Contents

1 Vector Spaces

1.1 Fields, Scalars and Vectors

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 $\mathcal{F}^2 \to \mathcal{F}$, 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. 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 $+ : V^2 \to V$, and
- scalar multiplication $(\quad)(\quad) : \mathcal{F} \times V \to V$,

such that

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

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 necessary the addition and scalar multiplication which we are used to in $\mathbb{R}^n$, but abstract mappings which satisfy the given axioms.

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 $\boldsymbol{u} \in V$ such that $\boldsymbol{v} + \boldsymbol{u} = \boldsymbol{v}$ for all $\boldsymbol{v} \in V$. Since $\mathbf{0} \in V$, we have

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

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

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

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 $\boldsymbol{v} \in V$ has a unique additive inverse.

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

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

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

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

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

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

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

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

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

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 $\mathbf{u} - \mathbf{v} = \boldsymbol{u} + (-1)\boldsymbol{v}$, so the above result allows us to write $\boldsymbol{u} - \boldsymbol{v} = \boldsymbol{u} + (-\boldsymbol{v})$.

1.1.1 Subspaces

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

$$U_1 \oplus U_2 \coloneqq \{\boldsymbol{u}_1 + \boldsymbol{u}_2 : \boldsymbol{u}_1 \in U_1, \boldsymbol{u}_2 \in U_2\}.$$

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 $\boldsymbol{v} \in V$ can be uniquely expressed as $\boldsymbol{u} + \boldsymbol{w}$ for some $\boldsymbol{u} \in U_1$ and $\boldsymbol{w} \in U_2$.

Proof. The existence of $\boldsymbol{u}$ and $\boldsymbol{w}$ is trivial by Definition 1.1.6. Suppose there exist $\boldsymbol{u}' \in U_1$ and $\boldsymbol{w}' \in U_2$ such that $\boldsymbol{u} + \boldsymbol{w} = \boldsymbol{u}' + \boldsymbol{w}'$, then we have $\boldsymbol{u} - \boldsymbol{u}' = \boldsymbol{w}' - \boldsymbol{w}$. Note that $\boldsymbol{u} - \boldsymbol{u}' \in U_1$ and $\boldsymbol{w}' - \boldsymbol{w} \in U_2$, so we have $\boldsymbol{u} - \boldsymbol{u}', \boldsymbol{w}' - \boldsymbol{w} \in U_1 \cap U_2 = \{\mathbf{0}\}$, i.e.,

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

Therefore, $\boldsymbol{u} = \boldsymbol{u}'$ and $\boldsymbol{w} = \boldsymbol{w}'$, i.e., $\boldsymbol{u}$ and $\boldsymbol{w}$ 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

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

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

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

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

Let $\boldsymbol{w} = \boldsymbol{v} - \boldsymbol{u}_0$, then clearly $\boldsymbol{v} = \boldsymbol{w} + \boldsymbol{u}_0$ where $\boldsymbol{u}_0 \in U$ and $\boldsymbol{w} \in U_\perp$. 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 $\phi : U \to V$ such that

$$\phi(\boldsymbol{u} + \boldsymbol{v}) = \phi(\boldsymbol{u}) + \phi(\boldsymbol{v})$$

for any $\boldsymbol{u}, \boldsymbol{v} \in U$.

Note that it suffices to only require $\phi(\boldsymbol{u} + \boldsymbol{v}) = \phi(\boldsymbol{u}) + \phi(\boldsymbol{v})$ but not $\phi(c\boldsymbol{u}) = c\phi(\boldsymbol{u})$. Here, note that a vector space is a connected set, so we can easily prove the above claim with some knowledge in mathematical analysis.

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 $\mathrm{id}_V : V \to V$. In fact, any mapping $\phi : \boldsymbol{v} \mapsto c\boldsymbol{v}$ 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 $\psi : V \to U$. We consider the following theorem:

Theorem 1.2.3 ▸ Composition Preserves Isomorphism

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

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

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

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

Using Theorem 1.2.3, we can immediately see that if $\phi : V \to V$ is any isomorphism and $\psi : V \to U$ 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 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

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

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

$$\operatorname{span}(V) \coloneqq \{a_i \boldsymbol{v}_i : \boldsymbol{v}_i \in V\},$$

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 $a_i \boldsymbol{v}_i = \boldsymbol{0}$ implies that $a_i = 0$ for all $i = 1, 2, \cdots$. A *basis* of a vector space V is a linearly independent set S such that $\operatorname{span}(S) = V$. One may check that if S is a basis for V, then any $\boldsymbol{v} \in V$ 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 $\boldsymbol{v}$ 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 $\{\boldsymbol{e}_i : i = 1, 2, \cdots, n\}$, where each $\boldsymbol{e}_i$ is a column vector with $1_\mathcal{F}$ in its i-th row and $0_\mathcal{F}$ 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 1.3.2 ▶ 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 $\boldsymbol{v} \in V$ be an arbitrary vector, then there is some $\boldsymbol{u} \in \mathcal{F}^n$ and a bijection $\phi : \mathcal{F}^n \to V$ such that

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

where $a_i \in \mathcal{F}$ for $i = 1, 2, \cdots, n$. 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 $\boldsymbol{w} \in \mathcal{F}^n$, there is some $\boldsymbol{r} \in V$ such that

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

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

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 two 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 $(a_1, a_2, \cdots, a_n) \in \mathcal{F}^n$, we can map the tuple to a vector in V whose components are exactly $a_1, a_2, \cdots, a_n$.

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

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

so a basis is nothing more but an isomorphism!

2

Linear Transformations

2.1 Linear Transformations

Definition 2.1.1 ▸ Linear Transformation

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

$$T(\boldsymbol{v} + \boldsymbol{u}) + T(\boldsymbol{v}) + T(\boldsymbol{u}), \qquad T(c\boldsymbol{v}) = cT(\boldsymbol{v})$$

for all $\boldsymbol{v}, \boldsymbol{u} \in V$ 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 $z : \mathcal{F}^n \to V$, 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 $\boldsymbol{v} \in V$ and define a mapping $\Theta_{\boldsymbol{v}} : \mathcal{F} \to V$ by $\Theta_{\boldsymbol{v}}(a) = a\boldsymbol{v}$. One may check that $\Theta_{\boldsymbol{v}}$ is a linear transformation, but $\Theta_{\boldsymbol{v}}$ is essentially $\boldsymbol{v}$, so vectors are linear transformations as well.

Definition 2.1.2 ▸ Range

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

$$\text{range}(T) \coloneqq \{T(\boldsymbol{v}) : \boldsymbol{v} \in V\}.$$

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 $T(\boldsymbol{v}), T(\boldsymbol{u}) \in \text{range}(T)$ for some $\boldsymbol{v} \neq \boldsymbol{u}$ in V, then we have

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

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

Definition 2.1.3 ▸ Kernel

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

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

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 $\boldsymbol{v}\in\ker(T)$ with $T(\boldsymbol{v})=\boldsymbol{0}$, so T is not injective.

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

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

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

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

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

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

Theorem 2.1.5 ▸ Fundamental Theorem of Linear Transformations

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

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

that $S(\boldsymbol{u}_1) = S(\boldsymbol{u}_2)$, then

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

Therefore, $\boldsymbol{u}_1 - \boldsymbol{u}_2 = \mathbf{0}_U$, and so $\boldsymbol{u}_1 = \boldsymbol{u}_2$, 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 $T : V \to V$ be a linear transformation, then T is injective if and only if it is surjective.

Proof. Suppose that T is injective, then $V \cong \text{range}(T)$ and so $\text{null}(T) = 0$, which implies that $\dim(V) = \text{rank}(T)$. However, $\text{range}(T) \subseteq V$, so $\text{range}(T) = V$. This means that T is surjective.

Suppose that T is surjective, then $\dim(V) = \text{rank}(T)$ and so $\text{null}(T) = 0$. Therefore, 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 $\hat{V} \coloneqq \mathcal{L}(V, \mathcal{F})$. The elements of $\hat{V}$ are called **dual vectors**.

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

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

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

In particular, we see that

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

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

Definition 2.2.2 ▸ Dual Basis

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

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

Let ζ 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 $\boldsymbol{v} \in V$ with respect to basis z, define a mapping $\beta : V \to \mathcal{F}$ by

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

Clearly, β is a linear combination of the elements of ζ. For any $\boldsymbol{z}_j \in z$, we have

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

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

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

However, since the a_i's are arbitrary, we have $p_i = 0$ for all $i = 1, 2, \cdots, n$. Therefore, ζ is linearly independent. This means that ζ 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 ▸ $\dim(V) = \dim\left(\hat{V}\right)$

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

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

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

Therefore, any dual vector with respect to basis z can be written as

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

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 $\hat{\mathcal{F}^n} \coloneqq \mathcal{L}(\mathcal{F}^n, \mathcal{F})$.

Definition 2.2.4 ▸ Canonical Dual Basis

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

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

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

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 $i = 1, 2, \cdots, n$ and ϵ^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 $\zeta : \hat{\mathcal{F}^n} \to \hat{V}$. We claim that this ζ is in fact just a mapping that satisfies

$$\zeta(p)\big(z(\boldsymbol{a})\big) = p(\boldsymbol{a})$$

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

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

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

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

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

then

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

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 $p \in \hat{\mathcal{F}^n}$. Therefore, ζ is surjective. One may check that ζ is injective. Moreover, note that for $p, q \in \hat{\mathcal{F}^n}$,

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

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

Naturally, it feels justified to define a "dual space of the dual space" of V, namely

$$\hat{\hat{V}} \coloneqq \mathcal{L}\left(\hat{V}, \mathcal{F}\right) = \mathcal{L}\big(\mathcal{L}(V, \mathcal{F}), \mathcal{F}\big).$$

We can in fact prove that this space is isomorphic to V itself. Therefore, this means that it is perfectly legal to view each $\boldsymbol{v} \in V$ as a mapping from $\hat{V}$ to $\mathcal{F}$. We can verify that $\boldsymbol{v}$ is linear, because

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

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

Definition 2.2.5 ▸ Transpose

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

$$\hat{T}(\alpha)(\boldsymbol{v}) = \alpha\big(T(\boldsymbol{v})\big)$$

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

2.3 Tensor Product

Definition 2.3.1 ▸ Tensor Product

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

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

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

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

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

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

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

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

By Definition 2.2.5, T is the transpose of $\boldsymbol{v} \otimes \alpha$ for any $\boldsymbol{v} \in V$ and $\alpha \in \hat{V}$. Furthermore, note that $\alpha(\boldsymbol{u}) \in \mathcal{F}$, we see that

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

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

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

Let ζ be the dual basis for V with respect to some basis z, then

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

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

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

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

Define a mapping

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

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

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

Therefore,

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

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

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

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

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

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

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

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 $\boldsymbol{M}$ be an $n \times m$ matrix and $\boldsymbol{N}$ be an $m \times p$ matrix, then their product $\boldsymbol{P} = \boldsymbol{MN}$ 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 $\times : V^2 \to V$ 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

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

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

Remark. Note that $T_j^i = \zeta^i\big(T(\boldsymbol{z}_j)\big)$.

Let $\hat{T}$ be the transpose of T, then by Definition 2.2.5 we have

$$\zeta^i\big(T(\boldsymbol{z}_j)\big) = \hat{T}(\zeta^i)(\boldsymbol{z}_j) = z_j\big(\hat{T}(\zeta^i)\big),$$

which we can use to verify that the matrix of $\hat{T}$ is indeed $\boldsymbol{T}^{\mathrm{T}}$.

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 $M_z : \mathcal{L}(V, V) \to \mathcal{M}_{n\times n}$ that maps each $T \in \mathcal{L}(V, V)$ to its matrix representation $\boldsymbol{T}$. 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 $z : \mathcal{F}^{n^2} \to \mathcal{L}(V, V)$, 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

$$T = \sum_{i=1}^{n} \sum_{j=1}^{n} M_z(T)_i^j \left(\boldsymbol{e}_i \otimes \epsilon^i\right).$$

Proposition 2.4.5 ▸ Matrices of Composite Mappings

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

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

Suppose $\boldsymbol{u} = T(\boldsymbol{v})$ and $z(\boldsymbol{a}) = \boldsymbol{v}$, $z(\boldsymbol{b}) = \boldsymbol{u}$, then

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

Suppose $\alpha \in \hat{V}$, we define some $\alpha_z^* \in \hat{\mathcal{F}^n}$ by $\alpha_z^* := \alpha \circ z$.

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

$$P \coloneqq z^{-1} \circ y \in \mathcal{L}\left(\mathcal{F}^n, \mathcal{F}^n\right),$$

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

$$\begin{aligned} y_i &= \epsilon^i(y) \\ &= \epsilon^i(z \circ P) \\ &= \sum_{k=1}^{n} \boldsymbol{P}_i^k z_k. \end{aligned}$$