Linear Algebra

Chapter 1

Vector Spaces

1.1 Abelian Groups and Fields

Definition 1.1. Let G be a set, and let $\star$ be a binary operation defined on G. We say that $(G, \star)$ is an **Abelian group** if it satisfies the following properties.

1. Closeness: If $a \in G$ and $b \in G$, then $a \star b \in G$.
2. Commutativity: We have $a \star b = b \star a$ for any $a, b \in G$.
3. Associativity: We have $(a \star b) \star c = a \star (b \star c)$ for any $a, b, c \in G$.
4. Existence of identity: There exists an element $e \in G$, called the **identity**, such that $a \star e = a$ for any $a \in G$.
5. Existence of inverses: For all $a \in G$, there exists an element $b \in G$, called the **inverse** of a, such that $a \star b = e$.

Definition 1.2. Let F be a field, and let $+$ and $\cdot$ be binary operations defined on F. We say that $(F, +, \cdot)$ is a **field** if it satisfies the following conditions.

1. Additive Abelian group: $(F, +)$ is an Abelian group with identity 0_F.
2. Multiplicative Abelian group: $(F \setminus \{0_F\}, \cdot)$ is an Abelian group with identity 1_F.
3. Distributivity: $a \cdot (b + c) = a \cdot b + a \cdot c$ for any $a, b, c \in F$.

Let $-a$ and a^{-1} denote the additive inverse and the multiplicative inverse of a, respectively. Subtraction and division are defined as follows.

- For $a \in F$ and $b \in F$, we define $a - b = a + (-b)$.
- For $a \in F$ and $b \in F \setminus \{0_F\}$, we define $a/b = a \cdot b^{-1}$.

Remark. We may use the underlying set F to represent the entire field if addition and multiplication are self-explanatory.

Examples. The set $\mathbb{Q}$ of rational numbers, the set $\mathbb{R}$ of real numbers, and the set $\mathbb{C}$ of complex numbers are all fields.

Theorem 1.3 (Cancellation Laws). Let F be a field with $a, b, c \in F$.

(a) If $a + c = b + c$, then $a = b$.

(b) If $a \cdot c = b \cdot c$ and $c \neq 0_F$, then $a = b$.

Proof. The proof of (a) follows from the definition of fields. We have

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

The proof of (b) is similar to the proof of (a). □

Corollary 1.4. The identity and inverse elements in a field are unique. That is, the following statements are true for any field F with $a, b \in F$.

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

(b) If $a \cdot b = a$ and $a \neq 0_F$, then $b = 1_F$. If $a \cdot b = 1_F$ and $a \neq 0_F$, then $b = a^{-1}$.

Theorem 1.5. Let F be a field and let $a \in F$. Then we have $-(-a) = a$. Also, if $a \neq 0_F$, we have $(a^{-1})^{-1} = a$.

Proof. Since $-(-a) + (-a) = 0_F = a + (-a)$, we have $-(-a) = a$. If $a \neq 0_F$, then we have $(a^{-1})^{-1} \cdot a^{-1} = 1_F = a \cdot a^{-1}$, and thus $(a^{-1})^{-1} = a$. □

Theorem 1.6. The following statements are true for any field F with $a, b \in F$.

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

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

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

Proof.

(a) It suffices to prove the first equality. Since

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

it follows from cancellation law (Theorem 1.3) that $a \cdot 0_F = 0_F$.

(b) It suffices to prove the first equality. We have

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

where the last equality follows from (a). Thus, $(-a) \cdot b = -(a \cdot b)$ due to the uniqueness of additive inverses.

(c) We have $(-a) \cdot (-b) = -(a \cdot (-b)) = -(-(a \cdot b)) = a \cdot b$ by applying (b) twice. □

1.2 Vector Spaces

Definition 1.7. 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.

- Closeness: If $a \in F$ and $x, y \in V$, then $x + y \in V$ and $a \cdot x \in V$.
- Commutativity: $x + y = y + x$ holds for any $x, y \in V$.
- Associativity: $(x + y) + z = x + (y + z)$ and $(a \cdot b) \cdot x = a \cdot (b \cdot x)$ hold for any $a, b \in F$ and $x, y, z \in V$.
- Identity elements: There is an element 0_V in V, called the **additive identity** of V, such that $x + 0_V = x$ for any $x \in V$. Also, $1_F \cdot x = x$ for any $x \in V$.
- Inverse elements: For each $x \in V$ there is an element $-x$ in V, called the **additive inverse** of x, such that $x + (-x) = 0_V$.
- Distributivity: $a \cdot (x + y) = a \cdot x + a \cdot y$ and $(a + b) \cdot x = a \cdot x + b \cdot x$ hold for any $a, b \in F$ and $x, y \in V$.

The elements of F and the elements of V are usually called **scalars** and **vectors**, respectively. Subtraction of vectors is defined by $x - y = x + (-y)$ for any $x, y \in V$.

Remark. For simplification, we usually write ax instead of $a \cdot x$ in this note.

Examples. Let F be a field.

- F is a vector space over F.
- The set of ***n*-tuples** with entries from F, denoted F^n, is a vector space over F.
- The set of all $m \times n$ **matrices** with entries from F, denoted $F^{m \times n}$, is a vector space over F.
- The set of **polynomials** with coefficients from F, denoted $\mathcal{P}(F)$, is a vector space over F.

1.3 Subspaces

Definition 1.8. Let V be a vector space over F. A **subspace** of V is a subset W of V such that W is a vector space over F, where addition and scalar multiplication are the same as those defined on V.

Theorem 1.9. Let V be a vector space over F and let $W \subseteq V$. Then W is a subspace of V if and only if $0_V \in W$ and $ax + y \in W$ for any $a \in F$ and $x, y \in W$.

Proof. $(\Rightarrow)$ Straightforward. $(\Leftarrow)$ It suffices to prove the closeness of addition and scalar multiplication, and the existence of additive inverses. For any $a \in F$ and $x, y \in W$, we have

$$a \cdot x = a \cdot x + 0_W \in W \quad \text{and} \quad x + y = 1_F \cdot x + y \in W.$$

Furthermore, we have

$$x + (-1_F) \cdot x = 1_F \cdot x + (-1_F) \cdot x = (1_F + (-1_F)) \cdot x = 0_F \cdot x = 0_V,$$

for any $x \in W$, which completes the proof. □

Example. For any vector space V, V and $\{0_V\}$ are subspaces of V.

Example. The set $\mathcal{P}_n(F)$ of polynomials in $\mathcal{P}(F)$ with degree less than or equal to n is a subspace of $\mathcal{P}(F)$.

Definition 1.10. Let V be a vector space and let S_1 and S_2 be subsets of V. Then the **sum** of S_1 and S_2 is defined by

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

Theorem 1.11. Let V be a vector space and let W_1 and W_2 be subspaces of V. Then $W_1 + W_2$ is the minimal subspace of V that contains $W_1 \cup W_2$.

Proof. First we prove that $W_1 + W_2$ is a subspace of V. We have $0_V = 0_V + 0_V \in W_1 + W_2$, and for any $a \in F$, $x_1, y_1 \in W_1$ and $x_2, y_2 \in W_2$, we have

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

Thus, it follows from Theorem 1.9 that $W_1 + W_2$ is a subspace of V.

Now we prove the minimality. Suppose that W is a subspace of V that contains $W_1 \cup W_2$. For any $x_1 \in W_1$ and $x_2 \in W_2$, we have $x_1 + x_2 \in W$. Thus, $W_1 + W_2 \subseteq W$, completing the proof. □

1.4 Spanning Sets

Definition 1.12. Let V be a vector space over F and let $S \subseteq V$. A vector $x \in V$ is called a **linear combination** of S if $x = 0_V$ or there exist scalars $a_1, \ldots, a_n \in F$ and vectors $x_1, \ldots, x_n \in S$ such that

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

The set of all linear combinations of S is called the **span** of S, denoted by $\operatorname{span}(S)$. If $W = \operatorname{span}(S)$, then we say that W is **spanned** by S, or S is a **spanning set** of W.

Theorem 1.13. Let V be a vector space over F and let $S \subseteq V$. Then $\operatorname{span}(S)$ is the minimal subspace of V that contains S.

Proof. First we prove that $\operatorname{span}(S)$ is a subspace of V. Obviously $0_V \in \operatorname{span}(S)$. For any $a_1, a_2, \ldots, a_n, b_1, b_2, \ldots, b_m, c \in F$ and for any $x_1, x_2, \ldots, x_n, y_1, y_2, \ldots, y_m \in S$, we have

$$c \sum_{i=1}^{n} a_i x_i + \sum_{j=1}^{m} b_j y_j = \sum_{i=1}^{n} c a_i x_i + \sum_{j=1}^{m} b_j y_j \in \operatorname{span}(S).$$

Thus, $\operatorname{span}(S)$ is a subspace of V.

Now we prove the minimality. Let W be a subspace of V such that $S \subseteq W$. Then each element of $\operatorname{span}(S)$ belongs to W due to the closeness of W. Thus, $\operatorname{span}(S) \subseteq W$, which completes the proof. $\square$

1.5 Linearly Independent Sets

Definition 1.14. Let V be a vector space over a field F and let $S \subseteq V$. We say that S is **linearly dependent** if there exist nonzero scalars $a_1, \ldots, a_n \in F$ and distinct vectors $x_1, \ldots, x_n \in S$ such that

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

We say that S is **linearly independent** is it is not linearly dependent. The empty set $\varnothing$ is considered to be linearly independent.

Lemma 1.15. Let V be a vector space over F and let $S \subseteq V$. Then S is linearly dependent if and only if $x \in \operatorname{span}(S \setminus \{x\})$ for some $x \in S$.

Proof. ($\Rightarrow$) Suppose that there exist nonzero scalars $a_1, \ldots, a_n \in F$ and distinct vectors $x_1, \ldots, x_n \in S$ such that

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

Then we have

$$x_1 = (-a_1)^{-1} \sum_{i=2}^{n} a_i x_i = \sum_{i=2}^{n} (-a_1)^{-1} a_i x_i \in \operatorname{span}(S \setminus \{x_1\}).$$

($\Leftarrow$) Suppose that $x \in \operatorname{span}(S \setminus \{x\})$. Then there exist nonzero scalars $a_1, \ldots, a_n \in F$ and distinct vectors $x_1, \ldots, x_n \in S \setminus \{x\}$ such that

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

implying

$$(-1_F)x + \sum_{i=1}^{n} a_i x_i = 0_V. \qquad \square$$

Lemma 1.16. Let V be a vector space over F and let $S \subseteq V$. For any $x \in S$, we have $x \in \operatorname{span}(S \setminus \{x\})$ if and only if $\operatorname{span}(S) = \operatorname{span}(S \setminus \{x\})$.

Proof. ($\Leftarrow$) Straightforward since $x \in S \subseteq \operatorname{span}(S) = \operatorname{span}(S \setminus \{x\})$. ($\Rightarrow$) Note that we have

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

Thus, $S \subseteq \operatorname{span}(S \setminus \{x\})$, and it follows that $\operatorname{span}(S) \subseteq \operatorname{span}(S \setminus \{x\})$. Obviously we have $\operatorname{span}(S \setminus \{x\}) \subseteq \operatorname{span}(S)$, and we conclude that $\operatorname{span}(S) = \operatorname{span}(S \setminus \{x\})$. $\square$

Theorem 1.17. 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 $x \in \operatorname{span}(S \setminus \{x\})$.

(c) There exists $x \in S$ with $\operatorname{span}(S) = \operatorname{span}(S \setminus \{x\})$.

Proof. Immediately from Lemma 1.15 and Lemma 1.16. □

Theorem 1.18. Let V be a vector space. Let R and S be subsets of V with $R \subseteq S$. If S is linearly independent, then R is linearly independent.

Proof. Suppose that R is linearly dependent, i.e., there exists a vector $x \in R$ such that $x \in \operatorname{span}(R \setminus \{x\})$. It follows that $x \in \operatorname{span}(S \setminus \{x\})$, implying that S is linearly dependent, contradiction. Thus, R is linearly independent. □

1.6 Bases and Dimension

Definition 1.19. A **basis** of a vector space V is a linearly independent subset of V that spans V.

Example. $\varnothing$ is a basis of the zero vector space.

Example. Let

$$e_j = \Big([\![i = j]\!] \Big)_{1 \le i \le n}$$

for each $j \in \{1, \dots, n\}$. Then the set $\{e_1, e_2, \dots, e_n\}$ is a basis of F^n.

Example. $\{1, t, t^2, \dots\}$ is a basis of $\mathcal{P}(F)$.

Proposition 1.20. 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., $S = \varnothing$. Then the proposition holds since one can choose $Q = \varnothing$ as a basis for V.

Now assume the induction hypothesis that the proposition holds for $|S| = n$ with $n \ge 0$. If S is linearly independent, then we can choose $Q = S$ as a basis for V. Otherwise, there exists $x \in S$ with $\operatorname{span}(S \setminus \{x\}) = \operatorname{span}(S)$, 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 1.21 (Replacement Theorem). Let V be a vector space over a field F. Let S be a finite set that spans V, and let $Q \subseteq V$ be a finite linearly independent set. Then $|Q| \le |S|$, and there exists $R \subseteq S \setminus Q$ such that both $|Q \cup R| = |S|$ and $\operatorname{span}(Q \cup R) = V$ hold.

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

Now suppose that the theorem is true for $|Q| = m$ with $m \ge 0$, 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 $\operatorname{span}(Q' \cup R') = V$. 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 $x_{m+1} \in \operatorname{span}(Q') = \operatorname{span}(Q \setminus \{x_{m+1}\})$, implying that Q is linearly dependent, contradiction. Thus, there must exist some $j \in \{1, \dots, k\}$ such that $b_j \ne 0_F$.s Without loss of generality, suppose that $b_k \ne 0_F$ with $k \ge 1$. Also, let $R = \{y_1, \dots, y_{k-1}\}$. Then $|Q \cup R| = (m+1) + (k-1) = |S|$, and we have $|Q| \le |S|$. It follows that

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

where the second inclusion holds because

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

Then, we have
$$V = \operatorname{span}(Q' \cup R') \subseteq \operatorname{span}(Q \cup R) \subseteq V.$$
by Theorem 1.13. Thus, $\operatorname{span}(Q \cup R) = V$, which completes the proof. □

Corollary 1.22. 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 **??**, we can conclude that Q' is also linearly independent. Thus, we have $|Q'| \leq |S|$ by replacement theorem (Theorem 1.21), contradiction. □

Corollary 1.23. 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 $V = \operatorname{span}(S)$ and Q is linearly independent, it follows that Q is finite by Corollary 1.22, and thus we have $|Q| \leq |S|$. Also, since $V = \operatorname{span}(Q)$ and S is linearly independent, we have $|S| \leq |Q|$. Thus, $|Q| = |S|$. □

Definition 1.24. 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 1.22.

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 1.25. 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 1.21), there exists $R \subseteq S \setminus Q$ such that $|Q \cup R| = |S|$ and $\operatorname{span}(Q \cup R) = V$. Since $|Q| = |S|$, we have $|R| = 0$, i.e., $R = \varnothing$. Thus, $\operatorname{span}(Q) = V$.

($\Leftarrow$) Suppose that S spans V with $|S| = n$. By Proposition 1.20, 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 1.26. Let V be a finite-dimensional vector space. Let $S = \{x_1, \ldots, x_n\}$ be a basis for V. Then for each $x \in V$, there exist a unique n-tuple $(a_1, \ldots, a_n) \in F^n$ with

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

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

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

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

$$x = b_1 x_1 + \cdots + b_n x_n.$$

Then we have

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

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

Proposition 1.27. 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 $V = \operatorname{span}(S)$, we have $|S'| \leq |S|$ by replacement theorem (Theorem 1.21). Thus, $\dim(V') \leq \dim(V)$.

(b) Since S' is linearly independent and $|S'| = \dim(V)$, we have $\operatorname{span}(S') = V$ by Proposition 1.25. Thus, $V' = \operatorname{span}(S') = V$. $\square$

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 $T : V \to W$ 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 $T : V \to W$ 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 $a_1, \ldots, a_n \in F$ and $x_1, \ldots, x_n \in V$.

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 **??** (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

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

which completes the proof. □

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

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

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 $\mathcal{F}(V, W)$. □

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

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

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

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

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

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

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

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

2.2 Rank and Nullity

Definition 2.5. Let V and W be vector spaces. The **range** of a transformation $T : V \to W$, 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 $T : V \to W$ 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. □

Definition 2.7. Let V and W be vector spaces. The **null space** of a transformation $T : V \to W$, 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 $T : V \to W$ 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. □

Definition 2.9. Let X and Y be sets. Let $f : X \to Y$ 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 $T : V \to W$ 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 of $T(S)$ 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 $x \in \operatorname{span}(S \setminus \{x\})$ for some $x \in S$, then

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

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

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

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

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

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

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

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

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

$$T'(ax + y) = T(ax + y) = aT(x) + T(y) = aT'(x) + T'(y).$$ □

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

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

Proof. Let S be a basis for V and Q a basis for $\mathcal{N}(T)$. By replacement theorem (Theorem 1.21), 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,

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

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 $x - x' \in \mathcal{N}(T) = \operatorname{span}(Q)$. It follows that x is a linear combination of $Q \cup \{x'\}$. If $x \neq x'$, then

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

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,

$$\operatorname{nullity}(T) + \operatorname{rank}(T) = |Q| + |T(R)| = |Q| + |R| = \dim(V).$$ □

2.3 Isomorphisms

Definition 2.15. Let X, Y, Z be sets. Let $f : X \to Y$ and $g : Y \to Z$ be functions. Then the **composition** of f and g is the function $gf : X \to Z$ 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 $I_X : X \to X$ with $I_X(x) = x$ for all $x \in X$.

Definition 2.17. Let X and Y be sets. A function $f : X \to Y$ is said to be **invertible** if there exists a function $f^{-1} : Y \to X$, 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 $f : X \to Y$ and $g : Y \to X$ 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 $g : Y \to X$ 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 $T : V \to W$ 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 1.27.

($\Leftarrow$) 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 $S = \{x_1, x_2, \ldots, x_n\}$ be a basis for V and let $y_1, y_2, \ldots, y_n$ be vectors in W. Then there exists a unique $T \in \mathcal{L}(V, W)$ with $T(x_i) = y_i$ for each $i \in \{1, \ldots, n\}$.

Proof. Let T be the transformation that satisfies

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

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

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

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

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

for any $a_1, \ldots, a_n \in F$. Thus, $T' = T$. □

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 $\mathrm{rank}(T) = \dim(W)$ since $\mathcal{R}(T) = W$. Furthermore, since T is injective, we have $\mathrm{nullity}(T) = 0$, and it follows that $\mathrm{rank}(T) = \dim(V)$ by rank-nullity theorem (Theorem 2.14). Thus, $\dim(V) = \mathrm{rank}(T) = \dim(W)$.

($\Leftarrow$) Suppose that $S = \{x_1, x_2, \ldots, x_n\}$ is a basis for V and $R = \{y_1, y_2, \ldots, y_n\}$ 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 $i \in \{1, \ldots, n\}$. Since R is a basis for W, for each $y \in W$ there exist scalars $a_1, \ldots, a_n \in F$ such that

$$y = \sum_{i=1}^{n} a_iy_i = \sum_{i=1}^{n} a_iT(x_i) = T\left(\sum_{i=1}^{n} a_ix_i\right).$$

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 $V \cong W$. □

2.4 Coordinates and Matrix Representations

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

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

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

Examples.

- The **standard ordered basis** for F^n is $(e_1, \ldots, e_n)$, 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 $(t^0, t^1, \ldots, t^n)$.

Definition 2.24. Let V be a finite-dimensional vector space over a field F. Let $\beta = (x_1, \ldots, x_n)$ be an ordered basis for V. Then we define $\phi_\beta : V \to F^n$ 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 $a_1, a_2, \ldots, a_n \in F$. For each vector x in V, $\phi_\beta(x)$ is called the **coordinate** of x with respect to β, denoted by $[x]_\beta$.

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

Proof. ϕ_β 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 $a_1, \ldots, a_n, b_1, \ldots, b_n, c \in F$. Also, ϕ_β is invertible since there exists $\phi_\beta^{-1} : F^n \to V$ with

$$\phi_\beta^{-1}\left(\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}\right) = a_1x_1 + a_2x_2 + \cdots + a_nx_n$$

for any $a_1, a_2, \ldots, a_n \in F$. Thus, ϕ_β 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 $\Phi_\beta^\gamma : \mathcal{L}(V, W) \to F^{m\times n}$ 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 $T : V \to W$, the matrix $\Phi_\beta^\gamma(T)$ is called the **matrix representation** of T with respect to β and γ, 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 β and γ be ordered bases for a vector spaces V and W over F, respectively. Then Φ_β^γ is an isomorphism from $\mathcal{L}(V, W)$ onto $F^{m\times n}$.

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

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

for any $i \in \{1, \ldots, m\}$ and $j \in \{1, \ldots, n\}$. To prove that Φ_β^γ 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 $T : V \to W$ such that

$$T(x_j) = \sum_{i=1}^{n} a_{ij} y_j$$

for each $j \in \{1, \ldots, n\}$, and it follows that $[T]_\beta^\gamma = A$. Thus, there exists $(\Phi_\beta^\gamma)^{-1} : F^{m \times n} \to \mathcal{L}(V, W)$ 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. □

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. □

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 $T_1 : U \to V$ and $T_2 : V \to W$ 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. □

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 $T_1 : U \to V$ and $T_2 : V \to W$ 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 $i \in \{1, \ldots, \ell\}$ and $k \in \{1, \ldots, n\}$. Thus, $C = AB$. $\square$

Corollary 2.33. Let V and W be finite-dimensional vector spaces with ordered bases β and γ over a field F, respectively. If $T : V \to W$ 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 $\varphi : F \to V$ 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

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

$\square$

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 α, β and γ be standard ordered bases for F^n, F^m and F^ℓ, 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) $L_{AB} = L_A L_B$ 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 $x, y \in F^n$,

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

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

(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 α and β 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

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

(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) $c(T_2 T_1) = (cT_2)T_1 = T_2(cT_1)$.

(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 $T_3(T_2T_1) = (T_3T_2)T_1$. □

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. □

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

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

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, $L_{AB} = L_A L_B$ is injective and surjective. Thus, $L_A : F^n \to F^n$ is injective and $L_B : F^n \to F^n$ 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 $B = A^{-1}$. $\square$

2.8 Direct Sums and Projections

Definition 2.40. Let V and W be subspaces of a vector space U. We say that U is the **direct sum** of V and W, denoted

$$U = V \oplus W,$$

if $V \cap W = \{0_U\}$ and $U = V + W$.

Theorem 2.41. Let U be a finite-dimensional vector space over F and let V and W be subspaces of U. Then the following statements are equivalent.

(a) $U = V \oplus W$.

(b) For any vector $x \in U$, there is a unique vector $y \in V$ and a unique vector $z \in W$ such that $x = y + z$.

(c) If R and S are bases of V and W, respectively, then $R \cup S$ is a basis of U with $R \cap S = \varnothing$.

Proof. First we asssume (a) and prove (b). Since $U = V + W$, for each $x \in U$ there are vectors $y \in V$ and $z \in W$ with $x = y + z$. For the uniqueness, let $y, y' \in V$ and $z, z' \in W$ be vectors with

$$x = y + z = y' + z'.$$

Note that $y - y' = z - z'$ is a vector in $V \cap W = \{0_V\}$. Thus, $y = y'$ and $z = z'$.

Now we assume (b) and prove (c). Let $R = \{x_1, \dots, x_m\}$ and $S = \{x_{m+1}, \dots, x_n\}$. Note that $R \cup S$ spans U since

$$\operatorname{span}(R \cup S) = \operatorname{span}(R) + \operatorname{span}(S) = V + W = U.$$

For the linear independence of $R \cup S$, suppose that $a_1, \dots, a_n \in F$ are scalars such that

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

Since $0_U = 0_V + 0_W$ holds and we have

$$\sum_{i=1}^{m} a_i x_i \in V \quad \text{and} \quad \sum_{i=m+1}^{n} a_i x_i \in W,$$

it follows that

$$\sum_{i=1}^{m} a_i x_i = 0_V \quad \text{and} \quad \sum_{i=m+1}^{n} a_i x_i = 0_W,$$

by (b), implying $a_i = 0_F$ for any $i \in \{1, \dots, n\}$ by the linear independence of R and S. Thus, $R \cup S$ are linearly independent. Since $R \cap S \subseteq V \cap W = \{0_V\}$, we have $R \cap S = \varnothing$.

Finally we assume (c) and prove (a). Let $R = \{x_1, \dots, x_m\}$ and $S = \{x_{m+1}, \dots, x_n\}$ are bases of V and W, respectively. Then $R \cup S$ is a basis of U, and thus

$$U = \operatorname{span}(R \cup S) = \operatorname{span}(R) + \operatorname{span}(S) = V + W.$$

If $x \in V \cap W$, then there exist scalars $a_1, \ldots, a_m, a'_{m+1}, \ldots, a'_n \in F$ such that

$$\sum_{i=1}^{m} a_i x_i = x = \sum_{i=m+1}^{n} a'_i x_i.$$

Let $a_i = -a'_i$ for all $i \in \{m+1, \ldots, n\}$. Then we have

$$\sum_{i=1}^{n} a_i x_i = \sum_{i=1}^{m} a_i x_i + \sum_{i=m+1}^{n} (-a'_i) x_i = x + (-x) = 0_U.$$

Since $R \cup S$ is linearly independent by (c) , it follows that $a_i = 0_F$ for all $i \in \{1, \ldots, n\}$, implying $x = 0_U$. Thus, $V \cap W = \{0_U\}$, which completes the proof. □

Definition 2.42. Let V and W be subspaces of a vector space U with $U = V \oplus W$. Then the **projection** on V along W is a transformation $T : U \to U$ such that

$$T(x) = y$$

holds for any $x \in U$ with

$$x = y + z,$$

where $y \in V$ and $z \in W$.

Theorem 2.43. Let V and W be subspaces of a vector space U with $U = V \oplus W$. Let $T : U \to U$ be the projection on V along W. Then T is linear.

Proof. Let $a \in F$ and $x, x' \in U$. Furthermore, let

$$y = T(x), \quad z = x - T(x)$$

and

$$y' = T(x'), \quad z' = x' - T(x').$$

Then we have

$$\begin{aligned} T(ax + x') &= T(a(y + z) + (y' + z')) \\ &= T((ay + y') + (az + z')) \\ &= ay + y' \\ &= aT(x) + T(x'). \end{aligned}$$

□

Theorem 2.44. Let V and W be subspaces of a vector space U with $U = V \oplus W$. Let $T : U \to U$ be linear. Then T is the projectioon on V along W if and only if $T(y) = y$ for any $y \in V$ and $T(z) = 0_U$ for any $z \in W$.

Proof. ($\Rightarrow$) Straightforward. ($\Leftarrow$) For any $x \in U$, let $y \in V$ and $w \in W$ be vectors with $x = y + z$. Then

$$T(x) = T(y + z) = T(y) + T(z) = y + 0_U = y.$$

Thus, T is the projection on V along W. □

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

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

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 $1 \le i \le m$ and $1 \le j \le n$.

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 $\operatorname{rank}(A)$ and $\operatorname{nullity}(A)$, respectively, are defined by

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

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

(a) $\mathcal{R}(L_A) = \operatorname{span}(\operatorname{col}(A))$.

(b) $\operatorname{rank}(A) = \dim(\operatorname{span}(\operatorname{col}(A)))$.

Proof.

(a) Let $\beta = (x_1, \ldots, x_n)$ and $\gamma = (y_1, \ldots, y_m)$ 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 ith column of $[L_A]_\beta^\gamma = A$. Thus, we have $L_A(\beta) = \operatorname{col}(A)$, and it follows that

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

(b) By (a), we have

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

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

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

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

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

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

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 $T : V \to W$ 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 $T(U) = T(\operatorname{span}(S)) = \operatorname{span}(T(S))$.

(a) Let Q be a basis for $T(U)$. By replacement theorem (Theorem 1.21),

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

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

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

□

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 $\operatorname{rank}(XA) = \operatorname{rank}(A)$.

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

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,

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

(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,

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

□

Theorem 3.10. Let V and W be finite-dimensional vector spaces with bases β and γ, respectively. If $T : V \to W$ is linear, then

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

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 ϕ_β and ϕ_γ are invertible, we have

$$\operatorname{rank}(T) = \operatorname{rank}(\phi_\gamma T) = \operatorname{rank}(L_A\phi_\beta) = \operatorname{rank}(L_A) = \operatorname{rank}(A).$$

□

Theorem 3.11. Let $A \in F^{m\times n}$ and let r be a nonnegative integer. Then $\operatorname{rank}(A) = r$ 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 $X_1, \ldots, X_p \in F^{m\times m}$ and $Y_1, \ldots, Y_q \in F^{n\times n}$ such that

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

Since elementary matrices are invertible,

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

($\Rightarrow$) 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 $2 \le i \le m$ and $2 \le j \le n$.

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 $2 \le i \le m$, if the $(i, 1)$-entry is nonzero, perform a type 3 row operation by adding a multiple of the first row to the ith row such that the $(i, 1)$-entry becomes zero.
    b. For $2 \le j \le n$, if the $(1, j)$-entry is nonzero, perform a type 3 column operation by adding a multiple of the first column to the jth column such that the $(1, j)$-entry becomes zero.

By Theorem 3.9, $\operatorname{rank}(B) = \operatorname{rank}(A) = r$ 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 $\operatorname{rank}(A) = 1$ 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 $k \ge 1$, and we prove that the theorem also holds for $\min(m, n) = k + 1$. Since $\min(m, n) \ge 2$, we have

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

where B' is an $(m-1) \times (n-1)$ matrix. Note that $\operatorname{rank}(B') = \operatorname{rank}(B) - 1 = r - 1$. 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

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

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. □

Corollary 3.12. Let $A \in F^{m\times n}$ and let r be a nonnegative integer. Then $\operatorname{rank}(A) = r$ 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

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

($\Rightarrow$) By Theorem 3.11, there exist elementary matrices $X_1, \ldots, X_p \in F^{m\times m}$ and $Y_1, \ldots, Y_q \in F^{n\times n}$ 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 $Y = Y_1 \cdots Y_q$. □

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

Proof. Let $r = \operatorname{rank}(A)$. 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,

$$\operatorname{rank}(A^t) = \operatorname{rank}(Y^t A^t X^t) = \operatorname{rank}(D^t) = r.$$ □

Theorem 3.14.

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

$$\operatorname{rank}(T_2T_1) \le \operatorname{rank}(T_1) \quad \text{and} \quad \operatorname{rank}(T_2T_1) \le \operatorname{rank}(T_2).$$

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

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

Proof.

(a) By Lemma 3.8, we have

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

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

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

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

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

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 $\operatorname{rank}(A) = n$, there exist elementary matrices $X_1, \ldots, X_p, Y_1, \ldots, Y_q \in F^{n\times n}$ 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 $x_1, x_2, \ldots, x_n$ are n variables that take values in F, is called a system of m **linear equations** in n unknowns over the field F. Furthremore, 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 $\operatorname{rank}(A) = \operatorname{rank}(A \mid b)$.

Proof. It is proved by

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

□

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. $\square$

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

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

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. $\square$

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

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

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 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 ith row is the first row that has not been processed, and find j such that the jth 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 jth column.
5. Mark the ith row and the jth 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 Permutations and Determinants

Definition 4.1. A **permutation** over a set R is a bijection from R to R. Let S_n denote the set of all permutation over $\{1, 2, \ldots, n\}$.

Remark. There are $n!$ permutations in S_n.

Example. Let $\sigma : \{1, 2, 3\} \to \{1, 2, 3\}$ be the function with

$$\sigma(1) = 2, \quad \sigma(2) = 3, \quad \sigma(3) = 1.$$

Then σ is a permutation and belongs to S_3.

Definition 4.2. For any permutation $\sigma \in S_n$, we say that a pair (i, j) of integers in $\{1, 2, \ldots, n\}$ is an **inversion** of σ if both $i < j$ and $\sigma(i) > \sigma(j)$ hold. Furthremore, we define the **sign** of σ by

$$\operatorname{sgn}(\sigma) = (-1)^{\rho(\sigma)},$$

where $\rho(\sigma)$ denotes the number of inversions of σ.

Example. If $\sigma : \{1, 2, 3\} \to \{1, 2, 3\}$ is a permutation with

$$\sigma(1) = 2, \quad \sigma(2) = 3, \quad \sigma(3) = 1,$$

then we have $\rho(\sigma) = 2$ since $(1, 3)$ and $(2, 3)$ are the only inversions of σ. It follows that $\operatorname{sgn}(\sigma) = (-1)^2 = 1$.

Definition 4.3. For any matrix $A \in F^{n\times n}$, we define the **determinant** of A by

$$\det(A) = \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) \prod_{(i,j)\in\sigma} A_{ij}.$$

Example. For each $A \in F^{3\times 3}$, we have

$$\begin{aligned}\det(A) &= \sum_{\sigma \in S_3} \operatorname{sgn}(\sigma) \prod_{(i,j)\in\sigma} A_{ij} \\ &= A_{11}A_{22}A_{33} - A_{11}A_{23}A_{32} + A_{12}A_{23}A_{31} - A_{12}A_{21}A_{33} + A_{13}A_{21}A_{32} - A_{13}A_{22}A_{31}.\end{aligned}$$

4.2 Characterization of the Determinant

Definition 4.4. A function $\delta : F^{n\times n} \to F$ 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 jth rows of A, B and C are identical for each $j \in \{1, \dots, n\} \setminus \{i\}$.
- The ith row of A is the sum of the ith row of B multiplied by k and the ith row of C.

Definition 4.5. An n-linear function $\delta : F^{n\times n} \to F$ is **alternating** if

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

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

Proposition 4.6. Let $\delta : F^{n\times n} \to F$ 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 $\delta(E_1A) = -\delta(A)$.

(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 $\delta(E_2A) = k\delta(A)$.

(c) If $E_3 \in F^{n\times n}$ is an elementary matrix of type 3, then $\delta(E_3A) = \delta(A)$.

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

(a) Let E_1 be obtained from I_n by interchanging the pth row and the qth row with $p < q$. Then we have

$$\begin{aligned}
0_F = \delta\begin{pmatrix} x_1 \\ \vdots \\ x_p + x_q \\ \vdots \\ x_p + x_q \\ \vdots \\ x_n \end{pmatrix} &= \delta\begin{pmatrix} x_1 \\ \vdots \\ x_p \\ \vdots \\ x_p + x_q \\ \vdots \\ x_n \end{pmatrix} + \delta\begin{pmatrix} x_1 \\ \vdots \\ x_q \\ \vdots \\ x_p + x_q \\ \vdots \\ x_n \end{pmatrix} \\
&= \delta\begin{pmatrix} x_1 \\ \vdots \\ x_p \\ \vdots \\ x_p \\ \vdots \\ x_n \end{pmatrix} + \delta\begin{pmatrix} x_1 \\ \vdots \\ x_p \\ \vdots \\ x_q \\ \vdots \\ x_n \end{pmatrix} + \delta\begin{pmatrix} x_1 \\ \vdots \\ x_q \\ \vdots \\ x_p \\ \vdots \\ x_n \end{pmatrix} + \delta\begin{pmatrix} x_1 \\ \vdots \\ x_q \\ \vdots \\ x_q \\ \vdots \\ x_n \end{pmatrix} \\
&= 0_F + \delta(A) + \delta(E_1A) + 0_F.
\end{aligned}$$

Thus, $\delta(E_1A) = -\delta(A)$.

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

$$\delta(E_2A) = \delta\begin{pmatrix} x_1 \\ \vdots \\ kx_p \\ \vdots \\ x_n \end{pmatrix} = k\delta\begin{pmatrix} x_1 \\ \vdots \\ x_p \\ \vdots \\ x_n \end{pmatrix} = k\delta(A).$$

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

$$\delta(E_3A) = \delta\begin{pmatrix} x_1 \\ \vdots \\ x_p \\ \vdots \\ kx_p + x_q \\ \vdots \\ x_n \end{pmatrix} = k\delta\begin{pmatrix} x_1 \\ \vdots \\ x_p \\ \vdots \\ x_p \\ \vdots \\ x_n \end{pmatrix} + \delta\begin{pmatrix} x_1 \\ \vdots \\ x_p \\ \vdots \\ x_q \\ \vdots \\ x_n \end{pmatrix} = k0_F + \delta(A) = \delta(A).$$

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

Theorem 4.7. Let $\delta : F^{n\times n} \to F$ be an alternating n-linear function and let $A \in F^{n\times n}$. If $\operatorname{rank}(A) < n$, then $\delta(A) = 0_F$.

Proof. Since

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

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

$$\delta(A) = \delta(E_p \cdots E_1A) = \delta(B) = 0_F,$$

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

Theorem 4.8. Let $\delta : F^{n\times n} \to F$ 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 $\operatorname{rank}(A) < n$. Then we have $\operatorname{rank}(AB) < n$. Thus,

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

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

$$\begin{aligned}
\delta(AB) &= \delta(E_k \cdots E_1B) \\
&= \delta(E_k) \cdots \delta(E_1)\delta(B) \\
&= \delta(E_k) \cdots \delta(E_1)\delta(I_n)\delta(B) && (\delta(I_n) = 1_F) \\
&= \delta(E_k \cdots E_1I_n)\delta(B) \\
&= \delta(A)\delta(B).
\end{aligned}$$

□

Theorem 4.9. There exists a unique alternating n-linear function $\delta : F^{n\times n} \to F$ with $\delta(I_n) = 1_F$.

Proof. Suppose that $\delta, \delta' : F^{n\times n} \to F$ 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 $\operatorname{rank}(A) < n$, then

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

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

$$A = E_p \cdots E_1$$

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

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

□

Definition 4.10. The **determinant** of $A \in F^{n\times n}$ is

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

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

Theorem 4.11. For any matrix $A \in F^{n\times n}$,

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

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

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

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

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

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

Then we have

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

Now we show that δ is alternating. Suppose that $D \in F^{n \times n}$ is a matrix whose pth row and qth 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 $i \in \{1, \ldots, n\} \setminus \{p, q\}$.

Then we have

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

Finally, we have

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

where σ_0 is the identity permutation. Therefore, δ is an alternating n-linear function with $\delta(I_n) = 1_F$, and by Theorem 4.9 we can conclude that it is exactly the determinant function. □

4.3 Properties of Determinants

Theorem 4.12. For any $A, B \in F^{n\times n}$, we have $\det(AB) = \det(A)\det(B)$.

Proof. Striaghtforward from Theorem 4.8. $\square$

Theorem 4.13. If $A \in F^{n\times n}$ is invertible, then $\det(A^{-1}) = (\det(A))^{-1}$.

Proof. It follows by

$$\det(A^{-1})\det(A) = \det(A^{-1}A) = \det(I_n) = 1_F. \qquad \square$$

Definition 4.14. Let $A, B \in F^{n\times n}$. We say that A and B are **similar**, denoted

$$A \sim B,$$

if there is an invertible matrix $Q \in F^{n\times n}$ such that $B = QAQ^{-1}$.

Theorem 4.15. For any $A, B \in F^{n\times n}$, if $A \sim B$, then $\det(A) = \det(B)$.

Proof. Suppose that Q is invertible such that $B = QAQ^{-1}$. Then

$$\begin{aligned}\det(B) &= \det(QAQ^{-1}) \\ &= \det(Q)\cdot\det(A)\cdot\det(Q^{-1}) \\ &= \det(Q)\cdot\det(Q^{-1})\cdot\det(A) \\ &= \det(I_n)\cdot\det(A) \\ &= \det(A).\end{aligned} \qquad \square$$

Definition 4.16. Let $n \geq 2$. For any $A \in F^{n\times n}$ and for any $i, j \in \{1, \dots, n\}$, let $\tilde{A}_{ij}$ denote the $(n-1)\times(n-1)$ matrix obtained from A by deleting the ith row and the jth column.

Theorem 4.17 (Laplace Expansion). Let $n \geq 2$. For any $A \in F^{n\times n}$ and $i \in \{1, \dots, n\}$, we have

$$\det(A) = \sum_{j=1}^{n} A_{ij}(-1)^{i+j}\det(\tilde{A}_{ij}).$$

Proof. For $j \in \{1, \dots, n\}$, let $B^{(j)}$ be the matrix obtained from A by replacing its ith row with e_j. Note that we can turn $B^{(j)}$ into a matrix

$$C^{(j)} = \begin{pmatrix} 1 & O \\ X & \tilde{A}_{ij} \end{pmatrix}$$

by $i-1$ row swaps and $j-1$ column swaps, where X is an $(n-1)\times 1$ matrix, and O is the $1\times(n-1)$ zero matrix. Thus, we have

$$\det(B^{(j)}) = (-1)^{(i-1)+(j-1)}\det(C^{j}) = (-1)^{i+j}\det(\tilde{A}_{ij}).$$

Since $\det(\cdot)$ is n-linear, we have

$$\det(A) = \sum_{j=1}^{n} A_{ij}\det(B^{(j)}) = \sum_{j=1}^{n} A_{ij}(-1)^{i+j}\det(\tilde{A}_{ij}). \qquad \square$$

Chapter 5

Diagonalization

5.1 Eigenvalues and Eigenvectors

Definition 5.1. Let $T : V \to V$ 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 (λ, x) is called an **eigenpair** of T, with λ and x called an **eigenvalue** and an **eigenvector** of T, respectively.

Definition 5.2. Let V be a finite-dimensional vector space over a field F. 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.

Theorem 5.3. Let V be a vector space over a field F and let $T : V \to V$ be linear. Let $\beta = (x_1, x_2, \ldots, x_n)$ be an ordered basis for T. Let $\lambda_1, \lambda_2, \ldots, \lambda_n \in F$ be scalars. Then

$$[T]_\beta^\beta = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}$$

if and only if $T(x_i) = \lambda_i x_i$ for each $i \in \{1, 2, \ldots, n\}$.

Proof. ($\Rightarrow$) For each $i \in \{1, 2, \ldots, n\}$, we have

$$[T(x_i)]_\beta = \lambda_i e_i = [\lambda_i x_i]_\beta.$$

Thus, $T(x_i) = \lambda_i x_i$. ($\Leftarrow$) For each $i \in \{1, 2, \ldots, n\}$, we have $[T(x_i)]_\beta = [\lambda_i x_i]_\beta = \lambda_i e_i$, and it follows that

$$[T]_\beta^\beta = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}. \qquad \square$$

Corollary 5.4. Let V be a finite-dimensional vector space and let $T : V \to V$ be linear. Let β be an ordered basis of T. Then $[T]_\beta^\beta$ is diagonal if and only if β is an eigenbasis of V for T.

Proof. Straightforward from Theorem 5.3. $\square$

Definition 5.5. Let $A \in F^{n \times n}$. If

$$Ax = \lambda x$$

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

Theorem 5.6. Let V be a finite-dimensional vector space with an ordered basis β. Let $\lambda \in F$ be a scalar and $x \in V$ be a vector. Then (λ, x) is an eigenpair of T if and only if $(\lambda, [x]_\beta)$ is an eigenpair of $[T]_\beta^\beta$.

Proof. ($\Rightarrow$) Suppose that $T(x) = \lambda x$. Then we have

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

($\Leftarrow$) Since

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

we can conclude that $T(x) = \lambda x$. $\square$

5.2 Characteristic Polynomials and Eigenspaces

Theorem 5.7. Let $A \in F^{n\times n}$ be a matrix and let $\lambda \in F$ be a scalar. Then λ is an eigenvalue of A if and only if $\det(A - \lambda I_n) = 0_F$.

Proof. The proof is as follows.

$$\begin{array}{rcl} \lambda \text{ is an eigenvalue of } A & \Leftrightarrow & Ax = \lambda x \text{ for some } x \in F^n \setminus \{0_{F^n}\} \\ & \Leftrightarrow & (A - \lambda I_n)x \text{ for some } x \in F^n \setminus \{0_{F^n}\} \\ & \Leftrightarrow & (A - \lambda I_n) \text{ is not invertible} \\ & \Leftrightarrow & \det(A - \lambda I_n) = 0_F. \end{array}$$

□

Theorem 5.8. Let $A, B \in F^{n\times n}$ and $\lambda \in F$. If $A \sim B$, then

$$\det(A - \lambda I_n) = \det(B - \lambda I_n).$$

Proof. Suppose that $Q \in F^{n\times n}$ is invertible such that $A = Q^{-1}BQ$. Then we have

$$\begin{aligned} \det(A - \lambda I_n) &= \det(Q^{-1}BQ - \lambda Q^{-1} I_n Q) \\ &= \det(Q^{-1}(B - \lambda I_n)Q) \\ &= \det(Q^{-1}) \det(B - \lambda I_n) \det(Q) \\ &= \det(B - \lambda I_n). \end{aligned}$$

□

Definition 5.9. Let V be a finite-dimensional vector space with $\dim(V) = n$.

- For any linear operator $T : V \to V$, the **characteristic polynomial** of T is

  $$f_T(t) = \det([T]_\beta^\beta - tI_n),$$

  where β is an arbitrary basis of V.

- For any $A \in F^{n\times n}$, the **characteristic polynomial** of A is

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

Remark. The characteristic polynomial of a linear operator $T : V \to V$ is well-defined, since $[T]_\beta^\beta \sim [T]_\gamma^\gamma$ holds for any bases β and γ of V.

Theorem 5.10. Let V be a vector space over a field F and let $T : V \to V$ be linear. For any scalar $\lambda \in F$ and for any nonzero vector $x \in V$, (λ, x) is an eigenpair of T if and only if $x \in N(T - \lambda I_V)$.

Proof. The proof is as follows.

$$\begin{array}{rcl} (\lambda, x) \text{ is an eigenpair of } T & \Leftrightarrow & T(x) = \lambda x \\ & \Leftrightarrow & T(x) = (\lambda I_V)(x) \\ & \Leftrightarrow & (T - \lambda I_V)(x) = 0_V \\ & \Leftrightarrow & x \in N(T - \lambda I_V). \end{array}$$

□

Definition 5.11. Let V be a vector space over F and let $T : V \to V$ be linear. For each scalar $\lambda \in F$, we define

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

If λ is an eigenvalue of T, then $E_T(\lambda)$ is called the **eigenspace** of T with respect to λ.

Theorem 5.12. Let V be a vector space over F and let $T : V \to V$ be linear. If $(\lambda_1, x_1), \ldots, (\lambda_k, x_k)$ are eigenpairs of T such that $\lambda_1, \ldots, \lambda_k$ are distinct, then $\{x_1, \ldots, x_k\}$ is linearly independent.

Proof. The proof is by induction on k. For $k = 1$, the theorem trivially holds. For the inductive step, let $k \geq 2$. Suppose that there are scalars $a_1, \ldots, a_k \in F$ such that

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

Applying $T - \lambda_k I_V$ to both sides, we have

$$0_V = \sum_{i=1}^{k} (T - \lambda_k I_V)(a_i x_i) = \sum_{i=1}^{k} a_i(\lambda_i - \lambda_k)x_i = \sum_{i=1}^{k-1} a_i(\lambda_i - \lambda_k)x_i.$$

Thus, we have $a_i = 0_F$ for each $i \in \{1, \ldots, k-1\}$ since $\{x_1, \ldots, x_{k-1}\}$ is linearly independent by induction hypothesis. It follows that $a_k = 0_F$ since

$$a_k x_k = 0_V - \sum_{i=1}^{k-1} a_i x_i = 0_V.$$

Thus, $\{x_1, \ldots, x_k\}$ is linearly independent, completing the proof. $\square$

5.3 Diagonalizability

Definition 5.13. Let V be a finite-dimensional vector space over F and let $T : V \to V$ be linear. For any scalar $\lambda \in F$, the **multiplicity** of λ with respect to T is the largest nonnegative integer m such that

$$(t-\lambda)^m \mid f_T(t).$$

Theorem 5.14. Let V be a finite-dimensional vector space over F and let $T : V \to V$ be linear. For any $\lambda \in F$, if m is the multiplicity of λ with respect to T and d is the dimension of $E_T(\lambda)$, then

$$d \leq m.$$

Proof. Let $\{x_1, \ldots, x_d\}$ be a basis of $E_T(\lambda)$. By replacement theorem, there exists an ordered basis $\beta = \{x_1, \ldots, x_n\}$ of V. Note that we have

$$[T]_\beta^\beta = \begin{pmatrix} \lambda I_d & X \\ O & Y \end{pmatrix},$$

where O is an $(n-d) \times d$ zero matrix. It follows that

$$f_T(t) = \det \begin{pmatrix} (\lambda - t)I_d & X \\ O & Y - tI_{n-d} \end{pmatrix} = (\lambda - t)^d \det(Y - tI_{n-d}),$$

implying

$$(t-\lambda)^d \mid f_T(t).$$

Thus, $d \leq m$. □

Theorem 5.15. Let V be a finite-dimensional vector space with $\dim(V) = n$ and let $T : V \to V$ be linear. Let $\lambda_1, \ldots, \lambda_k$ be the distinct eigenvalues of T, and let $d_i = \dim(E_T(\lambda_i))$ for $i \in \{1, \ldots, k\}$. Then V has an eigenbasis of T if and only if

$$\sum_{i=1}^{k} d_i = n.$$

Proof. ($\Leftarrow$) For each $i \in \{1, \ldots, k\}$ let

$$S_i = \{x_{ij} : 1 \leq j \leq d_i\}$$

be a basis of $E_T(\lambda_i)$. Suppose that there are scalars $a_{ij} \in F$ for each $i \in \{1, \ldots, k\}$ and for each $j \in \{1, \ldots, d_i\}$ such that

$$\sum_{i=1}^{k} \sum_{j=1}^{d_i} a_{ij} x_{ij} = 0_V,$$

and we define

$$y_i = \sum_{j=1}^{d_i} a_{ij} x_{ij}$$

for each $i \in \{1, \ldots, k\}$. We claim that $y_i = 0_V$ for each $i \in \{1, \ldots, k\}$, which is proved as follows.

Let π be a permutation over $\{1, \ldots, k\}$ such that $y_{\pi(1)}, \ldots, y_{\pi(\ell)}$ are nonzero and $y_{\pi(\ell+1)}, \ldots, y_{\pi(k)}$ are zero, where $0 \leq \ell \leq k$. Assume for contradiction that $\ell \neq 0$. It is obvious that $\{y_{\pi(1)}, y_{\pi(2)}, \ldots, y_{\pi(\ell)}\}$ is linearly dependent. However,

$$(\lambda_{\pi(1)}, y_{\pi(1)}), (\lambda_{\pi(2)}, y_{\pi(2)}), \ldots, (\lambda_{\pi(\ell)}, y_{\pi(\ell)})$$

are eigenpairs of T, implying that $\{y_{\pi(1)}, y_{\pi(2)}, \ldots, y_{\pi(\ell)}\}$ is linearly independent, contradiction.

It follows that for each $i \in \{1, \ldots, k\}$ we have $y_i = 0_V$, and thus $a_{ij} = 0_F$ for each $j \in \{1, \ldots, d_i\}$ since S_i is linearly independent. Therefore,

$$S = \bigcup_{i=1}^{k} S_i$$

is linearly independent, and thus is a basis of V.

($\Rightarrow$) Let S be an eigenbasis of V, and let $S_i = S \cap E_T(\lambda_i)$ for each $i \in \{1, \ldots, k\}$. Let m_i is the multiplicity of λ_i. Then we have

$$n = \sum_{i=1}^{k} |S_i| \leq \sum_{i=1}^{k} d_i \leq \sum_{i=1}^{k} m_i \leq n,$$

implying

$$\sum_{i=1}^{k} d_i = n. \qquad \square$$

Theorem 5.16. Let V be a finite-dimensional vector space over F with $\dim(V) = n$ and let $T : V \to V$ be linear. If $\lambda_1, \lambda_2, \ldots, \lambda_k$ are the eigenvalues of T, then

$$V = E_T(\lambda_1) \oplus E_T(\lambda_2) \oplus \cdots \oplus E_T(\lambda_k)$$

if and only if V has an eigenbasis for T.

Proof. ($\Rightarrow$) By Theorem 2.41, there is an ordered basis β_i of $E_T(\lambda_i)$ for each $i \in \{1, \ldots, k\}$ such that $\beta = \beta_1 \cup \beta_2 \cup \cdots \cup \beta_k$ is an ordered basis of V.

($\Leftarrow$) By Theorem 2.41, it suffices to show that there is an ordered basis β_i of $E_T(\lambda_i)$ for each $i \in \{1, \ldots, k\}$ such that $\beta_1 \cup \cdots \cup \beta_k$ is an ordered basis of V. Let β be an eigenbasis of V for T. For each $i \in \{1, \ldots, k\}$, let $\beta_i = \beta \cap E_T(\lambda_i)$ and $d_i = \dim(E_T(\lambda_i))$. Note that $|\beta_i| \leq d_i$ holds by the linear independence of β_i, and we have

$$\sum_{i=1}^{k} d_i = n = \sum_{i=1}^{k} |\beta_i|.$$

It follows that for each $i \in \{1, \ldots, k\}$, we have $|\beta_i| = d_i$, and thus β_i is an ordered basis of $E_T(\lambda_i)$ for each $i \in \{1, \ldots, k\}$. $\square$

5.4 Cayley-Hamilton Theorem

Definition 5.17. Let V be a vector space and let $T \in \mathcal{L}(V)$. A subspace W of V is a ***T*-invariant subspace** of V if

$$T(W) \subseteq W.$$

Theorem 5.18. Let V be a finite-dimensional vector space and let $T : V \to V$ be linear. Let W be a T-invariant subspace of V and define $T' : W \to W$ as the transformation such that $T'(x) = T(x)$ for any $x \in W$. Then we have

$$f_{T'}(t) \mid f_T(t).$$

Proof. Let $\gamma = (x_1, \ldots, x_k)$ be an ordered basis of W. By replacement theorem (Theorem 1.21), there is an ordered basis $\beta = (x_1, \ldots, x_k, x_{k+1}, \ldots, x_n)$ of V. It can be shown that

$$[T]_\beta^\beta = \begin{pmatrix} [T']_\gamma^\gamma & X \\ O & Y \end{pmatrix}$$

for some $X \in F^{k\times(n-k)}$ and $Y \in F^{(n-k)\times(n-k)}$. Thus, we have

$$\begin{aligned} f_T(t) &= \det([T]_\beta^\beta - tI_n) \\ &= \det \begin{pmatrix} [T']_\gamma^\gamma - tI_k & X \\ O & Y - tI_{n-k} \end{pmatrix} \\ &= \det([T']_\gamma^\gamma - tI_k) \cdot \det(Y - tI_{n-k}) \\ &= f_{T'}(t) \cdot \det(Y - tI_{n-k}). \end{aligned}$$

□

Definition 5.19. Let V be a vector space and let $T \in \mathcal{L}(V)$. The ***T*-cyclic subspace** of V generated by $x \in V$ is defined as

$$C_T(x) = \operatorname{span}\left(\bigcup_{i=0}^{\infty} \{T^i(x)\}\right).$$

Theorem 5.20. Let V be a vector space and let $T \in \mathcal{L}(V)$. Then the following statements hold for any $x \in V$.

(a) $C_T(x)$ is a T-invariant subspace of V.

(b) If W is a T-invariant subspace of V with $x \in W$, then $C_T(x) \subseteq W$.

Proof.

(a) Suppose that $y \in C_T(x)$ with

$$y = \sum_{i=0}^{k} a_i T^i(x).$$

Then we have

$$T(y) = T\left(\sum_{i=0}^{k} a_i T^i(x)\right) = \sum_{i=0}^{k} a_i T^{i+1}(x) \in C_T(x).$$

It follows that $T(C_T(x)) \subseteq C_T(x)$, and thus $C_T(x)$ is T-invariant.

(b) Since $x \in U$ and $T(U) \subseteq U$, we can conclude that $T^i(x) \in U$ holds for any nonnegative integer i. Thus, we have

$$\bigcup_{i=0}^{\infty}\{T^i(x)\} \subseteq U,$$

implying

$$C_T(x) = \operatorname{span}\left(\bigcup_{i=0}^{\infty}\{T^i(x)\}\right) \subseteq U. \qquad \square$$

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

$$\langle \cdot \mid \cdot \rangle : V \times V \to F$$

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

(a) $\langle ax + x' \mid y \rangle = a \langle x \mid y \rangle + \langle x' \mid y \rangle$.

(b) $\langle x \mid y \rangle = \overline{\langle y \mid x \rangle}$.

(c) $\langle x \mid x \rangle \in \mathbb{R}^+$ 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) $\langle x \mid ay + y' \rangle = \overline{a} \langle x \mid y \rangle + \langle x \mid y' \rangle$.

(b) $\langle x \mid 0_V \rangle = 0_F = \langle 0_V \mid x \rangle$.

(c) $\langle x \mid x \rangle = 0_F$ if and only if $x = 0_V$.

(d) If $\langle x \mid y \rangle = \langle x \mid y' \rangle$ holds for all $x \in V$, then $y = y'$.

Proof.

(a) It is proved by

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

(b) By

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

and

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

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

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

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

(d) Note that

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

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

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

- For $x, y \in V$, we say that x and y are **orthogonal** if

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

- 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 $x_1, \ldots, x_n$ be distinct vectors in S. Then for $y \in V$, if

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

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

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

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

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

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

implying

$$a_i = \frac{\langle y \mid x_i \rangle}{\langle x_i \mid x_i \rangle}.$$

□

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 $a_1, \ldots, a_n \in F$ and distinct vectors $x_1, \ldots, x_n \in S$ such that

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

Then we have

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

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

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

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

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

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, $\langle y_j \mid y_i \rangle = 0_F$ for distinct $i, j \in \{1, \ldots, n-1\}$. Then since for $i \in \{1, \ldots, n-1\}$, we have

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

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

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

because

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

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 $|S| = \dim(\operatorname{span}(R))$, we have $\operatorname{span}(S) = \operatorname{span}(R)$. □

Definition 6.7. Let V be an inner product space. For each vector $x \in S$, the **norm** of x is a nonnegative real number, defined as

$$\|x\| = \sqrt{\langle x \mid x \rangle}.$$

Proposition 6.8. Let V be an inner product space over a field $F \in \{\mathbb{R}, \mathbb{C}\}$. Then the following statements are true for any vectors $x, y \in V$ and any scalar $a \in F$.

(a) $\|ax\| = |a| \cdot \|x\|$.

(b) $\|x\| = 0_F$ if and only if $x = 0_V$.

Proof.

(a) We have

$$\|ax\| = \sqrt{\langle ax \mid ax \rangle} = \sqrt{a\bar{a} \langle x \mid x \rangle} = \sqrt{|a|^2 \langle x \mid x \rangle} = |a| \cdot \|x\|.$$

(b) We have
$$\|x\| = 0_F \quad \Leftrightarrow \quad \langle x \mid x \rangle = 0_F \quad \Leftrightarrow \quad x = 0_V.$$
$\square$

Theorem 6.9 (Pythagorean Theorem). Let V be an inner product space over a field $F \in \{\mathbb{R}, \mathbb{C}\}$. Then for any vectors $x, y \in V$ with $\langle x \mid y \rangle = 0_F$, we have
$$\|x + y\|^2 = \|x\|^2 + \|y\|^2 .$$

Proof. We have
$$\begin{aligned} \|x + y\|^2 &= \langle x + y \mid x + y \rangle \\ &= \langle x \mid x + y \rangle + \langle y \mid x + y \rangle \\ &= \langle x \mid x \rangle + \langle x \mid y \rangle + \langle y \mid x \rangle + \langle y \mid y \rangle \\ &= \langle x \mid x \rangle + 0_F + 0_F + \langle y \mid y \rangle \\ &= \|x\|^2 + \|y\|^2 . \end{aligned}$$
$\square$

Definition 6.10. Let V be an inner product space over a field $F \in \{\mathbb{R}, \mathbb{C}\}$. We say that a subset S of V is **orthonormal** if S is orthogonal and $\|x\| = 1_F$ for each $x \in S$.

Theorem 6.11. Let V be an inner product space over a field $F \in \{\mathbb{R}, \mathbb{C}\}$. Let S be an orthonormal subset of V and let $x_1, \ldots, x_n$ be distinct vectors in S. Then for $y \in V$, if
$$y = \sum_{i=1}^{n} a_i x_i$$
for some $a_1, \ldots, a_n \in F$, then
$$a_i = \langle y \mid x_i \rangle$$
for each $i \in \{1, \ldots, n\}$.

Proof. Since S is orthonormal, we have $0_V \notin S$. It follows that
$$a_i = \frac{\langle y \mid x_i \rangle}{\langle x_i \mid x_i \rangle} = \frac{\langle y \mid x_i \rangle}{1_F} = \langle y \mid x_i \rangle$$
for each $i \in \{1, \ldots, n\}$ by Theorem 6.4. $\square$

Definition 6.12. Let V be an inner product space over F and let S be a subspace of V. The **orthogonal complement** of S, denoted $S^\perp$, is the set of vectors that are orthogonal to every vector in S, i.e.,
$$S^\perp = \{x \in V : \langle x \mid y \rangle = 0_F \text{ for all } y \in S\}.$$

Theorem 6.13. Let V be an inner product space over F. For any subset S of V, $S^\perp$ is a subspace of V.

Proof. We have $0_V \in S^\perp$ since $\langle 0_V \mid z \rangle = 0_F$ for any $z \in S$. For any $a \in F$ and $x, y \in S^\perp$, we have
$$\begin{aligned} \langle ax + y \mid z \rangle &= a \langle x \mid z \rangle + \langle y \mid z \rangle \\ &= a0_F + 0_F \\ &= 0_F \end{aligned}$$
for any $z \in S$, implying $ax + y \in S^\perp$. Thus, $S^\perp$ is a subspace of V by **??**. $\square$

Theorem 6.14. Let V be a finite-dimensional inner product space over F. If W is a subspace of V, then $W \oplus W^{\perp} = V$.

Proof. Let $R = \{y_1, \ldots, y_k\}$ be an orthonormal basis of W. We have $W + W^{\perp} \subseteq V$ since W and $W^{\perp}$ are subspaces of V. To prove $V \subseteq W + W^{\perp}$, suppose that $x \in V$, and let

$$y = \sum_{i=1}^{k} \langle x \mid y_i \rangle \, y_i$$

be a vector in W. Then $x - y \in V^{\perp}$ since

$$\begin{aligned}
\langle x - y \mid y_j \rangle &= \left\langle x - \sum_{i=1}^{k} \langle x \mid y_i \rangle \, y_i \;\middle|\; y_j \right\rangle \\
&= \langle x \mid y_j \rangle - \sum_{i=1}^{k} \langle x \mid y_i \rangle \langle y_i \mid y_j \rangle \\
&= \langle x \mid y_j \rangle - \langle x \mid y_j \rangle \\
&= 0_F.
\end{aligned}$$

Thus,

$$x = y + (x - y) \in V + V^{\perp},$$

implying $V \subseteq W + W^{\perp}$, and thus $W + W^{\perp} = V$.

Furthermore, for any $x \in W \cap W^{\perp}$, we have $\langle x \mid x \rangle = 0_F$, implying $x = 0_V$. Thus, we have $W \cap W^{\perp} = \{0_V\}$, which implies $W \oplus W^{\perp} = V$. $\square$

6.2 The Adjoint of a Linear Operator

Theorem 6.15. Let V be a finite-dimensional inner product space over $F \in \{\mathbb{R}, \mathbb{C}\}$. Let $f : V \to F$ be a linear transformation. Then there exists a unique vector $y \in V$ such that

$$f(x) = \langle x \mid y \rangle$$

for all $x \in V$.

Proof. Let $S = \{x_1, x_2, \ldots, x_n\}$ be an orthonormal basis for V. Then we have

$$\begin{aligned} f(x) &= f\left(\sum_{i=1}^{n} \langle x \mid x_i \rangle \cdot x_i\right) \\ &= \sum_{i=1}^{n} \langle x \mid x_i \rangle \cdot f(x_i) \\ &= \left\langle x \,\middle|\, \sum_{i=1}^{n} \overline{f(x_i)} \cdot x_i \right\rangle. \end{aligned}$$

Thus, there exists

$$y = \sum_{i=1}^{n} \overline{f(x_i)} \cdot x_i$$

such that $f(x) = \langle x \mid y \rangle$ for all $x \in V$.

Furthermore, if there exists $y' \in V$ such that $f(x) = \langle x \mid y' \rangle$ for all $x \in V$, then we have $y' = y$ by Proposition 6.2 (d), which completes the proof. □

Theorem 6.16. Let V be a finite-dimensional inner product space over $F \in \{\mathbb{R}, \mathbb{C}\}$. For any linear operator $T : V \to V$, there exists a unique operator $T' : V \to V$ such that

$$\langle T(x) \mid y \rangle = \langle x \mid T'(y) \rangle$$

for all $x, y \in V$. Also, T' is linear.

Proof. Suppose that $y \in V$ is an arbitrary vector. Let $f : V \to F$ be a function such that $f(x) = \langle T(x) \mid y \rangle$ for each $x \in V$. Then f is linear since

$$\begin{aligned} f(ax_1 + x_2) &= \langle T(ax_1 + x_2) \mid y \rangle \\ &= \langle aT(x_1) + T(x_2) \mid y \rangle \\ &= a \langle T(x_1) \mid y \rangle + \langle T(x_2) \mid y \rangle \\ &= af(x_1) + f(x_2) \end{aligned}$$

holds for each $a \in F$ and for each $x_1, x_2 \in V$. Since f is linear, there exists a vector $y' \in V$ such that $f(x) = \langle x \mid y' \rangle$ by Theorem 6.15. Thus, we can define $T' : V \to V$ as the functoin with $T'(y) = y'$, implying $\langle T(x) \mid y \rangle = \langle x \mid T'(y) \rangle$ for each $x, y \in V$.

Now we prove that T' is linear. For any $a \in F$ and $x, y_1, y_2 \in V$, we have

$$\begin{aligned} \langle x \mid T'(ay_1 + y_2) \rangle &= \langle T(x) \mid ay_1 + y_2 \rangle \\ &= \overline{a} \langle T(x) \mid y_1 \rangle + \langle T(x) \mid y_2 \rangle \\ &= \overline{a} \langle x \mid T'(y_1) \rangle + \langle x \mid T'(y_2) \rangle \\ &= \langle x \mid aT'(y_1) + T'(y_2) \rangle. \end{aligned}$$

Thus, we can conclude that $T'(ay_1+y_2) = aT'(y_1)+T'(y_2)$ for any $a \in F$ and $y_1, y_2 \in V$ by Proposition 6.2 (d).

To show that T' is unique, suppose that $T'' : V \to V$ is linear and satisfies $\langle T(x) \mid y\rangle = \langle x \mid T''(y)\rangle$ for any $x, y \in V$. Then we have

$$\langle x \mid T''(y)\rangle = \langle T(x) \mid y\rangle = \langle x \mid T'(y)\rangle ,$$

implying $T''(y) = T'(y)$ for any $y \in V$ by Proposition 6.2 (d). Thus, $T'' = T'$. □

Definition 6.17. Let V be a finite-dimensional inner product space over $F \in \{\mathbb{R}, \mathbb{C}\}$ and let $T : V \to V$ be linear. The **adjoint** of T, denoted T^*, is the linear operator satisfying

$$\langle T(x) \mid y\rangle = \langle x \mid T^*(y)\rangle$$

for all $x, y \in V$.

Theorem 6.18. Let V be a finite-dimensional inner product space and let β be an ordered orthonormal basis of V. If $T : V \to V$ is linear, then

$$[T^*]_\beta^\beta = \left([T]_\beta^\beta\right)^*.$$

Proof. Suppose that $\dim(V) = n$ and $\beta = (x_1, x_2, \ldots, x_n)$. Let $A = [T^*]_\beta^\beta$ and $B = [T]_\beta^\beta$ be $n \times n$ matrices. Then for any $i, j \in \{1, \ldots, n\}$, we have

$$A_{ij} = \langle T^*(x_j) \mid x_i\rangle = \langle x_j \mid T(x_i)\rangle = \overline{\langle T(x_i) \mid x_j\rangle} = \overline{B_{ji}},$$

and thus $A = B^*$. □

Theorem 6.19. Let V be a finite-dimensional inner product space over F. Then the following statements hold for any $a \in F$ and $T_1, T_2, T \in \mathcal{L}(V)$.

(a) $(a \cdot T_1 + T_2)^* = \overline{a} \cdot T_1^* + T_2^*$.

(b) $(T_1T_2)^* = T_2^*T_1^*$.

(c) $(T^*)^* = T$.

(d) $I_V^* = I_V$.

Proof. To be completed. □

Corollary 6.20. Let $F \in \{\mathbb{R}, \mathbb{C}\}$ be a field. Then the following statements are true for any $c \in F$ and $A, B \in F^{n\times n}$.

(a) $(cA + B)^* = \overline{c}A^* + B^*$.

(b) $(AB)^* = B^*A^*$.

(c) $(A^*)^* = A$.

(d) $I_n^* = I_n$.

Proof. Straightforward from Theorem 6.19. □

Theorem 6.21. Let V be a inner product space over F and let $T : V \to V$ be linear. Then $\mathcal{R}(T^*)^\perp = \mathcal{N}(T)$.

Proof. The theorem is proved by

$$\begin{aligned} x \in \mathcal{R}(T^*)^\perp \quad &\Leftrightarrow \quad \langle x \mid T^*(y)\rangle = 0_F \text{ for all } y \in V \\ &\Leftrightarrow \quad \langle T(x) \mid y\rangle = 0_F \text{ for all } y \in V \\ &\Leftrightarrow \quad T(x) = 0_V \\ &\Leftrightarrow \quad x \in \mathcal{N}(T). \end{aligned}$$

□

Theorem 6.22. Let V be a finite-dimensional inner product space over F and let $T \in \mathcal{L}(V)$. Then $\overline{\lambda}$ is an eigenvalue of T^* if and only if λ is an eigenvalue of T.

Proof. The theorem is proved by

$$\begin{aligned} \mathcal{N}(T^* - \overline{\lambda} I_V) = \{0_V\} \quad &\Leftrightarrow \quad \mathcal{R}(T^* - \overline{\lambda} I_V) = V \\ &\Leftrightarrow \quad \mathcal{R}(T^* - \overline{\lambda} I_V)^\perp = \{0_V\} \\ &\Leftrightarrow \quad \mathcal{N}(T - \lambda I_V) = \{0_V\}. \end{aligned}$$

□

6.3 Normal and Self-Adjoint Operators

Definition 6.23. A polynomial f in $\mathcal{P}(F)$ **splits** if there are scalars $c, a_1, \ldots, a_n$ in F such that

$$f(t) = c\prod_{i=1}^{n}(t - a_i).$$

Theorem 6.24 (Schur's Theorem). Let V be a finite-dimensional inner product space over $F \in \{\mathbb{R}, \mathbb{C}\}$ and let $T : V \to V$ be linear. If f_T splits, then there is an orthonormal ordered basis β of V such that $[T]_\beta^\beta$ is upper triangular.

Proof. The proof is by induction on $n = \dim(V)$. The theorem holds trivially for $n = 1$. For $n \geq 2$, since f_T splits, T has an eigenvalue, and thus T^* has an eigenvalue by Theorem 6.22.

Suppose that (λ, x) is an eigenpair of T^* with $\|x\| = 1_F$. Let $W = \{x\}^\perp$. Then $\dim(W) = n - 1$, and we can conclude that W is T-invariant since

$$\langle x \mid T(y)\rangle = \langle T^*(x) \mid y\rangle = \langle \lambda x \mid y\rangle = \lambda \langle x \mid y\rangle = 0_F$$

for any $y \in W$. Define $T' : W \to W$ with $T'(y) = T(y)$ for each $y \in W$. It follows that $f_{T'}(t) \mid f_T(t)$, and thus $f_{T'}(t)$ splits. By induction hypothesis, there is an orthonormal ordered basis

$$\beta' = (x_1, \ldots, x_{n-1})$$

of W such that $A = [T']_{\beta'}^{\beta'}$ is upper triangular. We can conclude that

$$\beta = (x_1, \ldots, x_{n-1}, x)$$

is an orthonormal ordered basis of V, and it follows that $B = [T]_\beta^\beta$ is upper triangular since $B_{ij} = A_{ij}$ for all $i, j \in \{1, \ldots, n-1\}$, which completes the proof. □

Definition 6.25. Let V be an inner product space over $F \in \{\mathbb{R}, \mathbb{C}\}$.

- We say that $T \in \mathcal{L}(V)$ is **normal** if $TT^* = T^*T$.
- We say that $A \in F^{n\times n}$ is **normal** if $AA^* = A^*A$.

Theorem 6.26. Let V be an inner product space over $F \in \{\mathbb{R}, \mathbb{C}\}$, and let $T : V \to V$ be normal. Then the following statements hold.

(a) $\|T(x)\| = \|T^*(x)\|$ for any $x \in V$.

(b) $T - cI_V$ is normal for any $c \in F$.

(c) If (λ, x) is an eigenpair of T, then $(\overline{\lambda}, x)$ is an eigenpair of T^*.

(d) If λ_1 and λ_2 are distinct eigenvalues of T, then for any $x \in E_T(\lambda_1)$ and $y \in E_T(\lambda_2)$ we have $\langle x \mid y\rangle = 0_F$.

Proof. To be completed. □

Theorem 6.27. Let V be a finite-dimensional inner product space over $\mathbb{C}$ and let $T : V \to V$ be linear. Then T is normal if and only if there is an orthonormal eigenbasis of V for T.

Proof. ($\Rightarrow$) It can be shown that $f_T(t)$ splits by fundamental theorem of algebra. Thus, by Schur's theorem (Theorem 6.24), there is an orthonormal ordered basis $\beta = (x_1, \ldots, x_n)$ such that $A = [T]_\beta^\beta$ is upper triangular. By induction on $j \in \{1, \ldots, n\}$, we show that (A_{jj}, x_j) is an eigenpair of T. The induction basis with $j = 1$ holds trivially since A is upper triangular, implying $T(x_1) = A_{11}x_1$. For $j \geq 2$, we have

$$T(x_j) = \sum_{i=1}^{j} A_{ij} x_i,$$

and since

$$A_{ij} = \langle T(x_j) \mid x_i \rangle = \langle x_j \mid T^*(x_i) \rangle = \left\langle x_j \mid \overline{A_{ii}} x_i \right\rangle = A_{ii} \langle x_j \mid x_i \rangle = 0_F$$

holds for any $i \in \{1, \ldots, j-1\}$, it follows that $T(x_j) = A_{jj}x_j$. Thus, $x_1, \ldots, x_n$ are eigenvectors of T, implying β is an orthonormal eigenbasis of V.

($\Leftarrow$) Suppose that β is an orthonormal eigenbasis of T. Then $[T]_\beta^\beta$ is diagonal, implying

$$[T^*]_\beta^\beta = \left([T]_\beta^\beta\right)^*$$

is diagonal. It follows that

$$[TT^*]_\beta^\beta = [T]_\beta^\beta [T^*]_\beta^\beta = [T^*]_\beta^\beta [T]_\beta^\beta = [T^*T]_\beta^\beta,$$

implying $TT^* = T^*T$. $\square$

Definition 6.28. Let V be a finite-dimensional inner product space over $F \in \{\mathbb{R}, \mathbb{C}\}$.

- We say that $T \in \mathcal{L}(V)$ is **self-adjoint** if $T^* = T$.
- We say that $A \in F^{n \times n}$ is **self-adjoint** if $A^* = A$.

Theorem 6.29. Let V be a finite-dimensional inner product space over $F \in \{\mathbb{R}, \mathbb{C}\}$ and let $T : V \to V$ be self-adjoint. Then the following statements hold.

(a) Every eigenvalue of T is real.

(b) $f_T(t)$ splits.

Proof.

(a) Suppose that (λ, x) is an eigenpair of T. Note that T is normal since T is self-adjoint. By Theorem 6.26, $(\overline{\lambda}, x)$ is an eigenpair of $T^* = T$. Thus, $\overline{\lambda} = \lambda$, implying λ is real.

(b) Define $g_T(t) \in \mathcal{P}(\mathbb{C})$ such that $g_T(t) = f_T(t)$. By the fundamental theorem of algebra, we have

$$g_T(t) = \prod_{i=1}^{n} (t - \lambda_i)$$

for some $\lambda_1, \ldots, \lambda_n \in \mathbb{C}$. By (a), we can conclude that $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$. Thus, $f_T(t)$ splits even if $F = \mathbb{R}$. $\square$

Theorem 6.30. Let V be a finite-dimensional inner product space over $\mathbb{R}$ and let $T : V \to V$ be linear. Then T is self-adjoint if and only if there is an orthonormal eigenbasis of V for T.

Proof. ($\Rightarrow$) By Theorem 6.29 (b), $f_T(t)$ splits. Thus, by Schur's theorem (Theorem 6.24), there is an orthonormal ordered basis β such that $A = [T]_\beta^\beta$ is upper triangular. Moreover, $A^t = A$ since T is self-adjoint. Thus, A is diagonal, implying β is an orthonormal eigenbasis of V for T.

($\Leftarrow$) Suppose that β is an orthonormal eigenbasis of V for T. It follows that $A = [T]_\beta^\beta$ is diagonal, implying that A is self-adjoint. Thus, T is self-adjoint. $\square$

6.4 Unitary and Orthogonal Operators

Definition 6.31. Let V be an inner product space over F. Let $T : V \to V$ be linear.

- We say that T is **unitary** if $F = \mathbb{C}$ and $\|T(x)\| = \|x\|$ for any $x \in V$.
- We say that T is **orthogonal** if $F = \mathbb{R}$ and $\|T(x)\| = \|x\|$ for any $x \in V$.

Theorem 6.32. Let V be a finite-dimensional inner product space over F. Then the following statements are equivalent.

(a) $\|T(x)\| = \|x\|$ for any $x \in V$.

(b) $T^*T = I_V$.

(c) $\langle T(x) \mid T(y)\rangle = \langle x \mid y\rangle$ for any $x, y \in V$.

(d) If S is an orthonormal basis of V, so is $T(S)$.

(e) There is a subset S of V such that both S and $T(S)$ are orthonormal bases of V.

Proof. First we prove (b) from (a). Note that T^*T is self-adjoint and normal since $(T^*T)^* = T^*(T^*)^* = T^*T$. Thus, there exists an orthonormal basis $S = \{x_1, \ldots, x_n\}$ of V such that for any $i \in \{1, \ldots, n\}$, $T^*T(x_i) = \lambda_i x_i$ holds for some $\lambda_i \in F$. Since

$$\lambda_i = \lambda_i \langle x_i \mid x_i\rangle = \langle \lambda_i x_i \mid x_i\rangle = \langle T^*T(x_i) \mid x_i\rangle = \langle T(x_i) \mid T(x_i)\rangle = \langle x_i \mid x_i\rangle = 1_F$$

holds for each $i \in \{1, \ldots, n\}$, we have $T^*T(x) = x$ for any $x \in V$ by Lemma 2.21. Thus, $T^*T = I_V$.

Now we prove (c) from (b). The proof is given by

$$\langle T(x) \mid T(y)\rangle = \langle x \mid T^*T(y)\rangle = \langle x \mid y\rangle$$

for any $x, y \in V$.

Now we prove (d) from (c). Let $S = \{x_1, \ldots, x_n\}$ be an orthonormal basis of V. Then for any $i, j \in \{1, \ldots, n\}$ we have

$$\langle T(x_i) \mid T(x_j)\rangle = \langle x_i \mid x_j\rangle = [\![i = j]\!],$$

implying $T(S)$ is an orthonormal basis of V.

The proof of (e) from (d) is trivial. To prove (e) from (a), let $S = \{x_1, \ldots, x_n\}$. Then for any $x \in V$ with $x = a_1x_1 + \cdots + a_nx_n$ for some $a_1, \ldots, a_n \in F$, we have

$$\begin{aligned}
\langle T(x) \mid T(x)\rangle &= \left\langle \sum_{i=1}^{n} a_i T(x_i) \,\middle|\, \sum_{j=1}^{n} a_j T(x_j) \right\rangle \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} a_i \overline{a_j} \langle T(x_i) \mid T(x_j)\rangle \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} a_i \overline{a_j} \langle x_i \mid x_j\rangle \\
&= \left\langle \sum_{i=1}^{n} a_i x_i \,\middle|\, \sum_{j=1}^{n} a_j x_j \right\rangle \\
&= \langle x \mid x\rangle,
\end{aligned}$$

completing the proof. $\square$

Theorem 6.33. Let V be a finite-dimensional inner product space over F and let $T : V \to V$ be linear.

(a) Let $F = \mathbb{C}$. Then T is unitary if and only if V has an orthonormal eigenbasis for T and $|\lambda| = 1_F$ holds for each eigenvalue λ of T.

(b) Let $F = \mathbb{R}$. Then T is orthogonal and self-adjoint if and only if V has an orthonormal eigenbasis for T and $|\lambda| = 1_F$ holds for each eigenvalue λ of T.

Proof. We prove both statements simultaneously.

($\Rightarrow$) V has an orthonormal eigenbasis by Theorem 6.27 and Theorem 6.30. For each eigenpair (λ, x) of T, $(\overline{\lambda}, x)$ is an eigenpair of T^* by Theorem 6.26 (c), and we have

$$|\lambda|^2 x = \lambda\overline{\lambda}x = \lambda T^*(x) = T^*(\lambda x) = T^*(T(x)) = x$$

by Theorem 6.32, implying $|\lambda| = 1_F$.

($\Leftarrow$) Let $\beta = (x_1, \ldots, x_n)$ be an orthonormal eigenbasis of V for T. For each $i \in \{1, \ldots, n\}$, let λ_i be the corresponding eigenvalue of x_i for T. Since T is normal by Theorem 6.27 and Theorem 6.30, we have

$$T^*(T(x_i)) = T^*(\lambda_i x_i) = \lambda_i T^*(x_i) = \lambda_i\overline{\lambda_i}x_i = |\lambda_i|^2 x_i = x_i$$

for any $i \in \{1, \ldots, n\}$, implying $T^*T = I_V$. The proof is completed due to Theorem 6.32. $\square$

Definition 6.34. Let $Q \in F^{n\times n}$ with $F \in \{\mathbb{C}, \mathbb{R}\}$.

- We say that Q is **unitary** if $Q^*Q = I_n$.
- We say that Q is **orthogonal** if $Q^tQ = I_n$.

Definition 6.35. Let $A, B \in F^{n\times n}$ with $F \in \{\mathbb{C}, \mathbb{R}\}$.

- We say that A and B are **unitarily equivalent** if $B = QAQ^*$ for some unitary $Q \in F^{n\times n}$.
- We say that A and B are **orthogonally equivalent** if $B = QAQ^t$ for some orthogonal $Q \in F^{n\times n}$.

Theorem 6.36. Let $A \in F^{n\times n}$.

(a) If $F = \mathbb{C}$, then A is normal if and only if A is unitarily equivalent to a diagonal matrix in $F^{n\times n}$.

(b) If $F = \mathbb{R}$, then A is self-adjoint if and only if A is orthogonally equivalent to a diagonal matrix in $F^{n\times n}$.

Proof. We prove both statements simultaneously.

($\Rightarrow$) Let α be the standard ordered basis of F^n, and let $\beta = (x_1, \ldots, x_n)$ be an orthonormal eigenbasis of F^n for L_A. Then $B = [L_A]_\beta^\beta$ is diagonal. Define $Q = [I_{F^n}]_\beta^\alpha$, and we have

$$(Q^*Q)_{ij} = \sum_{k=1}^{n} \overline{Q_{ki}}Q_{kj} = \sum_{k=1}^{n} \overline{(x_i)_k}(x_j)_k = x_i^* x_j = [\![i = j]\!]$$

for any $i, j \in \{1, \ldots, n\}$, implying $Q^*Q = I_n$. Thus, A and B are unitarily equivalent since

$$B = [L_A]_\beta^\beta = [I_{F^n}]_\alpha^\beta [L_A]_\alpha^\alpha [I_{F^n}]_\beta^\alpha = Q^{-1}AQ = Q^*AQ.$$

($\Leftarrow$) Let Q be a unitary matrix such that $B = QAQ^*$ is diagonal. We have $A = Q^*BQ$. Thus, A is normal since

$$A^*A = (Q^*B^*Q)(Q^*BQ) = Q^*B^*BQ = Q^*BB^*Q = (Q^*BQ)(Q^*B^*Q) = AA^*.$$

If $F = \mathbb{R}$, then $B^* = B$, and it follows that A is self-adjoint since

$$A^* = Q^*B^*Q = Q^*BQ = A.$$

□

6.5 The Singular Value Decomposition

Definition 6.37. Let V be an inner product space over F.

- A self-adjoint operator $T : V \to V$ is said to be **positive semidefinite** and **positive definite** if $\langle T(x) \mid x \rangle$ is nonnegative and positive for any $x \in V$, respectively.

- A self-adjoint matrix $A \in F^{n\times n}$ is said to be **positive semidefinite** and **positive definite** if L_A is positive semidefinite and positive definite, respectively.

Theorem 6.38. Let V and W be finite-dimensional inner product spaces over F with $\dim(V) = n$ and $\dim(W) = m$. Let $T : V \to W$ be linear with $\operatorname{rank}(T) = r$. Then there exist positive real numbers $\sigma_1, \ldots, \sigma_r$ and orthonormal bases $\{x_1, \ldots, x_n\}$ and $\{y_1, \ldots, y_m\}$ of V and W, respectively, such that the following statements hold with $\sigma_k = 0_F$ for $k > r$, $x_k = 0_V$ for $k > n$, and $y_k = 0_W$ for $k > m$.

(a) $T(x_i) = \sigma_i y_i$ for any $i \in \{1, \ldots, n\}$.

(b) $T^*(y_j) = \sigma_j x_j$ for any $j \in \{1, \ldots, m\}$.

(c) (σ_i^2, x_i) is an eigenpair of T^*T for any $i \in \{1, \ldots, n\}$.

(d) (σ_j^2, y_j) is an eigenpair of TT^* for any $j \in \{1, \ldots, m\}$.

Proof. To be completed. □

Chapter 7

Canonical Forms

7.1 Generalized Eigenspaces

Definition 7.1. Let V be a vector space over F and let $T : V \to V$ be linear. We say that $\lambda \in F$ and $x \in V \setminus \{0_V\}$ form a **generalized eigenpair** if

$$(T - \lambda I_V)^{\ell}(x) = 0_V$$

holds for some positive integer ℓ.

Theorem 7.2. Let V be a vector space over F and let $T : V \to V$ be linear. If (λ, x) is a generalized eigenpair of T, then λ is an eigenvalue of T.

Proof. Let ℓ be the smallest positive integer such that $(T - \lambda I_V)^{\ell}(x) = 0_V$. Let

$$y = (T - \lambda I_V)^{\ell-1}(x).$$

Since $(T - \lambda I_V)(y) = (T - \lambda I_V)^{\ell}(x) = 0_V$, (λ, y) is an eigenpair of T, and thus λ is an eigenvalue of T. $\square$

Definition 7.3. Let V be a vector space over F and let $T : V \to V$ be linear. For any scalar $\lambda \in F$, we define

$$G_T(\lambda) = \{x \in V : (T - \lambda I_V)^{\ell}(x) = 0_V \text{ for some nonnegative integer } \ell\}.$$

If λ is an eigenvalue of T, then $G_T(\lambda)$ is called the **generalized eigenspace** of T corresponding to λ.

Theorem 7.4. Let V be a vector space over F and let $T : V \to V$ be linear. If scalars $\lambda_1, \lambda_2 \in F$ are distinct, then

$$G_T(\lambda_1) \cap G_T(\lambda_2) = \{0_V\}.$$

Proof. Assume $x \in (G_T(\lambda_1) \cap G_T(\lambda_2)) \setminus \{0_V\}$ for contradiction. Let ℓ_1 be the smallest positive integer with

$$(T - \lambda_1 I_V)^{\ell_1}(x) = 0_V.$$

Let $y = (T - \lambda_1 I_V)^{\ell_1 - 1}(x)$, and we have $(T - \lambda_1 I_V)(y) = 0_V$. Note that there is a positive integer ℓ_2 such that

$$(T - \lambda_2 I_V)^{\ell_2}(x) = 0_V,$$

and it follows that

$$\begin{aligned}(T-\lambda_2 I_V)^{\ell_2}(y) &= (T-\lambda_2 I_V)^{\ell_2}(T-\lambda_2 I_V)^{\ell_1-1}(x)\\ &= (T-\lambda_1 I_V)^{\ell_1-1}(T-\lambda_2 I_V)^{\ell_2}(x)\\ &= 0_V.\end{aligned}$$

Thus we can define ℓ_2' as the smallest positive integer such that

$$(T-\lambda_2 I_V)^{\ell_2'}(y) = 0_V.$$

Let $z = (T-\lambda_2 I_V)^{\ell_2'-1}(y)$, and we have $(T-\lambda_2 I_V)(z) = 0_V$. Furthermore,

$$\begin{aligned}(T-\lambda_1 I_V)(z) &= (T-\lambda_1 I_V)(T-\lambda_2 I_V)^{\ell_2'}(y)\\ &= (T-\lambda_2 I_V)^{\ell_2'}(T-\lambda_1 I_V)(y)\\ &= 0_V.\end{aligned}$$

Thus, $z \in (E_T(\lambda_1) \cap E_T(\lambda_2)) \setminus \{0_V\}$, contradiction. $\square$

7.2 The Jordan Canonical Form

Definition 7.5. Let V be a vector space over F and let $T : V \to V$ be linear. If (λ, x) is a generalized eigenpair and ℓ is the smallest positive integer such that

$$(T - \lambda I_V)^{\ell}(x) = 0_V,$$

then the ordered set

$$((T - \lambda I_V)^{\ell-1}(x), (T - \lambda I_V)^{\ell-2}(x), \ldots, (T - \lambda I_V)^{2}(x), (T - \lambda I_V)(x), x)$$

is a **chain** of generalized eigenvectors of T corresponding to λ, where

- ℓ is called the **length** of the chain, and
- $(T - \lambda I_V)^{\ell-1}(x)$ and x are called the **initial vector** and the **end vector** of the chain, respectively.