathbf{w}$$

is in Obj *in Definition 11.2 (i.e. i is a bilinear map). Moreover, by taking $\mathcal{T} := V \otimes W$ in Definition 11.2, it satisfies the universal property of tensor products.*

Proof. By Equations (11.1) and (11.2), i satisfies:

$$i(\mathbf{v}_1 + \mathbf{v}_2, \mathbf{w}) = i(\mathbf{v}_1, \mathbf{w}) + i(\mathbf{v}_2, \mathbf{w}),$$
$$i(\mathbf{v}, \mathbf{w}_1 + \mathbf{w}_2) = i(\mathbf{v}, \mathbf{w}_1) + i(\mathbf{v}, \mathbf{w}_2),$$
$$i(a\mathbf{v}, \mathbf{w}) = a \cdot i(\mathbf{v}, \mathbf{w}), \quad i(\mathbf{v}, b\mathbf{w}) = b \cdot i(\mathbf{v}, \mathbf{w}),$$
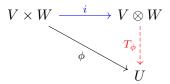
so $i \in$ Obj as in Definition 11.2.

Now, let $(\phi : V \times W \to U) \in$ Obj be any bilinear map into some vector space U. Since ϕ is bilinear, it respects the relations that generate $\mathfrak{Y}$. Therefore, we can define a linear map

$$T_\phi : V \otimes W \to U, \quad T_\phi(\mathbf{v} \otimes \mathbf{w}) := \phi(\mathbf{v}, \mathbf{w}),$$

and extend linearly to all of $V \otimes W$ (recall all elements of $V \otimes W$ are of the form $\mathbf{v}^{(1)} \otimes \mathbf{w}^{(1)} + \cdots + \mathbf{v}^{(k)} \otimes \mathbf{w}^{(k)}$). This map is well-defined because if two elements $(\mathbf{v}, \mathbf{w})$ and $(\mathbf{v}', \mathbf{w}')$ are equivalent in $\mathfrak{X}/\mathfrak{Y}$, i.e. $(\mathbf{v}, \mathbf{w}) - (\mathbf{v}', \mathbf{w}') \in \mathfrak{Y}$, then their images under ϕ are equal due to bilinearity.

It follows directly from the definition that the diagram

$$V \times W \xrightarrow{\quad i \quad} V \otimes W$$
$$\phi \searrow \quad \downarrow T_\phi$$
$$U$$

commutes, i.e., $\phi = T_\phi \circ i$.

For uniqueness: suppose there exists another linear map $T' : V \otimes W \to U$ such that $\phi = T' \circ i$. Then both T_ϕ and T' agree on all pure tensors $\mathbf{v} \otimes \mathbf{w}$, and hence, by linearity, on all of $V \otimes W$. So $T' = T_\phi$.

Therefore, the pair $(V \otimes W, i)$ satisfies the universal property of the tensor product. $\qquad \square$

11.3   Basis of $V \otimes W$

Given that $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is a basis of V, and $\{\mathbf{w}_1, \ldots, \mathbf{w}_m\}$ a basis of W, we aim to find a basis of $V \otimes W$ using $\mathbf{v}_i$'s and $\mathbf{w}_i$'s.

Proposition 11.9. *The set $\{\mathbf{v}_i \otimes \mathbf{w}_j \mid 1 \leq i \leq n, 1 \leq j \leq m\}$ spans the tensor product space $V \otimes W$.*

Proof. Consider any $\mathbf{v} \in V$ and $\mathbf{w} \in W$, and we want to express $\mathbf{v} \otimes \mathbf{w}$ in terms of $\mathbf{v}_i, \mathbf{w}_j$. Suppose that $\mathbf{v} = \alpha_1 \mathbf{v}_1 + \cdots + \alpha_n \mathbf{v}_n$ and $\mathbf{w} = \beta_1 \mathbf{w}_1 + \cdots + \beta_m \mathbf{w}_m$.

Substituting $\mathbf{v} = \alpha_1 \mathbf{v}_1 + \cdots + \alpha_n \mathbf{v}_n$ into the expression $\mathbf{v} \otimes \mathbf{w}$, we imply

$$
\begin{aligned}
\mathbf{v} \otimes \mathbf{w} &= (\alpha_1 \mathbf{v}_1 + \cdots + \alpha_n \mathbf{v}_n) \otimes \mathbf{w} \\
&= (\alpha_1 \mathbf{v}_1) \otimes \mathbf{w}_1 + \cdots + (\alpha_n \mathbf{v}_n) \otimes \mathbf{w}_n \\
&= \alpha_1 (\mathbf{v}_1 \otimes \mathbf{w}) + \cdots + \alpha_n (\mathbf{v}_n \otimes \mathbf{w})
\end{aligned}
$$

Similarly, for each $\mathbf{v}_i \otimes \mathbf{w}, i = 1, \ldots, n$,

$$
\mathbf{v}_i \otimes \mathbf{w} = \beta_1 (\mathbf{v}_i \otimes \mathbf{w}_1) + \cdots + \beta_m (\mathbf{v}_i \otimes \mathbf{w}_m).
$$

Therefore,

$$
\mathbf{v} \otimes \mathbf{w} = \sum_{i=1}^{n} \sum_{j=1}^{m} \alpha_i \beta_j (\mathbf{v}_i \otimes \mathbf{w}_j) \tag{11.4}
$$

By Equation 11.3, any vector in $V \otimes W$ is of the form $\mathbf{v}^{(1)} \otimes \mathbf{w}^{(1)} + \cdots + \mathbf{v}^{(k)} \otimes \mathbf{w}^{(k)}$, and by Equation 11.4, each $\mathbf{v}^{(\ell)} \otimes \mathbf{w}^{(\ell)}, \ell = 1, \ldots, k$, can be expressed as

$$
\mathbf{v}^{(\ell)} \otimes \mathbf{w}^{(\ell)} = \sum_{i=1}^{n} \sum_{j=1}^{m} \alpha_i^{(\ell)} \beta_j^{(\ell)} (\mathbf{v}_i \otimes \mathbf{w}_j)
$$

Therefore,

$$
\mathbf{v}^{(1)} \otimes \mathbf{w}^{(1)} + \cdots + \mathbf{v}^{(k)} \otimes \mathbf{w}^{(k)} = \sum_{\ell=1}^{k} \sum_{i=1}^{n} \sum_{j=1}^{m} \alpha_i^{(k)} \beta_j^{(k)} (\mathbf{v}_i \otimes \mathbf{w}_j)
$$

In other words, $\{\mathbf{v}_i \otimes \mathbf{w}_j \mid 1 \leq i \leq n, 1 \leq j \leq m\}$ spans $V \otimes W$. $\square$

Theorem 11.10. *A basis of $V \otimes W$ is $\{\mathbf{v}_i \otimes \mathbf{w}_j \mid 1 \leq i \leq n, 1 \leq j \leq m\}$*

Proof. By Proposition 11.9, it suffices to show that the set $\{\mathbf{v}_i \otimes \mathbf{w}_j \mid 1 \leq i \leq n, 1 \leq j \leq m\}$ is linear independent. Suppose that

$$
\sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_{ij} (\mathbf{v}_i \otimes \mathbf{w}_j) = \mathbf{0} \tag{13.3}
$$

Suppose that $\{\phi_1, \ldots, \phi_n\}$ is a dual basis of V^*, and $\{\psi_1, \ldots, \psi_m\}$ is a dual basis of W^*. Define the mapping

$$
\pi_{p,q} : V \times W \to \mathbb{F}
$$

by

$$
\pi_{p,q}(\mathbf{v}, \mathbf{w}) := \phi_p (\mathbf{v}) \, \psi_q (\mathbf{w}).
$$

Then one can check that $\pi_{p,q}$ is bilinear: for instance,

$$\pi_{p,q}\left(a\mathbf{v}_1 + b\mathbf{v}_2, \mathbf{w}\right) = \phi_p\left(a\mathbf{v}_1 + b\mathbf{v}_2\right)\psi_q\left(\mathbf{w}\right) = \left(a\phi_p\left(\mathbf{v}_1\right) + b\phi_p\left(\mathbf{v}_2\right)\right)\psi_q\left(\mathbf{w}\right)$$
$$= a\phi_p\left(\mathbf{v}_1\right)\psi_q\left(\mathbf{w}\right) + b\phi_p\left(\mathbf{v}_2\right)\psi_q\left(\mathbf{w}\right)$$
$$= a\pi_{p,q}\left(\mathbf{v}_1, \mathbf{w}\right) + b\pi_{p,q}\left(\mathbf{v}_2, \mathbf{w}\right).$$

similarly, we can check that $\pi_{p,q}\left(\mathbf{v}, a\mathbf{w}_1 + b\mathbf{w}_2\right) = a\pi_{p,q}\left(\mathbf{v}, \mathbf{w}_1\right) + b\pi_{p,q}\left(v, \mathbf{w}_2\right)$. Therefore, $\pi_{p,q} \in$ Obj. By the universal property of the tensor product (Theorem 11.8), $\pi_{p,q}$ induces the unique *linear transformation*

$$\Pi_{p,q}: \ V \otimes W \to \mathbb{F} \quad \text{with} \ \Pi_{p,q}\left(\mathbf{v} \otimes \mathbf{w}\right) = \pi_{p,q}\left(\mathbf{v}, \mathbf{w}\right)$$

In other words, $\Pi_{p,q}\left(\mathbf{v} \otimes \mathbf{w}\right) = \phi_p\left(\mathbf{v}\right)\psi_q\left(\mathbf{w}\right)$.

- Applying the mapping $\Pi_{p,q}$ on both sides of Equation 13.3, we imply

$$\Pi_{p,q}\left(\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_{ij}\left(\mathbf{v}_i \otimes \mathbf{w}_j\right)\right) = \Pi_{p,q}\left(\mathbf{0}\right)$$

Or equivalently,

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_{ij}\Pi_{p,q}\left(\mathbf{v}_i \otimes \mathbf{w}_j\right) = 0,$$

i.e.,

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_{ij}\phi_p\left(\mathbf{v}_i\right)\psi_q\left(\mathbf{w}_j\right) = \alpha_{p,q} = 0$$

Following this procedure, we can argue that $\alpha_{ij} = 0$ for all i and j. $\quad\square$

Corollary 11.11. *If* $\dim\left(V\right), \dim\left(W\right) < \infty$, *then* $\dim\left(V \otimes W\right) = \dim\left(V\right)\dim\left(W\right)$

Proof. Check dimension of the basis of $V \otimes W$. $\quad\square$

11.4 An Application of Universal Property

As we saw in the proof of Theorem 11.10, the universal property (Theorem 11.8) is very helpful. In particular, whenever $\phi : V \times W \to U$ is a *bilinear map*, Theorem 11.8 implies that we can induce an unique *linear transformation* $\psi : V \otimes W \to U$. Let's try another example for applying universal property:

Theorem 11.12. *For any vector spaces* U *and* V,

$$V \otimes U \cong U \otimes V.$$

Proof. Construct the mapping

$$\phi : \; V \times U \to U \otimes V$$

by $\phi(\mathbf{v}, \mathbf{u}) = \mathbf{u} \otimes \mathbf{v}$. Then ϕ is bilinear: for instance,

$$\phi(a\mathbf{v}_1 + b\mathbf{v}_2, \mathbf{u}) = u \otimes (a\mathbf{v}_1 + b\mathbf{v}_2) = a(\mathbf{u} \otimes \mathbf{v}_1) + b(u \otimes \mathbf{v}_2) = a\phi(\mathbf{v}_1, \mathbf{u}) + b\phi(\mathbf{v}_2, \mathbf{u}).$$

Therefore, $\phi \in \mathrm{Obj}$. By Theorem Theorem 11.8, we induce a unique linear transformation

$$\Phi : \; V \otimes U \to U \otimes V$$

satisfying $\Phi(\mathbf{v} \otimes \mathbf{u}) = \mathbf{u} \otimes \mathbf{v}$.

Similarly, we may induce the linear transformation

$$\Psi : \; U \otimes V \to V \otimes U$$

satisfying $\Psi(\mathbf{u} \otimes \mathbf{v}) = \mathbf{v} \otimes \mathbf{u}$.

Given any $\sum_i \mathbf{u}_i \otimes \mathbf{v}_i \in U \otimes V$, observe that

$$(\Phi \circ \Psi)\left(\sum_i \mathbf{u}_i \otimes \mathbf{v}_i\right) = \Phi\left(\sum_i \Psi(\mathbf{u}_i \otimes \mathbf{v}_i)\right) = \Phi\left(\sum_i \mathbf{v}_i \otimes \mathbf{u}_i\right) = \sum_i \Phi(\mathbf{v}_i \otimes \mathbf{u}_i) = \sum_i \mathbf{u}_i \otimes \mathbf{v}_i$$

Therefore, $\Phi \circ \Psi = \mathrm{id}_{U \otimes V}$. Similarly, $\Psi \circ \Phi = \mathrm{id}_{V \otimes U}$. Therefore, $U \otimes V \cong V \otimes U$. $\square$

11.5   Tensor Product of Linear Transformation

Given two linear transformations $T : V \to V'$ and $S : W \to W'$, we want to use the universal property of tensor product to construct the linear transformation between tensor product spaces:

$$T \otimes S : V \otimes W \to V' \otimes W'$$

Proposition 11.13. *Suppose that $T : V \to V'$ and $S : W \to W'$ are linear transformations, then there exists a unique linear transformation*

$$T \otimes S : V \otimes W \to V' \otimes W'$$

satisfying $(T \otimes S)(\mathbf{v} \otimes \mathbf{w}) = T(\mathbf{v}) \otimes S(\mathbf{w})$ for all $\mathbf{v} \in V$ and $\mathbf{w} \in W$.

Proof. Define the map

$$T \times S : \; V \times W \to V' \otimes W'$$

by $(T \times S)(\mathbf{v}, \mathbf{w}) := T(\mathbf{v}) \otimes S(\mathbf{w})$.

Then $T \times S$ is bilinear: for instance, we can show that

$$(T \times S)(a\mathbf{v}_1 + b\mathbf{v}_2, \mathbf{w}) = a(T \times S)(\mathbf{v}_1, \mathbf{w}) + b(T \times S)(\mathbf{v}_2, \mathbf{w})$$

Therefore, $T \times S \in \mathrm{Obj}$. By Theorem 11.8, there exists a unique *linear transformation*

$$T \otimes S \; : V \otimes W \to V' \otimes W'$$

with the properties as described in the proposition. $\qquad \square$

Notation Warning. We use the notation $T \otimes S$ for the linear transformation defined above. Does it possess the properties of tensor product? For instance, do we have the following equality:

$$(aT_1 + bT_2) \otimes S = a(T_1 \otimes S) + b(T_2 \otimes S)?$$

This will be proved in Proposition 11.16 below, which justifies our notation of $T \otimes S$.

Example 11.14.

Let $V = V' = \mathbb{F}^2$ and $W = W' = \mathbb{F}^3$. Define the linear transformations

$$T : V \to V, \qquad S : W \to W$$

by:

$$T(\mathbf{v}) := A\mathbf{v} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \mathbf{v}, \qquad S(\mathbf{w}) := B\mathbf{w} = \begin{pmatrix} p & q & r \\ s & t & u \\ v & w & x \end{pmatrix} \mathbf{w}.$$

How does $T \otimes S : V \otimes W \to V \otimes W$ look like? Suppose $\{e_1, \mathbf{e}_2\}, \{\mathbf{f}_1, \mathbf{f}_2, \mathbf{f}_3\}$ are the usual bases of V and W, respectively. By Theorem 11.10, a basis of $V \otimes W$ is given by:

$$\mathcal{C} = \{\mathbf{e}_1 \otimes \mathbf{f}_1, \mathbf{e}_1 \otimes \mathbf{f}_2, \mathbf{e}_1 \otimes \mathbf{f}_3, \mathbf{e}_2 \otimes \mathbf{f}_1, \mathbf{e}_2 \otimes \mathbf{f}_2, \mathbf{e}_2 \otimes \mathbf{f}_3\}.$$

Now compute $(T \otimes S)(\mathbf{e}_i \otimes \mathbf{f}_j)$ for $i = 1, 2$ and $j = 1, 2, 3$. For instance,

$$(T \otimes S)(\mathbf{e}_1 \otimes \mathbf{f}_1)$$
$$= T(\mathbf{e}_1) \otimes S(\mathbf{f}_1)$$
$$= (a\mathbf{e}_1 + c\mathbf{e}_2) \otimes (p\mathbf{f}_1 + s\mathbf{f}_2 + v\mathbf{f}_3)$$
$$= (ap)\,\mathbf{e}_1 \otimes \mathbf{f}_1 + (as)\,\mathbf{e}_1 \otimes \mathbf{f}_2 + (av)\,\mathbf{e}_1 \otimes \mathbf{f}_3 + (cp)\,\mathbf{e}_2 \otimes \mathbf{f}_1 + (cs)\,\mathbf{e}_2 \otimes \mathbf{f}_2 + (cv)\,\mathbf{e}_2 \otimes \mathbf{f}_3.$$

Therefore, the linear transformation $(T \otimes S)$ has a 6×6 matrix representation:

$$(T \otimes S)_{\mathcal{C},\mathcal{C}} = \begin{pmatrix} aB & bB \\ cB & dB \end{pmatrix},$$

Th above matrix is called the **Kronecker Tensor Product** of A and B (see the command kron in MATLAB for details).

Proposition 11.15. *More generally, given the linear operator $T : V \to V$ and $S : W \to W$, let $\mathcal{A} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}, \mathcal{B} = \{\mathbf{w}_1, \ldots, \mathbf{w}_m\}$ be a basis of V, W respectively, with*

$$(T)_{\mathcal{A},\mathcal{A}} = (a_{ij}) \quad := \quad A, \ (S)_{\mathcal{B},\mathcal{B}} = (b_{ij}) \quad := \quad B$$

As a result, $(T \otimes S)_{C,C} = A \otimes B$, where $C = \{\mathbf{v}_1 \otimes \mathbf{w}_1, \ldots, \mathbf{v}_n \otimes \mathbf{w}_m\}$, and $A \otimes B$ denotes the Kronecker tensor product, defined as the matrix

$$\begin{pmatrix} a_{1,1}B & \cdots & a_{1,n}B \\ \vdots & \ddots & \vdots \\ a_{n,1}B & \cdots & a_{n,n}B \end{pmatrix}.$$

Proof. Following the similar procedure as in Example 11.14 and applying the relation

$$(T \otimes S)(\mathbf{v}_i \otimes \mathbf{w}_j) = T(\mathbf{v}_i) \otimes S(\mathbf{w}_j) = \left(\sum_{k=1}^{n} a_{ki}\mathbf{v}_k\right) \otimes \left(\sum_{\ell=1}^{m} b_{\ell j}\mathbf{w}_\ell\right) = \sum_{k=1}^{n}\sum_{\ell=1}^{m} (a_{ki}b_{\ell j})\,\mathbf{v}_k \otimes \mathbf{w}_\ell$$

$\square$

Proposition 11.16. *The operation $T \otimes S$ satisfies all the properties of tensor product. For example,*

$$(aT_1 + bT_2) \otimes S = a(T_1 \otimes S) + b(T_2 \otimes S)$$

$$T \otimes (cS_1 + dS_2) = c(T \otimes S_1) + d(T \otimes S_2)$$

Therefore, the usage of the notion " $\otimes$ " is justified for the definition of $T \otimes S$.

Proof. using matrix multiplication. For instance, consider the operation $(T + T') \otimes S$, with

$(T)_{\mathcal{A},\mathcal{A}} = (a_{ij}), (T')_{\mathcal{A},\mathcal{A}} = (c_{ij}), (S)_{\mathcal{B},\mathcal{B}} = (b_{ij})$.

We compute its matrix representation directly:

$$\begin{aligned}
((T + T') \otimes S)_{C,C} &= (T + T')_{\mathcal{A},\mathcal{A}} \otimes (S)_{\mathcal{B},\mathcal{B}} \\
&= \left[(T)_{\mathcal{A},\mathcal{A}} + (T')_{\mathcal{A},\mathcal{A}}\right] \otimes (S)_{\mathcal{B},\mathcal{B}} \\
&= (T)_{\mathcal{A},\mathcal{A}} \otimes (S)_{\mathcal{B},\mathcal{B}} + (T')_{\mathcal{A},\mathcal{A}} \otimes (S)_{\mathcal{B},\mathcal{B}}
\end{aligned}$$

where the last equality is by the additive rule for Kronecker product for matrices. Therefore,

$$((T + T') \otimes S)_{C,C} = (T \otimes S)_{C,C} + (T' \otimes S)_{C,C} \Rightarrow (T + T') \otimes S = T \otimes S + T' \otimes S$$

$\square$

Alternatively, we study the linear transformations by applying them on a basis of $V \otimes W$. Another way of the proof is by computing $((T + T') \otimes S)(\mathbf{v}_i \otimes \mathbf{w}_j)$, where $\{\mathbf{v}_i \otimes \mathbf{w}_j \mid 1 \leq i \leq n, 1 \leq j \leq m\}$ forms a basis of $(T + T') \otimes S$:

$$
\begin{aligned}
((T + T') \otimes S)(\mathbf{v}_i \otimes \mathbf{w}_j) &= (T + T')(\mathbf{v}_i) \otimes S(\mathbf{w}_j) \\
&= (T(\mathbf{v}_i) + T'(\mathbf{v}_i)) \otimes S(\mathbf{w}_j) \\
&= T(\mathbf{v}_i) \otimes S(\mathbf{w}_j) + T'(\mathbf{v}_i) \otimes S(\mathbf{w}_j) \\
&= (T \otimes S)(\mathbf{v}_i \otimes \mathbf{w}_j) + (T' \otimes S)(\mathbf{v}_i \otimes \mathbf{w}_j)
\end{aligned}
$$

Since $((T + T') \otimes S)(\mathbf{v}_i \otimes \mathbf{w}_j)$ coincides with $(T \otimes S + T' \otimes S)(\mathbf{v}_i \otimes \mathbf{w}_j)$ for all basis vectors $\mathbf{v}_i \otimes \mathbf{w}_j \in C$, we imply $(T + T') \otimes S = T \otimes S + T' \otimes S$. $\square$

Proposition 11.17. *Let $A, C : V \to V$ be linear operators of V, and $B, D : W \to W$ be linear operators on W, then*
$$
(A \otimes B) \circ (C \otimes D) = (AC) \otimes (BD).
$$

Proposition 11.18. *Define linear operators $A : V \to V$ and $B : W \to W$ with $\dim(V), \dim(W) < \infty$. Then*
$$
\det(A \otimes B) = (\det(A))^{\dim(W)} (\det(B))^{\dim(V)}
$$

Proof. Homework question. $\square$

Corollary 11.19. *There exists a linear transformation*

$$
\Phi : \mathrm{Hom}(V, V) \otimes \mathrm{Hom}(W, W) \to \mathrm{Hom}(V \otimes W, V \otimes W)
$$

with $\Phi(A \otimes B) := A \otimes B$, where the input of Φ is a linear transformation $A : V \to V$ tensored with another linear transformation $B : W \to W$, and the output is the tensored linear transformation on $V \otimes W$ (c.f. Proposition 11.13).

 Moreover, if $\dim(V), \dim(W) < \infty$, then Φ is an isomorphism.

Proof. Construct the mapping

$$
\Phi : \mathrm{Hom}(V, V) \times \mathrm{Hom}(W, W) \to \mathrm{Hom}(V \otimes W, V \otimes W)
$$

with $\Phi(A, B) = A \otimes B$. Then Φ is indeed bilinear: for instance,

$$
\Phi(pA + qC, B) = (pA + qC) \otimes B = p(A \otimes B) + q(C \otimes B) = p\Phi(A, B) + q\Phi(C, B)
$$

So the first statement follows from the universal property of tensor product.

If $\dim(V), \dim(W) < \infty$, we imply

$$\begin{aligned}
\dim(\mathrm{Hom}\,(V,V) \otimes \mathrm{Hom}\,(W,W)) &= \dim(\mathrm{Hom}\,(V,V)) \cdot \dim(\mathrm{Hom}\,(W,W)) \\
&= [\dim(V)\dim(V)] \cdot [\dim(W)\dim(W)] \\
&= [\dim(V)\dim(W)]^2 \\
&= [\dim(V \otimes W)]^2 \\
&= \dim(\mathrm{Hom}\,(V \otimes W, V \otimes W))
\end{aligned}$$

Therefore, is Φ is an isomorphism. $\square$

As a result, every linear operator $\alpha : V \otimes W \to V \otimes W$ can be expressed as

$$\alpha = A_1 \otimes B_1 + \cdots + A_k \otimes B_k$$

where $A_i : V \to V$ and $B_j : W \to W$.

Chapter 12

Multilinear & Exterior Products

12.1 Multilinear Tensor Product

In this section, we will generalize the tensor product of more than 2 vector spaces:

Definition 12.1 (Tensor Product among multiple vector spaces)**.** Let $V_1, \ldots, V_p$ be vector spaces over $\mathbb{F}$. Let

$$S = \{(\mathbf{v}_1, \ldots, \mathbf{v}_p) \mid \mathbf{v}_i \in V_i\}$$

(we assume no relations among distinct elements in S), and define $\mathfrak{X} = \mathrm{span}(S)$. Then the tensor product space $V_1 \otimes \cdots \otimes V_p = \mathfrak{X}/\mathfrak{Y}$, where $\mathfrak{Y}$ is the vector subspace of $\mathfrak{X}$ spanned by vectors of the form

$$(\mathbf{v}_1, \ldots, \mathbf{v}_i + \mathbf{v}_i', \ldots, \mathbf{v}_p) - (\mathbf{v}_1, \ldots, \mathbf{v}_i, \ldots, \mathbf{v}_p) - (\mathbf{v}_1, \ldots, \mathbf{v}_i', \ldots, \mathbf{v}_p),$$

and

$$(\mathbf{v}_1, \ldots, \alpha\mathbf{v}_i, \ldots, \mathbf{v}_p) - \alpha(\mathbf{v}_1, \ldots, \mathbf{v}_i, \ldots, \mathbf{v}_p)$$

for $i = 1, 2, \ldots, p$.

Moreover, we define

$$\mathbf{v}_1 \otimes \cdots \otimes \mathbf{v}_p := [(\mathbf{v}_1, \ldots, \mathbf{v}_p) + \mathcal{Y}] \in V_1 \otimes \cdots \otimes V_p.$$

Remark 12.2. Similar to the tensor product of two vector spaces, we have:

$$\mathbf{v}_1 \otimes \cdots \otimes (\alpha\mathbf{v}_i + \beta\mathbf{v}_i') \otimes \cdots \otimes \mathbf{v}_p = \alpha(\mathbf{v}_1 \otimes \cdots \otimes \mathbf{v}_i \otimes \cdots \otimes \mathbf{v}_p) + \beta(\mathbf{v}_1 \otimes \cdots \otimes \mathbf{v}_i' \otimes \cdots \otimes \mathbf{v}_p).$$

Also, a general vector in $V_1 \otimes \cdots \otimes V_p$ is:

$$\sum_{i=1}^{n} \left(\mathbf{w}_1^{(i)} \otimes \cdots \otimes \mathbf{w}_p^{(i)} \right), \quad \text{where } \mathbf{w}_j^{(i)} \in V_j, j = 1, \ldots, p.$$

Finally, let $\mathcal{B}_i = \{\mathbf{v}_i^{(1)}, \ldots, \mathbf{v}_i^{(d_i)}\}$ be a basis of V_i, $i = 1, \ldots, p$, then

$$\mathcal{B} = \left\{ \mathbf{v}_1^{(\alpha_1)} \otimes \cdots \otimes \mathbf{v}_p^{(\alpha_p)} \mid 1 \leq \alpha_i \leq d_i \right\}$$

is a basis of $V_1 \otimes \cdots \otimes V_p$. As a result,

$$\dim(V_1 \otimes \cdots \otimes V_p) = d_1 \times \cdots \times d_p = \dim(V_1) \times \cdots \times \dim(V_p).$$

We now introduce the universal property of multilinear tensor product:

Theorem 12.3 (Universal Property of Multilinear Tensor). *Let*

$$\mathrm{Obj} = \{\phi : V_1 \times \cdots \times V_p \to W \mid \phi \text{ is a p-linear map}\},$$

i.e. for any $\phi \in \mathrm{Obj}$:

$$\phi(\mathbf{v}_1, \ldots, \alpha \mathbf{v}_i + \beta \mathbf{v}_i', \ldots, \mathbf{v}_p) = \alpha \phi(\mathbf{v}_1, \ldots, \mathbf{v}_i, \ldots, \mathbf{v}_p) + \beta \phi(\mathbf{v}_1, \ldots, \mathbf{v}_i', \ldots, \mathbf{v}_p),$$

for all $\mathbf{v}_i, \mathbf{v}_i' \in \mathbf{v}_i$, $i = 1, \ldots, p$, $\forall \alpha, \beta \in \mathbb{F}$. Then

$$i : V_1 \times \cdots \times V_p \to V_1 \otimes \cdots \otimes V_p$$

defined by

$$(\mathbf{v}_1, \ldots, \mathbf{v}_p) \mapsto \mathbf{v}_1 \otimes \cdots \otimes \mathbf{v}_p,$$

is in Obj. *Moreover, $(i : V_1 \times \cdots \times V_p \to V_1 \otimes \cdots \otimes V_p) \in$ Obj satisfies the* universal property *of multilinear tensor product, that is, for all $(\phi : V_1 \times \cdots \times V_p \to W) \in$ Obj, there exists a unique linear transformation*

$$\bar{\phi} : V_1 \otimes \cdots \otimes V_p \to W$$

such that the diagram commutes:

In other words, $\phi = \bar{\phi} \circ i$.

As an application of the universal property of multilinear tensor product, we can construct linear transformation of multilinear tensor product spaces:

Corollary 12.4. *Let $T_i : V_i \to V_i'$ be a linear transformation for $1 \leq i \leq p$. Then there is a unique linear transformation*

$$(T_1 \otimes \cdots \otimes T_p) : V_1 \otimes \cdots \otimes V_p \to V_1' \otimes \cdots \otimes V_p'$$

such that

$$(T_1 \otimes \cdots \otimes T_p)(\mathbf{v}_1 \otimes \cdots \otimes \mathbf{v}_p) = T_1(\mathbf{v}_1) \otimes \cdots \otimes T_p(\mathbf{v}_p).$$

Proof. Construct the mapping

$$\phi : V_1 \times \cdots \times V_p \to V_1' \otimes \cdots \otimes V_p'$$

defined by

$$(\mathbf{v}_1, \ldots, \mathbf{v}_p) \mapsto T_1(\mathbf{v}_1) \otimes \cdots \otimes T_p(\mathbf{v}_p),$$

which is indeed p-linear. By the universal property, we obtain the unique linear transformation

$$\bar{\phi} : V_1 \otimes \cdots \otimes V_p \to V_1' \otimes \cdots \otimes V_p'$$

satisfying our desired properties. $\qquad\square$

12.2   Exterior Products

From now on, we only consider $V_1 = \cdots = V_p = V$. Then for any linear transformation $T : V \to W$, we define

$$T^{\otimes p} : V \otimes \cdots \otimes V \to W \otimes \cdots \otimes W.$$

We use the shorthand notation $V^{\otimes p}$ to denote

$$\underbrace{V \otimes \cdots \otimes V}_{p \text{ terms in total}}.$$

Definition 12.5 (Alternating Map). A p-linear map $\phi : V \times \cdots \times V \to W$ is called **alternating** if

$$\phi(\mathbf{v}_1, \ldots, \mathbf{v}_i, \ldots, \mathbf{v}_j, \ldots, \mathbf{v}_p) = 0_W, \quad \text{whenever } \mathbf{v}_i = \mathbf{v}_j \text{ for } i \neq j.$$

We also say ϕ is *p-alternating.*

Example 12.6. The cross product map is an alternating bilinear map:

$$\phi : \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}^3, \quad (\mathbf{v}, \mathbf{w}) \mapsto \mathbf{v} \times \mathbf{w}.$$

It satisfies:

- ϕ is bilinear,

- $\phi(\mathbf{v}, \mathbf{v}) = \mathbf{v} \times \mathbf{v} = \mathbf{0}$.

Example 12.7. The determinant map is also alternating:

$$\phi : \underbrace{\mathbb{F}^n \times \cdots \times \mathbb{F}^n}_{n \text{ terms in total}} \to \mathbb{F}, \quad (\mathbf{v}_1, \ldots, \mathbf{v}_n) \mapsto \det([\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n])$$

It satisfies:

- ϕ is n-linear (MAT 2040 knowledge),

- ϕ is alternating (again, from MAT 2040 since the determinant of a matrix with repeated columns is zero).

Theorem 12.8 (Universal Property for Exterior Power)**.** *Let*

$$\mathrm{Obj} := \{\phi : V \times \cdots \times V \to W \mid \phi \text{ is } p\text{-alternating}\}.$$
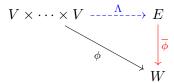
Then there exists a universal object

$$(\Lambda : V \times \cdots \times V \longrightarrow E) \in \mathrm{Obj}$$

such that for every $(\phi : V \times \cdots \times V \to W) \in \mathrm{Obj}$, there exists a unique linear map

$$\overline{\phi} : E \longrightarrow W$$

making the following diagram commute:

$$
\begin{array}{ccc}
V \times \cdots \times V & \xdashrightarrow{\ \ \Lambda\ \ } & E \\
& \searrow{\scriptstyle\phi} & \downarrow{\scriptstyle\overline{\phi}} \\
& & W
\end{array}
$$

In other words, $\phi = \overline{\phi} \circ \Lambda$.

We now construct the universal object $(\Lambda : V \times \cdots \times V \to E) \in \mathrm{Obj}$:

Definition 12.9 (Wedge Products)**.** Let V be a vector space. Consider the subspace $U \leq V^{\otimes p}$ spanned by the vectors

$$U = \mathrm{span}\{\cdots \otimes \mathbf{v} \otimes \cdots \otimes \mathbf{v} \otimes \cdots \mid \mathbf{v} \in V\}$$

with a repeated entry. Then the p-**exterior power** is defined by the quotient space:

$$\wedge^p V := V^{\otimes p}/U,$$

Similar to the case of $V^{\otimes p}$, we define the vector

$$\mathbf{v}_1 \wedge \cdots \wedge \mathbf{v}_p := \mathbf{v}_1 \otimes \cdots \otimes \mathbf{v}_p + U \in \wedge^p V.$$

We claim that the mapping

$$\Lambda : V \times \cdots \times V \to \wedge^p V,$$

defined by

$$(\mathbf{v}_1, \ldots, \mathbf{v}_p) \mapsto \mathbf{v}_1 \wedge \cdots \wedge \mathbf{v}_p$$

is in Obj, and it satisfies the universal property of exterior power (c.f. Theorem 12.8).

Proposition 12.10. *The p-exterior power $\wedge^p V$ is p-linear:*

$$\mathbf{v}_1 \wedge \cdots \wedge (a\mathbf{v}_i + b\mathbf{v}_i') \wedge \cdots \wedge \mathbf{v}_p = a\left(\mathbf{v}_1 \wedge \cdots \wedge \mathbf{v}_i \wedge \cdots \wedge \mathbf{v}_p\right) + b\left(\mathbf{v}_1 \wedge \cdots \wedge \mathbf{v}_i' \wedge \cdots \wedge \mathbf{v}_p\right)$$

and alternating:

$$\mathbf{v}_1 \wedge \cdots \wedge \mathbf{v} \wedge \cdots \wedge \mathbf{v} \wedge \cdots \wedge \mathbf{v}_p = \mathbf{0}_{\wedge^p V}$$

Consequently, $\Lambda : V \times \cdots \times V \to \wedge^p V$ is in Obj.

Proof. The proof is trivial - as an example, for the second statement:

$$\mathbf{v}_1 \wedge \cdots \wedge \mathbf{v} \wedge \cdots \wedge \mathbf{v} \wedge \cdots \wedge \mathbf{v}_p := \mathbf{v}_1 \otimes \cdots \otimes \mathbf{v} \otimes \cdots \otimes \mathbf{v} \otimes \cdots \otimes \mathbf{v}_p + U.$$

Since $\mathbf{v}_1 \otimes \cdots \otimes \mathbf{v} \otimes \cdots \otimes \mathbf{v} \in U$ by definition, it is equal to $\mathbf{0} + U = \mathbf{0}_{\wedge^p V}$. □

We leave it to the readers to check that $\Lambda : V \times \cdots \times V \to \wedge^p V$ satisfies the universal property for exterior power (Theorem 12.8).

Remark 12.11. The wedge product satisfies the sign reversal property:

$$\mathbf{v}_1 \wedge \cdots \wedge \mathbf{v} \wedge \cdots \wedge \mathbf{w} \wedge \cdots \wedge \mathbf{v}_p = -\mathbf{v}_1 \wedge \cdots \wedge \mathbf{w} \wedge \cdots \wedge \mathbf{v} \wedge \cdots \wedge \mathbf{v}_p. \tag{12.1}$$

For instance, one has:

$$(\mathbf{v} + \mathbf{w}) \wedge (\mathbf{v} + \mathbf{w}) = \mathbf{0} \;\Rightarrow\; \mathbf{v} \wedge \mathbf{v} + \mathbf{v} \wedge \mathbf{w} + \mathbf{w} \wedge \mathbf{v} + \mathbf{w} \wedge \mathbf{w} = \mathbf{0} \;\Rightarrow\; \mathbf{v} \wedge \mathbf{w} + \mathbf{w} \wedge \mathbf{v} = \mathbf{0}.$$

One can easily generalize the above proof to obtain Equation 12.1.

Proposition 12.12. *If $\dim(V) = n$ and $0 \le p \le n$, then*

$$\dim(\wedge^p V) = \binom{n}{p}.$$

Proof. Let $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ be a basis of V, then $\{\mathbf{v}_{i_1} \otimes \cdots \otimes \mathbf{v}_{i_p} \mid 1 \le i_k \le n\}$ forms a basis of $V^{\otimes p}$. Since $\pi_{\wedge^p V} : V^{\otimes p} \to V^{\otimes p}/U = \wedge^p V$ is surjective, the image

$$\left\{\pi_{\wedge^p V}\left(\mathbf{v}_{i_1} \otimes \cdots \otimes \mathbf{v}_{i_p}\right) = \mathbf{v}_{i_1} \wedge \cdots \wedge \mathbf{v}_{i_p} \mid 1 \le i_k \le n\right\}$$

spans $\wedge^p V$.

By Proposition 12.10 and Remark 12.11, many of the terms in the above expression are zero or repeated. For instance, using Remark 12.11, one can reorder the expression $\mathbf{v}_{i_1} \wedge \cdots \wedge \mathbf{v}_{i_p}$, so that (up to $\pm$) the indices $i_1 < \cdots < i_p$ are in ascending order. After removing the repeated entries, one

conclude that

$$\mathcal{B} := \{\mathbf{v}_{i_1} \wedge \cdots \wedge \mathbf{v}_{i_p} \mid 1 \leq i_1 < \cdots < i_k \leq n\}$$

spans $\wedge^p V$.

We omit the proof that $\mathcal{B}$ is linearly independent - its proof is similar to that of Theorem 11.10. Consequently, $\mathcal{B}$ is a basis of $\wedge^p V$ which has cardinality equal to $\binom{n}{p}$. $\qquad\square$

Let $T : V \to V$ be a linear operator. By applying the universal property for exterior power (Theorem 12.8), one can obtain as in Corollary 12.4:

Corollary 12.13. *Let V be a vector space, and $T : V \to V$ be a linear operator. For all $0 \leq p \leq \dim(V)$, there exists a unique linear operator*

$$T^{\wedge^p} : \wedge^p V \to \wedge^p V$$

defined by:

$$\mathbf{v}_1 \wedge \cdots \wedge \mathbf{v}_p \mapsto T(\mathbf{v}_1) \wedge \cdots \wedge T(\mathbf{v}_p).$$

12.3   The Determinant

To end this chapter, we define the determinant for all linear operators $T : V \to V$ on finite-dimensional vector space V directly. As a consequence, we will prove that

$$\det(T) = \det(T_{\mathcal{B},\mathcal{B}}) \text{ for any basis } \mathcal{B} \text{ of } T.$$

In other words, if $T(\mathbf{x}) = A\mathbf{x}$ is a matrix transformation, then one has $\det(T) = \det(A)$.

Moreover, we will have a natural way of proving $\det(T \circ S) = \det(T)\det(S)$, which in term will imply the well-known formula of $\det(AB) = \det(A)\det(B)$ in MAT2040.

In short, our definition will give generalize the notion of determinant of matrices to linear operators, and give a natural explanation of why determinant is multiplicative.

Definition 12.14 (Determinant for Linear Operators)**.** Let $\dim(V) = n$. Then $\dim(\wedge^n V) = \binom{n}{n} = 1$. By Corollary 12.13 we have a linear operator

$$T^{\wedge n} : \wedge^n V \to \wedge^n V.$$

on the 1-dimensional vector space $\wedge^n V$.

Therefore, for all $\tau \in \wedge^n V$, there exists $\alpha_T \in \mathbb{F}$ such that

$$T^{\wedge n}(\tau) = \alpha_T \tau,$$

and we define the *determinant* of T by: $\det(T) := \alpha_T$, so that $T^{\wedge n}(\tau) = \det(T)\tau$ Note that this definition does not depend on any choice of basis of V.

To compute the determinant, one can take **any** basis $\mathbf{v}_1, \ldots, \mathbf{v}_n$ of V, and

$$\wedge^n V = \mathrm{span}(\mathbf{v}_1 \wedge \cdots \wedge \mathbf{v}_n).$$

Then one computes

$$T^{\wedge n}(\mathbf{v}_1 \wedge \cdots \wedge \mathbf{v}_n) := T(\mathbf{v}_1) \wedge \cdots \wedge T(\mathbf{v}_n) = \det(T)\mathbf{v}_1 \wedge \cdots \wedge \mathbf{v}_n.$$

Example 12.15.   1. Suppose that $T = I : V \to V$ be the identity operator. Then

$$I^{\wedge n}(\mathbf{v}_1 \wedge \cdots \wedge \mathbf{v}_n) = I(\mathbf{v}_1) \wedge \cdots \wedge I(\mathbf{v}_n) = \mathbf{v}_1 \wedge \cdots \wedge \mathbf{v}_n = 1 \cdot (\mathbf{v}_1 \wedge \cdots \wedge \mathbf{v}_n).$$

Therefore, $\det(I) = 1$.

2. Suppose that $T : V \to V$ is diagonalizable with $\{\mathbf{w}_1, \ldots, \mathbf{w}_n\}$ forming an eigen-basis of T. As a result,

$$T^{\wedge n}(\mathbf{w}_1 \wedge \cdots \wedge \mathbf{w}_n) = T(\mathbf{w}_1) \wedge T(\mathbf{w}_2) \cdots \wedge T(\mathbf{w}_n) = (\lambda_1 \mathbf{w}_1) \wedge \cdots \wedge (\lambda_n \mathbf{w}_n) = (\lambda_1 \cdots \lambda_n)\mathbf{w}_1 \wedge \cdots \wedge \mathbf{w}_n$$

Therefore, $\det(T) = \lambda_1 \cdots \lambda_n$.

Proposition 12.16. *Let $T, S : V \to V$ be linear transformations, then*

$$(T \circ S)^{\wedge^p} : \wedge^p V \to \wedge^p V$$

with $T^{\wedge^p}, S^{\wedge^p} : \wedge^p V \to \wedge^p V$ satisfies

$$(T \circ S)^{\wedge^p} = (T^{\wedge^p}) \circ (S^{\wedge^p})$$

Proof. Pick any basis $\{\mathbf{v}_{i_1} \wedge \cdots \wedge \mathbf{v}_{i_p} \mid 1 \le i_1 < \cdots < i_p \le n\}$ of $\wedge^p V$. Then

$$(T \circ S)^{\wedge^p}(\mathbf{v}_{i_1} \wedge \cdots \wedge \mathbf{v}_{i_p}) = (T \circ S)(\mathbf{v}_{i_1}) \wedge \cdots \wedge (T \circ S)(\mathbf{v}_{i_p})$$

On the other hand,

$$(T^{\wedge^p}) \circ (S^{\wedge^p})(\mathbf{v}_{i_1} \wedge \cdots \wedge \mathbf{v}_{i_p}) = T^{\wedge^p}(S(\mathbf{v}_{i_1}) \wedge \cdots \wedge S(\mathbf{v}_{i_p})) = (T \circ S)(\mathbf{v}_{i_1}) \wedge \cdots \wedge (T \circ S)(\mathbf{v}_{i_p})$$

$$\square$$

$$T : \mathbb{R}^2 \to \mathbb{R}^2 \quad \text{with} \quad T\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}.$$

Example 12.17. Let $T : \mathbb{F}^2 \to \mathbb{F}^2$ be the matrix transformation given by $T(\mathbf{x}) = A\mathbf{x}$, where

$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. We take the usual basis $\{\mathbf{e}_1, \mathbf{e}_2\}$ to compute the determinant of T:

$$\det(T) \cdot (\mathbf{e}_1 \wedge \mathbf{e}_2) = T(\mathbf{e}_1) \wedge T(\mathbf{e}_2)$$
$$= \begin{pmatrix} a \\ c \end{pmatrix} \wedge \begin{pmatrix} b \\ d \end{pmatrix}$$
$$= (a\mathbf{e}_1 + c\mathbf{e}_2) \wedge (b\mathbf{e}_1 + d\mathbf{e}_2)$$
$$= (ab)\mathbf{e}_1 \wedge \mathbf{e}_1 + (ad)\mathbf{e}_1 \wedge \mathbf{e}_2 + (cb)\mathbf{e}_2 \wedge \mathbf{e}_1 + (cd)\mathbf{e}_2 \wedge \mathbf{e}_2$$
$$= (ad)\mathbf{e}_1 \wedge \mathbf{e}_2 + (cb)\mathbf{e}_2 \wedge \mathbf{e}_1$$
$$= (ad - bc)\mathbf{e}_1 \wedge \mathbf{e}_2.$$

where we have used $\mathbf{e}_1 \wedge \mathbf{e}_1 = \mathbf{e}_2 \wedge \mathbf{e}_2 = \mathbf{0}$ and $\mathbf{e}_1 \wedge \mathbf{e}_2 = -(\mathbf{e}_2 \wedge \mathbf{e}_1)$ by Proposition 12.10 and Remark 12.11. Therefore, we conclude:

$$\det(T) = \Delta(A) = ad - bc.$$

(to avoid confusion, we denote by $\Delta(A)$ the determinant of A as defined in MAT2040.)

For general $n \times n$-matrices A, one also has:

Theorem 12.18. *Let $V = \mathbb{F}^n$, and let*

$$T : V \to V, \quad \text{with } T(\mathbf{v}) = A\mathbf{v}, \quad A \in M_{n \times n}(\mathbb{F}).$$

Then $\det(T) = \Delta(A)$.

Proof. Let $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ be the usual basis of $V \cong \mathbb{F}^n$, and $\{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$ be the columns of A. Then:

$$\det(T) \cdot \mathbf{e}_1 \wedge \cdots \wedge \mathbf{e}_n = T^{\wedge n}(\mathbf{e}_1 \wedge \cdots \wedge \mathbf{e}_n)$$
$$= T(\mathbf{e}_1) \wedge \cdots \wedge T(\mathbf{e}_n) \tag{12.2}$$
$$= \mathbf{a}_1 \wedge \cdots \wedge \mathbf{a}_n.$$

We wish to show $\Delta(T) = \det(A)$. Indeed, by MAT2040, $\Delta(A)$ is uniquely determined by the following:

- $\Delta(I_{n \times n}) = 1$;
- $\Delta(\cdots |\alpha \mathbf{a}_i + \beta \mathbf{a}_i'| \cdots) = \alpha \Delta(\cdots |\mathbf{a}_i| \cdots) + \beta \Delta(\cdots |\mathbf{a}_i'| \cdots).$
- $\Delta(\cdots |\mathbf{v}| \cdots |\mathbf{v}| \cdots) = 0$

By the multilinearity and alternating property of the wedge product, one can easily see that $\det(T)$ satisfies all the above properties. For instance, for the last point, one can apply Equation 12.2 and get:

$$\det(T)\mathbf{e}_1 \wedge \cdots \wedge \mathbf{e}_n = \cdots \wedge \mathbf{v} \wedge \cdots \wedge \mathbf{v} \wedge \cdots = \mathbf{0} = 0(\mathbf{e}_1 \wedge \cdots \wedge \mathbf{e}_n).$$

Therefore, $\det(T) = 0$ as stated in the third point. And the result follows. $\qquad\square$

So we have just proved that our definition of det matches with that of the usual definition of determinant of matrices. Moreover, we have:

Corollary 12.19.
$$\det(T \circ S) = \det(T)\det(S)$$

Proof. Let $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ be a basis of V, so that $\mathbf{v}_1 \wedge \cdots \wedge \mathbf{v}_n$ is a basis of $\wedge^n V$. Then:

$$
\begin{aligned}
\det(T \circ S) \cdot \mathbf{v}_1 \wedge \cdots \wedge \mathbf{v}_n &= (T \circ S)^{\wedge n}(\mathbf{v}_1 \wedge \cdots \wedge \mathbf{v}_n) \\
&= T^{\wedge n}\left(S^{\wedge n}(\mathbf{v}_1 \wedge \cdots \wedge \mathbf{v}_n)\right) \\
&= T^{\wedge n}\left(\det(S) \cdot \mathbf{v}_1 \wedge \cdots \wedge \mathbf{v}_n\right) \\
&= \det(S) \cdot T^{\wedge n}(\mathbf{v}_1 \wedge \cdots \wedge \mathbf{v}_n) \\
&= \det(S) \cdot \det(T) \cdot \mathbf{v}_1 \wedge \cdots \wedge \mathbf{v}_n.
\end{aligned}
$$

Since $\mathbf{v}_1 \wedge \cdots \wedge \mathbf{v}_n \neq 0$, it follows that

$$\det(T \circ S) = \det(T)\det(S).$$

$\qquad\square$