

# Contents

<b>1</b>	<b>Vector Spaces</b>	<b>2</b>
1.1	Vector Space . . . . .	2
1.2	Isomorphism . . . . .	7
1.3	Basis . . . . .	9
<b>2</b>	<b>Linear Transformations</b>	<b>12</b>
2.1	Linear Transformations . . . . .	12
2.2	Duality . . . . .	15
2.3	Tensor Product . . . . .	21
2.4	Matrix . . . . .	23
2.5	Change of Basis . . . . .	27
<b>3</b>	<b>Operators</b>	<b>30</b>
3.1	Eigenspaces . . . . .	30
3.2	Diagonalisation . . . . .	33
3.3	Jordan Matrix . . . . .	36
3.4	Operators in Euclidean Spaces . . . . .	39
<b>4</b>	<b>Inner Product Spaces</b>	<b>42</b>
4.1	Bilinear Forms . . . . .	42
4.2	Geometry in Vector Spaces . . . . .	46
4.3	Orthogonalisation . . . . .	49
4.4	Riesz Representation . . . . .	52
4.5	Sesquilinear Forms . . . . .	54
<b>5</b>	<b>Determinant</b>	<b>60</b>
5.1	$m$ -Forms . . . . .	60
5.2	Determinant . . . . .	63
5.3	Volume . . . . .	65

# Vector Spaces

## 1.1 Vector Space

In elementary mathematics, we often refer to a vector as an ordered tuple of numbers with a direction and a magnitude. However, there is a much more abstract aspect to the notion of vectors. In fact, let us first generalise the notion of *scalars*, which are taken as complex constants in an elementary level.

In general, we have the following algebraic structure:

### Definition 1.1.1 ► Field

A **field** is a set  $\mathbb{F}$  with two binary operations  $\mathbb{F}^2 \rightarrow \mathbb{F}$ , namely addition and multiplication, such that

1.  $u + v = v + u$  for all  $u, v \in \mathbb{F}$ ;
2.  $(u + v) + w = u + (v + w)$  for all  $u, v, w \in \mathbb{F}$ ;
3.  $uv = vu$  for all  $u, v \in \mathbb{F}$ ;
4.  $(uv)w = u(vw)$  for all  $u, v, w \in \mathbb{F}$ ;
5.  $u(v + w) = uv + uw$  for all  $u, v, w \in \mathbb{F}$ ;
6. there exists  $0 \in \mathbb{F}$  such that  $u + 0 = u$  for all  $u \in \mathbb{F}$ ;
7. there exists  $1 \in \mathbb{F}$  such that  $1u = u$  for all  $u \in \mathbb{F}$ ;
8. for every  $u \in \mathbb{F}$ , there exists some  $v \in \mathbb{F}$  such that  $u + v = 0$ ;
9. for every  $u \in \mathbb{F}$ , there exists some  $v \in \mathbb{F}$  such that  $uv = 1$ .

One may check that both  $\mathbb{R}$  and  $\mathbb{C}$  are fields.

Now, in an elementary level, we say that a *vector* is “a quantity which has both a magnitude and a direction”. However, **that is not a real definition!** In the first place, we have not defined “magnitude” and “direction” yet, and it is also extremely unclear what the term “quantity” is supposed to mean.

Let us take the usual approach: instead of saying **what they are**, we shall focus on **what they do**. It turns out that we can also generalise the concept of vectors as any objects which possess properties similar to that of Euclidean vectors, i.e., we can view a vector as a mathematical quantity which can be added up and multiplied by another quantity called a scalar with some axioms which they follow. Rigorously, we define the notion of a *vector space*.

**Definition 1.1.2 ► Vector Space**

A **vector space** is a set  $V$  over a field  $\mathbb{F}$  with two binary operations, namely

- addition  $+: V^2 \rightarrow V$ , and
- scalar multiplication  $(\cdot)(\cdot): \mathbb{F} \times V \rightarrow V$ ,

such that

1.  $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$  for all  $\mathbf{u}, \mathbf{v} \in V$ ;
2.  $(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$  for all  $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$ ;
3.  $a\mathbf{b}\mathbf{v} = a(\mathbf{b}\mathbf{v})$  for all  $a, b \in \mathbb{F}$  and  $\mathbf{v} \in V$ ;
4. there exists an **additive identity** or **zero vector**  $\mathbf{0} \in V$  such that  $\mathbf{v} + \mathbf{0} = \mathbf{v}$  for all  $\mathbf{v} \in V$ ;
5. every  $\mathbf{v} \in V$  has an **additive inverse**  $\mathbf{w} \in V$  with  $\mathbf{v} + \mathbf{w} = \mathbf{0}$ ;
6. there exists a **multiplicative identity**  $1 \in \mathbb{F}$  such that  $1\mathbf{v} = \mathbf{v}$  for all  $\mathbf{v} \in V$ ;
7.  $a(\mathbf{u} + \mathbf{v}) = a\mathbf{u} + a\mathbf{v}$  and  $(a + b)\mathbf{u} = a\mathbf{u} + b\mathbf{u}$  for all  $a, b \in \mathbb{F}$  and  $\mathbf{u}, \mathbf{v} \in V$ .

Notice that here, the definitions of addition in scalar multiplication in a vector space imply that any vector space must be **closed** under these two operations. Notice also that the operations “addition” and “scalar multiplication” are not necessarily the addition and scalar multiplication which we are used to in  $\mathbb{R}^n$ , but abstract mappings which satisfy the given axioms.

Now, it is very easy to state what a vector really is: **any element in a certain vector space is a vector!** In fact, one can easily check that any field  $\mathbb{F}$  is a vector space over itself, so every scalar is a vector (if this sentence is confusing to you, read the previous paragraphs again).

Furthermore,  $\mathbb{R}$  with addition defined as real number addition and scalar multiplication defined as real number multiplication is a vector space over  $\mathbb{R}$  itself, but not the other way round, i.e., if I define addition as real number multiplication and scalar multiplication as real number addition, then  $\mathbb{R}$  is no longer a vector space over itself (check which axiom it violates).

We shall prove a few basic properties regarding vector spaces.

**Theorem 1.1.3 ► Uniqueness of Additive Identity**

*Let  $V$  be a vector space with  $\mathbf{0} \in V$  as an additive identity, then  $\mathbf{0}$  is unique.*

*Proof.* Suppose on contrary that there exists  $\mathbf{u} \in V$  such that  $\mathbf{v} + \mathbf{u} = \mathbf{v}$  for all  $\mathbf{v} \in V$ . Since  $\mathbf{0} \in V$ , we have

$$\mathbf{0} + \mathbf{u} = \mathbf{0}.$$

However,  $\mathbf{0}$  is the additive identity, so

$$\mathbf{u} = \mathbf{u} + \mathbf{0} = \mathbf{0} + \mathbf{u} = \mathbf{0},$$

i.e.  $\mathbf{0}$  is unique. □

Similarly, we can also prove the uniqueness of additive inverse.

#### Theorem 1.1.4 ► Uniqueness of Additive Inverse

*Let  $V$  be a vector space, then every  $\mathbf{v} \in V$  has a unique additive inverse.*

*Proof.* Suppose on contrary that there exist  $\mathbf{u}, \mathbf{w} \in V$  both being additive inverse of  $\mathbf{v}$ , then  $\mathbf{u} + \mathbf{v} = \mathbf{0}$  and  $\mathbf{w} + \mathbf{v} = \mathbf{0}$ . Therefore,

$$\mathbf{u} = (\mathbf{u} + \mathbf{v}) + \mathbf{u} = (\mathbf{w} + \mathbf{v}) + \mathbf{u} = \mathbf{w} + (\mathbf{u} + \mathbf{v}) = \mathbf{w},$$

i.e.,  $\mathbf{v}$  has a unique additive inverse. □

Theorem 1.1.4 justifies the notation  $-\mathbf{u}$  to denote the additive inverse of  $\mathbf{u}$ . However, so far we have not ascertained the fact that  $-\mathbf{u} = (-1)\mathbf{u}$  (note that the former means the inverse of  $\mathbf{u}$  while the latter means  $\mathbf{u}$  multiplied by the scalar  $-1$ )! While seemingly innocent, this result is not as easily proven as it looks.

First, we shall justify that  $0\mathbf{u} = \mathbf{0}$  for all  $\mathbf{u} \in V$ . Notice that

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

Adding  $-(0\mathbf{u})$  to both sides of the equation yields  $0\mathbf{u} = \mathbf{0}$  as desired. From this result we see that

$$(-1)\mathbf{u} + \mathbf{u} = (-1 + 1)\mathbf{u} = 0\mathbf{u} = \mathbf{0}.$$

By uniqueness of additive inverse, we must have  $(-1)\mathbf{u} = -\mathbf{u}$ .

Note that by using a similar technique we can prove that  $a\mathbf{0} = \mathbf{0}$  for all  $a \in \mathbb{F}$ , and so  $\mathbf{0} = -\mathbf{0}$  as a consequence.

Additionally, note that subtraction is defined as  $\mathbf{u} - \mathbf{v} = \mathbf{u} + (-1)\mathbf{v}$ , so the above result allows us to write  $\mathbf{u} - \mathbf{v} = \mathbf{u} + (-\mathbf{v})$ .

Note that a vector space is extended based on a set of vectors, so we can define *subspaces* similarly to the notion of subsets.

**Definition 1.1.5 ▶ Subspace**

Let  $V$  be a vector space.  $U \subseteq V$  is called a **subspace** if  $U$  is a vector space under addition and scalar multiplication in  $V$ .

It is easy to see that the intersection of any number of subspaces of a vector space  $V$  is still a subspace of  $V$ , but the union might not be so. Therefore, when we try to “merge” different subspaces, we usually use the notion of a *sum* instead of a union.

**Definition 1.1.6 ▶ Sum of Subspaces**

Let  $V_1, V_2$  be subspaces of a vector space  $V$ , then their **sum** is defined as

$$V_1 + V_2 := \{\mathbf{v}_1 + \mathbf{v}_2 : \mathbf{v}_1 \in V_1, \mathbf{v}_2 \in V_2\}.$$

We would like to verify that the space in Definition 1.1.6 is indeed still a subspace of  $V$ . First, the following proposition will be handy:

**Proposition 1.1.7 ▶ Subspaces Share a Unique Zero Vector**

Let  $U$  be any subspace of a vector space  $V$ . If  $\mathbf{0}_U$  and  $\mathbf{0}_V$  are the zero vectors in  $U$  and  $V$  respectively, then  $\mathbf{0}_U = \mathbf{0}_V$ .

*Proof.* Since  $U \subseteq V$ , we have  $\mathbf{0}_U \in V$ . Therefore,  $\mathbf{0}_U + \mathbf{0}_V = \mathbf{0}_U = \mathbf{0}_U + \mathbf{0}_U$ , which implies that  $\mathbf{0}_U = \mathbf{0}_V$ . □

Now, it is legal to state  $\mathbf{0} \in V$  as “the” zero vector of  $V_1, V_2$  and  $V$ . So clearly  $V_1 + V_2 \neq \emptyset$  because  $\mathbf{0} \in V_1 + V_2$ . Take any  $\mathbf{w}_1 = \mathbf{v}_1 + \mathbf{u}_1$  and  $\mathbf{w}_2 = \mathbf{v}_2 + \mathbf{u}_2$  in  $V_1 + V_2$ . Without loss of generality, let  $\mathbf{v}_1, \mathbf{v}_2 \in V_1$  and  $\mathbf{u}_1, \mathbf{u}_2 \in V_2$ . Then,

$$\mathbf{w}_1 + \alpha \mathbf{w}_2 = (\mathbf{v}_1 + \alpha \mathbf{v}_2) + (\mathbf{u}_1 + \alpha \mathbf{u}_2) \in V_1 + V_2,$$

which implies that  $V_1 + V_2$  is a subspace of  $V$ . In fact, not only is  $V_1 + V_2$  a subspace of  $V$ , it is the minimal subspace containing both  $V_1$  and  $V_2$ !

**Proposition 1.1.8 ▶ Sum as the Minimal Containing Subspace**

Let  $V_1, V_2, \dots, V_k$  be subspaces of  $V$ , then  $\sum_{i=1}^k V_i$  is the minimal subspace of  $V$  containing all of  $V_1, V_2, \dots, V_k$ .

*Proof.* Let  $U$  be the minimal subspace of  $V$  containing all of  $V_1, V_2, \dots, V_k$ . Clearly,

$V_i \subseteq U$  for all  $i = 1, 2, \dots, k$ . Take any  $\mathbf{w} \in \sum_{i=1}^k V_i$ , then we can write

$$\mathbf{w} = \sum_{i=1}^k \mathbf{v}_i$$

where  $\mathbf{v}_i \in V_i$ . Clearly,  $\mathbf{v}_i \in U$  for every  $i = 1, 2, \dots, k$ , so  $\mathbf{w} \in U$ , which means that

$$\sum_{i=1}^k V_i \subseteq U.$$

However, it is clear that  $V_i \subseteq \sum_{i=1}^k V_i$  for all  $i = 1, 2, \dots, k$ , so by the minimality of  $U$ , we have  $\sum_{i=1}^k V_i = U$ .  $\square$

In particular, we would like to consider a special construct known as *direct sum*.

#### Definition 1.1.9 ► Direct Sum

Let  $V$  be a vector space and  $U_1, U_2$  be subspaces of  $V$ . The sum  $U_1 + U_2$  is said to be a **direct sum**, denoted as  $U_1 \oplus U_2$ , if for every  $\mathbf{u} \in U_1 + U_2$  there is a unique way to express it as  $\mathbf{u} = \mathbf{u}_1 + \mathbf{u}_2$  for  $\mathbf{u}_1 \in U_1$  and  $\mathbf{u}_2 \in U_2$ .

In some sense, a direct sum of  $V$  can be viewed as a “partition” of  $V$  into two subsets with a minimal overlap. Note that unlike partition in its real definition, the subspaces  $U_1$  and  $U_2$  here cannot be disjoint sets as both of them have to contain the zero vector in  $V$ . In fact, we can prove that if  $U_1 \oplus U_2$  exists, then  $\mathbf{0}$  is the only thing in the intersection  $U_1 \cap U_2$ .

#### Theorem 1.1.10 ► Equivalent Definition of Direct Sums

Let  $V$  be a vector space and  $U_1, U_2$  be subspaces of  $V$ , then  $U_1 + U_2$  is a direct sum if and only if  $U_1 \cap U_2 = \{\mathbf{0}\}$ .

*Proof.* We shall prove that  $U_1 + U_2$  being a direct sum implies that  $U_1 \cap U_2 = \{\mathbf{0}\}$  by considering the contrapositive. Suppose on contrary that there exists some  $\mathbf{u}_0 \neq \mathbf{0}$  in  $U_1 \cap U_2$ , then clearly  $-\mathbf{u}_0 \in U_1 \cap U_2$  and so  $\mathbf{0} = \mathbf{u}_0 + (-\mathbf{u}_0) = \mathbf{0} + \mathbf{0}$  is expressible in two distinct ways. Therefore,  $U_1 + U_2$  is not a direct sum.

Suppose conversely that  $U_1 \cap U_2 = \{\mathbf{0}\}$ . Suppose that there exist some  $\mathbf{u}' \in U_1$  and  $\mathbf{w}' \in U_2$  such that  $\mathbf{u} + \mathbf{w} = \mathbf{u}' + \mathbf{w}'$ , then we have  $\mathbf{u} - \mathbf{u}' = \mathbf{w}' - \mathbf{w}$ . Notice that  $\mathbf{u} - \mathbf{u}' \in U_1$  and  $\mathbf{w}' - \mathbf{w} \in U_2$ , so we have  $\mathbf{u} - \mathbf{u}', \mathbf{w}' - \mathbf{w} \in U_1 \cap U_2 = \{\mathbf{0}\}$ , i.e.,

$$\mathbf{u} - \mathbf{u}' = \mathbf{w}' - \mathbf{w} = \mathbf{0}.$$

Therefore,  $\mathbf{u} = \mathbf{u}'$  and  $\mathbf{w} = \mathbf{w}'$ , i.e.,  $\mathbf{u}$  and  $\mathbf{w}$  are unique. □

*Remark.* Another interpretation here is that  $U_1 + U_2$  is a direct sum if and only if  $\mathbf{0}$  is uniquely expressible as  $\mathbf{u}_1 + \mathbf{u}_2$  for some  $\mathbf{u}_1 \in U_1$  and  $\mathbf{u}_2 \in U_2$ .

In particular, we would like to consider  $\mathbb{F}^n$  for a general field  $\mathbb{F}$ . We can define the dot product operation over  $\mathbb{F}^n$  in the same way as  $\mathbb{R}^n$ . Take any subspace  $U \subseteq \mathbb{F}^n$  and define the set

$$U_{\perp} := \{\mathbf{u} \in \mathbb{F}^n : \mathbf{u} \cdot \mathbf{v} = 0 \text{ for all } \mathbf{v} \in U\},$$

then  $\mathbb{F}^n = U \oplus U_{\perp}$ .

To justify this, we first take any  $\mathbf{v} \in \mathbb{F}^n$ . Using some calculus, we can show that there exists

$$\mathbf{u}_0 = \operatorname{argmin}_{\mathbf{u} \in U} |\mathbf{u} \cdot \mathbf{v}|.$$

Let  $\mathbf{w} = \mathbf{v} - \mathbf{u}_0$ , then clearly  $\mathbf{v} = \mathbf{w} + \mathbf{u}_0$  where  $\mathbf{u}_0 \in U$  and  $\mathbf{w} \in U_{\perp}$ . This implies that  $V = U + U_{\perp}$ . Note that  $\mathbf{0}$  is the only vector in  $\mathbb{F}^n$  which is orthogonal to itself, so we have  $U \cap U_{\perp} = \{\mathbf{0}\}$ . It follows that  $V = U \oplus U_{\perp}$ . This is known as the *orthogonal decomposition* of  $\mathbb{F}^n$ .

## 1.2 Isomorphism

Since the underlying structure of a vector space is still a set, the notion of a mapping between two vector spaces is well-defined. However, note that a vector space possesses unique algebraic structures and properties, namely that the linear combinations of any members of the space are still in the space, so we would like to focus on mapping which preserves such properties.

### Definition 1.2.1 ► Homomorphism

Let  $U$  and  $V$  be vector spaces, a **homomorphism** is a mapping  $\phi : U \rightarrow V$  such that

$$\phi(\mathbf{u} + \mathbf{v}) = \phi(\mathbf{u}) + \phi(\mathbf{v}), \quad \phi(c\mathbf{u}) = c\phi(\mathbf{u})$$

for any  $\mathbf{u}, \mathbf{v} \in U$  and  $c \in \mathbb{F}$ .

Naturally, if a homomorphism is bijective, then it means that the elements in two vector spaces have a one-to-one correspondence. In practice, this means we can treat them as equivalent spaces in some sense.

**Definition 1.2.2 ▶ Isomorphism**

An **isomorphism** between vector spaces  $U$  and  $V$  is a homomorphism between them which is bijective.

An interesting fact here is that an isomorphism between any vector spaces is not unique. To see this, let us first consider an arbitrary vector space  $V$ . Now, we can always find the trivial isomorphism  $\text{id}_V : V \rightarrow V$ . In fact, any mapping  $\phi : v \mapsto cv$  where  $c$  is a scalar is clearly an isomorphism from  $V$  to  $V$ . This means that there are infinitely many isomorphisms from  $V$  to itself.

Let  $U$  be an arbitrary vector space such that there exists some isomorphism  $\psi : V \rightarrow U$ . We consider the following theorem:

**Theorem 1.2.3 ▶ Composition Preserves Isomorphism**

Let  $U, V, W$  be vector spaces. If  $\phi : U \rightarrow V$  and  $\psi : V \rightarrow W$  are isomorphisms, then the composite mapping  $\psi \circ \phi : U \rightarrow W$  is an isomorphism.

*Proof.* Since both  $\phi$  and  $\psi$  are bijective, it is clear that  $\psi \circ \phi$  is bijective. Take any  $u, v \in U$ , then

$$\psi(\phi(u + v)) = \psi(\phi(u) + \phi(v)) = \psi(\phi(u)) + \psi(\phi(v))$$

since  $\phi(u), \phi(v) \in V$ . Therefore,  $\psi \circ \phi$  is an isomorphism. □

Using Theorem 1.2.3, we can immediately see that if  $\phi : V \rightarrow V$  is any isomorphism and  $\psi : V \rightarrow U$  is an isomorphism, then  $\psi \circ \phi$  is an isomorphism between  $V$  and  $U$ . Therefore, there are infinitely many isomorphisms between  $V$  and  $U$ .

*Remark.* If  $V$  is isomorphic to  $U$ , we write  $V \cong U$ . Clearly,  $\cong$  is an equivalence relation.

Now, observe that for any field  $\mathbb{F}$ , the set  $\mathbb{F}^n$  for any  $n \in \mathbb{N}$  is a vector space over  $\mathbb{F}$ .

**Definition 1.2.4 ▶ Finite-Dimensional Vector Space**

A vector space  $V$  is said to be **finite-dimensional** over a field  $\mathbb{F}$  if it is isomorphic to  $\mathbb{F}^n$  for some  $n \in \mathbb{N}$ .  $n$  is called the **dimension** of  $V$ .

Obviously, a vector space which is not finite-dimensional is called *infinite-dimensional*. For example, the set of all polynomials is a vector space of infinite dimension.



**Definition 1.2.5 ► Polynomial**

A function  $p : \mathbb{F} \rightarrow \mathbb{F}$  is said to be a **polynomial** with coefficients in  $\mathbb{F}$  if there exists scalars  $a_0, a_1, \dots, a_n \in \mathbb{F}$  such that

$$p(x) = \sum_{i=0}^n a_i x^i, \quad \text{for all } x \in \mathbb{F}.$$

The largest  $m$  such that  $a_m \neq 0$  is known as the **degree** of  $p$ , denoted by  $\deg p$ . We say that  $\deg p = -\infty$  if  $p = 0$ . We denote the set of all polynomials with coefficients in  $\mathbb{F}$  by  $\mathcal{P}(\mathbb{F})$ , and the set of all such polynomials with degrees at most  $m$  by  $\mathcal{P}_m(\mathbb{F})$ .

Clearly, both  $\mathcal{P}(\mathbb{F})$  and  $\mathcal{P}_m(\mathbb{F})$  are vector spaces over  $\mathbb{F}$ . One can also verify that they are subspaces of  $\mathbb{F}^{\mathbb{F}}$  (the set of all functions from  $\mathbb{F}$  to  $\mathbb{F}$ ). However  $\mathcal{P}_m(\mathbb{F})$  is finite-dimensional while  $\mathcal{P}(\mathbb{F})$  is clearly infinite-dimensional.

### 1.3 Basis

Recall that a *linear combination* of vectors is in the form of

$$\sum a_i \mathbf{v}_i = a_1 \mathbf{v}_1 + a_2 \mathbf{v}_2 + \dots.$$

In case where no confusion is caused, this can be abbreviated as  $a_i \mathbf{v}_i$ . Recall also that the *span* of a set of vectors is defined as

$$\text{span}(V) := \{a_i \mathbf{v}_i : \mathbf{v}_i \in V\},$$

where  $a_i$ 's are scalars. A span of a subset of a vector space  $V$  is clearly a subspace of  $V$ . By deploying a similar proof technique as Proposition 1.1.8, we can prove the following result:

**Proposition 1.3.1 ► Linear Span as the Minimal Containing Subspace**

*Let  $U \subseteq V$  be a set of vectors, then  $\text{span}(U)$  is the minimal subspace of  $V$  containing  $U$ .*

We also know that a set of vectors  $S$  is said to be *linearly independent* if and only if  $a_i \mathbf{v}_i = \mathbf{0}$  implies that  $a_i = 0$  for all  $i = 1, 2, \dots$ . A *basis* of a vector space  $V$  is a linearly independent set  $S$  such that  $\text{span}(S) = V$ . One may check that if  $S$  is a basis for  $V$ , then any  $\mathbf{v} \in V$  can be **uniquely** expressed as a linear combination of the members of  $S$ , but  $S$  itself is not unique. In particular, the coefficients in this linear combination is known as the *components* of  $\mathbf{v}$  relative to  $S$ .

We will see that the basis is closely related to the dimension of vector spaces. First, let us consider the trivial basis for  $\mathbb{F}^n$ .

### Definition 1.3.2 ► Canonical Basis

The **canonical basis** for  $\mathbb{F}^n$  is defined as  $\{\mathbf{e}_i : i = 1, 2, \dots, n\}$ , where each  $\mathbf{e}_i$  is a column vector with  $1_{\mathbb{F}}$  in its  $i$ -th row and  $0_{\mathbb{F}}$  in the other rows.

It is easy to see that the number of vectors in any basis of a finite-dimensional vector space  $V$  is uniquely equal to its dimension.

### Proposition 1.3.3 ► Dimension as Cardinality of Basis

*Let  $V$  be a finite-dimensional vector space with dimension  $n$  and basis  $S$ , then  $n = |S|$ .*

*Proof.* Note that  $V \cong \mathbb{F}^n$ . Let  $\mathbf{v} \in V$  be an arbitrary vector, then there is some  $\mathbf{u} \in \mathbb{F}^n$  and a bijection  $\phi : \mathbb{F}^n \rightarrow V$  such that

$$\begin{aligned}\mathbf{v} &= \phi(\mathbf{u}) \\ &= \phi\left(\sum_{i=1}^n a_i \mathbf{e}_i\right) \\ &= \sum_{i=1}^n a_i \phi(\mathbf{e}_i),\end{aligned}$$

where  $a_i \in \mathbb{F}$  for  $i = 1, 2, \dots, n$ . This means that  $V$  is spanned by at most  $n$  vectors and so its basis is finite. Suppose on contrary that  $|S| = m < n$ , then for any  $\mathbf{w} \in \mathbb{F}^n$ , there is some  $\mathbf{r} \in V$  such that

$$\begin{aligned}\mathbf{w} &= \phi^{-1}(\mathbf{r}) \\ &= \phi^{-1}\left(\sum_{i=1}^m b_i \mathbf{s}_i\right) \\ &= \sum_{i=1}^m b_i \phi^{-1}(\mathbf{s}_i),\end{aligned}$$

where  $b_i \in \mathbb{F}$  for  $i = 1, 2, \dots, m$ . This means that  $\mathbb{F}^n$  is spanned by  $m$  vectors, which is a contradiction. Therefore,  $|S| = n$ . □

An immediate corollary from Proposition 1.3.3 is that a finite-dimensional vector space always has a unique dimension, as otherwise it will have bases with different cardinalities.

Note that since every vector in a vector space  $V$  can be uniquely expressed as a linear combination of a basis  $S$  for  $V$ , this really means that we can view the notion of a basis equivalently

as a bijection between  $\mathbb{F}^n$  and  $V$ , i.e., for any  $(a_1, a_2, \dots, a_n) \in \mathbb{F}^n$ , we can map the tuple to a vector in  $V$  whose components are exactly  $a_1, a_2, \dots, a_n$ .

Now, let us denote a basis for  $V$  by the mapping  $z$ . Notice that for any  $\mathbf{c}_1, \mathbf{c}_2 \in \mathbb{F}^n$ , we have

$$z(\alpha \mathbf{c}_1 + \mathbf{c}_2) = \alpha z(\mathbf{c}_1) + z(\mathbf{c}_2),$$

so a basis is nothing more but an isomorphism! In fact, we could derive our old definition for a basis using this novel definition.

#### Definition 1.3.4 ► Basis

Let  $V$  be an  $n$ -dimensional vector space and  $z : \mathbb{F}^n \rightarrow V$  be any isomorphism, then the set

$$\{z(\mathbf{e}_i) : i = 1, 2, \dots, n\}$$

is known as a **basis** for  $V$ .

Let us do the verification. First, take any  $\mathbf{v} \in V$ , then  $z^{-1}(\mathbf{v}) \in \mathbb{F}^n$  is the coordinate vector of  $\mathbf{v}$ . Note that there are scalars  $a_1, a_2, \dots, a_n$  such that

$$z^{-1} = \sum_{i=1}^n a_i \mathbf{e}_i.$$

Therefore,

$$\mathbf{v} = z(z^{-1}(\mathbf{v})) = z\left(\sum_{i=1}^n a_i \mathbf{e}_i\right) = \sum_{i=1}^n a_i z(\mathbf{e}_i).$$

This implies that  $V = \text{span}\{z(\mathbf{e}_i) : i = 1, 2, \dots, n\}$ . Let  $\sum_{i=1}^n a_i z(\mathbf{e}_i) = \mathbf{0}$ , then since  $z^{-1}$  is bijective, we have

$$z^{-1}\left(\sum_{i=1}^n a_i z(\mathbf{e}_i)\right) = \sum_{i=1}^n a_i \mathbf{e}_i = \mathbf{0}.$$

Since the  $\mathbf{e}_i$ 's are linearly independent, we have  $a_i = 0$  for all  $i = 1, 2, \dots, n$ , and so  $\{z(\mathbf{e}_i) : i = 1, 2, \dots, n\}$  is a linearly independent set and is therefore a basis for  $V$ .

Therefore, every isomorphism  $z : \mathbb{F}^n \rightarrow V$  defines a basis for  $V$ .

# Linear Transformations

## 2.1 Linear Transformations

A common misconception is that linear algebra is a study of matrices — it is not. The real subject studied in linear algebra is **mappings** between vector spaces which preserves linearity.

### Definition 2.1.1 ► Linear Transformation

A **linear transformation** is a mapping  $T : V \rightarrow W$ , where  $V$  and  $W$  are vector spaces over  $\mathbb{F}$ , such that

$$T(\mathbf{v} + \mathbf{u}) = T(\mathbf{v}) + T(\mathbf{u}), \quad T(c\mathbf{v}) = cT(\mathbf{v})$$

for all  $\mathbf{v}, \mathbf{u} \in V$  and all  $c \in \mathbb{F}$ .

In essence, a linear transformation is a function between vector spaces which preserves the vector structure of its domain. We will see that many notions discussed so far can actually be abstracted into a linear transformation.

Let  $V \cong \mathbb{F}^n$  be a finite-dimensional vector space. Recall that a basis for  $V$  is essentially a bijective mapping  $z : \mathbb{F}^n \rightarrow V$ , so a basis is a linear transformation. In particular, if  $y$  is another basis for  $V$ , then clearly  $y \circ z^{-1}$  is a mapping from  $V$  to itself. Let  $Q = y \circ z^{-1}$ , then we have  $y = Q \circ z$ . This is known as a change of basis.

Now, consider a vector space  $V$  over  $\mathbb{F}$ . Fix some  $\mathbf{v} \in V$  and define a mapping  $\Theta_{\mathbf{v}} : \mathbb{F} \rightarrow V$  by  $\Theta_{\mathbf{v}}(a) = a\mathbf{v}$ . One may check that  $\Theta_{\mathbf{v}}$  is a linear transformation, but it can be easily verified that  $\mathbf{v} \mapsto \Theta_{\mathbf{v}}$  is an isomorphism between  $V$  and  $\{\Theta_{\mathbf{v}} : \mathbf{v} \in V\}$ . Therefore,  $\Theta_{\mathbf{v}}$  is essentially  $\mathbf{v}$ , so vectors are linear transformations as well.

Now, notice that given a linear transformation  $T : A \rightarrow B$ , it is guaranteed that every  $\mathbf{a} \in A$  as an image  $f(\mathbf{a}) \in B$ , but not the other way round, i.e., there could be some  $\mathbf{b} \in B$  without a pre-image in  $A$ . Obviously, what we care more are those elements in  $B$  which actually have pre-images in  $A$ .

**Definition 2.1.2 ► Range**

Let  $T : V \rightarrow W$  be a linear transformation. The **range** of  $T$  is defined as

$$\text{range}(T) := \{T(\mathbf{v}) : \mathbf{v} \in V\}.$$

The dimension of  $\text{range}(T)$  is called the **rank** of  $T$ .

It is immediate from the definition that  $T$  is surjective if and only if  $\text{range}(T) = W$ . Consider  $T(\mathbf{v}), T(\mathbf{u}) \in \text{range}(T)$  for some  $\mathbf{v} \neq \mathbf{u}$  in  $V$ , then we have

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

Clearly,  $\alpha \mathbf{v} + \mathbf{u} \in V$ , so  $\alpha T(\mathbf{v}) + T(\mathbf{u}) \in \text{range}(T)$  and  $\text{range}(T) \subseteq W$  is a vector space.

**Definition 2.1.3 ► Kernel**

Let  $T : V \rightarrow W$  be a linear transformation. The **kernel** of  $T$  is defined as

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

The dimension of  $\ker(T)$  is called the **nullity** of  $T$ .

Obviously,  $\ker(T)$  is a vector space. We claim that  $\text{null}(T)$  is related to the injectivity of  $T$  by the following result:

**Proposition 2.1.4 ► Injectivity Test**

*A linear transformation  $T$  is injective if and only if  $\text{null}(T) = 0$ .*

*Proof.* Suppose that  $T$  is injective. We shall prove that  $\text{null}(T) = 0$  by considering the contrapositive. Suppose that  $\text{null}(T) \neq 0$ , then there is some non-zero  $\mathbf{v} \in \ker(T)$  with  $T(\mathbf{v}) = \mathbf{0}$ , so  $T$  is not injective.

Suppose conversely that  $\text{null}(T) = 0$ , then  $\ker(T) = \{\mathbf{0}\}$ . Let  $T(\mathbf{v}) = T(\mathbf{u})$ , then

$$\mathbf{0} = T(\mathbf{v}) - T(\mathbf{u}) = T(\mathbf{v} - \mathbf{u}),$$

so  $\mathbf{v} - \mathbf{u} \in \ker(T)$ . This means that  $\mathbf{v} - \mathbf{u} = \mathbf{0}$  so  $\mathbf{v} = \mathbf{u}$ . Therefore,  $T$  is injective.  $\square$

For any linear transformation  $T : V \rightarrow W$ , we have

$$T(\mathbf{0}_V) = T(\mathbf{0}_V + \mathbf{0}_V) = 2T(\mathbf{0}_V).$$

Cancelling  $T(\mathbf{0}_V)$  on both sides yields  $\mathbf{0}_W = T(\mathbf{0}_V)$ . Therefore,  $\ker(T) - \{\mathbf{0}_V\}$  consists of all non-zero vectors which are mapped to zero under  $T$ , i.e., for all  $\mathbf{u} \in U := V - (\ker(T) - \{\mathbf{0}_V\})$ , we know that  $T(\mathbf{u}) \neq \mathbf{0}_W$ .

### Theorem 2.1.5 ► Fundamental Theorem of Linear Transformations

Let  $T : V \rightarrow W$  be a linear transformation, then  $\dim(V) = \text{rank}(T) + \text{null}(T)$ .

*Proof.* Define  $U := V - (\ker(T) - \{\mathbf{0}_V\})$ , then  $V = U \oplus \ker(T)$ . Let  $S : U \rightarrow \text{range}(T)$  be defined by  $S(\mathbf{u}) = T(\mathbf{u})$ . Note that for all  $\mathbf{v} \in \text{range}(T)$ , there exists some  $\mathbf{u} \in U$  such that  $S(\mathbf{u}) = \mathbf{v}$ , so  $S$  is surjective. Suppose there exist vectors  $\mathbf{u}_1, \mathbf{u}_2 \in U$  such that  $S(\mathbf{u}_1) = S(\mathbf{u}_2)$ , then

$$\mathbf{0}_W = S(\mathbf{u}_1) - S(\mathbf{u}_2) = S(\mathbf{u}_1 - \mathbf{u}_2).$$

This means that  $\mathbf{u}_1 - \mathbf{u}_2 = \mathbf{0}_U$ , and so  $\mathbf{u}_1 = \mathbf{u}_2$ , which implies that  $S$  is injective. Therefore,  $S$  is a bijection and so  $U \cong \text{range}(T)$ , which means  $\dim(U) = \text{rank}(T)$ . Therefore,

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

□

An interesting application of Theorem 2.1.5 gives the following corollary:

### Corollary 2.1.6 ► Bijectivity of Reflexive Mappings

Let  $T : V \rightarrow W$  be a linear transformation, where  $\dim(V) = \dim(W)$ , then  $T$  is injective if and only if it is surjective.

*Proof.* Suppose that  $T$  is injective, then  $\text{null}(T) = 0$ . By Theorem 2.1.5, we have

$$\text{rank}(T) = \dim(V) - \text{null}(T) = \dim(W).$$

Let  $z_1 : \mathbb{F}^n \rightarrow W$  be an isomorphism for some  $n \in \mathbb{Z}^+$ . Note that there exists an isomorphism  $z_2 : \mathbb{F}^n \rightarrow \text{range}(T)$ . Therefore, the mapping  $z_1 \circ z_2^{-1} : \text{range}(T) \rightarrow W$  is an isomorphism. However,  $\text{range}(T) \subseteq W$ , which implies that  $\text{range}(T) = W$ . This means that  $T$  is surjective.

Suppose conversely that  $T$  is surjective, then we must have  $W = \text{range}(T)$  and so  $\text{rank}(T) = \dim(W)$ . Therefore, by Theorem 2.1.5,

$$\text{null}(T) = \dim(V) - \text{rank}(T) = \dim(W) - \text{rank}(T) = 0.$$

By Proposition 2.1.4,  $T$  is injective. □

## 2.2 Duality

Let  $V$  and  $W$  be vector spaces and define  $\mathcal{L}(V, W)$  to be the set of all linear transformations from  $V$  to  $W$ . One may check that  $\mathcal{L}(V, W)$  is a vector space.

### Definition 2.2.1 ► Dual Space

Let  $V$  be a vector space over  $\mathbb{F}$ . The **dual space** of  $V$  is defined as  $\mathcal{V} := \mathcal{L}(V, \mathbb{F})$ . The elements of  $\mathcal{V}$  are called **dual vectors**.

**Author's Note:** in almost all fields of mathematics, the conventional notation for the dual space of  $V$  is  $V^*$ , but somehow the lecturer of this course decides (yet again) to use a confusing syntax by writing it as  $\mathcal{V}$ .

An intuitive example for an element of  $\mathcal{V}$  is as follows: let  $\{\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_n\}$  be a basis for  $V$ . For any  $\mathbf{v} \in V$ , we can write

$$\mathbf{v} = \sum_{i=1}^n a_i \mathbf{z}_i$$

for  $a_1, a_2, \dots, a_n \in \mathbb{F}$ . Now, define a mapping  $\zeta^i : V \rightarrow \mathbb{F}$  as  $\zeta^i(\mathbf{v}) = a_i$ , then clearly  $\zeta^i \in \mathcal{V}$  for  $i = 1, 2, \dots, n$ .

In particular, we see that

$$\zeta^i(\mathbf{z}_j) = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}.$$

Intuitively, this means that for each  $\mathbf{v} \in V$ , we have  $\mathbf{v} = \sum_{i=1}^n a_i \mathbf{z}_i$ .

### Definition 2.2.2 ► Dual Basis

Let  $V$  be a vector space with a basis  $\{\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_n\}$ . The **dual basis** of  $\{\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_n\}$  is defined as the set of all  $\zeta^i \in \mathcal{V}$  such that

$$\zeta^i(\mathbf{z}_j) = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}.$$

Let  $\zeta$  be a dual basis and  $\alpha \in \mathcal{V}$  be a linear mapping over a finite-dimensional vector space

$V$  with  $\dim(V) = n$ . For any  $\mathbf{v} \in V$  with respect to basis  $\mathbf{z}$ , define a mapping  $\beta : V \rightarrow \mathbb{F}$  by

$$\beta(\mathbf{v}) = \left( \sum_{i=1}^n \alpha(\mathbf{z}_i) \zeta^i \right) (\mathbf{v}).$$

Clearly,  $\beta$  is a linear combination of the elements of  $\zeta$ . For any  $\mathbf{z}_j \in \mathbf{z}$ , we have

$$\begin{aligned} \beta(\mathbf{z}_j) &= \left( \sum_{i=1}^n \alpha(\mathbf{z}_i) \zeta^i \right) (\mathbf{z}_j) \\ &= \sum_{i=1}^n \alpha(\mathbf{z}_i) \zeta^i(\mathbf{z}_j) \\ &= \alpha(\mathbf{z}_j). \end{aligned}$$

Since  $\mathbf{z}$  is a basis for  $V$ , this implies that for any  $\mathbf{v} \in V$ , we have  $\beta(\mathbf{v}) = \alpha(\mathbf{v})$ . Therefore,  $\alpha \in \text{span}(\zeta)$ .

Consider

$$\sum_{i=1}^n p_i \zeta^i = 0_{\mathfrak{p}},$$

which is the zero mapping from  $V$  to  $\{0\}$ , this means that for any  $\mathbf{v} \in V$ , we have

$$\begin{aligned} 0 &= \left( \sum_{i=1}^n p_i \zeta^i \right) (\mathbf{v}) \\ &= \left( \sum_{i=1}^n p_i \zeta^i \right) \left( \sum_{j=1}^n a_j \mathbf{z}_j \right) \\ &= \sum_{i=1}^n \left( p_i \sum_{j=1}^n a_j \zeta^i(\mathbf{z}_j) \right) \\ &= \sum_{i=1}^n p_i a_i. \end{aligned}$$

However, since the  $a_i$ 's are arbitrary, we have  $p_i = 0$  for all  $i = 1, 2, \dots, n$ . Therefore,  $\zeta$  is linearly independent. This means that  $\zeta$  is indeed a basis. Clearly, this also implies that any finite-dimensional vector space has the same dimension as its dual space.

### Proposition 2.2.3 ► Dimension of Dual Space

*Let  $V$  be any finite-dimensional vector space, then  $\dim(V) = \dim(\mathfrak{p})$ .*



Furthermore, note that for any dual vector  $\phi = \sum_{i=1}^n p_i \zeta^i$ , we have

$$\phi(\mathbf{z}_j) = \sum_{i=1}^n p_i \zeta^i(\mathbf{z}_j) = p_j.$$

Therefore, any dual vector with respect to basis  $\mathbf{z}$  can be written as

$$\phi = \sum_{i=1}^n \phi(\mathbf{z}_i) \zeta^i.$$

Recall that in Definition 1.3.2, we defined the canonical basis of  $\mathbb{F}^n$  to be the set of unit vectors with a single non-zero entry. Similarly, we can define the canonical basis for the dual space  $\mathbb{F}^n := \mathcal{L}(\mathbb{F}^n, \mathbb{F})$ .

#### Definition 2.2.4 ► Canonical Dual Basis

Write each  $\mathbf{v} \in \mathbb{F}^n$  as

$$\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}.$$

The **canonical dual basis** of  $\mathbb{F}^n$  is defined as  $\{\epsilon^i : i = 1, 2, \dots, n\}$  such that  $\epsilon^i(\mathbf{v}) = v_i$  for all  $\mathbf{v} \in \mathbb{F}^n$ .

Consider any  $p \in \mathbb{F}^n$ , then we can write

$$p = \sum_{i=1}^n q_i \epsilon^i$$

where  $q_i \in \mathbb{F}$  for  $i = 1, 2, \dots, n$  and  $\epsilon^i$ 's are the dual canonical basis. Recall that a basis of an  $n$ -dimensional vector space  $V$  is actually a mapping from  $\mathbb{F}^n$  to  $V$ , so it follows that the dual basis is in fact a mapping  $\zeta : \mathbb{F}^n \rightarrow \mathcal{V}$ . We claim that this  $\zeta$  is in fact just a mapping that satisfies

$$\zeta(p)(\mathbf{z}(\mathbf{a})) = p(\mathbf{a})$$

for any  $p \in \mathbb{F}^n$  and  $\mathbf{a} \in \mathbb{F}^n$ .

To prove that our new definition is consistent with Definition 2.2.2, it suffices to show that

$\zeta(\epsilon^i) = \zeta^i$ . Notice that

$$\begin{aligned}\zeta(\epsilon^i)(z(\mathbf{e}_j)) &= \epsilon^i(\mathbf{e}_j) \\ &= \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}.\end{aligned}$$

However,  $z(\mathbf{e}_j) = \mathbf{z}_j$ , so by Definition 2.2.2 we have  $\zeta(\epsilon^i)(\mathbf{z}_j) = \zeta^i(\mathbf{z}_j)$ . Let  $\mathbf{v} \in V$  be any vector with

$$\mathbf{v} = \sum_{i=1}^n a_i \mathbf{z}_i,$$

then

$$\begin{aligned}\zeta(\epsilon^i)(\mathbf{v}) &= \zeta(\epsilon^i)\left(\sum_{j=1}^n a_j \mathbf{z}_j\right) \\ &= \sum_{j=1}^n a_j \zeta(\epsilon^i)(\mathbf{z}_j) \\ &= \sum_{j=1}^n a_j \zeta^i(\mathbf{z}_j) \\ &= a_i \\ &= \zeta^i(\mathbf{v}).\end{aligned}$$

Therefore,  $\zeta(\epsilon^i) = \zeta^i$ . Let  $\phi \in \mathfrak{P}$  be a mapping, then

$$\begin{aligned}\phi &= \sum_{i=1}^n a_i \zeta^i \\ &= \sum_{i=1}^n a_i \zeta(\epsilon^i) \\ &= \zeta\left(\sum_{i=1}^n a_i \epsilon^i\right) \\ &= \zeta(p)\end{aligned}$$

for some  $p \in \mathbb{P}^n$ . Therefore,  $\zeta$  is surjective. One may check that  $\zeta$  is injective. Moreover,

note that for  $p, q \in \mathbb{F}^n$ ,

$$\begin{aligned}(m\zeta(p) + n\zeta(q))(z(\mathbf{a})) &= m\zeta(p)(z(\mathbf{a})) + n\zeta(q)(z(\mathbf{a})) \\ &= mp(\mathbf{a}) + nq(\mathbf{a}) \\ &= (mp + nq)(\mathbf{a}) \\ &= \zeta(mp + nq)(z(\mathbf{a})).\end{aligned}$$

Therefore,  $\zeta$  is a linear mapping and so it is an isomorphism. Indeed, this implies that  $\zeta$  is a basis for  $\mathfrak{V}$ .

Naturally, it feels justified to define a “dual space of the dual space” of  $V$ , namely

$$\mathfrak{V} := \mathcal{L}(\mathfrak{V}, \mathbb{F}) = \mathcal{L}(\mathcal{L}(V, \mathbb{F}), \mathbb{F}).$$

We can in fact prove that this space is isomorphic to  $V$  itself.

### Proposition 2.2.5 ► Double Dual Space

*Let  $V$  be a finite-dimensional vector space, then the double-dual space*

$$\mathfrak{V} := \mathcal{L}(\mathfrak{V}, \mathbb{F}) = \mathcal{L}(\mathcal{L}(V, \mathbb{F}), \mathbb{F})$$

*is isomorphic to  $V$ .*

*Proof.* Define a mapping  $K: V \rightarrow \mathfrak{V}$  by  $K(\mathbf{v}) = f$  where  $f: \mathfrak{V} \rightarrow \mathbb{F}$  is such that  $f(\alpha) = \alpha(\mathbf{v})$ . It suffices to prove that  $K$  is an isomorphism. First, note that for any  $\mathbf{u}, \mathbf{v} \in V$  and any  $c \in \mathbb{F}$ ,

$$\begin{aligned}K(c\mathbf{v} + \mathbf{u})(\alpha) &= \alpha(c\mathbf{v} + \mathbf{u}) \\ &= c\alpha(\mathbf{v}) + \alpha(\mathbf{u}) \\ &= cK(\mathbf{v})(\alpha) + K(\mathbf{u})(\alpha) \\ &= (cK(\mathbf{v}) + K(\mathbf{u}))(\alpha)\end{aligned}$$

for all  $\alpha \in \mathfrak{V}$ . Therefore,  $K$  is a homomorphism. Let  $K(\mathbf{v}) = 0_{\mathfrak{V} \rightarrow \mathbb{F}}$  be the zero mapping, then

$$K(\mathbf{v})(\alpha) = \alpha(\mathbf{v}) = 0$$

for all  $\alpha \in \mathfrak{V}$ . This implies that  $\mathbf{v} = \mathbf{0}$ , and so  $\text{null}(K) = \{0\}$  since  $V$  is finite-dimensional. By Proposition 2.1.4,  $K$  is injective, and so by Corollary 2.1.6,  $K$  is bijective. Therefore,  $K$  is an isomorphism and so  $\mathfrak{V} \cong V$ .  $\square$

Therefore, this means that it is perfectly legal to view each  $\mathbf{v} \in V$  as a mapping from  $\mathfrak{V}$  to

$\mathbb{F}$ . We can verify that  $\mathbf{v}$  is linear, because

$$\begin{aligned}\mathbf{v}(m\alpha + n\beta) &= m\alpha(\mathbf{v}) + n\beta(\mathbf{v}) \\ &= m\mathbf{v}(\alpha) + n\mathbf{v}(\beta).\end{aligned}$$

Therefore, we can in some sense view a vector space and its dual space the *dual* of each other. Consider a mapping  $T : V \rightarrow V$  for some vector space  $V$ , the duality between  $V$  and  $\mathcal{V}$  tempts us to think that there exists some mapping  $S : \mathcal{V} \rightarrow \mathcal{V}$  with a certain correspondence to  $T$ .

#### Definition 2.2.6 ▶ Transpose

Let  $T : V \rightarrow V$  be a linear transformation. The **transpose** of  $T$  is defined as the mapping  $\mathfrak{T} : \mathcal{V} \rightarrow \mathcal{V}$  such that

$$\mathfrak{T}(\alpha)(\mathbf{v}) = \alpha(T(\mathbf{v}))$$

for any  $\alpha \in \mathcal{V}$ .

We can verify that the transpose is still linear, because for any  $\alpha, \beta \in \mathcal{V}$  and any  $c \in \mathbb{F}$ , we have

$$\begin{aligned}\mathfrak{T}(c\alpha + \beta)(\mathbf{v}) &= (c\alpha + \beta)(T(\mathbf{v})) \\ &= c\alpha(T(\mathbf{v})) + \beta(T(\mathbf{v})) \\ &= c\mathfrak{T}(\alpha)(\mathbf{v}) + \mathfrak{T}(\beta)(\mathbf{v}) \\ &= (c\mathfrak{T}(\alpha) + \mathfrak{T}(\beta))(\mathbf{v})\end{aligned}$$

for all  $\mathbf{v} \in V$ , which implies that  $\mathfrak{T}(c\alpha + \beta) = c\mathfrak{T}(\alpha) + \mathfrak{T}(\beta)$ .

In fact, the very action of taking the transpose is a linear transformation!

#### Proposition 2.2.7 ▶ Linearity of Transposition

Let  $\tau$  be a mapping over  $\mathcal{L}(V, V)$  for some finite-dimensional vector space  $V$  such that  $\tau(T) = \mathfrak{T}$ , then  $\tau$  is a linear transformation.

*Proof.* Take any  $\alpha, \beta \in \mathcal{V}, \mathcal{V}$  and any  $c \in \mathbb{F}$ . For any  $\gamma \in \mathcal{L}(V, V)$  and any  $\mathbf{v} \in V$ ,

consider

$$\begin{aligned}
 \tau(c\alpha + \beta)(\gamma)(\mathbf{v}) &= \widehat{(c\alpha + \beta)}(\gamma)(\mathbf{v}) \\
 &= \gamma((c\alpha + \beta)(\mathbf{v})) \\
 &= c\gamma(\alpha(\mathbf{v})) + \gamma(\beta(\mathbf{v})) \\
 &= c\tau(\alpha)(\gamma)(\mathbf{v}) + \tau(\beta)(\gamma)(\mathbf{v}) \\
 &= c\tau(\alpha)(\gamma)(\mathbf{v}) + \tau(\beta)(\gamma)(\mathbf{v}).
 \end{aligned}$$

Since  $\mathbf{v}$  is arbitrary, this means that  $\tau(c\alpha + \beta)(\gamma) = c\tau(\alpha)(\gamma) + \tau(\beta)(\gamma)$  for all  $\gamma \in \mathcal{L}(V, V)$ , and so  $\tau(c\alpha + \beta) = c\tau(\alpha) + \tau(\beta)$ . Therefore,  $\tau$  is a linear transformation.  $\square$

The above is essentially saying that the transpose of a linear combination equals the linear combination of the transposes of each terms.

## 2.3 Tensor Product

An important concept related to dual spaces is the *tensor product*. It is sometimes referred to as an operation and other times as a space itself.

### Definition 2.3.1 ► Tensor Product between Vector Spaces and Their Dual

Let  $V$  be a vector space with the dual space  $\mathfrak{V}$ . For any  $\mathbf{v} \in V$  and  $\alpha \in \mathfrak{V}$ , their **tensor product** is defined as a mapping  $\mathbf{v} \otimes \alpha \in \mathcal{L}(V, V)$  such that

$$(\mathbf{v} \otimes \alpha)(\mathbf{u}) = \alpha(\mathbf{u})\mathbf{v}.$$

We can see that  $\mathbf{v} \otimes \alpha$  is linear with respect to both  $\mathbf{v}$  and  $\alpha$ , because

$$\begin{aligned}
 ((\mathbf{u} + \mathbf{v}) \otimes \alpha)(\mathbf{w}) &= \alpha(\mathbf{w})(\mathbf{u} + \mathbf{v}) = \alpha(\mathbf{w})\mathbf{u} + \alpha(\mathbf{w})\mathbf{v} = (\mathbf{u} \otimes \alpha + \mathbf{v} \otimes \alpha)(\mathbf{w}), \\
 (\mathbf{v} \otimes (\alpha + \beta))(\mathbf{w}) &= (\alpha + \beta)(\mathbf{w})\mathbf{v} = \alpha(\mathbf{w})\mathbf{v} + \beta(\mathbf{w})\mathbf{v} = (\mathbf{v} \otimes \alpha + \mathbf{v} \otimes \beta)(\mathbf{w}).
 \end{aligned}$$

In particular, let  $\beta \in \mathfrak{V}$ , then clearly  $\beta \circ (\mathbf{v} \otimes \alpha) \in \mathfrak{V}$ . Define a mapping  $T$  over  $\mathfrak{V}$  by

$$T(\beta) = \beta \circ (\mathbf{v} \otimes \alpha),$$

then we can see that  $T \in \mathcal{L}(\mathfrak{V}, \mathfrak{V})$ . In fact, for any  $\mathbf{u} \in V$ , consider

$$\begin{aligned}
 T(\beta)(\mathbf{u}) &= (\beta \circ (\mathbf{v} \otimes \alpha))(\mathbf{u}) \\
 &= \beta((\mathbf{v} \otimes \alpha)(\mathbf{u})).
 \end{aligned}$$

By Definition 2.2.6,  $T$  is the transpose of  $\mathbf{v} \otimes \alpha$  for any  $\mathbf{v} \in V$  and  $\alpha \in \mathcal{V}$ . Furthermore, note that  $\alpha(\mathbf{u}) \in \mathbb{F}$ , we see that

$$\begin{aligned} T(\beta)(\mathbf{u}) &= \beta((\mathbf{v} \otimes \alpha)(\mathbf{u})) \\ &= \beta(\alpha(\mathbf{u})\mathbf{v}) \\ &= \alpha(\mathbf{u})\beta(\mathbf{v}) \\ &= (\beta(\mathbf{v})\alpha)(\mathbf{u}). \end{aligned}$$

Therefore, actually  $\beta \circ (\mathbf{v} \otimes \alpha) = \beta(\mathbf{v})\alpha$ .

**Proposition 2.3.2 ▶ Tensor Products Form A Basis For  $\mathcal{L}(V, V)$**

Let  $\zeta$  be the dual basis for  $V$  with respect to some basis  $\mathbf{z}$ , then

$$\{\mathbf{z}_i \otimes \zeta^j : i, j = 1, 2, \dots, n\}$$

is a basis for  $\mathcal{L}(V, V)$ .

*Proof.* Let  $T \in \mathcal{L}(V, V)$ . For each  $i = 1, 2, \dots, n$ , we have

$$T(\mathbf{z}_i) = \sum_{j=1}^n a_{ij} \mathbf{z}_j.$$

Define a mapping

$$S := \sum_{i=1}^n \sum_{j=1}^n a_{ij} (\mathbf{z}_j \otimes \zeta^i).$$

Note that  $S \in \mathcal{L}(V, V)$ . For any  $\mathbf{z}_k$ , we have

$$\begin{aligned} S(\mathbf{z}_k) &= \left( \sum_{i=1}^n \sum_{j=1}^n a_{ij} (\mathbf{z}_j \otimes \zeta^i) \right) (\mathbf{z}_k) \\ &= \sum_{i=1}^n \sum_{j=1}^n a_{ij} \zeta^i(\mathbf{z}_k) \mathbf{z}_j \\ &= \sum_{j=1}^n a_{kj} \mathbf{z}_j \\ &= T(\mathbf{z}_k). \end{aligned}$$

Therefore,

$$T = S = \sum_{i=1}^n \sum_{j=1}^n a_{ij} (\mathbf{z}_j \otimes \zeta^i),$$

so  $\mathcal{L}(V, V) = \text{span}\{\mathbf{z}_i \otimes \zeta^j : i, j = 1, 2, \dots, n\}$ . Suppose  $T = 0_{\mathcal{L}}$  is the zero mapping, then for each  $\mathbf{z}_k$  with  $k = 1, 2, \dots, n$ , we have  $S(\mathbf{z}_k) = T(\mathbf{z}_k) = \mathbf{0}$ . This implies that

$$\sum_{j=1}^n a_{kj} \mathbf{z}_j = \mathbf{0}$$

for each  $k = 1, 2, \dots, n$ . Since  $\mathbf{z}$  is a basis, we have  $a_{kj} = 0$  for all  $k, j = 1, 2, \dots, n$ . Therefore, the set  $\{\mathbf{z}_i \otimes \zeta^j : i, j = 1, 2, \dots, n\}$  is linearly independent and so is a basis for  $\mathcal{L}(V, V)$ .  $\square$

Therefore, we see that for every vector space  $V$ , if we fix any basis  $\{\mathbf{z}_i : i = 1, 2, \dots, n\}$ , then any  $T \in \mathcal{L}(V, V)$  can be expressed as a linear combination of  $\mathbf{z}_i \otimes \zeta^j$  where  $\zeta$  is the dual basis with respect to  $\mathbf{z}$ . Proposition 2.3.2 also implies that for any finite-dimensional vector space  $V$ , we have  $\dim(\mathcal{L}(V, V)) = \dim(V)^2$ .

Let  $T \in \mathcal{L}(V, V)$ . Fix a basic vector  $\mathbf{z}_j$  of  $V$  with dual vector  $\zeta^i$ , we can “extract” the  $(i, j)$  component of  $T$  by

$$\zeta^i(T(\mathbf{z}_j)) = \zeta^i\left(\sum_{k=1}^n a_{kj} \mathbf{z}_k\right) = \sum_{k=1}^n a_{kj} \zeta^i(\mathbf{z}_k) = a_{ij}.$$

## 2.4 Matrix

### Definition 2.4.1 ► Matrix

An  $n \times m$  **matrix** is defined as an  $m$ -tuple of column vectors in  $\mathbb{F}^n$ .

Clearly, the set of all  $n \times m$  matrices for any  $m, n \in \mathbb{N}$  is a vector space. We denote this vector space as  $\mathcal{M}_{n \times m}$ . However, as compared to other vectors, we can also define the notion of multiplication between some matrices.

### Definition 2.4.2 ► Matrix Multiplication

Let  $\mathbf{M}$  be an  $n \times m$  matrix and  $\mathbf{N}$  be an  $m \times p$  matrix, then their product  $\mathbf{P} = \mathbf{MN}$  is a matrix whose entries are given by

$$P_j^i = \sum_{k=1}^m M_k^i N_i^k.$$

Consider the vector space  $\mathcal{M}_{n \times n}$  of square matrices. It is easy to see that this space is **closed under multiplication**.

**Definition 2.4.3 ► Algebra**

An **algebra** is a vector space  $V$  with a binary mapping  $\times : V^2 \rightarrow V$  known as multiplication.

It is not difficult to see that from Proposition 2.3.2, the components of a mapping  $T \in \mathcal{L}(V, V)$  can be essentially seen as the entries of a matrix.

**Definition 2.4.4 ► Matrix of A Linear Transformation**

Let  $T \in \mathcal{L}(V, V)$ . If for some basis  $z$  of  $V$ , we have

$$T = \sum_{i=1}^n \sum_{j=1}^n a_{ij} z_i \otimes \zeta^j,$$

then the matrix  $\mathbf{T}$  with  $T_j^i = a_{ij}$  is called the matrix of  $T$  relative to  $z$ .

*Remark.* Note that  $T_j^i = \zeta^i(T(z_j))$ .

Recall that we have defined transpose without using matrices. Now let us verify that our definition stays consistent with the elementary definition of matrix transposition.

**Proposition 2.4.5 ► Matrix Representation of Transpose Mappings**

Let  $T : V \rightarrow V$  be a linear mapping with matrix representation  $\mathbf{T}$ . Let  $\mathfrak{T}$  be the transpose of  $T$ , then the matrix representation of  $\mathfrak{T}$  is such that  $\mathfrak{T}_j^i = T_i^j$ .

*Proof.* Define  $K : V \rightarrow \mathfrak{V}$  by  $K(\mathbf{v})(\alpha) = \alpha(\mathbf{v})$ , then  $K(\mathbf{v}) = \alpha \mapsto \alpha(\mathbf{v})$ . Consider the following lemma:



**Lemma 2.4.6 ▶ Transpose of Tensor Products**

Let  $V$  be a finite-dimensional space. For any  $\mathbf{v} \in V$  and  $\alpha \in \mathfrak{V}$ , we have

$$\widehat{\mathbf{v} \otimes \alpha} = \alpha \otimes K(\mathbf{v}).$$

*Proof.* Let  $f : \mathfrak{V} \rightarrow \mathfrak{V}$ , for any  $\beta \in \mathfrak{V}$  and any  $\mathbf{w} \in V$ , consider

$$\begin{aligned} f(\beta)(\mathbf{w}) &= \beta(f(\mathbf{w})) \\ &= \beta(\alpha(\mathbf{w})\mathbf{v}) \\ &= \alpha(\mathbf{w})\beta(\mathbf{v}), \\ (\alpha \otimes K(\mathbf{v}))(\beta)(\mathbf{w}) &= (K(\mathbf{v})(\beta)\alpha)(\mathbf{w}) \\ &= \alpha(\mathbf{w})\beta(\mathbf{v}). \end{aligned}$$

Therefore,  $f(\beta) = (\alpha \otimes K(\mathbf{v}))(\beta)$  for all  $\beta \in \mathfrak{V}$ , which implies that

$$\mathbf{v} \otimes \alpha = f = \alpha \otimes K(\mathbf{v}).$$

□

Let  $\mathbf{z}_1, \dots, \mathbf{z}_n$  be a basis for  $V$ , then

$$K(\mathbf{z}_i)(\zeta^j) = \zeta^j(\mathbf{z}_i) = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}.$$

Therefore, all of the  $K(\mathbf{z}_i)$ 's forms a dual basis for  $\mathfrak{V}$ . By Proposition 2.3.2 and Definition 2.4.4, we know that

$$\mathfrak{P} = \sum_{i=1}^n \sum_{j=1}^n \mathfrak{P}_j^i (\zeta^i \otimes K(\mathbf{z}_j)), \quad T = \sum_{j=1}^n \sum_{i=1}^n T_i^j (\mathbf{z}_j \otimes \zeta^i).$$

By Proposition 2.2.7, we know that

$$\mathfrak{P} = \sum_{j=1}^n \sum_{i=1}^n T_i^j (\widehat{\mathbf{z}_j \otimes \zeta^i}).$$

By Lemma 2.4.6,  $\widehat{\mathbf{z}_j \otimes \zeta^i} = \zeta^i \otimes K(\mathbf{z}_j)$ , so  $\mathfrak{P}_j^i = T_i^j$ .

□

Let  $V$  be a finite-dimensional vector space with  $\dim(V) = n$  and let  $\mathcal{M}_{n \times n}$  be the vector space of all  $n \times n$  matrices. Define  $M_{\mathbf{z}} : \mathcal{L}(V, V) \rightarrow \mathcal{M}_{n \times n}$  that maps each  $T \in \mathcal{L}(V, V)$  to

its matrix representation  $\mathbf{T}$ . Intuitively, it is an isomorphism.

*Remark.* Note that actually  $\mathcal{M}_{n \times n} \cong \mathbb{F}^{n^2}$  and that a basis for  $\mathcal{L}(V, V)$  is just an isomorphism  $z: \mathbb{F}^{n^2} \rightarrow \mathcal{L}(V, V)$ , so in some sense  $M_z$  is just  $z^{-1}$ .

Consider  $\mathcal{L}(\mathbb{F}^n, \mathbb{F}^n)$ , it is clear that every  $T \in \mathcal{L}(\mathbb{F}^n, \mathbb{F}^n)$  can be expressed as

$$T = \sum_{i=1}^n \sum_{j=1}^n M_e(T)_i^j (\mathbf{e}_i \otimes \epsilon^j).$$

When the context is clear, we can be a bit sloppy and regard  $M_z$  just as  $\mathbf{T}$ .

We can augment  $\mathcal{L}(V, V)$  with the  $\circ$  operation. For any  $S, T \in \mathcal{L}(V, V)$ , clearly  $S \circ T \in \mathcal{L}(V, V)$ , so  $\mathcal{L}(V, V)$  is an algebra. This means that it is justified to write  $S \circ T$  as a product  $ST$ .

Let  $\mathbf{u}, \mathbf{v} \in V$  and  $\alpha, \beta \in \mathcal{V}$ , for any  $\mathbf{w} \in V$ , consider

$$\begin{aligned} ((\mathbf{u} \otimes \alpha) \circ (\mathbf{v} \otimes \beta))(\mathbf{w}) &= (\mathbf{u} \otimes \alpha)(\beta(\mathbf{w})\mathbf{v}) \\ &= \alpha(\beta(\mathbf{w})\mathbf{v})\mathbf{u} \\ &= \beta(\mathbf{w})\alpha(\mathbf{v})\mathbf{u} \\ &= (\alpha(\mathbf{v})\mathbf{u} \otimes \beta)(\mathbf{w}). \end{aligned}$$

Therefore,  $(\mathbf{u} \otimes \alpha) \circ (\mathbf{v} \otimes \beta) = \alpha(\mathbf{v})\mathbf{u} \otimes \beta$ .

Intuitively, the matrix of a composite map can be obtained via matrix multiplication.

#### Proposition 2.4.7 ► Matrices of Composite Mappings

Let  $S, T \in \mathcal{L}(V, V)$ , then for any basis  $z$  of  $V$ ,

$$M_z(ST) = M_z(S)M_z(T).$$

*Proof.* Let  $\zeta^i$  be the dual vector of  $\mathbf{z}_h$ . By Definition 2.4.4,

$$S = \sum_{i=1}^n \sum_{h=1}^n S_i^h (\mathbf{z}_h \otimes \zeta^i), \quad T = \sum_{k=1}^n \sum_{j=1}^n T_k^j (\mathbf{z}_j \otimes \zeta^k).$$

Therefore,

$$\begin{aligned}
 ST &= \left( \sum_{i=1}^n \sum_{h=1}^n S_i^h(\mathbf{z}_h \otimes \zeta^i) \right) \left( \sum_{k=1}^n \sum_{j=1}^n T_k^j(\mathbf{z}_j \otimes \zeta^k) \right) \\
 &= \sum_{i=1}^n \sum_{h=1}^n \sum_{k=1}^n \sum_{j=1}^n S_i^h T_k^j(\mathbf{z}_h \otimes \zeta^i)(\mathbf{z}_j \otimes \zeta^k) \\
 &= \sum_{i=1}^n \sum_{h=1}^n \sum_{k=1}^n \sum_{j=1}^n S_i^h T_k^j \zeta^i(\mathbf{z}_j) \mathbf{z}_h \otimes \zeta^k \\
 &= \sum_{i=1}^n \sum_{h=1}^n \sum_{k=1}^n S_i^h T_k^i \mathbf{z}_h \otimes \zeta^k \\
 &= \sum_{h=1}^n \sum_{k=1}^n (M_z(S)M_z(T))_k^h \mathbf{z}_h \otimes \zeta^k.
 \end{aligned}$$

Therefore,  $M_z(ST) = M_z(S)M_z(T)$ . □

## 2.5 Change of Basis

Let  $z$  be a basis for a vector space  $V$ . For any  $\mathbf{v} \in V$ , define  $\mathbf{v}_z^* := z^{-1}(\mathbf{v})$ . Clearly  $\mathbf{v}_z^* \in \mathbb{F}^n$ . We can view  $\mathbf{v}_z^*$  as the “coordinates” of  $\mathbf{v}$  with respect to  $z$ . Suppose  $\alpha \in \mathfrak{V}$ , we define some  $\alpha_z^* \in \mathbb{F}^n$  by  $\alpha_z^* := \alpha \circ z$ . In particular, suppose the  $i$ -th component of  $\alpha$  is  $\alpha_i$ , we notice that

$$\alpha_z^*(\mathbf{e}_i) = \alpha(z(\mathbf{e}_i)) = \alpha(\mathbf{z}_i) = \alpha_i.$$

So  $\alpha_z^*$  and  $\alpha$  can be represented with the same row vector. Similarly, consider  $T \in \mathcal{L}(V, V)$ . Fix any basis  $z$  of  $V$ . Define  $T_z^* := z^{-1} \circ T \circ z$ . We claim that  $M_z(T_z^*) = M_z(T)$ , because

$$\begin{aligned}
 M_z(T_z^*)_j^i &= \epsilon^i(T_z^*(\mathbf{e}_j)) \\
 &= \epsilon^i((z^{-1} \circ T \circ z)(\mathbf{e}_j)) \\
 &= \epsilon^i(z^{-1}(T(\mathbf{z}_j))) \\
 &= \epsilon^i \circ z^{-1} \left( \sum_{k=1}^n T_j^k \mathbf{z}_k \right) \\
 &= \epsilon^i \left( \sum_{k=1}^n T_j^k (z^{-1}(\mathbf{z}_k)) \right) \\
 &= \left( \sum_{k=1}^n T_j^k \epsilon^i(\mathbf{e}_k) \right) \\
 &= T_j^i.
 \end{aligned}$$

Note that if  $\mathbf{u} = T(\mathbf{v})$  and  $z(\mathbf{a}) = \mathbf{v}$ ,  $z(\mathbf{b}) = \mathbf{u}$ , then

$$\mathbf{b} = (z^{-1} \circ T \circ z)(\mathbf{a}).$$

Let  $z, y$  be bases for some finite-dimensional vector space  $V$ . Define a mapping

$$P := z^{-1} \circ y \in \mathcal{L}(\mathbb{F}^n, \mathbb{F}^n),$$

then if  $\mathbf{v} \in V$  is such that  $\mathbf{v} = z(\mathbf{a}) = y(\mathbf{b})$ , clearly  $\mathbf{a} = P(\mathbf{b})$ . Note that  $y = z \circ P$ , so

$$\begin{aligned} \mathbf{y}_i &= y(\mathbf{e}_i) \\ &= z(P(\mathbf{e}_i)) \\ &= \sum_{k=1}^n M_z(P)_i^k \mathbf{z}_k. \end{aligned}$$

This means that we can obtain the  $i$ -th column of  $M_z(P)$  by expressing  $\mathbf{y}_i$  in terms of the  $\mathbf{z}_k$ 's and fill the coefficients into the column. In fact, this can be summarised into the following:

**Proposition 2.5.1 ► Conversion between Different Bases**

Let  $\{\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_n\}$  be a basis for some vector space  $V$ , then the set  $\{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n\}$  is a basis for  $V$  if and only if there exists some non-singular matrix  $\mathbf{P}$  such that

$$\begin{bmatrix} \mathbf{y}_1 & \mathbf{y}_2 & \cdots & \mathbf{y}_n \end{bmatrix} = \begin{bmatrix} \mathbf{z}_1 & \mathbf{z}_2 & \cdots & \mathbf{z}_n \end{bmatrix} \mathbf{P}.$$

*Proof.* Let  $\{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n\}$  be a basis for  $V$ . Since  $\mathbf{y}_i \in V$ , we have

$$\mathbf{y}_i = \sum_{j=1}^n P_j^i \mathbf{z}_j.$$

Define

$$\mathbf{p}_j := \begin{bmatrix} P_j^1 \\ P_j^2 \\ \vdots \\ P_j^n \end{bmatrix}, \quad \mathbf{P} := \begin{bmatrix} \mathbf{p}_1 & \mathbf{p}_2 & \cdots & \mathbf{p}_n \end{bmatrix}$$

then clearly,

$$\begin{bmatrix} \mathbf{y}_1 & \mathbf{y}_2 & \cdots & \mathbf{y}_n \end{bmatrix} = \begin{bmatrix} \mathbf{z}_1 & \mathbf{z}_2 & \cdots & \mathbf{z}_n \end{bmatrix} \mathbf{P}.$$

Similarly, since  $\{z_1, z_2, \dots, z_n\}$  is also a basis, we can find some matrix  $Q$  such that

$$\begin{bmatrix} z_1 & z_2 & \cdots & z_n \end{bmatrix} = \begin{bmatrix} y_1 & y_2 & \cdots & y_n \end{bmatrix} Q.$$

However, this means that

$$\begin{bmatrix} y_1 & y_2 & \cdots & y_n \end{bmatrix} = \begin{bmatrix} y_1 & y_2 & \cdots & y_n \end{bmatrix} QP,$$

so clearly  $P$  has an inverse  $Q$ .

Suppose conversely that there is some non-singular matrix  $P$  such that

$$\begin{bmatrix} y_1 & y_2 & \cdots & y_n \end{bmatrix} = \begin{bmatrix} z_1 & z_2 & \cdots & z_n \end{bmatrix} P.$$

Let  $a_1, a_2, \dots, a_n$  be scalars such that  $\sum_{i=1}^n a_i y_i = \mathbf{0}$ , then

$$\sum_{i=1}^n \sum_{j=1}^n a_i P_i^j z_j = \mathbf{0}.$$

Therefore,  $a_i P_i^j = 0$  for every  $i, j$ . However,  $P$  is non-singular, so for any fixed  $i$ , not all  $P_i^j$ 's are zero. Therefore,  $a_i = 0$  for all  $i = 1, 2, \dots, n$ . This implies that  $\{y_1, y_2, \dots, y_n\}$  is a set of  $n$  linearly independent vectors and so it must be a basis for  $V$ .  $\square$

Let  $\mathbf{v} \in V$  be a vector with coordinate vector  $\mathbf{v}_y^* \in \mathbb{F}^n$  such that  $y(\mathbf{v}_y^*) = \mathbf{v}$ , then

$$\mathbf{v}_y^* = y^{-1}(\mathbf{v}) = (y^{-1} \circ z)(z^{-1}(\mathbf{v})) = P^{-1}(\mathbf{v}_z^*).$$

Combining everything, we have

### Definition 2.5.2 ► Change of Basis Matrix

Let  $T_z^*$  be the matrix of a linear transformation  $T : V \rightarrow V$  with respect to the basis  $z$  and let  $P$  be the matrix of the mapping  $z^{-1} \circ y$  for some basis  $y$  for  $V$ , then the matrix of  $T$  with respect to  $y$  is

$$T_y^* := P^{-1} T_z^* P.$$

$P$  is called a **change of basis matrix**.

# Operators

## 3.1 Eigenspaces

Since there are infinitely many linear transformations between two vector spaces, it is not of much help to study each one of them individually. Instead, it is useful to consider categorising all linear transformations into different families according to some well-defined properties.

### Definition 3.1.1 ► Operator

A linear transformation  $T : V \rightarrow V$  is called an **operator** on  $V$ .

Let us recall the notion of eigenvectors. In the old-fashioned way, we define a non-zero vector  $\mathbf{v}$  to be an eigenvector if there is some scalar  $\lambda$  such that  $\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$ . We can see that the left-multiplication by  $\mathbf{A}$  actually corresponds to a linear transformation from a vector space to itself, so we can re-write this definition in terms of linear transformations.

### Definition 3.1.2 ► Eigenvector and Eigenvalue

Let  $T : V \rightarrow V$  be an operator on a finite-dimensional vector space  $V$  over  $\mathbb{F}$ . A vector  $\mathbf{v} \in V$  is called an **eigenvector** of  $T$  if  $T(\mathbf{v}) = \lambda\mathbf{v}$  for some  $\lambda \in \mathbb{F}$ , which is known as the **eigenvalue** associated with  $\mathbf{v}$ .

Note that the operator  $T$  is a well-defined function, so for every eigenvector, its associated eigenvalue is unique. The uniqueness of eigenvalues guarantees the following contrapositive statement:

### Proposition 3.1.3 ► Uniqueness of Eigenvalue

*If  $\mathbf{u}, \mathbf{v}$  are eigenvectors of an operator  $T : V \rightarrow V$  associated to eigenvalues  $\lambda \neq \mu$  respectively, then  $\mathbf{u}$  and  $\mathbf{v}$  are linearly independent.*

*Proof.* We shall prove the contrapositive statement. Suppose conversely that  $\mathbf{u}$  and  $\mathbf{v}$  are linearly dependent, then there exists some scalar  $\alpha$  such that  $\mathbf{v} = \alpha\mathbf{u}$ . Note that

$$\mu\mathbf{v} = T(\mathbf{v}) = T(\alpha\mathbf{u}) = \alpha\lambda\mathbf{u} = \lambda\mathbf{v},$$

so  $\mu = \lambda$ . □

However, the converse is not true, because if  $T(\mathbf{v}) = \lambda\mathbf{v}$ , then  $T(\mu\mathbf{v}) = \lambda(\mu\mathbf{v})$  for all  $\mu \in \mathbb{F}$ . In fact, two linearly independent vectors can have the same eigenvalue. A silly example: for any finite-dimensional space  $V$ , consider the identity operator  $\text{id}_V$ . Clearly, every vector in  $V$  is an eigenvector with the associated eigenvalue of 1.

#### Definition 3.1.4 ► Eigenspace

Let  $T$  be an operator on a finite-dimensional vector space  $V$  over  $\mathbb{F}$ . For every eigenvalue  $\lambda \in \mathbb{F}$  associated to some vector in  $V$ , the **eigenspace** associated with  $\lambda$  is

$$E_\lambda := \{\mathbf{v} \in V : T(\mathbf{v}) = \lambda\mathbf{v}\}.$$

One can easily verify that every eigenspace is a subspace. Intuitively, eigenspaces associated to different eigenvalues should be different. To see this more clearly, consider the following proposition, which is a generalisation of Proposition 3.1.3:

#### Proposition 3.1.5 ► Linear Independence between Eigenspaces

Let  $T : V \rightarrow V$  be an operator with distinct eigenvalues  $\lambda_1, \lambda_2, \dots, \lambda_k$  associated with eigenvectors  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k$  respectively, then  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k$  are linearly independent.

*Proof.* The case where  $k = 2$  is exactly Proposition 3.1.3. Suppose that there is some  $m \in \mathbb{Z}^+$  with  $m \geq 2$  such that eigenvectors  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m$  associated with distinct eigenvalues  $\lambda_1, \lambda_2, \dots, \lambda_m$  respectively are linearly independent. Let  $\lambda_{m+1}$  be an eigenvalue associated to  $\mathbf{v}_{m+1}$ , we shall prove that  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_{m+1}$  are linearly independent by considering the contrapositive statement. Suppose conversely that  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_{m+1}$  are linearly dependent vectors. Since  $\mathbf{v}_{m+1} \neq \mathbf{0}$ , there exist scalars  $\alpha_1, \alpha_2, \dots, \alpha_m$  not all zero such that

$$\mathbf{v}_1 = \sum_{i=1}^m \alpha_i \mathbf{v}_i.$$

Therefore,

$$\begin{aligned} T(\mathbf{v}_{m+1}) &= T\left(\sum_{i=1}^m \alpha_i \mathbf{v}_i\right) \\ &= \sum_{i=1}^m \alpha_i T(\mathbf{v}_i) \\ &= \sum_{i=1}^m \alpha_i \lambda_i \mathbf{v}_i, \end{aligned}$$

but  $T(\mathbf{v}_{m+1}) = \lambda_{m+1}\mathbf{v}_{m+1} = \lambda_{m+1} \sum_{i=1}^m \alpha_i \mathbf{v}_i$ , so

$$\sum_{i=1}^m \alpha_i (\lambda_i - \lambda_{m+1}) \mathbf{v}_i = \mathbf{0}.$$

Since  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m$  are linearly independent, this implies that  $\alpha_i (\lambda_i - \lambda_{m+1}) = 0$  for all  $i = 1, 2, \dots, m$ . Note that there exists some integer  $j \in [1, m]$  such that  $\alpha_j \neq 0$ , so  $\lambda_j = \lambda_{m+1}$ , which means that  $\lambda_1, \lambda_2, \dots, \lambda_{m+1}$  are not distinct eigenvalues.  $\square$

Proposition 3.1.5 reveals an important fact that any finite-dimensional vector space has finitely many eigenvalues. In fact, for any  $V \cong \mathbb{F}^n$ , any linearly independent set contains at most  $n$  distinct vectors, so it means that  $V$  can have at most  $\dim(V)$  different eigenvalues.

The next natural question is: given any operator on a finite-dimensional vector space, does the operator always has an eigenvector? By instinct, one may answer “no” because it seems that the rotation transformation in  $\mathbb{R}^2$ , namely

$$(x, y) \mapsto (x \cos \theta - y \sin \theta, x \sin \theta + y \cos \theta),$$

does not have any eigenvector because there is no vector in  $\mathbb{R}^2$  being mapped to a parallel image. However, we notice something interesting when we extend the domain of the rotation transformation to  $\mathbb{C}^2$ , because

$$\begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} 1 \\ -i \end{bmatrix} = \begin{bmatrix} \cos \theta + i \sin \theta \\ -i(\cos \theta + i \sin \theta) \end{bmatrix} = \begin{bmatrix} e^{i\theta} \\ -ie^{i\theta} \end{bmatrix} = e^{i\theta} \begin{bmatrix} 1 \\ -i \end{bmatrix}.$$

Therefore, it is important to note that the eigenvalues and eigenvectors of a real vector space might not be real! The following proposition states this result rigorously:

### Proposition 3.1.6 ► Existence of Eigenvalue

*Every operator on a finite-dimensional vector space over  $\mathbb{C}$  has at least one eigenvalue.*

*Proof.* Let  $V$  be an  $n$ -dimensional vector space over  $\mathbb{C}$  and let  $T : V \rightarrow V$  be an operator. Take any non-zero vector  $\mathbf{v} \in V$ , then  $\mathbf{v}, T(\mathbf{v}), T^2(\mathbf{v}), \dots, T^n(\mathbf{v})$  are  $(n+1)$  vectors in  $V$  which must be linearly dependent. Therefore, there exists  $c_0, c_1, \dots, c_n \in \mathbb{C}$  not all zero such that

$$\sum_{i=0}^n c_i T^i(\mathbf{v}) = \mathbf{0}.$$

This implies that  $\sum_{i=0}^n c_i T^i$  is the zero transformation, i.e.,  $\sum_{i=0}^n c_i T^i = 0$ . Note that this is a polynomial equation with complex coefficients, so it has exactly  $(n+1)$  roots.



Therefore,

$$\sum_{i=0}^n c_i T^i(\mathbf{v}) = a_0 \left( \prod_{i=1}^n (T - a_i \text{id}_V) \right) (\mathbf{v}) = \mathbf{0}$$

for some  $a_0, a_1, \dots, a_n \in \mathbb{C}$ . Note that not all of the  $(T - a_i \text{id}_V)$ 's are injective, because otherwise  $\mathbf{v} = \mathbf{0}$  by Proposition 2.1.4. Therefore, there is some  $T - a_j \text{id}_V$  which is not injective. This means that there exists a non-zero vector  $\mathbf{w} \in V$  such that  $(T - a_j \text{id}_V)(\mathbf{w}) = \mathbf{0}$ , but this implies that  $T(\mathbf{w}) = a_j \mathbf{w}$ , so  $a_j$  is an eigenvalue of  $T$  associated with  $\mathbf{w}$ .  $\square$

## 3.2 Diagonalisation

To better classify operators, we consider a special type of matrices known as *triangular matrices*.

### Definition 3.2.1 ▶ Upper Triangular Matrix

A matrix  $\mathbf{M}$  is called **upper triangular** if  $M_j^i = 0$  whenever  $i > j$ .

We consider the following nice operator on a finite-dimensional vector space  $V$ : the operator  $T : V \rightarrow V$  is such that there exists a basis  $t$  of  $V$  such that  $T(\mathbf{t}_i)$  can be expressed as a linear combination of the  $\mathbf{t}_j$ 's for all  $j \leq i$ . Intuitively, if such an operator exists, then its matrix representation with respect to  $t$  must be upper-triangular. We will introduce several useful preliminary results.

### Definition 3.2.2 ▶ Similar Matrices

Matrices  $\mathbf{A}$  and  $\mathbf{B}$  are **similar** if there exists a matrix  $\mathbf{P}$  such that  $\mathbf{A} = \mathbf{P}^{-1}\mathbf{B}\mathbf{P}$ .

Let  $\lambda$  be an eigenvalue of  $\mathbf{A}$ , then there exists some non-zero vector  $\mathbf{x}$  such that  $\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$ . If  $\mathbf{A} = \mathbf{P}^{-1}\mathbf{B}\mathbf{P}$ , this means that

$$\mathbf{P}^{-1}\mathbf{B}\mathbf{P}\mathbf{x} = \lambda\mathbf{x},$$

and so

$$\mathbf{B}(\mathbf{P}\mathbf{x}) = \lambda\mathbf{P}\mathbf{x}.$$

Therefore,  $\lambda$  is also an eigenvalue of  $\mathbf{B}$  with associated eigenvector  $\mathbf{P}\mathbf{x} \neq \mathbf{0}$  (because  $\mathbf{P} \neq \mathbf{0}$ ). This implies that similar matrices always have the same eigenvalues.

Now, let  $T$  be any operator on  $V$  and  $z : \mathbb{F}^n \rightarrow V$  be a basis, then we know that the operator  $T_z^* := z^{-1} \circ T \circ z$  on  $\mathbb{F}^n$  has the same eigenvalues as  $T$ . Notice also that if  $T$  satisfies the property that  $T(\mathbf{t}_i)$  can be expressed as a linear combination of the  $\mathbf{t}_j$ 's for all  $j \leq i$ , then so does  $T_z^*$ . However, as we have discussed before, the matrix representation of  $T_z^*$  is upper-

triangular. Therefore, we might want to conclude that the matrix representation of any operator on  $V$  is similar to some upper-triangular matrix.

### Theorem 3.2.3 ► Triangularisability

*For every operator  $T : V \rightarrow V$  where  $V$  is a finite-dimensional vector space over  $\mathbb{C}$ , there exists a basis  $t$  of  $V$  such that  $T(t_i)$  can be expressed as a linear combination of the  $t_j$ 's for all  $j \leq i$ .*

*Proof.* Let  $\dim(V) = n$ , then it suffices to prove that every complex-valued  $n \times n$  matrix is similar to some upper-triangular matrix.

The case where  $n = 1$  is trivial because every  $1 \times 1$  matrix is upper-triangular. Suppose that there is some  $k \in \mathbb{Z}^+$  such that every complex-valued  $k \times k$  matrix is similar to some upper-triangular matrix. Let  $\mathbf{M}$  be any  $(k+1) \times (k+1)$  complex-valued matrix. By Proposition 3.1.6, there is some  $\lambda \in \mathbb{C}$  which is an eigenvalue of  $\mathbf{M}$  associated to some non-zero  $\mathbf{v} \in \mathbb{C}^{k+1}$ . Note that we can find  $\mathbf{u}_1, \dots, \mathbf{u}_k \in \mathbb{C}^{k+1}$  such that the columns of

$$\mathbf{N} = \begin{bmatrix} \mathbf{v} & \mathbf{u}_1 & \dots & \mathbf{u}_k \end{bmatrix}$$

is linearly independent. Note that  $\mathbf{M}\mathbf{v} = \lambda\mathbf{v}$ , so

$$\mathbf{N}^{-1}\mathbf{M}\mathbf{N} = \begin{bmatrix} \lambda & \alpha \\ 0 & \mathbf{P} \end{bmatrix},$$

where  $\mathbf{P}$  is a  $k \times k$  complex-valued matrix. By the inductive hypothesis, there exists a matrix  $\mathbf{Q}$  and an upper-triangular matrix  $\mathbf{R}$  such that  $\mathbf{Q}^{-1}\mathbf{P}\mathbf{Q} = \mathbf{R}$ . Consider

$$\begin{bmatrix} 1 & 0 \\ 0 & \mathbf{Q}^{-1} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & \mathbf{P} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & \mathbf{I} \end{bmatrix},$$

so  $\begin{bmatrix} 1 & 0 \\ 0 & \mathbf{Q}^{-1} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & \mathbf{Q} \end{bmatrix}^{-1}$ . Therefore,

$$\begin{bmatrix} 1 & 0 \\ 0 & \mathbf{Q} \end{bmatrix}^{-1} \begin{bmatrix} \lambda & \alpha \\ 0 & \mathbf{P} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & \mathbf{Q} \end{bmatrix} = \begin{bmatrix} \lambda & \alpha\mathbf{Q} \\ 0 & \mathbf{Q}^{-1}\mathbf{P}\mathbf{Q} \end{bmatrix} = \begin{bmatrix} \lambda & \alpha\mathbf{Q} \\ 0 & \mathbf{R} \end{bmatrix},$$

which is upper-triangular. This means that  $\mathbf{M}$  is similar to an upper-triangular matrix. □

Note that the matrix  $\mathbf{N}$  in the proof for Theorem 3.2.3 is basically the matrix representation of some basis mapping  $z$ . Let  $T$  be an operator with matrix representation  $\mathbf{M}$ , then  $z^{-1} \circ T \circ z$

is an operator whose matrix representation is upper-triangular. However, in Section 2.5, we have shown that any operator  $T : V \rightarrow V$  with respect to any basis  $z$  has the same matrix representation as the operator  $T_z^* := z^{-1} \circ T \circ z$  on the coordinate vectors of  $V$ . Therefore, we can indeed write the matrix representation of any operator as an upper-triangular matrix by choosing an appropriate basis  $z$ .

Interestingly, we realise that if  $T$  is made to have an upper-triangular matrix representation by the method described in Theorem 3.2.3, then the  $(1, 1)$  entry of that matrix will be an eigenvalue of  $T$ . The following result is trivial:

#### Proposition 3.2.4 ► Invertibility of Upper Triangular Matrices

*An upper-triangular matrix is invertible if and only if every entry along its diagonal is non-zero.*

To characterise triangularised mapping, we consider the following proposition:

#### Proposition 3.2.5 ► Eigenvalues of Upper Triangular Matrices

*Let  $N$  be an upper-triangular matrix, then the diagonal entries of  $N$  are its eigenvalues.*

*Proof.* Let  $\lambda$  be any of the diagonal entries of  $N$  and consider  $N - \lambda I$  which is also upper-triangular. Note that by Proposition 3.2.4,  $N - \lambda I$  is not invertible, and so there exists some  $v \neq 0$  such that  $(N - \lambda I)v = 0$ . Therefore,  $Nv = \lambda v$  and so  $\lambda$  is an eigenvalue of  $N$ .  $\square$

We have seen in the previous section that every operator has an upper-triangular matrix representation, but can we improve this further?

#### Definition 3.2.6 ► Diagonalisability

An operator  $T$  is said to be **diagonalisable** if there is a basis  $z$  such that  $M_z(T)$  is diagonal.

*Remark.* Another way of stating the definition is that a matrix is diagonalisable if it is similar to a diagonal matrix.

By Proposition 3.2.5, since a diagonal matrix is upper-triangular, its diagonal entries must be the eigenvalues. Denote the  $(i, i)$  entry of the matrix by  $\lambda_i$ , then there is some basic vector  $u_i$  such that  $T(u_i) = \lambda_i u_i$ . Therefore, it is clear that an operator is diagonalisable if and only if there exists a basis consisting of its eigenvectors.

An easy conclusion is that if a vector space  $V$  has  $\dim(V)$  distinct eigenvalues, then it has  $\dim(V)$  linearly independent eigenvectors which form a basis, so any  $V$  with  $\dim(V)$  dis-

tinct eigenvalues is diagonalisable.

### 3.3 Jordan Matrix

Of course, we know that it is not possible for every operator to be diagonalisable, i.e., not all operators have some basis  $z$  such that the operator under  $z$  maps every basic vector to a multiple of itself. Hence, we attempt to make a compromise here: that is, can we always find a basis for every operator such that the operator, under this basis, maps every basic vector to a linear combination of at most 2 distinct basic vectors?

#### Definition 3.3.1 ► Jordan Block

Let  $J$  be an  $m \times m$  matrix such that

$$J_j^i = \begin{cases} 1 & \text{if } i = j - 1 \\ 0 & \text{otherwise} \end{cases}.$$

A **Jordan block** of size  $m$  is the matrix  $\lambda I + J$  for any  $\lambda \in \mathbb{C}$ .

The reason that we consider the Jordan block is that any Jordan block has only one eigenspace which is one-dimensional.

#### Proposition 3.3.2 ► Eigenspace of Jordan Block

Let  $\lambda I + J$  be a Jordan block, then it has a unique eigenspace which is one-dimensional.

*Proof.* By Propositions 3.2.4 and 3.2.5, we know that  $\lambda$  is the only eigenvalue, so  $\lambda I + J$  has only one eigenspace. Let the eigenspace be  $E_\lambda$  and take any eigenvector

$$\mathbf{v} = \begin{bmatrix} x_1 & x_2 & \cdots & x_m \end{bmatrix}^T \in E_\lambda.$$

Note that

$$(\lambda I + J)\mathbf{v} = \begin{bmatrix} \lambda x_1 + x_2 \\ \lambda x_2 + x_3 \\ \vdots \\ \lambda x_{m-1} + x_m \\ \lambda x_m \end{bmatrix} = \begin{bmatrix} \lambda x_1 \\ \lambda x_2 \\ \vdots \\ \lambda x_{m-1} \\ \lambda x_m \end{bmatrix},$$

so clearly  $x_2 = x_3 = \cdots = x_m = 0$ . Therefore,  $E_\lambda$  is spanned by  $\mathbf{e}_1$  and so is one-dimensional. □

We can build up an upper-triangular matrix using Jordan blocks, which by triangularis-

ability gives the following result:

### Definition 3.3.3 ► Jordan Basis

Let  $T : V \rightarrow V$  be an operator. A **Jordan basis** of  $V$  is a basis  $z$  such that

$$M_z(T) = \begin{bmatrix} J_1 & 0 & \cdots & 0 \\ 0 & J_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & J_m \end{bmatrix},$$

where  $J_1, \dots, J_m$  are Jordan blocks.

It can be shown that every matrix is similar to some Jordan matrix, and so every operator can be represented by a Jordan basis.

### Definition 3.3.4 ► Jordan Canonical Form

Let  $j$  be a Jordan basis for some operator  $T$ , then  $M_j(T)$  is called the **Jordan canonical form** of  $T$ .

Jordan matrices have a few interesting applications. We can relate them to the eigenvalues of a matrix by considering the following definition:

### Definition 3.3.5 ► Multiplicity

Let  $J$  be the Jordan form of a matrix  $A$ . For any eigenvalue  $\lambda$  of  $A$ , the **multiplicity** of  $\lambda$  is defined as the sum of the sizes of Jordan blocks corresponding to  $\lambda$ .

In other words, we can interpret the multiplicity of  $\lambda$  as the number of occurrences of it along the diagonal of the Jordan form. Recall that previously we define the *characteristic polynomial* of a matrix  $A$  to be  $\det(A - \lambda I)$ . Now we propose another definition using the Jordan form.

### Definition 3.3.6 ► Characteristic Polynomial

Let  $T$  be a linear operator with eigenvalues  $\lambda_1, \lambda_2, \dots, \lambda_p$ , each with multiplicity  $m_1, m_2, \dots, m_p$  respectively, then the **characteristic polynomial** of  $T$  is defined as

$$\chi_T(x) = \prod_{i=1}^p (x - \lambda_i)^{m_i}.$$

Obviously,  $\chi_T(\lambda_i) = 0$  for all  $i = 1, 2, \dots, p$ , which leads us to the famous theorem below:

**Theorem 3.3.7 ► Cayley-Hamilton Theorem**

For any linear operator  $T$  with characteristic polynomial  $\chi_T$ , we have  $\chi_T(T) = 0$ .

*Proof.* Let  $T : V \rightarrow V$  for some vector space  $V$ . Notice that

$$\chi_T(T) = \prod_{i=1}^p (T - \lambda_i \text{id}_V)^{m_i}$$

where  $\lambda_1, \lambda_2, \dots, \lambda_p$  are the eigenvalues of  $T$  with multiplicity  $m_1, m_2, \dots, m_p$  respectively. Note that under a Jordan basis  $z$ ,

$$M_z(T - \lambda_i \text{id}_V) = \begin{bmatrix} 0 & 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & 1 & \cdots & 0 & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 \end{bmatrix}.$$

By direct calculation, one may check that  $M_z(T - \lambda_i \text{id}_V)^{m_i} = \mathbf{0}$ , and so  $(T - \lambda_i \text{id}_V)^{m_i}$  is the zero transformation for all  $i = 1, 2, \dots, p$ . Therefore,  $\chi_T(T) = 0$ .  $\square$

An important implication of Theorem 3.3.7 is that the operator obtained from composing some operator  $T$  on  $V$  with itself for  $n$  times can be “generated” with lower powers of  $T$ . Essentially, we have

$$\chi_T(T) = \sum_{i=0}^k c_i T^i = 0,$$

where  $T^0$  is just the identity transformation  $\text{id}_V$  and  $c_i$ 's are constants. However, this means that

$$T^k = -\frac{1}{c_k} \sum_{i=0}^{k-1} c_i T^i.$$

Clearly, from now on, for any integer  $n \geq k$ , we can express  $T^n$  as a linear combination of the transformations in  $\{\text{id}_V, T, T^2, \dots, T^{k-1}\}$ . In fact, we can work out  $T^n$  for all  $n \in \mathbb{Z}^-$  as well because

$$T^{-1} = T^{-1} \circ \text{id}_V = T^{-1} \circ \left( -\frac{1}{c_0} \sum_{i=1}^k c_i T^i \right).$$

### 3.4 Operators in Euclidean Spaces

Let  $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$  be an operator in some Euclidean space. If  $T$  is diagonalisable, then it has a basis  $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n\}$  consisting of distinct eigenvectors. Let  $\lambda_i$  be the eigenvalue associated to  $\mathbf{v}_i$ , then for any  $\mathbf{u} \in \mathbb{R}^n$ , we have

$$\begin{aligned} T(\mathbf{u}) &= T\left(\sum_{i=1}^n c_i \mathbf{v}_i\right) \\ &= \sum_{i=1}^n c_i \lambda_i \mathbf{v}_i. \end{aligned}$$

If the  $\lambda_i$ 's are all real, then for each  $\lambda_i \neq 0$ ,  $T$  will stretch  $\mathbf{u}$  by a factor of  $\lambda_i$  in the component  $\mathbf{v}_i$ . If some  $\lambda_j = 0$ , then the component  $\mathbf{v}_j$  is annihilated, i.e., the image of  $\mathbf{u}$  will lose one dimension in the component of  $\mathbf{v}_j$ . We call such transformations as a *crushing* transformation.

Before we discuss the other cases, first let us examine the following property of eigenvalues:

**Proposition 3.4.1 ► Eigenvalues Are in Conjugate Pairs**

*If  $\lambda \in \mathbb{C}$  is an eigenvalue of some operator  $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$ , then the complex conjugate  $\bar{\lambda}$  is also an eigenvalue of  $T$ .*

The proof of the proposition is immediate by considering the trace of the matrix representation for  $T$ , and is left to the reader as an exercise. Suppose that the operator  $T$  has all non-real eigenvalues  $\lambda_1, \bar{\lambda}_1, \lambda_2, \bar{\lambda}_2, \dots, \lambda_k, \bar{\lambda}_k$ , where  $\lambda_i$  and  $\bar{\lambda}_i$  are conjugate pairs. Write  $\lambda_i = r_i e^{i\theta_i}$ . Define

$$\mathbf{R}_i := \begin{bmatrix} \cos \theta_i & -\sin \theta_i \\ \sin \theta_i & \cos \theta_i \end{bmatrix}, \quad \mathbf{A} := \begin{bmatrix} 1 & -1 \\ -i & -i \end{bmatrix}$$

to be the matrix representation of the rotation transformation and its eigenvector matrix.

Denote

$$M_z(T) = \begin{bmatrix} \lambda_1 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & \bar{\lambda}_1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \lambda_2 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & \bar{\lambda}_k & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & \lambda_k & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & \bar{\lambda}_k \end{bmatrix},$$

then clearly,

$$\begin{aligned} M_z(T) &= \begin{bmatrix} r_1 & 0 & \cdots & 0 \\ 0 & r_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & r_k \end{bmatrix} \begin{bmatrix} \Lambda & 0 & \cdots & 0 \\ 0 & \Lambda & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \Lambda \end{bmatrix} \begin{bmatrix} \mathbf{R}_1 & 0 & \cdots & 0 \\ 0 & \mathbf{R}_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{R}_k \end{bmatrix} \begin{bmatrix} \Lambda & 0 & \cdots & 0 \\ 0 & \Lambda & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \Lambda \end{bmatrix}^{-1} \\ &= \begin{bmatrix} \Lambda & 0 & \cdots & 0 \\ 0 & \Lambda & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \Lambda \end{bmatrix} \begin{bmatrix} r_1 & 0 & \cdots & 0 \\ 0 & r_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & r_k \end{bmatrix} \begin{bmatrix} \mathbf{R}_1 & 0 & \cdots & 0 \\ 0 & \mathbf{R}_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{R}_k \end{bmatrix} \begin{bmatrix} \Lambda & 0 & \cdots & 0 \\ 0 & \Lambda & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \Lambda \end{bmatrix}^{-1}. \end{aligned}$$

Therefore, there exists some basis such that the matrix representation of  $T$  is

$$\begin{bmatrix} r_1 & 0 & \cdots & 0 \\ 0 & r_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & r_k \end{bmatrix} \begin{bmatrix} \mathbf{R}_1 & 0 & \cdots & 0 \\ 0 & \mathbf{R}_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{R}_k \end{bmatrix}.$$

Therefore, for a vector  $(a_1, a_2, \dots, a_n)^T$ , its image under  $T$  is equivalent to rotating the vector in its  $i$ - and  $(i+1)$ -th dimensions by  $\mathbf{R}_i$ , and stretching the  $i$ -th component by a factor of  $r_i$  (or crushing the component if  $r_i = 0$ ).

However, if  $T$  is not diagonalisable, we need to consider its Jordan canonical form. Let  $\lambda_i$



be any eigenvalue of  $T$  and consider its corresponding Jordan block

$$J_i := \begin{bmatrix} \lambda_i & 1 & 0 & \cdots & 0 & 0 \\ 0 & \lambda_i & 1 & \cdots & 0 & 0 \\ 0 & 0 & \lambda_i & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda_i & 1 \\ 0 & 0 & 0 & \cdots & 0 & \lambda_i \end{bmatrix} = \begin{bmatrix} \lambda_i & 0 & 0 & \cdots & 0 & 0 \\ 0 & \lambda_i & 0 & \cdots & 0 & 0 \\ 0 & 0 & \lambda_i & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda_i & 0 \\ 0 & 0 & 0 & \cdots & 0 & \lambda_i \end{bmatrix} \begin{bmatrix} 1 & \frac{1}{\lambda_i} & 0 & \cdots & 0 & 0 \\ 0 & 1 & \frac{1}{\lambda_i} & \cdots & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & \frac{1}{\lambda_i} \\ 0 & 0 & 0 & \cdots & 0 & 1 \end{bmatrix}.$$

Notice that we can set  $\frac{1}{\lambda_i} = \tan(\phi_i)$  for some  $\phi \in \mathbb{C}$ . In the case where  $\lambda_i \in \mathbb{R}$ , we impose a further condition requiring  $\phi \in [-\frac{\pi}{2}, \frac{\pi}{2}]$ . It is not hard to see that each Jordan block corresponds to a transformation obtained by shearing in every two adjacent components, followed by a combination of rotations and stretching or crushing as in the previous case.

To summarise, any operator  $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$  is

- a series of stretching and/or crushing transformations if it has  $n$  distinct real eigenvalues;
- a series of rotations followed by a series of stretching and/or crushing transformations if it has  $n$  distinct eigenvalues not all real;
- a series of shearing transformations followed by a series of stretching and/or crushing transformations if it has less than  $n$  distinct real eigenvalues;
- a series of shearing transformations, followed by a series of rotations, and then followed by another series of stretching and/or crushing transformations if it has less than  $n$  distinct eigenvalues not all real.

# Inner Product Spaces

## 4.1 Bilinear Forms

In the previous chapter, we have made an attempt to classify all operators over the Euclidean spaces. In this chapter, we further discuss geometric properties of operators. Now, this task would be quite trivial if we restrict our scope to the Euclidean spaces only, because we already have many well-defined notions such as length, area, angle, volume, etc., in Euclidean geometry. However, our ultimate aim is to generalise such geometry to any kind of vector spaces. We will first take the *dot product* as an example to see how we can apply an axiomatic approach to abstract Euclidean geometry for general vector spaces.

### Definition 4.1.1 ► Dot Product

Let  $\mathbb{F}$  be a field, the **dot product** is a binary operation  $\cdot : \mathbb{F}^n \times \mathbb{F}^n \rightarrow \mathbb{F}$  such that

1.  $\mathbf{u} \cdot \mathbf{u} \geq 0$  for all  $\mathbf{u} \in \mathbb{F}^n$ ;
2.  $\mathbf{u} \cdot \mathbf{u} = 0$  if and only if  $\mathbf{u} = \mathbf{0}$ ;
3.  $\mathbf{u} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{u}$  for all  $\mathbf{u}, \mathbf{v} \in \mathbb{F}^n$ ;
4.  $(\alpha \mathbf{u}_1 + \beta \mathbf{u}_2) \cdot \mathbf{v} = \alpha(\mathbf{u}_1 \cdot \mathbf{v}) + \beta(\mathbf{u}_2 \cdot \mathbf{v})$  and  $\mathbf{u} \cdot (\alpha \mathbf{v}_1 + \beta \mathbf{v}_2) = \alpha(\mathbf{u} \cdot \mathbf{v}_1) + \beta(\mathbf{u} \cdot \mathbf{v}_2)$  for all  $\mathbf{u}, \mathbf{v} \in \mathbb{F}^n$  and  $\alpha, \beta \in \mathbb{F}$ .

The dot product is an example of what is known as *bilinear maps*.

### Definition 4.1.2 ► Bilinear Map

**bilinear map** is a mapping

$$b : U \times V \rightarrow \mathbb{F}$$

where  $U, V, W$  are vector spaces over  $\mathbb{F}$  such that

$$b(\alpha \mathbf{u}_1 + \beta \mathbf{u}_2, \mathbf{v}) = \alpha b(\mathbf{u}_1, \mathbf{v}) + \beta b(\mathbf{u}_2, \mathbf{v}),$$

$$b(\mathbf{u}, \alpha \mathbf{v}_1 + \beta \mathbf{v}_2) = \alpha b(\mathbf{u}, \mathbf{v}_1) + \beta b(\mathbf{u}, \mathbf{v}_2)$$

for all  $\mathbf{u} \in U, \mathbf{v} \in V$  and  $\alpha, \beta \in \mathbb{F}$ .

In fact, the dot product is a special type of bilinear maps as its domain is the cartesian product of a vector space with itself, and its co-domain is a field. Such maps are known as

bilinear forms.

#### Definition 4.1.3 ► Bilinear Form

A **bilinear form** on a finite-dimensional space  $V$  over  $\mathbb{F}$  is a bilinear map

$$b : V \times V \rightarrow \mathbb{F}$$

Let  $\mathbb{B}(V)$  denote the set of all bilinear forms over a finite-dimensional vector space  $V$ , then  $\mathbb{B}(V)$  itself is also a vector space with the usual definitions of addition and scalar multiplication over mappings.

Using bilinear maps, we shall state the real definition of tensor products.

#### Definition 4.1.4 ► Tensor Product

Let  $V$  and  $W$  be vector spaces. The **tensor product** of  $V$  and  $W$  is a vector space  $V \otimes W$  with a bilinear map  $\otimes : V \times W \rightarrow V \otimes W$  such that for every bilinear map  $h : V \times W \rightarrow Z$ , there is a unique map  $\tilde{h} : V \otimes W \rightarrow Z$  with  $h = \tilde{h} \circ \otimes$ .

Now looking back at our Definition 2.3.1, we consider the vector space  $V \otimes \mathfrak{V}$ . For every  $\mathbf{u} \in V$ , define  $h_{\mathbf{u}} : V \times \mathfrak{V} \rightarrow V$  to be a bilinear map such that  $h_{\mathbf{u}}(\mathbf{v}, \alpha) = \alpha(\mathbf{v})\mathbf{u}$ . By Definition 4.1.4, there is a unique map  $\tilde{h}_{\mathbf{u}} : V \otimes \mathfrak{V} \rightarrow V$  such that

$$\tilde{h}_{\mathbf{u}}(\mathbf{v} \otimes \alpha) = h_{\mathbf{u}}(\mathbf{v}, \alpha) = \alpha(\mathbf{v})\mathbf{u}.$$

Clearly, by taking  $\mathbf{v} \otimes \alpha \in \mathcal{L}(V, V)$  and define  $\tilde{h}_{\mathbf{u}}(\mathbf{v} \otimes \alpha) = (\mathbf{v} \otimes \alpha)(\mathbf{u})$ , the only possible way for the definition to stay consistent is  $(\mathbf{v} \otimes \alpha)(\mathbf{u}) = \alpha(\mathbf{v})\mathbf{u}$ .

#### Definition 4.1.5 ► Tensor Product between Dual Vectors

Let  $V$  be a finite-dimensional vector space. For any  $\alpha, \beta \in \mathfrak{V}$ , their **tensor product** is defined as the mapping

$$\alpha \otimes \beta : V \times V \rightarrow \mathbb{F}$$

such that  $(\alpha \otimes \beta)(\mathbf{u}, \mathbf{v}) = \alpha(\mathbf{u})\beta(\mathbf{v})$ .

We shall also check that this definition stays consistent with the universal property in Definition 4.1.4. For each  $\mathbf{u}, \mathbf{v} \in V$ , define  $h_{\mathbf{u}, \mathbf{v}} : \mathfrak{V} \times \mathfrak{V} \rightarrow \mathbb{F}$  by  $h_{\mathbf{u}, \mathbf{v}}(\alpha, \beta) = \alpha(\mathbf{u})\beta(\mathbf{v})$ . Now, there is a unique map  $\tilde{h}_{\alpha, \beta} : \mathfrak{V} \otimes \mathfrak{V} \rightarrow \mathbb{F}$  such that

$$\tilde{h}_{\mathbf{u}, \mathbf{v}}(\alpha \otimes \beta) = h_{\mathbf{u}, \mathbf{v}}(\alpha, \beta) = \alpha(\mathbf{u})\beta(\mathbf{v}).$$

If we take  $\otimes : \mathfrak{V} \times \mathfrak{V} \rightarrow \mathcal{L}(V^2, \mathbb{F})$ , clearly  $\tilde{h}_{\mathbf{u}, \mathbf{v}}(\alpha \otimes \beta) = (\alpha \otimes \beta)(\mathbf{u}, \mathbf{v}) = \alpha(\mathbf{u})\beta(\mathbf{v})$ .

It is easy to see that for any basic vectors  $\zeta^i, \zeta^j$  for  $\mathcal{B}$ , the mapping  $\zeta^i \otimes \zeta^j$  is a bilinear form over  $V$ .

#### Proposition 4.1.6 ► Basis of the Bilinear Form Space

Let  $\mathbb{B}(V)$  be the set of all bilinear forms over a vector space  $V$  with a basis  $\{\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_n\}$  and a dual basis  $\{\zeta^1, \zeta^2, \dots, \zeta^n\}$ , then the set

$$\{\zeta^i \otimes \zeta^j : i \neq j\}$$

is a basis for  $\mathbb{B}(V)$ .

*Proof.* Take any  $b \in \mathbb{B}(V)$ . For any  $\mathbf{u}, \mathbf{v} \in V$ , we have

$$\begin{aligned} b(\mathbf{u}, \mathbf{v}) &= b\left(\sum_{i=1}^n u_i \mathbf{z}_i, \sum_{i=1}^n v_i \mathbf{z}_i\right) \\ &= \sum_{i=1}^n u_i b\left(\mathbf{z}_i, \sum_{j=1}^n v_j \mathbf{z}_j\right) \\ &= \sum_{i=1}^n \sum_{j=1}^n u_i v_j b(\mathbf{z}_i, \mathbf{z}_j). \end{aligned}$$

Let  $b(\mathbf{z}_i, \mathbf{z}_j) = b_{ij}$  and define

$$h = \sum_{i=1}^n \sum_{j=1}^n b_{ij} \zeta^i \otimes \zeta^j$$

to be a bilinear form. Notice that

$$h(\mathbf{z}_p, \mathbf{z}_q) = \sum_{i=1}^n \sum_{j=1}^n b_{ij} \zeta^i(\mathbf{z}_p) \zeta^j(\mathbf{z}_q) = b_{pq} = b(\mathbf{z}_p, \mathbf{z}_q),$$

for any  $\mathbf{z}_p, \mathbf{z}_q$ , so  $h = b$ , which implies that  $\mathbb{B}(V) = \text{span}\{\zeta^i \otimes \zeta^j : i \neq j\}$ . One may easily check that  $h$  is the zero mapping if and only if  $b_{ij} = 0$  for all  $i, j$ , so  $\{\zeta^i \otimes \zeta^j : i \neq j\}$  is a basis for  $\mathbb{B}(V)$ .  $\square$

In this sense,  $\mathbb{B}(V) = \mathcal{L}(V^2, \mathbb{F})$ . However, intuitively for each mapping  $(\mathbf{u}, \mathbf{v}) \mapsto c$ , this is equivalent to finding some  $T_{\mathbf{u}} \in \mathcal{B}$  for each  $\mathbf{u} \in V$  and apply the mapping to  $\mathbf{v}$ .

**Proposition 4.1.7 ▶ Every Bilinear Maps to a Unique Dual Vector**

For any finite-dimensional vector space  $V$ , we have  $\mathbb{B}(V) \cong \mathcal{L}(V, \mathbb{F})$ .

*Proof.* Take any  $b \in \mathbb{B}(V)$ . Define a mapping  $f: \mathbb{B}(V) \rightarrow \mathcal{L}(V, \mathbb{F})$  by  $f(b) = b^\#$ , where  $b^\#(\mathbf{v})(\mathbf{u}) = b(\mathbf{u}, \mathbf{v})$ . Suppose there are  $b_1, b_2 \in \mathbb{B}(V)$  with  $b_1 \neq b_2$ , then there exists  $\mathbf{u}, \mathbf{v} \in V$  such that

$$b_1^\#(\mathbf{v})(\mathbf{u}) = b_1(\mathbf{u}, \mathbf{v}) \neq b_2(\mathbf{u}, \mathbf{v}) = b_2^\#(\mathbf{v})(\mathbf{u}),$$

which implies that  $b_1^\#(\mathbf{v}) \neq b_2^\#(\mathbf{v})$ , and so  $b_1 \neq b_2$ . Therefore, the  $f$  is injective. Now, take any  $b^\# \in \mathcal{L}(V, \mathbb{F})$ , let  $b$  be a bilinear form with  $b(\mathbf{u}, \mathbf{v}) = b^\#(\mathbf{v})(\mathbf{u})$ , then by definition clearly  $f(b) = b^\#$ . Therefore,  $f$  is surjective, and so  $\mathbb{B}(V) \cong \mathcal{L}(V, \mathbb{F})$ .  $\square$

We can write

$$b^\#(\mathbf{v}) = \alpha = \sum_{i=1}^n p_i \zeta^i$$

in the above proof where the  $\zeta^i$ 's are the dual basis. Then, we know that

$$p_i = \alpha(\mathbf{z}_i) = b(\mathbf{z}_i, \mathbf{v}) = \sum_{j=1}^n b_{ij} v_j$$

where  $\mathbf{v} = \sum_{j=1}^n v_j \mathbf{z}_j$ . This means that we can use the bilinear form  $b$  to transform the components of  $\mathbf{v}$  into the components of  $b^\#(\mathbf{v})$ .

Let  $b$  be any bilinear form, then just like unary transformations,  $b$  has a matrix representation which we can transform by a change of basis. Let  $\{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n\}$  and  $\{\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_n\}$  be 2 different bases, then by Proposition 2.5.1, there exists a non-singular matrix  $\mathbf{P}$  such that  $\mathbf{y}_i = \sum_{j=1}^n P_i^j \mathbf{z}_j$  for every  $i = 1, 2, \dots, n$ . Therefore,

$$\begin{aligned} M_y(b)_j^i &= b(\mathbf{y}_i, \mathbf{y}_j) \\ &= b\left(\sum_{p=1}^n P_i^p \mathbf{z}_p, \sum_{q=1}^n P_j^q \mathbf{z}_q\right) \\ &= \sum_{p=1}^n \sum_{q=1}^n P_i^p b(\mathbf{z}_p, \mathbf{z}_q) P_j^q \\ &= \sum_{p=1}^n \sum_{q=1}^n P_i^p M_z(b)_q^p P_j^q. \end{aligned}$$

The above formula translates to  $M_y(b) = \mathbf{P}^T M_z(b) \mathbf{P}$ .

## 4.2 Geometry in Vector Spaces

Observe that the structure of a bilinear form over some vector space  $V$  matches exactly that of a “distance” between points in a space — it sends a pair of vectors to some scalar. This motivates us to generalise the dot product into the following form:

### Definition 4.2.1 ► Inner Product Space

An **inner product space** is a finite-dimensional vector space  $V$  over  $\mathbb{F}$ , where  $\mathbb{F}$  is a well-ordered field, augmented with a bilinear form  $g : V \times V \rightarrow \mathbb{F}$ , known as an **inner product** on  $V$ , such that

1.  $g(\mathbf{u}, \mathbf{u}) \geq 0$  for all  $\mathbf{u} \in V$ ;
2.  $g(\mathbf{u}, \mathbf{u}) = 0$  if and only if  $\mathbf{u} = \mathbf{0}$ ;
3.  $g(\mathbf{u}, \mathbf{v}) = g(\mathbf{v}, \mathbf{u})$  for all  $\mathbf{u}, \mathbf{v} \in V$ .

*Remark.* The inner product is not unique because any positive multiple of an inner product is still an inner product.

Obviously, the set of all inner products over some vector space  $V$  is not a subspace because the zero bilinear form is not an inner product.

The matrix representation of an inner product is said to be *symmetric*, *positive* and *definite*. Indeed, we can easily check that it is a positive definite matrix which is symmetric.

Recall that in  $\mathbb{R}^n$ , we define the *length* of a vector  $\mathbf{x}$  by  $\|\mathbf{x}\| = \sqrt{\mathbf{x} \cdot \mathbf{x}}$ . With Definition 4.2.1, we are able to generalise length to any vector space where an inner product is defined.

### Definition 4.2.2 ► Length

Let  $(V, g)$  be an inner product space, then for any  $\mathbf{v} \in V$ , its **length** is defined as

$$|\mathbf{v}| = \sqrt{g(\mathbf{v}, \mathbf{v})}.$$

Similarly, we can generalise orthogonality in Euclidean spaces to general vector spaces.

### Definition 4.2.3 ► Orthogonality

Let  $(V, g)$  be an inner product space.  $\mathbf{u}, \mathbf{v} \in V$  are **orthogonal** if  $g(\mathbf{u}, \mathbf{v}) = 0$ .

$\mathbf{0}$  is apparently the only vector orthogonal to itself. Furthermore,

$$|\mathbf{u} + \mathbf{v}|^2 = g(\mathbf{u} + \mathbf{v}, \mathbf{u} + \mathbf{v}) = g(\mathbf{u}, \mathbf{u}) + g(\mathbf{u}, \mathbf{v}) + g(\mathbf{v}, \mathbf{u}) + g(\mathbf{v}, \mathbf{v}),$$

so of course  $|\mathbf{u} + \mathbf{v}|^2 = |\mathbf{u}|^2 + |\mathbf{v}|^2$  if and only if  $\mathbf{u}$  and  $\mathbf{v}$  are orthogonal. Obviously this is a generalisation of the famous Pythagoras' Theorem. Both of the facts above align with the case in Euclidean spaces. Additionally, we can do *orthogonal decomposition* just like in Euclidean spaces as well!

#### Theorem 4.2.4 ► Orthogonal Decomposition

Let  $(V, g)$  be an inner product space and  $\mathbf{v} \in V$  be non-zero, then for any vector  $\mathbf{u} \in V$ , there exist some  $\mathbf{w}_1, \mathbf{w}_2 \in V$  with  $\mathbf{w}_1 \parallel \mathbf{v}$  and  $\mathbf{w}_2 \perp \mathbf{v}$  such that  $\mathbf{u} = \mathbf{w}_1 + \mathbf{w}_2$ .

*Proof.* Take

$$\mathbf{w}_1 = \frac{g(\mathbf{u}, \mathbf{v})}{|\mathbf{v}|^2} \mathbf{v}, \quad \mathbf{w}_2 = \mathbf{u} - \frac{g(\mathbf{u}, \mathbf{v})}{|\mathbf{v}|^2} \mathbf{v}.$$

Clearly,  $\mathbf{w}_1 \parallel \mathbf{v}$  and  $\mathbf{u} = \mathbf{w}_1 + \mathbf{w}_2$ . Observe that

$$g\left(\mathbf{u} - \frac{g(\mathbf{u}, \mathbf{v})}{|\mathbf{v}|^2} \mathbf{v}, \mathbf{v}\right) = g(\mathbf{u}, \mathbf{v}) - \frac{g(\mathbf{u}, \mathbf{v})}{|\mathbf{v}|^2} g(\mathbf{v}, \mathbf{v}) = 0,$$

so  $\mathbf{w}_2 \perp \mathbf{v}$ . □

We can also make use of an inner product to prove some interesting results, e.g., the *Cauchy-Schwarz inequality*.

#### Theorem 4.2.5 ► Cauchy-Schwarz Inequality

Let  $(V, g)$  be an inner product space, then  $g(\mathbf{u}, \mathbf{v}) \leq |\mathbf{u}||\mathbf{v}|$  for any  $\mathbf{u}, \mathbf{v} \in V$ , where the equality is attained if and only if  $\mathbf{u} = \lambda \mathbf{v}$  for some  $\lambda \in \mathbb{F}$ .

*Proof.* Notice that  $g(\mathbf{u}, \mathbf{0}) = 2g(\mathbf{u}, \mathbf{0})$ , so it suffices to prove for  $\mathbf{v} \neq \mathbf{0}$ . By Theorem 4.2.4, we have

$$\begin{aligned} |\mathbf{u}|^2 &= \left( \frac{g(\mathbf{u}, \mathbf{v})}{|\mathbf{v}|^2} \mathbf{v} \right)^2 + \left( \mathbf{u} - \frac{g(\mathbf{u}, \mathbf{v})}{|\mathbf{v}|^2} \mathbf{v} \right)^2 \\ &\geq \left( \frac{g(\mathbf{u}, \mathbf{v})}{|\mathbf{v}|^2} \mathbf{v} \right)^2 \\ &= g\left( \frac{g(\mathbf{u}, \mathbf{v})}{|\mathbf{v}|^2} \mathbf{v}, \frac{g(\mathbf{u}, \mathbf{v})}{|\mathbf{v}|^2} \mathbf{v} \right) \\ &= \left( \frac{g(\mathbf{u}, \mathbf{v})}{|\mathbf{v}|^2} \right)^2 g(\mathbf{v}, \mathbf{v}) \\ &= \frac{g(\mathbf{u}, \mathbf{v})^2}{|\mathbf{v}|^2}. \end{aligned}$$

Taking square root on both sides yields  $g(\mathbf{u}, \mathbf{v}) \leq |\mathbf{u}||\mathbf{v}|$ . Note that the equality is attained if and only if

$$\left(\mathbf{u} - \frac{g(\mathbf{u}, \mathbf{v})}{|\mathbf{v}|^2} \mathbf{v}\right)^2 = 0,$$

which implies that  $\mathbf{u} \parallel \mathbf{v}$ . □

Notice that if a metric space is also a vector space, then we get an inner product space for free by using the metric as an inner product. The metric satisfies something known as the *triangle inequality*, which can be proved as an immediate consequence of Theorem 4.2.5.

#### Corollary 4.2.6 ▶ Triangle Inequality

Let  $V$  be a finite-dimensional vector space, then for any  $\mathbf{u}, \mathbf{v} \in V$ ,  $|\mathbf{u} + \mathbf{v}| \leq |\mathbf{u}| + |\mathbf{v}|$ .

*Proof.* Notice that

$$\begin{aligned} |\mathbf{u} + \mathbf{v}|^2 &= g(\mathbf{u} + \mathbf{v}, \mathbf{u} + \mathbf{v}) \\ &= g(\mathbf{u}, \mathbf{u}) + g(\mathbf{u}, \mathbf{v}) + g(\mathbf{v}, \mathbf{u}) + g(\mathbf{v}, \mathbf{v}) \\ &= |\mathbf{u}|^2 + 2g(\mathbf{u}, \mathbf{v}) + |\mathbf{v}|^2. \end{aligned}$$

By Theorem 4.2.5,  $g(\mathbf{u}, \mathbf{v}) \leq |\mathbf{u}||\mathbf{v}|$ , so  $|\mathbf{u} + \mathbf{v}|^2 \leq |\mathbf{u}|^2 + 2|\mathbf{u}||\mathbf{v}| + |\mathbf{v}|^2$ . Taking square root on both sides yields  $|\mathbf{u} + \mathbf{v}| \leq |\mathbf{u}| + |\mathbf{v}|$ . □

*Remark.* Clearly, equality is attained in the triangle inequality if and only if  $\mathbf{u} \parallel \mathbf{v}$  as well due to the use of Cauchy-Schwarz inequality.

Recall that in the Euclidean space, a point is located by an origin, an angle and a distance from the origin. In general inner product spaces, we can easily take the zero vector as the origin, and define the distance of a vector  $\mathbf{x}$  to the origin as  $|\mathbf{x}|$ . The only thing remaining to be defined is the notion of an *angle*.

#### Definition 4.2.7 ▶ Angle

Let  $(V, g)$  be an inner product space and  $\mathbf{u}, \mathbf{v} \in V$  be any vectors. The **angle** between  $\mathbf{u}$  and  $\mathbf{v}$  is defined as a scalar  $\theta \in [0, \pi]$  such that

$$\cos \theta = \frac{g(\mathbf{u}, \mathbf{v})}{|\mathbf{u}||\mathbf{v}|}.$$

Note that this definition is only solid with Theorem 4.2.5 established, which ensures that  $\cos \theta \in [-1, 1]$  here.



## 4.3 Orthogonalisation

Recall that in  $\mathbb{R}^n$ , one can construct a coordinate system using a set of  $n$  pair-wise orthogonal unit vectors as a basis. The following definition helps generalise this to any vector space:

### Definition 4.3.1 ► Orthonormal Basis

Let  $(V, g)$  be an inner product space. A basis  $z$  of  $V$  is said to be an **orthonormal basis** if for any  $\mathbf{u}, \mathbf{v} \in V$ , we have  $g(\mathbf{u}, \mathbf{v}) = z^{-1}(\mathbf{u}) \cdot z^{-1}(\mathbf{v})$ .  $z$  is said to be an **isometry**.

*Remark.* In general, for any inner product spaces  $(V, g)$  and  $(U, h)$ , an isometry is a vector space isomorphism  $f : V \rightarrow U$  that preserves the inner product, i.e.,

$$h(f(\mathbf{x}), f(\mathbf{y})) = g(\mathbf{x}, \mathbf{y})$$

for all  $\mathbf{x}, \mathbf{y} \in V$ .

We can easily observe that an orthonormal basis  $z$  preserves the inner product between the inner product space  $(V, g)$  and  $(\mathbb{F}^n, \cdot)$ , so lengths and angles between any vectors in  $V$  are the same as the lengths and angles between their respective coordinate vectors in  $\mathbb{F}^n$ . Clearly, there are always infinitely many orthonormal bases for any inner product space.

Observe that by writing the basis  $z$  as  $\{\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_n\}$ , we have

$$g(\mathbf{z}_i, \mathbf{z}_j) = \mathbf{e}_i \cdot \mathbf{e}_j = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}.$$

This justifies that the basic vectors in an orthonormal basis are indeed pair-wise orthogonal unit vectors. Using Proposition 4.1.6, it is easy to check that under any orthonormal basis, the matrix representation of the inner product is the identity matrix. Therefore, if  $\mathbf{P}$  is a change-of-basis matrix between any orthonormal bases,

$$\mathbf{P}^T \mathbf{P} = \mathbf{P}^T \mathbf{I} \mathbf{P} = \mathbf{I},$$

which means that  $\mathbf{P}^{-1} = \mathbf{P}^T$ .

### Definition 4.3.2 ► Orthogonal Matrix

A matrix  $\mathbf{P}$  is said to be an **orthogonal matrix** if  $\mathbf{P}^T \mathbf{P} = \mathbf{I}$ .

The name “orthogonal” is appropriately chosen because if we let  $\mathbf{p}_i$  to be the  $i$ -th column

of an orthogonal matrix  $\mathbf{P}$ , Definition 4.3.2 implies that

$$\mathbf{p}_i^T \mathbf{p}_j = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}.$$

Therefore, the columns of  $\mathbf{P}$  are pair-wise orthonormal. However, this clearly means that the columns of every orthogonal matrix form an orthonormal basis on their own! By such change of basis, it is also obvious that the product of orthogonal matrices is still an orthogonal matrix, which we can verify algebraically easily. An advantage of an orthonormal basis is obvious:

### Proposition 4.3.3 ► Components under Orthonormal Bases

If  $\{\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_n\}$  is an orthonormal basis for an inner product space  $V$ , then for any vector  $\mathbf{v} := \sum_{i=1}^n a_i \mathbf{z}_i \in V$ ,

$$|\mathbf{v}|^2 = \sum_{i=1}^n a_i^2.$$

The above is immediate from Pythagoras' Theorem. A trivial corollary is:

### Corollary 4.3.4 ► Linear Independence of Orthonormal Vectors

Any set of pair-wise orthonormal vectors is linearly independent.

Of course, this implies that we can construct an orthonormal basis from an orthonormal spanning set easily because we get linear independence for free.

### Corollary 4.3.5 ► Construction of Orthonormal Bases

For any inner product space  $V$ , any orthonormal set of  $\dim(V)$  vectors in  $V$  is an orthonormal basis for  $V$ .

While any orthonormal spanning set is automatically an orthonormal basis, things are not so easy when we do not have an orthonormal set in the first place. In such situations, we make the following attempt to convert a non-orthonormal basis to an orthonormal one: first, let  $\{\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_n\}$  be any basis for an inner product space  $(V, g)$ . Consider

$$\mathbf{z}_1^+ := \frac{\mathbf{z}_1}{|\mathbf{z}_1|},$$

which is a unit vector. Notice that by applying Theorem 4.2.4, we can decompose  $\mathbf{z}_2$  into

$$\mathbf{z}_2 = \frac{g(\mathbf{z}_2, \mathbf{z}_1^+)}{|\mathbf{z}_1^+|^2} \mathbf{z}_1^+ + \left( \mathbf{z}_2 - \frac{g(\mathbf{z}_2, \mathbf{z}_1^+)}{|\mathbf{z}_1^+|^2} \mathbf{z}_1^+ \right) = g(\mathbf{z}_2, \mathbf{z}_1^+) \mathbf{z}_1^+ + \left( \mathbf{z}_2 - g(\mathbf{z}_2, \mathbf{z}_1^+) \mathbf{z}_1^+ \right).$$

Since  $\mathbf{z}_2 \perp \mathbf{z}_1$ , the component  $\mathbf{z}_2 - g(\mathbf{z}_2, \mathbf{z}_1^+) \mathbf{z}_1^+$  orthogonal to  $\mathbf{z}_1^+$  is non-zero, and so we can normalise it by

$$\mathbf{z}_2^+ := \frac{\mathbf{z}_2 - g(\mathbf{z}_2, \mathbf{z}_1^+) \mathbf{z}_1^+}{|\mathbf{z}_2 - g(\mathbf{z}_2, \mathbf{z}_1^+) \mathbf{z}_1^+|}$$

to obtain a unit vector orthogonal to  $\mathbf{z}_1^+$ . Applying the above steps repeatedly will convert all  $n$  vectors in the original basis into  $n$  pair-wise orthogonal unit vectors. This is justified because we can recursively applying orthogonal decomposition to remove the components in  $\mathbf{z}_k$  parallel to each of  $\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_{k-1}$  and obtain

$$\mathbf{z}_k^+ := \frac{\mathbf{z}_k - \sum_{i=1}^{k-1} g(\mathbf{z}_k, \mathbf{z}_i^+) \mathbf{z}_i^+}{|\mathbf{z}_k - \sum_{i=1}^{k-1} g(\mathbf{z}_k, \mathbf{z}_i^+) \mathbf{z}_i^+|}.$$

#### Technique 4.3.6 ▶ Gram-Schmidt Orthogonalisation

For any inner product space  $(V, g)$  with a basis  $\{\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_n\}$ , an orthonormal basis  $\{\mathbf{z}_1^+, \mathbf{z}_2^+, \dots, \mathbf{z}_n^+\}$  can be produced by

$$\mathbf{z}_1^+ = \mathbf{z}_1, \quad \mathbf{z}_k^+ := \frac{\mathbf{z}_k - \sum_{i=1}^{k-1} g(\mathbf{z}_k, \mathbf{z}_i^+) \mathbf{z}_i^+}{|\mathbf{z}_k - \sum_{i=1}^{k-1} g(\mathbf{z}_k, \mathbf{z}_i^+) \mathbf{z}_i^+|}.$$

Such a basis is known as a **Gram-Schmidt basis**.

In fact, let  $\mathcal{B}(V)$  be the set of all bases and  $\mathcal{B}_\perp(V)$  be the set of all orthonormal bases of some inner product space  $(V, g)$ , we can view Gram-Schmidt orthogonalisation process as a mapping  $g : \mathcal{B}(V) \rightarrow \mathcal{B}_\perp(V)$  defined by

$$g(\{\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_n\}) = \left\{ \mathbf{z}_k^+ \in V : \mathbf{z}_k^+ := \frac{\mathbf{z}_k - \sum_{i=1}^{k-1} g(\mathbf{z}_k, \mathbf{z}_i^+) \mathbf{z}_i^+}{|\mathbf{z}_k - \sum_{i=1}^{k-1} g(\mathbf{z}_k, \mathbf{z}_i^+) \mathbf{z}_i^+|}, k = 1, 2, \dots, n \right\}.$$

Observe that a Gram-Schmidt basis is such that  $\mathbf{z}_k^+$  is a linear combination of  $\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_{k-1}$ , so by Theorem 3.2.3, this clearly means that the change-of-basis matrix for the Gram-Schmidt process is upper-triangular. Moreover, this implies that Technique 4.3.6 always yields a valid output for any inner product space because every vector space has a basis.

#### Proposition 4.3.7 ▶ Existence of Orthonormal Bases

*Every inner product space has an orthonormal basis.*

## 4.4 Riesz Representation

Recall that in Proposition 4.1.7, we have demonstrated that every bilinear form over a vector space  $V$  defines a mapping from  $V$  to  $\mathcal{V}$ . This mapping becomes very special if we construct it with respect to an inner product.

### Theorem 4.4.1 ► Riesz Representation Theorem

For any inner product space  $(V, g)$ , let the mapping  $\Gamma : V \rightarrow \mathcal{V}$  be such that

$$\Gamma(\mathbf{u})(\mathbf{v}) = g(\mathbf{v}, \mathbf{u}),$$

then for every  $\alpha \in \mathcal{V}$ , there is a unique  $\mathbf{u}_\alpha \in V$  such that  $\alpha = \Gamma(\mathbf{u}_\alpha)$ .

*Proof.*  $\Gamma$  is clearly a homomorphism, so this is equivalent to proving that  $\Gamma$  is an isomorphism. Let  $\Gamma(\mathbf{u}_0) = 0$ , then  $g(\mathbf{v}, \mathbf{u}_0) = 0$  for all  $\mathbf{v} \in V$ . Therefore,  $g(\mathbf{0}, \mathbf{u}_0) = 0$ , which implies that  $\mathbf{u}_0 = \mathbf{0}$ . This means that  $\text{null}(\Gamma) = \{0\}$  and so  $\Gamma$  is injective. By Theorem 2.1.5,

$$\text{rank}(\Gamma) = \dim(V) = \dim(\mathcal{V}),$$

and so  $\text{range}(\Gamma) = \mathcal{V}$ , which means that  $\Gamma$  is surjective. Therefore,  $\Gamma$  is an isomorphism. □

The above theorem essentially says that in any inner product space  $(V, g)$ , the dual vectors can be expressed in terms of the vectors in  $V$  through  $g$ . Immediately, we see that if  $\{\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_n\}$  is an orthonormal basis, then the corresponding dual basis can be expressed as  $\zeta^i = \Gamma(\mathbf{z}_i)$  because

$$\Gamma(\mathbf{z}_i)(\mathbf{z}_j) = g(\mathbf{z}_j, \mathbf{z}_i) = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}.$$

Recall that in  $\mathbb{R}^n$ , we can extract the  $i$ -th component of a vector by computing the dot product between the  $i$ -th vector in the orthonormal basis and the vector. Using Theorem 4.4.1, this gets generalised to any vector space because

$$\Gamma(\mathbf{z}_i)(\mathbf{v}) = \Gamma(\mathbf{z}_i)\left(\sum_{j=1}^n a_j \mathbf{z}_j\right) = \zeta^i\left(\sum_{j=1}^n a_j \mathbf{z}_j\right) = a_i.$$

Note that  $\Gamma(\mathbf{v})$  is a dual vector, so it has a matrix representation. Notice that by definition,

the  $i$ -th component of  $\Gamma(\mathbf{v})$  is just

$$\Gamma(\mathbf{v})(\mathbf{z}_i) = g\left(\mathbf{z}_i, \sum_{j=1}^n a_j \mathbf{z}_j\right) = \sum_{j=1}^n G_j^i a_j,$$

where  $G_j^i$  is the  $(i, j)$  entry of the matrix representation of  $g$ . Therefore,

$$\Gamma(\mathbf{v}) = \sum_{i=1}^n \sum_{j=1}^n G_j^i a_j \zeta^i.$$

One important application of Theorem 4.4.1 is to convert an operator to a bilinear form over an inner product space.

**Proposition 4.4.2 ► Conversion from Operators to Bilinear Forms**

*Let  $T$  be an operator on an inner product space  $(V, g)$ , then  $\Gamma \circ T$  defines a bilinear form over  $V$ , where  $\Gamma$  is the Riesz isomorphism.*

*Proof.* By Proposition 4.1.7, since  $\Gamma \circ T \in \mathcal{L}(V, \mathfrak{V})$ , there is a unique bilinear form  $b_T : V \times V \rightarrow V$  defined by

$$b_T(\mathbf{u}, \mathbf{v}) = (\Gamma \circ T)(\mathbf{v})(\mathbf{u}) = \Gamma(T(\mathbf{v}))(\mathbf{u}) = g(\mathbf{u}, T(\mathbf{v})).$$

Conversely, for any bilinear form  $b \in \mathbb{B}(V)$ , we can similarly construct  $b_T^\# : V \rightarrow \mathfrak{V}$  by  $b_T^\#(\mathbf{v})(\mathbf{u}) = b(\mathbf{u}, \mathbf{v})$ . Therefore,  $\Gamma^{-1} \circ b_T^\# \in \mathcal{L}(V, V)$  is an operator over  $V$ .  $\square$

In fact, for any operator  $T$  over an inner product space  $(V, g)$ , we can write down the components of  $b_T$  explicitly. Fix a basis  $\{\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_n\}$  for  $V$ , then

$$\begin{aligned} M_z(b_T)_j^i &= b_T(\mathbf{z}_i, \mathbf{z}_j) \\ &= g(\mathbf{z}_i, T(\mathbf{z}_j)) \\ &= g\left(\mathbf{z}_i, \sum_{k=1}^n M_z(T)_j^k \mathbf{z}_k\right) \\ &= \sum_{k=1}^n M_z(T)_j^k g(\mathbf{z}_i, \mathbf{z}_k) \\ &= \sum_{k=1}^n M_z(g)_k^i M_z(T)_j^k. \end{aligned}$$

Therefore,  $M_z(b_T) = M_z(g)M_z(T)$ . Since  $g$  is an inner product,  $M_z(g)$  is invertible, which means  $M_z(T) = M_z(b_T)M_z(g)^{-1}$ . So here we obtain another way to convert between operators and bilinear forms via the inner product using matrix multiplication.

**Definition 4.4.3 ▶ Riesz Equivalence**

Let  $(V, g)$  be any inner product space with Riesz isomorphism  $\Gamma$ . An operator  $T$  over  $V$  and a bilinear form  $b \in \mathbb{B}(V)$  are said to be **Riesz-equivalent** if

$$b(\mathbf{u}, \mathbf{v}) = (\Gamma \circ T)(\mathbf{v})(\mathbf{u}) \quad \text{or} \quad b(\mathbf{u}, \mathbf{v}) = g(\mathbf{u}, T(\mathbf{v})).$$

Notice that  $M_z(g) = \mathbf{I}$  if  $z$  is an orthonormal basis for the inner product space  $(V, g)$ , so it is easy to check the following proposition holds:

**Proposition 4.4.4 ▶ Matrix Representation of Riesz-Equivalent Maps**

*If  $\tau$  is a bilinear form that is Riesz-equivalent to an operator  $T$ , then  $\tau$  and  $T$  have the same matrix representation under an orthonormal basis.*

## 4.5 Sesquilinear Forms

A peculiar case to take special care is vector spaces over  $\mathbb{C}$ . Note that  $\mathbb{C}$  is not a well-order, so some of the definitions from previous sections might need to get tweaked in order to stay consistent in the complex case.

**Definition 4.5.1 ▶ Dot Product over  $\mathbb{C}^n$** 

The **dot product** or **scalar product** between any  $\mathbf{x}, \mathbf{y} \in \mathbb{C}^n$  is defined as  $\mathbf{x} \cdot \mathbf{y} = \mathbf{x}^T \bar{\mathbf{y}}$ .

However, we immediately realise that following this new definition, the dot product is no longer commutative nor associative under scalar multiplication, because

$$\mathbf{u} \cdot \mathbf{v} = \overline{\mathbf{v} \cdot \mathbf{u}}, \quad \mathbf{u} \cdot c\mathbf{v} = \bar{c}\mathbf{u} \cdot \mathbf{v}.$$

Therefore, we need to re-define an equivalent notion of bilinear forms for complex-valued fields and vector spaces.

**Definition 4.5.2 ▶ Sesquilinear Form**

Let  $V$  be a vector space over  $\mathbb{C}$ . A **sesquilinear form** is a binary map  $s : V \times V \rightarrow \mathbb{C}$  such that

$$\begin{aligned} s(\alpha \mathbf{u}_1 + \beta \mathbf{u}_2, \mathbf{v}) &= \alpha s(\mathbf{u}_1, \mathbf{v}) + \beta s(\mathbf{u}_2, \mathbf{v}), \\ s(\mathbf{u}, \alpha \mathbf{v}_1 + \beta \mathbf{v}_2) &= \bar{\alpha} s(\mathbf{u}, \mathbf{v}_1) + \bar{\beta} s(\mathbf{u}, \mathbf{v}_2) \end{aligned}$$

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

If we denote the set of all sesquilinear forms over a vector space  $V$  as  $\mathbb{S}(V)$ , then just like Proposition 4.1.7,  $\mathbb{S}(V) \cong \text{Maps}(V, \mathbb{V})$ . This is because for every  $s \in \mathbb{S}(V)$ , we can also define some  $s^\# \in \text{Maps}(V, \mathbb{V})$  such that  $s^\#(\mathbf{v})(\mathbf{u}) = s(\mathbf{u}, \mathbf{v})$ . However, notice that now,  $s^\#$  is no longer linear. Instead, it is said to be *conjugate-linear*.

#### Definition 4.5.3 ► Conjugate-Linear Map

Let  $V$  and  $W$  be vector spaces over  $\mathbb{C}$ , then a map  $f : V \rightarrow W$  is said to be **conjugate-linear** if

$$f(a\mathbf{v} + b\mathbf{w}) = \bar{a}f(\mathbf{v}) + \bar{b}f(\mathbf{w})$$

for all  $a, b \in \mathbb{C}$  and all.

Therefore, every sesquilinear map over a complex vector space  $V$  defines a unique conjugate-linear map from  $V$  to  $\mathbb{V}$ . In the other direction, given any conjugate-linear map  $s^\# : V \rightarrow \mathbb{V}$ , we can define some sesquilinear form  $s : V \times V \rightarrow \mathbb{C}$  by  $s(\mathbf{u}, \mathbf{v}) = s^\#(\mathbf{v})(\mathbf{u})$ .

Luckily, the matrix representation of a sesquilinear form can be computed in the exact same way as a bilinear form by taking  $s(\mathbf{z}_i, \mathbf{z}_j)$  as the  $(i, j)$  entry. However, under a change of basis by  $\mathbf{y}_i = \sum_{j=1}^n P_i^j \mathbf{z}_j$ , the new components are given by

$$\begin{aligned} s(\mathbf{y}_i, \mathbf{y}_j) &= s\left(\sum_{k=1}^n P_i^k \mathbf{z}_k, \sum_{k=1}^n P_j^k \mathbf{z}_k\right) \\ &= \sum_{p=1}^n \sum_{q=1}^n P_i^p s(\mathbf{z}_p, \mathbf{z}_q) \overline{P_j^q}. \end{aligned}$$

Therefore, the change-of-basis procedure in the complex case is given by

$$M_y(s) = \mathbf{P}^T M_z(s) \bar{\mathbf{P}}.$$

We now construct the inner product in the complex case. However, we need to re-define symmetric-ness here because otherwise it would conflict with the properties of sesquilinear maps. A natural choice here is to just replace symmetric-ness with conjugate-symmetric-ness, i.e.,  $s(\mathbf{u}, \mathbf{v}) = \overline{s(\mathbf{v}, \mathbf{u})}$ . A nice consequence of this is that  $s(\mathbf{u}, \mathbf{u})$  is always real.

#### Definition 4.5.4 ► Complex Inner Product Space

A **complex inner product space** is a vector space  $V$  over  $\mathbb{C}$  augmented with a sesquilinear form  $g$  such that

1.  $g(\mathbf{u}, \mathbf{u}) \in \mathbb{R}_0^+$  for all  $\mathbf{u} \in V$ ;
2.  $g(\mathbf{u}, \mathbf{u}) = 0$  if and only if  $\mathbf{u} = \mathbf{0}$ ;
3.  $g(\mathbf{u}, \mathbf{v}) = \overline{g(\mathbf{v}, \mathbf{u})}$  for all  $\mathbf{u}, \mathbf{v} \in V$ .

The use of conjugate-symmetric-ness ensures consistency with sesquilinear form because

$$g(\mathbf{u}, c\mathbf{v}) = \overline{g(c\mathbf{v}, \mathbf{u})} = \overline{cg(\mathbf{v}, \mathbf{u})} = \bar{c}\overline{g(\mathbf{v}, \mathbf{u})} = \bar{c}g(\mathbf{u}, \mathbf{v}).$$

In terms of the matrix representation, a conjugate-symmetric matrix means that  $\mathbf{M}^T = \overline{\mathbf{M}}$ .

Now, most of the definitions can work similarly in the complex vector spaces with the modified definitions. The only remaining modification here is that when changing between orthonormal bases, we need to choose a *unitary* matrix instead of an orthogonal one.

#### Definition 4.5.5 ► Unitary Matrix

A matrix  $\mathbf{P}$  is said to be a **unitary matrix** if  $\overline{\mathbf{P}^T}\mathbf{P} = \mathbf{I}$ .

Similar to  $s^\# \in \text{Maps}(V, \mathfrak{V})$  defined uniquely by each sesquilinear form  $s$  over  $V$ , the Riesz isomorphism  $\Gamma: V \rightarrow \mathfrak{V}$  is now a conjugate-linear map instead of a linear one. In fact, we say that it is now a *conjugate isomorphism*. However, note that unlike in a well-ordered field, now  $\Gamma(\mathbf{u})(\mathbf{v}) = \overline{\Gamma(\mathbf{v})(\mathbf{u})}$  due to conjugate-symmetric-ness. Similarly, we can convert an operator  $T$  to a sesquilinear form  $s_T$  over a complex vector space, but now the matrix representation is given by

$$M_z(s_T)_j^i = \sum_{k=1}^n M_z(g)_k^i \overline{M_z(T)_j^k}.$$

We are now ready to state a bigger theorem.

#### Theorem 4.5.6 ► Schur's Triangularisation Theorem

*Let  $T$  be an operator on a finite-dimensional complex inner product space, then there exists an orthonormal basis  $z$  such  $M_z(T)$  is upper-triangular.*

*Proof.* Let  $(V, g)$  be a finite-dimensional complex inner product space with a basis  $y$  and let  $T$  be an operator on  $V$ . By Technique 4.3.6, we can obtain an orthonormal basis  $y^+$  such that  $M_{y^+}(T) = \mathbf{G}^{-1}M_y(T)\mathbf{G}$ , where  $\mathbf{G}$  is an upper-triangular unitary matrix. By Theorem 3.2.3, there exists some change-of-basis matrix  $\mathbf{M}$  such that

$$\mathbf{N} := \mathbf{M}^{-1}\mathbf{G}^{-1}M_y(T)\mathbf{G}\mathbf{M}$$

is upper-triangular. Using Gram-Schmidt process again gives an orthogonal matrix

$$\mathbf{Q} = \mathbf{G}^{-1}\mathbf{N}\mathbf{G},$$

which is an upper-triangular matrix of  $T$  under some orthonormal basis. □



*Remark.* An alternative way to state Schur's Triangularisation Theorem is that every square complex-valued matrix  $\mathbf{Q}$  can be expressed as  $\mathbf{G}^{-1}\mathbf{N}\mathbf{G}$  for some upper-triangular matrix  $\mathbf{N}$  and unitary matrix  $\mathbf{G}$ .

In Theorem 4.5.6, if our initial basis  $y$  is already orthonormal, then it is clear that  $\mathbf{Q} = \mathbf{G}^{-1}\mathbf{M}^{-1}M_y(T)\mathbf{M}\mathbf{G}$ , and so  $\mathbf{M}\mathbf{G}$  is unitary.

This theorem also has a version specialised for sesquilinear forms.

#### Corollary 4.5.7 ▶ Schur's Triangularisation Theorem for Sesquilinear Forms

*Let  $\tau$  be a sesquilinear form on a finite-dimensional complex inner product space, then there exists an orthonormal basis  $z$  such  $M_z(\tau)$  is upper-triangular.*

*Proof.* Let  $V$  be a finite-dimensional complex inner product space. Take  $T : V \rightarrow V$  to be a Riesz-equivalent operator to  $\tau$ . By Theorem 4.5.6, there exists some orthonormal basis  $z$  for  $T$  such that  $M_z(T)$  is upper-triangular. By Proposition 4.4.4, we know that  $M_z(\tau) = \overline{M_z(T)}$ , so  $M_z(\tau)$  is upper triangular.  $\square$

*Remark.* Another consequence of Corollary 4.5.7 is that the eigenvalues of  $\tau$  is exactly the conjugates of the eigenvalues of  $T$ .

Recall that every square matrix can be expressed as a sum of symmetric and anti-symmetric matrices. We have defined the analogue of symmetric matrix as *conjugate-symmetric* in complex vector spaces. Now let us discuss the counterpart of anti-symmetric matrices.

#### Definition 4.5.8 ▶ Hermitianity

A sesquilinear form  $s : V \times V \rightarrow \mathbb{C}$  is said to be **Hermitian** if it is conjugate-symmetric, i.e.,  $s(\mathbf{u}, \mathbf{v}) = \overline{s(\mathbf{v}, \mathbf{u})}$ , and **anti-Hermitian** if  $s(\mathbf{u}, \mathbf{v}) = -\overline{s(\mathbf{v}, \mathbf{u})}$ .

It is clear that all inner products in complex vector spaces are Hermitian. The following result is easy to verify:

#### Proposition 4.5.9 ▶ Hermitianity of Riesz-Equivalent Sesquilinear Forms

*Let  $\tau$  be the Riesz-equivalent sesquilinear form to an operator  $T$  in some complex inner product space  $(V, g)$ , then  $\tau$  is Hermitian if and only if  $g(\mathbf{u}, T(\mathbf{v})) = g(T(\mathbf{u}), \mathbf{v})$ .*

*Remark.* Sometimes we say that an operator is Hermitian if it is Riesz-equivalent to a Hermitian sesquilinear form.

In terms of matrices, if  $\mathbf{S}$  is the matrix representation for  $s$ , then  $S_j^i = \overline{S_i^j}$  if  $s$  is Hermitian and  $S_j^i = -\overline{S_i^j}$  if  $s$  is anti-Hermitian, which leads to the following definition:

#### Definition 4.5.10 ► Hermitian Matrix

A complex-valued matrix  $\mathbf{M}$  is said to be a **Hermitian matrix** if  $\mathbf{M}^T = \overline{\mathbf{M}}$ , and an **anti-Hermitian matrix** if  $\mathbf{M}^T = -\overline{\mathbf{M}}$ .

Combining Definitions 4.5.8 and 4.5.10 leads to the following result:

#### Theorem 4.5.11 ► Spectral Theorem

*Every Hermitian sesquilinear form over a complex inner product space has a real diagonal matrix representation under some orthonormal basis.*

*Proof.* Without loss of generality, let  $\mathbf{M}$  be the matrix representation of the sesquilinear form under an orthonormal basis. By Corollary 4.5.7, there exists some unitary matrix  $\mathbf{P}$  such that  $\mathbf{U} := \mathbf{P}^T \mathbf{M} \overline{\mathbf{P}}$  is upper-triangular. Since  $\mathbf{M}$  is Hermitian,

$$\overline{\mathbf{U}^T} = \overline{(\mathbf{P}^T \mathbf{M} \overline{\mathbf{P}})^T} = \mathbf{P}^T \mathbf{M} \overline{\mathbf{P}} = \mathbf{U}.$$

Therefore,  $\mathbf{U}$  is Hermitian. However, this means that  $\mathbf{U}$  is both upper- and lower-triangular, which means that  $\mathbf{U}$  is diagonal. Notice that  $U_i^i = \overline{U_i^i}$ , so  $\mathbf{U}$  must be a real matrix. □

Using Riesz-equivalence, we can easily translate this theorem to implications in terms of operators.

#### Corollary 4.5.12 ► Hermitian Operators Have Real Eigenvalues

*Let  $T$  be a Riesz-equivalent operator to a Hermitian sesquilinear form  $s$ , then all eigenvalues of  $T$  are real.*

*Proof.* By Theorem 4.5.11, there exists some basis  $z$  such that  $M_z(s)$  is real and diagonal. By Proposition 4.4.4,  $M_z(T) = M_z(s)$  is also real and diagonal, whose diagonal entries are exactly the eigenvalues of  $T$ . □

Of course, symmetric bilinear forms are just a special case of Hermitian sesquilinear forms, so we can extend Theorem 4.5.11 to the non-complex case.

**Corollary 4.5.13 ► Spectral Theorem in Non-complex Spaces**

*Every symmetric bilinear form  $b$  over a real inner product space  $V$  has a real diagonal matrix representation under some orthonormal basis such that there is a real eigenvector associated to each eigenvalue.*

*Proof.* By Theorem 4.5.11, it is clear that there is an orthonormal basis  $z$  such that the matrix  $M_z(b)$  is real and diagonal, so it suffices to prove that every eigenvalue of  $M_z(b)$  is associated to at least one real eigenvector. Let  $\lambda$  be any eigenvalue of  $M_z(b)$ , then there is some  $\mathbf{v} \neq \mathbf{0}$  such that  $M_z(b)\mathbf{v} = \lambda\mathbf{v}$ . Therefore,

$$M_z(b)\bar{\mathbf{v}} = \overline{M_z(b)\mathbf{v}} = \overline{\lambda\mathbf{v}} = \lambda\bar{\mathbf{v}}$$

because both  $M_z(b)$  and  $\lambda$  are real, and so  $\bar{\mathbf{v}}$  is also an eigenvector associated to  $\lambda$ . Clearly,

$$M_z(b)(\mathbf{v} + \bar{\mathbf{v}}) = \lambda(\mathbf{v} + \bar{\mathbf{v}}), \quad M_z(b)i(\mathbf{v} - \bar{\mathbf{v}}) = \lambda i(\mathbf{v} - \bar{\mathbf{v}}),$$

which implies that  $\mathbf{v} + \bar{\mathbf{v}}, i(\mathbf{v} - \bar{\mathbf{v}}) \in V$  are both real. Since  $\mathbf{v} \neq \mathbf{0}$ , they are not both zero, so at least one of them is a real eigenvector associated to  $\lambda$ .  $\square$

Notice that all of our discussion so far requires us to start with some sesquilinear form or operator over some well-defined inner product space. However, in practice it is possible to work with solely matrices. For example, if  $\mathbf{M}$  is an  $n \times n$  symmetric real matrix, it is of course the matrix representation for the bilinear form

$$\sum_{i=1}^n \sum_{j=1}^n M_j^i \epsilon^i \otimes \epsilon^j.$$

# Determinant

In the previous chapter, we briefly discuss the notions of lengths and angles generalised to any vector space with the aid of an inner product. Now, we will continue our discussion to introduce the notion of *volume* in the generalised sense (clearly, area is just a volume in 2-dimensional spaces).

## 5.1 $m$ -Forms

We would like to generalise the notion of bilinear forms.

### Definition 5.1.1 ► Multilinear Form

Let  $V$  be a vector space over  $\mathbb{F}$ . A **multilinear form** over  $V$  is a map  $T : V^k \rightarrow \mathbb{F}$  such that  $T$  is linear in each of its  $k$  arguments.  $k$  is known as the **degree** of  $T$ .

Obviously, the set of all multilinear form of degree  $k$  over  $V$  is a vector space itself. Bilinear forms are just multilinear forms of degree 2.

### Definition 5.1.2 ► Two-Form

A bilinear form  $\Psi$  over a finite-dimensional vector space  $V$  is said to be a **two-form** if it is anti-symmetric, i.e.,  $\Psi(\mathbf{u}, \mathbf{v}) = -\Psi(\mathbf{v}, \mathbf{u})$  for all  $\mathbf{u}, \mathbf{v} \in V$ .

Again, the set of all two-forms over a vector space  $V$  is a vector space itself. Recall that the inner product is just a special type of symmetric bilinear forms. Correspondingly here, the two-form is related to something called the *exterior product*.

### Definition 5.1.3 ► Wedge Product

Let  $V$  be a finite-dimensional vector space. For every  $\alpha, \beta \in \mathcal{V}$ , their **wedge product** or **exterior product** is defined as

$$\alpha \wedge \beta := \alpha \otimes \beta - \beta \otimes \alpha.$$

Notice that  $\beta \wedge \alpha = -\alpha \wedge \beta$ , so the wedge product is clearly a two-form. Apparently, this implies that  $\alpha \wedge \alpha$  is just the zero mapping. Using the wedge product, we can derive the dimension of the space containing all two-forms over a vector space.

**Proposition 5.1.4 ► Dimension of the Two-Form Space**

Let  $\mathcal{F}_2(V)$  be the set of all two-forms over a finite-dimensional vector space  $V$ , then

$$\dim(\mathcal{F}_2(V)) = \binom{\dim(V)}{2}$$

*Proof.* Let  $\{\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_n\}$  be a basis for  $V$  with dual basis  $\{\zeta^1, \zeta^2, \dots, \zeta^n\}$ . For any two-form  $\Psi \in \mathcal{F}_2(V)$ , define

$$\Phi := \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \psi_{ij} \zeta^i \wedge \zeta^j$$

where  $\psi_{ij} = \Psi(\mathbf{z}_i, \mathbf{z}_j)$ . Note that  $\psi_{ji} = -\psi_{ij}$ . Therefore,

$$\begin{aligned} \Phi(\mathbf{z}_p, \mathbf{z}_q) &= \frac{1}{2} \left( \sum_{i=1}^n \sum_{j=1}^n \psi_{ij} \zeta^i \wedge \zeta^j \right)(\mathbf{z}_p, \mathbf{z}_q) \\ &= \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \psi_{ij} (\zeta^i \otimes \zeta^j - \zeta^j \otimes \zeta^i)(\mathbf{z}_p, \mathbf{z}_q) \\ &= \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \psi_{ij} (\zeta^i \otimes \zeta^j)(\mathbf{z}_p, \mathbf{z}_q) + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \psi_{ji} (\zeta^j \otimes \zeta^i)(\mathbf{z}_p, \mathbf{z}_q) \\ &= \sum_{i=1}^n \sum_{j=1}^n \psi_{ij} (\zeta^i \otimes \zeta^j)(\mathbf{z}_p, \mathbf{z}_q). \end{aligned}$$

By Proposition 4.1.6, we know that

$$\Psi = \sum_{i=1}^n \sum_{j=1}^n \Psi(\mathbf{z}_i, \mathbf{z}_j) (\zeta^i \otimes \zeta^j) = \sum_{i=1}^n \sum_{j=1}^n \psi_{ij} (\zeta^i \otimes \zeta^j) = \Phi.$$

Therefore,  $\{\zeta^i \wedge \zeta^j : i \neq j\}$  is a spanning set of  $\mathcal{F}_2(V)$ . Since  $\zeta^i \wedge \zeta^j = -\zeta^j \wedge \zeta^i$ , this implies that

$$\mathcal{F}_2(V) = \text{span}\{\zeta^i \wedge \zeta^j : i < j\}.$$

Notice that this spanning set is minimal, so it is a basis for  $\mathcal{F}_2(V)$ . Therefore,

$$\dim(\mathcal{F}_2(V)) = |\{\zeta^i \wedge \zeta^j : i < j\}| = \binom{\dim(V)}{2}.$$

□

We can easily generalise two-forms to  $m$ -forms.

**Definition 5.1.5** ▶ *m*-Form

An ***m*-form** in a finite-dimensional vector space  $V$  over  $F$  is an anti-symmetric multilinear form  $\Psi : V^m \rightarrow \mathbb{F}$ , i.e.,

$$\Psi(v_1, v_2, \dots, v_i, \dots, v_j, \dots, v_m) = -\Psi(v_1, v_2, \dots, v_j, \dots, v_i, \dots, v_m)$$

for any  $i \neq j$ .

Similarly, the wedge product between  $m$  dual vectors can be written as

$$\bigwedge_{i=1}^m \alpha_i := \sum_{i=1}^{m!} \text{sgn}(\pi_i) \bigotimes_{j=1}^m \alpha_{\pi_i(j)}$$

where each  $\pi_i$  is a distinct permutation of the  $\alpha_i$ 's and

$$\text{sgn}(\pi_i) = \begin{cases} 1 & \text{if an even number of swaps occur} \\ -1 & \text{otherwise} \end{cases}.$$

Intuitively, we can extend Proposition 5.1.4 to general  $m$ -forms.

**Proposition 5.1.6** ▶ Dimension of the *m*-Form Space

Let  $\mathcal{F}_m(V)$  be the space of all  $m$ -forms over  $V$ , then

$$\dim(\mathcal{F}_m(V)) = \binom{\dim(V)}{m}$$

and if  $\{z_1, z_2, \dots, z_n\}$  be a basis for  $V$  with dual basis  $\{\zeta^1, \zeta^2, \dots, \zeta^n\}$ , each  $\Psi \in \mathcal{F}_m(V)$  can be written as

$$\Psi = \frac{1}{m!} \sum_{1 \leq i_1 < i_2 < \dots < i_m \leq \dim(V)} \Psi(z_{i_1}, z_{i_2}, \dots, z_{i_m}) \bigwedge_{j=1}^m \zeta^{i_j}$$

Two results are immediate from the above proposition:

1. For an  $m$ -form to be well-defined over an  $n$ -dimensional vector space, we require  $0 < m \leq n$ .
2. The space of  $n$ -forms over an  $n$ -dimensional vector space is 1-dimensional.

Furthermore, the space of  $n$ -forms over an  $n$ -dimensional vector space  $V$  can be explicitly

written as

$$\mathcal{F}_n(V) := \left\{ \lambda \bigwedge_{i=1}^n \zeta^i : \lambda \in \mathbb{F} \right\}$$

for some dual basis  $\{\zeta^1, \zeta^2, \dots, \zeta^n\}$ .

Recall that in Definition 2.2.6, we defined the transposition of linear transformations. Applying this to  $m$ -forms, we can similarly define, for each linear operator  $T$  on  $V$ , an operator  $\mathfrak{p}$  on the space of  $m$ -forms over  $V$  by

$$\mathfrak{p}(\Psi)(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m) = \Psi(T(\mathbf{v}_1), T(\mathbf{v}_2), \dots, T(\mathbf{v}_m)).$$

Now, if  $V$  happens to be  $m$ -dimensional, then  $\mathfrak{p}(\Psi)$  has to be some scalar multiple of  $\Psi$  because the space of  $m$ -forms now becomes one-dimensional.

## 5.2 Determinant

### Definition 5.2.1 ► Determinant

For each operator  $T$  on a finite-dimensional vector space  $V$ , we define the **determinant** of  $T$  to be a scalar  $\Delta(T)$  such that

$$\mathfrak{p}(\Omega) = \Delta(T)\Omega$$

for every  $n$ -form  $\Omega$  over  $V$ .

*Remark.* Clearly,  $\Delta(T)$  is the unique eigenvalue of  $\mathfrak{p}$ .

Now we propose the following way to compute  $\Delta(T)$  for any operator:

### Proposition 5.2.2 ► Computing Determinant of Operators

For each operator  $T$  with eigenvalues  $\lambda_1, \lambda_2, \dots, \lambda_k$  on an  $n$ -dimensional vector space  $V$ , where  $m_i$  is the multiplicity of  $\lambda_i$ , we have

$$\Delta(T) = \prod_{i=1}^k \lambda_i^{m_i}.$$

*Proof.* By Theorem 3.2.3 and Proposition 3.2.5, there exists a basis  $\{\mathbf{t}_1, \mathbf{t}_2, \dots, \mathbf{t}_n\}$  for  $V$  such that

$$T(\mathbf{t}_i) = \lambda_{I(i)}\mathbf{t}_i + \sum_{j=1}^{i-1} a_j\mathbf{t}_j$$

for each  $i = 1, 2, \dots, n$  where  $a_1, a_2, \dots, a_{i-1}$  are constants and  $I(i)$  is the index of the eigenvalue associated to  $\mathbf{t}_i$ . Take any  $n$ -form  $\Omega$  over  $V$  and consider

$$\begin{aligned}\Delta(T)\Omega &= \mathfrak{p}(\Omega)(\mathbf{t}_1, \mathbf{t}_2, \dots, \mathbf{t}_n) \\ &= \Omega(T(\mathbf{t}_1), T(\mathbf{t}_2), \dots, T(\mathbf{t}_n)) \\ &= \Omega(\lambda_{I(1)}\mathbf{t}_1, \lambda_{I(2)}\mathbf{t}_2, \dots, \lambda_{I(n)}\mathbf{t}_n) \\ &= \prod_{i=1}^k \lambda_i^{m_i} \Omega(\mathbf{t}_1, \mathbf{t}_2, \dots, \mathbf{t}_n).\end{aligned}$$

Therefore,

$$\Delta(T) = \prod_{i=1}^k \lambda_i^{m_i}.$$

□

Combining with previous results, the followings are trivial:

1. If  $T$  is Hermitian, then  $\Delta(T) \in \mathbb{R}$ .
2. If  $T$  is a linear transformation on  $V \cong \mathbb{F}^n$  with a basis  $\mathbf{z}$ , then the coordinate transformation  $T_{\mathbf{z}}^*$  is such that  $\Delta(T) = \Delta(T_{\mathbf{z}}^*)$ .
3. For a square matrix  $\mathbf{M}$ , define  $\Delta(\mathbf{M}) = \Delta(T)$  if  $\mathbf{M}$  is the matrix representation of  $T$  under some basis, then for any matrix  $\mathbf{N}$  similar to  $\mathbf{M}$ , we have  $\Delta(\mathbf{N}) = \Delta(\mathbf{M})$ .

We can show that the determinant stays consistent under operator composition.

### Proposition 5.2.3 ► Operator Composition Preserves Determinant

For any operators  $S, T$  over  $V$ , we have  $\Delta(TS) = \Delta(T)\Delta(S)$ .

*Proof.* Let  $\Omega$  be any  $\dim(V)$ -form on  $V$ . Notice that

$$\begin{aligned}\Delta(TS)\Omega(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_{\dim(V)}) &= \mathfrak{p}(\Omega)(TS(\mathbf{v}_1), TS(\mathbf{v}_2), \dots, TS(\mathbf{v}_{\dim(V)})) \\ &= \Omega(TS(\mathbf{v}_1), TS(\mathbf{v}_2), \dots, TS(\mathbf{v}_{\dim(V)})) \\ &= \mathfrak{p}(\Omega)(S(\mathbf{v}_1), S(\mathbf{v}_2), \dots, S(\mathbf{v}_{\dim(V)})) \\ &= \Delta(T)\mathfrak{p}(\Omega)(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_{\dim(V)}) \\ &= \Delta(T)\Delta(S)(\Omega)(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_{\dim(V)}).\end{aligned}$$

Therefore,  $\Delta(TS) = \Delta(T)\Delta(S)$ .

□



Now we have a better justification for  $\Delta(\mathbf{N}) = \Delta(\mathbf{M})$  if  $\mathbf{N}$  and  $\mathbf{M}$  are similar, because

$$\Delta(\mathbf{N}) = \Delta(\mathbf{P}^{-1}\mathbf{M}\mathbf{P}) = \Delta(\mathbf{P}^{-1}\mathbf{P})\Delta(\mathbf{M}) = \Delta(\mathbf{M}).$$

A clever application of the determinant allows us to determine if an operator is bijective.

#### Proposition 5.2.4 ► Bijectivity Test with Determinant

*An operator  $T$  is bijective if and only if  $\Delta(T) \neq 0$ .*

*Proof.* If  $T$  is bijective, then for every  $\mathbf{v} \neq \mathbf{0}$ , we have  $T(\mathbf{v}) \neq \mathbf{0}$ . Therefore, all eigenvalues of  $T$  are non-zero. By Proposition 5.2.2, this means that  $\Delta(T) \neq 0$ .

Conversely, if  $\Delta(T) \neq 0$ , we shall prove that  $T$  is bijective by considering the contrapositive statement. If  $T$  is not a bijection, then by Corollary 2.1.6,  $T$  is not injective and so there exists some  $\mathbf{v} \neq \mathbf{0}$  such that  $T(\mathbf{v}) = \mathbf{0}$ . This implies that 0 is an eigenvalue of  $T$  and so by Proposition 5.2.2,  $\Delta(T) = 0$ .  $\square$

The proof of the above proposition motivates the following:

#### Corollary 5.2.5 ► Determinant of $T - \lambda \text{id}$

*If  $\lambda$  is an eigenvalue of an operator  $T$ , then  $\Delta(T - \lambda \text{id}) = 0$ .*

*Proof.* Let  $\mathbf{v}$  be an eigenvector associated to  $\lambda$ , then  $\mathbf{v} \neq \mathbf{0}$  and

$$(T - \lambda \text{id})(\mathbf{v}) = \mathbf{0},$$

so  $T - \lambda \text{id}$  is not bijective. By Proposition 5.2.4,  $\Delta(T - \lambda \text{id}) = 0$ .  $\square$

## 5.3 Volume

Recall that the determinant of an operator reflects the scaling factor of the volume for a unit hypercube under the transformation. Therefore, we can abstract the notion of volume by using the determinant.

#### Definition 5.3.1 ► Volume

Let  $P$  be a parallelotope spanned by  $\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n\}$  in an  $n$ -dimensional inner product space over a well-ordered field  $\mathbb{F}$ . The **volume** of  $P$  is defined as a mapping

$$\Theta : V^n \rightarrow \mathbb{F}$$

such that

1.  $\Theta(\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n) \geq 0$ ;
2.  $\Theta(\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n) \neq 0$  if and only if  $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n$  are linearly independent;
3.  $\Theta(\mathbf{u}_1, \mathbf{u}_2, \dots, c\mathbf{u}_i, \dots, \mathbf{u}_n) = c\Theta(\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n)$  for any  $c \in \mathbb{F}$ ;
4.  $\Theta(\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_i, \dots, \mathbf{u}_j + c\mathbf{u}_i, \dots, \mathbf{u}_n) = \Theta(\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n)$  for any  $c \in \mathbb{F}$ ;
5.  $\Theta(\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n) = 1$  whenever  $\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n\}$  is an orthonormal basis.

We can easily check that volume is multilinear. Take any basis  $\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n\}$  for  $V$ , then for any vectors  $\mathbf{s}, \mathbf{t} \in V$ , we can write

$$\mathbf{s} = s_1\mathbf{u}_1 + \sum_{i=2}^n s_i\mathbf{u}_i, \quad \mathbf{t} = t_1\mathbf{u}_1 + \sum_{i=2}^n t_i\mathbf{u}_i.$$

Therefore,

$$\begin{aligned} \Theta(\mathbf{s} + \mathbf{t}, \mathbf{u}_2, \mathbf{u}_3, \dots, \mathbf{u}_n) &= \Theta\left(s_1\mathbf{u}_1 + \sum_{i=2}^n s_i\mathbf{u}_i + t_1\mathbf{u}_1 + \sum_{i=2}^n t_i\mathbf{u}_i, \mathbf{u}_2, \mathbf{u}_3, \dots, \mathbf{u}_n\right) \\ &= \Theta(s_1\mathbf{u}_1 + t_1\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \dots, \mathbf{u}_n) \\ &= s_1\Theta(\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \dots, \mathbf{u}_n) + t_1\Theta(\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \dots, \mathbf{u}_n) \\ &= \Theta(s_1\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \dots, \mathbf{u}_n) + \Theta(t_1\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \dots, \mathbf{u}_n) \\ &= \Theta\left(s_1\mathbf{u}_1 + \sum_{i=2}^n s_i\mathbf{u}_i, \mathbf{u}_2, \dots, \mathbf{u}_n\right) + \Theta\left(t_1\mathbf{u}_1 + \sum_{i=2}^n t_i\mathbf{u}_i, \mathbf{u}_2, \dots, \mathbf{u}_n\right) \\ &= \Theta(\mathbf{s}, \mathbf{u}_2, \mathbf{u}_3, \dots, \mathbf{u}_n) + \Theta(\mathbf{t}, \mathbf{u}_2, \mathbf{u}_3, \dots, \mathbf{u}_n). \end{aligned}$$

We can apply this trick to each argument of  $\Theta$  and eventually realise that  $\Theta$  is linear in every single slot. Furthermore, in the 2-dimensional case, we have

$$\begin{aligned} 0 &= \Theta(\mathbf{x} + \mathbf{y}, \mathbf{x} + \mathbf{y}) - \Theta(\mathbf{x}, \mathbf{x}) - \Theta(\mathbf{y}, \mathbf{y}) \\ &= \Theta(\mathbf{x}, \mathbf{y}) + \Theta(\mathbf{y}, \mathbf{x}). \end{aligned}$$

This shows that  $\Theta$  is anti-symmetric and is therefore an  $n$ -form (the verification for the  $n$ -dimensional general case is left to the reader as an exercise).

Now we make the following attempt to construct this volume map explicitly:

### Definition 5.3.2 ► Volume Form

Let  $V$  be an  $n$ -dimensional inner product space over a well-ordered field  $\mathbb{F}$ . Take any orthonormal basis  $\{\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_n\}$  with dual basis  $\{\zeta^1, \zeta^2, \dots, \zeta^n\}$ . The **volume form** of

$V$  is defined as

$$\Theta := |\omega_z| = \left| \bigwedge_{i=1}^n \zeta^i \right|.$$

We shall verify that the volume form truly produces the volume. It is trivial that  $|\omega_z| \geq 0$ . Since  $\Theta$  defined this way is also multilinear, it satisfies axioms 3 and 4 in Definition 5.3.1. It is also clear that  $\Theta(\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_n) = 1$ . Let  $\mathbf{y}$  be any different orthonormal basis, then there exists a transformation such that  $\mathbf{y}_i = Q(\mathbf{z}_i)$  for each  $i = 1, 2, \dots, n$ . Therefore, since  $\omega_z$  is an  $n$ -form,

$$\omega_z(\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n) = \omega_z(Q(\mathbf{z}_1), Q(\mathbf{z}_2), \dots, Q(\mathbf{z}_n)) = \Delta(Q)(\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_n) = \Delta(Q).$$

Let  $\mathbf{Q}$  be the matrix representation of  $Q$ , then  $\mathbf{Q}$  is orthogonal. Therefore,

$$\Delta(Q)^2 = \Delta(\mathbf{Q}^T) \Delta(\mathbf{Q}) = \Delta(\mathbf{Q}^T \mathbf{Q}) = \Delta(\mathbf{I}) = 1$$

Therefore,  $\Theta(\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n) = |\Delta(Q)| = 1$ .

Let  $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n$  be any  $n$  vectors in  $V$ . We first check that if the vectors are linearly dependent, then  $\Theta(\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n) = 0$ . Note that it suffices to check that the volume form evaluates to zero whenever the arguments are repeated. Consider

$$\begin{aligned} \Theta(\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n) &= \left| \left( \bigwedge_{i=1}^n \zeta^i \right) (\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n) \right| \\ &= \left| \sum_{i=1}^{m!} \text{sgn}(\pi_i) \left( \bigotimes_{j=1}^m \zeta^{\pi_i(j)} \right) (\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n) \right|. \end{aligned}$$

If there are some  $\mathbf{u}_p = \mathbf{u}_q$ , then it is clear that one of  $\zeta^{\pi_i(p)}(\mathbf{u}_p)$  and  $\zeta^{\pi_i(q)}(\mathbf{u}_q)$  is zero for every  $i$ , so  $\Theta(\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n) = 0$  as desired. Conversely, if  $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n$  are linearly independent, then there exists some operator  $Q$  such that  $\mathbf{u}_i = Q(\mathbf{z}_i)$  for each  $i$  because the vectors form a basis. Therefore,

$$\omega_z(\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n) = \Delta(Q)\omega_z(\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_n) = \Delta(Q).$$

$Q$  is clearly invertible, so  $\Delta(Q) \neq 0$ . Therefore, indeed the volume form computes the volume. It is obvious that the volume form is not unique and depends on our choice of the orthonormal basis  $\mathbf{z}$ .

**Definition 5.3.3 ▶ Orientation**

An **oriented inner product space**  $(V, g, \omega_z)$  is an inner product space  $(V, g)$  augmented with a mapping

$$\omega_z := \bigwedge_{i=1}^n \zeta^i$$

where  $\{\zeta^1, \zeta^2, \dots, \zeta^n\}$  is the dual basis with respect to  $z$  such that  $\Theta := |\omega_z|$  is a volume in  $(V, g)$ .  $\omega_z$  is said to be an **orientation**. A basis  $x$  is said to be **positively oriented** if  $\omega_z(x_1, x_2, \dots, x_n) > 0$ , and **negatively oriented** otherwise.

By convention, we use the positively oriented canonical basis in  $\mathbb{R}^n$  to construct

$$\omega_\epsilon := \bigwedge_{i=1}^n \epsilon^i.$$

Correspondingly, the standard canonical basis is also called *right-handed*, and the negatively oriented basis  $\{-e_1, -e_2, \dots, -e_n\}$  is said to be *left-handed*. With a little bit of computation, one can verify that

$$\Theta(u, v, w) = |u \cdot v \times w|$$

under the right-handed basis. Lastly, take any linear operator  $T$  over an oriented inner product space  $V$ , we have

$$\Theta(T(u_1), T(u_2), \dots, T(u_n)) = |\Delta(T)| \Theta(u_1, u_2, \dots, u_n).$$

Therefore, indeed the determinant of an operator represents the scaling factor in volume. Note that so far we have been working with orthonormal bases in computing the volume, but in general it may not be the case.

**Proposition 5.3.4 ▶ Volume Form under Non-orthonormal Bases**

Let  $(V, g, \omega_z)$  be an oriented inner product space where  $z$  is an orthonormal basis with dual basis  $\{\zeta^1, \zeta^2, \dots, \zeta^n\}$ . Under any dual basis  $\{\eta^1, \eta^2, \dots, \eta^n\}$ , the volume form of  $V$  is

$$\Theta = \sqrt{\Delta(g)} \left| \bigwedge_{i=1}^n \eta^i \right|.$$

*Proof.* Let  $\{y_1, y_2, \dots, y_n\}$  be the basis corresponding to the  $\eta^i$ 's, then there exists an operator  $Q: V \rightarrow V$  such that  $y_i = Q(z_i)$ . Therefore,

$$\Theta(y_1, y_2, \dots, y_n) = \Delta(Q) \Theta(z_1, z_2, \dots, z_n) = \Delta(Q).$$

Note that  $M_z(g) = I$  because  $z$  is an orthonormal basis, so we have

$$M_y(g) = M_z(Q)^T I M_z(Q) = M_z(Q)^T M_z(Q)$$

Therefore,  $\Delta(M_z(Q))^2 = \Delta(M_y(g)) = \Delta(g)$ , and so

$$\Theta(y_1, y_2, \dots, y_n) = \sqrt{\Delta(g)}.$$

Therefore,

$$\Theta = \sqrt{\Delta(g)} \left| \bigwedge_{i=1}^n \eta^i \right|.$$

□