

# Contents

<b>1</b>	<b>Vector Spaces</b>	<b>2</b>
1.1	Vector Space . . . . .	2
1.2	Isomorphism . . . . .	6
1.3	Basis . . . . .	7
<b>2</b>	<b>Linear Transformations</b>	<b>10</b>
2.1	Linear Transformation . . . . .	10
2.2	Duality . . . . .	13
2.3	Tensor Product . . . . .	19
2.4	Matrix . . . . .	21
2.5	Change of Basis . . . . .	25
2.6	Jordan Matrix . . . . .	27
<b>3</b>	<b>Linear Algebra and Geometry</b>	<b>37</b>
3.1	Bilinear Forms . . . . .	37
3.2	Geometry in Vector Spaces . . . . .	38

# 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  $\mathcal{F}$  with two binary operations  $\mathcal{F}^2 \rightarrow \mathcal{F}$ , namely addition and multiplication, such that

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

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

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  $a, b \in \mathcal{F}$  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**  $1 \in \mathcal{F}$  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  $a, b \in \mathcal{F}$  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  $\mathcal{F}$  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  $a \in \mathcal{F}$ , 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. In particular, we would like to consider a special construct known as *direct sum*.

**Definition 1.1.6 ▶ Direct Sum**

Let  $V$  be a vector space and  $U_1, U_2 \subseteq V$  such that  $U_1 \cap U_2 = \{\mathbf{0}\}$ , then their **direct sum** is defined as

$$U_1 \oplus U_2 := \{\mathbf{u}_1 + \mathbf{u}_2 : \mathbf{u}_1 \in U_1, \mathbf{u}_2 \in U_2\}.$$

More generally, we can let  $U_1$  and  $U_2$  be any subsets of  $V$  and define  $U_1 + U_2$  in the same manner, which is known as the *sum* of  $U_1$  and  $U_2$ .

It can be easily proven that for any vector space  $V$ , the direct sum of any two subspaces of  $V$  is still a subspace of  $V$ . A nice property of direct sum can be proven as follows:

**Proposition 1.1.7 ▶ Unique Decomposition with Direct Sums**

Let  $V = U_1 \oplus U_2$ , then every  $\mathbf{v} \in V$  can be uniquely expressed as  $\mathbf{u} + \mathbf{w}$  for some  $\mathbf{u} \in U_1$  and  $\mathbf{w} \in U_2$ .

*Proof.* The existence of  $\mathbf{u}$  and  $\mathbf{w}$  is trivial by Definition 1.1.6. Suppose there exist  $\mathbf{u}' \in U_1$  and  $\mathbf{w}' \in U_2$  such that  $\mathbf{u} + \mathbf{w} = \mathbf{u}' + \mathbf{w}'$ , then we have  $\mathbf{u} - \mathbf{u}' = \mathbf{w}' - \mathbf{w}$ . Note 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. □

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  $V$ . More generally, for any subspace  $U \subseteq V$ , we have  $\mathbf{0}_U = \mathbf{0}_V$ , the proof of which should be trivial enough as an exercise to the reader.

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

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

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

To justify this, we first take any  $\mathbf{v} \in \mathcal{F}^n$ . 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  $\mathcal{F}^n$  which is orthogonal to itself, so we have  $U \cap U_\perp = \{\mathbf{0}\}$ . It follows that  $V = U \oplus U_\perp$ .

## 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  $c \in \mathcal{F}$ .

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  $\operatorname{id}_V : V \rightarrow V$ . In fact, any mapping  $\phi : \mathbf{v} \mapsto c\mathbf{v}$  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  $\mathbf{u}, \mathbf{v} \in U$ , then

$$\psi(\phi(\mathbf{u} + \mathbf{v})) = \psi(\phi(\mathbf{u}) + \phi(\mathbf{v})) = \psi(\phi(\mathbf{u})) + \psi(\phi(\mathbf{v}))$$

since  $\phi(\mathbf{u}), \phi(\mathbf{v}) \in V$ . 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  $\mathcal{F}$ , the set  $\mathcal{F}^n$  for any  $n \in \mathbb{N}$  is a vector space over  $\mathcal{F}$ .

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

A vector space  $V$  is said to be **finite-dimensional** over a field  $\mathcal{F}$  if it is isomorphic to  $\mathcal{F}^n$  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.

## 1.3 Basis

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

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

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  $V$ . 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  $\mathcal{F}^n$ .

### Definition 1.3.1 ► Canonical Basis

The **canonical basis** for  $\mathcal{F}^n$  is defined as  $\{\mathbf{e}_i : i = 1, 2, \dots, n\}$ , where each  $\mathbf{e}_i$  is a column vector with 1 in its  $i$ -th row and 0 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 1.3.2 ► 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  $V \cong \mathcal{F}^n$ . Let  $\mathbf{v} \in V$  be an arbitrary vector, then there is some  $\mathbf{u} \in \mathcal{F}^n$  and a bijection  $\phi : \mathcal{F}^n \rightarrow V$  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  $a_i \in \mathcal{F}$  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  $\mathbf{w} \in \mathcal{F}^n$ , 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  $b_i \in \mathcal{F}$  for  $i = 1, 2, \dots, m$ . This means that  $\mathcal{F}^n$  is spanned by  $m$  vectors, which is a contradiction. Therefore,  $|S| = n$ .  $\square$

An immediate corollary from Proposition 1.3.2 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  $\mathcal{F}^n$  and  $V$ , i.e., for any  $(a_1, a_2, \dots, a_n) \in \mathcal{F}^n$ , 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  $\mathbf{c}_1, \mathbf{c}_2 \in \mathcal{F}^n$ , 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 1.3.3 ► Basis

Let  $V$  be an  $n$ -dimensional vector space and  $z : \mathcal{F}^n \rightarrow V$  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  $z^{-1}(\mathbf{v}) \in \mathcal{F}^n$  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  $z : \mathcal{F}^n \rightarrow V$  defines a basis for  $V$ .

# Linear Transformations

## 2.1 Linear Transformation

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  $\mathcal{F}$ , 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  $c \in \mathcal{F}$ .

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  $V \cong \mathcal{F}^n$  be a finite-dimensional vector space. Recall that a basis for  $V$  is essentially a bijective mapping  $z : \mathcal{F}^n \rightarrow V$ , 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  $\mathcal{F}$ . Fix some  $\mathbf{v} \in V$  and define a mapping  $\Theta_{\mathbf{v}} : \mathcal{F} \rightarrow V$  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  $z_1 : \mathcal{F}^n \rightarrow W$  be an isomorphism for some  $n \in \mathbb{Z}^+$ . Note that there exists an isomorphism  $z_2 : \mathcal{F}^n \rightarrow \text{range}(T)$ . 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  $\mathcal{F}$ . The **dual space** of  $V$  is defined as  $\hat{V} := \mathcal{L}(V, \mathcal{F})$ . The elements of  $\hat{V}$  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  $\hat{V}$ .

An intuitive example for an element of  $\hat{V}$  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  $a_1, a_2, \dots, a_n \in \mathcal{F}$ . Now, define a mapping  $\zeta^i : V \rightarrow \mathcal{F}$  as  $\zeta^i(\mathbf{v}) = a_i$ , then clearly  $\zeta^i \in \hat{V}$  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  $\zeta^i \in \hat{V}$  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  $\alpha \in \hat{V}$  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  $\beta : V \rightarrow \mathcal{F}$  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

$$\sum_{i=1}^n p_i \zeta^i = 0_{\hat{V}},$$

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**  $\triangleright \dim(V) = \dim(\hat{V})$

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

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

#### Definition 2.2.4 ► Canonical Dual Basis

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

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

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

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

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

where  $q_i \in \mathcal{F}$  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  $\mathcal{F}^n$  to  $V$ , so it follows that the dual basis is in fact a mapping  $\zeta : \hat{\mathcal{F}}^n \rightarrow \hat{V}$ . We claim that this  $\zeta$  is in fact just a mapping that satisfies

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

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

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  $\phi \in \hat{V}$  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  $p \in \mathcal{F}^n$ . Therefore,  $\zeta$  is surjective. One may check that  $\zeta$  is injective. Moreover,



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

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

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

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

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

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

**Proposition 2.2.5**  $\hat{\hat{V}} \cong V$

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

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

*is isomorphic to  $V$ .*

*Proof.* Define a mapping  $K: V \rightarrow \hat{\hat{V}}$  by  $K(\mathbf{v}) = f$  where  $f: \hat{V} \rightarrow V$  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  $c \in \mathcal{F}$ ,

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

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

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

for all  $\alpha \in \hat{V}$ . 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  $\hat{\hat{V}} \cong V$ .  $\square$

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

$\mathcal{F}$ . 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  $\hat{V}$  tempts us to think that there exists some mapping  $S : \hat{V} \rightarrow \hat{V}$  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  $\hat{T} : \hat{V} \rightarrow \hat{V}$  such that

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

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

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

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

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

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  $\tau(T) = \hat{T}$ , then  $\tau$  is a linear transformation.

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

consider

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

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

### Definition 2.3.1 ► Tensor Product

Let  $V$  be a vector space with the dual space  $\hat{V}$ . For any  $\mathbf{v} \in V$  and  $\alpha \in \hat{V}$ , 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  $\beta \in \hat{V}$ , then clearly  $\beta \circ (\mathbf{v} \otimes \alpha) \in \hat{V}$ . Define a mapping  $T$  over  $\hat{V}$  by

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

then we can see that  $T \in \mathcal{L}(\hat{V}, \hat{V})$ . 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  $\alpha \in \hat{V}$ . Furthermore,

note that  $\alpha(\mathbf{u}) \in \mathcal{F}$ , 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

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

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  $\mathcal{F}^n$ .

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  $\hat{T}$  be the transpose of  $T$ , then the matrix representation of  $\hat{T}$  is such that  $\hat{T}_j^i = T_i^j$ .

*Proof.* Define  $K : V \rightarrow \hat{V}$  by  $K(v)(\alpha) = \alpha(v)$ , then  $K(v) = \alpha \mapsto \alpha(v)$ . 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  $\alpha \in \hat{V}$ , we have

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

*Proof.* Let  $f : \mathbf{v} \otimes \alpha$ , for any  $\beta \in \hat{V}$  and any  $\mathbf{w} \in V$ , consider

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

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

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

□

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  $\hat{V}$ . By Proposition 2.3.2 and Definition 2.4.4, we know that

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

By Proposition 2.2.7, we know that

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

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

□

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  $M_z : \mathcal{L}(V, V) \rightarrow \mathcal{M}_{n \times n}$  that maps each  $T \in \mathcal{L}(V, V)$  to

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

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

Consider  $\mathcal{L}(\mathcal{F}^n, \mathcal{F}^n)$ , it is clear that every  $T \in \mathcal{L}(\mathcal{F}^n, \mathcal{F}^n)$  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  $\alpha, \beta \in \hat{V}$ , 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  $\mathbf{v}_z^* \in \mathcal{F}^n$ . We can view  $\mathbf{v}_z^*$  as the “coordinates” of  $\mathbf{v}$  with respect to  $z$ .

Suppose  $\alpha \in \hat{V}$ , we define some  $\alpha_z^* \in \hat{\mathcal{F}}^n$  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

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

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.

Let  $\mathbf{v} \in V$  be a vector with coordinate vector  $\mathbf{v}_y^* \in \mathcal{F}^n$  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 2.5.1 ▶ 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**.

## 2.6 Jordan Matrix

Since there are infinitely many linear transformations between two vectors 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 2.6.1 ▶ 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  $A\mathbf{v} = \lambda\mathbf{v}$ . We can see that the left-multiplication by  $A$  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 2.6.2 ▶ Eigenvector and Eigenvalue**

Let  $T : V \rightarrow V$  be an operator on a finite-dimensional vector space  $V$  over  $\mathcal{F}$ . A vector  $\mathbf{v} \in V$  is called an **eigenvector** of  $T$  if  $T(\mathbf{v}) = \lambda\mathbf{v}$  for some  $\lambda \in \mathcal{F}$ , 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 2.6.3 ▶ 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  $\mu \in \mathcal{F}$ . 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 2.6.4 ► Eigenspace

Let  $T$  be an operator on a finite-dimensional vector space  $V$  over  $\mathcal{F}$ . For every eigenvalue  $\lambda \in \mathcal{F}$  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 2.6.3:

#### Proposition 2.6.5 ► 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 2.6.3. 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 2.6.5 reveals an important fact that any finite-dimensional vector space has finitely many eigenvalues. In fact, for any  $V \cong \mathcal{F}^n$ , 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 2.6.6 ▶ 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  $\mathbf{w}$ .  $\square$

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

**Definition 2.6.7 ▶ 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 2.6.8 ▶ 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

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

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

Now, let  $T$  be any operator on  $V$  and  $z : \mathcal{F}^n \rightarrow V$  be a basis, then we know that the operator  $T_z^* := z^{-1} \circ T \circ z$  on  $\mathcal{F}^n$  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 2.6.9 ► 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  $T(\mathbf{t}_i)$  can be expressed as a linear combination of the  $\mathbf{t}_j$ '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  $M$  be any  $(k+1) \times (k+1)$  complex-valued matrix. By Proposition 2.6.6, there is some  $\lambda \in \mathbb{C}$  which is an eigenvalue of  $M$  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

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

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

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

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

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

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

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

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

Note that the matrix  $\mathbf{N}$  in the proof for Theorem 2.6.9 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 2.6.9, then the  $(1, 1)$  entry of that matrix will be an eigenvalue of  $T$ . The following result is trivial:

#### Proposition 2.6.10 ► 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 2.6.11 ► Eigenvalues of Upper Triangular Matrices

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

*Proof.* Let  $\lambda$  be any of the diagonal entries of  $\mathbf{N}$  and consider  $\mathbf{N} - \lambda \mathbf{I}$  which is also upper-triangular. Note that by Proposition 2.6.10,  $\mathbf{N} - \lambda \mathbf{I}$  is not invertible, and so there exists some  $\mathbf{v} \neq \mathbf{0}$  such that  $(\mathbf{N} - \lambda \mathbf{I})\mathbf{v} = \mathbf{0}$ . Therefore,  $\mathbf{N}\mathbf{v} = \lambda\mathbf{v}$  and so  $\lambda$  is an eigenvalue of  $\mathbf{N}$ .  $\square$

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



**Definition 2.6.12 ► 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 2.6.11, 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  $\mathbf{u}_i$  such that  $T(\mathbf{u}_i) = \lambda_i \mathbf{u}_i$ . 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  $\dim(V)$  distinct eigenvalues is diagonalisable.

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 2.6.13 ► Jordan Block**

Let  $\mathbf{J}$  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  $\lambda \mathbf{I} + \mathbf{J}$  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 2.6.14 ► Eigenspace of Jordan Block**

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

*Proof.* By Propositions 2.6.10 and 2.6.11, 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.  $\square$

We can build up an upper-triangular matrix using Jordan blocks, which by triangularisability gives the following result:

#### Definition 2.6.15 ► 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 2.6.16 ► 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 2.6.17 ▶ Multiplicity**

Let  $\mathbf{J}$  be the Jordan form of a matrix  $\mathbf{A}$ . For any eigenvalue  $\lambda$  of  $\mathbf{A}$ , 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  $\mathbf{A}$  to be  $\det(\mathbf{A} - \lambda \mathbf{I})$ . Now we propose another definition using the Jordan form.

**Definition 2.6.18 ▶ 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 2.6.19 ▶ 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,  $\chi_T(T) = 0$ . □

An important implication of Theorem 2.6.19 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

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

# Linear Algebra and Geometry

## 3.1 Bilinear Forms

### Definition 3.1.1 ► Dot Product

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

1.  $\mathbf{u} \cdot \mathbf{u} \geq 0$  for all  $\mathbf{u} \in \mathcal{F}^n$ ;
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  $\mathbf{u}, \mathbf{v} \in \mathcal{F}^n$ ;
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  $\mathbf{u}, \mathbf{v} \in \mathcal{F}^n$  and  $\alpha, \beta \in \mathcal{F}$ .

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

### Definition 3.1.2 ► Bilinear Form

A **bilinear form** on a finite-dimensional space  $V$  is defined to be a mapping  $b : V \times V \rightarrow \mathcal{F}$  such that

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

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

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.

### Definition 3.1.3 ► Tensor Product between Dual Vectors

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

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

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

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

over  $V$ .

#### Definition 3.1.4 ► Inner Product Space

An **inner product space** is a finite-dimensional vector space  $V$  with a bilinear form  $g : V \times V \rightarrow \mathcal{F}$ , 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  $\mathbf{u}, \mathbf{v} \in V$ .

## 3.2 Geometry in Vector Spaces

#### Definition 3.2.1 ► Length

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

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

#### Definition 3.2.2 ► Orthogonality

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

#### Theorem 3.2.3 ► Cauchy-Schwarz Inequality

Let  $(V, g)$  be an inner product space, then  $g(\mathbf{u}, \mathbf{v}) \leq |\mathbf{u}||\mathbf{v}|$  for any  $\mathbf{u}, \mathbf{v} \in V$ .

#### Corollary 3.2.4 ► Triangle Inequality

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