Contents

1

Vector Spaces

1.1 Vector Spaces

In this section, we introduce the notion of **vector spaces**.

> **Definition 1.1.1 ▶ Vector space**
>
> A vector space is a **non-empty** set V, whose elements are known as **vectors**, over a field K, whose elements are known as **scalars**, with two binary operations, namely addition and scalar multiplication, which satisfies the following axioms:
> - Closure under vector addition: for all $\boldsymbol{u}, \boldsymbol{v} \in V, \boldsymbol{u} + \boldsymbol{v} \in V$.
> - Closure under scalar multiplication: for all $\boldsymbol{v} \in V$ and $c \in K, c\boldsymbol{v} \in V$.
> - Associativity of vector addition: for all $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w} \in V, \boldsymbol{u} + (\boldsymbol{v} + \boldsymbol{w}) = (\boldsymbol{u} + \boldsymbol{v}) + \boldsymbol{w}$.
> - Commutativity of vector addition: for all $\boldsymbol{u}, \boldsymbol{v} \in V, \boldsymbol{u} + \boldsymbol{v} = \boldsymbol{v} + \boldsymbol{u}$.
> - Existence of additive identity: there exists a zero vector $\boldsymbol{0}_V \in V$ known as the the additive identity such that $\forall \boldsymbol{v} \in V, \boldsymbol{v} + \boldsymbol{0}_V = \boldsymbol{v}$.
> - Existence of additive inverse: for all $\boldsymbol{v} \in V$, there exists a vector $-\boldsymbol{v} \in V$ known as the additive inverse of $\boldsymbol{v}$ such that $\boldsymbol{v} + (-\boldsymbol{v}) = 0$.
> - Associativity of scalar multiplication: for all $\boldsymbol{v} \in V$ and $a, b \in K, a(b\boldsymbol{v}) = (ab)\boldsymbol{v}$.
> - Distributivity of scalar multiplication: for all $\boldsymbol{v}, \boldsymbol{u} \in V$ and $a, b \in K, (a + b)\boldsymbol{v} = a\boldsymbol{v} + b\boldsymbol{v}$ and $a(\boldsymbol{u} + \boldsymbol{v}) = a\boldsymbol{u} + a\boldsymbol{v}$.
> - Existence of multiplicative identity: there exists a scalar 1_V known as the multiplicative identity such that for all $\boldsymbol{v} \in V, 1_V\boldsymbol{v} = \boldsymbol{v}$.

It is worth noting that the "vectors" here might not refer to actual vectors in $\mathbb{R}^n$, i.e., these could be abstract structures which behave like vectors. For example, consider the set of all polynomials of degrees at most n. One can check that this set fullfils all axioms from Definition 1.1.1.

Without establishing further theorems, there are already some interesting properties we can derive from the above definition. For instance:

Theorem 1.1.2 ▶ Uniqueness of the zero vector

For every vector space V, there is a unique zero vector $\mathbf{0}_V \in V$.

Proof. Suppose there exist $\mathbf{0}_V, \mathbf{0}'_V \in V$ both being zero vectors in V, then we have

$$\mathbf{v} + \mathbf{0}_V = \mathbf{v},$$
$$\mathbf{v} + \mathbf{0}'_V = \mathbf{v}$$

for all $\mathbf{v} \in V$. Therefore, $\mathbf{v} + \mathbf{0}_V = \mathbf{v} + \mathbf{0}'_V$ which means $\mathbf{0}_V = \mathbf{0}'_V$, i.e., $\mathbf{0}_V$ is unique. $\square$

Theorem 1.1.3 ▶ Uniqueness of the inverse element

For every vector space V and every vector $\mathbf{v} \in V$, the inverse of $\mathbf{v}$ is unique.

Proof. Suppose there exist $\mathbf{u}_1, \mathbf{u}_2 \in V$ such that both are inverses of $\mathbf{v}$, then we have

$$\mathbf{v} + \mathbf{u}_1 = 0_V = \mathbf{v} + \mathbf{u}_2.$$

Therefore, $\mathbf{u}_1 = \mathbf{u}_2 = -\mathbf{v}$, i.e., $-\mathbf{v}$ is unique. $\square$

One might be tempted to think that the multiplicative identity of a vector space is also unique, and produce the following proof:

Proof?

Suppose there exist scalars a and b such that $a\mathbf{v} = b\mathbf{v} = \mathbf{v}$ for all $\mathbf{v} \in V$, then we have

$$(a - b)\mathbf{v} = \mathbf{0}_V.$$

Therefore, $a = b$, and so the multiplicative identity is unique.

Unfortunately, the above proof is wrong, but can you see why?

Remark. Note that for the above proof to hold, one crucial condition is that there exists some non-zero $\mathbf{v} \in V$, which might not be true! What if $V = \{\mathbf{0}_V\}$, meaning that every element in V is now effectively $\mathbf{0}_V$? In that case, a and b can take arbitrary values, which makes the proof fail.

A Special Case: the Euclidean Space

We focus on a special type of vector spaces which is known as the Euclidean Space. We shall first give the definition.

Definition 1.1.4 ▶ Euclidean space

The Euclidean n-space is the set of all **ordered** n-tuple of **real numbers** $x = (x_1, x_2, x_3, \cdots, x_n)$, known as an n-vector, where x_i is called the i-th component or the i-th coordinate of x. Denoting the Euclidean n-space by $\mathbb{R}^n$, we have

$$\mathbb{R}^n = \{\, (x_1, x_2, x_3, \cdots, x_n) \; : \; x_i \in \mathbb{R}, \, i \in \mathbb{N}^+, \, i \leq n \,\}.$$

Note that we can write x either as a row vector:

$$\begin{bmatrix} x_1 & x_2 & x_3 & \cdots & x_n \end{bmatrix},$$

or a column vector:

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{bmatrix}.$$

We also define the following relations for Euclidean spaces:

Relations and operations in the Euclidean n-space

- Equality: $x = y$ if and only if $x_i = y_i$ for all i.
- Zero vector: the n-vector $0 = (0, 0, 0, \cdots, 0)$ is the zero vector of the Euclidean n-space.
- Scalar multiplication: for $c \in \mathbb{R}$, $cx = (cx_1, cx_2, cx_3, \cdot, cx_n)$.
- Inverse: $(-1)x = -x$ is the inverse of x.
- Vector Addition: $x + y = (x_1 + y_1, x_2 + y_2, x_3 + y_3, \cdots, x_n + y_n)$.

1.2   Vector Subspaces

Notice that a vector space is just an infinite set which satisfies the axioms listed in Definition 1.1.1. Therefore, just like any sets, we can define the notion of **vector subspaces** as the analogue of subsets.

> **Definition 1.2.1 ▶ Vector subspace**
>
> Let V be a vector space over the field K and let $W \subseteq V$. W is known as a **vector subspace** (or simply subspace) of V if it is a vector space over K under the operations defined in V.

The definition above may sound like a mouthful, but in reality, it suffices to check only two conditions for W to determine whether it is a subspace: first, check that W is **non-empty**; second, check that W is **closed under vector addition and scalar multiplication**.

To check that W is non-empty, we simply check that $\mathbf{0}_V \in W$. In fact, if $\mathbf{0}_V \notin W$, then W cannot be a subspace of V, because $\boldsymbol{w} + (-\boldsymbol{w}) \notin W$!

Concisely, the above conditions are written as:

> **Theorem 1.2.2 ▶ Conditions for a subspace**
>
> *set $W \subseteq V$ is a subspace if and only if*
> - $\mathbf{0}_V \in W$*, and*
> - *for all $\boldsymbol{w}_1, \boldsymbol{w}_2 \in W$, $\alpha\boldsymbol{w}_1 + \boldsymbol{w}_2 \in W$ for an arbitrary scalar α.*

Additionally, we can see that for any vector space V, there exist two trivial subspaces to it: the space V itself and the space containing only the zero vector, i.e., $\{\mathbf{0}_V\}$.

2

Orthogonality

2.1   Dot Product

Recall that in a Euclidean n-space, we can define the **dot product** between two vectors:

Definition 2.1.1 ▸ Dot Product

Let $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^n$, then the dot product between $\boldsymbol{u}$ and $\boldsymbol{v}$ is defined as

$$\boldsymbol{u} \cdot \boldsymbol{v} = \|\boldsymbol{u}\| \, \|\boldsymbol{v}\| \cos \theta,$$

where θ is the acute angle between $\boldsymbol{u}$ and $\boldsymbol{v}$.

In particular, if $\boldsymbol{u} = (u_1, u_2, u_3, \cdots, u_n)$, $\boldsymbol{v} = (v_1, v_2, v_3, \cdots, v_n)$, we have

$$\boldsymbol{u} \cdot \boldsymbol{v} = \sum_{i=1}^{n} u_i v_i.$$

Note that we can either view $\boldsymbol{u}$ and $\boldsymbol{v}$ as both row vectors or as both column vectors. Either way, we are able to obtain the following matrix representations:

Matrix representations of the dot product

If we write $\boldsymbol{u}$ and $\boldsymbol{v}$ both as row vectors, we have

$$\boldsymbol{u} \cdot \boldsymbol{v} = \begin{bmatrix} u_1 & u_2 & u_3 & \cdots & u_n \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ \vdots \\ v_n \end{bmatrix} = \boldsymbol{u}\boldsymbol{v}^{\mathrm{T}};$$

---

if we write $\boldsymbol{u}$ and $\boldsymbol{v}$ both as column vectors, we have

$$\boldsymbol{u} \cdot \boldsymbol{v} = \begin{bmatrix} u_1 & u_2 & u_3 & \cdots & u_n \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ \vdots \\ v_n \end{bmatrix} = \boldsymbol{u}^{\mathrm{T}} \boldsymbol{v}.$$

2.2   Orthogonal and Orthonormal Bases

Definition 2.2.1 ▶ Orthogonality

Let $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^n$, they are said to be **orthogonal** if $\boldsymbol{u} \cdot \boldsymbol{v} = \boldsymbol{0}$, denoted as $\boldsymbol{u} \perp \boldsymbol{v}$.

Definition 2.2.2 ▶ Orthogonal and orthonormal sets

Let $S \subseteq \mathbb{R}^n$, if the elements in S are **pairwise orthogonal**, then S is called an **orthogonal set**.

In particular, an orthogonal set S is called an **orthonormal set** if every element in S is a **unit vector**.

There are some trivial conclusions regarding orthogonal and orthonormal sets which shall be stated without proof:

Some trivial results

- Every orthonormal set is orthogonal.
- If S is an orthogonal set, then every set $T \in \mathcal{P}(S)$ is orthogonal.
- If S is an orthonormal set, the every set $T \in \mathcal{P}(S)$ is orthonormal.
- If S is an orthogonal set, then $S \cup \{\boldsymbol{0}\}$ is orthogonal.
- If S is an orthonormal set, then $\boldsymbol{0} \notin S$.

We can easily convert an orthogonal set of **non-zero** vectors into an orthonormal set:

Theorem 2.2.3 ▶ Normalisation

If S is an orthogonal set and $\boldsymbol{0} \notin S$, then the set

$$\left\{ \boldsymbol{t} \;:\; \boldsymbol{t} = \frac{\boldsymbol{s}}{\|\boldsymbol{s}\|}, \, \boldsymbol{s} \in S \right\}$$

is an orthonormal set.

It is obvious that a **non-zero** orthogonal set must also be a linearly independent set:

Theorem 2.2.4 ▶ Linear independency of orthonormal sets

Let $S \subseteq \mathbb{R}^n$ be a non-zero orthogonal set, then S is linearly independent.

Proof. Let s_i $(i = 1, 2, 3, \cdots, k)$ be the elements of S.
Suppose $\sum_{i=1}^{k} c_i s_i = \mathbf{0}$ for some sequence of real numbers $\{c_j\}_{j=1}^{k}$, then for all n $(1 \leq n \leq k)$, we have

$$0 = s_n \cdot \mathbf{0}$$
$$= s_n \cdot \left(\sum_{i=1}^{k} c_i s_i \right)$$
$$= \sum_{i=1}^{k} c_i (s_n \cdot s_i)$$
$$= c_n (s_n \cdot s_n)$$

Note that $s_n \neq \mathbf{0}$, so $c_n = 0$ for all n $(1 \leq n \leq k)$. Therefore, S is linearly independent.

$\square$

Corollary 2.2.5 ▶ Linear independency of orthogonal sets

Any orthonormal set is linearly independent.

With these preliminary definitions and theorems, we now define the notion of orthogonal and orthonormal bases.

Definition 2.2.6 ▶ Orthogonal and orthonormal bases

A basis S for a vector space V is called an orthogonal basis if it is orthogonal, and orthonormal basis if it is orthornormal.

There are several advantages of using orthogonal bases, one of which is the ease of finding coordinate vectors.

Theorem 2.2.7 ▶ Coordinate vectors under orthogonal bases

Let $S = \{ s_1, s_2, s_3, \cdots, s_k \}$ be an orthogonal basis for a vector space V. Let $v \in V$ be an

---

arbitrary vector, then

$$[\boldsymbol{v}]_S = \begin{bmatrix} \frac{\boldsymbol{v}\cdot\boldsymbol{s}_1}{\|\boldsymbol{s}_1\|^2} \\ \frac{\boldsymbol{v}\cdot\boldsymbol{s}_2}{\|\boldsymbol{s}_2\|^2} \\ \frac{\boldsymbol{v}\cdot\boldsymbol{s}_3}{\|\boldsymbol{s}_3\|^2} \\ \vdots \\ \frac{\boldsymbol{v}\cdot\boldsymbol{s}_k}{\|\boldsymbol{s}_k\|^2} \end{bmatrix}.$$

Proof. Since $V = \mathrm{span}(S)$, we can write $\boldsymbol{v} = \sum_{i=1}^{k} c_i \boldsymbol{s}_i$.

Let n be an integer such that $1 \leq n \leq k$, consider

$$\begin{aligned} \boldsymbol{s}_n \cdot \boldsymbol{v} &= \boldsymbol{s}_n \cdot \sum_{i=1}^{k} c_i \boldsymbol{s}_i \\ &= \sum_{i=1}^{k} c_i (\boldsymbol{s}_n \cdot \boldsymbol{s}_i) \\ &= c_n \|\boldsymbol{s}_n\|^2. \end{aligned}$$

Therefore, $c_n = \frac{\boldsymbol{v}\cdot\boldsymbol{s}_n}{\|\boldsymbol{s}_n\|^2}$ for all n, and so

$$[\boldsymbol{v}]_S = \begin{bmatrix} \frac{\boldsymbol{v}\cdot\boldsymbol{s}_1}{\|\boldsymbol{s}_1\|^2} \\ \frac{\boldsymbol{v}\cdot\boldsymbol{s}_2}{\|\boldsymbol{s}_2\|^2} \\ \frac{\boldsymbol{v}\cdot\boldsymbol{s}_3}{\|\boldsymbol{s}_3\|^2} \\ \vdots \\ \frac{\boldsymbol{v}\cdot\boldsymbol{s}_k}{\|\boldsymbol{s}_k\|^2} \end{bmatrix}.$$

$\square$

Based on Theorem 2.2.7, for every orthogonal set $S = \{\boldsymbol{s}_1, \boldsymbol{s}_2, \boldsymbol{s}_3, \cdots, \boldsymbol{s}_k\}$ and any $\boldsymbol{v} \in \mathrm{span}(S)$, we can write $\boldsymbol{v} = \sum_{i=1}^{k} \frac{\boldsymbol{v}\cdot\boldsymbol{s}_i}{\|\boldsymbol{s}_i\|^2} \boldsymbol{s}_i$.

2.3 Orthogonality in n-Dimensional Spaces

In 2D and 3D Euclidean spaces, we define the notion of perpendicularity as having an angle of $\frac{\pi}{2}$ between two vectors. Being perpendicular is actually a special case of orthogonality in 2D and 3D Euclidean spaces, which we can generalise to any vector spaces of dimension n.

> **Definition 2.3.1 ▶ Orthogonality in n-dimenional vector spaces**
>
> Let V be a subspace of U and $\boldsymbol{u} \in U$ be an arbitrary vector. We say that $\boldsymbol{u}$ is orthogonal to V if it is orthogonal to every vector in V, i.e.,
>
> $$\boldsymbol{u} \perp V \Leftrightarrow \forall \boldsymbol{v} \in V(\boldsymbol{u} \perp \boldsymbol{v}).$$

The following statement follows trivially from the definition above:

> **Theorem 2.3.2 ▶ A trivial observation**
>
> *Let V be a vector space and $W \subseteq V$. If $\boldsymbol{u} \perp V$, then $\boldsymbol{u} \perp W$.*

Definition 2.3.1 may seem not very useful, as it is not realistically possible to literally iterate through every vector in V (Note that V is infinite!) and check whether it is orthogonal to $\boldsymbol{u}$.

However, notice that for every $\boldsymbol{v} \in V$, we can write $\boldsymbol{v}$ as a linear combination of the vectors in some spanning set of V. Therefore, to check whether a vector $\boldsymbol{u}$ is orthogonal to V, it suffices to just check the orthogonality between $\boldsymbol{u}$ and each vector in any spanning set of V.

> **Theorem 2.3.3 ▶ Condition for orthogonality**
>
> *Let $V = \text{span}(S)$ and $\boldsymbol{u}$ be any vector, then $\boldsymbol{u} \perp V$ if and only if $\boldsymbol{u} \perp \boldsymbol{s}$ for all $\boldsymbol{s} \in S$.*
>
> *Proof.* Note that if $\boldsymbol{u} \perp V$, then by Theorem 2.3.2 it holds trivially that $\boldsymbol{u} \perp \boldsymbol{s}$ for all $\boldsymbol{s} \in S \subseteq V$.
> We then prove that if $\boldsymbol{u} \perp \boldsymbol{s}$ for all $\boldsymbol{s} \in S$, then $\boldsymbol{u} \perp V$.
> Write $S = \{\boldsymbol{s}_1, \boldsymbol{s}_2, \boldsymbol{s}_3, \cdots, \boldsymbol{s}_n\}$. Let $\boldsymbol{v}$ be any vector in V. Note that we can write $\boldsymbol{v} = \sum_{i=1}^{n} c_i \boldsymbol{s}_i$ for some sequence of real numbers $\{c_i\}_{i=1}^{n}$. Consider
>
> $$
> \begin{aligned}
> \boldsymbol{u} \cdot \boldsymbol{v} &= \boldsymbol{u} \cdot \sum_{i=1}^{n} c_i \boldsymbol{s}_i \\
> &= \sum_{i=1}^{n} c_i \boldsymbol{u} \cdot \boldsymbol{s}_i \\
> &= 0.
> \end{aligned}
> $$
>
> Therefore, $\boldsymbol{u} \perp V$. $\qquad\square$

2.4   Projection

2.4.1   Projections in n-Dimensional Spaces

Similar to perpendicularity, we can also extend and generalise the notion of projections from 2D and 3D Euclidean spaces to any n-dimensional vector spaces.

Definition 2.4.1 ▶ Projection

Let V be a subspace of $\mathbb{R}^n$ and $\boldsymbol{w} \in \mathbb{R}^n$ be a vector. The **projection** of $\boldsymbol{w}$ in V is the vector $\boldsymbol{p} \in V$ such that $(\boldsymbol{w} - \boldsymbol{p}) \perp V$.

Just like how we can compute projections in 2D and 3D Euclidean spaces, we also want to derive a general formula to compute the projection of some vector onto a vector space. We achieve this by utilising the **orthogonal basis**.

Theorem 2.4.2 ▶ General formula for projections

Let $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \cdots, \boldsymbol{v}_k\}$ be an orthogonal basis for a vector space $V \subseteq \mathbb{R}^n$. The projection of $\boldsymbol{w} \in \mathbb{R}^n$ onto V is given by

$$\sum_{i=1}^{k} \frac{\boldsymbol{w} \cdot \boldsymbol{v}_i}{\|\boldsymbol{v}_i\|^2} \boldsymbol{v}_i.$$

Proof. Let $\boldsymbol{p} \in V$ be the projection of $\boldsymbol{w}$ onto V, then there is some $\boldsymbol{n} \perp V$ such that $\boldsymbol{w} = \boldsymbol{p} + \boldsymbol{n}$.

For $i = 1, 2, 3, \cdots, k$, note that $(\boldsymbol{w} - \boldsymbol{p}) \cdot \boldsymbol{v}_i = \boldsymbol{n} \cdot \boldsymbol{v}_i = 0$, so we have $\boldsymbol{w} \cdot \boldsymbol{v}_i = \boldsymbol{p} \cdot \boldsymbol{v}_i$.

By Theorem 2.2.7, we have

$$\begin{aligned}
\boldsymbol{p} &= \sum_{i=1}^{k} \frac{\boldsymbol{p} \cdot \boldsymbol{v}_i}{\|\boldsymbol{v}_i\|^2} \boldsymbol{v}_i \\
&= \sum_{i=1}^{k} \frac{\boldsymbol{w} \cdot \boldsymbol{v}_i}{\|\boldsymbol{v}_i\|^2} \boldsymbol{v}_i.
\end{aligned}$$

$\square$

Corollary 2.4.3 ▶ Alternative formula using orthonormal bases

Let $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \cdots, \boldsymbol{v}_k\}$ be an orthonormal basis for a vector space $V \subseteq \mathbb{R}^n$. The projection of $\boldsymbol{w} \in \mathbb{R}^n$ onto V is given by

$$\sum_{i=1}^{k} (\boldsymbol{w} \cdot \boldsymbol{v}_i) \boldsymbol{v}_i.$$

2.4.2   Gram-Schmidt Process

Now we have derived a useful formula to compute the projection of any vector onto a vector space given an orthogonal basis for that vector space. The remaining question is: can we find a way to acquire an orthogonal basis for any vector space?

We can consider some simple cases to get some intuition:

- Suppose $V_1 = \operatorname{span}\{\boldsymbol{u}_1\}$, then $\{\boldsymbol{u}_1\}$ is already orthogonal.

- Suppose $V_2 = \operatorname{span}\{\boldsymbol{u}_1, \boldsymbol{u}_2\}$. We can first find the projection of $\boldsymbol{u}_2$ onto $\boldsymbol{u}_1$ as $\boldsymbol{p} = \frac{\boldsymbol{u}_1 \cdot \boldsymbol{u}_2}{\|\boldsymbol{u}_1\|^2}\boldsymbol{u}_1$. Note that $\{\boldsymbol{u}_1, \boldsymbol{u}_2 - \boldsymbol{p}\}$ is an orthogonal set, so it is an orthogonal basis for V_2.

- Suppose $V_3 = \operatorname{span}\{\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3\}$. Similar to the previous case, we can first find the projection of $\boldsymbol{u}_3$ onto $\operatorname{span}\{\boldsymbol{u}_1, \boldsymbol{u}_2\} = \operatorname{span}\{\boldsymbol{u}_1, \boldsymbol{u}_2 - \boldsymbol{p}\}$ as $\boldsymbol{q} = \frac{\boldsymbol{u}_1 \cdot \boldsymbol{u}_3}{\|\boldsymbol{u}_1\|^2}\boldsymbol{u}_3 + \frac{(\boldsymbol{u}_2 - \boldsymbol{p}) \cdot \boldsymbol{u}_3}{\|\boldsymbol{u}_2 - \boldsymbol{p}\|^2}$. Note that $\{\boldsymbol{u}_1, \boldsymbol{u}_2 - \boldsymbol{p}, \boldsymbol{u}_3 - \boldsymbol{q}\}$ is an orthogonal set, so it is an orthogonal basis for V_3.

Notice that a recursive pattern starts to form, which we can describe as follows:

A recursive algorithm (with wishful thinking)

To find an orthogonal basis for $\operatorname{span}\{\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3, \cdots, \boldsymbol{u}_k\}$:

1. Suppose we already have an orthogonal basis $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \cdots, \boldsymbol{v}_{k-1}\}$ for $\operatorname{span}\{\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3, \cdots, \boldsymbol{u}_{k-1}\}$.

2. We obtain the projection of $\boldsymbol{u}_k$ onto $\operatorname{span}\{\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3, \cdots, \boldsymbol{u}_k\}$ by Theorem 2.4.2:

$$\boldsymbol{p} = \sum_{i=1}^{k-1} \frac{\boldsymbol{u}_k \cdot \boldsymbol{v}_i}{\|\boldsymbol{v}_i\|^2}\boldsymbol{v}_i.$$

3. Now the vector $\boldsymbol{u}_k - \boldsymbol{p}$ is orthogonal to $\operatorname{span}\{\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \cdots, \boldsymbol{v}_{k-1}\}$. Setting $\boldsymbol{v}_k = \boldsymbol{u}_k - \boldsymbol{p}$, we have $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \cdots, \boldsymbol{v}_k\}$ to be an orthogonal basis for $\operatorname{span}\{\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3, \cdots, \boldsymbol{u}_k\}$.

4. However, in order to obtain the orthogonal set $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \cdots, \boldsymbol{v}_{k-1}\}$, we need to find $\boldsymbol{v}_{k-1}$ while assuming we already have an orthogonal basis $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \cdots, \boldsymbol{v}_{k-2}\}$ for $\operatorname{span}\{\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3, \cdots, \boldsymbol{u}_{k-2}\}$.

5. The recursion continues until we reach $\{\boldsymbol{u}_1\}$ which is trivially orthogonal, where we stop and compute backwards to arrive at the final result.

The above algorithm can be described more rigorously as the **Gram-Schmidt process**.

> ### Theorem 2.4.4 ▶ Gram-Schmidt process
>
> *Let $V_k = \operatorname{span}\{\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3, \cdots, \boldsymbol{u}_k\}$. Define*
>
> $$\boldsymbol{v}_1 = \boldsymbol{u}_1,$$
>
> $$\boldsymbol{v}_n = \boldsymbol{u}_n - \sum_{i=1}^{n-1} \frac{\boldsymbol{u}_n \cdot \boldsymbol{v}_i}{\|\boldsymbol{v}_i\|^2},$$
>
> *then $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \cdots, \boldsymbol{v}_k\}$ is an orthogonal basis for V_k.*
> *In particular, take $\boldsymbol{w}_n = \frac{\boldsymbol{v}_n}{\|\boldsymbol{v}_n\|^2}$, then $\{\boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{w}_3, \cdots, \boldsymbol{w}_k\}$ is an orthonormal basis for V_k.*

Note that the Gram-Schmidt process can be abstracted as a transition from an arbitrary basis to an orthonormal basis for a given vector space. This means we can also find some transition matrix between any basis for a given vector space and an orthonormal basis for said space, which leads to the following theorem.

> ### Theorem 2.4.5 ▶ Decomposition
>
> *Let $\boldsymbol{A}$ be an $m \times n$ matrix whose columns are linearly independent, then there exist an $m \times n$ matrix $\boldsymbol{Q}$ whose columns form an orthonormal set and an $n \times n$ **upper-triangular** matrix $\boldsymbol{R}$ such that $\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{R}$.*
>
> ---
>
> *Proof.* Let the i-th column of $\boldsymbol{A}$ be $\boldsymbol{u}_i$, then $\{\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3, \cdots, \boldsymbol{u}_n\}$ is a basis for some vector space V. By Gram-Schmidt process, there is some orthonormal basis $\{\boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{w}_3, \cdots, \boldsymbol{w}_n\}$ that spans V such that $\operatorname{span}\{\boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{w}_3, \cdots, \boldsymbol{w}_j\} = \operatorname{span}\{\boldsymbol{u}_1, \boldsymbol{u}_2, \boldsymbol{u}_3, \cdots, \boldsymbol{u}_j\}$ for $j = 1, 2, 3, \cdots, n$. Therefore, we can write
>
> $$\boldsymbol{u}_j = \sum_{i=1}^{j} c_{ij} \boldsymbol{w}_i \qquad (*)$$

for $j = 1, 2, 3, \cdots, n$. Notice that equation $(*)$ can be written as matrix multiplication:

$$u_j = \begin{bmatrix} w_1 & w_2 & w_3 & \cdots & w_n \end{bmatrix} \begin{bmatrix} c_{1j} \\ c_{2j} \\ \vdots \\ c_{jj} \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

Therefore, we obtain the following expression:

$$A = \begin{bmatrix} u_1 & u_2 & u_3 & \cdots & u_n \end{bmatrix}$$

$$= \begin{bmatrix} w_1 & w_2 & w_3 & \cdots & w_n \end{bmatrix} \begin{bmatrix} c_{11} & c_{12} & c_{13} & \cdots & c_{1n} \\ 0 & c_{22} & c_{23} & \cdots & c_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & c_{nn} \end{bmatrix}$$

$$= QR,$$

where Q is an matrix whose columns form an orthonormal set and R is an upper-triangular matrix. $\qquad\square$

2.4.3   Least Squares Solutions

Recall that in 2D and 3D Euclidean spaces, we can find the distance between two points using the Pythagoras' Theorem. In terms of vectors, the distance can be viewed as the **modulus of the difference** between the position vectors of the corresponding points. With that, now we are ready to abstract further the notion of distance to any n-dimensional vector space.

Definition 2.4.6 ▸ Distance

Let S be a vector space with $\dim(S) = n$, and let $u, v \in S$. The **distance** between u and v is defined as

$$d(u, v) = \|u - v\|.$$

A classic problem in mathematics is the optimisation problem, one of the scenarios of

which is the minimisation of the distance between to objects. We introduce the following definition regarding minimising distances:

> **Definition 2.4.7 ▶ Best approximation**
>
> Let S, V be vectors spaces such that $V \subseteq S$. The **best approximation** of $\boldsymbol{u} \in S$ in V is the vector $\boldsymbol{p} \in V$ such that
>
> $$d(\boldsymbol{u}, \boldsymbol{p}) \leq d(\boldsymbol{u}, \boldsymbol{v}) \quad \text{for all } \boldsymbol{v} \in V.$$

A result that follows immediately from the above definition is:

> **Theorem 2.4.8 ▶ Uniqueness of best approximation**
>
> *If $\boldsymbol{p} \in V \subseteq S$ is a best approximation of $\boldsymbol{u} \in S$ in V, then $\boldsymbol{p}$ is unique.*
>
> *Proof.* Suppose there is some $\boldsymbol{p}' \in V$ which is another best approximation of $\boldsymbol{u}$ in V. By Definition 2.4.7, we have
>
> $$d(\boldsymbol{u}, \boldsymbol{p}) \leq d(\boldsymbol{u}, \boldsymbol{v}),$$
> $$d(\boldsymbol{u}, \boldsymbol{p}') \leq d(\boldsymbol{u}, \boldsymbol{v})$$
>
> for all $\boldsymbol{v} \in V$. Since $\boldsymbol{p}, \boldsymbol{p}' \in V$, it follows that
>
> $$d(\boldsymbol{u}, \boldsymbol{p}) \leq d(\boldsymbol{u}, \boldsymbol{p}'),$$
> $$d(\boldsymbol{u}, \boldsymbol{p}') \leq d(\boldsymbol{u}, \boldsymbol{p}),$$
>
> which means $d(\boldsymbol{u}, \boldsymbol{p}) = d(\boldsymbol{u}, \boldsymbol{p}')$.
> Since $\boldsymbol{p} \neq \boldsymbol{p}'$, we have $\boldsymbol{u} - \boldsymbol{p}' = (\boldsymbol{u} - \boldsymbol{p}) + (\boldsymbol{p} - \boldsymbol{p}')$. It then follows that
>
> $$\begin{aligned}
> d(\boldsymbol{u}, \boldsymbol{p}') &= \|\boldsymbol{u} - \boldsymbol{p}'\| \\
> &= \|(\boldsymbol{u} - \boldsymbol{p}) + (\boldsymbol{p} - \boldsymbol{p}')\| \\
> &= \sqrt{[(\boldsymbol{u} - \boldsymbol{p}) + (\boldsymbol{p} - \boldsymbol{p}')] \cdot [(\boldsymbol{u} - \boldsymbol{p}) + (\boldsymbol{p} - \boldsymbol{p}')]} \\
> &= \sqrt{\|\boldsymbol{u} - \boldsymbol{p}\|^2 + 2(\boldsymbol{u} - \boldsymbol{p}) \cdot (\boldsymbol{p} - \boldsymbol{p}') + \|\boldsymbol{p} - \boldsymbol{p}'\|^2} \\
> &= \sqrt{\|\boldsymbol{u} - \boldsymbol{p}\|^2 + \|\boldsymbol{p} - \boldsymbol{p}'\|^2} \\
> &\geq \|\boldsymbol{u} - \boldsymbol{p}\| \\
> &= d(\boldsymbol{u}, \boldsymbol{p}),
> \end{aligned}$$
>
> which is a contradiction, so $\boldsymbol{p}' = \boldsymbol{p}$, i.e., $\boldsymbol{p}$ is unique. $\qquad\square$

Suppose there exists a $p \in V$ which is the best approximation of u in V, what we are interested now is to compute this best approximation. With some manipulation, we realise that this best approximation p is exactly the **projection** of u onto V!

Theorem 2.4.9 ▶ Projection as best approximation

Let S, V be vectors spaces such that $V \subseteq S$. Let $u \in S$ and $p \in V$ be vectors. If p is the projection of u onto V, then p is the best approximation of u in V.

Proof. This is equivalent to proving that $d(u, p) \leq d(u, v)$ for all $v \in V$.
Let $n = u - p$ and $x = u - v$ for an arbitrary $v \in V$. Note that $n \perp V$. Set $w = p - v \in V$, then $n \cdot w = 0$ and $n = w = x$. Consider

$$
\begin{aligned}
d(u, v) &= \|x\| \\
&= \sqrt{(n + w) \cdot (n + w)} \\
&= \sqrt{\|n\|^2 + 2(n \cdot w) + \|w\|^2} \\
&\geq \|n\| \\
&= d(u, p).
\end{aligned}
$$

Therefore, p is the best approximation of u in V. □

Best approximations are useful in dealing with situations where an exact solution is not attainable. For example, suppose the linear system $Ax = b$ is consistent, then this is equvalent to saying that there is some x such that $Ax - b = 0$, i.e., $\|b - Ax\| = 0$. Note that this also means that b is in the column space of A. Set this column space as V.

Now, consider the situation where $b \notin V$. In this case, the linear system $Ax = b$ is inconsistent and thus unsolvable. However, can we still find an x such that Ax produces a vector which is **closest** to b? In other words, what is the vector x for the error term $\|b - Ax\|$ to be minimised?

Definition 2.4.10 ▶ Least Squares Solution

Let A be an $m \times n$ matrix and b be an $m \times 1$ matrix. A **least squares solution** to the linear system $Ax = b$ is an $n \times 1$ matrix u such that $\|b - Au\| \leq \|b - Av\|$ for all $n \times 1$ matrices v.

Note that Ax is a vector in the column space of A and $\|b - Ax\| = d(b, Ax)$. Naturally, this least squares solution u is such that Au is the best approximation of b in the column space of A!

Theorem 2.4.11 ▸ Finding the least squares solution to $Ax = b$

Let A be an $m \times n$ matrix with column space V and b be an $m \times 1$ matrix. Let p be the projection of b onto V, then u is a least squares solution to the linear system $Ax = b$ if and only if u is a solution to the linear system $Ax = p$.

Proof. Since p is the projection of b onto V, by Theorem 2.4.9 and Definition 2.4.7 we have:

$$d(b, p) \leq d(b, w) \quad \text{for all } w \in V.$$

Note that $V = \{Av : v \in \mathcal{M}_{n \times 1}\}$, where $\mathcal{M}_{n \times 1}$ is the set of all $n \times 1$ matrices, so for all $w \in V$, there is some $v \in \mathcal{M}_{n \times 1}$ such that $w = Av$. Therefore, we have:

$$d(b, p) \leq d(b, Av) \quad \text{for all } v \in \mathcal{M}_{n \times 1}.$$

We first prove that if u is a solution to the linear system $Ax = p$, then u is a least squares solution to the linear system $Ax = b$.
Write $Au = p$, then we have:

$$
\begin{aligned}
\|b - Au\| &= d(b, p) \\
&\leq d(b, Av) \\
&= \|b - Av\|,
\end{aligned}
$$

which means that u is the least squares solution to the linear system $Ax = b$.
We then prove that if u is a least squares solution to the linear system $Ax = b$, then u is a solution to the linear system $Ax = p$.
By Definition 2.4.10, $\|b - Au\| \leq \|b - Av\|$ for all $v \in \mathcal{M}_{n \times 1}$. Note that $Au, Av \in V$, so Au is the projection of b onto V, i.e. $Au = p$. Therefore, u is a solution to the linear system $Ax = p$. □

Thus, we can conclude the following algorithm for finding least squares solutions:

A general algorithm for finding a least squares solution to $Ax = b$

1. Check if the linear system $Ax = b$ is consistent. If the system is consistent, then any solution to it is a least squares solution.
2. If the system is inconsistent, find an orthogonal basis for the column space of A.
3. Find the projection p of b onto the column space of A. Note that the linear system $Ax = p$ is consistent as p is in the column space of A.
4. Solve the linear system $Ax = p$ to find a least squares solution to $Ax = b$.

Note that in the above algorithm, we have established a way to derive a least square solution

using the projection. This gives us motivation to consider the reverse: can we derive the projection of some vector $\boldsymbol{b}$ onto the column space of some matrix $\boldsymbol{A}$ given a least squares solution to $\boldsymbol{Ax} = \boldsymbol{b}$? To do that, we first introduce the following alternative method to compute a least squares solution:

Theorem 2.4.12 ▶ An alternative way to find the least squares solution to $\boldsymbol{Ax} = \boldsymbol{b}$

Let $\boldsymbol{A}$ be an $m \times n$ matrix with column space V and $\boldsymbol{b}$ be an $m \times 1$ matrix. A vector $\boldsymbol{u}$ is the least squares solution to the linear system $\boldsymbol{Ax} = \boldsymbol{b}$ if and only if $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{Au} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}$.

Proof. Let $\boldsymbol{a}_i$ be the i-th column of $\boldsymbol{A}$, so $V = \mathrm{span}\{\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3, \cdots, \boldsymbol{a}_n\}$ is the column space of $\boldsymbol{A}$.

By Theorem 2.4.11, $\boldsymbol{u}$ is a least squares solution to $\boldsymbol{Ax} = \boldsymbol{b}$ if and only if $\boldsymbol{Au}$ is the projection of $\boldsymbol{b}$ onto V, i.e., $\boldsymbol{Au} - \boldsymbol{b} \perp V$. By Theorem 2.3.3, this is equivalent to

$$\boldsymbol{a}_i \cdot (\boldsymbol{Au} - \boldsymbol{b}) = 0 \quad \text{for } i = 1, 2, 3, \cdots, n.$$

The above dot product is equivalent to

$$\boldsymbol{0} = \begin{bmatrix} \boldsymbol{a}_1^{\mathrm{T}} \\ \boldsymbol{a}_2^{\mathrm{T}} \\ \vdots \\ \boldsymbol{a}_n^{\mathrm{T}} \end{bmatrix} (\boldsymbol{Au} - \boldsymbol{b})$$

$$= \boldsymbol{A}^{\mathrm{T}}(\boldsymbol{Au} - \boldsymbol{b}),$$

which simplifies to

$$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{Au} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}.$$

□

Remark. Be mindful that in Theorem 2.4.12, $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{Au} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}$ **does not imply** that $\boldsymbol{Au} = \boldsymbol{b}$ because $\boldsymbol{A}$ may be **uninvertible**!

Recall that if $\boldsymbol{u}$ is a least squares solution to $\boldsymbol{Ax} = \boldsymbol{b}$, then $\boldsymbol{Au}$ is the projection of $\boldsymbol{b}$ onto the column space of $\boldsymbol{A}$, so we have found another way to compute the projection of a vector onto a vector space.

> **An alternative algorithm to compute the projection of b onto V**
>
> 1. Write $V = \operatorname{span}\{a_1, a_2, a_3, \cdots, a_n\}$ for some arbitrary spanning set.
> 2. Construct a matrix A with a_i as its i-th column.
> 3. Find a solution u to the linear system $A^{\mathrm{T}}Au = A^{\mathrm{T}}b$.
> 4. The projection of b onto V is given by $p = Au$.

Furthermore, we can also consider a special case where the columns of A forms an **orthonormal** set.

> **Theorem 2.4.13 ▶ Least squares solution with orthonormal sets**
>
> *If A is a matrix whose columns form an orthonormal set, then a least squares solution to $Ax = b$ is given by $u = A^{\mathrm{T}}b$.*
>
> *Proof.* Let $A = \begin{bmatrix} a_1 & a_2 & a_3 & \cdots & a_n \end{bmatrix}$, then
>
> $$a_i \cdot a_j = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}.$$
>
> Consider
>
> $$A^{\mathrm{T}}A = \begin{bmatrix} a_1^{\mathrm{T}} \\ a_2^{\mathrm{T}} \\ a_3^{\mathrm{T}} \\ \vdots \\ a_n^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} a_1 & a_2 & a_3 & \cdots & a_n \end{bmatrix}$$
>
> $$= \begin{bmatrix} a_1 \cdot a_1 & a_1 \cdot a_2 & \cdots & a_1 \cdot a_n \\ a_2 \cdot a_1 & a_2 \cdot a_2 & \cdots & a_2 \cdot a_n \\ \vdots & \vdots & \ddots & \vdots \\ a_n \cdot a_1 & a_n \cdot a_2 & \cdots & a_n \cdot a_n \end{bmatrix}$$
>
> $$= I.$$
>
> Let u be a least squares solution to $Ax = b$, then $u = Iu = A^{\mathrm{T}}Au = A^{\mathrm{T}}b$. $\qquad\square$

The above theorem demonstrates yet another advantage of an orthonormal set, which leads us to the next section.

2.5   Orthogonal Matrices

> **Definition 2.5.1 ▶ Orthogonal matrix**
>
> A **square** matrix A is called an **orthogonal matrix** if $A^{\mathrm{T}}A = I$.

> *Remark.* There are a few things worth noting from the above definition.
>   1. The columns of an orthogonal matrix forms an **orthonormal** set.
>   2. Equivalently, a square matrix A is an orthogonal matrix if $A^{\mathrm{T}} = A^{-1}$.
>   3. An $n \times n$ orthogonal matrix A is always full-rank, i.e., $\operatorname{rank}(A) = n$.

We also have some simple properties regarding orthogonal matrices:

> **Theorem 2.5.2 ▶ Transposes of orthogonal matrices are orthogonal**
>
> *If A is an orthogonal matrix, then A^{T} is an orthogonal matrix.*
>
> *Proof.* Let $B = A^{\mathrm{T}}$, then $BA = I$. This means $A = B^{-1}$. However, $A = B^{\mathrm{T}}$, so $B^{\mathrm{T}} = B^{-1}$, i.e., A^{T} is an orthogonal matrix. □

> **Theorem 2.5.3 ▶ Product of orthogonal matrices is orthogonal**
>
> *If A, B are orthogonal matrices, then AB is an orthogonal matrix.*
>
> *Proof.* Consider $(AB)^{\mathrm{T}}(AB) = B^{\mathrm{T}}A^{\mathrm{T}}AB = B^{\mathrm{T}}B = I$, so AB is an orthogonal matrix. □

Notice that the both the rows and the columns of an orthogonal matrix are linearly independent, i.e., both form an **orthonormal basis** for a vector space.

> **Theorem 2.5.4 ▶ Rows and columns of orthogonal matrices**
>
> *If A is an $n \times n$ orthogonal matrix, then the rows and columns of A respectively form an orthonormal set.*
>
> *Proof.* Write
> $$A = \begin{bmatrix} c_1 & c_2 & \cdots & c_n \end{bmatrix}$$
> where c_i is the i-th column of A. Consider
> $$A^{\mathrm{T}}A = (c_i \cdot c_j)_{n \times n} = I,$$

which is equivalent to

$$c_i \cdot c_j = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}.$$

Therefore, the columns of A forms an orthonormal set.

Write

$$A = \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_n \end{bmatrix}$$

where r_i is the i-th row of A. Consider

$$AA^{\mathrm{T}} = (r_i \cdot r_j)_{n \times n} = I,$$

which is equivalent to

$$r_i \cdot r_j = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}.$$

Therefore, the rows of A forms an orthonormal set. $\square$

Corollary 2.5.5 ▶ A generalised statement

Let A be an $m \times n$ matrix. The columns of A form an orthonormal set if and only if $A^{\mathrm{T}}A = I_n$, and the rows of A form an orthonormal set if and only if $AA^{\mathrm{T}} = I_m$.

Using the above theorems, we can easily establish a bijection the matches each orthonormal basis to a matrix. The next question is how do we then establish a transition matrix between two orthonormal bases.

Theorem 2.5.6 ▶ Transition matrix between orthonormal bases

Let A and B be matrices whose columns form the orthonormal bases for S and T respectively for the same vector space V, then $P = B^{\mathrm{T}}A$ is the transition matrix from S to T and $Q = A^{\mathrm{T}}B$ is the transition matrix from T to S.

Proof. Take $v \in V$ to be an arbitrary vector. Note that $v = A[v]_S = B[v]_T$.
By Corollary 2.5.5, $B^{\mathrm{T}}B = A^{\mathrm{T}}A = I$, so we have:

$$[v]_T = B^{\mathrm{T}}B[v]_T = B^{\mathrm{T}}A[v]_S$$
$$[v]_S = A^{\mathrm{T}}A[v]_S = A^{\mathrm{T}}B[v]_T$$

Let $P = B^{\mathrm{T}}A$ and $Q = A^{\mathrm{T}}B$, then P, Q are the transition matrices from S to T and from T to S respectively. □

Theorem 2.5.7 ▶ Orthogonal transition matrices

Let S and T be two orthonormal bases for a vector space V, and let P be the transition matrix from S to T, then P is an orthogonal matrix.

Proof. By Theorem 2.5.6, we can find matrices A and B whose columns form the bases S and T respectively such that $P = B^{\mathrm{T}}A$. Consider

$$P^{\mathrm{T}}P = \left(B^{\mathrm{T}}A\right)^{\mathrm{T}}\left(B^{\mathrm{T}}A\right) = A^{\mathrm{T}}BB^{\mathrm{T}}A = I,$$

so P is an orthogonal matrix. □

3

Linear Transformations

3.1   Linear Transformations in n-Dimensional Spaces

So far we have encountered many types of transformations. For example, translation, scaling, reflection and rotation are geometric transformations applied to shapes in Euclidean spaces; in the previous chapters we have also seen how to transform the coordinate system (i.e. the basis of a vector space) by means of transition matrices; in fact, integration and differentiation are also transformations which are applied to functions.

All those many instances where transformations appear share one thing in common — the transformations involved are **linear**. Though stemming from the root word "line", the notion of linearity goes far beyond the geometric representations of lines, which we shall define rigorously as follows:

Definition 3.1.1 ▶ Linear Transformation

A mapping between to vector spaces is called a **linear transformation** if it preserves the operations of vector addition and scalar multiplication, i.e., $T : V \to W$ is a linear transformation if

$$T(\alpha \boldsymbol{u} + \beta \boldsymbol{v}) = \alpha T(\boldsymbol{u}) + \beta T(\boldsymbol{v}) \quad \text{for } \boldsymbol{u}, \boldsymbol{v} \in V.$$

In particular, a linear transformation $T : V \to V$ is also called a **linear operator** on V.

Remark. The above definition can be concisely described as: a transformation is a linear transformation if **the image of linear combinations equals the linear combinations of the images**.

We also define two special linear transformations for every vector space:

Definition 3.1.2 ▶ Identity transformation

For every vector space V, there is a **identity transformation** $I : V \to V$ defined by

$$I(\boldsymbol{v}) = \boldsymbol{v} \quad \text{for all } \boldsymbol{v} \in V.$$

> **Definition 3.1.3 ▶ Zero transformation**
>
> For every vector space V, there is a **zero transformation** $O : V \to W$ defined by
>
> $$O(\boldsymbol{v}) = \mathbf{0}_W \quad \text{for all } \boldsymbol{v} \in V.$$

It is trivial that the identity transformation is **unique** for a vector space V.

> *Remark.* Note that the zero transformation is **not unique** because its image can be the zero vector in **any** vector space.

So far, the notion of linear transformations still appear a bit abstract. Let us put it into more concrete forms with some manipulations.

Let $T : V \to W$ be a linear transformation where V and W are finite-dimensional. Let S, U be the bases for V and W respectively. Note that there is a **bijection** between V and the set $\{\,[\boldsymbol{v}]_S : \boldsymbol{v} \in V = \text{span}(S)\,\}$, as well as between W and the set $\{\,[\boldsymbol{w}]_U : \boldsymbol{w} \in W = \text{span}(U)\,\}$. This means that we can view T as the mapping $[\boldsymbol{v}]_S \mapsto [\boldsymbol{w}]_U$.

Based on Definition 3.1.1, we can immediately arrive at the following equation:

$$T(\boldsymbol{v}) = T\left(\sum_{j=1}^{n} c_j \boldsymbol{v}_j\right) = \sum_{j=1}^{n} c_j T(\boldsymbol{v}_j), \tag{$*$}$$

where the $\boldsymbol{v}_j$'s are the vectors in S and $\{c_j\}_{j=1}^{n}$ is a sequence of scalar coefficients. Note that the right-hand side of the equation is a linear combination, which means that the image of V under T is **spanned** by the **images of the vectors in the basis of V**. In other words, the image of the basis of V under T, i.e., the set $\{\,T(\boldsymbol{v}_1), T(\boldsymbol{v}_2), \cdots, T(\boldsymbol{v}_n)\,\}$ completely determines the output of $T(\boldsymbol{v})$ for any $\boldsymbol{v} \in V$.

Let $U = \{\,\boldsymbol{w}_1, \boldsymbol{w}_2, \cdots, \boldsymbol{w}_m\,\}$, then for each of the $\boldsymbol{v}_j$'s in the basis of V, we can write

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

for some sequence of scalar coefficients $\{a_{ij}\}_{i=1}^{m}$. Notice that now Equation $(*)$ can be fur-

ther expanded to

$$T(\boldsymbol{v}) = \sum_{j=1}^{n} c_j T(\boldsymbol{v}_j)$$

$$= \sum_{j=1}^{n} \left(c_j \sum_{i=1}^{m} a_{ij} \boldsymbol{w}_i \right)$$

$$= \sum_{j=1}^{n} \left(\sum_{i=1}^{m} a_{ij} c_j \boldsymbol{w}_i \right)$$

$$= \sum_{i=1}^{m} \left(\sum_{j=1}^{n} a_{ij} c_j \right) \boldsymbol{w}_i.$$

Suppose $T(\boldsymbol{v}) = \boldsymbol{w}$. Construct a matrix $\boldsymbol{M}$ in the form of

$$\boldsymbol{M} = \left[\begin{array}{cccc} [T(\boldsymbol{v}_1)]_U & [T(\boldsymbol{v}_2)]_U & \cdots & [T(\boldsymbol{v}_n)]_U \end{array} \right].$$

Note that $\boldsymbol{M}[\boldsymbol{v}]_S = \sum_{j=1}^{n} c_j [T(\boldsymbol{v}_j)]_U$ is precisely $[T(\boldsymbol{v})]_U$. Therefore, T can be described as $[\boldsymbol{v}]_S \mapsto \boldsymbol{M}[\boldsymbol{v}]_S$.

In particular, if we set both S and U to be the standard bases, we would have $\boldsymbol{v} = [\boldsymbol{v}]_S$ and $\boldsymbol{w} = \boldsymbol{w}_U$. In this case, we have:

$$T(\boldsymbol{v}) = \boldsymbol{M}\boldsymbol{v}.$$

In other words, every linear transformation can be represented with a matrix, which is summarised into the following theorem:

> **Theorem 3.1.4 ▶ Standard matrix**
>
> *Let $T : V \to W$ be a linear transformation between finite-dimensional vector spaces, and let S and U be bases for V and W respectively, then there exists a $\dim(W) \times \dim(V)$ matrix $\boldsymbol{M}$, known as the* **standard matrix**, *given by*
>
> $$\boldsymbol{M} = \left[\begin{array}{cccc} [T(\boldsymbol{v}_1)]_U & [T(\boldsymbol{v}_2)]_U & \cdots & [T(\boldsymbol{v}_n)]_U \end{array} \right],$$
>
> *where the $\boldsymbol{v}_i$'s are vectors in S, such that $[T(\boldsymbol{v})]_U = \boldsymbol{M}[\boldsymbol{v}]_S$ for all $\boldsymbol{v} \in V$. More specifically, if S and U are the standard bases, then $T(\boldsymbol{v}) = \boldsymbol{M}\boldsymbol{v}$.*

The above argument has proved the existence of the standard matrix for every linear transformation between finite-dimensional vector spaces. We can further prove the uniqueness of the standard matrix.

> **Theorem 3.1.5 ▶ Uniqueness of standard matrix**
>
> *Let M be the standard matrix for the linear transformation $T : V \to W$, then M is unique.*
>
> ---
>
> *Proof.* Let $\dim(V) = n$ and $\dim(W) = m$. Suppose there exist $m \times n$ matrices M and N such that $T(v) = Mv = Nv$, then we have:
>
> $$(M - N)v = 0_{m \times 1}$$
>
> for all $v \in V$. Specifically, let v_i ($i = 1, 2, 3, \cdots, n$) be the vectors which form the standard basis for V, i.e., the i1-th entry of v_i is 1 but all other entries are 0, then we have
>
> $$(M - N)v_i = 0_{m \times 1}$$
>
> for $i = 1, 2, 3, \cdots, n$. Re-writing the equations into a single matrix multiplication, we have
>
> $$(M - N)I_{n \times n} = 0_{m \times n}$$
>
> However, $(M - N)I_{n \times n} = M - N$, which means $M - N = 0_{m \times n}$, i.e., $M = N$. Therefore, M is unique. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

3.2   Change of Bases

In the previous section, we have used the coordinate vectors under bases for different vector spaces to establish the matrix representation of linear transformation. We can further expand on the notion of "changing the bases" onto some other applications.