

# Chapter 1

## Vector Spaces

### 1.1 Fields

**Definition 1.1.** A **field** is a set  $F$  with two operations, called **addition** (denoted by  $+$ ) and **multiplication** (denoted by  $\cdot$ ), which satisfy the following axioms.

- (A 1) If  $a \in F$  and  $b \in F$ , then  $a + b \in F$ .
- (A 2)  $a + b = b + a$  for all  $a, b \in F$ .
- (A 3)  $(a + b) + c = a + (b + c)$  for all  $a, b, c \in F$ .
- (A 4) There is an element  $0_F$  in  $F$  such that  $0_F + a = a$  for all  $a \in F$ .
- (A 5) For each  $a \in F$  there is an element  $-a$  in  $F$  such that  $a + (-a) = 0_F$ .
- (M 1) If  $a \in F$  and  $b \in F$ , then  $a \cdot b \in F$ .
- (M 2)  $a \cdot b = b \cdot a$  for all  $a, b \in F$ .
- (M 3)  $(a \cdot b) \cdot c = a \cdot (b \cdot c)$  for all  $a, b, c \in F$ .
- (M 4) There is an element  $1_F$  in  $F \setminus \{0_F\}$  such that  $1_F \cdot a = a$  for all  $a \in F$ .
- (M 5) For each  $a \in F \setminus \{0_F\}$  there is an element  $a^{-1}$  in  $F$  such that  $a \cdot a^{-1} = 1_F$ .
- (D)  $a \cdot (b + c) = a \cdot b + a \cdot c$  for all  $a, b, c \in F$ .

**Remark.**

- For simplification, we usually write  $ab$  instead of  $a \cdot b$ .
- The axioms labeled with “A” and “M” are usually called the **axioms of addition** and the **axioms of multiplication**, respectively. The axiom labeled with “D” is the **distributive law**.
- The elements  $0_F$  and  $1_F$  are usually called the **additive identity** and the **multiplicative identity** of  $F$ , respectively. Also,  $-a$  and  $a^{-1}$  are called the **additive inverse** and the **multiplicative inverse** of  $a$ , respectively.
- **Subtraction** and **division** can be defined using additive and multiplicative inverses.

**Example.**  $\mathbb{Q}$ ,  $\mathbb{R}$  and  $\mathbb{C}$  are fields.

**Example.** Let  $\mathbb{B} = \{0, 1\}$  and the operations  $\oplus$  and  $\odot$  are defined as follows.

$$\begin{array}{c|cc} \oplus & 0 & 1 \\ \hline 0 & 0 & 1 \\ 1 & 1 & 0 \end{array} \quad \begin{array}{c|cc} \odot & 0 & 1 \\ \hline 0 & 0 & 0 \\ 1 & 0 & 1 \end{array}$$

Then  $\mathbb{B}$  is a field with  $\oplus$  and  $\odot$  as addition and multiplication, respectively.

**Proposition 1.2.** Let  $F$  be a field with  $a, b, c \in F$ .

- (a) If  $a + b = a + c$ , then  $b = c$ .
- (b) If  $a + b = a$ , then  $b = 0_F$ .
- (c) If  $a + b = 0_F$ , then  $b = -a$ .
- (d)  $-(-a) = a$ .

*Proof.*

- (a) It can be proved by

$$\begin{aligned} b &= 0_F + b \\ &= (-a + a) + b \\ &= -a + (a + b) \\ &= -a + (a + c) \\ &= (-a + a) + c \\ &= 0_F + c \\ &= c. \end{aligned}$$

- (b) By applying (a), it follows from  $a + b = a + 0_F$  that  $b = 0_F$ .
- (c) By applying (a), it follows from  $a + b = a + (-a)$  that  $b = -a$ .
- (d) Since  $-a + a = 0_F$ , we have  $a = -(-a)$  by (c). □

**Proposition 1.3.** Let  $F$  be a field with  $a, b, c \in F$  and  $a \neq 0_F$ .

- (a) If  $a \cdot b = a \cdot c$ , then  $b = c$ .
- (b) If  $a \cdot b = a$ , then  $b = 1_F$ .
- (c) If  $a \cdot b = 1_F$ , then  $b = a^{-1}$ .
- (d)  $(a^{-1})^{-1} = a$ .

*Proof.* The proof is omitted since it is similar to that of Proposition 1.2. □

**Proposition 1.4.** Let  $F$  be a field with  $a, b \in F$ .

- (a)  $0_F \cdot a = 0_F$ .

$$(b) \quad (-a) \cdot b = -(a \cdot b) = a \cdot (-b).$$

$$(c) \quad (-a) \cdot (-b) = a \cdot b.$$

*Proof.*

(a) Since

$$0_F \cdot a + 0_F \cdot a = (0_F + 0_F) \cdot a = 0_F \cdot a,$$

we have  $0_F \cdot a = 0_F$  by Proposition 1.2 (b).

(b) Since

$$(-a) \cdot b + a \cdot b = (-a + a) \cdot b = 0_F \cdot b = 0_F,$$

we have  $(-a) \cdot b = -(a \cdot b)$  by Proposition 1.2 (c). The other half can be proved similarly.

(c) By applying (b) twice, we have

$$(-a) \cdot (-b) = -(a \cdot (-b)) = -(-(a \cdot b)) = a \cdot b. \quad \square$$

## 1.2 Vector Spaces

**Definition 1.5.** A **vector space** over a field  $F$  is a set  $V$  with two operations, called **addition** (denoted by  $+$ ) and **scalar multiplication** (denoted by  $\cdot$ ), which satisfy the following axioms.

(V 1) If  $x \in V$  and  $y \in V$ , then  $x + y \in V$ .

(V 2)  $x + y = y + x$  for all  $x, y \in V$ .

(V 3)  $(x + y) + z = x + (y + z)$  for all  $x, y, z \in V$ .

(V 4) There is an element  $0_V$  in  $V$  such that  $0_V + x = x$  for all  $x \in V$ .

(V 5) For each  $x \in V$  there is an element  $-x$  such that  $x + (-x) = 0_V$ .

(V 6) If  $a \in F$  and  $x \in V$ , then  $a \cdot x \in V$ .

(V 7)  $(a \cdot b) \cdot x = a \cdot (b \cdot x)$  for all  $a, b \in F$  and  $x \in V$ .

(V 8)  $1_F \cdot x = x$  for all  $x \in V$ .

(V 9)  $a \cdot (x + y) = a \cdot x + a \cdot y$  for all  $a \in F$  and  $x, y \in V$ .

(V 10)  $(a + b) \cdot x = a \cdot x + b \cdot x$  for all  $a, b \in F$  and  $x \in V$ .

**Remark.**

- For simplification, we usually write  $ax$  instead of  $a \cdot x$ .
- The element  $0_V$  is usually called the **additive identity** of  $V$ , and  $-x$  is called the **additive inverse** of  $x$  in  $V$ .
- **Subtraction** can be defined using additive inverses.

**Examples.**

- A field is a vector space over itself, e.g.,  $\mathbb{R}$  is a vector space over  $\mathbb{R}$ .
- $\mathbb{C}$  is a vector space over  $\mathbb{R}$ .
- $\mathbb{R}$  is a vector space over  $\mathbb{Q}$ .

**Examples.**

- The set of  **$n$ -tuples** with elements from a field  $F$  is denoted by  $F^n$ . For  $x = (x_1, \dots, x_n) \in F^n$ ,  $y = (y_1, \dots, y_n) \in F^n$ , and  $c \in F$ , we define the operations of addition and scalar multiplication by

$$x + y = (x_1 + y_1, \dots, x_n + y_n) \quad \text{and} \quad c \cdot x = (c \cdot x_1, \dots, c \cdot x_n).$$

Then  $F^n$  is a vector space over  $F$ .

- The set of all  $m \times n$  **matrices** with elements from a field  $F$  is denoted by  $F^{m \times n}$ . For  $A, B \in F^{m \times n}$  and  $c \in F$ , we define the operations of addition and scalar multiplication by

$$(A + B)_{ij} = A_{ij} + B_{ij} \quad \text{and} \quad (c \cdot A)_{ij} = c \cdot A_{ij}$$

for  $i \in \{1, \dots, m\}$  and  $j \in \{1, \dots, n\}$ . Then  $F^{m \times n}$  is a vector space over  $F$ .

- The set of **functions** from a nonempty set  $S$  to a field  $F$  is denoted by  $\mathcal{F}(S, F)$ . For  $f, g \in \mathcal{F}(S, F)$  and  $c \in F$ , we define the operations of addition and scalar multiplication by

$$(f + g)(s) = f(s) + g(s) \quad \text{and} \quad (c \cdot f)(s) = c \cdot f(s)$$

for all  $s \in S$ . Then  $\mathcal{F}(S, F)$  is a vector space over  $F$ .

- The set of **polynomials** with coefficients from a field  $F$  is denoted by  $\mathcal{P}(F)$ . For  $f, g \in \mathcal{P}(F)$  and  $c \in F$  with

$$f(t) = \sum_{i=0}^n a_i t^i \quad \text{and} \quad g(t) = \sum_{i=0}^n b_i t^i,$$

we define the operations of addition and scalar multiplication by

$$(f + g)(t) = \sum_{i=0}^n (a_i + b_i) t^i \quad \text{and} \quad (c \cdot f)(t) = \sum_{i=0}^n (c \cdot a_i) t^i.$$

Then  $\mathcal{P}(F)$  is a vector space over  $F$ .

**Proposition 1.6.** Let  $V$  be a vector space with  $x, y, z \in V$ .

- (a) If  $x + y = x + z$ , then  $y = z$ .
- (b) If  $x + y = x$ , then  $y = 0_V$ .
- (c) If  $x + y = 0_V$ , then  $y = -x$ .
- (d)  $-(-x) = x$ .

*Proof.* The proof is omitted since it is similar to that of Proposition 1.2. □

**Proposition 1.7.** Let  $V$  be a vector space over a field  $F$  with  $x \in V$  and  $a \in F$ .

- (a)  $0_F \cdot x = 0_V$ .
- (b)  $a \cdot 0_V = 0_V$ .
- (c)  $(-a) \cdot x = -(a \cdot x) = a \cdot (-x)$ .

*Proof.* The proof is omitted since it is similar to that of Proposition 1.4. □

## 1.3 Subspaces

**Definition 1.8.** Let  $V$  be a vector space over a field  $F$ . Then a subset  $W$  of  $V$  is called a **subspace** of  $V$  if  $W$  is a vector space over  $F$  with the operations of addition and scalar multiplication defined on  $V$ .

**Theorem 1.9.** Let  $V$  be a vector space over a field  $F$  and  $W \subseteq V$ . Then  $W$  is a subspace of  $V$  if the following conditions hold.

- (a)  $0_V \in W$ .
- (b)  $x + y \in W$  for all  $x, y \in W$ .
- (c)  $ax \in W$  for all  $x \in W$  and  $a \in F$ .

*Proof.* Since a vector in  $W$  is also in  $V$ , (V 2), (V 3), (V 7), (V 8), (V 9) and (V 10) in Definition 1.5 hold trivially. Furthermore, (a) implies (V 4), (b) implies (V 1), (c) implies (V 6), and (V 5) is also true since

$$-x = -(1_F x) = (-1_F)x \in W$$

holds for all  $x \in W$ . Thus,  $W$  is a vector space over  $F$ . □

**Corollary 1.10.** Let  $V$  be a vector space over a field  $F$  and  $W \subseteq V$ . Then  $W$  is a subspace of  $V$  if and only if the following conditions hold.

- (a)  $0_V \in W$ .
- (b)  $ax + y \in W$  for all  $x, y \in W$  and  $a \in F$ .

*Proof.* ( $\Rightarrow$ ) Straightforward. ( $\Leftarrow$ ) For all  $x, y \in W$  and  $a \in F$ , we have

$$x + y = 1_F x + y \in W \quad \text{and} \quad ax = ax + 0_V \in W.$$

Thus,  $W$  is a subspace of  $V$  by Theorem 1.9. □

**Example.** The set of polynomials in  $\mathcal{P}(F)$  with degree not greater than  $n$  is denoted by  $\mathcal{P}_n(F)$ , where the **degree** of a nonzero polynomial

$$f(t) = a_0 + a_1 t + a_2 t^2 + \cdots + a_m t^m$$

is defined to be the largest integer  $n$  such that  $a_n \neq 0_F$ , and the degree of zero polynomial is defined to be  $-1$ . Then one can verify that  $\mathcal{P}_n(F)$  is a subspace of  $\mathcal{P}(F)$ .

**Examples.**

- An  $n \times n$  matrix  $A$  is called **diagonal** if  $A_{ij} = 0_F$  for all  $i, j \in \{1, \dots, n\}$  with  $i \neq j$ . Then one can verify that the set of  $n \times n$  diagonal matrices is a subspace of  $F^{n \times n}$ .
- The **trace** of an  $n \times n$  matrix  $A$ , denoted by  $\text{tr}(A)$ , is defined by

$$\text{tr}(A) = \sum_{i=1}^n A_{ii}.$$

Then one can verify that the set of  $n \times n$  matrices that have trace equal to  $0_F$  is a subspace of  $F^{n \times n}$ .

**Proposition 1.11.** Let  $V$  be a vector space and let  $W_1$  and  $W_2$  be subspaces of  $V$ . Then  $W_1 \cap W_2$  is a subspace of  $V$ .

*Proof.* Since  $W_1$  and  $W_2$  are subspaces of  $V$ , we have  $0_V \in W_1 \cap W_2$ . Furthermore, for each  $x, y \in W_1 \cap W_2$  and for each  $a \in F$ , we have  $ax + y \in W_1 \cap W_2$  by Corollary 1.10. Thus,  $W_1 \cap W_2$  is a subspace of  $V$ .  $\square$

**Example.** Let  $W_1$  be the set of  $n \times n$  diagonal matrices. Let  $W_2$  be the set of  $n \times n$  matrices that have trace equal to  $0_F$ . Then since both  $W_1$  and  $W_2$  are subspaces of  $F^{n \times n}$ , we can conclude that  $W_1 \cap W_2$  is also a subspace of  $F^{n \times n}$ .

**Definition 1.12.** Let  $V$  be a vector space and let  $S_1, S_2 \subseteq V$ . Then the **sum** of  $S_1$  and  $S_2$ , denoted by  $S_1 + S_2$ , is the set

$$\{x + y : x \in S_1 \text{ and } y \in S_2\}.$$

**Proposition 1.13.** Let  $V$  be a vector space and let  $W_1$  and  $W_2$  be subspaces of  $V$ . Then the following statements are true.

- (a)  $W_1 + W_2$  is a subspace of  $V$ .
- (b) If  $U$  is a subspace of  $V$  with  $W_1 \cup W_2 \subseteq U$ , then  $W_1 + W_2 \subseteq U$ .

*Proof.*

- (a) We have  $0_V = 0_V + 0_V \in W_1 + W_2$ . For each  $x, y \in W_1 + W_2$  and for each  $a \in F$ , by Definition 1.12 there exist  $x_1, y_1 \in W_1$  and  $x_2, y_2 \in W_2$  such that  $x = x_1 + x_2$  and  $y = y_1 + y_2$ . Thus,

$$\begin{aligned} ax + y &= a(x_1 + x_2) + (y_1 + y_2) \\ &= (ax_1 + ax_2) + (y_1 + y_2) \\ &= (ax_1 + y_1) + (ax_2 + y_2) \\ &\in W_1 + W_2. \end{aligned}$$

- (b) Let  $x$  be a vector in  $W_1 + W_2$ . Then by Definition 1.12 there exists  $x_1 \in W_1$  and  $x_2 \in W_2$  such that  $x = x_1 + x_2$ . We have  $x_1 \in U$  since  $W_1 \subseteq U$ . Also, we have  $x_2 \in U$  since  $W_2 \subseteq U$ . It follows that  $x = x_1 + x_2 \in U$ , and thus  $W_1 + W_2 \subseteq U$ .  $\square$

## 1.4 Spanning Sets

**Definition 1.14.** Let  $V$  be a vector space over a field  $F$  and let  $S \subseteq V$ . Then a vector  $x \in V$  is called a **linear combination** of  $S$  if there exist scalars  $a_1, \dots, a_n \in F$  and vectors  $x_1, \dots, x_n \in S$  for some nonnegative integer  $n$  such that

$$x = \sum_{i=1}^n a_i x_i.$$

**Remark.**

- If  $n = 0$ , then the sum in the right hand side is  $0_V$  since nothing are added up. Thus,  $0_V$  is a linear combination of any subset of  $V$ .
- Note that  $n$  should be finite. Thus, in the vector space  $\mathbb{R}$  over the field  $\mathbb{Q}$ ,  $e$  is not a linear combination of  $\mathbb{Q}$  even if we have

$$e = \sum_{i=0}^{\infty} \frac{1}{i!}.$$

**Definition 1.15.** Let  $V$  be a vector space over a field  $F$  and let  $S \subseteq V$ . Then the **span** of  $S$ , denoted  $\text{span}(S)$ , is defined as the set of all linear combinations of  $S$ .

**Theorem 1.16.** Let  $V$  be a vector space over  $F$  and let  $S \subseteq V$ . Then the following statements are true.

- $\text{span}(S)$  is a subspace of  $V$ .
- If  $U$  is a subspace of  $V$  such that  $S \subseteq U$ , then  $\text{span}(S) \subseteq U$ .

*Proof.*

- Let  $c \in F$  and  $x, y \in \text{span}(S)$ . Then there exist scalars  $a_1, \dots, a_n \in F$  and vectors  $x_1, \dots, x_n \in S$  such that

$$x = a_1 x_1 + \dots + a_n x_n.$$

Also, there exist scalars  $b_1, \dots, b_m \in F$  and vectors  $y_1, \dots, y_m \in S$  such that

$$y = b_1 y_1 + \dots + b_m y_m.$$

Thus, we have

$$\begin{aligned} cx + y &= c(x_1 + \dots + x_n) + (y_1 + \dots + y_m) \\ &= cx_1 + \dots + cx_n + y_1 + \dots + y_m \\ &\in \text{span}(S). \end{aligned}$$

Furthermore,  $0_V \in \text{span}(S)$ . Hence,  $\text{span}(S)$  is a subspace of  $V$  by Corollary 1.10.

- Let  $x \in \text{span}(S)$ . Then there exist scalars  $a_1, \dots, a_n \in F$  and vectors  $x_1, \dots, x_n \in S$  such that

$$x = a_1 x_1 + \dots + a_n x_n.$$

Since  $S \subseteq U$ , we have  $x_1, \dots, x_n \in U$ , and it follows that  $x = a_1 x_1 + \dots + a_n x_n \in U$  due to the closeness of  $U$ . Thus,  $\text{span}(S) \subseteq U$ .  $\square$



**Definition 1.17.** Let  $V$  be a vector space and let  $S \subseteq V$ . If  $\text{span}(S) = V$ , then  $S$  is called a **spanning set** of  $V$ , and we also say  $S$  **spans**  $V$ .

**Example.**  $\{(0, 1, 1), (1, 0, 1), (1, 1, 0)\}$  is a spanning set of  $\mathbb{R}^3$  since for any  $x, y, z \in \mathbb{R}$ ,

$$(x, y, z) = \frac{-x + y + z}{2} \cdot (0, 1, 1) + \frac{x - y + z}{2} \cdot (1, 0, 1) + \frac{x + y - z}{2} \cdot (1, 1, 0).$$

**Proposition 1.18.** Let  $V$  be a vector space and let  $R, S \subseteq V$ .

- (a)  $S \subseteq \text{span}(S)$ .
- (b) If  $R \subseteq S$ , then  $\text{span}(R) \subseteq \text{span}(S)$ .
- (c)  $S = \text{span}(S)$  if and only if  $S$  is a subspace of  $V$ .
- (d)  $\text{span}(R \cup S) = \text{span}(R) + \text{span}(S)$ .

*Proof.*

- (a) Straightforward.
- (b) It is true since a linear combination of a subset of  $S$  is also a linear combination of  $S$ .
- (c)  $(\Rightarrow)$  Straightforward from Theorem 1.16 (a).  
 $(\Leftarrow)$  Note that any linear combination of  $S$  is in  $S$  due to closeness of addition and scalar multiplication in  $S$ . Thus,  $\text{span}(S) \subseteq S$ , and it follows that  $S = \text{span}(S)$ .
- (d) Since  $R \subseteq \text{span}(R)$  and  $S \subseteq \text{span}(S)$ , we have  $R \cup S \subseteq \text{span}(R) + \text{span}(S)$ . Thus, by Theorem 1.16, we have  $\text{span}(R \cup S) \subseteq \text{span}(R) + \text{span}(S)$ . On the other side, since

$$\text{span}(R) \subseteq \text{span}(R \cup S) \quad \text{and} \quad \text{span}(S) \subseteq \text{span}(R \cup S),$$

we can conclude that  $\text{span}(R) + \text{span}(S) \subseteq \text{span}(R \cup S)$ . Thus,  $\text{span}(R) + \text{span}(S) \subseteq \text{span}(R \cup S)$  by Proposition 1.13.  $\square$

## 1.5 Linearly Independent Sets

**Definition 1.19.** Let  $V$  be a vector space over a field  $F$  and let  $S \subseteq V$ .

- $S$  is **linearly dependent** if there exist scalars  $a_1, a_2, \dots, a_n \in F \setminus \{0_F\}$  and distinct vectors  $x_1, x_2, \dots, x_n \in S$  for some positive integer  $n$  such that

$$a_1x_1 + a_2x_2 + \dots + a_nx_n = 0_V.$$

- $S$  is **linearly independent** if it is not linearly dependent.

**Remark.**

- Note that  $\emptyset$  is linearly independent.

**Theorem 1.20.** Let  $V$  be a vector space over a field  $F$  and let  $S \subseteq V$ . Then the following statements are equivalent.

- (a)  $S$  is linearly dependent.
- (b) There exists  $x \in S$  with  $x \in \text{span}(S \setminus \{x\})$ .
- (c) There exists  $x \in S$  with  $\text{span}(S) = \text{span}(S \setminus \{x\})$ .

*Proof.*

- (i) First we assume (a) and prove (b). Suppose that

$$a_0x_0 + a_1x_1 + \dots + a_nx_n = 0_V,$$

where  $a_0, a_1, \dots, a_n$  are nonzero scalars and  $x_0, x_1, \dots, x_n$  are distinct vectors. Then

$$\begin{aligned} x_0 &= (-a_0)^{-1}(a_1x_1 + \dots + a_nx_n) \\ &= ((-a_0)^{-1}a_1)x_1 + \dots + ((-a_0)^{-1}a_n)x_n \\ &\in \text{span}(S \setminus \{x_0\}). \end{aligned}$$

- (ii) Then we assume (b) and prove (c). Since

$$x \in \text{span}(S \setminus \{x\}) \quad \text{and} \quad S \setminus \{x\} \subseteq \text{span}(S \setminus \{x\}),$$

we have  $S \subseteq \text{span}(S \setminus \{x\})$ . Thus,  $\text{span}(S) \subseteq \text{span}(S \setminus \{x\})$  by Theorem 1.16, and we can conclude that  $\text{span}(S) = \text{span}(S \setminus \{x\})$ .

- (iii) Then we assume (c) and prove (b). It is straightforward since  $x \in S \subseteq \text{span}(S) = \text{span}(S \setminus \{x\})$ .
- (iv) Finally we assume (b) and prove (a). Without loss of generality, let  $a_1, \dots, a_n \in F$  be nonzero scalars and  $x_1, \dots, x_n \in S \setminus \{x\}$  be distinct vectors such that  $x = a_1x_1 + \dots + a_nx_n$ . Then we have

$$(-1_F)x + a_1x_1 + \dots + a_nx_n = 0_V,$$

which completes the proof. □

**Example.** Let  $S = \{(0, 1, 1), (1, 0, 1), (1, 1, 0)\}$  be a subset of  $\mathbb{R}^3$ . Suppose that  $a_1, a_2, a_3 \in \mathbb{R}$  are scalars such that

$$a_1(0, 1, 1) + a_2(1, 0, 1) + a_3(1, 1, 0) = (0, 0, 0).$$

Then we have the following system of equations.

$$\begin{aligned} a_2 + a_3 &= 0 \\ a_1 + a_3 &= 0 \\ a_1 + a_2 &= 0 \end{aligned}$$

Since the only solution to this system of equations is  $a_1 = a_2 = a_3 = 0$ , we can conclude that  $S$  is linearly independent by Definition 1.19.

**Example.** Let  $S = \{(1, 1, 1), (0, 1, 1), (1, 0, 1), (1, 1, 0)\}$  be a subset of  $\mathbb{R}^3$ . We can conclude that  $S$  is linearly dependent since

$$(1, 1, 1) = \frac{1}{2} \cdot (0, 1, 1) + \frac{1}{2} \cdot (1, 0, 1) + \frac{1}{2} \cdot (1, 1, 0).$$

**Proposition 1.21.** Let  $V$  be a vector space and let  $R, S$  be subsets of  $V$  with  $R \subseteq S$ .

- (a) If  $R$  is linearly dependent, then so is  $S$ .
- (b) If  $S$  is linearly independent, then so is  $R$ .

*Proof.*

- (a) Suppose that  $R$  is linearly dependent. Then by Definition 1.19 there exists  $x \in R$  such that  $x \in \text{span}(R \setminus \{x\})$ . Also, we have  $R \setminus \{x\} \subseteq S \setminus \{x\}$  since  $R \subseteq S$ . Thus,  $x \in \text{span } S \setminus \{x\}$ , and it follows that  $S$  is linearly dependent.
- (b) Straightforward from (a). □

## 1.6 Bases and Dimension

**Definition 1.22.** A **basis** for a vector space  $V$  is a linearly independent subset of  $V$  that spans  $V$ .

**Examples.**

- $\emptyset$  is a basis for  $\{0_V\}$ .
- $\{e_1, \dots, e_n\}$  is a basis for  $F^n$ , where  $e_i$  is the  $n$ -tuple whose  $i$ -th component is  $1_F$  and the other components are all  $0_F$ .
- $\{E_{ij} : 1 \leq i \leq m \text{ and } 1 \leq j \leq n\}$  is a basis for  $F^{m \times n}$ , where  $E_{ij}$  is the matrix whose  $(i, j)$ -entry is  $1_F$  and the other entries are all  $0_F$ .
- $\{t^0, t^1, t^2, \dots, t^n\}$  is a basis for  $\mathcal{P}_n(F)$ .
- $\{t^0, t^1, t^2, \dots\}$  is a basis for  $\mathcal{P}(F)$ .

**Proposition 1.23.** Let  $V$  be a vector space. If there exists a finite set  $S$  that spans  $V$ , then there is a subset  $Q$  of  $S$  that is a finite basis of  $V$ .

*Proof.* The proof is by induction on  $|S|$ . For the induction basis, suppose that  $|S| = 0$ , i.e.,  $S = \emptyset$ . Then the proposition holds since one can choose  $Q = \emptyset$  as a basis for  $V$ .

Now assume the induction hypothesis that the proposition holds for  $|S| = n$  with  $n \geq 0$ . If  $S$  is linearly independent, then we can choose  $Q = S$  as a basis for  $V$ . Otherwise, there exists  $x \in S$  with  $\text{span}(S \setminus \{x\}) = \text{span}(S)$ , i.e.,  $S \setminus \{x\}$  spans  $V$ . Thus, by induction hypothesis there is a subset  $Q$  of  $S \setminus \{x\}$  that is a basis for  $V$ , which completes the proof.  $\square$

**Theorem 1.24 (Replacement Theorem).** Let  $V$  be a vector space over a field  $F$ . Let  $S$  be a finite set that spans  $V$ , and let  $Q \subseteq S$  be a finite linearly independent set. Then  $|Q| \leq |S|$ , and there exists  $R \subseteq S \setminus Q$  such that both  $|Q \cup R| = |S|$  and  $\text{span}(Q \cup R) = V$  hold.

*Proof.* The proof is based on induction on  $|Q|$ . The theorem holds for  $|Q| = 0$ , i.e.,  $Q = \emptyset$ , since we have  $|\emptyset| \leq |S|$ ,  $|\emptyset \cup S| = |S|$  and  $\text{span}(\emptyset \cup S) = V$ .

Now suppose that the theorem is true for  $|Q| = m$  with  $m \geq 0$ , and we prove that the theorem holds for  $|Q| = m + 1$ . Let  $Q = \{x_1, \dots, x_{m+1}\}$  and let  $Q' = \{x_1, \dots, x_m\}$ . By induction hypothesis, there exists  $R' = \{y_1, \dots, y_k\} \subseteq S \setminus Q'$  such that  $|Q'| + |R'| = |S|$  and  $\text{span}(Q' \cup R') = V$ . Since  $Q' \cup R'$  spans  $V$ , there exists  $a_1, \dots, a_m, b_1, \dots, b_k \in F$  such that

$$x_{m+1} = \sum_{i=1}^m a_i x_i + \sum_{j=1}^k b_j y_j.$$

If  $b_j = 0_F$  for all  $j \in \{1, \dots, k\}$ , then  $x_{m+1} \in \text{span}(Q') = \text{span}(Q \setminus \{x_{m+1}\})$ , implying that  $Q$  is linearly dependent, contradiction. Thus, there must exist some  $j \in \{1, \dots, k\}$  such that  $b_j \neq 0_F$ . Without loss of generality, suppose that  $b_k \neq 0_F$  with  $k \geq 1$ . Also, let  $R = \{y_1, \dots, y_{k-1}\}$ . Then  $|Q \cup R| = (m+1) + (k-1) = |S|$ , and we have  $|Q| \leq |S|$ . It follows that

$$Q' \cup R' \subseteq Q \cup R \cup \{y_k\} \subseteq \text{span}(Q \cup R),$$

where the second inclusion holds because

$$y_k = (-b_k)^{-1} \left( \sum_{i=1}^m a_i x_i + (-1_F) x_{m+1} + \sum_{j=1}^{k-1} b_j y_j \right) \in \text{span}(Q \cup R).$$

Then, we have

$$V = \text{span}(Q' \cup R') \subseteq \text{span}(Q \cup R) \subseteq V.$$

by Theorem 1.16. Thus,  $\text{span}(Q \cup R) = V$ , which completes the proof.  $\square$

**Corollary 1.25.** Let  $V$  be a vector space and  $Q$  be a linearly independent subset of  $V$  that is infinite. Then each spanning set of  $V$  is infinite.

*Proof.* Suppose that there is a finite set  $S$  that spans  $V$ . Let  $Q'$  be a subset of  $Q$  with  $|Q'| = |S| + 1$ . By Proposition 1.21, we can conclude that  $Q'$  is also linearly independent. Thus, we have  $|Q'| \leq |S|$  by replacement theorem (Theorem 1.24), contradiction.  $\square$

**Corollary 1.26.** Let  $V$  be a vector space. If  $V$  has a finite basis, then each basis for  $V$  has the same size.

*Proof.* Let  $S$  be a finite basis for  $V$  and  $Q$  an arbitrary basis for  $V$ . Since  $V = \text{span}(S)$  and  $Q$  is linearly independent, it follows that  $Q$  is finite by Corollary 1.25, and thus we have  $|Q| \leq |S|$ . Also, since  $V = \text{span}(Q)$  and  $S$  is linearly independent, we have  $|S| \leq |Q|$ . Thus,  $|Q| = |S|$ .  $\square$

**Definition 1.27.** Let  $V$  be a vector space.

- $V$  is **finite-dimensional** if it has a finite basis. In this case, the number of vectors in each basis for  $V$  is called the **dimension** of  $V$ , denoted by  $\dim(V)$ .
- $V$  is **infinite-dimensional** if it is not finite-dimensional.

**Remark.**

- If a vector space has a linearly independent subset that is infinite, we can conclude that it is infinite-dimensional by Corollary 1.25.

**Examples.** One can find the dimension of a vector space by any basis it admits.

- $\dim(\{0_V\}) = 0$ .
- $\dim(F^n) = n$ .
- $\dim(F^{m \times n}) = mn$ .
- $\dim(\mathcal{P}_n(F)) = n + 1$ .
- $\mathcal{P}(F)$  is infinite-dimensional.

**Examples.** Note that the dimension of a vector space depends on its field of scalars.

- Let  $V = \mathbb{C}$  be a vector space over  $\mathbb{R}$ . Then we have  $\dim(V) = 2$  since  $\{1, i\}$  is a basis for  $V$ .
- Let  $W = \mathbb{C}$  be a vector space over  $\mathbb{C}$ . Then we have  $\dim(W) = 1$  since  $\{1\}$  is a basis for  $V$ .

**Proposition 1.28.** Let  $V$  be a vector space. Then a subset of  $V$  of  $n = \dim(V)$  vectors is linearly independent if and only if it is a spanning set of  $V$ .

*Proof.* ( $\Rightarrow$ ) Suppose that  $Q$  is linearly independent with  $|Q| = n$ . By replacement theorem (Theorem 1.24), there exists  $R \subseteq S \setminus Q$  such that  $|Q \cup R| = |S|$  and  $\text{span}(Q \cup R) = V$ . Since  $|Q| = |S|$ , we have  $|R| = 0$ , i.e.,  $R = \emptyset$ . Thus,  $\text{span}(Q) = V$ .

( $\Leftarrow$ ) Suppose that  $S$  spans  $V$  with  $|S| = n$ . By Proposition 1.23, there is a subset  $Q$  of  $S$  that is a basis of  $V$ . Then we have  $|Q| = n$ , implying  $Q = S$ . Thus,  $S$  is a basis for  $V$ .  $\square$

**Proposition 1.29.** Let  $V$  be a finite-dimensional vector space. Let  $S = \{x_1, \dots, x_n\}$  be a basis for  $V$ . Then for each  $x \in V$ , there exist a unique  $n$ -tuple  $(a_1, \dots, a_n) \in F^n$  with

$$x = a_1x_1 + \dots + a_nx_n.$$

*Proof.* Since  $x \in \text{span}(S)$ , there exist scalars  $a_1, \dots, a_n \in F$  such that

$$x = a_1x_1 + \dots + a_nx_n.$$

Now we prove the uniqueness. Let  $b_1, \dots, b_n \in F$  be scalars with

$$x = b_1x_1 + \dots + b_nx_n.$$

Then we have

$$0_V = (a_1 - b_1)x_1 + \dots + (a_n - b_n)x_n,$$

and it follows that  $(a_1 - b_1, a_2 - b_2, \dots, a_n - b_n) = 0_{F^n}$  since  $S$  is linearly independent. Thus,  $(a_1, \dots, a_n) = (b_1, \dots, b_n)$ .  $\square$

**Proposition 1.30.** Let  $V$  be a finite-dimensional vector space. Let  $V'$  be a subspace of  $V$ . Then the following statements are true.

- (a)  $\dim(V') \leq \dim(V)$ .
- (b) If  $\dim(V') = \dim(V)$ , then  $V' = V$ .

*Proof.* Let  $S$  and  $S'$  be bases for  $V$  and  $V'$ , respectively.

- (a) Since  $S'$  is linearly independent and  $V = \text{span}(S)$ , we have  $|S'| \leq |S|$  by replacement theorem (Theorem 1.24). Thus,  $\dim(V') \leq \dim(V)$ .
- (b) Since  $S'$  is linearly independent and  $|S'| = \dim(V)$ , we have  $\text{span}(S') = V$  by Proposition 1.28. Thus,  $V' = \text{span}(S') = V$ .  $\square$

**Example.** Let  $W$  be the set of  $n \times n$  diagonal matrices, which is a subspace of  $F^{n \times n}$ . Then one can verify that  $\{E_{ii} : 1 \leq i \leq n\}$  is a basis for  $W$ , where  $E_{ij}$  is the matrix whose  $(i, j)$ -entry is  $1_F$  and the other entries are  $0_F$ . Thus,  $\dim(W) = n$ .

# Chapter 2

## Linear Transformations

### 2.1 Linear Transformations

**Definition 2.1.** Let  $V$  and  $W$  be vector spaces over a field  $F$ . A transformation  $T : V \rightarrow W$  is said to be **linear** if

$$T(ax + y) = aT(x) + T(y)$$

holds for any scalar  $a \in F$  and any vectors  $x, y \in V$ . The set of all linear transformations from  $V$  to  $W$  is denoted by  $\mathcal{L}(V, W)$ , and  $\mathcal{L}(V)$  for short if  $V = W$ .

**Proposition 2.2.** Let  $V$  and  $W$  be vector spaces over a common field  $F$ . Let  $T : V \rightarrow W$  be linear. Then we have the following properties.

- (a)  $T(0_V) = 0_W$ .
- (b) For nonnegative integer  $n$ ,

$$T\left(\sum_{i=1}^n a_i x_i\right) = \sum_{i=1}^n a_i T(x_i)$$

hold for any  $a_1, \dots, a_n \in F$  and  $x_1, \dots, x_n \in V$ .

*Proof.*

- (a) Since

$$T(0_V) + T(0_V) = 1_F T(0_V) + T(0_V) = T(1_F 0_V + 0_V) = T(0_V),$$

we have  $T(0_V) = 0_W$  by Proposition 1.6 (b).

- (b) The proof is by induction on  $n$ . The induction basis with  $n = 0$  is proved by

$$T\left(\sum_{i=1}^0 a_i x_i\right) = T(0_V) = 0_W = \sum_{i=1}^0 a_i T(x_i).$$

Now assume the induction hypothesis that the property holds for  $n = k$ . Then it follows that

$$\begin{aligned}
T\left(\sum_{i=1}^{k+1} a_i x_i\right) &= T\left(a_{k+1} x_{k+1} + \sum_{i=1}^k a_i x_i\right) \\
&= a_{k+1} T(x_{k+1}) + T\left(\sum_{i=1}^k a_i x_i\right) && \text{(linearity of } T) \\
&= a_{k+1} T(x_{k+1}) + \sum_{i=1}^k a_i T(x_i) && \text{(induction hypothesis)} \\
&= \sum_{i=1}^{k+1} a_i T(x_i),
\end{aligned}$$

which completes the proof.  $\square$

**Theorem 2.3.** If  $V$  and  $W$  are vector spaces over a field  $F$ , then  $\mathcal{L}(V, W)$  is also a vector space over  $F$ .

*Proof.* For any  $c \in F$  and  $T_1, T_2 \in \mathcal{L}(V, W)$ , since

$$\begin{aligned}
(cT_1 + T_2)(ax + y) &= cT_1(ax + y) + T_2(ax + y) && \text{(linearity of } cT_1 + T_2) \\
&= c(aT_1(x) + T_1(y)) + (aT_2(x) + T_2(y)) && \text{(linearity of } T_1 \text{ and } T_2) \\
&= acT_1(x) + cT_1(y) + aT_2(x) + T_2(y) \\
&= a(cT_1(x) + T_2(x)) + (cT_1(y) + T_2(y)) \\
&= a(cT_1 + T_2)(x) + (cT_1 + T_2)(y) && \text{(linearity of } cT_1 + T_2)
\end{aligned}$$

holds for each  $a \in F$  and  $x, y \in V$ , we have  $cT_1 + T_2 \in \mathcal{L}(V, W)$ . Furthermore,  $0_{\mathcal{F}(V, W)} \in \mathcal{L}(V, W)$ . Thus,  $\mathcal{L}(V, W)$  is a subspace of  $\mathcal{F}(V, W)$ .  $\square$

**Theorem 2.4.** Let  $V$  and  $W$  be vector spaces and let  $T : V \rightarrow W$  be linear. Then for any subset  $S$  of  $V$ , we have

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

*Proof.* If  $y \in T(\text{span}(S))$ , then there exist  $a_i \in F$  and  $x_i \in S$  for each  $1 \leq i \leq n$  such that

$$y = T\left(\sum_{i=1}^n a_i x_i\right) = \sum_{i=1}^n a_i T(x_i) \in \text{span}(T(S)).$$

Thus,  $T(\text{span}(S)) \subseteq \text{span}(T(S))$ .

On the other hand, if  $y \in \text{span}(T(S))$ , then there exist  $a_i \in F$  and  $x_i \in S$  for each  $1 \leq i \leq n$  such that

$$y = \sum_{i=1}^n a_i T(x_i) = T\left(\sum_{i=1}^n a_i x_i\right) \in T(\text{span}(S)).$$

Thus,  $\text{span}(T(S)) \subseteq T(\text{span}(S))$ , which completes the proof.  $\square$



## 2.2 Rank and Nullity

**Definition 2.5.** Let  $V$  and  $W$  be vector spaces. The **range** of a transformation  $T : V \rightarrow W$ , denoted by  $\mathcal{R}(T)$ , is defined by

$$\mathcal{R}(T) = \{y \in W : y = T(x) \text{ for some } x \in V\}.$$

**Proposition 2.6.** Let  $V$  and  $W$  be vector spaces over a field  $F$ . If  $T : V \rightarrow W$  is linear, then  $\mathcal{R}(T)$  is a subspace of  $W$ .

*Proof.* For each  $a \in F$  and  $y, y' \in \mathcal{R}(T)$ , there exist  $x, x' \in V$  such that  $y = T(x)$  and  $y' = T(x')$ . Since

$$ay + y' = aT(x) + T(x') = T(ax + x'),$$

we have  $ay + y' \in \mathcal{R}(T)$ . Furthermore,  $0_W = T(0_V) \in \mathcal{R}(T)$ . Thus,  $\mathcal{R}(T)$  is a subspace of  $W$ .  $\square$

**Definition 2.7.** Let  $V$  and  $W$  be vector spaces. The **null space** of a transformation  $T : V \rightarrow W$ , denoted by  $\mathcal{N}(T)$ , is defined by

$$\mathcal{N}(T) = \{x \in V : T(x) = 0_W\}.$$

**Proposition 2.8.** Let  $V$  and  $W$  be vector spaces over a field  $F$ . If  $T : V \rightarrow W$  is linear, then  $\mathcal{N}(T)$  is a subspace of  $V$ .

*Proof.* For each  $a \in F$  and  $x, x' \in \mathcal{N}(T)$ , we have

$$T(ax + x') = aT(x) + T(x') = a0_W + 0_W = 0_W.$$

Thus,  $ax + x' \in \mathcal{N}(T)$ . Furthermore,  $0_V \in \mathcal{N}(T)$  since  $T(0_V) = 0_W$ . Thus,  $\mathcal{N}(T)$  is a subspace of  $V$ .  $\square$

**Definition 2.9.** Let  $X$  and  $Y$  be sets. Let  $f : X \rightarrow Y$  be a function.

- $f$  is **injective** if  $T(x) = T(x')$  implies  $x = x'$  for all  $x, x' \in X$ .
- $f$  is **surjective** if there exists  $x \in X$  with  $T(x) = y$  for each  $y \in Y$ .
- $f$  is **bijective** if  $f$  is injective and surjective.

**Proposition 2.10.** Let  $V$  and  $W$  be vector spaces and let  $T : V \rightarrow W$  be linear. Let  $S$  be a subset of  $V$ . Then the following statements are true.

- (a)  $T$  is injective if and only if  $\mathcal{N}(T) = \{0_V\}$ .
- (b) If  $T$  is injective, then  $S$  is linearly dependent if and only if  $T(S)$  is linearly dependent.

*Proof.*

- (a) ( $\Rightarrow$ ) We have  $T(0_V) = 0_W$  since  $T$  is linear. If  $T(x) = 0_W$ , then  $x = 0_V$  since  $T$  is injective. Thus,  $\mathcal{N}(T) = \{0_V\}$ .
- ( $\Leftarrow$ ) Suppose that  $x, y \in V$  be vectors with  $T(x) = T(y)$ . Since

$$T(x - y) = T(x) - T(y) = 0_W,$$

we have  $x - y \in \mathcal{N}(T)$ , and thus  $x - y = 0_V$ , implying  $x = y$ . Thus,  $T$  is injective.

(b) ( $\Rightarrow$ ) If  $x \in \text{span}(S \setminus \{x\})$  for some  $x \in S$ , then

$$\begin{aligned} T(x) &\in T(\text{span}(S \setminus \{x\})) \\ &= \text{span}(T(S \setminus \{x\})) && (T \text{ is linear}) \\ &= \text{span}(T(S) \setminus \{T(x)\}). && (T \text{ is injective}) \end{aligned}$$

( $\Leftarrow$ ) If  $T(x) \in \text{span}(T(S) \setminus \{T(x)\})$  for some  $x \in S$ , then

$$\begin{aligned} T(x) &\in \text{span}(T(S) \setminus \{T(x)\}) \\ &= \text{span}(T(S \setminus \{x\})) && (T \text{ is injective}) \\ &= T(\text{span}(S \setminus \{x\})). && (T \text{ is linear}) \end{aligned}$$

Thus,  $x \in \text{span}(S \setminus \{x\})$  since  $T$  is injective.  $\square$

**Definition 2.11.** Let  $V$  and  $W$  be vector spaces. Let  $T : V \rightarrow W$  be linear.

- The **rank** of  $T$ , denoted by  $\text{rank}(T)$ , is the dimension of  $\mathcal{R}(T)$ .
- The **nullity** of  $T$ , denoted by  $\text{nullity}(T)$ , is the dimension of  $\mathcal{N}(T)$ .

**Definition 2.12.** Let  $f : X \rightarrow Y$  be a function. Let  $D$  be a subset of  $X$ . Then the **restriction** of  $f$  to  $D$  is the function  $f' : D \rightarrow Y$  with  $f'(x) = f(x)$  for each  $x \in D$ .

**Proposition 2.13.** Let  $V$  and  $W$  be vector spaces and let  $T : V \rightarrow W$  be linear. Let  $U$  be a subspace of  $V$ . Then the restriction of  $T$  to  $U$  is linear.

*Proof.* Let  $T' : U \rightarrow W$  be the restriction of  $T$  to  $U$ . Then  $T'$  is linear since for each  $a \in F$  and  $x, y \in U$ , we have

$$T'(ax + y) = T(ax + y) = aT(x) + T(y) = aT'(x) + T'(y). \quad \square$$

**Theorem 2.14 (Rank-nullity Theorem).** Let  $V$  and  $W$  be finite-dimensional vector spaces over  $F$ . Let  $T : V \rightarrow W$  be linear. Then we have

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

*Proof.* Let  $S$  be a basis for  $V$  and  $Q$  a basis for  $\mathcal{N}(T)$ . By replacement theorem (Theorem 1.24), there is  $R \subseteq S \setminus Q$  such that  $Q \cup R$  is a basis for  $V$ .

We prove that  $T(R)$  is a basis for  $\mathcal{R}(T)$ . First,

$$\begin{aligned} \mathcal{R}(T) &= T(\text{span}(Q \cup R)) \\ &= \text{span}(T(Q \cup R)) \\ &= \text{span}(T(Q) \cup T(R)) \\ &= \text{span}(T(R)). && (T(Q) = \{0_V\}) \end{aligned}$$

Now we prove that  $T(R)$  is linearly independent. Let  $T'$  be the restriction of  $T$  to  $R$ . Since  $R$  is linearly independent, it suffices to prove that  $T'$  is injective. Suppose that  $T(x) = T(x')$  for some  $x, x' \in R$ . Then we have  $T(x - x') = T(x) - T(x') = 0_W$ , and thus  $x - x' \in \mathcal{N}(T) = \text{span}(Q)$ . It follows that  $x$  is a linear combination of  $Q \cup \{x'\}$ . If  $x \neq x'$ , then

$$x \in \text{span}(Q \cup \{x'\}) \subseteq \text{span}(Q \cup R \setminus \{x\}),$$

contradiction to the fact that  $Q \cup R$  is linearly independent. Thus,  $T'$  is injective, implying  $T(R)$  is linearly independent.

Note that  $|T(R)| = |R|$  since  $T'$  is injective. Thus,

$$\text{nullity}(T) + \text{rank}(T) = |Q| + |T(R)| = |Q| + |R| = \dim(V). \quad \square$$

## 2.3 Isomorphisms

**Definition 2.15.** Let  $X, Y, Z$  be sets. Let  $f : X \rightarrow Y$  and  $g : Y \rightarrow Z$  be functions. Then the **composition** of  $f$  and  $g$  is the function  $gf : X \rightarrow Z$  such that

$$(gf)(x) = g(f(x))$$

for all  $x \in X$ .

**Definition 2.16.** The **identity function** over a set  $X$  is a function  $I_X : X \rightarrow X$  with  $I_X(x) = x$  for all  $x \in X$ .

**Definition 2.17.** Let  $X$  and  $Y$  be sets. A function  $f : X \rightarrow Y$  is said to be **invertible** if there exists a function  $f^{-1} : Y \rightarrow X$ , called the **inverse** of  $f$ , such that

$$f^{-1}f = I_X \quad \text{and} \quad ff^{-1} = I_Y.$$

**Proposition 2.18.** Let  $X$  and  $Y$  be sets. Let  $f : X \rightarrow Y$  and  $g : Y \rightarrow X$  be functions.

- (a) If  $f$  is invertible, then  $f^{-1}$  is invertible.
- (b) If  $f$  is invertible, then  $f^{-1}$  is linear.
- (c) If  $f$  is invertible, then either  $gf = I_X$  or  $fg = I_Y$  implies  $g = f^{-1}$ .
- (d)  $f$  is invertible if and only if  $f$  is bijective.

*Proof.*

- (a) Straightforward from Definition 2.17.
- (b) For  $a \in F$  and  $y, y' \in Y$ , we have

$$\begin{aligned} f^{-1}(ay + y') &= f^{-1}(af(f^{-1}(y)) + f(f^{-1}(y'))) && (ff^{-1} = I_Y) \\ &= f^{-1}(f(af^{-1}(y) + f^{-1}(y'))) && (\text{linearity of } f) \\ &= af^{-1}(y) + f^{-1}(y'). && (f^{-1}f = I_X) \end{aligned}$$

Thus,  $f^{-1}$  is linear.

- (c) If  $gf = I_X$ , then

$$g = gI_Y = g(ff^{-1}) = (gf)f^{-1} = I_X f^{-1} = f^{-1}.$$

If  $fg = I_Y$ , then

$$g = I_X g = (f^{-1}f)g = f^{-1}(fg) = f^{-1}I_Y = f^{-1}.$$

- (d) ( $\Rightarrow$ ) Suppose that  $f$  is invertible. Then  $f$  is injective since for each  $x, x' \in X$  with  $f(x) = f(x')$ , we have

$$x = (f^{-1}f)(x) = f^{-1}(f(x)) = f^{-1}(f(x')) = (f^{-1}f)(x') = x'.$$

Also,  $f$  is surjective since for each  $y \in Y$ , we have

$$y = (ff^{-1})y = f(f^{-1}(y)).$$

( $\Leftarrow$ ) If  $f$  is bijective, then for each  $y \in Y$  there exists a unique element  $x \in X$  with  $f(x) = y$ . Thus, there exists a function  $g : Y \rightarrow X$  such that

$$g(f(x)) = x$$

for each  $x \in X$ . For any  $y \in Y$ , if  $x \in X$  is the element such that  $f(x) = y$ , then we have

$$f(g(y)) = f(g(f(x))) = f(x) = y.$$

Thus,  $f$  is invertible since  $gf = I_X$  and  $fg = I_Y$ .  $\square$

**Definition 2.19.** Let  $V$  and  $W$  be vector spaces. An **isomorphism** from  $V$  onto  $W$  is a invertible linear transformation from  $V$  to  $W$ . If there is an isomorphism from  $V$  onto  $W$ , then  $V$  and  $W$  are said to be **isomorphic**, denoted by  $V \cong W$ .

**Lemma 2.20.** Let  $V$  and  $W$  be finite-dimensional vector spaces with  $\dim(V) = \dim(W)$ . Let  $T : V \rightarrow W$  be linear. Then  $T$  is injective if and only if  $T$  is surjective.

*Proof.* ( $\Rightarrow$ ) If  $T$  is injective, then  $\mathcal{N}(T) = \{0_V\}$ , implying  $\text{nullity}(T) = 0$ . Then we have

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

Since  $\mathcal{R}(T)$  is a subspace of  $W$  with  $\dim(\mathcal{R}(T)) = \dim(W)$ , we can conclude that  $\mathcal{R}(T) = W$  by Proposition 1.30.

( $\Leftarrow$ ) If  $T$  is surjective, then  $\mathcal{R}(T) = W$ . Thus,

$$\text{nullity}(T) = \dim(V) - \text{rank}(T) = \dim(W) - \dim(W) = 0,$$

implying  $\mathcal{N}(T) = \{0_V\}$ . It follows that  $T$  is injective.  $\square$

**Lemma 2.21.** Let  $V$  and  $W$  be finite-dimensional vector spaces over a field  $F$ . Let  $S = \{x_1, x_2, \dots, x_n\}$  be a basis for  $V$  and let  $y_1, y_2, \dots, y_n$  be vectors in  $W$ . Then there exists a unique  $T \in \mathcal{L}(V, W)$  with  $T(x_i) = y_i$  for each  $i \in \{1, \dots, n\}$ .

*Proof.* Let  $T$  be the transformation that satisfies

$$T(a_1x_1 + a_2x_2 + \dots + a_nx_n) = a_1y_1 + a_2y_2 + \dots + a_ny_n$$

for any  $a_1, a_2, \dots, a_n \in F$ . It is obvious that  $T(x_i) = y_i$  for each  $i \in \{1, \dots, n\}$ , and  $T$  is linear since

$$\begin{aligned} T\left(c \sum_{i=1}^n a_i x_i + \sum_{i=1}^n b_i x_i\right) &= T\left(\sum_{i=1}^n (ca_i + b_i) x_i\right) \\ &= \sum_{i=1}^n (ca_i + b_i) y_i \\ &= c \sum_{i=1}^n a_i y_i + \sum_{i=1}^n b_i y_i \\ &= cT\left(\sum_{i=1}^n a_i x_i\right) + T\left(\sum_{i=1}^n b_i x_i\right) \end{aligned}$$

holds for any scalars  $a_1, a_2, \dots, a_n, b_1, b_2, \dots, b_n, c \in F$ . To see the uniqueness, if  $T' \in \mathcal{L}(V, W)$  satisfies  $T'(x_i) = y_i$  for each  $i \in \{1, \dots, n\}$ , then we have

$$\begin{aligned} T'(a_1x_1 + \dots + a_nx_n) &= a_1T'(x_1) + \dots + a_nT'(x_n) \\ &= a_1T(x_1) + \dots + a_nT(x_n) \\ &= T(a_1x_1 + \dots + a_nx_n). \end{aligned}$$

for any  $a_1, \dots, a_n \in F$ . Thus,  $T' = T$ .  $\square$

**Theorem 2.22.** Let  $V$  and  $W$  be finite-dimensional vector spaces over a field  $F$ . Then  $V \cong W$  if and only if  $\dim(V) = \dim(W)$ .

*Proof.* ( $\Rightarrow$ ) Let  $T$  be an isomorphism from  $V$  onto  $W$ . Since  $T$  is invertible,  $T$  is bijective. Then we have  $\text{rank}(T) = \dim(W)$  since  $\mathcal{R}(T) = W$ . Furthermore, since  $T$  is injective, we have  $\text{nullity}(T) = 0$ , and it follows that  $\text{rank}(T) = \dim(V)$  by rank-nullity theorem (Theorem 2.14). Thus,  $\dim(V) = \text{rank}(T) = \dim(W)$ .

( $\Leftarrow$ ) Suppose that  $S = \{x_1, x_2, \dots, x_n\}$  is a basis for  $V$  and  $R = \{y_1, y_2, \dots, y_n\}$  is a basis for  $W$ . Then by Lemma 2.21 there exists  $T \in \mathcal{L}(V, W)$  such that  $T(x_i) = y_i$  for each  $i \in \{1, \dots, n\}$ . Since  $R$  is a basis for  $W$ , for each  $y \in W$  there exist scalars  $a_1, \dots, a_n \in F$  such that

$$y = \sum_{i=1}^n a_i y_i = \sum_{i=1}^n a_i T(x_i) = T\left(\sum_{i=1}^n a_i x_i\right).$$

It follows that  $T$  is surjective, and we can conclude that  $T$  is bijective by Lemma 2.20. Thus,  $T$  is an isomorphism from  $V$  onto  $W$ , implying  $V \cong W$ .  $\square$

## 2.4 Coordinates and Matrix Representations

**Definition 2.23.** Let  $V$  be a finite-dimensional vector space over a field  $F$  with  $\dim(V) = n$ . An **ordered basis** for  $V$  is a finite sequence

$$\beta = (x_1, x_2, \dots, x_n)$$

of vectors in  $V$  such that the set  $S = \{x_1, x_2, \dots, x_n\}$  is a basis for  $V$ .

**Examples.**

- The **standard ordered basis** for  $F^n$  is  $(e_1, \dots, e_n)$ , where  $e_i$  is the  $n$ -tuple whose  $i$ -th component is  $1_F$  and the other components are all  $0_F$ .
- The **standard ordered basis** for  $\mathcal{P}_n(F)$  is  $(t^0, t^1, \dots, t^n)$ .

**Definition 2.24.** Let  $V$  be a finite-dimensional vector space over a field  $F$ . Let  $\beta = (x_1, \dots, x_n)$  be an ordered basis for  $V$ . Then we define  $\phi_\beta : V \rightarrow F^n$  such that

$$\phi_\beta(x) = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} \quad \text{for each } x \in V \text{ with } x = \sum_{i=1}^n a_i x_i,$$

where  $a_1, a_2, \dots, a_n \in F$ . For each vector  $x$  in  $V$ ,  $\phi_\beta(x)$  is called the **coordinate** of  $x$  with respect to  $\beta$ , denoted by  $[x]_\beta$ .

**Proposition 2.25.** Let  $\beta = (x_1, \dots, x_n)$  be an ordered basis for a vector space  $V$  over  $F$ . Then  $\phi_\beta$  is an isomorphism from  $V$  onto  $F^n$ .

*Proof.*  $\phi_\beta$  is linear since

$$\begin{aligned} \phi_\beta \left( c \sum_{i=1}^n a_i x_i + \sum_{i=1}^n b_i x_i \right) &= \phi_\beta \left( \sum_{i=1}^n (ca_i + b_i) x_i \right) = \begin{pmatrix} ca_1 + b_1 \\ \vdots \\ ca_n + b_n \end{pmatrix} = c \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} + \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix} \\ &= c \cdot \phi_\beta \left( \sum_{i=1}^n a_i x_i \right) + \phi_\beta \left( \sum_{i=1}^n b_i x_i \right) \end{aligned}$$

holds for any  $a_1, \dots, a_n, b_1, \dots, b_n, c \in F$ . Also,  $\phi_\beta$  is invertible since there exists  $\phi_\beta^{-1} : F^n \rightarrow V$  with

$$\phi_\beta^{-1} \left( \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} \right) = a_1 x_1 + a_2 x_2 + \dots + a_n x_n$$

for any  $a_1, a_2, \dots, a_n \in F$ . Thus,  $\phi_\beta$  is an isomorphism.  $\square$

**Definition 2.26.** Let  $V$  and  $W$  be finite-dimensional vector spaces over a field  $F$ . Let

$$\beta = (x_1, \dots, x_n) \quad \text{and} \quad \gamma = (y_1, \dots, y_m)$$

be ordered basis for  $V$  and  $W$ , respectively. Then we define  $\Phi_\beta^\gamma : \mathcal{L}(V, W) \rightarrow F^{m \times n}$  by

$$\Phi_\beta^\gamma(T) = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix},$$

for each  $T \in \mathcal{L}(V, W)$ , where

$$\begin{aligned} T(x_1) &= a_{11}y_1 + a_{21}y_2 + \cdots + a_{m1}y_m \\ T(x_2) &= a_{12}y_1 + a_{22}y_2 + \cdots + a_{m2}y_m \\ &\vdots \\ T(x_n) &= a_{1n}y_1 + a_{2n}y_2 + \cdots + a_{mn}y_m \end{aligned}$$

hold. For each linear  $T : V \rightarrow W$ , the matrix  $\Phi_\beta^\gamma(T)$  is called the **matrix representation** of  $T$  with respect to  $\beta$  and  $\gamma$ , denoted by  $[T]_\beta^\gamma$ .

**Proposition 2.27.** Let  $\beta = (x_1, \dots, x_n)$  and  $\gamma = (y_1, \dots, y_m)$  be ordered bases for a vector spaces  $V$  and  $W$  over  $F$ , respectively. Then for any  $T \in \mathcal{L}(V, W)$ , we have

$$\left([T]_\beta^\gamma\right)_{ij} = \left([T(x_j)]_\gamma\right)_i$$

for any  $i \in \{1, \dots, m\}$  and  $j \in \{1, \dots, n\}$ .

*Proof.* Let

$$[T]_\beta^\gamma = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}.$$

Since  $T(x_j) = a_{1j}y_1 + a_{2j}y_2 + \cdots + a_{mj}y_m$ , we have

$$[T(x_j)]_\gamma = \begin{pmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{pmatrix}.$$

Thus,

$$\left([T(x_j)]_\gamma\right)_i = a_{ij}$$

holds, which completes the proof.  $\square$

**Theorem 2.28.** Let  $\beta$  and  $\gamma$  be ordered bases for a vector spaces  $V$  and  $W$  over  $F$ , respectively. Then  $\Phi_\beta^\gamma$  is an isomorphism from  $\mathcal{L}(V, W)$  onto  $F^{m \times n}$ .

*Proof.* Let  $\beta = (x_1, \dots, x_n)$  and  $\gamma = (y_1, \dots, y_m)$ . Note that  $\Phi_\beta^\gamma$  is linear since for any  $c \in F$  and  $T_1, T_2 \in \mathcal{L}(V, W)$ , we have

$$\begin{aligned}
\left([cT_1 + T_2]_\beta^\gamma\right)_{ij} &= \left([(cT_1 + T_2)(x_j)]_\gamma\right)_i && \text{(Proposition 2.27)} \\
&= \left([cT_1(x_j) + T_2(x_j)]_\gamma\right)_i \\
&= \left(c[T_1(x_j)]_\gamma + [T_2(x_j)]_\gamma\right)_i && (\phi_\gamma \text{ is linear}) \\
&= c\left([T_1(x_j)]_\gamma\right)_i + \left([T_2(x_j)]_\gamma\right)_i \\
&= c\left([T_1]_\beta^\gamma\right)_{ij} + \left([T_2]_\beta^\gamma\right)_{ij} && \text{(Proposition 2.27)} \\
&= \left(c[T_1]_\beta^\gamma + [T_2]_\beta^\gamma\right)_{ij}
\end{aligned}$$

for any  $i \in \{1, \dots, m\}$  and  $j \in \{1, \dots, n\}$ . To prove that  $\Phi_\beta^\gamma$  is invertible, let

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

be an arbitrary matrix in  $F^{m \times n}$ . By Lemma 2.21, there exists a unique linear transformation  $T : V \rightarrow W$  such that

$$T(x_j) = \sum_{i=1}^m a_{ij} y_i$$

for each  $j \in \{1, \dots, n\}$ , and it follows that  $[T]_\beta^\gamma = A$ . Thus, there exists  $(\Phi_\beta^\gamma)^{-1} : F^{m \times n} \rightarrow \mathcal{L}(V, W)$  such that  $(\Phi_\beta^\gamma)^{-1}(A) = T$  with  $[T]_\beta^\gamma = A$  for each  $A \in F^{m \times n}$ , which completes the proof.  $\square$

**Corollary 2.29.** If  $V$  and  $W$  are finite-dimensional vector spaces over  $F$  with  $\dim(V) = n$  and  $\dim(W) = m$ , then  $\mathcal{L}(V, W)$  is finite-dimensional with  $\dim(\mathcal{L}(V, W)) = mn$ .

*Proof.* Straightforward from Theorem 2.22 and Theorem 2.28.  $\square$



## 2.5 Matrix Multiplication

**Definition 2.30.** Let  $F$  be a field and let  $A \in F^{\ell \times m}$  and  $B \in F^{m \times n}$  be matrices. The **product** of  $A$  and  $B$ , denoted by  $AB$ , is a matrix in  $F^{\ell \times n}$  that satisfies

$$(AB)_{ik} = \sum_{j=1}^m A_{ij}B_{jk}$$

for  $i \in \{1, \dots, \ell\}$  and  $k \in \{1, \dots, n\}$ .

**Proposition 2.31.** Let  $U, V, W$  be vector spaces over  $F$ . If  $T_1 : U \rightarrow V$  and  $T_2 : V \rightarrow W$  are linear, then so is  $T_2T_1$ .

*Proof.* For  $a \in F$  and  $x, y \in U$ , we have

$$\begin{aligned} (T_2T_1)(ax + y) &= T_2(T_1(ax + y)) && \text{(composition of } T_1 \text{ and } T_2) \\ &= T_2(aT_1(x) + T_1(y)) && \text{(linearity of } T_1) \\ &= aT_2(T_1(x)) + T_2(T_1(y)) && \text{(linearity of } T_2) \\ &= a(T_2T_1)(x) + (T_2T_1)(y). && \text{(composition of } T_1 \text{ and } T_2) \end{aligned}$$

Thus,  $T_2T_1$  is linear.  $\square$

**Theorem 2.32.** Let  $U, V, W$  be finite-dimensional vector spaces with ordered bases

$$\alpha = (x_1, \dots, x_n), \quad \beta = (y_1, \dots, y_m), \quad \text{and} \quad \gamma = (z_1, \dots, z_\ell),$$

respectively. If  $T_1 : U \rightarrow V$  and  $T_2 : V \rightarrow W$  are linear, then

$$[T_2T_1]_\alpha^\gamma = [T_2]_\beta^\gamma [T_1]_\alpha^\beta.$$

*Proof.* Let  $A = [T_2]_\beta^\gamma$ ,  $B = [T_1]_\alpha^\beta$  and  $C = [T_2T_1]_\alpha^\gamma$ . Then

$$T_2(y_j) = \sum_{i=1}^{\ell} A_{ij}z_i, \quad T_1(x_k) = \sum_{j=1}^m B_{jk}y_j, \quad \text{and} \quad (T_2T_1)(x_k) = \sum_{i=1}^{\ell} C_{ik}z_i$$

hold for any  $j \in \{1, \dots, m\}$  and  $k \in \{1, \dots, n\}$ . Since for each  $k \in \{1, \dots, n\}$ ,

$$\begin{aligned} \sum_{i=1}^{\ell} C_{ik}z_i &= (T_2T_1)(x_k) \\ &= T_2(T_1(x_k)) \\ &= T_2\left(\sum_{j=1}^m B_{jk}y_j\right) \\ &= \sum_{j=1}^m B_{jk}T_2(y_j) \\ &= \sum_{j=1}^m B_{jk} \sum_{i=1}^{\ell} A_{ij}z_i \\ &= \sum_{i=1}^{\ell} \left(\sum_{j=1}^m A_{ij}B_{jk}\right) z_i, \end{aligned}$$

we have

$$C_{ik} = \sum_{j=1}^m A_{ij} B_{jk}$$

for each  $i \in \{1, \dots, \ell\}$  and  $k \in \{1, \dots, n\}$ . Thus,  $C = AB$ .  $\square$

**Corollary 2.33.** Let  $V$  and  $W$  be finite-dimensional vector spaces with ordered bases  $\beta$  and  $\gamma$  over a field  $F$ , respectively. If  $T : V \rightarrow W$  is linear, then

$$[T(x)]_\gamma = [T]_\beta^\gamma [x]_\beta$$

for each  $x \in V$ .

*Proof.* Let  $\alpha = (1_F)$  be an ordered basis for  $F$ . For each  $x \in V$ , let  $\varphi : F \rightarrow V$  be the linear transformation with  $\varphi(c) = cx$  for each  $c \in F$ . By Definition 2.26, we have

$$[\varphi]_\alpha^\beta = [\varphi(1_F)]_\beta \quad \text{and} \quad [T\varphi]_\alpha^\gamma = [(T\varphi)(1_F)]_\gamma.$$

Thus, it follows that

$$\begin{aligned} [T(x)]_\gamma &= [T(\varphi(1_F))]_\gamma \\ &= [(T\varphi)(1_F)]_\gamma \\ &= [T\varphi]_\alpha^\gamma \\ &= [T]_\beta^\gamma [\varphi]_\alpha^\beta && \text{(Theorem 2.32)} \\ &= [T]_\beta^\gamma [\varphi(1_F)]_\beta \\ &= [T]_\beta^\gamma [x]_\beta. \end{aligned} \quad \square$$

## 2.6 Left-Multiplication Transformations

**Definition 2.34.** Let  $A \in F^{m \times n}$  be a matrix. The **left-multiplication transformation** of  $A$ , denoted by  $L_A$ , is the transformation from  $F^n$  to  $F^m$  with

$$L_A(x) = Ax$$

for each  $x \in F^n$ .

**Proposition 2.35.** Let  $\alpha$ ,  $\beta$  and  $\gamma$  be standard ordered bases for  $F^n$ ,  $F^m$  and  $F^\ell$ , respectively. Then the following statements are true.

- (a)  $L_A$  is linear for each  $A \in F^{m \times n}$ .
- (b)  $[L_A]_\alpha^\beta = A$  for each  $A \in F^{m \times n}$ .
- (c)  $L_{cA+B} = cL_A + L_B$  for each  $c \in F$  and  $A, B \in F^{m \times n}$ .
- (d)  $L_{AB} = L_AL_B$  for each  $A \in F^{\ell \times m}$  and  $B \in F^{m \times n}$ .
- (e)  $L_{I_n} = I_{F^n}$ .

*Proof.*

- (a)  $L_A$  is linear since for any  $c \in F$  and  $x, y \in F^n$ ,

$$\begin{aligned} [L_A(cx + y)]_i &= [A(cx + y)]_i \\ &= \sum_{j=1}^n A_{ij} [cx + y]_j \\ &= \sum_{j=1}^n A_{ij} (cx_j + y_j) \\ &= c \sum_{j=1}^n A_{ij} x_j + \sum_{j=1}^n A_{ij} y_j \\ &= c[Ax]_i + [Ay]_i \\ &= [cAx + Ay]_i \\ &= [cL_A(x) + L_A(y)]_i \end{aligned}$$

holds for each  $i \in \{1, \dots, m\}$ .

- (b) Let  $T \in \mathcal{L}(V, W)$  be the transformation with  $[T]_\alpha^\beta = A$ . Then we have

$$T(x) = [T(x)]_\beta = [T]_\alpha^\beta [x]_\alpha = Ax$$

for each  $x \in F^n$  since  $\alpha$  and  $\beta$  are standard ordered bases. Thus,  $T = L_A$ .

- (c) It is proved by

$$[L_{cA+B}]_\alpha^\beta = cA + B = c[L_A]_\alpha^\beta + [L_B]_\alpha^\beta = [cL_A + L_B]_\alpha^\beta.$$

(d) It is proved by

$$[L_{AB}]_{\alpha}^{\gamma} = AB = [L_A]_{\beta}^{\gamma}[L_B]_{\alpha}^{\beta} = [L_AL_B]_{\alpha}^{\gamma}.$$

(e) Since

$$L_{I_n}(x) = I_n x = x = I_{F^n}(x)$$

holds for each  $x \in F^n$ ,  $L_{I_n} = I_{F^n}$ . □

**Lemma 2.36.** Let  $U, V, W, X$  be vector spaces. Let

$$T_1, T'_1 \in \mathcal{L}(U, V), \quad T_2, T'_2 \in \mathcal{L}(V, W), \quad \text{and} \quad T_3 \in \mathcal{L}(W, X)$$

be linear transformations and let  $c \in F$  be a scalar. Then the following statements are true.

- (a)  $T_1 I_U = T_1 = I_V T_1$ .
- (b)  $T_2(T_1 + T'_1) = T_2 T_1 + T_2 T'_1$ .
- (c)  $(T_2 + T'_2)T_1 = T_2 T_1 + T'_2 T_1$ .
- (d)  $c(T_2 T_1) = (c T_2)T_1 = T_2(c T_1)$ .
- (e)  $T_3(T_2 T_1) = (T_3 T_2)T_1$ .

*Proof.*

(a) Since

$$(T_1 I_U)(x) = T_1(I_U(x)) = T_1(x) = I_V(T_1(x)) = (I_V T_1)(x)$$

holds for each  $x \in U$ , we have  $T_1 I_U = T_1 = I_V T_1$ .

(b) Since

$$\begin{aligned} (T_2(T_1 + T'_1))(x) &= T_2((T_1 + T'_1)(x)) && \text{(composition)} \\ &= T_2(T_1(x) + T'_1(x)) && \text{(addition)} \\ &= T_2(T_1(x)) + T_2(T'_1(x)) && \text{(linearity)} \\ &= (T_2 T_1)(x) + (T_2 T'_1)(x) && \text{(composition)} \\ &= (T_2 T_1 + T_2 T'_1)(x) && \text{(addition)} \end{aligned}$$

holds for each  $x \in U$ , we have  $T_2(T_1 + T'_1) = T_2 T_1 + T_2 T'_1$ .

(c) Since

$$\begin{aligned} ((T_2 + T'_2)T_1)(x) &= (T_2 + T'_2)(T_1(x)) && \text{(composition)} \\ &= T_2(T_1(x)) + T'_2(T_1(x)) && \text{(addition)} \\ &= (T_2 T_1)(x) + (T'_2 T_1)(x) && \text{(composition)} \\ &= (T_2 T_1 + T'_2 T_1)(x) && \text{(addition)} \end{aligned}$$

holds for each  $x \in U$ , we have  $(T_2 + T'_2)T_1 = T_2 T_1 + T'_2 T_1$ .

(d) Since

$$\begin{aligned} (c(T_2T_1))(x) &= c(T_2T_1)(x) = cT_2(T_1(x)) \\ ((cT_2)T_1)(x) &= (cT_2)(T_1(x)) = cT_2(T_1(x)) \\ (T_2(cT_1))(x) &= T_2(cT_1(x)) = cT_2(T_1(x)) \end{aligned}$$

hold for each  $x \in U$ , we have  $c(T_2T_1) = (cT_2)T_1 = T_2(cT_1)$ .

(e) Since

$$\begin{aligned} (T_3(T_2T_1))(x) &= T_3((T_2T_1)(x)) && \text{(composition of } T_3 \text{ and } T_2T_1) \\ &= T_3(T_2(T_1(x))) && \text{(composition of } T_2 \text{ and } T_1) \\ &= (T_3T_2)(T_1(x)) && \text{(composition of } T_3 \text{ and } T_2) \\ &= ((T_3T_2)T_1)(x) && \text{(composition of } T_3T_2 \text{ and } T_1) \end{aligned}$$

holds for each  $x \in U$ , we have  $T_3(T_2T_1) = (T_3T_2)T_1$ .  $\square$

**Theorem 2.37.** Let  $A, A' \in F^{k \times \ell}$ ,  $B, B' \in F^{\ell \times m}$  and  $C \in F^{m \times n}$  be matrices and let  $c \in F$  be a scalar. Then the following statements are true.

(a)  $AI_\ell = A = I_kA$ .

(b)  $A(B + B') = AB + AB'$ .

(c)  $(A + A')B = AB + A'B$ .

(d)  $c(AB) = (cA)B = A(cB)$ .

(e)  $A(BC) = (AB)C$ .

*Proof.* Straightforward from Lemma 2.36.  $\square$

## 2.7 Invertible Matrices

**Definition 2.38.** A matrix  $A \in F^{n \times n}$  is **invertible** if  $L_A$  is invertible. If  $A$  is invertible, then it has an **inverse**, denoted by  $A^{-1}$ , which is the matrix in  $F^{n \times n}$  such that

$$L_{A^{-1}} = (L_A)^{-1}.$$

**Proposition 2.39.** The following statements are true for matrices  $A, B \in F^{n \times n}$ .

- (a) If  $A$  is invertible, then  $AA^{-1} = I_n = A^{-1}A$ .
- (b) If  $AB = I_n$ , then  $A$  and  $B$  are invertible. Furthermore,  $A = B^{-1}$  and  $B = A^{-1}$ .

*Proof.*

- (a) We have

$$L_{AA^{-1}} = L_AL_{A^{-1}} = L_A(L_A)^{-1} = I_{F^n} = L_{I_n}$$

and

$$L_{A^{-1}A} = L_{A^{-1}}L_A = (L_A)^{-1}L_A = I_{F^n} = L_{I_n},$$

implying  $AA^{-1} = I_n = A^{-1}A$ .

- (b) Since  $AB$  is invertible,  $L_{AB} = L_AL_B$  is injective and surjective. Thus,  $L_A : F^n \rightarrow F^n$  is injective and  $L_B : F^n \rightarrow F^n$  is surjective. It follows that  $L_A$  and  $L_B$  are bijective by Lemma 2.20, and thus are invertible, implying  $A$  and  $B$  are invertible. By Proposition 2.18 (c), we have  $L_A = (L_B)^{-1}$  and  $L_B = (L_A)^{-1}$ . Thus, we have  $A = B^{-1}$  and  $B = A^{-1}$ .  $\square$

# Chapter 3

## Systems of Linear Equations

### 3.1 Elementary Matrices

**Definition 3.1.** Any one of the following three operations on matrices is called an **elementary row operation**.

- (Type 1) Exchanging two different rows.
- (Type 2) Multiplying a row by a nonzero scalar.
- (Type 3) Adding a scalar multiple of a row to another row.

Similarly, any one of the following three operations on matrices is called an **elementary column operation**.

- (Type 1) Exchanging two different columns.
- (Type 2) Multiplying a column by a nonzero scalar.
- (Type 3) Adding a scalar multiple of a column to another column.

Furthermore, an **elementary operation** is either an elementary row operation or an elementary column operation.

**Definition 3.2.** A matrix  $X \in F^{n \times n}$  is **elementary** if it can be obtained from  $I_n$  by applying an elementary operation. We say that an elementary matrix is of type 1, 2, or 3 if its corresponding elementary operation is a type 1, 2, or 3 operation, respectively.

**Proposition 3.3.** Let  $X \in F^{m \times m}$  and  $Y \in F^{n \times n}$  be elementary matrices. Then the following statements hold for any matrix  $A \in F^{m \times n}$ .

- (a)  $XA$  is the matrix obtained from  $A$  by applying the elementary row operation corresponding to  $X$ .
- (b)  $AY$  is the matrix obtained from  $A$  by applying the elementary column operation corresponding to  $Y$ .

*Proof.* We will prove (a), and the proof of (b) is similar to that of (a) so that we omit it.

Let  $\gamma = (e_1, e_2, \dots, e_m)$  be the standard basis for  $F^m$ . Also, let

$$\text{row}(X) = (x_1, x_2, \dots, x_m) \quad \text{and} \quad \text{col}(A) = (c_1, c_2, \dots, c_n).$$

Then we have

$$(XA)_{ij} = \sum_{k=1}^m X_{ik}A_{kj} = \sum_{k=1}^m (x_i)_k(c_j)_k$$

for each  $1 \leq i \leq m$  and  $1 \leq j \leq n$ .

First, suppose that  $X$  is of type 1, obtained from  $I_m$  by exchanging the  $p$ -th row and the  $q$ -th row. It follows that  $x_p = e_q$ ,  $x_q = e_p$ , and  $x_i = e_i$  for each  $i \in \{1, \dots, m\} \setminus \{p, q\}$ . Thus,

$$\begin{aligned} (XA)_{pj} &= \sum_{k=1}^m (e_q)_k(c_j)_k = (c_j)_q = A_{qj} \\ (XA)_{qj} &= \sum_{k=1}^m (e_p)_k(c_j)_k = (c_j)_p = A_{pj} \\ (XA)_{ij} &= \sum_{k=1}^m (e_i)_k(c_j)_k = (c_j)_i = A_{ij} \quad \text{for } i \in \{1, \dots, m\} \setminus \{p, q\} \end{aligned}$$

hold for any  $j \in \{1, \dots, n\}$ , implying  $XA$  is the matrix obtained from  $A$  by exchanging the  $p$ -th row and the  $q$ -th row.

Secondly, suppose that  $X$  is of type 2, obtained from  $I_m$  by multiplying the  $p$ -th row by a scalar  $a$ . It follows that  $x_p = ae_p$  and  $x_i = e_i$  for  $i \in \{1, \dots, m\} \setminus \{p\}$ . Thus,

$$\begin{aligned} (XA)_{pj} &= \sum_{k=1}^m (ae_p)_k(c_j)_k = a(c_j)_p = aA_{pj} \\ (XA)_{ij} &= \sum_{k=1}^m (e_i)_k(c_j)_k = (c_j)_i = A_{ij} \quad \text{for } i \in \{1, \dots, m\} \setminus \{p\} \end{aligned}$$

hold for any  $j \in \{1, \dots, n\}$ , implying  $XA$  is the matrix obtained from  $A$  by multiplying the  $p$ -th row by a scalar  $a$ .

Finally, suppose that  $X$  is of type 3, obtained from  $I_m$  by adding the  $p$ -th row multiplied by  $a$  to the  $q$ -th row. It follows that  $x_q = ae_p + e_q$  and  $x_i = e_i$  for each  $i \in \{1, \dots, m\} \setminus \{q\}$ . Thus,

$$\begin{aligned} (XA)_{qj} &= \sum_{k=1}^m (ae_p + e_q)_k(c_j)_k = a(c_j)_p + (c_j)_q = aA_{pj} + A_{qj} \\ (XA)_{ij} &= \sum_{k=1}^m (e_i)_k(c_j)_k = (c_j)_i = A_{ij} \quad \text{for } i \in \{1, \dots, m\} \setminus \{q\} \end{aligned}$$

hold for any  $j \in \{1, \dots, n\}$ , implying  $XA$  is the matrix obtained from  $A$  by adding the  $p$ -th row multiplied by  $a$  to the  $q$ -th row.  $\square$

**Proposition 3.4.** Let  $X \in F^{n \times n}$  be an elementary matrix. Then  $X$  is invertible, and  $X^{-1}$  is also an elementary matrix.

*Proof.* There exists an elementary matrix  $Y \in F^{n \times n}$  with  $YX = I_n$  as follows.

- If  $X$  is of type 1 obtained from  $I_n$  by exchanging the  $p$ -th row and the  $q$ -th row, then  $Y$  is also of type 1 obtained from  $I_n$  by exchanging the  $p$ -th row and the  $q$ -th row.



- If  $X$  is of type 2 obtained from  $I_n$  by multiplying the  $p$ -th row by a scalar  $a$ , then  $Y$  is also of type 2 obtained from  $I_n$  by multiplying the  $p$ -th row by  $(1/a)$ .
- If  $X$  is of type 3 obtained from  $I_n$  by adding the  $p$ -th row multiplied by a scalar  $a$  to the  $q$ -th row, then  $Y$  is also of type 3 obtained from  $I_n$  by adding the  $p$ -th row multiplied by  $(-a)$  to the  $q$ -th row.

Thus, by Proposition 2.39 (b) we can conclude that  $X$  is invertible and  $Y = X^{-1}$ , which completes the proof.  $\square$

## 3.2 Rank and Nullity of Matrices

**Definition 3.5.** The **rank** and **nullity** of a matrix  $A \in F^{m \times n}$ , denoted by  $\text{rank}(A)$  and  $\text{nullity}(A)$ , respectively, are defined by

$$\begin{aligned}\text{rank}(A) &= \text{rank}(L_A) \\ \text{nullity}(A) &= \text{nullity}(L_A).\end{aligned}$$

**Theorem 3.6.** The following statements are true for any matrix  $A \in F^{m \times n}$ .

- (a)  $\mathcal{R}(L_A) = \text{span}(\text{col}(A))$ .
- (b)  $\text{rank}(A) = \dim(\text{span}(\text{col}(A)))$ .

*Proof.*

- (a) Let  $\beta = (x_1, \dots, x_n)$  and  $\gamma = (y_1, \dots, y_m)$  be the standard ordered basis for  $F^n$  and  $F^m$ , respectively. Then we have

$$Ax_i = [L_A(x_i)]_\gamma,$$

which is the  $i$ th column of  $[L_A]_\beta^\gamma = A$ . Thus, we have  $L_A(\beta) = \text{col}(A)$ , and it follows that

$$\mathcal{R}(L_A) = L_A(F^n) = L_A(\text{span}(\beta)) = \text{span}(L_A(\beta)) = \text{span}(\text{col}(A)).$$

- (b) By (a), we have

$$\text{rank}(A) = \text{rank}(L_A) = \dim(\mathcal{R}(L_A)) = \dim(\text{span}(\text{col}(A))). \quad \square$$

**Theorem 3.7.** If  $A \in F^{n \times n}$ , then  $A$  is invertible if and only if  $\text{rank}(A) = n$ .

*Proof.* ( $\Rightarrow$ ) Suppose that  $A$  is invertible. It follows that  $L_A : F^n \rightarrow F^n$  is also invertible, and thus is bijective. Therefore,

$$\text{rank}(A) = \text{rank}(L_A) = \dim(\mathcal{R}(L_A)) = \dim(F^n) = n.$$

( $\Leftarrow$ ) Suppose that  $\text{rank}(A) = n$ . Then we can conclude that  $\mathcal{R}(L_A) = F^n$  since  $\mathcal{R}(L_A)$  is a subspace of  $F^n$  with

$$\dim(\mathcal{R}(L_A)) = \text{rank}(L_A) = \text{rank}(A) = n = \dim(F^n).$$

Thus,  $L_A$  is surjective. It follows that  $L_A$  is bijective by Lemma 2.20, and thus  $L_A$  is invertible. Therefore,  $A$  is invertible.  $\square$

**Lemma 3.8.** Let  $V$  and  $W$  be vector spaces and let  $T : V \rightarrow W$  be linear. Let  $U$  be a subspace of  $V$ .

- (a)  $\dim(T(U)) \leq \dim(U)$ .
- (b) If  $T$  is injective, then  $\dim(T(U)) = \dim(U)$ .

*Proof.* Let  $S$  be a basis for  $U$ . Then we have  $T(U) = T(\text{span}(S)) = \text{span}(T(S))$ .

(a) Let  $Q$  be a basis for  $T(U)$ . By replacement theorem (Theorem 1.24),

$$\dim(T(U)) = |Q| \leq |T(S)| \leq |S| = \dim(U).$$

(b) If  $T$  is injective, then  $T(S)$  is linearly independent. Thus,  $T(S)$  is a basis for  $T(U)$ , implying

$$\dim(T(U)) = |T(S)| = |S| = \dim(U). \quad \square$$

**Theorem 3.9.** The following statements hold for any matrix  $A \in F^{m \times n}$ .

(a) If  $X \in F^{m \times m}$  is invertible, then  $\text{rank}(XA) = \text{rank}(A)$ .

(b) If  $Y \in F^{n \times n}$  is invertible, then  $\text{rank}(AY) = \text{rank}(A)$ .

*Proof.*

(a) Since  $X$  is invertible,  $L_X$  is invertible, and thus is bijective. It follows that  $\dim(L_X(U)) = \dim(U)$  for any subspace  $U$  of  $F^n$  since  $L_X$  is injective. Thus,

$$\begin{aligned} \text{rank}(XA) &= \text{rank}(L_X A) \\ &= \dim(L_X(L_A(F^n))) \\ &= \dim(L_A(F^n)) \\ &= \text{rank}(L_A) \\ &= \text{rank}(A). \end{aligned}$$

(b) Since  $Y$  is invertible,  $L_Y$  is invertible, and thus is bijective. It follows that  $L_Y(F^n) = F^n$  since  $L_Y$  is surjective. Thus,

$$\begin{aligned} \text{rank}(AY) &= \text{rank}(L_{AY}) \\ &= \dim(L_A(L_Y(F^n))) \\ &= \dim(L_A(F^n)) \\ &= \text{rank}(L_A) \\ &= \text{rank}(A). \end{aligned} \quad \square$$

**Theorem 3.10.** Let  $V$  and  $W$  be finite-dimensional vector spaces with bases  $\beta$  and  $\gamma$ , respectively. If  $T : V \rightarrow W$  is linear, then

$$\text{rank}(T) = \text{rank}([T]_{\beta}^{\gamma}).$$

*Proof.* Let  $A = [T]_{\beta}^{\gamma}$ . Since  $[T(x)]_{\gamma} = [T]_{\beta}^{\gamma}[x]_{\beta}$  holds for any  $x \in V$ , we have

$$\phi_{\gamma}T = L_A\phi_{\beta}.$$

Thus, since  $\phi_{\beta}$  and  $\phi_{\gamma}$  are invertible, we have

$$\text{rank}(T) = \text{rank}(\phi_{\gamma}T) = \text{rank}(L_A\phi_{\beta}) = \text{rank}(L_A) = \text{rank}(A). \quad \square$$

**Theorem 3.11.** Let  $A \in F^{m \times n}$  and let  $r$  be a nonnegative integer. Then  $\text{rank}(A) = r$  if and only if  $A$  can be transformed into

$$D = \begin{pmatrix} I_r & O_1 \\ O_2 & O_3 \end{pmatrix}$$

by performing a finite number of elementary operations.

*Proof.* ( $\Leftarrow$ ) Since  $A$  can be transformed into  $D$  by a finite number of elementary operations, there exist elementary matrices  $X_1, \dots, X_p \in F^{m \times m}$  and  $Y_1, \dots, Y_q \in F^{n \times n}$  such that

$$X_p \cdots X_1 A Y_1 \cdots Y_q = D.$$

Since elementary matrices are invertible,

$$\text{rank}(A) = \text{rank}(X_p \cdots X_1 A Y_1 \cdots Y_q) = \text{rank}(D) = r.$$

( $\Rightarrow$ ) If  $A$  is the zero matrix, then we have  $r = 0$ , and thus the theorem holds in this case with  $D = A$ . Now suppose that  $A$  is not the zero matrix. The proof is by induction on  $k = \min(m, n)$ .

First, we show that  $A$  can be transformed into some matrix  $B$  by a finite number of elementary operations as follows, where  $B_{11} = 1$ ,  $B_{1j} = 0$  and  $B_{i1} = 0$  for  $2 \leq i \leq m$  and  $2 \leq j \leq n$ .

1. First, we turn the  $(1, 1)$ -entry into a nonzero number by performing type 1 elementary operations.
  - a. If the first row contains only zeros, perform a type 1 row operation by exchanging the first row and a nonzero row.
  - b. If the  $(1, 1)$ -entry is zero, perform a type 1 column operation by exchanging the first column and a column whose first entry is not zero.
2. Then we turn the  $(1, 1)$ -entry into 1 by performing a type 2 operation.
3. Finally, we eliminate all nonzero entries in the first row and the first column except the  $(1, 1)$ -entry by performing type 3 operations.
  - a. For  $2 \leq i \leq m$ , if the  $(i, 1)$ -entry is nonzero, perform a type 3 row operation by adding a multiple of the first row to the  $i$ th row such that the  $(i, 1)$ -entry becomes zero.
  - b. For  $2 \leq j \leq n$ , if the  $(1, j)$ -entry is nonzero, perform a type 3 column operation by adding a multiple of the first column to the  $j$ th column such that the  $(1, j)$ -entry becomes zero.

By Theorem 3.9,  $\text{rank}(B) = \text{rank}(A) = r$  since  $B$  can be obtained from  $A$  by performing a finite number of elementary operations.

Now we prove the theorem by induction on  $\min(m, n)$ . For the induction basis, assume that  $m = 1$  or  $n = 1$  holds. Then  $\text{rank}(A) = 1$  since  $A$  is not the zero matrix, and thus the theorem holds with  $D = B$ .

Now assume that the theorem holds for  $\min(m, n) = k$  with  $k \geq 1$ , and we prove that the theorem also holds for  $\min(m, n) = k + 1$ . Since  $\min(m, n) \geq 2$ , we have

$$B = \left( \begin{array}{c|ccc} 1 & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & & \\ 0 & & & \end{array} \middle| \begin{array}{ccc} & & \\ & & \\ & & \\ & & \end{array} B' \right),$$

where  $B'$  is an  $(m-1) \times (n-1)$  matrix. Note that  $\text{rank}(B') = \text{rank}(B) - 1 = r - 1$ . By induction hypothesis,  $B'$  can be transformed into

$$D' = \begin{pmatrix} I_{r-1} & O_1 \\ O_2 & O_3 \end{pmatrix}$$

by a finite number of elementary row and column operations. It follows that

$$D = \left( \begin{array}{c|ccc} 1 & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & D' & \\ 0 & & & \end{array} \right)$$

is obtained from  $B$  by performing these operations. Thus,  $A$  can be transformed into  $D$  by a finite number of elementary operations, which completes the proof.  $\square$

**Corollary 3.12.** Let  $A \in F^{m \times n}$  and let  $r$  be a nonnegative integer. Then  $\text{rank}(A) = r$  if and only if there exist invertible  $X \in F^{m \times m}$  and  $Y \in F^{n \times n}$  such that

$$XAY = \begin{pmatrix} I_r & O_1 \\ O_2 & O_3 \end{pmatrix}.$$

*Proof.* ( $\Leftarrow$ ) It is proved by

$$\text{rank}(A) = \text{rank}(XAY) = r.$$

( $\Rightarrow$ ) By Theorem 3.11, there exist elementary matrices  $X_1, \dots, X_p \in F^{m \times m}$  and  $Y_1, \dots, Y_q \in F^{n \times n}$  such that

$$X_p \cdots X_1 A Y_1 \cdots Y_q = \begin{pmatrix} I_r & O_1 \\ O_2 & O_3 \end{pmatrix}.$$

Thus, the theorem holds by assigning  $X = X_p \cdots X_1$  and  $Y = Y_1 \cdots Y_q$ .  $\square$

**Theorem 3.13.** For any  $A \in F^{m \times n}$ ,  $\text{rank}(A^t) = \text{rank}(A)$ .

*Proof.* Let  $r = \text{rank}(A)$ . By Corollary 3.12, there exist invertible matrices  $X \in F^{m \times m}$  and  $Y \in F^{n \times n}$  such that

$$XAY = D = \begin{pmatrix} I_r & O_1 \\ O_2 & O_3 \end{pmatrix}$$

implying

$$Y^t A^t X^t = D^t.$$

Thus,

$$\text{rank}(A^t) = \text{rank}(Y^t A^t X^t) = \text{rank}(D^t) = r. \quad \square$$

**Theorem 3.14.**

- (a) Let  $U, V, W$  be finite-dimensional vector spaces over  $F$ . For any linear transformations  $T_1 : U \rightarrow V$  and  $T_2 : V \rightarrow W$ , we have

$$\text{rank}(T_2 T_1) \leq \text{rank}(T_1) \quad \text{and} \quad \text{rank}(T_2 T_1) \leq \text{rank}(T_2).$$

(b) For any matrices  $A \in F^{\ell \times m}$  and  $B \in F^{m \times n}$ , we have

$$\text{rank}(AB) \leq \text{rank}(A) \quad \text{and} \quad \text{rank}(AB) \leq \text{rank}(B).$$

*Proof.*

(a) By Lemma 3.8, we have

$$\text{rank}(T_2T_1) = \dim(T_2(T_1(U))) \leq \dim(T_1(U)) = \text{rank}(T_1).$$

Furthermore, since  $T_1(U) \subseteq V$ , we have  $T_2(T_1(U)) \subseteq T_2(V)$ . Thus,

$$\text{rank}(T_2T_1) = \dim(T_2(T_1(U))) \leq \dim(T_2(V)) = \text{rank}(T_2).$$

(b) By (a), we can conclude that

$$\begin{aligned} \text{rank}(AB) &= \text{rank}(L_{AB}) = \text{rank}(L_AL_B) \leq \text{rank}(L_A) = \text{rank}(A) \\ \text{rank}(AB) &= \text{rank}(L_{AB}) = \text{rank}(L_AL_B) \leq \text{rank}(L_B) = \text{rank}(B). \end{aligned}$$

□

### 3.3 Matrix Inverses

**Theorem 3.15.** Every invertible matrix is a product of elementary matrices.

*Proof.* Suppose  $A$  is an invertible  $n \times n$  matrix. Since  $\text{rank}(A) = n$ , there exist elementary matrices  $X_1, \dots, X_p, Y_1, \dots, Y_q \in F^{n \times n}$  such that

$$X_p \cdots X_1 A Y_1 \cdots Y_q = I_n,$$

implying

$$A = X_1^{-1} \cdots X_p^{-1} Y_q^{-1} \cdots Y_1^{-1}.$$

Since the inverses of elementary matrices are elementary matrices, we can conclude that  $A$  is a product of elementary matrices.  $\square$

### 3.4 Systems of Linear Equations

**Definition 3.16.** The system  $E$  of equations

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m, \end{aligned}$$

where  $a_{ij}$  and  $b_i$  are scalars in a field  $F$  and  $x_1, x_2, \dots, x_n$  are  $n$  variables that take values in  $F$ , is called a system of  $m$  **linear equations** in  $n$  unknowns over the field  $F$ . Furthermore, it can be rewritten as a matrix equation

$$E : Ax = b$$

with

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}, \quad x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \quad \text{and} \quad b = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix},$$

and the matrices

$$A \in F^{m \times n} \quad \text{and} \quad (A \mid b) \in F^{m \times (n+1)}$$

are called the **coefficient matrix** and the **augmented matrix** of  $E$ , respectively.

**Definition 3.17.** For any system  $E : Ax = b$  of linear equations with  $A \in F^{m \times n}$ , the **solution set** of  $E$ , denoted by  $S(E)$ , is defined by

$$S(E) = \{s \in F^n : As = b\}.$$

Each element of  $S(E)$  is called a **solution** to  $E$ .

**Theorem 3.18.** If  $E : Ax = b$  is a system of linear equations, then  $S(E)$  is nonempty if and only if  $\text{rank}(A) = \text{rank}(A \mid b)$ .

*Proof.* It is proved by

$$\begin{aligned} S(E) \neq \emptyset &\Leftrightarrow Ax = b \text{ for some } x \in F^n \\ &\Leftrightarrow b \in \mathcal{R}(L_A) \\ &\Leftrightarrow b \in \text{span}(\text{col}(A)) \\ &\Leftrightarrow \text{span}(\text{col}(A)) = \text{span}(\text{col}(A \mid b)) \\ &\Leftrightarrow \text{rank}(A) = \text{rank}(A \mid b). \end{aligned}$$

□

**Definition 3.19.** A system  $E : Ax = b$  of linear equations with  $A \in F^{m \times n}$  is said to be **homogeneous** if  $b = 0_{F^m}$ .

**Proposition 3.20.** The following statements are true for any homogeneous system  $E : Ax = 0_{F^m}$  of linear equations with  $A \in F^{m \times n}$ .

- (a)  $S(E) = \mathcal{N}(L_A)$ .



(b)  $S(E)$  is a subspace of  $A$  with  $\dim(S(E)) = \text{nullity}(A)$ .

*Proof.* Straightforward. □

**Definition 3.21.** For any system

$$E : Ax = b$$

of linear equations with  $A \in F^{m \times n}$ , the system

$$E_H : Ax = 0_{F^m}$$

of linear equations is called the **homogeneous system** corresponding to  $E$ .

**Proposition 3.22.** For any system  $E : Ax = b$  of linear equations with  $A \in F^{m \times n}$ ,

$$S(E) = \{s\} + S(E_H)$$

holds for any solution  $s \in S(E)$ .

*Proof.* For any  $r \in F^n$ , we have

$$\begin{aligned} r \in S(E) &\Leftrightarrow Ar = b \\ &\Leftrightarrow A(r - s) = 0_{F^m} \\ &\Leftrightarrow r - s \in S(E_H) \\ &\Leftrightarrow r \in \{s\} + S(E_H). \end{aligned}$$

□

**Theorem 3.23.** Let  $E : Ax = b$  be a system of linear equations with  $A \in F^{n \times n}$ . Then  $A$  is invertible if and only if  $E$  has exactly one solution.

*Proof.* ( $\Rightarrow$ ) Suppose that  $s \in F^n$  is a solution to  $E$ . Then we have  $As = b$ , implying  $s = A^{-1}b$ . Thus,  $S(E) = \{A^{-1}b\}$ .

( $\Leftarrow$ ) Let  $s \in F^n$  be the unique solution to  $E$ . Since  $S(E) = \{s\} + S(E_H)$ , we can conclude that  $S(E_H) = \{0_{F^n}\}$ , implying

$$\text{rank}(A) = n - \text{nullity}(A) = n - \dim(S(E_H)) = n - 0 = n.$$

Thus,  $A$  is invertible. □

**Theorem 3.24.** Let  $E : Ax = b$  and  $E' : A'x = b'$  be systems of linear equations with  $A, A' \in F^{m \times n}$ . If there is an invertible matrix  $X \in F^{m \times m}$  with

$$X(A \mid b) = (A' \mid b'),$$

then  $S(E) = S(E')$ .

*Proof.* For any  $s \in F^n$ , we have

$$\begin{aligned} s \in S(E) &\Leftrightarrow As = b \\ &\Leftrightarrow X(As) = Xb \\ &\Leftrightarrow A's = b' \\ &\Leftrightarrow s \in S(E'). \end{aligned}$$

□

**Definition 3.25.** A matrix is said to be in **reduced row echelon form** if it satisfies the following conditions.

- (a) Any nonzero rows are above rows with all zeros.
- (b) The first nonzero entry in each row is  $1_F$  and it occurs to the right of the first nonzero entry above it.
- (c) The first nonzero entry in each row is the only nonzero entry in its column.

**Theorem 3.26.** Any matrix can be transformed into a matrix in reduced row echelon form by a finite number of elementary row operations.

*Proof.* One can repeat the following steps until all rows are processed or all nonzero columns are processed. At first, all rows and all columns has not been processed.

1. Find  $i$  such that the  $i$ th row is the first row that has not been processed, and find  $j$  such that the  $j$ th column is the first nonzero column that has not been processed.
2. If  $(i, j)$ -entry is zero, perform a type 1 row operation such that the  $(i, j)$ -entry becomes nonzero.
3. Perform a type 2 row operation to turn the  $(i, j)$ -entry into  $1_F$ .
4. Perform type 3 row operations such that the  $(i, j)$ -entry becomes the only nonzero entry in the  $j$ th column.
5. Mark the  $i$ th row and the  $j$ th column as processed.

After the process above, any matrix should be transformed into a matrix in reduced row echelon form.  $\square$

**Remark.** The algorithm in the proof above is called **Gaussian-Jordan elimination**.

# Chapter 4

## Determinants

### 4.1 Characterization of the Determinant

**Definition 4.1.** A function  $\delta : F^{n \times n} \rightarrow F$  is  **$n$ -linear** if

$$\delta(A) = k\delta(B) + \delta(C)$$

holds for any matrices  $A, B, C \in F^{n \times n}$  satisfying the following properties for any  $i \in \{1, \dots, n\}$  and for any  $k \in F$ .

- The  $j$ th rows of  $A, B$  and  $C$  are identical for each  $j \in \{1, \dots, n\} \setminus \{i\}$ .
- The  $i$ th row of  $A$  is the sum of the  $i$ th row of  $B$  multiplied by  $k$  and the  $i$ th row of  $C$ .

**Definition 4.2.** An  $n$ -linear function  $\delta : F^{n \times n} \rightarrow F$  is **alternating** if

$$\delta(A) = 0_F$$

holds for any matrix  $A \in F^{n \times n}$  that has two identical rows.

**Proposition 4.3.** Let  $\delta : F^{n \times n} \rightarrow F$  be an alternating  $n$ -linear function and let  $A \in F^{n \times n}$ . Then the following statements are true.

- (a) If  $E_1 \in F^{n \times n}$  is an elementary matrix of type 1, then  $\delta(E_1 A) = -\delta(A)$ .
- (b) If  $E_2 \in F^{n \times n}$  is an elementary matrix of type 2 obtained by multiplying one row of  $I_n$  by scalar  $k \in F$ , then  $\delta(E_2 A) = k\delta(A)$ .
- (c) If  $E_3 \in F^{n \times n}$  is an elementary matrix of type 3, then  $\delta(E_3 A) = \delta(A)$ .

*Proof.* Let  $\text{row}(A) = (x_1, \dots, x_n)$ .

- (a) Let  $E_1$  be obtained from  $I_n$  by interchanging the  $p$ th row and the  $q$ th row with

$p < q$ . Then we have

$$\begin{aligned}
0_F &= \delta \begin{pmatrix} x_1 \\ \vdots \\ x_p + x_q \\ \vdots \\ x_p + x_q \\ \vdots \\ x_n \end{pmatrix} = \delta \begin{pmatrix} x_1 \\ \vdots \\ x_p \\ \vdots \\ x_p + x_q \\ \vdots \\ x_n \end{pmatrix} + \delta \begin{pmatrix} x_1 \\ \vdots \\ x_q \\ \vdots \\ x_p + x_q \\ \vdots \\ x_n \end{pmatrix} \\
&= \delta \begin{pmatrix} x_1 \\ \vdots \\ x_p \\ \vdots \\ x_p \\ \vdots \\ x_n \end{pmatrix} + \delta \begin{pmatrix} x_1 \\ \vdots \\ x_p \\ \vdots \\ x_q \\ \vdots \\ x_n \end{pmatrix} + \delta \begin{pmatrix} x_1 \\ \vdots \\ x_q \\ \vdots \\ x_p \\ \vdots \\ x_n \end{pmatrix} + \delta \begin{pmatrix} x_1 \\ \vdots \\ x_q \\ \vdots \\ x_q \\ \vdots \\ x_n \end{pmatrix} \\
&= 0_F + \delta(A) + \delta(E_1 A) + 0_F.
\end{aligned}$$

Thus,  $\delta(E_1 A) = -\delta(A)$ .

- (b) Let  $E_2$  be obtained from  $I_n$  by multiplying the  $p$ th row by some scalar  $k$ . Then we have

$$\delta(E_2 A) = \delta \begin{pmatrix} x_1 \\ \vdots \\ kx_p \\ \vdots \\ x_n \end{pmatrix} = k\delta \begin{pmatrix} x_1 \\ \vdots \\ x_p \\ \vdots \\ x_n \end{pmatrix} = k\delta(A).$$

- (c) Let  $E_3$  be obtained from  $I_n$  by adding the  $p$ th row multiplied by some scalar  $k$  to the  $q$ th row. If  $p < q$ , then we have

$$\delta(E_3 A) = \delta \begin{pmatrix} x_1 \\ \vdots \\ x_p \\ \vdots \\ kx_p + x_q \\ \vdots \\ x_n \end{pmatrix} = k\delta \begin{pmatrix} x_1 \\ \vdots \\ x_p \\ \vdots \\ x_p \\ \vdots \\ x_n \end{pmatrix} + \delta \begin{pmatrix} x_1 \\ \vdots \\ x_p \\ \vdots \\ x_q \\ \vdots \\ x_n \end{pmatrix} = k0_F + \delta(A) = \delta(A).$$

The case that  $q < p$  can be proved similarly.  $\square$

**Theorem 4.4.** Let  $\delta : F^{n \times n} \rightarrow F$  be an alternating  $n$ -linear function and let  $A \in F^{n \times n}$ . If  $\text{rank}(A) < n$ , then  $\delta(A) = 0_F$ .

*Proof.* Since

$$\dim(\text{span}(\text{row}(A))) = \text{rank}(A^t) = \text{rank}(A) < n,$$

the rows of  $A$  is not a spanning set of  $F^n$ , and thus is linearly dependent, implying that there exists a row which is a linear combination of the other rows.

Therefore,  $A$  can be transformed into a matrix  $B$  that has two identical rows by a finite number of elementary row operations. It follows that

$$\delta(A) = \delta(E_p \cdots E_1 A) = \delta(B) = 0_F,$$

where  $E_1, \dots, E_p \in F^{n \times n}$  are elementary matrices.  $\square$

**Theorem 4.5.** Let  $\delta : F^{n \times n} \rightarrow F$  be an alternating  $n$ -linear function such that  $\delta(I_n) = 1_F$ . Then for any  $A, B \in F^{m \times n}$ , we have

$$\delta(AB) = \delta(A)\delta(B).$$

*Proof.* First, suppose that  $\text{rank}(A) < n$ . Then we have  $\text{rank}(AB) < n$ . Thus,

$$\delta(AB) = 0_F = \delta(A)\delta(B).$$

Now suppose that  $\text{rank}(A) = n$ . That is,  $A$  is invertible, and thus  $A = E_k \cdots E_1$  for some elementary matrices  $E_1, \dots, E_k \in F^{n \times n}$ . Then we have

$$\begin{aligned} \delta(AB) &= \delta(E_k \cdots E_1 B) \\ &= \delta(E_k) \cdots \delta(E_1) \delta(B) \\ &= \delta(E_k) \cdots \delta(E_1) \delta(I_n) \delta(B) & (\delta(I_n) = 1_F) \\ &= \delta(E_k \cdots E_1 I_n) \delta(B) \\ &= \delta(A) \delta(B). \end{aligned} \quad \square$$

**Theorem 4.6.** There exists a unique alternating  $n$ -linear function  $\delta : F^{n \times n} \rightarrow F$  with  $\delta(I_n) = 1_F$ .

*Proof.* Suppose that  $\delta, \delta' : F^{n \times n} \rightarrow F$  are alternating  $n$ -linear functions with  $\delta(I_n) = 1_F = \delta'(I_n)$ . We prove that  $\delta(A) = \delta'(A)$  for any  $A \in F^{n \times n}$ . If  $\text{rank}(A) < n$ , then

$$\delta(A) = 0_F = \delta'(A).$$

If  $\text{rank}(A) = n$ , then  $A$  is invertible, and thus

$$A = E_p \cdots E_1$$

for some elementary matrices  $E_1, \dots, E_p \in F^{n \times n}$ . It follows that

$$\begin{aligned} \delta(A) &= \delta(E_p \cdots E_1 I_n) \\ &= \delta(E_p) \cdots \delta(E_1) \delta(I_n) \\ &= \delta'(E_p) \cdots \delta'(E_1) \delta(I_n) \\ &= \delta'(E_p \cdots E_1 I_n) \\ &= \delta'(A). \end{aligned} \quad \square$$

**Definition 4.7.** The **determinant** of  $A \in F^{n \times n}$  is

$$\det(A) = \delta(A),$$

where  $\delta : F^{n \times n} \rightarrow F$  is the alternating  $n$ -linear function with  $\delta(I_n) = 1_F$ .

## 4.2 Permutations

**Definition 4.8.**

- A function  $\sigma : \{1, 2, \dots, n\} \rightarrow \{1, 2, \dots, n\}$  is a **permutation** over  $\{1, 2, \dots, n\}$  if  $\sigma$  is bijective. The set of all permutations over  $\{1, 2, \dots, n\}$  is denoted by  $S_n$ .
- An inversion of  $\sigma \in S_n$  is a pair  $(i, j)$  with  $1 \leq i < j \leq n$  and  $\sigma(i) > \sigma(j)$ . The number of inversions of  $\sigma$  is denoted by  $\rho(\sigma)$ .
- The **sign** of  $\sigma \in S_n$  is defined by

$$\text{sgn}(\sigma) = (-1)^{\rho(\sigma)}.$$

**Theorem 4.9.** For any matrix  $A \in F^{n \times n}$ ,

$$\det(A) = \sum_{\sigma \in S_n} \text{sgn}(\sigma) \prod_{i=1}^n A_{i, \sigma(i)}.$$

*Proof.* Let  $\delta : F^{n \times n} \rightarrow F$  be the function

$$\delta(A) = \sum_{\sigma \in S_n} \text{sgn}(\sigma) \prod_{i=1}^n A_{i, \sigma(i)}.$$

We prove that  $\delta$  is an alternating  $n$ -linear function with  $\delta(I_n) = 1_F$ .

First, we show that  $\delta$  is  $n$ -linear. Suppose that  $A, B, C \in F^{n \times n}$  are matrices satisfying the following properties for any  $p \in \{1, \dots, n\}$  and for any  $k \in F$ .

- The  $i$ th rows of  $A, B$  and  $C$  are identical for each  $i \in \{1, \dots, n\} \setminus \{p\}$ .
- The  $p$ th row of  $A$  is the sum of the  $p$ th row of  $B$  multiplied by  $k$  and the  $p$ th row of  $C$ .

Then we have

$$\begin{aligned} \delta(A) &= \sum_{\sigma \in S_n} \text{sgn}(\sigma) A_{p, \sigma(p)} \prod_{\substack{1 \leq i \leq n \\ i \neq p}} A_{i, \sigma(i)} \\ &= \sum_{\sigma \in S_n} \text{sgn}(\sigma) (kB_{p, \sigma(p)} + C_{p, \sigma(p)}) \prod_{\substack{1 \leq i \leq n \\ i \neq p}} A_{i, \sigma(i)} \\ &= k \sum_{\sigma \in S_n} \text{sgn}(\sigma) B_{p, \sigma(p)} \prod_{\substack{1 \leq i \leq n \\ i \neq p}} A_{i, \sigma(i)} + \sum_{\sigma \in S_n} \text{sgn}(\sigma) C_{p, \sigma(p)} \prod_{\substack{1 \leq i \leq n \\ i \neq p}} A_{i, \sigma(i)} \\ &= k \sum_{\sigma \in S_n} \text{sgn}(\sigma) B_{p, \sigma(p)} \prod_{\substack{1 \leq i \leq n \\ i \neq p}} B_{i, \sigma(i)} + \sum_{\sigma \in S_n} \text{sgn}(\sigma) C_{p, \sigma(p)} \prod_{\substack{1 \leq i \leq n \\ i \neq p}} C_{i, \sigma(i)} \\ &= k\delta(B) + \delta(C). \end{aligned}$$

Now we show that  $\delta$  is alternating. Suppose that  $D \in F^{n \times n}$  is a matrix whose  $p$ th row and  $q$ th row are identical with  $p \neq q$ . For each  $\sigma \in S_n$ , let  $\sigma' \in S_n$  be the permutation that satisfies the following properties.

- $\sigma'(p) = \sigma(q)$  and  $\sigma'(q) = \sigma(p)$ .
- $\sigma'(i) = \sigma(i)$  for each  $i \in \{1, \dots, n\} \setminus \{p, q\}$ .

Then we have

$$\begin{aligned}
\delta(D) &= \sum_{\sigma \in S_n} \text{sgn}(\sigma) \prod_{1 \leq i \leq n} D_{i, \sigma(i)} \\
&= \sum_{\substack{\sigma \in S_n \\ \sigma(p) < \sigma(q)}} \text{sgn}(\sigma) \prod_{1 \leq i \leq n} D_{i, \sigma(i)} + \sum_{\substack{\sigma \in S_n \\ \sigma(p) > \sigma(q)}} \text{sgn}(\sigma) \prod_{1 \leq i \leq n} D_{i, \sigma(i)} \\
&= \sum_{\substack{\sigma \in S_n \\ \sigma(p) < \sigma(q)}} \text{sgn}(\sigma) \prod_{1 \leq i \leq n} D_{i, \sigma(i)} + \sum_{\substack{\sigma \in S_n \\ \sigma(p) < \sigma(q)}} \text{sgn}(\sigma') \prod_{1 \leq i \leq n} D_{i, \sigma'(i)} \\
&= \sum_{\substack{\sigma \in S_n \\ \sigma(p) < \sigma(q)}} (\text{sgn}(\sigma) + \text{sgn}(\sigma')) \prod_{1 \leq i \leq n} D_{i, \sigma(i)} \\
&= 0_F.
\end{aligned}$$

Finally, we have

$$\delta(I_n) = \text{sgn}(\sigma_0) = 1_F,$$

where  $\sigma_0$  is the identity permutation. Therefore,  $\delta$  is an alternating  $n$ -linear function with  $\delta(I_n) = 1_F$ , and by Theorem 4.6 we can conclude that it is exactly the determinant function.  $\square$

# Chapter 5

## Diagonalization

### 5.1 Eigenvalues and Eigenvectors

**Definition 5.1.** Let  $T : V \rightarrow V$  be a linear operator on a vector space  $V$  over a field  $F$ . If

$$T(x) = \lambda x$$

holds for some scalar  $\lambda \in F$  and some vector  $x \in V \setminus \{0_V\}$ , then  $(\lambda, x)$  is called an **eigenpair** of  $T$ , with  $\lambda$  and  $x$  called an **eigenvalue** and an **eigenvector** of  $T$ , respectively.

**Definition 5.2.** Let  $V$  be a vector space over a field  $F$  with  $\dim(V) = n$ . Let  $T \in \mathcal{L}(V)$ . An **eigenbasis** of  $V$  for  $T$  is an ordered basis of  $V$  in which every vector is an eigenvector of  $T$ .



## 5.2 Characteristic Polynomials and Eigenspaces

**Definition 5.3.** The **characteristic polynomial** of  $A \in F^{n \times n}$  is the polynomial  $f_A \in \mathcal{P}(F)$  with

$$f_A(t) = \det(A - tI_n).$$

**Definition 5.4.** Let  $V$  be a vector space over a field  $F$ . Let  $T : V \rightarrow V$  be linear and let  $\lambda$  be an eigenvalue of  $T$ . Then

$$E_T(\lambda) = \mathcal{N}(T - \lambda I_V).$$

is called the **eigenspace** of  $T$  with respect to  $\lambda$ .

# Chapter 6

## Inner Product Spaces

### 6.1 Inner Products and Norms

**Definition 6.1.** Let  $V$  be a vector space over a field  $F \in \{\mathbb{R}, \mathbb{C}\}$ . A function

$$\langle \cdot | \cdot \rangle : V \times V \rightarrow F$$

is called an **inner product** on  $V$  if it satisfies the following properties for all  $x, x', y \in V$ .

- (a)  $\langle ax + x' | y \rangle = a \langle x | y \rangle + \langle x' | y \rangle$ .
- (b)  $\langle x | y \rangle = \overline{\langle y | x \rangle}$ .
- (c)  $\langle x | x \rangle \in \mathbb{R}^+$  for any  $x \in V \setminus \{0_V\}$ .

A vector space equipped with an inner product is called an **inner product space**.

**Proposition 6.2.** Let  $V$  be an inner product space over a field  $F \in \{\mathbb{R}, \mathbb{C}\}$ . Then the following statements are true for  $x, y, y' \in V$  and  $a \in F$ .

- (a)  $\langle x | ay + y' \rangle = \bar{a} \langle x | y \rangle + \langle x | y' \rangle$ .
- (b)  $\langle x | 0_V \rangle = 0_F = \langle 0_V | x \rangle$ .
- (c)  $\langle x | x \rangle = 0_F$  if and only if  $x = 0_V$ .
- (d) If  $\langle x | y \rangle = \langle x | y' \rangle$  holds for all  $x \in V$ , then  $y = y'$ .

*Proof.*

- (a) It is proved by

$$\langle x | ay + y' \rangle = \overline{\langle ay + y' | x \rangle} = \overline{a \langle y | x \rangle + \langle y' | x \rangle} = \bar{a} \langle x | y \rangle + \langle x | y' \rangle.$$

- (b) By

$$\langle x | x \rangle = \langle x | 1_F x + 0_V \rangle = \overline{1_F} \langle x | x \rangle + \langle x | 0_V \rangle = \langle x | x \rangle + \langle x | 0_V \rangle$$

and

$$\langle x | x \rangle = \langle 1_F x + 0_V | x \rangle = 1_F \langle x | x \rangle + \langle 0_V | x \rangle = \langle x | x \rangle + \langle 0_V | x \rangle,$$

we have  $\langle x | 0_V \rangle = 0_F = \langle 0_V | x \rangle$ .

(c) ( $\Leftarrow$ ) If  $x = 0_V$ , then  $\langle x | x \rangle = 0_F$  by (b).

( $\Rightarrow$ ) If  $\langle x | x \rangle = 0_F$ , then  $x = 0_V$  by Definition 6.1 (c).

(d) Note that

$$\langle x | y - y' \rangle = \langle x | y \rangle + \overline{(-1_F)} \langle x | y' \rangle = 0_F$$

holds for all  $x \in V$ . Since  $\langle y - y' | y - y' \rangle = 0_F$ , we have  $y - y' = 0_V$ , and thus  $y = y'$ .  $\square$

**Definition 6.3.** Let  $V$  be an inner product space over a field  $F$ .

- Vectors  $x$  and  $y$  in  $V$  are **orthogonal** if

$$\langle x | y \rangle = 0_F.$$

- A subset  $S$  of  $V$  is **orthogonal** if any two distinct vectors in  $S$  are orthogonal.

**Theorem 6.4.** Let  $V$  be an inner product space over a field  $F$ . Let  $S$  be an orthogonal subset of  $V \setminus \{0_V\}$  and let  $x_1, \dots, x_n$  be distinct vectors in  $S$ . Then for  $y \in V$ , if

$$y = \sum_{i=1}^n a_i x_i$$

for some  $a_1, \dots, a_n \in F$ , then

$$a_i = \frac{\langle y | x_i \rangle}{\langle x_i | x_i \rangle}$$

for each  $i \in \{1, \dots, n\}$ .

*Proof.* For each  $i \in \{1, \dots, n\}$ , we have

$$\langle y | x_i \rangle = \left\langle \sum_{j=1}^n a_j x_j \middle| x_i \right\rangle = \sum_{j=1}^n a_j \langle x_j | x_i \rangle = a_i \langle x_i | x_i \rangle,$$

implying

$$a_i = \frac{\langle y | x_i \rangle}{\langle x_i | x_i \rangle}. \quad \square$$

**Corollary 6.5.** Let  $V$  be an inner product space over a field  $F$ . If  $S$  is an orthogonal subset of  $V \setminus \{0_V\}$ , then  $S$  is linearly independent.

*Proof.* Suppose that there exist scalars  $a_1, \dots, a_n \in F$  and distinct vectors  $x_1, \dots, x_n \in S$  such that

$$\sum_{i=1}^n a_i x_i = 0_V.$$

Then we have

$$a_i = \frac{\langle 0_V | x_i \rangle}{\langle x_i | x_i \rangle} = 0_F$$

for each  $i \in \{1, \dots, n\}$ . Thus,  $S$  is linearly independent.  $\square$

**Theorem 6.6 (Gram-Schmidt Process).** Let  $V$  be a finite-dimensional inner product space over a field  $F$ . Let  $R = \{x_1, \dots, x_n\}$  be a linearly independent subset of  $V$ . Then the set  $S = \{y_1, \dots, y_n\}$  with

$$y_i = x_i - \sum_{j=1}^{i-1} \frac{\langle x_i | y_j \rangle}{\langle y_j | y_j \rangle} y_j$$

for  $1 \leq i \leq n$  is an orthogonal set of nonzero vectors satisfying  $\text{span}(S) = \text{span}(R)$ .

*Proof.* The proof is by induction on  $n$ . The theorem holds for  $n = 0$ . To show the induction step, let  $n \geq 1$ . By the induction hypothesis,  $\langle y_j | y_i \rangle = 0_F$  for distinct  $i, j \in \{1, \dots, n-1\}$ . Then since for  $i \in \{1, \dots, n-1\}$ , we have

$$\begin{aligned} \langle y_n | y_i \rangle &= \left\langle x_n - \sum_{j=1}^{n-1} \frac{\langle x_n | y_j \rangle}{\langle y_j | y_j \rangle} y_j \middle| y_i \right\rangle \\ &= \langle x_n | y_i \rangle - \sum_{j=1}^{n-1} \frac{\langle x_n | y_j \rangle}{\langle y_j | y_j \rangle} \langle y_j | y_i \rangle \\ &= \langle x_n | y_i \rangle - \frac{\langle x_n | y_i \rangle}{\langle y_i | y_i \rangle} \langle y_i | y_i \rangle \\ &= 0_F, \end{aligned}$$

we can conclude that  $S$  is orthogonal. Furthermore, if  $y_n = 0_V$ , then

$$x_n \in \text{span}(\{y_1, \dots, y_{n-1}\}) = \text{span}(\{x_1, \dots, x_{n-1}\})$$

because

$$x_n = y_n + \sum_{j=1}^{n-1} \frac{\langle x_n | y_j \rangle}{\langle y_j | y_j \rangle} y_j,$$

contradiction to the fact that  $R$  is linearly independent. Thus,  $y_n \neq 0_V$ , implying  $0_V \notin S$ . It follows that  $S$  is linearly independent by Corollary 6.5. Therefore, since  $|S| = \dim(\text{span}(R))$ , we have  $\text{span}(S) = \text{span}(R)$ .  $\square$