Notes for

Math 331

Abstract Linear Algebra

using Friedberg-Insel-Spence, 4th ed.

Ethan D. Bloch

Revised draft
October 10, 2017

Contents

2 Binary Operations

2.1 Binary Operations

Definition 2.1.1. Let A be a set. A **binary operation** on A is a function $A \times A \to A$. A **unary operation** on A is a function $A \to A$. △

Definition 2.1.2. Let A be a set, let $*$ be a binary operation on A and let $H \subseteq A$. The subset H is **closed** under $*$ if $a * b \in H$ for all $a, b \in H$. △

Definition 2.1.3. Let A be a set, and let $*$ be a binary operation on A. The binary operation $*$ satisfies the **Commutative Law** (an alternative expression is that $*$ is **commutative**) if $a * b = b * a$ for all $a, b \in A$. △

Definition 2.1.4. Let A be a set, and let $*$ be a binary operation on A. The binary operation $*$ satisfies the **Associative Law** (an alternative expression is that $*$ is **associative**) if $(a * b) * c = a * (b * c)$ for all $a, b, c \in A$. △

Definition 2.1.5. Let A be a set, and let $*$ be a binary operation on A.

1. Let $e \in A$. The element e is an **identity element** for $*$ if $a * e = a = e * a$ for all $a \in A$.

2. If $*$ has an identity element, the binary operation $*$ satisfies the **Identity Law**. △

Lemma 2.1.6. *Let A be a set, and let $*$ be a binary operation on A. If $*$ has an identity element, the identity element is unique.*

Proof. Let $e, \hat{e} \in A$. Suppose that e and $\hat{e}$ are both identity elements for $*$. Then $e = e * \hat{e} = \hat{e}$, where in the first equality we are thinking of $\hat{e}$ as an identity element, and in the second equality we are thinking of e as an identity element. Therefore the identity element is unique. □

Definition 2.1.7. Let A be a set, and let $*$ be a binary operation of A. Let $e \in A$. Suppose that e is an identity element for $*$.

1. Let $a \in A$. An **inverse** for a is an element $a' \in A$ such that $a * a' = e$ and $a' * a = e$.

2. If every element in A has an inverse, the binary operation $*$ satisfies the **Inverses Law**. △

Definition 2.1.8. Let A be a set, and let $+$ and $\cdot$ be binary operations on A.

1. The binary operations $+$ and $\cdot$ satisfy the **Left Distributive Law** (an alternative expression is that $\cdot$ is **left distributive over** $+$) if $a \cdot (b + c) = (a \cdot b) + (a \cdot c)$ for all $a, b, c \in A$.

2. The binary operations $+$ and $\cdot$ satisfy the **Right Distributive Law** (an alternative expression is that $\cdot$ is **right distributive over** $+$) if $(b + c) \cdot a = (b \cdot a) + (c \cdot a)$ for all $a, b, c \in A$.

3. The binary operations $+$ and $\cdot$ satisfy the **Distributive Law** (an alternative expression is that $\cdot$ is **distributive over** $+$) if they satisfy both the Left Distributive Law and the Right Distributive Law. $\triangle$

Exercises

Exercise 2.1.1. Which of the following formulas defines a binary operation on the given set?

(1) Let $*$ be defined by $x * y = xy$ for all $x, y \in \{-1, -2, -3, \ldots\}$.

(2) Let $\diamond$ be defined by $x \diamond y = \sqrt{xy}$ for all $x, y \in [2, \infty)$.

(3) Let $\oplus$ be defined by $x \oplus y = x - y$ for all $x, y \in \mathbb{Q}$.

(4) Let $\circ$ be defined by $(x, y) \circ (z, w) = (x + z, y + w)$ for all $(x, y), (z, w) \in \mathbb{R}^2 - \{(0, 0)\}$.

(5) Let $\odot$ be defined by $x \odot y = |x + y|$ for all $x, y \in \mathbb{N}$.

(6) Let $\otimes$ be defined by $x \otimes y = \ln(|xy| - e)$ for all $x, y \in \mathbb{N}$.

Exercise 2.1.2. For each of the following binary operations, state whether the binary operation is associative, whether it is commutative, whether there is an identity element and, if there is an identity element, which elements have inverses.

(1) The binary operation $\oplus$ on $\mathbb{Z}$ defined by $x \oplus y = -xy$ for all $x, y \in \mathbb{Z}$.

(2) The binary operation $\star$ on $\mathbb{R}$ defined by $x \star y = x + 2y$ for all $x, y \in \mathbb{R}$.

(3) The binary operation $\otimes$ on $\mathbb{R}$ defined by $x \otimes y = x + y - 7$ for all $x, y \in \mathbb{R}$.

(4) The binary operation $*$ on $\mathbb{Q}$ defined by $x * y = 3(x + y)$ for all $x, y \in \mathbb{Q}$.

(5) The binary operation $\circ$ on $\mathbb{R}$ defined by $x \circ y = x$ for all $x, y \in \mathbb{R}$.

(6) The binary operation $\diamond$ on $\mathbb{Q}$ defined by $x \diamond y = x + y + xy$ for all $x, y \in \mathbb{Q}$.

(7) The binary operation $\odot$ on $\mathbb{R}^2$ defined by $(x, y) \odot (z, w) = (4xz, y + w)$ for all $(x, y), (z, w) \in \mathbb{R}^2$.

Exercise 2.1.3. For each of the following binary operations given by operation tables, state whether the binary operation is commutative, whether there is an identity element and, if there is an identity element, which elements have inverses. (Do not check for associativity.)

(1)
$$\begin{array}{c|ccc} \otimes & 1 & 2 & 3 \\ \hline 1 & 1 & 2 & 1 \\ 2 & 2 & 3 & 2 \\ 3 & 1 & 2 & 3 \end{array}.$$

(2)

$\odot$	j	k	l	m
j	k	j	m	j
k	j	k	l	m
l	k	l	j	l
m	j	m	l	m

(4)

$\star$	a	b	c	d	e
a	d	e	a	b	b
b	e	a	b	a	d
c	a	b	c	d	e
d	b	a	d	e	c
e	b	d	e	c	a

(3)

$*$	x	y	z	w
x	x	z	w	y
y	z	w	y	x
z	w	y	x	z
w	y	x	z	w

(5)

$\diamond$	i	r	s	a	b	c
i	i	r	s	a	b	c
r	r	s	i	c	a	b
s	s	i	r	b	c	a
a	a	b	c	i	s	r
b	b	c	a	r	i	s
c	c	a	b	s	r	i

Exercise 2.1.4. Find an example of a set and a binary operation on the set such that the binary operation satisfies the Identity Law and Inverses Law, but not the Associative Law, and for which at least one element of the set has more than one inverse. The simplest way to solve this problem is by constructing an appropriate operation table.

Exercise 2.1.5. Let $n \in \mathbb{N}$. Recall the definition of the set $\mathbb{Z}_n$ and the binary operation $\cdot$ on $\mathbb{Z}_n$. Observe that $[1]$ is the identity element for $\mathbb{Z}_n$ with respect to multiplication. Let $a \in \mathbb{Z}$. Prove that the following are equivalent.

a. The element $[a] \in \mathbb{Z}_n$ has an inverse with respect to multiplication.

b. The equation $ax \equiv 1 \pmod{n}$ has a solution.

c. There exist $p, q \in \mathbb{Z}$ such that $ap + nq = 1$.

(It turns out that the three conditions listed above are equivalent to the fact that a and n are relatively prime.)

Exercise 2.1.6. Let A be a set. A **ternary operation** on A is a function $A \times A \times A \to A$. A ternary operation $\star\colon A \times A \times A \to A$ is **left-induced** by a binary operation $\diamond\colon A \times A \to A$ if $\star((a, b, c)) = (a \diamond b) \diamond c$ for all $a, b, c \in A$.

Is every ternary operation on a set left-induced by a binary operation? Give a proof or a counterexample.

Exercise 2.1.7. Let A be a set, and let $*$ be a binary operation on A. Suppose that $*$ satisfies the Associative Law and the Commutative Law. Prove that $(a * b) * (c * d) = b * [(d * a) * c]$ for all $a, b, c, d \in A$.

Exercise 2.1.8. Let B be a set, and let $\diamond$ be a binary operation on B. Suppose that $\diamond$ satisfies the Associative Law. Let

$$P = \{b \in B \mid b \diamond w = w \diamond b \text{ for all } w \in B\}.$$

Prove that P is closed under $\diamond$.

Exercise 2.1.9. Let C be a set, and let $\star$ be a binary operation on C. Suppose that $\star$ satisfies the Associative Law and the Commutative Law. Let

$$Q = \{c \in C \mid c \star c = c\}.$$

Prove that Q is closed under $\star$.

Exercise 2.1.10. Let A be a set, and let $*$ be a binary operation on A. An element $c \in A$ is a **left identity element** for $*$ if $c * a = a$ for all $a \in A$. An element $d \in A$ is a **right identity element** for $*$ if $a * d = a$ for all $a \in A$.

(1) If A has a left identity element, is it unique? Give a proof or a counterexample.

(2) If A has a right identity element, is it unique? Give a proof or a counterexample.

(3) If A has a left identity element and a right identity element, do these elements have to be equal? Give a proof or a counterexample.

3 Vector Spaces

3.1 Fields

Friedberg-Insel-Spence, 4th ed. – Section Appendix C

Definition 3.1.1. A **field** is a non-empty set F with two elements denoted 0 and 1, and with two binary operations $+\colon F \times F \to F$ and $\cdot\colon F \times F \to F$ that satisfy the following properties. Let $a, b, c \in F$.

1. $(a + b) + c = a + (b + c)$ (Associative Law for $+$).
2. $a + b = b + a$ (Commutative Law for $+$).
3. $a + 0 = a$ (Identity Law for $+$).
4. There is an element $-a \in F$ such that $a + (-a) = 0$ (Inverses Law for $+$).
5. $(a \cdot b) \cdot c = a \cdot (b \cdot c)$ (Associative Law for $\cdot$).
6. $a \cdot b = b \cdot a$ (Commutative Law for $\cdot$).
7. $a \cdot 1 = a$ (Identity Law for $\cdot$).
8. If $a \neq 0$, there is an element $a^{-1} \in F$ such that $a \cdot a^{-1} = 1$ (Inverses Law for $\cdot$).
9. $a \cdot (b + c) = a \cdot b + a \cdot c$ (Distributive Law).
10. $0 \neq 1$ (Non-Triviality). △

Lemma 3.1.2. *Let F be a field, and let $a, b, c \in F$.*

1. *0 is unique.*
2. *1 is unique.*
3. *$-a$ is unique.*
4. *If $a \neq 0$, then a^{-1} is unique.*
5. *$a + b = a + c$ implies $b = c$.*
6. *If $a \neq 0$, then $a \cdot b = a \cdot c$ implies $b = c$.*
7. *$a \cdot 0 = 0$.*
8. *$-(-a) = a$.*
9. *If $a \neq 0$, then $(a^{-1})^{-1} = a$.*

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

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

12. $(-1) \cdot a = -a$.

13. 0 *has no multiplicative inverse.*

14. $ab = 0$ *if and only if* $a = 0$ *or* $b = 0$.

Proof. We prove Parts (1), (2), (3), (7) and (10); the remaining parts of this lemma are left to the reader in Exercise 3.1.1.

For the proof of each part, we can use any of the previous parts, but not any of the subsequent ones.

(1). and **(2).** These two parts follow immediately from Lemma 2.1.6.

(3). Let $g \in F$. Suppose that $a + g = 0$. We also know that $a + (-a) = 0$. Hence $a + g = a + (-a)$. Then $(-a) + (a + g) = (-a) + (a + (-a))$. By the Associate Law for $+$ we obtain $((-a) + a) + g = ((-a) + a) + (-a)$. By the Inverses Law for $+$ we deduce that $0 + g = 0 + (-a)$. By the Identity Law for $+$ it follows that $g = -a$, which means that $-a$ is unique.

(7). By the Identity Law for $+$ we know that $0 + 0 = 0$. Then $a \cdot (0 + 0) = a \cdot 0$. By the Distributive Law we see that $a \cdot 0 + a \cdot 0 = a \cdot 0$. By the Identity Law for $+$ again we deduce $a \cdot 0 + a \cdot 0 = a \cdot 0 + 0$. It then follows from Part (5) of this lemma that $a \cdot 0 = 0$.

(10). We will show that $a \cdot (-b) = -(a \cdot b)$. The other equality is similar, and the details are omitted. Using the Distributive Law, the Inverses Law for $+$ and Part (7) of this lemma, in that order, we see that $a \cdot b + a \cdot (-b) = a \cdot (b + (-b)) = a \cdot 0 = 0$. It now follows from Part (3) of this lemma that $a \cdot (-b) = -(a \cdot b)$. □

Exercises

Exercise 3.1.1. Prove Lemma 3.1.2 (4), (5), (6), (8), (9), (11), (13) and (14).

3.2 Vector Spaces

Friedberg-Insel-Spence, 4th ed. – Section 1.2

Definition 3.2.1. Let F be a field. A **vector space** (also called a **linear space**) over F is a set V with a binary operation $+ \colon V \times V \to V$ and scalar multiplication $F \times V \to V$ that satisfy the following properties. Let $x, y, z \in V$ and let $a, b \in F$.

1. $(x + y) + z = x + (y + z)$ (Associative Law).
2. $x + y = y + x$ (Commutative Law).
3. There is an element $0 \in V$ such that $x + 0 = x$ (Identity Law).
4. There is an element $-x \in V$ such that $x + (-x) = 0$ (Inverses Law).
5. $1x = x$.
6. $(ab)x = a(bx)$.
7. $a(x + y) = ax + ay$ (Distributive Law).
8. $(a + b)x = ax + bx$ (Distributive Law). △

Definition 3.2.2. Let F be a field, and let $m, n \in \mathbb{N}$. The set of all $m \times n$ matrices with entries in F is denoted $\mathsf{M}_{m\times n}(F)$. An element $A \in \mathsf{M}_{m\times n}(F)$ is abbreviated by the notation $A = \left[a_{ij}\right]$. △

Definition 3.2.3. Let F be a field, and let $m, n \in \mathbb{N}$.

1. The $m \times n$ **zero matrix** is the matrix O_{mn} defined by $O_{mn} = \left[c_{ij}\right]$, where $c_{ij} = 0$ for all $i \in \{1, \ldots, m\}$ and $j \in \{1, \ldots, n\}$.
2. The $n \times n$ **identity matrix** is the matrix I_n defined by $I_n = \left[\delta_{ij}\right]$, where

$$\delta_{ij} = \begin{cases} 1, & \text{if } i = j \\ 0, & \text{if } i \neq j \end{cases}$$

for all $i, j \in \{1, \ldots, n\}$. △

Definition 3.2.4. Let F be a field, and let $m, n \in \mathbb{N}$. Let $A, B \in \mathsf{M}_{m\times n}(F)$, and let $c \in F$. Suppose that $A = \left[a_{ij}\right]$ and $B = \left[b_{ij}\right]$.

1. The matrix $A + B \in \mathsf{M}_{m\times n}(F)$ is defined by $A + B = \left[c_{ij}\right]$, where $c_{ij} = a_{ij} + b_{ij}$ for all $i \in \{1, \ldots, m\}$ and $j \in \{1, \ldots, n\}$.

2. The matrix $-A \in \mathsf{M}_{m\times n}(F)$ is defined by $-A = \left[d_{ij}\right]$, where $d_{ij} = -a_{ij}$ for all $i \in \{1, \ldots, m\}$ and $j \in \{1, \ldots, n\}$.

3. The matrix $cA \in \mathsf{M}_{m\times n}(F)$ is defined by $cA = \left[s_{ij}\right]$, where $s_{ij} = ca_{ij}$ for all $i \in \{1, \ldots, m\}$ and $j \in \{1, \ldots, n\}$. △

Lemma 3.2.5. *Let F be a field, and let $m, n \in \mathbb{N}$. Let $A, B, C \in \mathsf{M}_{m\times n}(F)$, and let $s, t \in F$.*

1. *$A + (B + C) = (A + B) + C$.*

2. *$A + B = B + A$.*

3. *$A + O_{mn} = A$ and $A + O_{mn} = A$.*

4. *$A + (-A) = O_{mn}$ and $(-A) + A = O_{mn}$.*

5. *$1A = A$.*

6. *$(st)A = s(tA)$.*

7. *$s(A + B) = sA + sB$.*

8. *$(s + t)A = sA + tA$.*

Proof. The proofs of these facts about matrices are straightforward, and are material belonging to Elementary Linear Algebra; we omit the details. □

Corollary 3.2.6. *Let F be a field, and let $m, n \in \mathbb{N}$. Then $\mathsf{M}_{m\times n}(F)$ is a vector space over F.*

Lemma 3.2.7. *Let V be a vector space over a field F. let $x, y, z \in V$ and let $a \in F$.*

1. *$x + y = x + z$ implies $y = z$.*

2. *If $x + y = x$, then $y = 0$.*

3. *If $x + y = 0$, then $y = -x$.*

4. *$-(x + y) = (-x) + (-y)$.*

5. *$\mathsf{0}x = 0$.*

6. *$a0 = 0$.*

7. *$(-a)x = a(-x) = -(ax)$.*

8. *$(-1)x = -x$.*

9. *$ax = 0$ if and only if $a = \mathsf{0}$ or $x = 0$.*

Proof. We prove Parts (1), (4) and (9); the remaining parts of this lemma are left to the reader in Exercise 3.2.1.

For the proof of each part, we can use any of the previous parts, but not any of the subsequent ones.

(1). Suppose that $x + y = x + z$. Then $(-x) + (x + y) = (-x) + (x + z)$. By the Associate Law we obtain $((-x) + x) + y = ((-x) + x) + z$. By the Commutative Law we obtain $(x + (-x)) + y = (x + (-x)) + z$. By the Inverses Law we deduce that $0 + y = 0 + z$. By the Identity Law it follows that $y = z$

(4). Using the Associate Law and the Commutative Law repeatedly, and then the Inverses Law and the Identity Law, we compute $(x + y) + ((-x) + (-y)) = ((x + y) + (-x)) + (-y) = ((y + x) + (-x)) + (-y) = (y + (x + (-x)) + (-y) = (y + 0) + (-y) = y + (-y) = 0$. It now follows from Part (3) of this lemma that $(-x) + (-y) = -(x + y)$.

(9). First, suppose that $a = 0$ or $x = 0$. Then it follows from Parts (5) and (6) of this lemma that $ax = 0$.

Second, suppose that $ax = 0$. Suppose further that $a \neq 0$. Then there is an element $a^{-1} \in F$ such that $aa^{-1} = 1$. Then $a^{-1}(ax) = a^{-1}0$. By Property (6) of Definition 3.2.1, together with Part (6) of this lemma, we see that $(a^{-1}a)x = 0$. By Property (6) of Definition 3.1.1 it follows that $(aa^{-1})x = 0$. Therefore $1x = 0$. By Property (5) of Definition 3.2.1 we deduce that $x = 0$. □

Remark 3.2.8. Let V be a vector space over a field F. The additive identity element 0 of V is unique, which can be seen either from Lemma 3.2.7 (2) or from Lemma 2.1.6. Moreover, for each $x \in V$, its additive inverse $-x$ is unique, as can be seen from Lemma 3.2.7 (3). ◊

Exercises

Exercise 3.2.1. Prove Lemma 3.2.7 (2), (3), (5), (6), (7) and (8).

Exercise 3.2.2. Let V, W be vector spaces over a field F. Define addition and scalar multiplication on $V \times W$ as follows. For each $(v, w), (x, y) \in V \times W$ and $c \in F$, let

$$(v, w) + (x, y) = (v + x, w + y) \qquad \text{and} \qquad c(v, w) = (cv, cw).$$

Prove that $V \times W$ is a vector space over F with these operations. This vector space is called the **product vector space** of V and W.

Exercise 3.2.3. Let F be a field, and let S be a non-empty set. Let $\mathcal{F}(S, F)$ be the set of all functions $S \to F$. Define addition and scalar multiplication on $\mathcal{F}(S, F)$ as follows. For each $f, g \in \mathcal{F}(S, F)$ and $c \in F$, let $f + g, cf \in \mathcal{F}(S, F)$ be defined by $(f + g)(x) = f(x) + g(x)$ and $(cf)(x) = cf(x)$ for all $x \in S$.

Prove that $\mathcal{F}(S, F)$ is a vector space over F with these operations.

3.3 Subspaces

Friedberg-Insel-Spence, 4th ed. – Section 1.3

Definition 3.3.1. Let V be a vector space over a field F, and let $W \subseteq V$. The subset W is **closed** under scalar multiplication by F if $av \in W$ for all $v \in W$ and $a \in F$. △

Definition 3.3.2. Let V be a vector space over a field F, and let $W \subseteq V$. The subset W is a **subspace** of V if the following three conditions hold.

1. W is closed under $+$.

2. W is closed under scalar multiplication by F.

3. W is a vector space over F. △

Lemma 3.3.3. *Let V be a vector space over a field F, and let $W \subseteq V$ be a subspace.*

1. *The additive identity element of V is in W, and it is the additive identity element of W.*

2. *The additive inverse operation in W is the same as the additive inverse operation in V.*

Proof.

(1). Let $0 \in V$ be the identity element of V, and let $0' \in W$ be the identity element of W. Let $x \in W$. Then $x + 0' = x$. Also, note $x \in V$, so $x + 0 = x$. Hence $x + 0' = x + 0$, and therefore by Lemma 3.2.7 (1), we see that $0' = 0$.

(2). Let $x \in W$. Let $-x$ denote the additive inverse of x in V, and let $\neg x$ denote the additive inverse of x in W. Then $x + (-x) = 0 = x + (\neg x)$, and therefore by Lemma 3.2.7 (1), we see that $-x = \neg x$. □

Lemma 3.3.4. *Let V be a vector space over a field F, and let $W \subseteq V$. Then W is a subspace of V if and only if the following three conditions hold.*

1. *$0 \in W$.*

2. *W is closed under $+$.*

3. *W is closed under scalar multiplication by F.*

Proof. First, suppose that W is a subspace of V. Then $0 \in W$, and hence Property (1) holds. Properties (2) and (3) hold by definition.

Second, suppose that Properties (1), (2) and (3) hold. To show that W is a subspace of V, we need to show that W is a vector space over F. We know that $+$ is associative and commutative with respect to all the elements of V, so it certainly is associative and commutative with respect to the elements of V.

Let $x \in W$. Then $-x = (-1)x$ by Lemma 3.2.7 (9). It follows from Property (3) that $-x \in W$. Hence Parts (1), (2), (3) and (4) of Definition 3.2.1 hold for W. Parts (5), (6), (7) and (8) of that definition immediately hold for W because they hold for V. □

Lemma 3.3.5. *Let V be a vector space over a field F, and let $W \subseteq V$. Then W is a subspace of V if and only if the following three conditions hold.*

1. $W \neq \emptyset$.
2. *W is closed under $+$.*
3. *W is closed under scalar multiplication by F.*

Proof. First, suppose that W is a subspace. Then Properties (1), (2) and (3) hold by Lemma 3.3.4.

Second, suppose that Properties (1), (2) and (3) hold. Because $W \neq \emptyset$, there is some $v \in W$. By Property (3) we know that $(-1)v \in W$. By Lemma 3.2.7 (8) we deduce that $-v \in W$. By Property (2) we deduce that $v + (-v) \in W$, and hence $0 \in W$. We now use Lemma 3.3.4 to deduce that W is a subspace. □

Lemma 3.3.6. *Let V be a vector space over a field F, and and let $U \subseteq W \subseteq V$ be subsets. If U is a subspace of W, and W is a subspace of V, then U is a subspace of V.*

Proof. This proof is straightforward, and we omit the details. □

Lemma 3.3.7. *Let V be a vector space over a field F, and let $\{W_i\}_{i\in I}$ be a family of subspaces of V indexed by I. Then $\bigcap_{i\in I} W_i$ is a subspace of V.*

Proof. Note that $0 \in W_i$ for all $i \in I$ by Lemma 3.3.3. Hence $0 \in \bigcap_{i\in I} W_i$.

Let $x, y \in \bigcap_{i\in I} W_i$ and let $a \in F$. Let $k \in I$. Then $x, y \in W_k$, so $x + y \in W_k$ and $ax \in W_k$. Therefore $x + y \in \bigcap_{i\in I} W_i$ and $ax \in \bigcap_{i\in I} W_i$. Therefore $\bigcap_{i\in I}$ is a subspace of U by Lemma 3.3.4. □

Definition 3.3.8. Let V be a vector space over a field F, and let $S, T \subseteq V$. The **sum** of S and T, denoted $S + T$, is the subset of V defined by

$$S + T = \{s + t \mid s \in S \text{ and } t \in T\}. \qquad \triangle$$

Definition 3.3.9. Let V be a vector space over a field F, and let $X, Y \subseteq V$ be subspaces. The vector space V is the **direct sum** of X and Y, denoted $V = X \oplus Y$, if the following two conditions hold.

1. $X + Y = V$.
2. $X \cap Y = \{0\}$. △

Exercises

Exercise 3.3.1. Let
$$W = \{\begin{bmatrix} x \\ y \\ z \end{bmatrix} \in \mathbb{R}^3 \mid x + y + z = 0\}.$$
Prove that W is a subspace of $\mathbb{R}^3$.

Exercise 3.3.2. Let F be a field, and let S be a non-empty set. Let $\mathcal{F}(S, F)$ be as defined in Exercise 3.2.3. Let $C(S, F)$ be defined by

$$C(S, F) = \{f \in \mathcal{F}(S, F) \mid f(s) = 0 \text{ for all but a finite number of elements } s \in S\}.$$

Prove that $C(S, F)$ is a subspace of $\mathcal{F}(S, F)$.

Exercise 3.3.3. Let V be a vector space over a field F, and let $W \subseteq V$. Prove that W is a subspace of V if and only if the following conditions hold.

1. $W \neq \emptyset$.

2. If $x, y \in W$ and $a \in F$, then $ax + y \in W$.

Exercise 3.3.4. Let V be a vector space over a field F, and let $W \subseteq V$ be a subspace. Let $w_1, \ldots, w_n \in W$ and $a_1, \ldots, a_n \in F$. Prove that $a_1 w_1 + \cdots + a_n w_n \in W$.

Exercise 3.3.5. Let $X, Y \subseteq V$ be subspaces.

(1) Prove that $X \subseteq X + Y$ and $Y \subseteq X + Y$.

(2) Prove that $X + Y$ is a subspace of V.

(3) Prove that if W is a subspace of V such that $X \subseteq W$ and $Y \subseteq W$, then $X + Y \subseteq W$.

3.4 Linear Combinations and Span

Friedberg-Insel-Spence, 4th ed. – Section 1.4

Definition 3.4.1. Let V be a vector space over a field F, and let $S \subseteq V$ be a non-empty subset. Let $v \in V$. The vector v is a **linear combination** of vectors of S if

$$v = a_1 v_1 + a_2 v_2 + \cdots + a_n v_n$$

for some $n \in \mathbb{N}$ and some $v_1, v_2, \ldots, v_n \in S$ and $a_1, a_2, \ldots, a_n \in F$ △

Definition 3.4.2. Let V be a vector space over a field F.

1. Let $S \subseteq V$. Suppose that $S \neq \emptyset$. The **span** of S, denoted $\operatorname{span}(S)$, is the set of all linear combinations of the vectors in S.
2. Let $\operatorname{span}(\emptyset) = \{0\}$. △

Lemma 3.4.3. *Let V be a vector space over a field F, and let $S \subseteq V$ be a non-empty subset.*

1. $S \subseteq \operatorname{span}(S)$.
2. $\operatorname{span}(S)$ *is a subspace of V.*
3. *If $W \subseteq V$ is a subspace and $S \subseteq W$, then* $\operatorname{span}(S) \subseteq W$.
4. $\operatorname{span}(S) = \bigcap\{U \subseteq V \mid U \text{ is a subspace of } V \text{ and } S \subseteq U\}$.

Proof. We prove Parts (1) and (4); the remaining parts of this lemma are left to the reader in Exercise 3.4.2.

(1). Let $x \in S$. Then $x = 1x$ is a linear combination of vectors in S, so $x \in \operatorname{span}(S)$.

(4). Let $H = \bigcap\{U \subseteq V \mid U \text{ is a subspace of } V \text{ and } S \subseteq U\}$. By Parts (2) and (1) of this lemma, we know $\operatorname{span}(S)$ is a subspace of V and that $S \subseteq \operatorname{span}(S)$. We therefore see that $\operatorname{span}(S)$ is one of the subspaces of which H is the intersection. It follows that $H \subseteq \operatorname{span}(S)$.

Let $W \subseteq V$ be a subspace such that $S \subseteq W$. Then by Part (3) of this lemma we know that $\operatorname{span}(S) \subseteq W$. We therefore see that $\operatorname{span}(S)$ is a subset of all the subspaces of which H is the intersection. It follows that $\operatorname{span}(S) \subseteq H$. We conclude that $\operatorname{span}(S) = H$. □

Definition 3.4.4. Let V be a vector space over a field F, and let $S \subseteq V$ be a non-empty subset. The set S **spans** (also **generates**) V if $\operatorname{span}(S) = V$. △

Remark 3.4.5. There is a standard strategy for showing that a set S spans V, as follows.

Proof. Let $v \in V$.

$\vdots$

(argumentation)

$\vdots$

Let $v_1, \ldots, v_n \in S$ and $a_1, \ldots, a_n \in F$ be defined by ...

$\vdots$

(argumentation)

$\vdots$

Then $v = a_1 v_1 + \cdots + a_n v_n$. Hence S spans V. $\square$

In the above strategy, if S is finite, then we can take $v_1, \ldots, v_n$ to be all of S. $\Diamond$

Exercises

Exercise 3.4.1. Using only the definition of spanning, prove that $\{\left[\begin{smallmatrix}1\\2\end{smallmatrix}\right], \left[\begin{smallmatrix}3\\5\end{smallmatrix}\right]\}$ spans $\mathbb{R}^2$.

Exercise 3.4.2. Prove Lemma 3.4.3 (2) and (3).

Exercise 3.4.3. Let V be a vector space over a field F, and let $W \subseteq V$. Prove that W is a subspace of V if and only if $\operatorname{span}(W) = W$.

Exercise 3.4.4. Let V be a vector space over a field F, and let $S \subseteq V$. Prove that $\operatorname{span}(\operatorname{span}(S)) = \operatorname{span}(S)$.

Exercise 3.4.5. Let V be a vector space over a field F, and let $S, T \subseteq V$. Suppose that $S \subseteq T$.

(1) Prove that $\operatorname{span}(S) \subseteq \operatorname{span}(T)$.

(2) Prove that if $\operatorname{span}(S) = V$, then $\operatorname{span}(T) = V$.

Exercise 3.4.6. Let V be a vector space over a field F, and let $S, T \subseteq V$.

(1) Prove that $\operatorname{span}(S \cap T) \subseteq \operatorname{span}(S) \cap \operatorname{span}(T)$.

(2) Give an example of subsets $S, T \subseteq \mathbb{R}^2$ such that S and T are non-empty, not equal to each other, and $\operatorname{span}(S \cap T) = \operatorname{span}(S) \cap \operatorname{span}(T)$. A proof is not needed; it suffices to state what each of S, T, $S \cap T$, $\operatorname{span}(S)$, $\operatorname{span}(T)$, $\operatorname{span}(S \cap T)$ and $\operatorname{span}(S) \cap \operatorname{span}(T)$ are.

(3) Give an example of subsets $S, T \subseteq \mathbb{R}^2$ such that S and T are non-empty, not equal to each other, and $\operatorname{span}(S \cap T) \subsetneqq \operatorname{span}(S) \cap \operatorname{span}(T)$. A proof is not needed; it suffices to state what each of S, T, $S \cap T$, $\operatorname{span}(S)$, $\operatorname{span}(T)$, $\operatorname{span}(S \cap T)$ and $\operatorname{span}(S) \cap \operatorname{span}(T)$ are.

3.5 Linear Independence

Friedberg-Insel-Spence, 4th ed. – Section 1.5

Definition 3.5.1. Let V be a vector space over a field F, and let $S \subseteq V$. The set S is **linearly dependent** if there are $n \in \mathbb{N}$, distinct vectors $v_1, v_2, \dots v_n \in S$, and $a_1, a_2, \dots a_n \in F$ that are not all 0, such that $a_1 v_1 + \cdots + a_n v_n = 0$. △

Lemma 3.5.2. *Let V be a vector space over a field F, and let $S \subseteq V$. If $0 \in S$, then S is linearly dependent.*

Proof. Observe that $1 \cdot 0 = 0$. □

Lemma 3.5.3. *Let V be a vector space over a field F, and let $S \subseteq V$. Suppose that $S \neq \emptyset$ and $S \neq \{0\}$. The following are equivalent.*

a. *S is linearly dependent.*

b. *There is some $v \in S$ such that $v \in \operatorname{span}(S - \{v\})$.*

c. *There is some $v \in S$ such that $\operatorname{span}(S - \{v\}) = \operatorname{span}(S)$.*

Proof. (a) ⇒ (b) Suppose S is linearly dependent. Then there are $n \in \mathbb{N}$, distinct vectors $v_1, \dots, v_n \in S$, and $a_1, \dots, a_n \in F$ not all 0, such that $a_1 v_1 + \cdots + a_n v_n = 0$. Then there is some $k \in \{1, \dots, n\}$ such that $a_k \neq 0$. Therefore

$$v_k = -\frac{a_1}{a_k} v_1 - \cdots - \frac{a_{k-1}}{a_k} v_{k-1} - \frac{a_{k+1}}{a_k} v_{k+1} - \cdots - \frac{a_n}{a_k} v_n.$$

Hence $v_k \in \operatorname{span}(S - \{v_k\})$.

(b) ⇒ (c) Suppose that is some $v \in S$ such that $v \in \operatorname{span}(S - \{v\})$. Then there are $p \in \mathbb{N}$, and $w_1, w_2, \dots w_p \in S - \{v\}$ and $c_1, c_2, \dots c_p \in F$ such that $v = c_1 w_1 + \cdots + c_p w_p$

By Exercise 3.4.5 (1) we know that $\operatorname{span}(S - \{v\}) \subseteq \operatorname{span}(S)$.

Let $x \in \operatorname{span}(S)$. Then there are $m \in \mathbb{N}$, and $u_1, u_2, \dots u_m \in S$ and $b_1, b_2, \dots b_m \in F$ such that $x = b_1 u_1 + \cdots + b_m u_m$. First, suppose that v is not any of $u_1, u_2, \dots u_m$. Then clearly $x \in \operatorname{span}(S - \{v\})$. Second, suppose that v is one of $u_1, u_2, \dots u_m$. Without loss of generality, suppose that $v = u_1$. Then

$$\begin{aligned} x &= b_1(c_1 w_1 + \cdots + c_p w_p) + b_2 u_2 + \cdots + b_m u_m \\ &= b_1 c_1 w_1 + \cdots + b_1 c_p w_p + b_2 u_2 + \cdots + b_m u_m. \end{aligned}$$

Hence $x \in \operatorname{span}(S - \{v\})$. Putting the two cases together, we conclude that $\operatorname{span}(S) \subseteq \operatorname{span}(S - \{v\})$. Therefore $\operatorname{span}(S - \{v\}) = \operatorname{span}(S)$

(c) ⇒ (b) Suppose that there is some $w \in S$ such that $\operatorname{span}(S - \{w\}) = \operatorname{span}(S)$. Because $w \in S$, then $w \in \operatorname{span}(S)$, and hence $w \in \operatorname{span}(S - \{w\})$.

(b) $\Rightarrow$ (a) Suppose that there is some $u \in S$ such that $u \in \operatorname{span}(S - \{u\})$. Hence there are $r \in m$, and $x_1, \ldots, x_r \in S - \{u\}$ and $d_1, \ldots, d_r \in F$ such that $u = d_1x_1 + \cdots + d_rx_r$. Without loss of generality, we can assume that $x_1, \ldots, x_r$ are distinct. Therefore

$$1 \cdot u + (-d_1)x_1 + \cdots + (-d_m)x_m = 0.$$

Because $1 \neq 0$, and because $u, x_1, \ldots, x_r$ are distinct, we deduce that S is linearly dependent. □

Definition 3.5.4. Let V be a vector space over a field F, and let $S \subseteq V$. The set S is **linearly independent** if it is not linearly dependent. △

Remark 3.5.5. There is a standard strategy for showing that a set S in a vector space is linearly independent, as follows.

Proof. Let $v_1, \ldots, v_n \in S$ and $a_1, \ldots, a_n \in F$. Suppose that $v_1, \ldots, v_n$ are distinct, and that $a_1v_1 + \cdots + a_nv_n = 0$.

$\vdots$

(argumentation)

$\vdots$

Then $a_1 = 0, \ldots, a_n = 0$. Hence S is linearly independent. □

In the above strategy, if S is finite, then we simply take $v_1, \ldots, v_n$ to be all of S. ◊

Lemma 3.5.6. *Let V be a vector space over a field F.*

1. *$\emptyset$ is linearly independent.*
2. *If $v \in V$ and $v \neq 0$, then $\{v\}$ is linearly independent.*

Proof.

(1). To prove that a set of vectors S is linearly independent, we need to show that "if $v_1, \ldots, v_n \in S$ are distinct vectors and if $a_1v_1 + \cdots + a_nv_n = 0$ for some $a_1, \ldots, a_n \in F$, then $a_1 = 0, \ldots, a_n = 0$." However, when $S = \emptyset$, then the statement "$v_1, \ldots, v_n \in S$ are distinct vectors" is always false, which means that the logical implication "if $v_1, \ldots, v_n \in S$ are distinct vectors and if $a_1v_1 + \cdots + a_nv_n = 0$ for some $a_1, \ldots, a_n \in F$, then $a_1 = 0, \ldots, a_n = 0$" is always true, using the precise definition of if-then statements. We deduce that $\emptyset$ is linearly independent.

(2). Let $a \in F$. Suppose that $av = 0$. Because $v \neq 0$, we use Lemma 3.2.7 (9) to deduce that $a = 0$. It follows that $\{v\}$ is linearly independent. □

Lemma 3.5.7. *Let V be a vector space over a field F, and let $S_1 \subseteq S_2 \subseteq V$.*

1. *If S_1 is linearly dependent, then S_2 is linearly dependent.*

2. *If S_2 is linearly independent, then S_1 is linearly independent.*

Proof. We prove Part (1); observe that Part (2) is just the contrapositive of Part (1), so Part (2) will automatically hold.

(1). Suppose that S_1 is linearly dependent. Then there are $n \in \mathbb{N}$, distinct vectors $v_1, v_2, \ldots v_n \in S_1$, and $a_1, a_2, \ldots a_n \in F$ that are not all 0, such that $a_1v_1 + \cdots + a_nv_n = 0$. But it is also true that $v_1, v_2, \ldots v_n \in S_2$, which means that S_2 is linearly dependent. □

Lemma 3.5.8. *Let V be a vector space over a field F, let $S \subseteq V$ and let $v \in V - S$. Suppose that S is linearly independent. Then $S \cup \{v\}$ is linearly dependent if and only if $v \in \operatorname{span}(S)$.*

Proof. Suppose that $S \cup \{v\}$ is linearly dependent. Then there are $n \in \mathbb{N}$, and $v_1, v_2, \ldots, v_n \in S \cup \{v\}$ and $a_1, a_2, \ldots, a_n \in F$ not all equal to zero such that $a_1v_1 + \cdots + a_nv_n = 0$. Because S is linearly independent, it must be the case that v is one of the vectors $v_1, v_2, \ldots, v_n$. Without loss of generality, assume $v = v_1$. It must be the case that $a_1 \neq 0$, again because S is linearly independent. Then

$$v = -\frac{a_2}{a_1}v_2 - \cdots - \frac{a_n}{a_1}v_1$$

Because $v_2, \ldots, v_n \in S$, then $v \in \operatorname{span}(S)$.

Suppose that $v \in \operatorname{span}(S)$. Then v is a linear combination of the vectors of S. Thus $S \cup \{v\}$ is linearly independent by Lemma 3.5.3. □

Exercises

Exercise 3.5.1. Using only the definition of linear independence, prove that $\{x^2 + 1, x^2 + 2x, x + 3\}$ is a linearly independent subset of $\mathbb{R}_2[x]$.

Exercise 3.5.2. Let V be a vector space over a field F, and let $u, v \in V$. Suppose that $u \neq v$. Prove that $\{u, v\}$ is linearly dependent if and only if at least one of u or v is a multiple of the other.

Exercise 3.5.3. Let V be a vector space over a field F, and let $u_1, \ldots, u_n \in V$. Prove that the set $\{u_1, \ldots, u_n\}$ is linearly dependent if and only if $u_1 = 0$ or there is some $k \in \{1, \ldots, n-1\}$ such that $u_{k+1} \in \operatorname{span}(\{u_1, \ldots, u_k\})$.

3.6 Bases and Dimension

Friedberg-Insel-Spence, 4th ed. – Section 1.6

Definition 3.6.1. Let V be a vector space over a field F, and let $B \subseteq V$. The set B is a **basis** for V if B is linearly independent and B spans V. △

Theorem 3.6.2. *Let V be a vector space over a field F, and let $B \subseteq V$.*

1. *The set B is a basis for V if and only if every vector in V can be written as a linear combination of vectors in B, where the set of vectors in B with non-zero coefficients in any such linear combination, together with their non-zero coefficients, are unique.*

2. *Suppose that $B = \{u_1, \ldots, u_n\}$ for some $n \in \mathbb{N}$ and $u_1, \ldots, u_n \in V$. Then B is a basis for V if and only if for each vector $v \in V$, there are unique $a_1, \ldots, a_n \in F$ such that $v = a_1u_1 + \cdots + a_nu_n$.*

Proof.

(1). Suppose that B is a basis for V. Then B spans V, and hence every vector in V can be written as a linear combination of vectors in B. Let $v \in V$. Suppose that there are $n, m \in \mathbb{N}$, and $v_1, \ldots, v_n, u_1, \ldots, u_m \in B$ and $a_1, \ldots, a_n, b_1, \ldots, b_m \in F$ such that

$$v = a_1v_1 + a_2v_2 + \cdots + a_nv_n \quad \text{and} \quad v = b_1u_1 + b_2u_2 + \cdots + b_mu_m.$$

Without loss of generality, suppose that $n \geq m$. If might be the case that the sets $\{v_1, \ldots, v_n\}$ and $\{u_1, \ldots, u_m\}$ overlap. By renaming and reordering the vectors in these two sets appropriately, we may assume that $\{v_1, \ldots, v_n\}$ and $\{u_1, \ldots, u_m\}$ are both subsets of a set $\{z_1, \ldots, z_p\}$ for some $p \in \mathbb{N}$ and $z_1, \ldots, z_p \in B$. It will then suffice to show that if

$$v = c_1z_1 + c_2z_2 + \cdots + c_pz_p \quad \text{and} \quad v = d_1z_1 + d_2z_2 + \cdots + d_pz_p \tag{1}$$

for some $c_1, \ldots, c_p, d_1, \ldots, d_p \in F$, then $c_i = d_i$ for all $i \in \{1, \ldots, p\}$.

Suppose that Equation (1) holds. Then

$$(c_1 - d_1)z_1 + \cdots + (c_p - d_p)z_p = 0.$$

Because B is linearly independent, it follows that $c_i - d_i = 0$ for all $i \in \{1, \ldots, p\}$. Because $c_i = d_i$ for all $i \in \{1, \ldots, p\}$, we see in particular that $c_i = 0$ if and only if $d_i = 0$. Hence every vector in V can be written as a linear combination of vectors in B, where the set of vectors in B with non-zero coefficients in any such linear combination, together with their non-zero coefficients, are unique.

Next, suppose that every vector in V can be written as a linear combination of vectors in B, where the set of vectors in B with non-zero coefficients in any such linear combination, together with their non-zero coefficients, are unique. Clearly B spans V. Suppose that there are $n \in \mathbb{N}$, and $v_1, \ldots, v_n \in B$ and $a_1, \ldots, a_n \in F$ such that $a_1v_1 + a_2v_2 + \cdots + a_nv_n = 0$. It is also the case that $0 \cdot v_1 + 0 \cdot v_2 + \cdots + 0 \cdot v_n = 0$. By uniqueness, we deduce that $a_i = 0$ for all $i \in \{1, \ldots, n\}$. Hence B is linearly independent.

(2). This part of the theorem follows from the previous part. □

Lemma 3.6.3. *Let V be a vector space over a field F, and let $S \subseteq V$. The following are equivalent.*

a. *S is a basis for V.*

b. *S is linearly independent, and is contained in no linearly independent subset of V other than itself.*

Proof. Suppose that S is a basis for V. Then S is linearly independent. Suppose that $S \subsetneqq T$ for some linearly independent subset $T \subseteq V$. Let $v \in T - S$. Because S is a basis, then $\operatorname{span}(S) = V$, and hence $v \in \operatorname{span}(S)$. It follows from Lemma 3.5.8 that $S \cup \{v\}$ is linearly dependent. It follows from Lemma 3.5.7 (1) that T is linearly dependent, a contradiction. Hence S is contained in no linearly independent subset of V other than itself.

Suppose that S is linearly independent, and is contained in no linearly independent subset of V other than itself. Let $w \in V$. First, suppose that $w \in S$. Then $w \in \operatorname{span}(S)$ by Lemma 3.4.3 (1). Second, suppose that $w \in V - S$. By the hypothesis on S we see that $S \cup \{w\}$ is linearly dependent. Using Lemma 3.5.8 we deduce that $w \in \operatorname{span}(S)$. Combining the two cases, it follows that $V \subseteq \operatorname{span}(S)$. By definition $\operatorname{span}(S) \subseteq V$. Therefore $\operatorname{span}(S) = V$, and hence S is a basis. □

Theorem 3.6.4. *Let V be a vector space over a field F, and let $S \subseteq V$. Suppose that S is finite. If S spans V, then some subset of S is a basis for V.*

Proof. Suppose that S spans V. If S is linearly independent then S is a basis for V. Now suppose that S is linearly dependent.

Case One: Suppose $S = \{0\}$. Then $V = \operatorname{span}(S) = \{0\}$. This case is trivial because $\emptyset$ is a basis.

Case Two: Suppose S contains at least one non-zero vector. Let $v_1 \in S$ be such that $v_1 \neq 0$. Then $\{v_1\}$ is linearly independent by Lemma 3.5.6. By adding one vector from S at a time, we obtain a linearly independent subset $\{v_1, \ldots, v_n\} \subseteq S$ such that adding any more vectors from set S would render the subset linearly dependent.

Let $B = \{v_1, \ldots, v_n\}$. Because S is finite and $B \subseteq S$, we can write $S = \{v_1, \ldots, v_n, v_{n+1}, \ldots, v_p\}$ for some $p \in \mathbb{Z}$ such that $p \geq n + 1$.

Let $i \in \{n + 1, \ldots, p\}$. Then by the construction of B we know that $B \cup \{v_i\}$ is linearly dependent. It follows from Lemma 3.5.8 implies that $v_i \in \operatorname{span}(B)$.

Let $w \in V - B$. Because S spans V, there are $a_1, \ldots, a_p \in F$ such that $w = a_1v_1 + a_2v_2 + \cdots + a_pv_p$. Because each of $v_{n+1}, \ldots, v_p$ is a linear combination of the elements of B, it follows that w can be written as a linear combination of elements of B. We then use Lemma 3.5.3 (b) to deduce that $B \cup \{w\}$ is linearly dependent. It now follows from Lemma 3.6.3 that B is a basis. □

Theorem 3.6.5 (Replacement Theorem). *Let V be a vector space over a field F, and let $S, L \subseteq V$. Suppose that S and L are finite sets. Suppose that S spans V, and that L is linearly independent.*

1. $|L| \leq |S|$.
2. *There is a subset $H \subseteq S$ such that $|H| = |S| - |L|$, and such that $L \cup H$ spans V.*

Proof. Let $m = |L|$ and $n = |S|$. We will show that this theorem holds by induction on m.

Base Case: Suppose $m = 0$. Then $L = \emptyset$ and $m \leq n$. Let $H = S$. Then H and S have $n - m = n - 0 = n$ elements, and $L \cup H = \emptyset \cup S = S$, and so $L \cup H$ spans V.

Inductive Step: Suppose the result is true for m, and suppose L has $m + 1$ vectors. Suppose $L = \{v_1, \ldots, v_{m+1}\}$. Let $L' = \{v_1, \ldots, v_m\}$. By Lemma 3.5.7 we know that L' is linearly independent. Hence, by the inductive hypothesis, we know that $m \leq n$ and that there is a subset $H' \subseteq S$ such that H' has $n - m$ elements and $L' \cup H'$ spans V. Suppose $H' = \{u_1, \ldots, u_{n-m}\}$. Because $L' \cup H'$ spans V, there are $a_1, \ldots, a_m, b_1, \ldots, b_{m-m} \in F$ such that $v_{m+1} = a_1v_1 + \cdots + a_nv_n + b_1u_1 + \cdots + b_{n-m}u_{n-m}$. Because $v_1, \ldots, v_{m+1}$ is linearly independent, then v_{m+1} is not a linear combination of $\{v_1, \ldots, v_n\}$. Hence $n - m > 0$ and not all $b_1, \ldots, b_{n-m}$ are zero.

Because $n - m > 0$, then $n > m$, and therefore $n \geq m + 1$.

Without loss of generality, assume $b_1 \neq 0$. Then

$$u_1 = \frac{1}{b_1}v_{m+1} - \frac{a_1}{b_1}v_1 - \cdots - \frac{a_m}{b_1}v_m - \frac{b_2}{b_1}u_2 - \cdots - \frac{b_{n-m}}{b_1}u_{n-m}.$$

Let $H = \{u_2, \ldots, u_{n-m}\}$. Clearly H has $n - (m + 1)$ elements. Then

$$L \cup H = \{v_1, \ldots, v_{m+1}, u_2, \ldots, u_{n-m}\}.$$

We claim that $L \cup H$ spans V. Clearly, $v_1, \ldots, v_m, u_2, \ldots, u_{n-m} \in \operatorname{span}(L \cup H)$. Also $u_1 \in \operatorname{span}(L \cup H)$. Hence $L' \cup H' \subseteq \operatorname{span}(L \cup H)$. We know that $\operatorname{span}(L' \cup H') = V$, and hence by Exercise 3.4.5 (2) we see that $\operatorname{span}(\operatorname{span}(L \cup H)) = V$. It follows from Exercise 3.4.4 that $\operatorname{span}(L \cup H) = V$. □

Corollary 3.6.6. *Let V be a vector space over a field F. Suppose that V has a finite basis. Then all bases of V are finite, and all bases have the same number of vectors.*

Proof. Let B be a finite basis for V. Let $n = |B|$. Let K be some other basis for V. Suppose that K has more elements than B. Then K has at least $n + 1$ elements (it could be that K is infinite). In particular, let C be a subset of K that has precisely $n + 1$ elements. Then C is linearly independent by Lemma 3.5.7. Because B spans V, then by Theorem 3.6.5 (1) we deduce that $n + 1 \leq n$, which is a contradiction.

Next, suppose that K has fewer elements than B. Then K is finite. Let $m = |K|$. Then $m < n$. Because K spans V and B is linearly independent, then by Theorem 3.6.5 (1) we deduce that $n \leq m$, which is a contradiction.

We conclude that K has the same number of vectors as B. □

Definition 3.6.7. Let V be a vector space over a field F.

1. The vector space V is **finite-dimensional** if V has a finite basis.

2. The vector space V is **infinite-dimensional** if V does not have a finite basis.

3. If V is finite-dimensional, the **dimension** of V, denoted $\dim(V)$, is the number of elements in any basis. △

Lemma 3.6.8. *Let V be a vector space over a field F. Then $\dim(V) = 0$ if and only if $V = \{0\}$.*

Proof. By Lemma 3.5.6 (1) we know that $\emptyset$ is linearly independent. Using Definition 3.4.2 we see that $\dim(V) = 0$ if and only if $\emptyset$ is a basis for V if and only if $V = \text{span}(\emptyset)$ if and only if $V = \{0\}$. □

Corollary 3.6.9. *Let V be a vector space over a field F, and let $S \subseteq V$. Suppose that V is finite-dimensional. Suppose that S is finite.*

1. *If S spans V, then $|S| \geq \dim(V)$.*

2. *If S spans V and $|S| = \dim(V)$, then S is a basis for V.*

3. *If S is linearly independent, then $|S| \leq \dim(V)$.*

4. *If S is linearly independent and $|S| = \dim(V)$, then S is a basis for V.*

5. *If S is linearly independent, then it can be extended to a basis for V.*

Proof. We prove Parts (1) and (5), leaving the rest to the reader in Exercise 3.6.2.

Let $n = \dim(V)$.

(1). Suppose that S spans V. By Theorem 3.6.4 we know that there is some $H \subseteq S$ such that H is a basis for V. Corollary 3.6.6 implies that $|H| = n$. It follows that $|S| \geq n$.

(5). Suppose that S is linearly independent. Let B be a basis for V. Then $|B| = n$. Because B is a basis for V, then B spans V. By the Replacement Theorem (Theorem 3.6.5) there is a subset $K \subseteq B$ such that $|K| = |B| - |S|$, and such that $S \cup K$ spans V. Note that $|S \cup K| = |B| = n$. It follows from Part (2) of this corollary that $S \cup K$ is a basis. Therefore S can be extended to a basis. □

Theorem 3.6.10. *Let V be vector space over a field F, and let $W \subseteq V$ be a subspace. Suppose that V is finite-dimensional.*

1. *W is finite-dimensional.*

2. $\dim(W) \leq \dim(V)$.

3. *If* $\dim(W) = \dim(V)$, *then* $W = V$.

4. *Any basis for W can be extended to a basis for V.*

Proof. Let $n = \dim(V)$. We prove all four parts of the theorem together.

Case One: Suppose $W = \{0\}$. Then all four parts of the theorem hold.

Case Two: Suppose $W \neq \{0\}$. Then there is some $x_1 \in W$ such that $x_1 \neq 0$. Note that $\{x_1\}$ is linearly independent. It might be the case that there is some $x_2 \in W$ such that $\{x_1, x_2\}$ is linearly independent. Keep going, adding one vector at a time while maintaining linear independence. Because $W \subseteq V$, then there are at most n linearly independent vectors in W by Corollary 3.6.9 (3). Hence we can keep adding vectors until we get $\{x_1, \ldots, x_k\} \in W$ for some $k \in \mathbb{N}$ such that $k \leq n$, where adding any other vector in V would render the set linearly dependent. Hence, adding any vector in W would render it linearly dependent. By Lemma 3.6.3 we see that $\{x_1, \ldots, x_k\}$ is a basis for W. Therefore W is finite-dimensional and $\dim(W) \leq \dim(V)$.

Now suppose $\dim(W) = \dim(V)$. Then $k = n$ and $\{x_1, \ldots, x_n\}$ is a linearly independent set in V with n elements. By Corollary 3.6.9 (4), we know that $\{x_1, \ldots, x_n\}$ is a basis for V. Then $W = \operatorname{span}(\{x_1, \ldots, x_n\}) = V$.

From Corollary 3.6.9 (5) we deduce that any basis for W, which is a linearly independent set in V, can be extended to a basis for V. □

Exercises

Exercise 3.6.1. Let
$$W = \{\begin{bmatrix} x \\ y \\ z \end{bmatrix} \in \mathbb{R}^3 \mid x + y + z = 0\}.$$
It was proved in Exercise 3.3.1 that W is a subspace of $\mathbb{R}^3$. What is $\dim(W)$? Prove your answer.

Exercise 3.6.2. Prove Corollary 3.6.9 (2), (3) and (4).

Exercise 3.6.3. Let V be a vector space over a field F, and let $S, T \subseteq V$. Suppose that $S \cup T$ is a basis for V, and that $S \cap T = \emptyset$. Prove that $\operatorname{span}(S) \oplus \operatorname{span}(T) = V$. (See Definition 3.3.9 for the definition of $\operatorname{span}(S) \oplus \operatorname{span}(T)$.)

Exercise 3.6.4. Let V be a vector space over a field F, and let $X, Y \subseteq V$ be subspaces. Suppose that X and Y are finite-dimensional. Find necessary and sufficient conditions on X and Y so that $\dim(X \cap Y) = \dim(X)$.

Exercise 3.6.5. Let V, W be vector spaces over a field F. Suppose that V and W are finite-dimensional. Let $V \times W$ be the product vector space, as defined in Exercise 3.2.2. Express $\dim(V \times W)$ in terms of $\dim(V)$ and $\dim(W)$. Prove your answer.

Exercise 3.6.6. Let V be a vector space over a field F, and let $L \subseteq S \subseteq V$. Suppose that S spans V. Prove that the following are equivalent.

a. L is a basis for V.

b. L is linearly independent, and is contained in no linearly independent subset of S other than itself.

3.7 Bases for Arbitrary Vector Spaces

Friedberg-Insel-Spence, 4th ed. – Section 1.7

Definition 3.7.1. Let $\mathcal{P}$ be a non-empty family of sets, and let $M \in \mathcal{P}$. The set M is a **maximal element** of $\mathcal{P}$ if there is no $Q \in \mathcal{P}$ such that $M \subsetneq Q$. △

Lemma 3.7.2. *Let V be a vector space over a field F. Let $\mathcal{B}$ be the family of all linearly independent subsets of V. Let $S \in \mathcal{B}$. Then S is a basis for V if and only if S is a maximal element of $\mathcal{B}$.*

Proof. This lemma follows immediately from Lemma 3.6.3. □

Definition 3.7.3. Let $\mathcal{P}$ be a non-empty family of sets, and let $\mathcal{C} \subseteq \mathcal{P}$. The family $\mathcal{C}$ is a **chain** if $A, B \in \mathcal{C}$ implies $A \subseteq B$ or $A \supseteq B$. △

Theorem 3.7.4 (Zorn's Lemma). *Let $\mathcal{P}$ be a non-empty family of sets. Suppose that for each chain $\mathcal{C}$ in $\mathcal{P}$, the set $\bigcup_{C \in \mathcal{C}} C$ is in $\mathcal{P}$. Then $\mathcal{P}$ has a maximal element.*

Theorem 3.7.5. *Let V be a vector space over a field F. Then V has a basis.*

Proof. Let $\mathcal{B}$ be the family of all linearly independent subsets of V. We will show that $\mathcal{B}$ has a maximal element by using Zorn's Lemma (Theorem 3.7.4). The maximal element of $\mathcal{B}$ will be a basis for V by Lemma 3.7.2.

Because $\emptyset$ is a linearly independent subset of V, as stated in Lemma 3.5.6 (1), we see that $\emptyset \in \mathcal{B}$, and hence $\mathcal{B}$ is non-empty.

Let $\mathcal{C}$ be a chain in $\mathcal{B}$. Let $U = \bigcup_{C \in \mathcal{C}} C$. We need to show that $U \in \mathcal{B}$. That is, we need to show that U is linearly independent. Let $v_1, \ldots, v_n \in U$ and suppose $a_1 v_1 + \cdots + a_n v_n = 0$ for some $a_1, \ldots, a_n \in F$. By the definition of union, we know that for each $i \in \{1, \ldots, n\}$, there is some $C_i \in \mathcal{C}$ such that $v_i \in C_i$. Because $\mathcal{C}$ is a chain, we know that for any two of $C_1, \ldots, C_n$, one contains the other. Hence we can find $k \in \{1, \ldots, n\}$ such that $C_i \subseteq C_k$ for all $i \in \{1, \ldots, n\}$. Hence $v_1, \ldots, v_n \in C_k$. Because $C_k \in \mathcal{C} \subseteq \mathcal{B}$, then C_k is linearly independent, and so $a_1 v_1 + \cdots + a_n v_n = 0$ implies $a_i = 0$ for all $i \in \{1, \ldots, n\}$. Hence U is linearly independent, and therefore $U \in \mathcal{B}$.

We have now seen that $\mathcal{B}$ satisfies the hypotheses of Zorn's Lemma, and by that lemma we deduce that $\mathcal{B}$ has a maximal element. □

Exercises

Exercise 3.7.1. Let V be a vector space over a field F, and let $S \subseteq V$. Prove that if S spans V, then some subset of S is a basis for V.

4 Linear Maps

4.1 Linear Maps

Friedberg-Insel-Spence, 4th ed. – Section 2.1

Definition 4.1.1. Let V, W be vector spaces over a field F. Let $f : V \to W$ be a function. The function f is a **linear map** (also called **linear transformation** or **vector space homomorphism**) if the following two conditions hold. Let $x, y \in V$ and $c \in F$.

1. $f(x + y) = f(x) + f(y)$
2. $f(cx) = cf(x)$ △

Lemma 4.1.2. *Let V, W be vector spaces over a field F, and let $f : V \to W$ be a linear map.*

1. *$f(0) = 0$.*
2. *If $x \in V$, then $f(-x) = -f(x)$.*

Proof. We will prove Part (2), leaving the other part to the reader in Exercise 4.1.1.

(2). Let $x \in V$. Then $f(x) + f(-x) = f(x + (-x)) = f(0) = 0$, where the last equality uses Part (1) of this lemma, and the other two equalities use the fact that f is a linear map and that V is a vector space. By Lemma 3.2.7 (3), it follows that $f(-x) = -f(x)$. □

Lemma 4.1.3. *Let V, W be vector spaces over a field F, and let $f : V \to W$ be a function. The following are equivalent.*

a. *f is a linear map.*

b. *$f(cx + y) = cf(x) + f(y)$ for all $x, y \in V$ and $c \in F$.*

c. *$f(a_1x_1 + \cdots + a_nx_n) = a_1f(x_1) + \cdots + a_nf(x_n)$ for all $x_1, \ldots x_n \in V$ and $a_1, \ldots a_n \in F$.*

Proof. Left to the reader in Exercise 4.1.2. □

Lemma 4.1.4. *Let V, W, Z be vector spaces over a field F, and let $f : V \to W$ and $g : W \to Z$ be linear maps.*

1. *The identity map $1_V : V \to V$ is a linear map.*
2. *The function $g \circ f$ is a linear map.*

Proof.

(1). This part is straightforward.

(2). Let $x, y \in V$ and $c \in F$. Then

$$\begin{aligned}(g \circ f)(x + y) &= g(f(x + y)) = g(f(x) + f(y)) = g(f(x)) + g(f(y))\\ &= (g \circ f)(x) + (g \circ f)(y)\end{aligned}$$

and

$$(g \circ f)(cx) = g(f(cx)) = g(c(f(x))) = c(g(f(x))) = c(g \circ f)(x).$$

Hence $(g \circ f)$ is a linear map. □

Lemma 4.1.5. *Let V, W be vector spaces over a field F, and let $f: V \to W$ be a linear map.*

1. *If A is a subspace of V, then $f(A)$ is a subspace of W.*
2. *If B is a subspace of W, then $f^{-1}(B)$ is a subspace of V.*

Proof. We will prove Part (1), leaving the other part to the reader in Exercise 4.1.3.

(1). Let A be a subspace of V. By Lemma 3.3.4 (1) we know that $0 \in W$, and by Lemma 4.1.2 (1) we know that $0 = f(0) \in f(A)$.

Let $x, y \in f(A)$. Then there are $a, b \in A$ such that $x = f(a)$ and $y = f(b)$. Hence $x + y = f(a) + f(b) = f(a + b)$, because f is a linear map. Because A is a subspace of V we know that $a + b \in A$, and hence $x + y \in f(A)$. It follows that $f(A)$ is closed under $+$.

Let $s \in F$. Because f is a linear map, we see that $sx = sf(a) = f(sa)$. Because A is a subspace of V we know that $sa \in A$, and hence $sx \in f(A)$. It follows that $f(A)$ is closed under scalar multiplication by F.

We now use Lemma 3.3.4 to deduce that $f(A)$ is a subspace of V. □

Theorem 4.1.6. *Let V, W be vector spaces over a field F.*

1. *Let B be a basis for V. Let $g: B \to W$ be a function. Then there is a unique linear map $f: V \to W$ such that $f|_B = g$.*
2. *Let $\{v_1, \ldots, v_n\}$ be a basis for V, and let $w_1, \ldots, w_n \in W$. Then there is a unique linear map $f: V \to W$ such that $f(v_i) = w_i$ for all $i \in \{1, \ldots, n\}$.*

Proof. We prove Part (1); Part (2) follows immediately from Part (1).

Let $v \in V$. Then by Theorem 3.6.2 (1) we know that v can be written as $v = a_1x_1 + \cdots + a_nx_n$ for some $x_1, \ldots, x_n \in B$ and $a_1, \ldots a_n \in F$, where the set of vectors with non-zero coefficients, together with their non-zero coefficients, are unique. Then define $f(v) = a_1g(x_1) + \cdots + a_ng(x_n)$. If v is written in two different ways as linear combinations of elements of B, then the uniqueness of the vectors in B with non-zero coefficients, together with their non-zero coefficients, implies that $f(v)$ is well-defined.

Observe that if $v \in B$, then $v = 1 \cdot v$ is the unique way of expressing v as a linear combination of vectors in B, and therefore $f(v) = 1 \cdot g(v) = g(v)$. Hence $f|_B = g$.

Let $v, w, \in V$ and let $c \in F$. Then we can write $v = a_1x_1 + \cdots + a_nx_n$ and $w = b_1x_1 + \cdots + b_nx_n$ where $x_1, \ldots, x_n \in B$ and $a_1, \ldots, a_n, b_1, \ldots, b_n \in F$. Then $v + w = \sum_{i=1}^{n}(a_i + b_i)x_i$, and hence

$$f(v+w) = \sum_{i=1}^{n}(a_i + b_i)g(x_i) = \sum_{i=1}^{n} a_i g(x_i) + \sum_{i=1}^{n} b_i g(x_i) = f(v) + f(w).$$

A similar proof shows that $f(cv) = cf(v)$. Hence f is linear map.

Let $h \colon V \to W$ be a linear map such that $h|_B = g$. Let $v \in V$. Then $v = a_1x_1 + \cdots + a_nx_n$ for some $x_1, \ldots, x_n \in B$ and $a_1, \ldots a_n \in F$. Hence

$$h(v) = h(\sum_{i=1}^{n} a_i x_i) = \sum_{i=1}^{n} a_i h(x_i) = \sum_{i=1}^{n} a_i g(x_i) = f(v)$$

Therefore $h = f$. It follows that f is unique. □

Corollary 4.1.7. *Let V, W be vector spaces over a field F, and let $f, g \colon V \to W$ be linear maps. Let B be a basis for V. Suppose that $f(v) = g(v)$ for all $v \in B$. Then $f = g$.*

Proof. This corollary is an immediate consequence of Theorem 4.1.6, and we omit the details. □

Exercises

Exercise 4.1.1. Prove Lemma 4.1.2 (1).

Exercise 4.1.2. Prove Lemma 4.1.3.

Exercise 4.1.3. Prove Lemma 4.1.5 (2).

Exercise 4.1.4. Prove that there exists a linear map $f \colon \mathbb{R}^2 \to \mathbb{R}^3$ such that $f(\begin{bmatrix} 1 \\ 1 \end{bmatrix}) = \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}$ and $f(\begin{bmatrix} 2 \\ 3 \end{bmatrix}) = \begin{bmatrix} 1 \\ -1 \\ 4 \end{bmatrix}$. What is $f(\begin{bmatrix} 8 \\ 11 \end{bmatrix})$?

Exercise 4.1.5. Does there exist a linear map $g \colon \mathbb{R}^3 \to \mathbb{R}^2$ such that $g(\begin{bmatrix} 1 \\ 0 \\ 3 \end{bmatrix}) = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $g(\begin{bmatrix} -2 \\ 0 \\ -6 \end{bmatrix}) = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$? Explain why or why not.

4.2 Kernel and Image

Friedberg-Insel-Spence, 4th ed. – Section 2.1

Definition 4.2.1. Let V, W be vector spaces over a field F, and let $f: V \to W$ be a linear map.

1. The **kernel** (also called the **null space**) of f, denoted $\ker f$, is the set $\ker f = f^{-1}(\{0\})$.
2. The **image** of f, denoted $\operatorname{im} f$, is the set $\operatorname{im} f = f(V)$. △

Remark 4.2.2. Observe that

$$\ker f = \{v \in V \mid f(v) = 0\}$$

and

$$\operatorname{im} f = \{w \in W \mid w = f(v) \text{ for some } v \in V\}. \qquad \Diamond$$

Lemma 4.2.3. *Let V, W be vector spaces over a field F, and let $f: V \to W$ be a linear map.*

1. $\ker f$ *is a subspace of* V.
2. $\operatorname{im} f$ *is a subspace of* W.

Proof. This lemma follows immediately from Lemma 4.1.5. □

Lemma 4.2.4. *Let V, W be vector spaces over a field F, and let $f: V \to W$ be a linear map. Then f is injective if and only if* $\ker f = \{0\}$.

Proof. Suppose that f is injective. Because $f(0) = 0$ by Theorem 4.1.2 (1), it follows from the injectivity of f that $\ker f = f^{-1}(\{0\}) = \{0\}$.

Now suppose that $\ker f = \{0\}$. Let $v, w \in W$, and suppose that $f(v) = f(w)$. By Theorem 4.1.2 (2) and the definition of homomorphisms we see that

$$f(v + (-w)) = f(v) + f(-w) = f(v) + (-f(w)) = 0.$$

It follows that $v+(-w) \in f^{-1}(\{0\}) = \ker f$. Because $\ker f = \{0\}$, we deduce that $v+(-w) = 0$. Hence $v = w$. Hence f is injective. □

Lemma 4.2.5. *Let V, W be vector spaces over a field F, and let $f: V \to W$ be a linear map. Let $w \in W$. If $a \in f^{-1}(\{w\})$, then $f^{-1}(\{w\}) = a + \ker f$.*

Proof. Suppose that $a \in f^{-1}(\{w\})$. Then $f(a) = w$.

Let $y \in f^{-1}(\{w\})$. Then $f(y) = w$. Then $f(y + (-a)) = f(y) + f(-a) = f(y) + (-f(a)) = w + (-w) = 0$. Hence $y + (-a) \in \ker f$. Then there is some $q \in \ker f$ such that $y + (-a) = q$. Therefore $y = a + q \in a + \ker f$.

Let $x \in a + \ker f$. Then there is some $p \in \ker f$ such that $x = a + p$. Then $f(p) = 0$, and hence $f(x) = f(a + p) = f(a) + f(p) = w + 0 = w$. Therefore $x \in f^{-1}(\{w\})$, □

Lemma 4.2.6. *Let V, W be vector spaces over a field F, let $f: V \to W$ be a linear map and let B be a basis for V. Then* $\operatorname{im} f = \operatorname{span}(f(B))$.

Proof. Clearly $f(B) \subseteq \operatorname{im} f$. By Lemma 4.2.3 (2) and Lemma 3.4.3 (3), we deduce that $\operatorname{span}(f(B)) \subseteq \operatorname{im} f$.

Let $y \in \operatorname{im} f$. Then $y = f(v)$ for some $v \in V$. Then $v = a_1 v_1 + \cdots + a_n v_n$ for some $v_1, \ldots, v_n \in B$ and $a_1, \ldots, a_n \in F$. Then

$$y = f(v) = f(a_1 v_1 + \cdots + a_n v_n) = a_1 f(v_1) + \cdots + a_n f(v_n) \in \operatorname{span}(f(B)).$$

Therefore $\operatorname{im} f \subseteq \operatorname{span}(B)$, and hence $\operatorname{im} f = \operatorname{span}(f(B))$. □

Lemma 4.2.7. *Let V, W be vector spaces over a field F, and let $f: V \to W$ be a linear map. Suppose that V is finite-dimensional. Then* $\ker f$ *and* $\operatorname{im} f$ *are finite-dimensional.*

Proof. By Lemma 4.2.3 (1) we know that $\ker f$ is a subspace of V, and hence $\ker f$ is finite-dimensional by Theorem 3.6.10 (1).

Let B be a basis for V. By Corollary 3.6.6 we know that B is finite. Hence $f(B)$ is finite. By Lemma 4.2.6 we see that $\operatorname{im} f = \operatorname{span}(f(B))$. It follows from Theorem 3.6.4 that a subset of $f(B)$ is a basis for $\operatorname{im} f$, which implies that $\operatorname{im} f$ is finite-dimensional. □

Exercises

Exercise 4.2.1. Let $h: \mathbb{R}^3 \to \mathbb{R}^2$ be defined by $h\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) = \begin{bmatrix} x-y-z \\ 2x+y+3z \end{bmatrix}$ for all $\begin{bmatrix} x \\ y \\ z \end{bmatrix} \in \mathbb{R}^3$. Find $\ker h$.

Exercise 4.2.2. Let $G: \mathbb{R}_2[x] \to \mathbb{R}_2[x]$ be defined by $D(ax^2 + bx + c) = ax^2 + (a + 2b + c)x + (3a - 2b - c)$ for all $ax^2 + bx + c \in \mathbb{R}_2[x]$. Find $\ker G$.

Exercise 4.2.3. Let V, W be vector spaces over a field F, and let $f: V \to W$ be a linear map. Let $w_1, \ldots, w_k \in \operatorname{im} f$ be linearly independent vectors. Let $v_1, \ldots, v_k \in V$ be vectors such that $f(v_i) = w_i$ for all $i \in \{1, \ldots, k\}$. Prove that $v_1, \ldots, v_k$ are linearly independent.

Exercise 4.2.4. Let V, W be vector spaces over a field F, and let $f: V \to W$ be a linear map.

(1) Prove that f is injective if and only if for every linearly independent subset $S \subseteq V$, the set $f(S)$ is linearly independent.

(2) Supppose that f is injective. Let $T \subseteq V$. Prove that T is linearly independent if and only if $f(T)$ is linearly independent.

(3) Supppose that f is bijective. Let $B \subseteq V$. Prove that B is a basis for V if and only if $f(B)$ is a basis for W.

Exercise 4.2.5. Find an example of two linear maps $f, g: \mathbb{R}^2 \to \mathbb{R}^2$ such that $\ker f = \ker g$ and $\operatorname{im} f = \operatorname{im} g$, and none of these kernels and images is the trivial vector space, and $f \neq g$.

4.3 Rank-Nullity Theorem

Friedberg-Insel-Spence, 4th ed. – Section 2.1

Definition 4.3.1. Let V, W be vector spaces over a field F, and let $f\colon V \to W$ be a linear map.

1. If $\ker f$ is finite-dimensional, the **nullity** of f, denoted $\mathrm{nullity}(f)$, is defined by $\mathrm{nullity}(f) = \dim(\ker f)$.

2. If $\operatorname{im} f$ is finite-dimensional, the **rank** of f, denoted $\mathrm{rank}(f)$, is defined by $\mathrm{rank}(f) = \dim(\operatorname{im} f)$. △

Theorem 4.3.2 (**Rank-Nullity Theorem**). *Let V, W be vector spaces over a field F, and let $f\colon V \to W$ be a linear map. Suppose that V is finite-dimensional. Then*

$$\mathrm{nullity}(f) + \mathrm{rank}(f) = \dim(V).$$

Proof. Let $n = \dim(V)$. By Lemma 4.2.3 (1) we know that $\ker f$ is a subspace of V, and hence $\ker f$ is finite-dimensional by Theorem 3.6.10 (1), and $\mathrm{nullity}(f) = \dim(\ker f) \leq \dim(V)$ by Theorem 3.6.10 (2). Let $k = \mathrm{nullity}(f)$. Then $k \leq n$. Let $\{v_1, \ldots, v_k\}$ be a basis for $\ker f$. By Theorem 3.6.10 (4) $\{v_1, \ldots, v_k\}$ can be extended to a basis $\{v_1, \ldots, v_n\}$ for V. We will show that $\{f(v_{k+1}), \ldots, f(v_n)\}$ is a basis for $\operatorname{im} f$. It will then follow that the $\mathrm{rank}(f) = n - k$, which will prove the theorem.

By Lemma 4.2.6 we know that $\operatorname{im} f = \mathrm{span}(\{f(v_1), \ldots, f(v_n)\})$. Note that $v_1, \ldots, v_k \in \ker f$, and therefore $f(v_1) = \cdots = f(v_k) = 0$. It follows that $\operatorname{im} f = \mathrm{span}(\{f(v_{k+1}), \ldots, f(v_n)\})$.

Suppose $b_{k+1} f(v_{k+1}) + \cdots + b_n f(v_n) = 0$ for some $b_{k+1}, \ldots, b_n \in F$. Hence $f(b_{k+1}v_{k+1} + \cdots + b_n v_n) = 0$. Therefore $b_{k+1}v_{k+1} + \cdots + b_n v_n \in \ker f$. Because $\{v_1, \ldots, v_n\}$ is a basis for $\ker f$, then $b_{k+1}v_{k+1} + \cdots + b_n v_n = b_1 v_1 + \cdots + b_k v_k$ for some $b_1, \ldots, b_k \in F$. Then $b_1 v_1 + \cdots + b_k v_k + (-b_{k+1})v_{k+1} + \cdots + (-b_n)v_n = 0$. Because $\{v_1, \ldots, v_n\}$ is a basis for V, then $b_1 = \cdots = b_n = 0$. Therefore $f(v_{k+1}), \ldots, f(v_n)$ are linearly independent. □

Corollary 4.3.3. *Let V, W be vector spaces over a field F, and let $f\colon V \to W$ be a linear map. Suppose that V is finite-dimensional. Then* $\mathrm{rank}(f) \leq \dim(V)$.

Proof. This corollary is an immediate consequence of Rank-Nullity Theorem (Theorem 4.3.2), and we omit the details. □

Corollary 4.3.4. *Let V, W be vector spaces over a field F, and let $f\colon V \to W$ be a linear map. Suppose that V and W are finite-dimensional, and that* $\dim(V) = \dim(W)$. *The following are equivalent.*

a. *f is injective.*

b. *f is surjective*

c. *f is bijective.*

d. $\mathrm{rank}(f) = \dim(V)$.

Proof. Clearly (c) $\Rightarrow$ (a), and (c) $\Rightarrow$ (b). We will show below that (a) $\Leftrightarrow$ (d), and (b) $\Leftrightarrow$ (d). It will then follow that (a) $\Leftrightarrow$ (b), and from that we will deduce that (a) $\Rightarrow$ (c), and (b) $\Rightarrow$ (c).

(a) $\Leftrightarrow$ (d) By Lemma 4.2.4 we know that f is injective if and only if $\ker f = \{0\}$. By Lemma 3.6.8 we deduce that f is injective if and only if $\dim(\ker f) = 0$, and by definition that is true if and only if $\mathrm{nullity}(f) = 0$. By The Rank-Nullity Theorem (Theorem 4.3.2), we know that $\mathrm{nullity}(f) = \dim(V) - \mathrm{rank}(f)$. It follows that f is injective if and only if $\dim(V) - \mathrm{rank}(f) = 0$, which is the same as $\mathrm{rank}(f) = \dim(V)$.

(b) $\Leftrightarrow$ (d) By definition f is surjective if and only if $\mathrm{im}\, f = W$. By Lemma 4.2.3 (2) we know that $\mathrm{im}\, f$ is a subspace of W. If $\mathrm{im}\, f = W$ then clearly $\dim(\mathrm{im}\, f) = \dim(W)$; by Theorem 3.6.10 (3) we know that if $\dim(\mathrm{im}\, f) = \dim(W)$ then $\mathrm{im}\, f = W$. Hence f is surjective if and only if $\dim(\mathrm{im}\, f) = \dim(W)$, and by definition that is true if and only if $\mathrm{rank}(f) = \dim(W)$. By hypothesis $\dim(W) = \dim(V)$, and therefore f is surjective if and only if $\mathrm{rank}(f) = \dim(V)$. $\square$

Corollary 4.3.5. *Let V, W, Z be vector spaces over a field F, and let $f\colon V \to W$ and $g\colon W \to Z$ be linear maps. Suppose that V and W are finite-dimensional.*

1. $\mathrm{rank}(g \circ f) \le \mathrm{rank}(g)$.

2. $\mathrm{rank}(g \circ f) \le \mathrm{rank}(f)$.

Proof.

(1). Observe that $\mathrm{im}(g \circ f) = (g \circ f)(V) = g(f(V)) \subseteq g(W) = \mathrm{im}\, g$. By Lemma 4.2.3 (2) we know that $\mathrm{im}(g \circ f)$ and $\mathrm{im}\, g$ are subspaces of W. It is straightforward to see that $\mathrm{im}(g \circ f)$ is a subspace of $\mathrm{im}\, g$. It follows from Theorem 3.6.10 (2) that $\mathrm{rank}(g \circ f) = \dim(\mathrm{im}(g \circ f)) \le \dim(\mathrm{im}\, g) = \mathrm{rank}(g)$.

(2). By Corollary 4.3.3 we see that $\mathrm{rank}(g \circ f) = \dim(\mathrm{im}(g \circ f)) = \dim((g \circ f)(V)) = \dim(g(f(V))) = \dim(g|_{f(V)}(f(V))) = \mathrm{rank}(g|_{f(V)}) \le \dim(f(V)) = \dim(\mathrm{im}\, f) = \mathrm{rank}(f)$. $\square$

Exercises

Exercise 4.3.1. Let V, W be vector spaces over a field F, and let $f\colon V \to W$ be a linear map. Suppose that V and W are finite-dimensional.

(1) Prove that if $\dim(V) < \dim(W)$, then f cannot be surjective.

(2) Prove that if $\dim(V) > \dim(W)$, then f cannot be injective.

4.4 Isomorphisms

Friedberg-Insel-Spence, 4th ed. – Section 2.4

Definition 4.4.1. Let V and W be a vector space over a field F and let $f\colon V \to W$ be a function. The function f is an **isomorphism** if f is bijective and is a linear map. △

Definition 4.4.2. Let V, W be a vector space over a field F. The vector spaces V and W are **isomorphic** if there is an isomorphism $V \to W$. △

Lemma 4.4.3. *Let V, W and Z be vector spaces over a field F, and let $f\colon V \to W$ be and $g\colon W \to Z$ be isomorphisms.*

1. *The identity map $1_V\colon V \to V$ is an isomorphism.*
2. *The function f^{-1} is an isomorphism.*
3. *The function $g \circ f$ is an isomorphism.*

Proof. We prove Part (2); the remaining parts of this lemma follow immediately from Lemma 4.1.4 together with basic facts about bijective functions, and we omit the details.

(2). Using basic facts about bijective functions, we know that f^{-1} is bijective.

Let $x, y \in V$ and $c \in F$. Let $a = f^{-1}(x)$ and $b = f^{-1}(y)$. Then $f(a) = x$ and $f(b) = y$. Then

$$\begin{aligned} f^{-1}(x+y) &= f^{-1}(f(a)+f(b)) = f^{-1}(f(a+b)) \\ &= (f^{-1}\circ f)(a+b) = a+b = f^{-1}(x)+f^{-1}(y) \end{aligned}$$

and

$$f^{-1}(cx) = f^{-1}(cf(a)) = f^{-1}(f(ca)) = (f^{-1}\circ f)(ca) = ca = cf^{-1}(x).$$

Hence f^{-1} is a linear map. □

Corollary 4.4.4. *Let V, W be vector spaces over a field F, and let $f\colon V \to W$ be a linear map. Suppose that V and W are finite-dimensional, and that $\dim(V) = \dim(W)$. The following are equivalent.*

a. *f is injective.*
b. *f is surjective*
c. *f is an isomorphism.*

d. $\operatorname{rank}(f) = \dim(V)$.

Corollary 4.4.5. *Let V, W be vector spaces over a field F, and let $f : V \to W$ be a linear map. Suppose that V and W are finite-dimensional, and that* $\dim(V) = \dim(W)$.

1. *If a function $g : W \to V$ is a right inverse of f, then f is bijective and $g = f^{-1}$.*

2. *If a function $g : W \to V$ is a left inverse of f, then f is bijective and $g = f^{-1}$.*

Proof. This result follows immediately from Corollary 4.4.4, together with the fact, seen in Proofs and Fundamentals, that if a function has both a left inverse and a right inverse, then these two one-sided inverses are equal, and it is a full inverse. □

Lemma 4.4.6. *Let V, W be vector spaces over a field F, and let $f : V \to W$ be a linear map. Let B be a basis for V. Then f is an isomorphism if and only if $f(B)$ is a basis for W.*

Proof. Suppose that f is an isomorphism. Let $v_1, v_2, \ldots v_n \in f(B)$ and $a_1, a_2, \ldots a_n \in F$, and suppose that $v_1, \ldots, v_n$ are distinct, and that $a_1 v_1 + \cdots + a_n v_n = 0$. There are $w_1, \ldots, w_n \in B$ such that $f(w_i) = v_i$ for all $i \in \{1, \ldots, n\}$. Clearly $w_1, \ldots, w_n$ are distinct. Then $a_1 f(w_1) + \cdots + a_n f(w_n) = 0$. It follows that $f(a_1 w_1 + \cdots + a_n w_n) = 0$, which means that $a_1 w_1 + \cdots + a_n w_n \in \ker f$. Because f is injective, then by Lemma 4.2.4 we know that $\ker f = \{0\}$. Therefore $a_1 w_1 + \cdots + a_n w_n = 0$. Because $\{w_1, \ldots, w_n\} \subseteq B$, and because B is linearly independent, it follows from Lemma 3.5.7 (2) that $\{w_1, \ldots, w_n\}$ is linearly independent. Hence $a_1 = a_2 = \cdots = a_n = 0$. We deduce that $f(B)$ is linearly independent. Because f is surjective, we know that $\operatorname{im} f = W$. It follows from Lemma 4.2.6 that $\operatorname{span}(f(B)) = W$. We conclude that $f(B)$ is a basis for W.

Suppose that $f(B)$ is a basis for W. Then $\operatorname{span}(f(B)) = W$, and by Lemma 4.2.6 we deduce that $\operatorname{im} f = W$, which means that f is surjective. Let $v \in \ker f$. Because B is a basis for V, there are $m \in \mathbb{N}$, vectors $u_1, \ldots, u_m \in B$ and $c_1, \ldots, c_m \in F$ such that $v = c_1 u_1 + \cdots + c_m u_m$. Then $f(c_1 u_1 + \cdots + c_m u_m) = 0$, and hence $c_1 f(u_1) + \cdots + c_m (u_m) = 0$. Because $f(B)$ is linearly independent, it follows that $c_1 = \cdots = c_m = 0$. We deduce that $v = 0$. Therefore $\ker f = \{0\}$. By Lemma 4.2.4 we conclude that f is injective. □

Theorem 4.4.7. *Let V, W be vector spaces over a field F. Then V and W are isomorphic if and only if there is a basis B of V and a basis C of W such that B and C have the same cardinality.*

Proof. Suppose V and W are isomorphic. Let $f : V \to W$ be an isomorphism, and let D be a basis for V. Then by Lemma 4.4.6 we know that $f(D)$ is a basis for W, and clearly D and $f(D)$ have the same cardinality.

Suppose that there is a basis B of V and a basis C of W such that B and C have the same cardinality. Let $g : B \to C$ be a bijective map. Extend g to a linear map $h : V \to W$ by Theorem 4.1.6 (1). Then $h(B) = C$, so $h(B)$ is a basis for W, and it follows by Lemma 4.4.6 that h is an isomorphism. □

Corollary 4.4.8. *Let V, W be vector spaces over a field F. Suppose that V and W are isomorphic. Then V is finite-dimensional if and only if W is finite-dimensional. If V and W are both finite-dimensional, then* $\dim(V) = \dim(W)$.

Proof. This result follows immediately from Theorem 4.4.7, because a vector space is finite dimensional if and only if it has a finite basis, and the dimension of a finite-dimensional vector space is the cardinality of any basis for the vector space. □

Corollary 4.4.9. *Let V, W be vector spaces over a field F. Suppose that V and W are finite-dimensional. Then V and W are isomorphic if and only if* $\dim(V) = \dim(W)$*.*

Proof. This result follows immediately from Theorem 4.4.7, because the dimension of a finite-dimensional vector space is the cardinality of any basis for the vector space. □

Corollary 4.4.10. *Let V be a vector space over a field F. Suppose that V is finite-dimensional. Let* $n = \dim(V)$*. Then V is isomorphic to F^n.*

Proof. Observe that $\dim(F^n) = n$. The result then follows immediately from Corollary 4.4.9. □

Lemma 4.4.11. *Let V, W be vector spaces over a field F, let $X \subseteq V$ be a subspace and let $f : V \to W$ be an isomorphism. Suppose that V and W are finite-dimensional. Then* $\dim X = \dim f(X)$*.*

Proof. Observe that $f|_X$ is an isomorphism $X \to f(X)$, and then apply Corollary 4.4.8 to X and $f(X)$. □

Lemma 4.4.12. *Let V, W, Z be vector spaces over a field F, and let $f : V \to W$ and $g : W \to Z$ be linear maps. Suppose that V and W are finite-dimensional.*

1. *If f is an isomorphism, then* $\operatorname{rank}(g \circ f) = \operatorname{rank}(g)$*.*
2. *If g is an isomorphism, then* $\operatorname{rank}(g \circ f) = \operatorname{rank}(f)$*.*

Proof.

(1). Suppose that f is an isomorphism. Then f^{-1} is an isomorphism by Lemma 4.4.3. Observe that $\ker(g \circ f) = (g \circ f)^{-1}(\{0\}) = f^{-1}(g^{-1}(\{0\})) = f^{-1}(\ker g)$. Hence, by Lemma 4.4.11 applied to f^{-1}, we see that $\operatorname{nullity}(g) = \dim(\ker g) = \dim(f^{-1}(\ker g)) = \dim(\ker(g \circ f)) = \operatorname{nullity}(g \circ f)$. Next, we observe that $\operatorname{rank}(g) + \operatorname{nullity}(g) = \dim(W)$ and $\operatorname{rank}(g \circ f) + \operatorname{nullity}(g \circ f) = \dim(V)$. Because f is an isomorphism, we know by Lemma 4.4.9 that $\dim(V) = \dim(W)$. Then $\operatorname{rank}(g) = \dim(W) - \operatorname{nullity}(g) = \dim(V) - \operatorname{nullity}(g \circ f) = \operatorname{rank}(g \circ f)$.

(2). Suppose that g is an isomorphism. Observe that $\operatorname{im}(g \circ f) = (g \circ f)(V) = g(f(V)) = g(\operatorname{im} f)$. Hence, by Lemma 4.4.11 applied to g, we see that $\operatorname{rank}(f) = \dim(\operatorname{im} f) = \dim(g(\operatorname{im} f)) = \dim(\operatorname{im}(g \circ f)) = \operatorname{rank}(g \circ f)$. □

Exercises

Exercise 4.4.1. Let V be a vector space over a field F. Suppose that V non-trivial. Let B be a basis for V. Let $\mathcal{C}(B, F)$ be as defined in Exercise 3.3.2. It was seen in Exercise 3.3.2 that $\mathcal{C}(B, F)$ is a vector space over F. Let $\Psi\colon \mathcal{C}(B, F) \to V$ be defined by

$$\Psi(f) = \sum_{\substack{v \in B \\ f(v) \neq 0}} f(v)v,$$

for all $f \in \mathcal{C}(B, F)$. Prove that Ψ is an isomorphism. Hence every non-trivial vector space can be viewed as a space of functions.

4.5 Spaces of Linear Maps

Friedberg-Insel-Spence, 4th ed. – Section 2.2

Definition 4.5.1. Let V, W be vector spaces over a field F. The set of all linear maps $V \to W$ is denoted $\mathcal{L}(V, W)$. The set of all linear maps $V \to V$ is denoted $\mathcal{L}(V)$. △

Definition 4.5.2. Let A be a set, let W be a vector space over a field F, let $f, g\colon A \to W$ be functions and let $c \in F$.

1. Let $f + g\colon A \to W$ be defined by $(f+g)(x) = f(x) + g(x)$ for all $x \in A$.
2. Let $-f\colon A \to W$ be defined by $(-f)(x) = -f(x)$ for all $x \in A$.
3. Let $cf\colon A \to W$ be defined by $(cf)(x) = cf(x)$ for all $x \in A$.
4. Let $\mathbf{0}\colon A \to W$ be defined by $\mathbf{0}(x) = 0$ for all $x \in A$. △

Lemma 4.5.3. *Let V, W be vector spaces over a field F, let $f, g\colon V \to W$ be linear maps and let $c \in F$.*

1. *$f + g$ is a linear map.*
2. *$-f$ is a linear map.*
3. *cf is a linear map.*
4. *$\mathbf{0}$ is a linear map.*

Proof. We prove Part (1); the other parts are similar, and are left to the reader.

(1). Let $x, y \in V$ and let $d \in F$. Then

$$\begin{aligned}(f+g)(x+y) &= f(x+y) + g(x+y) = [f(x) + f(y)] + [g(x) + g(y)] \\ &= [f(x) + g(x)] + [f(y) + g(y)] = (f+g)(x) + (f+g)(y)\end{aligned}$$

and

$$(f+g)(dx) = f(dx) + g(dx) = df(x) + dg(x) = d[f(x) + g(x)] = d(f+g)(x).$$

□

Lemma 4.5.4. *Let V, W be vector spaces over a field F. Then $\mathcal{L}(V, W)$ is a vector space over F.*

Proof. We will show Property (7) in the definition of vector spaces; the other properties are similar. Let $f, g \in \mathcal{L}(V, W)$ and let $a \in F$. Let $x \in V$. Then

$$\begin{aligned}[a(f+g)](x) &= a[(f+g)(x)] = a[f(x)+g(x)] \\ &= af(x)+ag(x) = (af)(x)+(ag)(x) = [af+ag](x).\end{aligned}$$

Hence $a(f+g) = af + ag$. ☐

Lemma 4.5.5. *Let V, W, X, Z be vector spaces over a field F. Let $f, g\colon V \to W$ and $k\colon X \to V$ and $h\colon W \to Z$ be linear maps, and let $c \in F$.*

1. $(f+g)\circ k = (f\circ k) + (g \circ k)$.
2. $h\circ(f+g) = (h\circ f) + (h\circ g)$.
3. $c(h\circ f) = (ch)\circ f = h\circ(cf)$.

Proof. We prove Part (1); the other parts are similar, and are left to the reader.

(1). Let $x \in X$. Then

$$\begin{aligned}[(f+g)\circ k](x) &= (f+g)(k(x)) = f(k(x)) + g(k(x)) \\ &= (f\circ k)(x) + (g\circ k)(x) = [(f\circ k)+(g\circ k)](x).\end{aligned}$$

Hence $(f+g)\circ k = (f\circ k) + (g\circ k)$. ☐

Theorem 4.5.6. *Let V, W be vector spaces over a field F. Suppose that V and W are finite-dimensional. Then $\mathcal{L}(V, W)$ is finite-dimensional, and* $\dim(\mathcal{L}(V, W)) = \dim(V)\cdot\dim(W)$.

Proof. Let $n = \dim(V)$ and $m = \dim(W)$. Let $\{v_1, \ldots, v_n\}$ be a basis for V, and let $\{w_1, \ldots, w_m\}$ be a basis for W.

For each $i \in \{1, \ldots, n\}$ and $j \in \{1, \ldots, m\}$, let $e^{ij}\colon V \to W$ be defined as follows. First, let

$$e^{ij}(v_k) = \begin{cases} w_j, & \text{if } k = i \\ 0, & \text{if } k \in \{1, \ldots, n\} \text{ and } k \neq i. \end{cases}$$

Next, because $\{v_1, \ldots, v_n\}$ is a basis for V, we can use Theorem 4.1.6 (2) to extend e^{ij} to a unique linear map $V \to W$.

We claim that the set $T = \{e^{ij} \mid i \in \{1, \ldots, n\} \text{ and } j \in \{1, \ldots, m\}\}$ is a basis for $\mathcal{L}(V, W)$. Once we prove that claim, the result will follow, because T has nm elements.

Suppose that there is some $a_{ij} \in F$ for each $i \in \{1, \ldots, n\}$ and $j \in \{1, \ldots, m\}$ such that

$$\sum_{i=1}^{n}\sum_{j=1}^{m} a_{ij}e^{ij} = \mathbf{0}.$$

Let $k \in \{1, \ldots, n\}$. Then
$$\sum_{i=1}^{n}\sum_{j=1}^{m} a_{ij}e^{ij}(v_k) = \mathbf{0}(v_k),$$
which implies that
$$\sum_{j=1}^{m} a_{kj}w_j = 0.$$
Because $\{w_1, \ldots, w_m\}$ is linearly independent, it follows that $a_{kj} = 0$ for all $j \in \{1, \ldots, m\}$.

We deduce that $a_{ij} = 0$ for all $i \in \{1, \ldots, n\}$ and $j \in \{1, \ldots, m\}$. Hence T is linearly independent.

Let $f \in \mathcal{L}(V, W)$. Let $r \in \{1, \ldots, n\}$. Then $f(v_r) \in W$. Because $\{w_1, \ldots, w_m\}$ is spans W, there is some $c_{rp} \in F$ for each $p \in \{1, \ldots, m\}$ such that $f(v_r) = \sum_{j=1}^{m} c_{rj}w_j$.

Observe that
$$\sum_{i=1}^{n}\sum_{j=1}^{m} c_{ij}e^{ij}(v_r) = \sum_{j=1}^{m} c_{rj}wj = f(v_r).$$

Hence f and $\sum_{i=1}^{n}\sum_{j=1}^{m} c_{ij}e^{ij}$ agree on $\{v_1, \ldots, v_n\}$, and it follows from Corollary 4.1.7 that $f = \sum_{ij} c_{ij}e^{ij}$. Hence T spans $\mathcal{L}(V, W)$, and we conclude that T is a basis for $\mathcal{L}(V, W)$. □

Exercises

Exercise 4.5.1. Let V, W be vector spaces over a field F, and let $f, g\colon V \to W$ be non-zero linear maps. Suppose that $\operatorname{im} f \cap \operatorname{im} g = \{0\}$. Prove that $\{f, g\}$ is a linearly independent subset of $\mathcal{L}(V, W)$.

Exercise 4.5.2. Let V, W be vector spaces over a field F, and let $S \subseteq V$. Let $S^\circ \subseteq \mathcal{L}(V, W)$ be defined by
$$S^\circ = \{f \in \mathcal{L}(V, W) \mid f(x) = 0 \text{ for all } x \in S\}.$$

(1) Prove that S° is a subspace of $\mathcal{L}(V, W)$.

(2) Let $T \subseteq V$. Prove that if $S \subseteq T$, then $T^\circ \subseteq S^\circ$.

(3) Let $X, Y \subseteq V$ be subspaces. Prove that $(X + Y)^\circ = X^\circ \cap Y^\circ$. (See Definition 3.3.8 for the definition of $X + Y$.)

5 Linear Maps and Matrices

5.1 Review of Matrices—Multiplication

Friedberg-Insel-Spence, 4th ed. – Section 2.3

Definition 5.1.1. Let F be a field, and let $m, n, p \in \mathbb{N}$. Let $A \in \mathsf{M}_{m\times n}(F)$ and $B \in \mathsf{M}_{n\times p}(F)$. Suppose that $A = \left[a_{ij}\right]$ and $B = \left[b_{ij}\right]$. The matrix $AB \in \mathsf{M}_{m\times p}(F)$ is defined by $AB = \left[c_{ij}\right]$, where $c_{ij} = \sum_{k=1}^{n} a_{ik}b_{kj}$ for all $i \in \{1, \ldots, m\}$ and $j \in \{1, \ldots, p\}$. △

Lemma 5.1.2. *Let F be a field, and let $m, n, p, q \in \mathbb{N}$. Let $A \in \mathsf{M}_{m\times n}(F)$, let $B \in \mathsf{M}_{n\times p}(F)$ and let $C \in \mathsf{M}_{p\times q}(F)$.*

1. *$A(BC) = (AB)C$.*
2. *$AI_n = A$ and $I_m A = A$.*

Proof.

(1). Suppose that $A = \left[a_{ij}\right]$ and $B = \left[b_{ij}\right]$ and $C = \left[c_{ij}\right]$, and $AB = \left[s_{ij}\right]$ and $BC = \left[t_{ij}\right]$ and $A(BC) = \left[u_{ij}\right]$ and $(AB)C = \left[w_{ij}\right]$. Then $s_{ij} = \sum_{k=1}^{n} a_{ik}b_{kj}$ for all $i \in \{1, \ldots, m\}$ and $j \in \{1, \ldots, p\}$; and $t_{ij} = \sum_{z=1}^{p} b_{iz}c_{zj}$ for all $i \in \{1, \ldots, n\}$ and $j \in \{1, \ldots, q\}$. Then $u_{ij} = \sum_{x=1}^{n} a_{ix}t_{xj} = \sum_{x=1}^{n} a_{ix}\left(\sum_{z=1}^{p} b_{xz}c_{zj}\right)$ for all $i \in \{1, \ldots, m\}$ and $j \in \{1, \ldots, q\}$; and $w_{ij} = \sum_{y=1}^{p} s_{iy}c_{yj} = \sum_{y=1}^{p}\left(\sum_{k=1}^{n} a_{ik}b_{ky}\right)c_{yj}$ for all $i \in \{1, \ldots, m\}$ and $j \in \{1, \ldots, q\}$. Rearranging shows that $u_{ij} = w_{ij}$ for all $i \in \{1, \ldots, m\}$ and $j \in \{1, \ldots, q\}$.

(2). Straightforward. □

Lemma 5.1.3. *Let F be a field, and let $m, n, p \in \mathbb{N}$. Let $A, B \in \mathsf{M}_{m\times n}(F)$ and let $C, D \in \mathsf{M}_{n\times p}(F)$. Then $A(C + D) = AC + AD$ and $(A + B)C = AC + BC$.*

Proof. The proof of this fact about matrices is straightforward, and is material belonging to Elementary Linear Algebra; we omit the details. □

Definition 5.1.4. Let F be a field, and let $n \in \mathbb{N}$. Let $A \in \mathsf{M}_{n\times n}(F)$. The matrix A is **invertible** if there is some $B \in \mathsf{M}_{n\times n}(F)$ such that $BA = I_n$ and $AB = I_n$. Such a matrix B is an **inverse** of A. △

Lemma 5.1.5. *Let F be a field, and let $n \in \mathbb{N}$. Let $A \in \mathsf{M}_{n\times n}(F)$. If A has an inverse, then the inverse is unique.*

Proof. Suppose that A has two inverse matrices, say B and C. Then $AB = I_n = BA$ and $AC = I_n = CA$. Using standard properties of matrix multiplication, we then compute

$$B = BI_n = B(AC) = (BA)C = I_nC = C.$$

Because $B = C$, we deduce that A has a unique inverse. □

Definition 5.1.6. Let F be a field, and let $n \in \mathbb{N}$. Let $A \in \mathsf{M}_{n\times n}(F)$. If A has an inverse, then the inverse is denoted A^{-1}. △

Lemma 5.1.7. *Let F be a field, and let $n \in \mathbb{N}$. Let $A, B \in \mathsf{M}_{n\times n}(F)$. Suppose that A and B are invertible*

1. *A^{-1} is invertible, and $(A^{-1})^{-1} = A$.*

2. *AB is invertible, and $(AB)^{-1} = B^{-1}A^{-1}$.*

Proof. We prove Part (2), leaving the rest to the reader.

(2). By Lemma 5.1.5 we know that if AB has an inverse, then it is unique. If we can show that $(AB)(B^{-1}A^{-1}) = I_n$ and $(B^{-1}A^{-1})(AB) = I_n$, then it will follow that $B^{-1}A^{-1}$ is the unique inverse for AB, which means that $(AB)^{-1} = B^{-1}A^{-1}$. Using standard properties of matrix multiplication, we then compute

$$\begin{aligned}(AB)(B^{-1}A^{-1}) &= [(AB)B^{-1}]A^{-1} = [A(BB^{-1})]A^{-1} \\ &= [AI_n]A^{-1} = AA^{-1} = I_n.\end{aligned}$$

A similar computation shows that $(B^{-1}A^{-1})(AB) = I_n$. □

Definition 5.1.8. Let F be a field, and let $n \in \mathbb{N}$. The set of all $n \times n$ invertible matrices with entries in F is denoted $\mathrm{GL}_n(F)$. △

Definition 5.1.9. Let F be a field, and let $m, n \in \mathbb{N}$. Let $A \in \mathsf{M}_{m\times n}(F)$. Suppose that $A = \left[a_{ij}\right]$. The **transpose** of A is the matrix $A^t \in \mathsf{M}_{n\times m}(F)$ defined by $A^t = \left[c_{ij}\right]$, where $c_{ij} = a_{ji}$ for all $i \in \{1, \dots, n\}$ and $j \in \{1, \dots, m\}$. △

Remark 5.1.10. Let F be a field, and let $A \in \mathsf{M}_{n\times n}(F)$. Then A is symmetric if and only if $A^t = A$. ◊

Lemma 5.1.11. *Let F be a field, and let $m, n \in \mathbb{N}$. Let $A, B \in \mathsf{M}_{m\times n}(F)$, and let $s \in F$.*

1. *$(A + B)^t = A^t + B^t$.*

2. *$(sA)^t = sA^t$.*

3. *$A^{tt} = A$.*

Proof. The proofs of these facts about matrices are straightforward, and are material belonging to Elementary Linear Algebra; we omit the details. □

Lemma 5.1.12. *Let F be a field, and let $n \in \mathbb{N}$. Let $A, B \in \mathsf{M}_{n\times n}(F)$.*

1. *$(I_n)^t = I_n$.*

2. $(AB)^t = B^t A^t$.

3. *A is invertible if and only if A^t is invertible; if A is invertible, then $(A^t)^{-1} = (A^{-1})^t$.*

Proof. The proofs of the first two part are straightforward, and are material belonging to Elementary Linear Algebra; the third part follows from the first two parts. We omit the details. □

Exercises

Exercise 5.1.1. Let F be a field, and let $n \in \mathbb{N}$. Let $A, B \in \mathsf{M}_{n\times n}(F)$. The **trace** of A is defined by

$$\operatorname{tr} A = \sum_{i=1}^{n} a_{ii}.$$

Prove that $\operatorname{tr}(AB) = \operatorname{tr}(BA)$.

5.2 Linear Maps Given by Matrix Multiplication

Friedberg-Insel-Spence, 4th ed. – Section 2.3

Definition 5.2.1. Let F be a field, and let $m, n \in \mathbb{N}$. Let $A \in \mathsf{M}_{m\times n}(F)$. The **linear map induced** by A is the function $\mathsf{L}_A\colon F^n \to F^m$ defined by $\mathsf{L}_A(v) = Av$ for all $v \in F^n$. △

Lemma 5.2.2. *Let F be a field, and let $m, n, p \in \mathbb{N}$. Let $A, B \in \mathsf{M}_{m\times n}(F)$, let $C \in \mathsf{M}_{n\times p}(F)$, and let $s \in F$.*

1. *L_A is a linear map.*
2. *$\mathsf{L}_A = \mathsf{L}_B$ if and only if $A = B$.*
3. *$\mathsf{L}_{A+B} = \mathsf{L}_A + \mathsf{L}_B$.*
4. *$\mathsf{L}_{sA} = s\mathsf{L}_A$.*
5. *$\mathsf{L}_{AC} = \mathsf{L}_A \circ \mathsf{L}_C$.*
6. *Suppose $m = n$. Then $\mathsf{L}_{I_n} = 1_{F^n}$.*

Proof. Suppose that $A = \left[a_{ij}\right]$ and $B = \left[b_{ij}\right]$. Let $\{e_1, \ldots, e_n\}$ be the standard basis for F^n.

(1). Let $v, w \in F^n$. Then $\mathsf{L}_A(v + w) = A(v + w) = Av + Aw = \mathsf{L}_A(v) + \mathsf{L}_A(w)$, and $\mathsf{L}_A(sv) = A(sv) = s(Av) = s\mathsf{L}_A(v)$.

(2). If $A = B$, then clearly $\mathsf{L}_A = \mathsf{L}_B$.

Suppose $\mathsf{L}_A = \mathsf{L}_B$. Let $j \in \{1, \ldots, n\}$. Then $\mathsf{L}_A(e_i) = \mathsf{L}_B(e_i)$, and hence $Ae_j = Be_j$, which means that the j-th column of A equals the j-th column of B. Hence $A = B$.

(3). Let $v \in F^n$. Then $\mathsf{L}_{A+B}(v) = (A + B)(v) = Av + Bv = \mathsf{L}_A(v) + \mathsf{L}_B(v)$. Hence $\mathsf{L}_{A+B} = \mathsf{L}_A + \mathsf{L}_B$.

(4). The proof is similar to the proof of Part (3).

(5). Let $j \in \{1, \ldots, n\}$. Then $\mathsf{L}_{AC}(e_j) = (AC)(e_j)$, and $(\mathsf{L}_A \circ \mathsf{L}_C)(e_j) = \mathsf{L}_A(\mathsf{L}_C(e_j)) = A(C(e_j))$. Observe that $(AC)(e_j)$ is the j-th column of AC, and that $C(e_j)$ is the j-th column of C. However, the j-th column of AC is defined by A times the j-th column of C. Hence $\mathsf{L}_{AC}(e_j) = (\mathsf{L}_A \circ \mathsf{L}_C)(e_j)$. Therefore L_{AC} and $\mathsf{L}_A \circ \mathsf{L}_C$ agree on a basis, and by Corollary 4.1.7 we deduce that $\mathsf{L}_{AC} = \mathsf{L}_A \circ \mathsf{L}_C$.

(6). Trivial. □

Corollary 5.2.3. *Let F be a field, and let $m, n, p, q \in \mathbb{N}$. Let $A \in \mathsf{M}_{m\times n}(F)$, let $B \in \mathsf{M}_{n\times p}(F)$, and let $C \in \mathsf{M}_{p\times q}(F)$. Then $(AB)C = A(BC)$.*

Proof. Using Lemma 5.2.2 (3) together with the associativity of the composition of functions, we see that $\mathsf{L}_{A(BC)} = \mathsf{L}_A \circ \mathsf{L}_{BC} = \mathsf{L}_A \circ (\mathsf{L}_B \circ \mathsf{L}_C) = (\mathsf{L}_A \circ \mathsf{L}_B) \circ \mathsf{L}_C = \mathsf{L}_{AB} \circ \mathsf{L}_C = \mathsf{L}_{(AB)C}$. By Lemma 5.2.2 (2) we deduce that $A(BC) = (AB)C$. $\square$

5.3 All Linear Maps $F^n \to F^m$

Friedberg-Insel-Spence, 4th ed. – Section 2.2

Lemma 5.3.1. *Let F be a field. Let $n, m \in \mathbb{N}$, and let $f : F^n \to F^m$ be a linear map. Then $f = \mathsf{L}_A$, where $A \in \mathsf{M}_{m\times n}(F)$ is the matrix that has columns $f(e_1), \ldots, f(e_n)$.*

Proof. Let $i \in \{1, \ldots, n\}$. Let $\begin{bmatrix} a_{1i} \\ \vdots \\ a_{mi} \end{bmatrix} = f(e_i)$.

Let $v \in F^n$. Then $v = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$ for some $x_1, \ldots, x_n \in F$. Then

$$\begin{aligned} f(v) = f(\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}) &= f(x_1 e_1 + \cdots + x_n e_n) = x_1 f(e_1) + \cdots + x_n f(e_n) \\ &= x_1 \begin{bmatrix} a_{11} \\ \vdots \\ a_{m1} \end{bmatrix} + \cdots + x_n \begin{bmatrix} a_{1n} \\ \vdots \\ a_{mn} \end{bmatrix} = \begin{bmatrix} x_1 a_{11} + \cdots + x_n a_{1n} \\ \vdots \\ x_1 a_{m1} + \cdots + x_n a_{mn} \end{bmatrix} \\ &= \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = Av = \mathsf{L}_A(v). \end{aligned}$$

Hence $f = \mathsf{L}_A$. □

5.4 Coordinate Vectors with respect to a Basis

Friedberg-Insel-Spence, 4th ed. – Section 2.2

Definition 5.4.1. Let V be a vector space over a field F, and let $\beta \subseteq V$ be a basis for V. The set β is an **ordered basis** if the elements of β are given a specific order. △

Definition 5.4.2. Let V be a vector space over a field F. Suppose that V is finite-dimensional. Let $n = \dim(V)$. Let $\beta = \{v_1, \ldots, v_n\}$ be an ordered basis for V. Let $x \in V$. Then there are unique $a_1, \ldots, a_n \in F$ such that $x = a_1 v_1 + \cdots + a_n v_n$. The **coordinate vector** of x relative to β is $[x]_\beta = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} \in F^n$. △

Lemma 5.4.3. *Let F be a field, and let $n \in \mathbb{N}$. Let β be the standard ordered basis for F^n. If $v \in F^n$, then $[v]_\beta = v$*

Proof. Let $v \in F^n$. Suppose that $v = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}$. Let $\{e_1, \ldots, e_n\}$ be the standard basis for F^n. Then $v = a_1 e_1 + \cdots + a_n e_n$. It follows that $[v]_\beta = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} = v$. □

Definition 5.4.4. Let V be a vector space over a field F. Suppose that V is finite-dimensional. Let $n = \dim(V)$. Let β be an ordered basis for V. The **standard representation** of V with respect to β is the function $\phi_\beta \colon V \to F^n$ defined by $\phi_\beta(x) = [x]_\beta$ for all $x \in V$. △

Theorem 5.4.5. *Let V be a vector space over a field F. Suppose that V is finite-dimensional. Let $n = \dim(V)$. Let β be an ordered basis for V. Then ϕ_β is an isomorphism.*

Proof. Let $\{e_1, \ldots, e_n\}$ be the standard basis for F^n.

Let $\beta = \{u_1, \ldots, u_n\}$. Let $i \in \{1, \ldots, n\}$. Then $\phi_\beta(u_i) = e_i$. By Theorem 4.1.6 (2) there is a unique linear map $g \colon V \to F^n$ such that $g(u_i) = e_i$ for all $i \in \{1, \ldots, n\}$.

Let $v \in V$. Then there are unique $a_1, \ldots, a_n \in F$ such that $x = a_1 v_1 + \cdots + a_n v_n$. Hence

$$\begin{aligned} \phi_\beta(v) = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} &= a_1 e_1 + \cdots + a_n e_n = a_1 g(u_1) + \cdots + a_n g(u_n) \\ &= g(a_1 u_1 + \cdots + a_n u_n) = g(v). \end{aligned}$$

Hence $\phi_\beta = g$. It follows that ϕ_β is linear.

We know by Lemma 4.2.6 that $\operatorname{im} \phi_\beta = \operatorname{span}(\phi_\beta(\beta)) = \operatorname{span}\{e_1, \ldots, e_n\} = F^n$. Hence ϕ_β is surjective. Because $\dim(V) = n = \dim(F^n)$, it follows from Corollary 4.4.4 that ϕ_β is an isomorphism. □

5.5 Matrix Representation of Linear Maps—Basics

Friedberg-Insel-Spence, 4th ed. – Section 2.2

Definition 5.5.1. Let V, W be vector spaces over a field F, and let $f\colon V \to W$ be a linear map. Suppose that V and W are finite-dimensional. Let $n = \dim(V)$ and $m = \dim(W)$. Let $\beta = \{v_1, \ldots, v_n\}$ be an ordered basis for V and $\gamma = \{w_1, \ldots, w_n\}$ be an ordered basis for W. The **matrix representation** of f with respect to β and γ is the $m \times n$ matrix $[f]_\beta^\gamma$ with j-th column equal to $[f(v_j)]_\gamma$ for all $j \in \{1, \ldots, n\}$.

If $V = W$ and $\beta = \gamma$, the matrix $[f]_\beta^\gamma$ is written $[f]_\beta$. △

Remark 5.5.2. With the hypotheses of Definition 5.5.1, we see that $[f]_\beta^\gamma = \left[a_{ij}\right]$, where the elements $a_{ij} \in F$ are the elements such that

$$f(v_j) = \sum_{i=1}^{m} a_{ij} w_i$$

for all $j \in \{1, \ldots n\}$. ◊

Lemma 5.5.3. *Let V, W be vector spaces over a field F, let $f, g\colon V \to W$ be linear maps, and let $c \in F$. Suppose that V and W are finite-dimensional. Let $n = \dim(V)$. Let β be an ordered basis for V, and let γ be an ordered basis for W.*

1. *$[f]_\beta^\gamma = [g]_\beta^\gamma$ if and only if $f = g$.*
2. *$[f + g]_\beta^\gamma = [f]_\beta^\gamma + [g]_\beta^\gamma$.*
3. *$[cf]_\beta^\gamma = c[f]_\beta^\gamma$.*
4. *$[1_V]_\beta = I_n$.*

Proof. We prove Part (1); the other parts are straightforward.

(1). If $f = g$, then clearly $[f]_\beta^\gamma = [g]_\beta^\gamma$.

Suppose that $[f]_\beta^\gamma = [g]_\beta^\gamma$. Let $\beta = \{v_1, \ldots, v_n\}$. Let $j \in \{1, \ldots, n\}$. Then $[f(v_j)]_\gamma$ is the j-th column of $[f]_\beta^\gamma$, and $[g(v_j)]_\gamma$ is the j-th column of $[g]_\beta^\gamma$. It follows that $f(v_j)$ and $g(v_j)$ have the same coordinate vector relative to γ. Hence $f(v_j) = g(v_j)$. Therefore f and g agree on a basis, and by Corollary 4.1.7 we deduce that $f = g$. □

Exercises

Exercise 5.5.1. Let $\beta = \{\left[\begin{smallmatrix}1\\0\end{smallmatrix}\right], \left[\begin{smallmatrix}0\\1\end{smallmatrix}\right]\}$ and let $\gamma = \{\left[\begin{smallmatrix}2\\-1\end{smallmatrix}\right], \left[\begin{smallmatrix}1\\3\end{smallmatrix}\right]\}$; these are bases for $\mathbb{R}^2$. Let $f\colon \mathbb{R}^2 \to \mathbb{R}^2$ be defined by $f(\left[\begin{smallmatrix}x\\y\end{smallmatrix}\right]) = \left[\begin{smallmatrix}x-y\\3x+y\end{smallmatrix}\right]$ for all $\left[\begin{smallmatrix}x\\y\end{smallmatrix}\right] \in \mathbb{R}^2$. Then find $[f]_\beta$ and $[f]_\beta^\gamma$.

Exercise 5.5.2. Let $H\colon \mathbb{R}_3[x] \to \mathbb{R}_3[x]$ by defined by $H(f) = xf' - f$ for all $f \in \mathbb{R}_3[x]$. Let β be the standard ordered basis for $\mathbb{R}_3[x]$. Find $[H]_\beta$. We will use this example again.

Exercise 5.5.3. Let V, W be vector spaces over a field F. Suppose that V and W are finite-dimensional. Let $n = \dim(V)$ and $m = \dim(W)$. Let β be an ordered basis for V, and let γ be an ordered basis for W. Let $A \in \mathsf{M}_{m\times n}(F)$. Prove that there is a linear map $f\colon V \to W$ such that $[f]_\beta^\gamma = A$.

Exercise 5.5.4. Let V, W be vector spaces over a field F, and let $f\colon V \to W$ be a linear map. Suppose that V and W are finite-dimensional.

(1) Suppose that f is an isomorphism. Then there is an ordered basis α for V and an ordered basis δ for W such that $[f]_\alpha^\delta$ is the identity matrix.

(2) Suppose that f is an arbitrary linear map. Then there is an ordered basis α for V and an ordered basis δ for W such that $[f]_\alpha^\delta$ has the form

$$[f]_\beta^\gamma = \begin{bmatrix} I_r & O \\ O & O \end{bmatrix},$$

where O denotes the appropriate zero matrices, for some $r \in \{0, 1, \dots, n\}$.

5.6 Matrix Representation of Linear Maps—Composition

Friedberg-Insel-Spence, 4th ed. – Section 2.3

Theorem 5.6.1. *Let V, W, Z be vector spaces over a field F, and let $f\colon V \to W$ and $g\colon W \to Z$ be linear maps. Suppose that V, W and Z are finite-dimensional. Let β be an ordered basis for V, let γ be an ordered basis for W, and let δ be an ordered basis for Z. Then $[g \circ f]_\beta^\delta = [g]_\gamma^\delta [f]_\beta^\gamma$.*

Proof. Suppose that $[f]_\beta^\gamma = \left[a_{ij}\right]$, that $[g]_\gamma^\delta = \left[b_{ij}\right]$, that $[g \circ f]_\beta^\delta = \left[c_{ij}\right]$, and that $[g]_\gamma^\delta [f]_\beta^\gamma = \left[d_{ij}\right]$.

Let $n = \dim(V)$, let $m = \dim(W)$ and let $p = \dim(Z)$. Let $\beta = \{v_1, \ldots, v_n\}$, let $\gamma = \{w_1, \ldots, w_m\}$ and let $\delta = \{z_1, \ldots, z_p\}$.

By the definition of matrix multiplication, we see that $d_{ij} = \sum_{k=1}^n b_{ik} a_{kj}$ for all $i \in \{1, \ldots, p\}$ and $j \in \{1, \ldots, n\}$.

Let $j \in \{1, \ldots, n\}$. Then by Remark 5.5.2 we see that

$$(g \circ f)(v_j) = \sum_{r=1}^p c_{rj} z_r.$$

On the other hand, using Remark 5.5.2 again, we have

$$\begin{aligned} (g \circ f)(v_j) = g(f(v_j)) = g(\sum_{i=1}^m a_{ij} w_i) &= \sum_{i=1}^m a_{ij} g(w_i) \\ = \sum_{i=1}^m a_{ij} \left[\sum_{r=1}^p b_{ri} z_r \right] &= \sum_{r=1}^p \left[\sum_{i=1}^m b_{ri} a_{ij} \right] z_r. \end{aligned}$$

Because $\{z_1, \ldots, z_p\}$ is a basis, it follows Theorem 3.6.2 (2) that $\sum_{i=1}^m b_{ri} a_{ij} = c_{rj}$ for all $r \in \{1, \ldots, p\}$.

Hence $d_{ij} = c_{ij}$ for all $i \in \{1, \ldots, p\}$ and $j \in \{1, \ldots, n\}$, which means that $[g \circ f]_\beta^\delta = [g]_\gamma^\delta [f]_\beta^\gamma$. ☐

Theorem 5.6.2. *Let V, W be vector spaces over a field F, and let $f\colon V \to W$ be a linear map. Suppose that V and W are finite-dimensional. Let β be an ordered basis for V and let γ be an ordered basis for W. Let $v \in V$. Then $[f(v)]_\gamma = [f]_\beta^\gamma [v]_\beta$.*

Proof. Let $h\colon F \to V$ be defined by $h(a) = av$ for all $a \in F$. Let $g\colon F \to W$ be defined by $g(a) = af(v)$ for all $a \in F$. It can be verified that h and g are linear maps; the details are left to the reader.

Let $\alpha = \{1\}$ be the standard basis ordered basis for F as a vector space over itself. Observe that $f \circ h = g$, because $f(h(a)) = f(av) = af(v) = g(a)$ for all $a \in F$. Then

$$[f(v)]_\gamma = [g(1)]_\gamma = [g]_\alpha^\gamma = [f \circ h]_\alpha^\gamma = [f]_\beta^\gamma [h]_\alpha^\beta = [f]_\beta^\gamma [h(1)]_\beta = [f]_\beta^\gamma [v]_\beta \qquad \square$$

Lemma 5.6.3. *Let F be a field, and let $m, n \in \mathbb{N}$. Let β be the standard ordered basis for F^n, and let γ be the standard ordered basis for F^m.*

1. *Let $A \in \mathsf{M}_{m\times n}(F)$. Then $[\mathsf{L}_A]_\beta^\gamma = A$.*
2. *Let $f \colon F^n \to F^m$ be a linear map. Then $f = \mathsf{L}_C$, where $C = [f]_\beta^\gamma$.*

Proof.

(1). Let $\{e_1, \ldots, e_n\}$ be the standard basis for F^n. Let $j \in \{1, \ldots, n\}$. By Lemma 5.4.3, we see that $Ae_j = \mathsf{L}_A(e_j) = [\mathsf{L}_A(e_j)]_\gamma$. Observe that Ae_j is the j-th column of A, and $[\mathsf{L}_A(e_j)]_\gamma$ is the j-th column of $[\mathsf{L}_A]_\beta^\gamma$. Hence $A = [\mathsf{L}_A]_\beta^\gamma$.

(2). Let $v \in F^n$. Using Lemma 5.4.3 and Theorem 5.6.2, we see that $f(v) = [f(v)]_\gamma = [f]_\beta^\gamma [v]_\beta = Cv = \mathsf{L}_C(v)$. Hence $f = \mathsf{L}_C$. $\square$

5.7 Matrix Representation of Linear Maps—Isomorphisms

Friedberg-Insel-Spence, 4th ed. – Section 2.4

Theorem 5.7.1. *Let V, W be vector spaces over a field F, and let $f\colon V \to W$ be a linear map. Suppose that V and W are finite-dimensional, and that $\dim(V) = \dim(W)$. Let β be an ordered basis for V, and let γ be an ordered basis for W.*

1. *f is an isomorphism if and only if $[f]_\beta^\gamma$ is invertible.*
2. *If f is an isomorphism, then $[f^{-1}]_\gamma^\beta = \left([f]_\beta^\gamma\right)^{-1}$.*

Proof. Both parts of the theorem are proved together. Let $n = \dim(V) = \dim(W)$.

Suppose that f is an isomorphism. By definition of inverse maps we know that $f^{-1} \circ f = 1_V$ and $f \circ f^{-1} = 1_W$. By Lemma 4.4.3 we know that that f^{-1} is a linear map. Hence, using Theorem 5.6.1 and Lemma 5.5.3 (4), we deduce that

$$[f^{-1}]_\gamma^\beta [f]_\beta^\gamma = [f^{-1} \circ f]_\beta^\beta = [1_V]_\beta^\beta = [1_V]_\beta = I_n.$$

A similar argument shows that

$$[f]_\beta^\gamma [f^{-1}]_\gamma^\beta = I_n.$$

It follows that $[f]_\beta^\gamma$ is invertible and $\left([f]_\beta^\gamma\right)^{-1} = [f^{-1}]_\gamma^\beta$.

Suppose that $[f]_\beta^\gamma$ is invertible. Let $A = [f]_\beta^\gamma$. Then there is some $B \in \mathsf{M}_{n\times n}(F)$ such that $AB = I_n$ and $BA = I_n$. Suppose that $B = \left[b_{ij}\right]$.

Suppose that $\beta = \{v_1, \ldots, v_n\}$ and that $\gamma = \{w_1, \ldots, w_n\}$. By Theorem 4.1.6 (2) there is a unique linear map $g\colon W \to V$ such that $g(w_i) = \sum_{k=1}^n b_{ki} v_i$ for all $i \in \{1, \ldots, n\}$. Then by definition we have $[g]_\gamma^\beta = B$.

Using Theorem 5.6.1 and Lemma 5.5.3 (4), we deduce that

$$[g \circ f]_\beta^\beta = [g]_\gamma^\beta [f]_\beta^\gamma = BA = I_n = [1_V]_\beta^\beta.$$

A similar argument shows that

$$[f \circ g]_\gamma^\gamma = [1_W]_\gamma^\gamma.$$

It follows from Lemma 5.5.3 (1) that $g \circ f = 1_V$ and $f \circ g = 1_W$. Hence f has an inverse, and it is therefore bijective. We conclude that f is an isomorphism. □

Corollary 5.7.2. *Let F be a field, and let $n \in \mathbb{N}$. Let $A \in \mathsf{M}_{n\times n}(F)$.*

1. *A is invertible if and only if L_A is an isomorphism.*

2. *If A is invertible, then* $(\mathsf{L}_A)^{-1} = \mathsf{L}_{A^{-1}}$.

Proof. Left to the reader in Exercise 5.7.3. □

Exercises

Exercise 5.7.1. In this exercise, we will use the notation $f(\beta) = \gamma$ in the sense of ordered bases, so that f takes the first element of β to the first element of γ, the second element of β to the second element of γ, etc.

Let V, W be vector spaces over a field F, and let $f : V \to W$ be a linear map. Suppose that V and W are finite-dimensional.

(1) Let β be an ordered basis for V and let γ be an ordered basis for W. Then $[f]_\beta^\gamma$ is the identity matrix if and only if $f(\beta) = \gamma$.

(2) The map f is an isomorphism if and only if there is an ordered basis α for V and an ordered basis δ for W such that $[f]_\alpha^\delta$ is the identity matrix.

Exercise 5.7.2. Let V, W be vector spaces over a field F, and let $f : V \to W$ be a linear map. Suppose that V and W are finite-dimensional. Let β be an ordered basis for V, and let γ be an ordered basis for W. Let $A = [f]_\beta^\gamma$.

(1) Prove that $\operatorname{rank}(f) = \operatorname{rank}(\mathsf{L}_A)$.

(2) Prove that $\operatorname{nullity}(f) = \operatorname{nullity}(\mathsf{L}_A)$.

Exercise 5.7.3. Prove Corollary 5.7.2.

5.8 Matrix Representation of Linear Maps—The Big Picture

Friedberg-Insel-Spence, 4th ed. – Section 2.4

Theorem 5.8.1. *Let V, W be vector spaces over a field F. Suppose that V and W are finite-dimensional. Let $n = \dim(V)$ and let $m = \dim(W)$. Let β be an ordered basis for V, and let γ be an ordered basis for W. Let $\Phi\colon \mathcal{L}(V, W) \to \mathsf{M}_{m\times n}(F)$ be defined by $\Phi(f) = [f]_\beta^\gamma$ for all $f \in \mathcal{L}(V, W)$.*

1. *Φ is an isomorphism.*
2. *$\mathsf{L}_{\Phi(f)} \circ \phi_\beta = \phi_\gamma \circ f$ for all $f \in \mathcal{L}(V, W)$.*

Proof.

(1). The fact that Φ is a linear map is just a restatement of Lemma 5.5.3 (2) and (3). We know by Theorem 4.5.6 that $\dim(\mathcal{L}(V, W)) = nm$. We also know that $\dim(\mathsf{M}_{m\times n}(F)) = nm$. Hence $\dim(\mathcal{L}(V, W)) = \dim(\mathsf{M}_{m\times n}(F))$. The fact that Φ is injective is just a restatement of Lemma 5.5.3 (1). It now follows from Corollary 4.4.4 that Φ is an isomorphism.

(2). Let $f \in \mathcal{L}(V, W)$. Let $v \in V$. Using Theorem 5.6.2, we see that

$$\begin{aligned}(\phi_\gamma \circ f)(v) &= \phi_\gamma(f(v)) = [f(v)]_\gamma = [f]_\beta^\gamma [v]_\beta \\ &= \Phi(f)\phi_\beta(v) = \mathsf{L}_{\Phi(f)}(\phi_\beta(v)) = (\mathsf{L}_{\Phi(f)} \circ \phi_\beta)(v).\end{aligned}$$

Hence $\mathsf{L}_{\Phi(f)} \circ \phi_\beta = \phi_\gamma \circ f$. □

Remark 5.8.2. The equation $\mathsf{L}_{\Phi(f)} \circ \phi_\beta = \phi_\gamma \circ f$ in Theorem 5.8.1 (2) is represented by the following commutative diagram, where "commutative" here means that going around the diagram either way yields the same result.

$$\begin{tikzcd} V \arrow[r, "f"] \arrow[d, "\phi_\beta"'] & W \arrow[d, "\phi_\gamma"] \\ F^n \arrow[r, "\mathsf{L}_{\Phi(f)}"'] & F^m \end{tikzcd}$$

◊

5.9 Matrix Representation of Linear Maps—Change of Basis

Friedberg-Insel-Spence, 4th ed. – Section 2.5

Lemma 5.9.1. *Let V be a vector space over a field F. Suppose that V is finite-dimensional. Let β and β' be ordered bases for V.*

1. $[1_V]_{\beta'}^{\beta}$ *is invertible.*

2. *If $v \in V$, then $[v]_\beta = [1_V]_{\beta'}^{\beta}[v]_{\beta'}$.*

Proof.

(1). We know that 1_V is an isomorphism, and therefore Theorem 5.7.1 (1) implies that $[1_V]_{\beta'}^{\beta}$ is invertible.

(2). Let $v \in V$. Then $1_V(v) = v$, and hence $[1_V(v)]_\beta = [v]_\beta$. It follows from Theorem 5.6.2 that $[1_V]_{\beta'}^{\beta}[v]_{\beta'} = [v]_\beta$. □

Definition 5.9.2. Let V be a vector space over a field F. Suppose that V is finite-dimensional. Let β and β' be ordered bases for V. The **change of coordinate matrix** (also called the **change of basis matrix**) that changes β'-coordinates into β-coordinates is the matrix $[1_V]_{\beta'}^{\beta}$. △

Remark 5.9.3. Let V be a vector space over a field F. Suppose that V is finite-dimensional. Let β and β' be ordered bases for V. The change of coordinate matrix that changes β'-coordinates into β-coordinates is formed by writing the elements of β' in terms of β and putting the coordinates of each element of β' in terms of β into a column vector, and assembling these column vectors into a matrix. ◊

Lemma 5.9.4. *Let V be a vector space over a field F. Suppose that V is finite-dimensional. Let α, β and γ be ordered bases for V. Let Q be the change of coordinate matrix that changes α-coordinates into β-coordinates, and let R be the change of coordinate matrix that changes β-coordinates into γ-coordinates*

1. *RQ is the change of coordinate matrix that changes α-coordinates into γ-coordinates*

2. *Q^{-1} is the change of coordinate matrix that changes β-coordinates into α-coordinates*

Proof. Left to the reader in Exercise 5.9.1. □

Theorem 5.9.5. *Let V, W be vector spaces over a field F. Suppose that V and W are finite-dimensional. Let β and β' be ordered bases for V, and let γ and γ' be ordered bases for W. Let Q be the change of coordinate matrix that changes β'-coordinates into β-coordinates, and let P be the change of coordinate matrix that changes γ'-coordinates into γ-coordinates. If $f\colon V \to W$ is a linear map, then $[f]_{\beta'}^{\gamma'} = P^{-1}[f]_{\beta}^{\gamma}Q$.*

Proof. Let $f\colon V \to W$ be a linear map. Observe that $f = 1_W \circ f \circ 1_V$. Then $[f]_{\beta'}^{\gamma'} = [1_W \circ f \circ 1_V]_{\beta'}^{\gamma'}$. It follows from Theorem 5.6.1 that $[f]_{\beta'}^{\gamma'} = [1_W]_{\gamma}^{\gamma'}[f]_{\beta}^{\gamma}[1_V]_{\beta'}^{\beta}$. By Lemma 5.9.4, we deduce that $[f]_{\beta'}^{\gamma'} = P^{-1}[f]_{\beta}^{\gamma}Q$. □

Corollary 5.9.6. *Let V be a vector space over a field F. Suppose that V is finite-dimensional. Let β and β' be ordered bases for V. Let Q be the change of coordinate matrix that changes β'-coordinates into β-coordinates. If $f\colon V \to V$ is a linear map, then $[f]_{\beta'} = Q^{-1}[f]_{\beta}Q$.*

Corollary 5.9.7. *Let F be a field, and let $n \in \mathbb{N}$. Let $A \in \mathsf{M}_{n\times n}(F)$. Let $\gamma = \{v_1, \ldots, v_n\}$ be an ordered basis for F^n. Let $Q \in \mathsf{M}_{n\times n}(F)$ be the matrix whose j-th column is v_j. Then $[\mathsf{L}_A]_{\gamma} = Q^{-1}AQ$.*

Definition 5.9.8. Let F be a field, and let $n \in \mathbb{N}$. Let $A, B \in \mathsf{M}_{n\times n}(F)$. The matrices A and B are **similar** if there is an invertible matrix $Q \in \mathsf{M}_{n\times n}(F)$ such that $A = Q^{-1}BQ$. △

Lemma 5.9.9. *Let F be a field, and let $n \in \mathbb{N}$. The relation of matrices being similar is an equivalence relation on $\mathsf{M}_{n\times n}(F)$.*

Proof. Left to the reader in Exercise 5.9.2. □

Corollary 5.9.10. *Let V be a vector space over a field F, and let $f\colon V \to V$ be a linear map. Suppose that V is finite-dimensional. Let β and β' be ordered bases for V. Then $[f]_{\beta}$ and $[f]_{\beta'}$ are similar.*

Lemma 5.9.11. *Let V be a vector space over a field F. Suppose that V is finite-dimensional. Let $\beta = \{x_1, \ldots, x_n\}$ be an ordered basis for V. Let $Q \in \mathsf{M}_{n\times n}(F)$ be an invertible matrix. Define $\beta' = \{x'_1, \ldots, x'_n\}$ by $x'_j = \displaystyle\sum_{i=1}^{n} Q_{ij}x_i$ for all $j \in \{1, \ldots, n\}$. Then β' is a basis for V, and Q is the change of coordinate matrix that changes β'-coordinates into β-coordinates.*

Proof. It suffices to show that β' is linearly independent. Suppose $\sum_{j=1} a_j x'_j = 0$ for some $a_1, \ldots, a_n \in F$. Then plug in the definition of the x'_j, rearrange, and deduce from the linear independence of β that $\sum_{j=1}^{n} a_j Q_{ij} = 0$ for each $i \in \{1, \ldots, n\}$. Let A be the column vector with entries $a_1, \ldots, a_n$ going down. Then QA equals the zero column vector. Because Q is invertible, it follows that A is the zero column vector, which is what needed to be proved. □

Corollary 5.9.12. *Let F be a field, and let $n \in \mathbb{N}$. Let $A, B \in \mathsf{M}_{n\times n}(F)$, and suppose that $B = Q^{-1}AQ$ for some invertible $Q \in \mathsf{M}_{n\times n}(F)$. Then there exists a finite-dimensional vector space V over F, with $\dim(V) = n$, bases β and β' for V, and a linear map $f : V \to V$ such that $A = [f]_\beta$ and $B = [f]_{\beta'}$.*

Proof. Left to the reader in Exercise 5.9.3. □

Exercises

Exercise 5.9.1. Prove Lemma 5.9.4.

Exercise 5.9.2. Prove Lemma 5.9.9.

Exercise 5.9.3. Prove Corollary 5.9.12.

6 Applications of Linear Maps to Matrices and Systems of Linear Equations

6.1 Elementary Moves

Definition 6.1.1. Let F be a field. Let $A \in \mathsf{M}_{m\times n}(F)$. The **elementary row and column operations** on A are as follows.

1. interchanging any two columns
2. multiplying any column by a non-zero scalar
3. adding a scalar multiple of one column to another column △

Definition 6.1.2. Let V, W be vector spaces over a field F, let $\beta = \{v_1, \ldots, v_n\}$ be a finite ordered subset of V, and let $f : V \to W$ be a linear map. We will use the notation $f(\beta)$ to denote the ordered set $\{f(v_1), \ldots, f(v_n)\}$, where all n elements $f(v_1), \ldots, f(v_n)$ are thought of as distinct, and in that order. △

Definition 6.1.3. Let V be a finite-dimensional vector space over a field F, let β and γ be ordered subsets of V. The basis γ **can be obtained from** β **by an elementary move** of Type 1, Type 2 or Type 3 (respectively) if the following holds.

Type 1: γ is the same as β, except that two of the elements of β have switched places. If the i-th and k-th elements of β are switched, where $i \neq k$, we denote this elementary move by $\mathcal{E}_1(i, k)$.

Type 2: γ is the same as β, except that one elements of β has been multiplied by a non-zero scalar. If the i-th element of β is multiplied by $a \in F$, where $a \neq 0$, we denote this elementary move by $\mathcal{E}_2(i; a)$.

Type 3: γ is the same as β, except that a scalar multiple of one element of β has been added to another element of β. If a times the k-th element of β is added to the i-th element of β, for some $a \in F$, we denote this elementary move by $\mathcal{E}_3(k, i; a)$.

△

Remark 6.1.4. We can write out the three types of elementary moves explicitly as follows. Let V be a finite-dimensional vector space over a field F, and let $\beta = \{v_1, \ldots, v_n\}$ be an ordered subset of V. Suppose that γ can be obtained from β by an elementary move $\mathcal{E}$. We then have the following three cases.

Type 1: If $\mathcal{E} = \mathcal{E}_1(i, k)$ for some $i, k \in \{1, \ldots, n\}$ such that $i \neq k$, then $\gamma = \{v_1, \ldots, v_{i-1}, v_k, v_{i+1}, \ldots, v_{k-1}, v_i, v_k$

Type 2: If $\mathcal{E} = \mathcal{E}_2(i; a)$ for some $i \in \{1, \ldots, n\}$ and $a \in F$ such that $a \neq 0$, then $\gamma = \{v_1, \ldots, v_{i-1}, av_i, v_{i+1}, \ldots, v_n\}$.

Type 3: If $\mathcal{E} = \mathcal{E}_3(k, i; a)$ for some $i, k \in \{1, \ldots, n\}$ and $a \in F$, then $\gamma = \{v_1, \ldots, v_{i-1}, v_i + av_k, v_{i+1}, \ldots, v_n\}$ for some $i, k \in \{1, \ldots, n\}$ such that $i \neq k$, and some $a \in F$.

◊

Definition 6.1.5. Let F be a field, let $a \in F$, and let $i, k \in \mathbb{N}$. Let $\mathcal{E}$ be an elementary move.
(1) The **reverse elementary move** of $\mathcal{E}$, denoted $\mathcal{E}^R$, is the elementary move given by

$$\mathcal{E}^R = \begin{cases} \mathcal{E}_1(i,j), & \text{if } \mathcal{E} = \mathcal{E}_1(i,j) \\ \mathcal{E}_2(i;a^{-1}), & \text{if } \mathcal{E} = \mathcal{E}_2(i;a) \\ \mathcal{E}_3(k,i;-a), & \text{if } \mathcal{E} = \mathcal{E}_3(k,i;a). \end{cases}$$

(2) The **associate elementary move** of $\mathcal{E}$, denoted $\mathcal{E}^A$, is the elementary move given by

$$\mathcal{E}^A = \begin{cases} \mathcal{E}_1(i,j), & \text{if } \mathcal{E} = \mathcal{E}_1(i,j) \\ \mathcal{E}_2(i;a), & \text{if } \mathcal{E} = \mathcal{E}_2(i;a) \\ \mathcal{E}_3(i,k;a), & \text{if } \mathcal{E} = \mathcal{E}_3(k,i;a). \end{cases}$$

(3) The **obverse elementary move** of $\mathcal{E}$, denoted $\mathcal{E}^O$, is the elementary move given by

$$\mathcal{E}^O = \begin{cases} \mathcal{E}_1(i,j), & \text{if } \mathcal{E} = \mathcal{E}_1(i,j) \\ \mathcal{E}_2(i;a^{-1}), & \text{if } \mathcal{E} = \mathcal{E}_2(i;a) \\ \mathcal{E}_3(i,k;-a), & \text{if } \mathcal{E} = \mathcal{E}_3(k,i;a). \end{cases}$$ △

Lemma 6.1.6. *Let $\mathcal{E}$ be an elementary move.*

1. $(\mathcal{E}^R)^R = \mathcal{E}$.
2. $(\mathcal{E}^A)^A = \mathcal{E}$.
3. $(\mathcal{E}^O)^O = \mathcal{E}$.
4. $(\mathcal{E}^O)^R = \mathcal{E}^A = (\mathcal{E}^R)^O$.
5. $(\mathcal{E}^A)^R = \mathcal{E}^O = (\mathcal{E}^R)^A$.
6. $(\mathcal{E}^O)^A = \mathcal{E}^R = (\mathcal{E}^A)^O$.

Proof. This proof is straightforward, and simply involves looking at the three three types of elementary moves for each part of the lemma. We omit the details. □

Lemma 6.1.7. *Let V be a vector space over a field F, and let β and γ be finite ordered subsets of V. If γ is obtained from β by an elementary move $\mathcal{E}$, then β is obtained from γ by $\mathcal{E}^R$.*

Proof. This proof is straightforward, and the details are omitted. □

Lemma 6.1.8. *Let V be a finite-dimensional vector space over a field F, and let β and γ be finite ordered subsets of V. Suppose that γ can be obtained from β by an elementary move. Then γ is a basis for V if and only if β is a basis for V.*

Proof. First, suppose that β is a basis for V. Let $\beta = \{v_1, \ldots, v_n\}$. Suppose that γ is obtained from β by the elementary move $\mathcal{E}$. Because γ has the same number of elements as β, we know by Corollary 3.6.9 (4) that in order to prove that γ is a basis, it suffices to prove that γ is linearly independent. We have to examine each type of elementary move separately.

Type 1: Suppose that $\mathcal{E} = \mathcal{E}_1(i, k)$ for some $i, k \in \{1, \ldots, n\}$ such that $i \neq k$. In this case γ is the same set as β, though in a different order, and so clearly γ is a basis.

Type 2: Suppose $\mathcal{E} = \mathcal{E}_2(i; a)$ for some $i \in \{1, \ldots, n\}$ and $a \in F$ such that $a \neq 0$. Then $\gamma = \{v_1, \ldots, v_{i-1}, av_i, v_{i+1}, \ldots, v_n\}$. Suppose

$$b_1 v_1 + \cdots + b_{i-1} v_{i-1} + b_i a v_i + b_{i+1} v_{i+1} + \cdots + b_n v_n = 0$$

for some $b_1, \ldots, b_n \in F$. Then $b_j = 0$ for all $j \in \{1, \ldots, n\}$ such that $j \neq i$, and $b_i a = 0$. Because $a \neq 0$, we know by Lemma 3.1.2 (14) that $b_i = 0$. Hence γ is linearly independent.

Type 3: Suppose $\mathcal{E} = \mathcal{E}_3(k, i; a)$ for some $i, k \in \{1, \ldots, n\}$ and $a \in F$. Then $\gamma = \{v_1, \ldots, v_{i-1}, v_i + av_k, v_{i+1}, \ldots, v_n\}$. Suppose

$$b_1 v_1 + \cdots + b_{i-1} v_{i-1} + b_i(v_i + av_k) + b_{i+1} v_{i+1} + \cdots + b_n v_n = 0$$

for some $b_1, \ldots, b_n \in F$. Hence

$$\begin{aligned} b_1 v_1 + \cdots + b_{i-1} v_{i-1} + b_i v_i + b_{i+1} v_{i+1} + \cdots \\ + b_{k-1} v_{k-1} + (b_k + b_i a) v_k + b_{k+1} v_{k+1} + \cdots + b_n v_n = 0. \end{aligned}$$

Then $b_j = 0$ for all $j \in \{1, \ldots, n\}$ such that $j \neq k$, and $b_k + b_i a = 0$. Because $b_i = 0$, it follows that $b_k = 0$. Hence γ is linearly independent.

Now suppose that γ is a basis for V. By Lemma 6.1.7, we know that β can be obtained from γ by an elementary move. The same argument as above shows that β is a basis. □

Theorem 6.1.9. *Let V be a finite-dimensional vector space over a field F, and let β and γ be ordered bases for V. Then there is a finite collection of ordered bases $\beta = \alpha_0, \alpha_1, \ldots, \alpha_p = \gamma$ of V such that α_i is obtained from α_{i-1} by a single elementary move.*

Proof. Let $\beta = \{v_1, \ldots, v_n\}$ and $\gamma = \{w_1, \ldots, w_n\}$. Because γ is a basis, then for each $i \in \{1, \ldots, n\}$ we can write

$$v_i = \sum_{j=1}^{n} a_{ij} w_j,$$

where $a_{ij} \in F$ for all appropriate $j \in \{1, \ldots, n\}$.

We start by letting $\alpha_0 = \beta$.

Next, we claim that there is some $k \in \{1, \ldots, n\}$ such that the coefficient of w_1 in v_k is not zero; that is, we claim that $a_{k1} \neq 0$ for some $k \in \{1, \ldots, n\}$. To see why, assume to the contrary that $a_{i1} = 0$ for all $i \in \{1, \ldots, n\}$. Then each element of β can be written as a linear combination of $\{w_2, \ldots, w_n\}$. In other words, we see that $\beta \subseteq \operatorname{span}(\{w_2, \ldots, w_n\})$. Because β is a basis for V, we know $\operatorname{span}(\beta) = V$. It then follows from Exercise 3.4.5 (c) that $\operatorname{span}(\{w_2, \ldots, w_n\}) = V$. We now have a contradiction to Lemma 3.6.9 (1), because $\dim(V) = n$. We therefore deduce that there is some $k \in \{1, \ldots, n\}$ such that $a_{k1} \neq 0$ (if there is more than one such i, choose one).

We now define α_1 to be the result of taking α_0 and switching v_1 and v_k, which is a Type 1 elementary move. To avoid overly cumbersome notation, we will now redefine $\{v_1, \ldots, v_n\}$ so that they now denote the elements of α_1. At each stage of our process, where we define the α_r in terms of α_{r-1} for each $r \in \{1, \ldots, p\}$, we will at each stage redefine $\{v_1, \ldots, v_n\}$ so that they now denote the elements of α_r. (The alternative would be to write $\alpha_r = \{v_1^r, \ldots, v_n^r\}$, and the like, and that would be hard to read.)

By construction, we know that in α_1, the coefficient of w_1 in v_1 is non-zero; that is, we have $a_{11} \neq 0$. We then define α_2 to be the result of taking α_1 and multiplying v_1 by $(a_{11})^{-1}$, which is a Type 2 elementary move.

By construction, we know that in α_2, the coefficient of w_1 in v_1 is 1; that is, we have $a_{11} = 1$. We now look at the coefficient of w_1 in v_2. If the coefficent, which is a_{21}, is zero, then we do nothing to v_2 at this point. If $a_{21} \neq 0$, then we define α_3 to be the result of taking α_2 and adding $-a_{21}v_1$ to v_2, which is a Type 3 elementary move.

By construction, we know that in α_3, the coefficient of w_1 in v_2 is zero. We continue in this way, examining the coefficients of w_1 in all the v_i in turn, and doing Type 3 elementary moves as necessary until we obtain α_m, for some $m \in \mathbb{N}$, in which the coefficient of w_1 in v_1 is 1, and the coefficient of w_1 in all the other v_i is zero. That is, in α_m we have

$$v_1 = w_1 + \sum_{j=2}^{n} a_{ij} w_j,$$

and for all $i \in \{2, \ldots, n\}$ we have

$$v_i = \sum_{j=2}^{n} a_{ij} w_j.$$

We next turn to the coefficients of the w_2. We claim that there is some $k \in \{2, \ldots, n\}$ such that the coefficient of w_2 in v_k is not zero; that is, we claim that $a_{k2} \neq 0$ for some $k \in \{2, \ldots, n\}$. To see why, suppose to the contrary that $a_{i2} = 0$ for all $i \in \{2, \ldots, n\}$. Then

$$v_i = \sum_{j=3}^{n} a_{ij} w_j$$

and for all $i \in \{2, \ldots, n\}$.

We now claim that $w_2 \notin \operatorname{span}(\beta)$. Once we prove that, we will have reached a contradiction to the fact that β is a basis for V, and we will therefore have completed our proof of the fact that that there is some $i \in \{k, \ldots, n\}$ such that $a_{ik} \neq 0$.

To prove that $w_2 \notin \operatorname{span}(\beta)$, suppose to the contrary that there are $b_1, \ldots, b_n \in F$ such that $w_2 = b_1 v_1 + \cdots + b_n v_n$. Then we have

$$w_2 = b_1 w_1 + b_1 a_{12} w_2 + \left[\sum_{p=1}^{n} b_p a_{p(k+1)}\right] w_{k+1} + \cdots + \left[\sum_{p=1}^{n} b_p a_{pn}\right] w_n.$$

We thus have written w_2 as a linear combination of the members of γ. On the other hand, we also have $w_k = 0w_1 + 1w_2 + 0w_3 + \cdots 0w_n$. Theorem 3.6.2 (2) states that each element of V can be written uniquely as a linear combination of elements of γ, and hence we deduce that $b_1 = 0$ and $b_1 a_{12} = 1$. We have reached a contradiction, because Lemma 3.2.7 (5) says that $0a_{12} = 0$. We have therefore proved that $w_2 \notin \operatorname{span}(\beta)$, as claimed above. We have therefore completed the proof of the claim that there is some $k \in \{2, \ldots, n\}$ such that $a_{k2} \neq 0$.

We now continue analogously to what we did previously. We define α_{m+1} to be the result of taking α_m and switching v_2 and v_k, which is a Type 1 elementary move, so that in α_{m+1}, the coefficient of w_2 in v_2 is non-zero. We then define α_{m+2} to be the result of taking α_{m+1} and multiplying v_2 by $(a_{22})^{-1}$, which is a Type 2 elementary move. Next, we look at the coefficient of w_2 in each of the v_i other than v_2, and perform Type 3 elementary moves until we have the coefficient of w_2 in all the v_i other than v_2 is zero. Call the resulting basis α_r for some $r \in \mathbb{N}$. In α_r we then have

$$v_1 = w_1 + \sum_{j=3}^{n} a_{ij} w_j,$$

and

$$v_2 = w_2 + \sum_{j=3}^{n} a_{ij} w_j,$$

and for all $i \in \{3, \ldots, n\}$ we have

$$v_i = \sum_{j=3}^{n} a_{ij} w_j.$$

We continue in this way, performing one elementary move at a time, until we obtain a basis α_p for some $p \in \mathbb{N}$ such that $v_i = w_i$ for all $i \in \{1, \ldots, n\}$. Hence $\alpha_p = \gamma$, and the proof is complete. □

Lemma 6.1.10. *Let V and W be finite-dimensional vector spaces over a field F, let β be an ordered basis for V, let γ be an ordered basis for W, and let $f : V \to W$ be a linear map.*

1. *If β' is a basis for V that is obtained from β by a single elementary move $\mathcal{E}$, then $[f]_{\beta'}^{\gamma}$ can be obtained from $[f]_{\beta}^{\gamma}$ by $\mathcal{E}$ applied to the columns of $[f]_{\beta}^{\gamma}$.*

2. *If γ' is a basis for W that is obtained from γ by a single elementary move $\mathcal{G}$, then $[f]_{\beta}^{\gamma'}$ can be obtained from $[f]_{\beta}^{\gamma}$ by $\mathcal{G}^{O}$ applied to the rows of $[f]_{\beta}^{\gamma}$.*

Proof.

(1). Let $\beta = \{v_1, \ldots, v_n\}$, and let β' be an ordered basis for V that is obtained from β by a single elementary move $\mathcal{E}$. We have three cases, depending upon the type of elementary move used. Let $j \in \{1, \ldots, n\}$. We know by Remark 5.5.2 that the j-th column of $[f]_{\beta}^{\gamma}$ is just $[f(v_j)]_{\gamma}$, and similarly for $[f]_{\beta'}^{\gamma}$.

Type 1: Suppose $\mathcal{E} = \mathcal{E}_1(i,k)$ for some $i, k \in \{1, \ldots, n\}$ such that $i \neq k$. Then $\beta' = \{v_1, \ldots, v_{i-1}, v_k, v_{i+1}, \ldots, v_{k-1}, v_i, v_{k+1}, \ldots, v_n\}$. Let $j \in \{1, \ldots, n\}$. It is clear that if $j \neq i$ and $j \neq k$, then the j-th column of $[f]_{\beta'}^{\gamma}$ is the same as the j-th column of $[f]_{\beta}^{\gamma}$. It is also evident that the i-th column of $[f]_{\beta'}^{\gamma}$ is the same as the k-th column of $[f]_{\beta}^{\gamma}$, and that the k-th column of $[f]_{\beta'}^{\gamma}$ is the same as the i-th column of $[f]_{\beta}^{\gamma}$. Hence $[f]_{\beta'}^{\gamma}$ is obtained from $[f]_{\beta}^{\gamma}$ by $\mathcal{E}$ applied to the columns of $[f]_{\beta}^{\gamma}$.

Type 2: Suppose $\mathcal{E} = \mathcal{E}_2(i; a)$ for some $i \in \{1, \ldots, n\}$ and $a \in F$ such that $a \neq 0$. Then$\beta' = \{v_1, \ldots, v_{i-1}, av_i, v_{i+1}, \ldots, v_n\}$. Let $j \in \{1, \ldots, n\}$. It is clear that if $j \neq i$, then the j-th column of $[f]_{\beta'}^{\gamma}$ is the same as the j-th column of $[f]_{\beta}^{\gamma}$. By using Theorem 5.4.5, it is also seen that the i-th column of $[f]_{\beta'}^{\gamma}$ is $[f(av_i)]_{\gamma} = [af(v_i)]_{\gamma} = a[f(v_i)]_{\gamma}$, which is a times the i-th column of $[f]_{\beta}^{\gamma}$. Hence $[f]_{\beta'}^{\gamma}$ is obtained from $[f]_{\beta}^{\gamma}$ by $\mathcal{E}$ applied to the columns of $[f]_{\beta}^{\gamma}$.

Type 3: Suppose $\mathcal{E} = \mathcal{E}_3(k, i; a)$ for some $i, k \in \{1, \ldots, n\}$ and $a \in F$. Then $\gamma = \{v_1, \ldots, v_{i-1}, v_i + av_k, v_{i+1}, \ldots, v_n\}$ for some $i, k \in \{1, \ldots, n\}$ such that $i \neq k$, and some $a \in F$. Let $j \in \{1, \ldots, n\}$. It is clear that if $j \neq i$, then the j-th column of $[f]_{\beta'}^{\gamma}$ is the same as the j-th column of $[f]_{\beta}^{\gamma}$. By using Theorem 5.4.5, it is also seen that the i-th column of $[f]_{\beta'}^{\gamma}$ is $[f(v_i + av_k)]_{\gamma} = [f(v_i) + af(v_k)]_{\gamma} = [f(v_i)]_{\gamma} + a[f(v_k)]_{\gamma}$, which is a times the k-th column of $[f]_{\beta}^{\gamma}$ added to the i-th column of $[f]_{\beta}^{\gamma}$. Hence $[f]_{\beta'}^{\gamma}$ is obtained from $[f]_{\beta}^{\gamma}$ by $\mathcal{E}$ applied to the columns of $[f]_{\beta}^{\gamma}$.

(2). Let $\gamma = \{w_1, \ldots, w_m\}$, and let γ' be an ordered basis for W that is obtained from γ by a single elementary move $\mathcal{G}$. We have three cases, depending upon the type of elementary move used. Let $r \in \{1, \ldots, n\}$. We know by Remark 5.5.2 that the r-th column of $[f]_{\beta}^{\gamma}$ is just $[f(v_r)]_{\gamma}$, and similarly for $[f]_{\beta}^{\gamma'}$. Let $[f]_{\beta}^{\gamma} = \left(a_{ij}\right)$, and hence $f(v_r) = a_{1r}w_1 + \cdots + a_{mr}w_m$.

Type 1: Suppose $\mathcal{G} = \mathcal{E}_1(i,k)$ for some $i, k \in \{1, \ldots, n\}$ such that $i \neq k$. Then $\gamma' = \{w_1, \ldots, w_{i-1}, w_k, w_{i+1}, \ldots, w_{k-1}, w_i, w_{k+1}, \ldots, w_m\}$. Let $j \in \{1, \ldots, m\}$. It is clear that if $j \neq i$ and $j \neq k$, then the j-th row of $[f]_\beta^{\gamma'}$ is the same as the j-th row of $[f]_\beta^\gamma$. It is also evident that the i-th row of $[f]_\beta^{\gamma'}$ is the same as the k-th row of $[f]_\beta^\gamma$, and that the k-th row of $[f]_\beta^{\gamma'}$ is the same as the i-th row of $[f]_\beta^\gamma$. Hence $[f]_\beta^{\gamma'}$ is obtained from $[f]_\beta^\gamma$ by $\mathcal{G}$ applied to the columns of $[f]_\beta^\gamma$. For a Type 1 elementary move, observe that $\mathcal{G} = \mathcal{G}^O$.

Type 2: Suppose $\mathcal{G} = \mathcal{E}_2(i; a)$ for some $i \in \{1, \ldots, n\}$ and $a \in F$ such that $a \neq 0$. Then$\gamma' = \{w_1, \ldots, w_{i-1}, aw_i, w_{i+1}, \ldots, w_n\}$. Let $r \in \{1, \ldots, n\}$. Because $f(v_r) = a_{1r}w_1 + \cdots + a_{mr}w_m$, we therefore have $f(v_r) = a_{1r}w_1 + \cdots + a_{(i-1)r}w_{i-1} + (a_{ir}a^{-1})(aw_i) + a_{(i+1)r}w_{i+1} + \cdots + a_{mr}w_m$. Let $j \in \{1, \ldots, m\}$. It is now seen that if $j \neq i$, then the j-th row of $[f]_\beta^{\gamma'}$ is the same as the j-th row of $[f]_\beta^\gamma$. It is also seen that the i-th row of $[f]_\beta^{\gamma'}$ is a^{-1} times the i-th row of $[f]_\beta^\gamma$. Hence $[f]_\beta^{\gamma'}$ is obtained from $[f]_\beta^\gamma$ by $\mathcal{G}^O$ applied to the rows of $[f]_\beta^\gamma$.

Type 3: Suppose $\mathcal{G} = \mathcal{E}_3(k, i; a)$ for some $i, k \in \{1, \ldots, n\}$ and $a \in F$. Then $\gamma' = \{w_1, \ldots, w_{i-1}, w_i + aw_k, w_{i+1}, \ldots, w_m\}$. Let $r \in \{1, \ldots, n\}$. Because $f(v_r) = a_{1r}w_1 + \cdots + a_{mr}w_m$, we therefore have $f(v_r) = a_{1r}w_1 + \cdots + a_{(k-1)r}w_{k-1} + (a_{kr} - aa_{ir})w_k + a_{(i+1)r}w_{i+1} + \cdots + a_{(i-1)r}w_{i-1} + a_{ir}(w_i + aw_k) + a_{(i+1)r}w_{i+1} + \cdots + a_{mr}w_m$. Let $j \in \{1, \ldots, m\}$. It is now seen that if $j \neq k$, then the j-th row of $[f]_\beta^{\gamma'}$ is the same as the j-th row of $[f]_\beta^\gamma$. It is also seen that the k-th row of $[f]_\beta^{\gamma'}$ is $-a$ times the i-th row of $[f]_\beta^\gamma$ added to the k-th row of $[f]_\beta^\gamma$. Hence $[f]_\beta^{\gamma'}$ is obtained from $[f]_\beta^\gamma$ by $\mathcal{G}^O$ applied to the rows of $[f]_\beta^\gamma$. □

Exercises

Exercise 6.1.1. Let V be a finite-dimensional vector space over a field F, and let β and γ be finite ordered subsets of V. Suppose that γ can be obtained from β by a Type 1 elementary move. Prove that γ can be obtained from β by three Type 3 elementary moves followed by one Type 2 elementary move.

Exercise 6.1.2. Let V and W be finite-dimensional vector spaces over a field F, let β be an ordered basis for V, let γ be an ordered basis for W, and let $f : V \to W$ be a linear map.

(1) Let B be the matrix obtained from $[f]_\beta^\gamma$ by a single elementary column operation $\mathcal{E}$. If β' is the basis for V obtained from β by $\mathcal{E}$, then $B = [f]_{\beta'}^\gamma$.

(2) Let C be the matrix obtained from $[f]_\beta^\gamma$ by a single elementary row operation $\mathcal{G}$. If γ' is the basis for W obtained from γ by $\mathcal{G}^O$, then $C = [f]_\beta^{\gamma'}$.

6.2 Elementary Matrices

Definition 6.2.1. Let F be a field. Let $E \in \mathsf{M}_{n\times n}(F)$ be a matrix. The matrix E is an **elementary matrix** of Type 1, Type 2 or Type 3, respectively, if E can be obtained from the identity matrix by a single elementary column or row operation of Type 1, Type 2 or Type 3, respectively. △

Lemma 6.2.2. *Let V be a finite-dimensional vector space over a field F, and let β and γ be ordered bases for V. Suppose that β can be obtained from γ by a single elementary move $\mathcal{E}$.*

1. *The matrix $[1_V]^\gamma_\beta$ can be obtained from the identity matrix by $\mathcal{E}$ applied to the columns of the identity matrix.*
2. *The matrix $[1_V]^\gamma_\beta$ can be obtained from the identity matrix by $\mathcal{E}^A$ applied to the rows of the identity matrix.*

Proof.

(1). We know from Lemma 5.5.3 (4) that $[1_V]^\gamma_\gamma = I$. Because β is obtained from γ by $\mathcal{E}$, we can apply Lemma 6.1.10 (1) to deduce that he matrix $[1_V]^\gamma_\beta$ can be obtained from $[1_V]^\gamma_\gamma$ by $\mathcal{E}$ applied to the columns of $[1_V]^\gamma_\gamma$.

(2). We know from Lemma 5.5.34 that $[1_V]^\beta_\beta = I$. Because β is obtained from γ by $\mathcal{E}$, it follows from Lemma 6.1.7 that γ is obtained from β by $\mathcal{E}^R$. We can then apply Lemma 6.1.10 (2) to deduce that he matrix $[1_V]^\gamma_\beta$ can be obtained from $[1_V]^\beta_\beta$ by the $(\mathcal{E}^R)^O$ applied to the rows of $[1_V]^\beta_\beta$. By Lemma 6.1.6 (4) we know that $(\mathcal{E}^R)^O = \mathcal{E}^A$. □

Lemma 6.2.3. *Let F be a field. Let $E \in \mathsf{M}_{n\times n}(F)$ be a matrix. Let V be a vector space over F, and let γ be an ordered basis for V.*

1. *Suppose that E is obtained from the identity matrix by a single elementary column operation $\mathcal{E}$. If β is obtained from γ by $\mathcal{E}$, then $E = [1_V]^\gamma_\beta$.*
2. *Suppose that E is obtained from the identity matrix by a single elementary row operation $\mathcal{G}$. If β is obtained from γ by $\mathcal{G}^A$, then $E = [1_V]^\gamma_\beta$.*

Proof. This lemma follows immediately from Lemma 6.2.2, together with Lemma 6.1.6 (2). □

Corollary 6.2.4. *Let F be a field. Let $E \in \mathsf{M}_{n\times n}(F)$ be a matrix. Then E is an elementary matrix if and only if for any finite-dimensional vector space V over F, and any ordered basis γ for V, the matrix E is the change of basis matrix that changes β-coordinates into γ-coordinates, where β is obtained from γ by a single elementary move.*

Lemma 6.2.5. *Let F be a field. Let $A \in \mathsf{M}_{m\times n}(F)$.*

1. *Let $E \in \mathsf{M}_{n\times n}(F)$ be the matrix obtained by performing a single elementary column operation $\mathcal{E}$ to I_n. Let $B \in \mathsf{M}_{m\times n}(F)$. Then B is obtained from A by $\mathcal{E}$ applied to the columns of A if and only if $B = AE$.*

2. *Let $G \in \mathsf{M}_{n\times n}(F)$ be the matrix obtained by performing a single elementary row operation $\mathcal{G}$ to I_n. Let $C \in \mathsf{M}_{m\times n}(F)$. Then C is obtained from A by $\mathcal{G}$ applied to the rows of A if and only if $C = GA$.*

Proof. Let β be the standard ordered basis for F^n, and let γ be the standard ordered basis for F^m. By Lemma 5.6.3 (1) we know that $[\mathsf{L}_A]_\beta^\gamma = A$.

(1). Let β' be obtained from β by $\mathcal{E}$.

Clearly $\mathsf{L}_A = \mathsf{L}_A \circ 1_{F^n}$. By Theorem 5.6.1 we deduce that $[\mathsf{L}_A]_{\beta'}^\gamma = [\mathsf{L}_A]_\beta^\gamma [1_{F^n}]_{\beta'}^\beta$. By Lemma 6.2.2 (1) we know that the elementary matrix $[1_{F^n}]_{\beta'}^\beta$ is obtained from the identity matrix by $\mathcal{E}$ applied to the columns of the identity matrix. That is, we see that $[1_{F^n}]_{\beta'}^\beta = E$. Hence $[\mathsf{L}_A]_{\beta'}^\gamma = AE$.

By Lemma 6.1.10 (1) we know that $[\mathsf{L}_A]_{\beta'}^\gamma$ is obtained from $[\mathsf{L}_A]_\beta^\gamma$ by $\mathcal{E}$ applied to the columns of $[\mathsf{L}_A]_\beta^\gamma$. That is, we know $[\mathsf{L}_A]_{\beta'}^\gamma$ is obtained from A by $\mathcal{E}$ applied to the columns of A.

It follows that B is obtained from A by $\mathcal{E}$ applied to the columns of A if and only if $B = [\mathsf{L}_A]_{\beta'}^\gamma$ if and only if $B = AE$.

(2). Let γ' be obtained from γ by $\mathcal{G}^O$. Then by Lemma 6.1.7 we know that γ is obtained from γ' by $(\mathcal{G}^O)^R$.

Clearly $\mathsf{L}_A = 1_{F^m} \circ \mathsf{L}_A$. By Theorem 5.6.1 we deduce that $[\mathsf{L}_A]_\beta^{\gamma'} = [1_{F^m}]_\gamma^{\gamma'} [\mathsf{L}_A]_\beta^\gamma$. By Lemma 6.2.2 (2) we know that the elementary matrix $[1_{F^m}]_\gamma^{\gamma'}$ is obtained from the identity matrix by $((\mathcal{G}^O)^R)^A$ applied to the rows of the identity matrix, which is the same as $\mathcal{G}$ applied to the rows of the identity matrix by Lemma 6.1.6 (4) and (2). That is, we see that $[1_{F^m}]_\gamma^{\gamma'} = G$. Hence $[\mathsf{L}_A]_\beta^{\gamma'} = GA$.

By Lemma 6.1.10 (2) we know that $[\mathsf{L}_A]_\beta^{\gamma'}$ can be obtained from $[\mathsf{L}_A]_\beta^\gamma$ by $(\mathcal{G}^O)^O$ applied to the rows of $[\mathsf{L}_A]_\beta^\gamma$, which is the same as $\mathcal{G}$ applied to the rows of $[\mathsf{L}_A]_\beta^\gamma$ by Lemma 6.1.6 (3). That is, we know that $[\mathsf{L}_A]_\beta^{\gamma'}$ is obtained from A by $\mathcal{G}$ applied to the rows of A.

It follows that C is obtained from A by $\mathcal{G}$ applied to the rows of A if and only if $C = [\mathsf{L}_A]_\beta^{\gamma'}$ if and only if $C = GA$. □

Lemma 6.2.6. *Let F be a field. Let $E \in \mathsf{M}_{n\times n}(F)$ be an elementary matrix.*

1. *E is invertible.*

2. *If E is obtained from the identity matrix by an elementary column (respectively row) operation $\mathcal{E}$, then E^{-1} is the elementary matrix obtained from the identity matrix by the elementary column (respectively row) operation $\mathcal{E}^R$.*

Proof. Suppose that E is obtained from the identity matrix by an elementary column (respectively row) operation $\mathcal{E}$. Let V be a vector space over F, and let γ be an ordered basis for V, and let β be obtained from γ by $\mathcal{E}$ (respectively $\mathcal{E}^A$). Then by Lemma 6.2.3 we know that $E = [1_V]_\beta^\gamma$.

(1). Lemma 5.9.1 (1) implies that E is invertible.

(2). By Lemma 5.9.4 (2) we know that $E^{-1} = [1_V]_\gamma^\beta$. It follows from Lemma 6.1.7 that γ can be obtained from β by $\mathcal{E}^R$ (respectively $(\mathcal{E}^A)^R$). Then by Lemma 6.2.2 we know that E^{-1} is obtained from the identity matrix by $\mathcal{E}^R$ applied to the columns of the identity matrix (respectively $((\mathcal{E}^A)^R)^A$ applied to the rows of the identity matrix, and observe that $((\mathcal{E}^A)^R)^A = \mathcal{E}^R$ by Lemma 6.1.6 (5) and (2)). □

Exercises

Exercise 6.2.1. Let F be a field. Let $A \in \mathsf{M}_{m\times n}(F)$. Prove that A can be transformed into an upper triangular matrix by a finite sequence of Type 1 and Type 3 elementary row operations.

Exercise 6.2.2. Find a linear map $f : \mathbb{R}^2 \to \mathbb{R}^2$ such that there is a basis β such that $[f]_\beta^\beta$ is a Type 1 elementary matrix, and such that there is another basis γ such that $[f]_\gamma^\gamma$ is a Type 2 elementary matrix.

Exercise 6.2.3. Let V be a finite-dimensional vector space over a field F, let β be an ordered basis for V, and let $f : V \to V$ be a linear map. Then $[f]_\beta^\beta$ is an elementary matrix if and only if $f(\beta)$ can be obtained from β by an elementary move.

6.3 Rank of a Matrix

Friedberg-Insel-Spence, 4th ed. – Section 3.2

Definition 6.3.1. Let F be a field. Let $A \in \mathsf{M}_{m\times n}(F)$.

1. The **column rank** of A, denoted $\operatorname{columnrank} A$, is the dimension of the span of the columns of A in F^m
2. The **row rank** of A, denoted $\operatorname{rowrank} A$, is the dimension of the span of the rows of A in F^n. △

Definition 6.3.2. Let F be a field. Let $A \in \mathsf{M}_{m\times n}(F)$. The **rank** of A, denoted $\operatorname{rank} A$, is the column rank of A. △

Lemma 6.3.3. *Let F be a field. Let $A \in \mathsf{M}_{n\times n}(F)$. Then* $\operatorname{rank} \mathsf{L}_A = \operatorname{rank} A$.

Proof. Note that $\operatorname{rank} \mathsf{L}_A = \dim(\operatorname{im} A)$, and note that L_A is a map $F^n \to F^m$. Let $\beta = \{e_1, \ldots, e_n\}$be the standard ordered basis for F^n. Then $\operatorname{im} \mathsf{L}_A = \operatorname{span}\{\mathsf{L}_A(\beta)\} = \operatorname{span}\{Ae_1, \ldots, Ae_n\}$. Note that Ae_i is the ith column of A for all $i \in \{1, \ldots, n\}$. Then $\operatorname{im} \mathsf{L}_A$ is the span of the columns of A. Hence $\operatorname{rank} \mathsf{L}_A$ is the dimension of the span of the columns of A. □

Lemma 6.3.4. *Let V, W be vector spaces over a field F, and suppose that V and W are finite dimensional. Let β be an ordered basis for V, and let γ be an ordered basis for W. Let $f : V \to W$ be a linear map. Then* $\operatorname{rank} f = \operatorname{rank}[f]_\beta^\gamma$.

Proof. (We follow [Ber92, pp. 99–100].) Look at the commutative diagram in Remark 5.8.2. Using that notation, and by Theorem 5.8.1 (2), we have $\mathsf{L}_{\Phi(f)} \circ \phi_\beta = \phi_\gamma \circ f$. Then $\operatorname{rank}(\mathsf{L}_{\Phi(f)} \circ \phi_\beta) = \operatorname{rank}(\phi_\gamma \circ f)$. By Theorem 5.4.5 we know that ϕ_β and ϕ_γ are isomorphisms. It now follows from Lemma 4.4.12 that $\operatorname{rank} \mathsf{L}_{\Phi(f)} = \operatorname{rank} f$, and then use the definition of $\Phi(f)$ and Lemma 6.3.3. □

Lemma 6.3.5. *Let F be a field. Let $A \in \mathsf{M}_{m\times n}(F)$, let $B \in \mathsf{M}_{n\times p}(F)$, let $C \in \mathsf{M}_{q\times m}(F)$, let $P \in \mathsf{M}_{m\times m}(F)$ and let $Q \in \mathsf{M}_{n\times n}(F)$. Suppose that P and Q are invertible.*

1. $\operatorname{rank} AQ = \operatorname{rank} A$.
2. $\operatorname{rank} PA = \operatorname{rank} A$.
3. $\operatorname{rank} PAQ = \operatorname{rank} A$.
4. $\operatorname{rank} AB \leq \operatorname{rank} A$.
5. $\operatorname{rank} AB \leq \operatorname{rank} B$.

Proof. We prove Part (1); the remaining parts of this lemma are left to the reader in Exercise 6.3.2.

(1). By Corollary 5.7.2 (1) we see that L_Q is an isomorphism. We compute $\operatorname{rank} AQ = \operatorname{rank}(\mathsf{L}_{AQ}) = \operatorname{rank}(\mathsf{L}_A \circ \mathsf{L}_Q) = \operatorname{rank}(\mathsf{L}_A) = \operatorname{rank} A$, where the first equality is by Lemma 6.3.3, the second equality is by Lemma 5.2.2 (5), the third equality is by Lemma 4.4.12 (1), and the fourth equality is by Lemma 6.3.3. □

Lemma 6.3.6. *Let F be a field. Let $A \in \mathsf{M}_{m\times n}(F)$. Let $B \in \mathsf{M}_{m\times n}(F)$ be obtained from A by performing an elementary row or column operation. Then* $\operatorname{rank} B = \operatorname{rank} A$.

Proof. Combine Lemma 6.2.5, Lemma 6.2.6 and Lemma 6.3.5. □

Theorem 6.3.7. *Let F be a field. Let $A \in \mathsf{M}_{m\times n}(F)$. Suppose that* $\operatorname{rank} A = r$. *Then there exist matrices $P \in \mathsf{M}_{m\times m}(F)$ and $Q \in \mathsf{M}_{n\times n}(F)$ such that P and Q are invertible, and that*

$$PAQ = \begin{bmatrix} I_r & O \\ O & O \end{bmatrix},$$

where O denotes the appropriate zero matrices.

Proof. Let β be the standard ordered basis for F^n, and let γ be the standard ordered basis for F^m. Then by Lemma 5.6.3 (1) we know that $[\mathsf{L}_A]_\beta^\gamma = A$. By Exercise 5.5.4 there is an ordered basis α for F^n and an ordered basis δ for F^m such that $[\mathsf{L}_A]_\alpha^\delta$ has the form

$$[\mathsf{L}_A]_\alpha^\delta = \begin{bmatrix} I_r & O \\ O & O \end{bmatrix},$$

where O denotes the appropriate zero matrices, for some $r \in \{0, 1, \ldots, n\}$. Now, by Lemma 6.3.3 and Lemma 6.3.4, we know that

$$\operatorname{rank} A = \operatorname{rank} \mathsf{L}_A = \operatorname{rank}[\mathsf{L}_A]_\alpha^\delta = \operatorname{rank}\begin{bmatrix} I_r & O \\ O & O \end{bmatrix} = r.$$

Let Q be the change of coordinate matrix that changes α-coordinates into β-coordinates, and let P be the change of coordinate matrix that changes γ-coordinates into δ-coordinates. We know from Lemma 5.9.1 (1) that Q and P are invertible. By Lemma 5.9.4 (2) we know that P^{-1} is the change of coordinate matrix that changes δ-coordinates into γ-coordinates. It now follows from Theorem 5.9.5 that $[\mathsf{L}_A]_\alpha^\delta = (P^{-1})^{-1}[\mathsf{L}_A]_\beta^\gamma Q = P[\mathsf{L}_A]_\beta^\gamma Q$. Combining this last fact with previous observations, the proof is complete. □

Lemma 6.3.8. *Let F be a field. Let $A \in \mathsf{M}_{m\times n}(F)$. Then* $\operatorname{rank} A^t = \operatorname{rank} A$.

Proof. (This proof follows Friedberg-Insel-Spence, 4th ed.) By Theorem 6.3.7, we know that there are invertible matrices $P \in \mathsf{M}_{m\times m}(F)$ and $Q \in \mathsf{M}_{n\times n}(F)$ such that

$$PAQ = \begin{bmatrix} I_r & O \\ O & O \end{bmatrix}.$$

Let D denote the right hand side of the above equation. It is clear from the simple nature of D that $\operatorname{rank} D^t = \operatorname{columnrank} D^t = \operatorname{columnrank} D = \operatorname{rank} D$. We know that P^{-1} and Q^{-1} are invertible, and hence so are $(P^{-1})^t$ and $(Q^{-1})^t$ by using Lemma 5.1.7. Note that $A = P^{-1}DQ^{-1}$. Then, using Lemma 3.2.5 and Lemma 6.3.5, we have

$$\operatorname{rank} A^t = \operatorname{rank}(Q^{-1})^t D^t (P^{-1})^t = \operatorname{rank} D^t = \operatorname{rank} D = \operatorname{rank} P^{-1}DQ^{-1} = \operatorname{rank} A. \qquad \square$$

Theorem 6.3.9. *Let F be a field. Let $A \in \mathsf{M}_{m\times n}(F)$. Then* $\operatorname{columnrank} A = \operatorname{rowrank} A$.

Proof. By Lemma 6.3.8 we have $\operatorname{rowrank} A = \operatorname{columnrank} A^t = \operatorname{rank} A^t = \operatorname{rank} A = \operatorname{columnrank} A$. □

Remark 6.3.10. It follows from Theorem 6.3.9 that $\operatorname{rank} A = \operatorname{rowrank} A$. ◊

Exercises

Exercise 6.3.1. Let F be a field. Let $A \in \mathsf{M}_{m\times n}(F)$. Prove that $\operatorname{rank} A = 0$ if and only if A is the zero matrix.

Exercise 6.3.2. Prove Lemma 6.3.5 (2), (3), (4) and (5).

Exercise 6.3.3. Let V, W be vector spaces over a field F, and let $f, g \colon V \to W$ be linear maps.

(1) Prove that $\operatorname{im}(f + g) \subseteq \operatorname{im} f + \operatorname{im} g$. (See Definition 3.3.8 for the definition of the sum of two subsets.)

(2) Suppose that W is finite-dimensional. Prove that $\operatorname{rank}(f + g) \leq \operatorname{rank} f + \operatorname{rank} g$.

(3) Let F be a field. Let $A, B \in \mathsf{M}_{m\times n}(F)$. Prove that $\operatorname{rank}(A + B) \leq \operatorname{rank} A + \operatorname{rank} B$.

Exercise 6.3.4. Let F be a field. Let $A \in \mathsf{M}_{m\times n}(F)$. Suppose that $\operatorname{rank} A = m$. Prove that there exists $B \in \mathsf{M}_{n\times m}(F)$ such that $AB = I_m$.

6.4 Invertibility of Matrices

Friedberg-Insel-Spence, 4th ed. – Section 3.2

Corollary 6.4.1. *Let F be a field. Let $A \in \mathsf{M}_{n\times n}(F)$. Then A is invertible if and only if* $\operatorname{rank} A = n$.

Proof. Combine Corollaries 5.7.2, 4.4.4 and Lemma 6.3.3. □

Theorem 6.4.2. *Let F be a field. Let $A \in \mathsf{M}_{n\times n}(F)$.*

(1) The following are equivalent.

(a) There exists $B \in \mathsf{M}_{n\times n}(F)$ such that $AB = I_n$.

(b) There exists $C \in \mathsf{M}_{n\times n}(F)$ such that $CA = I_n$.

(c) A is invertible.

(2)

(a) If a matrix $B \in \mathsf{M}_{n\times n}(F)$ satisfies $AB = I_n$, then A is invertible and $B = A^{-1}$.

(b) If a matrix $C \in \mathsf{M}_{n\times n}(F)$ satisfies $CA = I_n$, then A is invertible and $C = A^{-1}$.

Proof. (We follow [Ber92, pp. 126–127].)

(1). It is clear that (c) implies each of (a) and (b). We will show that (a) implies (c); the proof that (b) implies (c) is similar. Let V be an n-dimensional vector space over F, and let β be an ordered basis for V. By Theorem 5.8.1 we know that there are unique linear maps $f, g\colon V \to V$ such that $\Phi(f) = A$ and $\Phi(g) = B$. By Theorem 5.6.1 and Lemma 5.5.3 (4) we deduce that $\Phi(f \circ g) = \Phi(f)\Phi(g) = AB = I = \Phi(1_V)$. It follows from Lemma 5.5.3 (1) that $f \circ g = 1_V$. Hence g is a right inverse of f. It follows from Corollary 4.4.5 that g is a left inverse of f, which means that $g \circ f = 1_V$. Be applying Φ to both sides of this equation, we deduce that $BA = I$, and hence A has an inverse, and hence is invertible.

(2). This part follows from the proof of Part (1). □

Theorem 6.4.3. *Let F be a field. Let $A \in \mathsf{M}_{n\times n}(F)$. Then A is invertible if and only if A is the product of finitely many elementary matrices.*

Proof. First, suppose that A is the product of finitely many elementary matrices. It follows immediately that A is invertible, because elementary matrices are invertible by Lemma 6.2.6 (1), and the product of finitely many invertible matrices is invertible by Lemma 5.1.7 (1) and induction.

Now suppose that A is invertible. Let β be the standard ordered basis for F^n. By Lemma 5.6.3 (1) we know that $[\mathsf{L}_A]_\beta^\beta = A$. By Corollary 5.7.2 (1) we know that L_A is an isomorphism. Let $\gamma = \mathsf{L}_A(\beta)$. It follows from Lemma 4.4.6 that γ is an ordered basis for F^n. From Exercise 5.7.1 (1) we know that $[\mathsf{L}_A]_\beta^\gamma$ is the identity matrix.

We now use Theorem 6.1.9 to see that there is a finite collection of bases $\beta = \alpha_0, \alpha_1, \ldots, \alpha_p = \gamma$ of F^n such that α_i is obtained from α_{i-1} by an elementary move. Clearly

$$\mathsf{L}_A = \underbrace{1_{F^n} \circ \cdots \circ 1_{F^n}}_{p \text{ times}} \circ \mathsf{L}_A.$$

We then use Theorem 5.6.1 to deduce that

$$[\mathsf{L}_A]_\beta^\beta = [1_{F^n}]_{\alpha_1}^{\alpha_0} \cdots [1_{F^n}]_{\alpha_p}^{\alpha_{p-1}} [\mathsf{L}_A]_\beta^\gamma,$$

and hence

$$A = [1_{F^n}]_{\alpha_1}^{\alpha_0} \cdots [1_{F^n}]_{\alpha_p}^{\alpha_{p-1}} I.$$

Finally, we know by definition that $[1_{F^n}]_{\alpha_i}^{\alpha_{i-1}}$ is an elementary matrix for all $i \in \{1, \ldots, p\}$. We have therefore expressed A as a product of finitely many elementary matrices. □

Corollary 6.4.4. *Let F be a field. Let $A \in \mathsf{M}_{m\times n}(F)$. Suppose that* $\operatorname{rank} A = r$. *The A can be transformed by a finite number of elementary row and column operations into the matrix $D \in \mathsf{M}_{m\times n}(F)$ given by*

$$D = \begin{bmatrix} I_r & O \\ O & O \end{bmatrix},$$

where O denotes the appropriate zero matrices.

Proof. Combine Theorem 6.3.7 and Theorem 6.4.3, and Lemma 6.2.5. □

Definition 6.4.5. Let F be a field. Let $A \in \mathsf{M}_{m\times n}(F)$, and let $B \in \mathsf{M}_{m\times p}(F)$. The **augmented matrix** formed by A and B, denoted $[A|B]$, is the $m \times (n+p)$ matrix formed by the columns of A and B, in that order. △

Remark 6.4.6. Let F be a field. Let $A \in \mathsf{M}_{m\times n}(F)$, let $B \in \mathsf{M}_{m\times p}(F)$, and let $C \in \mathsf{M}_{k\times m}(F)$. Then $C[A|B] = [CA|CB]$ ◊

Theorem 6.4.7. *Let F be a field. Let $A \in \mathsf{M}_{n\times n}(F)$.*

1. *A is invertible if and only if A can be transformed by a finite number of elementary row operations into I_n.*

2. *If A is invertible, then $[A|I_n]$ can be transformed by a finite number of elementary row operations into $[I_n|B]$ for some $B \in \mathsf{M}_{n\times n}(F)$, and then $B = A^{-1}$.*

Proof. We do both parts of the theorem together.

Suppose that A is invertible. By Theorem 6.4.3 we know that A is the product of finitely many elementary matrices. Let $A = E_1 E_2 \cdots E_k$, where $E_1, E_2, \ldots, E_k$ are $n \times n$ elementary matrices. Then $A = E_1 E_2 \cdots E_k I_n$. By Lemma 6.2.6 (1) we know that $E_1, E_2, \ldots, E_k$ are invertible. Hence, using Lemma 5.1.7 (2), we see that $(E_k)^{-1}(E_{k-1})^{-1}\cdots(E_1)^{-1}A = I_n$. By Lemma 6.2.6 (2) we know that $(E_k)^{-1}, (E_{k-1})^{-1} \ldots, (E_1)^{-1}$ are elementary matrices. By Lemma 6.2.2 we can think of each of $(E_k)^{-1}, (E_{k-1})^{-1} \ldots, (E_1)^{-1}$ as obtained by doing an elementary row operation applied to the identity matrix. By Lemma 6.2.5 we see that $(E_k)^{-1}(E_{k-1})^{-1}\cdots(E_1)^{-1}A$ is the result of doing k row operations to A. But $(E_k)^{-1}(E_{k-1})^{-1}\cdots(E_1)^{-1}A = I_n$, so we deduce that A can be transformed by a finite number of elementary row operations into I_n. That proves one of the directions of Part (1).

Moreover, let $B = (E_k)^{-1}(E_{k-1})^{-1}\cdots(E_1)^{-1}$. Hence $BA = I_n$. Then by Theorem 6.4.2(2) implies that $B = A^{-1}$. Using Remark 6.4.6, we see that

$$\begin{aligned}&(E_k)^{-1}(E_{k-1})^{-1}\cdots(E_1)^{-1}[A|I_n]\\&\qquad= [(E_k)^{-1}(E_{k-1})^{-1}\cdots(E_1)^{-1}A|(E_k)^{-1}(E_{k-1})^{-1}\cdots(E_1)^{-1}I_n] = [I_n|B].\end{aligned}$$

By Lemma 6.2.5 we see that $(E_k)^{-1}(E_{k-1})^{-1}\cdots(E_1)^{-1}[A|I_n]$ is the result of doing k row operations to $[A|I_n]$. That proves Part (2).

Next, suppose that A can be transformed by a finite number of elementary row operations into I_n. By Lemma 6.2.5 (2) there are elementary matrices $G_1, G_2, \ldots, G_p$ such that $G_1G_2\cdots G_pA = I_n$. Let $D = G_1G_2\cdots G_p$. Then $DA = I_n$. It follows from Theorem 6.4.2(2) that A is invertible. □

6.5 Linear Equations–Theory

Friedberg-Insel-Spence, 4th ed. – Section 3.3

Definition 6.5.1. Let F be a field. Let $m, n \in \mathbb{N}$. A **system of m linear equations in n unknowns over F** is a system of equations with unknowns $x_1, x_2, \ldots, x_n$ that can be written in the form

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

for some $a_{11}, a_{12}, \ldots, a_{mn} \in F$ and $b_1, b_2, \ldots, b_m \in F$. △

Remark 6.5.2. The system of linear equations given in Equation (1) can be rewritten via matrices as follows. Let

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & a_{m3} & \cdots & a_{mn} \end{bmatrix} \quad \text{and} \quad b = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix} \quad \text{and} \quad x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}.$$

Observe that $A \in \mathsf{M}_{m\times n}(F)$, and $b \in F^m$ and $x \in F^n$. The system of linear equations is equivalent to the single equation $Ax = b$. ◊

Definition 6.5.3. Let F be a field. Let $A \in \mathsf{M}_{m\times n}(F)$, and $b \in F^m$.

1. A **solution** to the equation $Ax = b$ is any vector $y \in F^n$ such that $Ay = b$.
2. The **solution set** of the equation $Ax = b$ is the set of all solutions of the equation.
3. The equation $Ax = b$ is **consistent** if the solution set is not empty.
4. The equation $Ax = b$ is **inconsistent** if the solution set is empty.
5. The equation $Ax = b$ is **homogeneous** if $b = 0$.
6. The equation $Ax = b$ is **non-homogeneous** if $b \neq 0$. △

Theorem 6.5.4. *Let F be a field. Let $A \in \mathsf{M}_{m\times n}(F)$. Let K be the solution set of the homogeneous system of linear equations $Ax = 0$.*

1. *K is a subspace of F^n.*
2. $\dim(K) = n - \operatorname{rank}(A)$.

3. *If $m < n$, then the system of equations has a non-zero solution.*

Proof. The proof is based upon the observation is that $K = \ker \mathsf{L}_A$.

(1). This part of the theorem follows immediately from Lemma 4.2.3 (1) applied to the linear map L_A.

(2). Observe that $\text{nullity}(\mathsf{L}_A) = \dim(K)$. By Lemma 6.3.3 we know that $\text{rank}(\mathsf{L}_A) = \text{rank}(A)$. The Rank-Nullity Theorem (Theorem 4.3.2) says that $\text{nullity}(\mathsf{L}_A) + \text{rank}(\mathsf{L}_A) = \dim(F^n)$, which implies that $\dim(K) + \text{rank}(A) = n$.

(3). Suppose that $m < n$. We know by Remark 6.3.10 that $\text{rank}\, A = \text{rowrank}\, A$. But $\text{rowrank}\, A \leq m < n$, so that $\text{rank}\, A < n$. It follows from Part (2) of this theorem that $\dim(K) > 0$, and therefore K has elements other than 0. □

Theorem 6.5.5. *Let F be a field. Let $A \in \mathsf{M}_{m\times n}(F)$ and let $b \in F^m$. Let K_H be the solution set of the homogeneous system of linear equations $Ax = 0$. If s is any solution to the system of linear equations $Ax = b$, then the solution set of $Ax = b$ is $s + K_H$.*

Theorem 6.5.6. *Let F be a field. Let $A \in \mathsf{M}_{n\times n}(F)$ and let $b \in F^n$.*

1. *If A is invertible, the system of linear equations $Ax = b$ has a unique solution.*

2. *If A is not invertible, the system of linear equations $Ax = b$ has either no solutions or infinitely many solutions*

Proof. By Corollary 5.7.2 (1) we know that A is invertible if and only if L_A is an isomorphism. By Corollary 4.4.4 we know that L_A is an isomorphism if and only if it is injective. By Lemma 4.2.4 we know that L_A is injective if and only if $\ker \mathsf{L}_A = \{0\}$.

Let K_H be the solution set of the homogeneous system of linear equations $Ax = 0$. Observe that $K_H = \ker \mathsf{L}_A$. By Theorem 6.5.5, we know that if s is any solution to the system of linear equations $Ax = b$, then the solution set of $Ax = b$ is $s + K_H$.

First, suppose that A is invertible. Then $K_H = \{0\}$. Moreover, because A is invertible, we know that $x = A^{-1}b$ is a solution. Hence the solution set is $A^{-1}b + \{0\} = \{A^{-1}b\}$. Hence there is a unique solution.

Second, suppose that A is not invertible. Then $K_H \neq \{0\}$. Because $K_H = \ker \mathsf{L}_A$ is a non-trivial subspace of F^n, then it is an infinite set.

If $Ax = b$ has no solution, then there is nothing to prove. Suppose that $Ax = b$ has a solution s. Then the solution set is $s + K_H$, which is infinite. □

Corollary 6.5.7. *Let F be a field. Let $A \in \mathsf{M}_{n\times n}(F)$ and let $b \in F^n$. The system of linear equations $Ax = b$ has a unique solution if and only if A is invertible.*

7 Determinants

7.1 Determinants—the 2×2 Case

Friedberg-Insel-Spence, 4th ed. – Section 4.1

Definition 7.1.1. Let F be a field. Let $A \in \mathsf{M}_{2\times 2}(F)$. Suppose A is given by $A = \left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right]$. The **determinant** of A, denoted $\det A$, is defined by $\det A = \det\left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right] = ad - bc$. △

Theorem 7.1.2. *Let F be a field. Let $A \in \mathsf{M}_{2\times 2}(F)$.*

1. *A is invertible if and only if $\det A \neq 0$.*
2. *If $A = \left[\begin{smallmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{smallmatrix}\right]$ is invertible, then*

$$A^{-1} = \frac{1}{\det A}\begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix}.$$

Exercises

Exercise 7.1.1. Let F be a field. Let $A \in \mathsf{M}_{2\times 2}(F)$. Let B be obtained from A by interchanging the two columns. Prove that $\det B = -\det A$.

Exercise 7.1.2. Let F be a field. Let $\delta\colon \mathsf{M}_{2\times 2}(F) \to F$ be a function that satisfies the following three properties.

1. The map δ is a linear function of each column, when the other column is held fixed.
2. If $A \in \mathsf{M}_{2\times 2}(F)$ and A has two identical columns, then $\delta(A) = 0$.
3. $\delta(I_2) = 1$.

Using only what has been discussed so far in these notes, prove that $\delta(A) = \det A$ for all $A \in \mathsf{M}_{2\times 2}(F)$. Do not use any theorems stated later in these notes.

7.2 Determinants—Axiomatic Characterization

Friedberg-Insel-Spence, 4th ed. – Section 4.2

Definition 7.2.1. Let F be a field, and let $n \in \mathbb{N}$. Let $\delta\colon \mathsf{M}_{n\times n}(F) \to F$ be a function. The function δ is ***n*-linear** if it is linear as a function of each column when the other columns are fixed. That is, if $(a_1|\ldots|a_n) \in \mathsf{M}_{n\times n}(F)$, if $i \in \{1,\ldots,n\}$, if $x \in F^n$ and if $c \in F$, then

$$\delta(a_1|\ldots|a_{i-1}|a_i + x|a_{i+1}|\ldots|a_n) = \delta(a_1|\ldots|a_{i-1}|a_i|a_{i+1}|\ldots|a_n) + \delta(a_1|\ldots|a_{i-1}|x|a_{i+1}|\ldots|a_n)$$

and

$$\delta(a_1|\ldots|a_{i-1}|ca_i|a_i|\ldots|a_n) = c\cdot\delta(a_1|\ldots|a_{i-1}|a_i|a_i|\ldots|a_n).$$ △

Definition 7.2.2. Let F be a field, and let $n \in \mathbb{N}$. Let $\delta\colon \mathsf{M}_{n\times n}(F) \to F$ be a function. The function δ is **alternating** if $\delta(A) = 0$ whenever $A \in \mathsf{M}_{n\times n}(F)$ has two identical adjacent columns. △

Theorem 7.2.3. *Let F be a field. Let $n \in \mathbb{N}$. Then there is a unique function $\delta\colon \mathsf{M}_{n\times n}(F) \to F$ satisfying the following three criteria.*

1. *δ is n-linear.*
2. *δ is alternating.*
3. *$\delta(I_n) = 1$.*

Lemma 7.2.4. *Let F be a field. Let $A \in \mathsf{M}_{n\times n}(F)$, and let $c \in F$.*

1. *If B is obtained from A by interchanging two columns, then $\det B = -\det A$.*
2. *If any two columns of A are identical, then $\det A = 0$.*
3. *If B is obtained from A by adding a scalar multiple of one column to another column, then $\det B = \det A$.*
4. *If A has a column that is entirely zero, then $\det A = 0$.*
5. *If the columns of A are linearly dependent, then $\det A = 0$.*
6. *If $\operatorname{rank} A < n$, then $\det A = 0$.*

Proof.

(1). We first prove the result for interchanging two adjacent columns. Suppose that $A = (a_1|a_2|\cdots|a_n)$, and we interchange columns i and $i+1$. Observe that the alternating property implies that $\det(a_1|\cdots|a_i + a_{i+1}|a_i + a_{i+1}|\cdots|a_n) = 0$, and then use linearity and alternating to derive the result. Next, we show that interchanging any two columns can be obtained by an odd number of interchanges of adjacent columns, which is proved by induction on the distance of the two columns to be interchanged.

(2). If we interchange the two identical columns, on the one hand we do not change the matrix, and on the other hand we obtain negative of the original determinant. The only way out is if the original determinant were zero.

(3). This part is relatively straightforward, using linearity, and Part (2) of this lemma.

(4). This part is straightforward, using linearity to factor out a zero.

(5). Suppose that the columns are linearly dependent. Then there is a column, say a_k, that is a linear combination of the other columns. Hence, we can subtract a linear combination of the other columns from this column to obtain a zero column, without changing the determinant.

(6). This part follows from Part (5) of this lemma and the definition of rank of a matrix. □

Lemma 7.2.5. *Let F be a field. Let $A \in \mathsf{M}_{n\times n}(F)$. If A is upper triangular or lower triangular, then* $\det A$ *is the product of the diagonal elements of A.*

Proof. We outline the proof, omitting some of the details.

Suppose that A is upper triangular or lower triangular.

First, suppose that A has a zero on the diagonal. It can then be seen that $\operatorname{rank} A < n$. By Lemma 7.2.4 (6) it then follows that $\det A = 0$, which is what the product of the diagonal elements equals.

Second, suppose that all the diagonal elements of A are non-zero. Let $d_1, \ldots, d_n$ be the diagonal elements. We then factor out the diagonal elements, resulting in a matrix B that has every diagonal element equal to 1. By the n-linearity of the determinant, we see that $\det A = d_1 \cdots d_n \det B$. It can be seen that by doing appropriate Type 3 column operations on B, we can transform B into I_n. By Lemma 7.2.4 (3) we deduce that $\det B = \det I_n = 1$. It follows that $\det A = d_1 \cdots d_n$. □

Exercises

Exercise 7.2.1. Let F be a field. Let $n \in \mathbb{N}$. Let $\delta, \gamma \colon \mathsf{M}_{n\times n}(F) \to F$ be functions, and let $k \in F$.

(1) Suppose that δ and γ are n-linear. Prove that $\delta + \gamma$ and $\delta - \gamma$ and $k\delta$ are n-linear.

(2) Suppose that δ and γ are alternating. Prove that $\delta+\gamma$ and $\delta-\gamma$ and $k\delta$ are alternating.

Exercise 7.2.2. Let F be a field. Let $n \in \mathbb{N}$. Let $A \in \mathsf{M}_{n\times n}(F)$. Let $\delta\colon \mathsf{M}_{n\times n}(F) \to F$ be defined by $\delta(X) = \det(AX)$ for all $X \in \mathsf{M}_{n\times n}(F)$. Prove that δ is n-linear and alternating.

Exercise 7.2.3. Let F be a field. Let $A \in \mathsf{M}_{n\times n}(F)$, and let $k \in F$. Prove that $\det(kA) = k^n \det A$.

Exercise 7.2.4. Let F be a field. Let $A \in \mathsf{M}_{n\times n}(F)$. For which values of n is it the case that $\det(-A) = \det A$?

Exercise 7.2.5. Let F be a field. Let $A \in \mathsf{M}_{n\times n}(F)$. Suppose that A is given by $A = (a_1|a_2|\cdots|a_n)$, where $a_i \in F^n$ is a column vector for all $\in \{1,\ldots,n\}i$. Let $B \in \mathsf{M}_{n\times n}(F)$ be given by $B = (a_n|a_{n-1}|\cdots|a_1)$. Calculate $\det B$ in terms of $\det A$.

7.3 Determinants—Elementary Matrices and Consequences

Friedberg-Insel-Spence, 4th ed. – Section 4.3

Lemma 7.3.1. *Let F be a field. Let $E \in \mathsf{M}_{n\times n}(F)$ be an elementary matrix.*

1. *If E is obtained from I_n by interchanging two columns, then* $\det E = -1$.
2. *If E is obtained from I_n by multiplying a column by a non-zero scalar k, then* $\det E = k$.
3. *If E is obtained from I_n by adding a scalar multiple of one column to another, then* $\det E = 1$.
4. $\det E^t = \det E$.

Proof. Left to the reader in Exercise 7.3.1. □

Theorem 7.3.2. *Let F be a field. Let $A, B \in \mathsf{M}_{n\times n}(F)$. Then* $\det(AB) = \det A \cdot \det B$.

First Proof of Theorem 7.3.2. We have three cases regarding B.

Case 1: Suppose that B is an elementary matrix. There are now three subcases, depending upon the type of elementary matrix that B is.

Type 1: Suppose that B is obtained from I_n by a Type 1 column operation, which means switching two columns. By Lemma 6.2.5 AB is the result of switching two columns of A. By Lemma 7.3.1 (1) we know that $\det B = -1$. By Lemma 7.2.4 (1) we know that $\det(AB) = -\det A$, and it follows that $\det(AB) = \det A \cdot \det B$.

Type 2: Suppose that B is obtained from I_n by a Type 2 column operation, which means one column is multiplied by a non-zero element $c \in F$. By Lemma 6.2.5 AB is the result of multiplying a column of A by c. By the axioms for the determinant function, we see that $\det B = c \det I_n = c$, and that $\det(AB) = c \det A$. It follows that $\det(AB) = \det A \cdot \det B$.

Type 3: Suppose that B is obtained from I_n by a Type 3 column operation, which means adding a scalar multiple of one column to another column. By Lemma 6.2.5 AB is the result of adding a scalar multiple of one column of A to another column of A. By Lemma 7.3.1 (3) we know that $\det B = 1$. By Lemma 7.2.4 (3) we know that $\det(AB) = \det A$, and it follows that $\det(AB) = \det A \cdot \det B$.

Case 2: Suppose that $\operatorname{rank} B < n$. Hence by Corollary 6.4.1 we know that B is not invertible. By Lemma 6.3.5 (5) we see that $\operatorname{rank} AB \leq \operatorname{rank} B < n$. By Lemma 7.2.4 (6) we deduce that $\det AB = 0$ and $\det B = 0$. Then $\det AB = 0 = \det A \cdot \det B$.

Case 3: Suppose that $\operatorname{rank} B = n$. Hence by Corollary 6.4.1 we know that B is invertible. By Theorem 6.4.3 we see that B is the product of finitely many elementary matrices. Let $B = E_1E_2\cdots E_k$, where $E_1, E_2, \ldots, E_k$ are $n \times n$ elementary matrices. Then by Case 1 we have

$$\begin{aligned}
\det(AB) &= \det(AE_1E_2\cdots E_k)\\
&= \det(AE_1E_2\cdots E_{k-1})\det(E_k)\\
&= \det(AE_1E_2\cdots E_{k-2})\det(E_{k-1})\det(E_k)\\
&\vdots\\
&= \det(A)\det(E_1)\det(E_2)\cdots\det(E_{k-1})\det(E_k)\\
&= \det(A)\det(E_1E_2)\det(E_3)\cdots\det(E_{k-1})\det(E_k)\\
&= \det(A)\det(E_1E_2E_3)\cdots\det(E_{k-1})\det(E_k)\\
&\vdots\\
&= \det(A)\det(E_1E_2E_3\cdots E_k) = \det A\cdot\det B.
\end{aligned}$$

□

Second Proof of Theorem 7.3.2. (We follow [Cur74, pp. 147–148].) There are two cases. First, suppose that $\det A = 1$. Let $\delta\colon \mathsf{M}_{n\times n}(F) \to F$ be defined by $\delta(X) = \det(AX)$ for all $X \in \mathsf{M}_{n\times n}(F)$. By Exercise 7.2.2 we know that δ is n-linear and alternating. Moreover, we have $\delta(I_n) = \det(AI_n) = \det A = 1$. Hence δ satisfies the three criteria in Theorem 7.2.3, and therefore $\delta = \det$. It follows that $\det(AB) = \delta(B) = \det B = 1\cdot\det B = \det A\cdot\det B$.

Next, suppose that $\det A \neq 1$. Let $\gamma\colon \mathsf{M}_{n\times n}(F) \to F$ be defined by

$$\gamma(X) = \frac{\det X - \det(AX)}{1 - \det A}$$

for all $X \in \mathsf{M}_{n\times n}(F)$. By Exercise 7.2.2 and Exercise 7.2.1 we know that γ is n-linear and alternating. Moreover, we have

$$\gamma(I_n) = \frac{\det I_n - \det(AI_n)}{1 - \det A} = \frac{1 - \det A}{1 - \det A} = 1.$$

Hence γ satisfies the three criteria in Theorem 7.2.3, and therefore $\gamma = \det$. It follows that

$$\det B = \gamma(B) = \frac{\det B - \det(AB)}{1 - \det A}.$$

Hence $\det B\cdot(1 - \det A) = \det B - \det(AB)$, and it follows that $\det B\cdot\det A = \det(AB)$. □

Corollary 7.3.3. *Let F be a field. Let $A \in \mathsf{M}_{n\times n}(F)$.*

1. *A is invertible if and only if* $\det A \neq 0$.

2. *If A is invertible, then*

$$\det(A^{-1}) = \frac{1}{\det A}.$$

Proof. First, suppose that A is invertible. Then there is a matrix $A^{-1} \in \mathsf{M}_{n\times n}(F)$ such that $AA^{-1} = I_n = A^{-1}A$. By Theorem 7.3.2 and the definition of the determinant function we deduce that $detA \cdot \det(A^{-1}) = \det(I_n) = 1$. Because $detA$ and $\det(A^{-1})$ are real numbers, it follows that $\det A \neq 0$ and that $\det(A^{-1}) = \frac{1}{\det A}$.

Now suppose that A is not invertible. By Corollary 6.4.1 we deduce that $\operatorname{rank} A < n$. By Lemma 7.2.4 (6) we deduce that $\det A = 0$. □

Corollary 7.3.4. *Let F be a field. Let $A \in \mathsf{M}_{n\times n}(F)$ and let $b \in F^n$. The system of linear equations $Ax = b$ has a unique solution if and only if* $\det A \neq 0$*.*

Proof. Combine Corollary 6.5.7 and Corollary 7.3.3 (1). □

Corollary 7.3.5. *Let F be a field. Let $A, B \in \mathsf{M}_{n\times n}(F)$. Suppose that A and B are similar. Then* $\det A = \det B$*.*

Proof. Because A and B are similar, there is an invertible matrix $Q \in \mathsf{M}_{n\times n}(F)$ such that $A = Q^{-1}BQ$. Using Theorem 7.3.2 and Corollary 7.3.3 we deduce that $\det Q \neq 0$, and that $\det A = \det(Q^{-1}BQ) = \det(Q^{-1}) \cdot \det B \cdot \det Q = \frac{1}{\det Q} \cdot \det B \cdot \det Q = \det B$. □

Theorem 7.3.6. *Let F be a field. Let $A \in \mathsf{M}_{n\times n}(F)$. Then* $\det A^t = \det A$*.*

Proof. First, suppose that A is not invertible. By Corollary 6.4.1 we deduce that $\operatorname{rank} A < n$. By Lemma 6.3.8 we see that $\operatorname{rank} A^t = \operatorname{rank} A < n$. It now follows from Lemma 7.2.4 (6) we deduce that $\det A^t = 0 = \det A$.

Second, suppose that A is invertible. By Theorem 6.4.3 we see that A is the product of finitely many elementary matrices. Let $A = E_1E_2\cdots E_k$, where $E_1, E_2, \ldots, E_k$ are $n \times n$ elementary matrices. By Lemma 7.3.1 (4) we see that $\det(E_i)^t = \det E_i$ for all $i \in \{1, \ldots, k\}$. It follows from Lemma 5.1.12 (2) and Theorem 7.3.2 that

$$\begin{aligned}
\det A^t &= \det(E_1E_2\cdots E_k)^t \\
&= \det[(E_k)^t(E_{k-1})^t\cdots(E_1)^t] \\
&= \det(E_k)^t \cdot \det(E_{k-1})^t \cdots \det(E_1)^t \\
&= \det(E_k) \cdot \det(E_{k-1}) \cdots \det(E_1) \\
&= \det(E_1) \cdot \det(E_2) \cdots \det(E_k) \\
&= \det(E_1E_2\cdots E_k) = \det A.
\end{aligned}$$

□

Definition 7.3.7. Let V be a vector space over a field F, and let $f\colon V \to V$ be a linear map. Suppose that V is finite-dimensional. The **determinant of the linear map** f is defined to be equal to $\det[f]_\beta$, for any ordered basis β for V. △

Exercises

Exercise 7.3.1. Prove Lemma 7.3.1.

Exercise 7.3.2. Let $Q \in \mathsf{M}_{n\times n}(\mathbb{R})$. The matrix Q is an **orthogonal** matrix if $QQ^t = I$.
Prove that if Q is orthogonal, then $\det Q = \pm 1$.

Exercise 7.3.3. Let F be a field. Let $B \in \mathsf{M}_{n\times n}(F)$. Suppose that B is given by $B = (b_1|b_2|\cdots|b_n)$, where $b_i \in F^n$ is a column vector for all i. Assume that $b_i \neq b_j$ when $i \neq j$. Let $\beta = \{b_1, \ldots, b_n\}$. Prove that β is a basis for F^n if and only if $\det B \neq 0$.

7.4 Determinants—Computing

Friedberg-Insel-Spence, 4th ed. – Section 4.2

Definition 7.4.1. Let F be a field. Let $A \in \mathsf{M}_{n\times n}(F)$. Let $i, j \in \{1, \ldots, n\}$.

1. Let $\tilde{A}_{ij}$ be the $(n-1)\times(n-1)$ matrix obtained by deleting the i^{th} row and j^{th} column of A.
2. The ij^{th} **cofactor** of A, denoted A_{ij}, is defined by $A_{ij} = (-1)^{i+j}\det\tilde{A}_{ij}$.
3. The **cofactor matrix** of A, denoted $\operatorname{cof} A$, is the matrix $\left[A_{ij}\right]$. △

Theorem 7.4.2. *Let F be a field. Let $n \in \mathbb{N}$ be such that $n \geq 2$. Let $A \in \mathsf{M}_{n\times n}(F)$. Let $i \in \{1, \ldots, n\}$. Then*

$$\det A = \sum_{k=1}^{n}(-1)^{i+k}a_{ik}\cdot\det(\tilde{A}_{ik}) = \sum_{k=1}^{n} a_{ik}A_{ik}$$

and

$$\det A = \sum_{k=1}^{n}(-1)^{i+k}a_{ki}\cdot\det(\tilde{A}_{ki}) = \sum_{k=1}^{n} a_{ki}A_{ki}.$$

Theorem 7.4.3 (Cramer's Rule). *Let F be a field. Let $A \in \mathsf{M}_{n\times n}(F)$ and let $b \in F^n$. If $\det A \neq 0$, then the system of linear equations $Ax = b$ has a unique solution, which is given by*

$$x_i = \frac{\det M_i}{\det A}$$

for each $i \in \{1, \ldots, n\}$, where $M_i \in \mathsf{M}_{n\times n}(F)$ is obtained by replacing the i^{th} column of A with b.

Proof. Suppose $\det A \neq 0$. By Corollary 7.3.4 we know that the system of linear equations $Ax = b$ has a unique solution. Let $x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$ be that unique solution.

Let $\{e_1, \ldots, e_n\}$ be the standard basis for F^n.

Let $k \in \{1, \ldots, n\}$. Let v_k be the k^{th} column of A. Observe that $Ae_k = v_k$. Let X_k be the result of taking I_n and replacing the k^{th} column by x. Observe that $AX_k = M_k$.

We can find $\det X_k$ by expanding along the k^{th} row, which yields $\det X_k = x_k$. Also, using Theorem 7.3.2, we see that $\det M_k = \det(AX_k) = \det A \cdot \det X_k = \det A \cdot x_k$, and that yields $x_k = \frac{\det M_k}{\det A}$. □

Theorem 7.4.4. *Let F be a field. Let $A \in \mathsf{M}_{n\times n}(F)$. If $\det A \neq 0$, then A is invertible and*

$$A^{-1} = \frac{1}{\det A}(\operatorname{cof} A)^t.$$

Proof. The proof is outlined in Exercise 7.4.1 □

Exercises

Exercise 7.4.1. Let F be a field. Let $A \in \mathsf{M}_{n\times n}(F)$. Suppose $\det A \neq 0$. The purpose of this exercise is to prove that

$$A^{-1} = \frac{1}{\det A}(\operatorname{cof} A)^t.$$

Let $\{e_1, \ldots, e_n\}$ be the standard basis for F^n.
Recall the definition of the cofactor matrix $\operatorname{cof} A = \left[A_{ij}\right]$ of A given in Definition 7.4.1.

(1) Let $j, k \in \{1, \ldots, n\}$. Let $B_k \in \mathsf{M}_{n\times n}(F)$ be obtained by replacing the k^{th} column of A with e_j. Prove that $\det B_k = A_{jk}$.

(2) Let $r \in \{1, \ldots, n\}$. Let D_r be the the r^{th} column of $(\operatorname{cof} A)^t$. Prove that $AD_r = \det A \cdot e_r$. (Hint: Use Cramer's Rule with the system of linear equations $Ax = e_r$.)

(3) Prove that $A(\operatorname{cof} A)^t = \det A \cdot I_n$.

(4) Deduce that

$$A^{-1} = \frac{1}{\det A}(\operatorname{cof} A)^t.$$

Exercise 7.4.2. Let F be a field. Let $A \in \mathsf{M}_{n\times n}(F)$. Suppose that there is some $p \in \{1, \ldots, n-1\}$, and there are matrices $B \in \mathsf{M}_{p\times p}(F)$, and $C \in \mathsf{M}_{p\times(n-p)}(F)$ and $D \in \mathsf{M}_{(n-p)\times(n-p)}(F)$, such that A can be written as

$$A = \begin{bmatrix} B & C \\ O & D \end{bmatrix},$$

where $O \in \mathsf{M}_{(n-p)\times p}(F)$ is the zero matrix. Prove that $\det A = \det B \cdot \det D$.

7.5 Determinants—Proof of Theorem 7.2.3 and Theorem 7.4.2

Friedberg-Insel-Spence, 4th ed. – Section 4.5

Proof of Theorem 7.2.3 and Theorem 7.4.2. **Step 1**: We start with the uniqueness part of Theorem 7.2.3. Here we follow [Cur74, pp. 140–141]. Let $n \in \mathbb{N}$, and let $\delta, \gamma \colon \mathsf{M}_{n\times n}(F) \to F$ satisfying the three criteria listed in Theorem 7.2.3. We will show that $\delta = \gamma$. Define $\Delta\colon \mathsf{M}_{n\times n}(F) \to F$ by $\Delta(A) = \delta(A) - \gamma(A)$ for all $A \in \mathsf{M}_{n\times n}(F)$. We will show that Δ is constantly zero, and that will imply that $\delta = \gamma$.

We can easily deduce some elementary properties of Δ. Because δ and γ are both n-linear and alternating, it is easy to see that Δ is also n-linear and alternating. Moreover, we can apply Lemma 7.2.4 to each of δ and γ, and we can then deduce that if $A \in \mathsf{M}_{n\times n}(F)$, and then if B is obtained from A by interchanging two colums, then $\Delta B = -\Delta A$, and if any two columns of A are identical, then $\Delta A = 0$. Finally, because $\delta(I_n) = 1 = \gamma(I_n)$, it follows that $\Delta(I_n) = 0$.

We can think of Δ as a function of n column vectors in F^n. If $A \in \mathsf{M}_{n\times n}(F)$, and if A can be written as columns $(a_1|\cdots|a_n)$, then we will write $\Delta(A)$ as $\Delta(a_1, \ldots, a_n)$. As always, let $e_1, \ldots, e_n$ denote the standard basis for F^n. We then see that $I_n = (e_1|\cdots|e_n)$, and hence $\Delta(e_1, \ldots, e_n) = 0$. Next, suppose that $k_1, \ldots, k_n \in \{1, \ldots, n\}$. We claim that $\Delta(e_{k_1}, \ldots, e_{k_n}) = 0$. There are two cases to look at. If the numbers $k_1, \ldots, k_n$ are not all distinct, then the matrix $(e_{k_1}|\cdots|e_{k_n})$ has at least two identical columns, and in that case we know $\Delta(e_{k_1}, \ldots, e_{k_n}) = 0$. On the other hand, suppose that the numbers $k_1, \ldots, k_n$ are all distinct. Then $k_1, \ldots, k_n$ can be obtained by rearranging the numbers $1, \ldots, n$. In that case, the matrix $(e_{k_1}|\cdots|e_{k_n})$ is obtained from the identity matrix by a finite number of column interchanges. If follows that $\Delta(e_{k_1}, \ldots, e_{k_n}) = \pm\Delta(e_1, \ldots, e_n) = 0$. Thus we have proved the claim.

Finally, suppose that we have $A \in \mathsf{M}_{n\times n}(F)$. We write A as $(a_1|\cdots|a_n)$. For each $j \in \{1, \ldots, n\}$, we can write $a_j = \sum_{k=1}^n c_{jk} e_k$, for some scalars c_{kj}. Then, using the n-linearity of Δ, we see that

$$\begin{aligned}\Delta(a_1, \ldots, a_n) &= \Delta(\sum_{k_1=1}^n c_{1k_1} e_{k_1}, \ldots, \sum_{k_n=1}^n c_{nk_n} e_{k_n})\\ &= \sum_{k_1=1}^n \cdots \sum_{k_n=1}^n c_{1k_1} \cdots c_{nk_n} \Delta(e_{k_1}, \ldots, e_{k_n}) = 0.\end{aligned}$$

We now see that Δ is constantly zero, and that proves uniqueness.

Step 2: We now simultaneously show the existence part of Theorem 7.2.3 and all of Theorem 7.4.2. Here we follow follow [Lan66, pp. 96–98]. For this part we will leave out the details. We proceed by induction on n.

Base Case: It is easy to define determinants in the 1×1 and the 2×2 cases. It is trivial to see that the definition of the determinant in the 1×1 case satisfies the three properties listed in Theorem 7.2.3, and we know from Section 7.1 that the definition of the determinant in the 2×2 case satisfies the three properties listed in Theorem 7.2.3.

Inductive Step: Let $n \in \mathbb{N}$. Suppose that $n \geq 3$, and that determinants have been defined in the $(n-1)\times(n-1)$ case, in a way that satisfies the three properties in Theorem 7.2.3, and also satisfies Theorem 7.4.2.

Let $i \in \{1, \ldots, n\}$. We then define maps $\delta_i, \gamma_i \colon \mathsf{M}_{n\times n}(F) \to F$ as follows. If $A \in \mathsf{M}_{n\times n}(F)$, then let

$$\delta_i(A) = \sum_{k=1}^{n} (-1)^{i+k} a_{ik} \cdot \det(\tilde{A}_{ik}) \quad \text{and} \quad \gamma_i(A) = \sum_{k=1}^{n} (-1)^{j+k} a_{kj} \cdot \det(\tilde{A}_{kj}).$$

With a bit of work, it can be shown that δ_i and γ_i satisfy the three properties listed in Theorem 7.2.3. We will skip those details, leaving them to the reader.

We now know by Step 1 of this proof that $\delta_1, \ldots, \delta_n, \gamma_1, \ldots, \gamma_n$ are all equal. We then define the $n \times n$ determinant to be the function $\det \colon \mathsf{M}_{n\times n}(F) \to F$ given by $\det(A) = \delta_i(A) = \gamma_i(A)$ for any $i \in \{1, \ldots, n\}$, where $A \in \mathsf{M}_{n\times n}(F)$. It now follows immediately that the $n \times n$ determinant satisfies all three properties of Theorem 7.2.3, and that Theorem 7.4.2 holds as well. $\square$

8 Eigenvalues

8.1 Eigenvalues

Friedberg-Insel-Spence, 4th ed. – Section 5.1

Definition 8.1.1. Let F be a field.

1. Let V be a vector space over F, and let $f : V \to V$ be a linear map. Let $v \in V$. The vector v is an **eigenvector** of f if $v \neq 0$ and $f(v) = \lambda v$ for some $\lambda \in F$; the scalar λ is the**eigenvalue** of f corresponding to v.

2. Let $A \in \mathsf{M}_{n\times n}(F)$. Let $v \in F^n$. The vector v is an **eigenvector** of A if $v \neq 0$ and $Av = \lambda v$ for some $\lambda \in F$; the scalar λ is the**eigenvalue** of A corresponding to v. △

Lemma 8.1.2. *Let F be a field. Let $A \in \mathsf{M}_{n\times n}(F)$. Then $\lambda \in F$ is an eigenvalue of A if and only if* $\det(A - \lambda I_n) = 0$.

Proof. Let $\lambda \in F$. Then λ is an eigenvalue of A if and only if there is some non-zero vector $v \in F^n$ such that $Av = \lambda v$, which is true if and only if $(A - \lambda I_n)v = 0$. But, we know that $(A - \lambda I_n)0 = 0$, so there is a non-zero vector $v \in F^n$ such that $(A - \lambda I_n)v = 0$ if and only if the system of linear equations $(A - \lambda I_n)x = 0$ has more than one solution, which, by Corollary 7.3.4, is true if and only if $\det(A - \lambda I_n) = 0$. □

Definition 8.1.3. Let F be a field. Let $A \in \mathsf{M}_{n\times n}(F)$. The **characteristic polynomial** of A is $\det(A - xI_n)$. △

Remark 8.1.4. Let F be a field. Let $A \in \mathsf{M}_{n\times n}(F)$. The eigenvalues of A are precisely the roots of the characteristic polynomial of A. ◊

Lemma 8.1.5. *Let F be a field. Let $A \in \mathsf{M}_{n\times n}(F)$.*

1. *The characteristic polynomial of A has degree n, and leading coefficient* $(-1)^n$.

2. *A has at most n distinct eigenvalues.*

Proof. The proof of this lemma is straightforward, and we omit the details. □

Theorem 8.1.6. *Let $n \in \mathbb{N}$. Let $A \in \mathsf{M}_{n\times n}(\mathbb{R})$. If n is odd, then A has at least one eigenvalue.*

Lemma 8.1.7. *Let F be a field. Let $A \in \mathsf{M}_{n\times n}(F)$. If A is upper-triangular or lower-triangular, then the eigenvalues of A are the diagonal elements of A.*

Proof. Observe that the matrix $A - \lambda I_n$ is upper-triangular or lower-triangular-. The result then follows straightforwardly from Lemma 7.2.5. □

Lemma 8.1.8. *Let F be a field. Let $A \in \mathsf{M}_{n\times n}(F)$, and let $\lambda \in F$ be an eigenvalue of A. Let $v \in F^n$. Then v is an eigenvector for λ if and only if $v \neq 0$ and $(A - \lambda I_n)v = 0$.*

Proof. Trivial. □

Lemma 8.1.9. *Let F be a field. Let $A, B \in \mathsf{M}_{n\times n}(F)$. Suppose that A and B are similar. Then A and B have the same characteristic polynomials, and the same eigenvalues.*

Proof. It is left to the reader in Exercise 8.1.5 to show that A and B have the same characteristic polynomials. Because the eigenvalues of a matrix are just the roots of the characteristic polynomial, then there is nothing more to prove. □

Lemma 8.1.10. *Let V be a vector space over a field F, and let $f : V \to V$ be a linear map. Suppose that V is finite-dimensional. Let β and β' be ordered bases for V. Then $[f]_\beta$ and $[f]_{\beta'}$ have the same characteristic polynomials, and the same eigenvalues.*

Proof. The result follows immediately from Corollary 5.9.10 and Lemma 8.1.9. □

Definition 8.1.11. Let V be a vector space over a field F, and let $f : V \to V$ be a linear map. Suppose that V is finite-dimensional. The **characteristic polynomial** of f is the characteristic polynomial of the matrix $[f]_\beta$ for any ordered basis β of V. △

Theorem 8.1.12. *Let V be a vector space over a field F. Suppose that V is finite-dimensional. Let $n = \dim(V)$. Let β be an ordered basis for V. Let $f : V \to V$ be a linear map. Let $\lambda \in F$.*

1. *Let $v \in V$. Then v is an eigenvector of f with eigenvalue λ if and only if $[v]_\beta$ is an eigenvector of the matrix $[f]_\beta$ with eigenvalue λ.*

2. *Let $y \in F^n$. There is a unique $u \in V$ such that $[u]_\beta = y$. Then y is an eigenvector of the matrix $[f]_\beta$ with eigenvalue λ if and only if u is an eigenvector of f with eigenvalue λ.*

Proof. We will prove Part (1); the other part is similar, and we omit the details.

(1). First, suppose that $v \in V$ is an eigenvalue of f with eigenvalue λ. Then $f(v) = \lambda v$. By Theorem 5.6.2 we see that $[f(v)]_\beta = [f]_\beta[v]_\beta$. Hence $[\lambda v]_\beta = [f]_\beta[v]_\beta$. By Theorem 5.4.5 we know that ϕ_β is a linear map, and from that we deduce that $\lambda[v]_\beta = [f]_\beta[v]_\beta$, and that means that $[v]_\beta$ is an eigenvector of the matrix $[f]_\beta$ with eigenvalue λ.

Second, suppose that $[v]_\beta$ is an eigenvector of the matrix $[f]_\beta$ with eigenvalue λ. Then $[f]_\beta[v]_\beta = \lambda[v]_\beta$. As before we deduce that $[f(v)]_\beta = [\lambda v]_\beta$. By Theorem 5.4.5 we know that ϕ_β is injective, and from that we deduce that we deduce that $f(v) = \lambda v$, and that means that v is an eigenvector of f with eigenvalue λ. □

Corollary 8.1.13. *Let V be a vector space over a field F, and let $f : V \to V$ be a linear map. Suppose that V is finite-dimensional. Let β be an ordered basis for V. Then the eigenvectors and eigenvalues of f are the same as the eigenvalues and eigenvectors of the matrix $[f]_\beta$.*

Lemma 8.1.14. *Let V be a vector space over a field F, let $f : V \to V$ be a linear map, and let $\lambda \in F$ be an eigenvalue of f. Let $v \in V$. Then v is an eigenvector for λ if and only if $v \neq 0$ and $v \in \ker(f - \lambda 1_V)$.*

Proof. Trivial. □

Exercises

Exercise 8.1.1. Let $A = \begin{bmatrix} 1 & 1 \\ -3 & 5 \end{bmatrix}$. Find the eigenvalues of A, and find an eigenvector for each eigenvalue.

Exercise 8.1.2. Let $B = \begin{bmatrix} 3 & -1 & 0 \\ 0 & 2 & 0 \\ -1 & 1 & 2 \end{bmatrix}$. Find the eigenvalues of B, and find an eigenvector for each eigenvalue.

Exercise 8.1.3. Let $\Omega\colon \mathbb{R}_2[x] \to \mathbb{R}_2[x]$ be defined by $\Omega(f) = (2x+1)f' + x^2 f''$ for all $f \in \mathbb{R}_2[x]$. Find the eigenvalues of Ω.

Exercise 8.1.4. Let V be a finite-dimensional vector space over a field F, and let $f\colon V \to V$ be a linear map.

(1) Prove that f is an isomorphism if and only if $\mathbf{0}$ is not an eigenvalue of f.

(2) Suppose that f is an isomorphism. Prove that $\lambda \in F$ is an eigenvalue of f if and only if λ^{-1} is an eigenvalue of f^{-1}.

Exercise 8.1.5. Let F be a field. Let $A, B \in \mathsf{M}_{n\times n}(F)$. Suppose that A and B are similar. Prove that A and B have the same characteristic polynomial.

8.2 Multiplicity of Eigenvalues

Friedberg-Insel-Spence, 4th ed. – Section 5.2

Definition 8.2.1. Let F be a field. Let $f \in F[x]$, and let $r \in F$.

1. The element r is a **root** of f (also called **zero** of f) if $f(r) = 0$.
2. Suppose that r is a root of f. Let $k \in \mathbb{N}$. The root r has **multiplicity** k if $(x - r)^k$ is a factor of f, and if $(x - r)^s$ is not a root of s for any $s \in \mathbb{N}$ such that $s > k$. △

△

Definition 8.2.2. Let F be a field.

1. Let V be a vector space over F, and let $f : V \to V$ be a linear map. Suppose that V is finite-dimensional. Let λ be an eigenvalue of f. The **multiplicity** of λ as an eigenvalue of f is its multiplicity as a root of the characteristic polynomial of f.
2. Let $A \in \mathsf{M}_{n\times n}(F)$. Let λ be an eigenvalue of A. The **multiplicity** of λ as an eigenvalue of A is its multiplicity as a root of the characteristic polynomial of A. △

Definition 8.2.3. Let F be a field.

1. Let V be a vector space over F, and let $f : V \to V$ be a linear map. Let λ be an eigenvalue of f. The **eigenspace** of λ, denoted E_λ, is the set
$$\mathsf{E}_\lambda = \{x \in V \mid x \text{ is an eigenvector for } \lambda \text{ or } x = 0\}.$$
2. Let $A \in \mathsf{M}_{n\times n}(F)$. Let λ be an eigenvalue of A. The **eigenspace** of λ, denoted E_λ, is the set
$$\mathsf{E}_\lambda = \{x \in F^n \mid x \text{ is an eigenvector for } \lambda \text{ or } x = 0\}.$$

△

Lemma 8.2.4. *Let F be a field.*

1. *Let V be a vector space over F, and let $f : V \to V$ be a linear map. Let λ be an eigenvalue of f.*
   1. $\mathsf{E}_\lambda = \ker(f - \lambda 1_V)$.
   2. E_λ *is a subspace of V.*
2. *Let $A \in \mathsf{M}_{n\times n}(F)$. Let λ be an eigenvalue of A.*
   1. E_λ *is the solution set of the homogeneous system of linear equations $(A - \lambda I_n)v = 0$.*

2. *E_λ is a subspace of F^n.*

Proof.

(1). This part of the lemma follows from Lemma 8.1.14 and Lemma 4.2.3 (1).

(2). This part of the lemma follows from Lemma 8.1.8 and Lemma 6.5.4 (1). □

Lemma 8.2.5. *Let F be a field.*

1. *Let V be a vector space over F, and let $f: V \to V$ be a linear map. Suppose that V is finite-dimensional. Let λ be an eigenvalue of f. If λ has multiplicity m, then $1 \le \dim(\mathsf{E}_\lambda) \le m$.*

2. *Let $A \in \mathsf{M}_{n\times n}(F)$. Let λ be an eigenvalue of A. If λ has multiplicity m, then $1 \le \dim(\mathsf{E}_\lambda) \le m$.*

Proof. We prove Part (2) of the lemma; the other part is very similar, but it uses the matrix representation of f, and we omit the details.

(2). Let $p = \dim(\mathsf{E}_\lambda)$. It is evident that $p \ge 1$, because λ must have an eigenvector, which is by definition not the zero vector. Let $\{v_1, \ldots, v_p\}$ be an ordered basis for E_λ. Then $\{v_1, \ldots, v_p\}$ is linear independent, and by Corollary 3.6.9 (5) $\{v_1, \ldots, v_p\}$ can be extended to a basis $\beta = \{v_1, \ldots, v_p, v_{p+1}, \ldots, v_n\}$ of F^n. Clearly $p \le n$. Note that $Av_i = \lambda v_i$ for all $i \in \{1, \ldots, p\}$.

First, suppose that $p = n$. Then $A = \lambda I_n$. Then the characteristic polynomial of A is $(\lambda - x)^n$, so that λ has multiplicity n. That is, we have $m = n$. It also follows that $A - \lambda I_n$ is the zero matrix. By Lemma 8.2.4 we deduce that $\mathsf{E}_\lambda = F^n$, and hence $\dim(\mathsf{E}_\lambda) = n = m$. Hence $\dim(\mathsf{E}_\lambda) = m$, so that $\dim(\mathsf{E}_\lambda) \le m$.

Now suppose $p < n$. It is then seen that the matrix A has the form

$$A = \begin{bmatrix} \lambda I_p & B \\ O & C \end{bmatrix},$$

where $B \in \mathsf{M}_{p\times(n-p)}(F)$ and $C \in \mathsf{M}_{(n-p)\times(n-p)}(F)$, and where $O \in \mathsf{M}_{(n-p)\times p}(F)$ is the zero matrix. Then $[f]_\beta - xI_n$ has the form

$$A - xI_n = \begin{bmatrix} \lambda I_p - xI_p & B \\ O & C - xI_{n-p} \end{bmatrix}.$$

It now follows from Exercise 7.4.2 that the characteristic polynomial of A is

$$\det(A - xI_n) = \det(\lambda I_p - xI_p) \cdot \det(C - xI_{n-p}) = (\lambda - x)^p \cdot \det(C - xI_{n-p}).$$

We deduce that $(\lambda - x)^p$ is a factor of the characteristic polynomial, which means that the multiplicity of λ is at least p. Hence $p \le m$. □

Theorem 8.2.6. *Let F be a field.*

1. *Let V be a vector space over F, and let $f\colon V \to V$ be a linear map. Let $\lambda_1, \ldots, \lambda_k \in F$ be distinct eigenvalue of f. Let $v_i \in \mathsf{E}_{\lambda_i} - \{0\}$ for all $i \in \{1, \ldots, k\}$. Then $v_1, \ldots, v_k$ are linearly independent.*

2. *Let $A \in \mathsf{M}_{n\times n}(F)$. Let $\lambda_1, \ldots, \lambda_k \in F$ be distinct eigenvalue of A. Let $v_i \in \mathsf{E}_{\lambda_i} - \{0\}$ for all $i \in \{1, \ldots, k\}$. Then $v_1, \ldots, v_k$ are linearly independent.*

Proof. We prove Part (2) of the lemma; the other part is very similar, but it uses the matrix representation of f, and we omit the details.

(2). The proof is by induction on k.

Base Case: Suppose that $k = 1$. It follows from Lemma 3.5.6 (2) that the single vector v_1 is linearly independent.

Inductive Step: suppose that $k \geq 2$, and that the result is true for $k-1$. Let $a_1, \ldots, a_k \in F$. Suppose that

$$a_1 v_1 + \cdots + a_k v_k = 0. \tag{1}$$

Observe that if $i \in \{1, \ldots, k-1\}$, then $(A-\lambda_k I_n)v_i = (\lambda_i - \lambda_k)v_i$, and that $(A-\lambda_k I_n)v_k = 0$. Multiplying both sides of Equation (1) by $A - \lambda_k I_n$ yields

$$a_1(\lambda_1 - \lambda_k)v_1 + \cdots + a_{k-1}(\lambda_{k-1} - \lambda_k)v_{k-1} + 0 = 0.$$

By the inductive hypothesis we know that $v_1, \ldots, v_{k-1}$ are linearly independent. It follows that $a_i(\lambda_i - \lambda_k) = 0$ for all $i \in \{1, \ldots, k-1\}$. Because $\lambda_1, \ldots, \lambda_k$ are distinct, we know that $\lambda_i - \lambda_k \neq 0$ for all $i \in \{1, \ldots, k-1\}$. It follows that $a_i = 0$ for all $i \in \{1, \ldots, k-1\}$. Equation (1) then reduces to $a_k v_k = 0$, and because $v_k \neq 0$, it follows that $a_k = 0$. We deduce that $v_1, \ldots, v_k$ are linearly independent. □

Corollary 8.2.7. *Let F be a field.*

1. *Let V be a vector space over F, and let $f\colon V \to V$ be a linear map. Let $\lambda_1, \ldots, \lambda_k \in F$ be distinct eigenvalue of f. Let $S_i \subseteq \mathsf{E}_{\lambda_i}$ be a finite linearly independent set for all $i \in \{1, \ldots, k\}$. Then $S_1 \cup \cdots \cup S_k$ is linearly independent.*

2. *Let $A \in \mathsf{M}_{n\times n}(F)$. Let $\lambda_1, \ldots, \lambda_k \in F$ be distinct eigenvalue of f. Let $S_i \subseteq \mathsf{E}_{\lambda_i}$ be a finite linearly independent set for all $i \in \{1, \ldots, k\}$. Then $S_1 \cup \cdots \cup S_k$ is linearly independent.*

Proof. We prove Part (2) of the lemma; the other part is very similar, and we omit the details.

(2). For each $i \in \{1, \dots, k\}$, let $S_i = \{v_1^i, \dots, v_{r_i}^i\}$. Then

$$S_1 \cup \cdots \cup S_k = \{v_1^1, \dots, v_{r_1}^1, \dots, v_1^k, \dots, v_{r_k}^k\}.$$

Let $c_1^1, \dots, c_{r_1}^1, \dots, c_1^k, \dots, c_{r_k}^k \in F$. Suppose that

$$c_1^1 v_1^1 + \cdots + c_{r_1}^1 v_{r_1}^1 + \cdots + c_1^k v_1^k + \cdots + c_{r_k}^k v_{r_k}^k = 0.$$

For each $i \in \{1, \dots, k\}$, let $w_i = c_1^i v_1^i + \cdots + c_{r_i}^i v_{r_i}^i$. Then $w_1 + \cdots + w_k = 0$. By Lemma 8.2.4 (2), we know that E_{λ_i} is a subspace of F^n. Because $S_i \subseteq \mathsf{E}_{\lambda_i}$, it follows that $w_i \in \mathsf{E}_{\lambda_i}$. We now use Exercise 8.2.5 to deduce that $w_i = 0$ for all $i \in \{1, \dots, k\}$.

Let $i \in \{1, \dots, k\}$. Because $w_i = 0$, we see that $c_1^i v_1^i + \cdots + c_{r_i}^i v_{r_i}^i = 0$. Because $\{v_1^i, \dots, v_{r_i}^i\}$ is linearly independent, then $c_1^i = 0, \dots, c_{r_i}^i = 0$.

It now follows that $S_1 \cup \cdots \cup S_k$ is linearly independent. ◻

Corollary 8.2.8. *Let F be a field.*

1. *Let V be a vector space over F, and let $f : V \to V$ be a linear map. If $\lambda_1, \dots, \lambda_k \in F$ are all the distinct eigenvalue of f, then*

$$\sum_{i=1}^{k} \dim(\mathsf{E}_{\lambda_i}) \le n.$$

2. *Let $A \in \mathsf{M}_{n\times n}(F)$. If $\lambda_1, \dots, \lambda_k \in F$ are all the distinct eigenvalue of A, then*

$$\sum_{i=1}^{k} \dim(\mathsf{E}_{\lambda_i}) \le n.$$

Proof. The proofs of the two parts are identical.

For each $i \in \{1, \dots, k\}$, let B_i be a basis for E_{λ_i}. Then B_i is linearly independent for all $i \in \{1, \dots, k\}$. By Corollary 8.2.7 (1) we know $B_1 \cup \cdots \cup B_k$ is linearly independent. It follows from Corollary 3.6.9 (3) that $|B_1 \cup \cdots \cup B_k| \le n$. However, we also see that $|B_1 \cup \cdots \cup B_k| = \sum_{i=1}^{k} |B_i| = \sum_{i=1}^{k} \dim(\mathsf{E}_{\lambda_i})$, which completes the proof. ◻

Exercises

Exercise 8.2.1. Let $A = \begin{bmatrix} 1 & 1 \\ -3 & 5 \end{bmatrix}$. Find the eigenspace for each eigenvalue of A.

Exercise 8.2.2. Let $B = \begin{bmatrix} 3 & -1 & 0 \\ 0 & 2 & 0 \\ -1 & 1 & 2 \end{bmatrix}$. Find the eigenspace for each eigenvalue of B.

Exercise 8.2.3. Let V be a finite-dimensional vector space over a field F, and let $f\colon V \to V$ be a linear map. Suppose that f has distinct eigenvalues $\lambda_1, \ldots, \lambda_k$ with multiplicities $m_1, \ldots, m_k$ respectively. Suppose that β is a basis for V such that $[f]_\beta$ is an upper triangular matrix. Prove that the diagonal entries of $[f]_\beta$ are $\lambda_1, \ldots, \lambda_k$, and that λ_i occurs m_i times on the diagonal for $i \in \{1, \ldots, k\}$.

Exercise 8.2.4. Let V be a finite-dimensional vector space over a field F, and let $f\colon V \to V$ be a linear map. Suppose that f is an isomorphism. Let $\lambda \in F$ be an eigenvalue of f. By Exercise 8.1.4, we know that λ^{-1} is an eigenvalue of f^{-1}. Prove that the eigenspace of f corresponding to λ is the same as the eigenspace of f^{-1} corresponding to λ^{-1}.

Exercise 8.2.5. Let F be a field. Let $A \in \mathsf{M}_{n\times n}(F)$. Let $\lambda_1, \ldots, \lambda_k \in F$ be distinct eigenvalue of A. Let $v_i \in \mathsf{E}_{\lambda_i}$ for all $i \in \{1, \ldots, k\}$. Prove that if $v_1 + \cdots + v_k = 0$, then $v_i = 0$ for all $i \in \{1, \ldots, k\}$.

8.3 Diagonalizability

Friedberg-Insel-Spence, 4th ed. – Section 5.2

Definition 8.3.1. Let V be a vector space over a field F, and let $f: V \to V$ be a linear map. Suppose that V is finite-dimensional. The linear map f is **diagonalizable** if there is a basis β of V such that $[f]_\beta$ is a diagonal matrix. △

Theorem 8.3.2. *Let V be a vector space over a field F, and let $f: V \to V$ be a linear map. Suppose that V is finite-dimensional. The following are equivalent.*

a. *f is diagonalizable.*

b. *There is an ordered basis for V consisting of eigenvectors of f.*

c. *If $\lambda_1, \ldots, \lambda_k \in F$ are all the distinct eigenvalue of f, then*

$$\sum_{i=1}^{k} \dim(\mathsf{E}_{\lambda_i}) = n.$$

Proof. The equivalence of Part (a) and Part (b) is trivial.

Suppose Part (b) is true. Let β be an ordered basis of eigenvectors of f.

Let $\lambda_1, \ldots, \lambda_k \in F$ be all the distinct eigenvalue of f. Then the ordered basis β can be written as a union $\beta = \beta_1 \cup \cdots \cup \beta_k$, where β_i consists of those elements of β that correspond to the eigenvalue λ_i, for all $i \in \{1, \ldots, k\}$. Clearly $n = |\beta| = |\beta_1 \cup \cdots \cup \beta_k| = \sum_{i=1}^{k} |\beta_i|$.

Let $i \in \{1, \ldots, k\}$. Then β_i is a subset of β, so β_i is linearly independent. Because β_i is a linearly independent subset of E_{λ_i}, then we know by Lemma 3.6.9 (3) that $|\beta_i| \leq \dim(\mathsf{E}_{\lambda_i})$. It then follows that $n = \sum_{i=1}^{k} |\beta_i| \leq \dim(\mathsf{E}_{\lambda_i})$.

On the other hand, we know by Corollary 8.2.8 (1) that $\sum_{i=1}^{k} \dim(\mathsf{E}_{\lambda_i}) \leq n$. We deduce that $\sum_{i=1}^{k} \dim(\mathsf{E}_{\lambda_i}) = n$, which is Part (c).

Now suppose that Part (c) is true. Hence $\sum_{i=1}^{k} \dim(\mathsf{E}_{\lambda_i}) = n$.

For each $i \in \{1, \ldots, k\}$, let γ_i be a basis for E_{λ_i}. Then $\sum_{i=1}^{k} |\gamma_i| = n$.

We know that γ_i is a linearly independent set for all $i \in \{1, \ldots, k\}$. By Corollary 8.2.7 (1) we know that $\gamma_1 \cup \cdots \cup \gamma_k$ is linearly independent. It follows from Lemma 3.6.9 (4) that $\gamma_1 \cup \cdots \cup \gamma_k$ is a basis for V. By definition every element in $\gamma_1 \cup \cdots \cup \gamma_k$ is an eigenvalue of f, and hence f has a basis of eigenvectors, which is Part (b). □

Theorem 8.3.3. *Let V be a vector space over a field F, and let $f: V \to V$ be a linear map. Suppose that V is finite-dimensional. Let $n = \dim(V)$. Suppose that f has n distinct eigenvalues. Then f is diagonalizable.*

Proof. Let $\lambda_1, \ldots, \lambda_n \in F$ be the distinct eigenvalue of f. Let $v_i \in \mathsf{E}_{\lambda_i} - \{0\}$ for all $i \in \{1, \ldots, n\}$. Then by Theorem 8.2.6 (1) we know that $\{v_1, \ldots, v_n\}$ is linearly independent. It follows from Corollary 3.6.9 (4) that $\{v_1, \ldots, v_n\}$ is a basis for V. It now follows from Theorem 8.3.2 that f is diagonalizable. □

Definition 8.3.4. Let F be a field. Let $f \in F[x]$. The polynomial f **splits over** F if there are $c, a_1, a_2 \ldots, a_k \in F$ such that $f = c(x - a_1)(x - a_2) \cdots (x - a_k)$. △

Remark 8.3.5. Let F be a field. Let $p \in F[x]$. Then p splits if and only if the sum of the multiplicities of the roots of p equals the degree of p. ◊

Lemma 8.3.6. *Let V be a vector space over a field F, and let $f : V \to V$ be a linear map. Suppose that V is finite-dimensional. If f is diagonalizable, then the characteristic polynomial of f splits.*

Proof. Suppose that f is diagonalizable. Then there is a basis β of V such that $[f]_\beta$ is a diagonal matrix. Let $n = \dim(V)$. Suppose the diagonal entries of this diagonal matrix are $\lambda_1, \ldots, \lambda_n$. Then the characteristic polynomial of f is the characteristic polynomial of $[f]_\beta$, and it is straightforward to see that that characteristic polynomial is $(\lambda_1 - x) \cdots (\lambda_n - x) = (-1)^n (x - \lambda_1) \cdots (x - \lambda_n)$. Hence the characteristic polynomial of f splits. □

Corollary 8.3.7. *Let V be a vector space over a field F, and let $f : V \to V$ be a linear map. Suppose that V is finite-dimensional. Then f is diagonalizable if and only if the following two conditions hold.*

(a) *The characteristic polynomial of f splits.*

(b) *The multiplicity of each eigenvalue λ of f equals* $\dim(\mathsf{E}_\lambda)$.

Proof. Let p_f denote the characteristic polynomial of f. Let $\lambda_1, \ldots, \lambda_k$ be all the distinct eigenvalue of f. For each $i \in \{1, \ldots, k\}$, let m_i denote the multiplicity of λ_i. By Lemma 8.2.5 (1), we know that $\dim(\mathsf{E}_{\lambda_i}) \leq m_i$ for all $i \in \{1, \ldots, k\}$.

First, suppose that f is diagonalizable. Then by Lemma 8.3.6 we know that p_f splits. Hence Part (a) holds.

By Theorem 8.3.2 we know that $\sum_{i=1}^k \dim(\mathsf{E}_{\lambda_i}) = n$. Because p_f splits, we know by Remark 8.3.5 that $\sum_{i=1}^k m_i$ equals the degree of p_f, and by Lemma 8.1.5 (1) we deduce that $\sum_{i=1}^k m_i = n$. Hence $\sum_{i=1}^k \dim(\mathsf{E}_{\lambda_i}) = \sum_{i=1}^k m_i$. This last equality, combined with the fact that $\dim(\mathsf{E}_{\lambda_i}) \leq m_i$ for all $i \in \{1, \ldots, k\}$, implies that in fact $\dim(\mathsf{E}_{\lambda_i}) = m_i$ for all $i \in \{1, \ldots, k\}$. Hence Part (b) holds.

Now suppose that Part (a) and Part (b) both hold.

By Part (b) we know that $\dim(\mathsf{E}_{\lambda_i}) = m_i$ for all $i \in \{1, \ldots, k\}$.

As before, Part (a) says that p_f splits, and we deduce that $\sum_{i=1}^k m_i = n$. It follows that $\sum_{i=1}^k \dim(\mathsf{E}_{\lambda_i}) = n$.

It now follows from Theorem 8.3.2 that f is diagonalizable. □

Definition 8.3.8. Let F be a field. Let $A \in \mathsf{M}_{n\times n}(F)$. The matrix A is **diagonalizable** if $\mathsf{L}_A : F^n \to F^n$ is diagonalizable. △

Lemma 8.3.9. *Let F be a field. Let $A \in \mathsf{M}_{n\times n}(F)$. Then A is diagonalizable if and only if there is an invertible matrix $Q \in \mathsf{M}_{n\times n}(F)$ such that $Q^{-1}AQ$ is a diagonal matrix.*

Proof. First, suppose that A is diagonalizable. Hence L_A is diagonalizable, which means that there is an ordered basis γ of F^n such that $[\mathsf{L}_A]_\gamma$ is a diagonal matrix. By Corollary 5.9.7 there is an invertible matrix $Q \in \mathsf{M}_{n\times n}(F)$ such that $[\mathsf{L}_A]_\gamma = Q^{-1}AQ$. Hence $Q^{-1}AQ$ is a diagonal matrix.

Second, suppose that there is an invertible matrix $P \in \mathsf{M}_{n\times n}(F)$ such that $P^{-1}AP$ is a diagonal matrix. Let β be the standard ordered basis for F^n. By Lemma 5.9.11 there is an ordered basis β' for F^n such that P is the change of coordinate matrix that changes β'-coordinates into β-coordinates. By Corollary 5.9.6 we know that $[\mathsf{L}_A]_{\beta'} = P^{-1}[\mathsf{L}_A]_\beta P$. We see from Lemma 5.6.3 (1) that $[\mathsf{L}_A]_\beta = A$. Hence $[\mathsf{L}_A]_{\beta'} = P^{-1}AP$, which means that $[\mathsf{L}_A]_{\beta'}$ is a diagonal matrix, which in turn means that L_A is diagonalizable, which means that A is diagonalizable. □

Corollary 8.3.10. *Let F be a field. Let $A \in \mathsf{M}_{n\times n}(F)$. Then A is diagonalizable if and only if the following two conditions hold.*

(a) *The characteristic polynomial of A splits.*

(b) *The multiplicity of each eigenvalue λ of A equals* $\dim(\mathsf{E}_\lambda)$.

Proof. This corollary is just a rephrasing of Corollary 8.3.7, which is straightforward using Lemma 5.6.3. □

Remark 8.3.11. Let F be a field. Let $A \in \mathsf{M}_{n\times n}(F)$. Suppose that A is diagonalizable. To find an invertible matrix $Q \in \mathsf{M}_{n\times n}(F)$ such that $Q^{-1}AQ$ is a diagonal matrix, use the following steps.

(1). Let β be the standard basis for F^n.

(2). Find the eigenvalues of A.

(3). For each eigenvalue λ, find a basis for E_λ.

(4). Assemble all the bases for the eigenspaces into a basis for F^n; call this basis β'.

(5). The matrix Q is the change of coordinate matrix that changes β'-coordinates into β-coordinates. As in Remark 5.9.3, that matrix is formed by writing the elements of β' in terms of β and putting the coordinates of each element of β' in terms of β into a column vector, and assembling these column vectors into a matrix. ◊

Lemma 8.3.12. *Let F be a field. Let $A \in \mathsf{M}_{n\times n}(F)$. Suppose that A is diagonalizable. Let $Q \in \mathsf{M}_{n\times n}(F)$ be an invertible matrix such that $Q^{-1}AQ$ is a diagonal matrix. Let $D = Q^{-1}AQ$. Then $A^n = QD^nQ^{-1}$.*

Proof. Because $D = Q^{-1}AQ$, then $A = QDQ^{-1}$. It follows that

$$A^n = (QDQ^{-1})(QDQ^{-1})\cdots(QDQ^{-1}) = QD^nQ^{-1}.$$ □

Theorem 8.3.13. *Let F be a field.*

1. *Let V be a vector space over F, and let $f: V \to V$ be a linear map. Suppose that V is finite-dimensional. Then f is diagonalizable if and only if V is the direct sum of the eigenspaces of f.*

2. *Let $A \in \mathsf{M}_{n\times n}(F)$. Then A is diagonalizable if and only if F^n is the direct sum of the eigenspaces of A.*

Proof. We omit the proof. It is on pp. 275–278 of Friedberg-Insel-Spence, 4th ed. □

Exercises

Exercise 8.3.1. Let $A = \begin{bmatrix} 1 & 1 \\ -3 & 5 \end{bmatrix}$.

(1) Find an invertible matrix $Q \in \mathsf{M}_{2\times 2}(\mathbb{R})$ such that $Q^{-1}AQ$ is a diagonal matrix.

(2) Use Part (1) of this exercise to find an expression for A^n, where $n \in \mathbb{N}$.

Exercise 8.3.2. For each of the following matrices, determine whether or not the matrix is diagonalizable, and explain why or why not.

(1) Let $A = \begin{bmatrix} 3 & -1 & 0 \\ 0 & 2 & 0 \\ -1 & 1 & 2 \end{bmatrix}$.

(2) Let $A = \begin{bmatrix} 7 & -8 & 6 \\ 8 & -9 & 6 \\ 0 & 0 & -1 \end{bmatrix}$.

Exercise 8.3.3. Use diagonalization to find the general solution of the system of linear ordinary differential equations

$$\begin{aligned} x' &= x + 4y \\ y' &= 2x + 3y. \end{aligned}$$

Exercise 8.3.4. Let F be a field. Let $A \in \mathsf{M}_{n\times n}(F)$. Suppose that A has two distinct eigenvalues $\lambda, \mu \in F$, and that $\dim(\mathsf{E}_\lambda) = n - 1$. Prove that A is diagonalizable.

Exercise 8.3.5. Let V be a vector space over a field F, and let $f, g: V \to V$ be linear maps. Suppose that V is finite-dimensional. We say that f and g are **simultaneously diagonalizable** if there exists an ordered basis β for V such that $[f]_\beta$ and $[g]_\beta$ are both diagonal matrices.

Suppose that f and g are simultaneously diagonalizable. Prove that $g \circ f = f \circ g$.

9 Inner Product Spaces

9.1 Inner Products

Friedberg-Insel-Spence, 4th ed. – Section 6.1

Definition 9.1.1. Let V be a vector space over $\mathbb{R}$. An **inner product** on V is a function $\langle , \rangle \colon V \times V \to \mathbb{R}$ that satisfies the following properties. Let $x, y, z \in V$ and let $c \in \mathbb{R}$

1. $\langle x + y, z\rangle = \langle x, z\rangle + \langle y, z\rangle$.
2. $\langle cx, y\rangle = c\langle x, y\rangle$.
3. $\langle x, y\rangle = \langle y, x\rangle$ (Symmetry Law).
4. if $x \neq 0$ then $\langle x, x\rangle > 0$ (Positive Definite Law). △

Definition 9.1.2. An **inner product space** is a vector space over $\mathbb{R}$ with a specific choice of inner product. △

Lemma 9.1.3. *Let V be an inner product space over $\mathbb{R}$, let $x, y, z \in V$, and let $c \in \mathbb{R}$.*

1. $\langle x, y + z\rangle = \langle x, y\rangle + \langle x, z\rangle$.
2. $\langle x, cy\rangle = c\langle x, y\rangle$.
3. $\langle x, 0\rangle = 0 = \langle 0, x\rangle$.
4. $\langle x, x\rangle = 0$ *if and only if* $x = 0$.
5. *If* $\langle w, y\rangle = \langle w, z\rangle$ *for all* $w \in V$, *then* $y = z$.

Proof. Part (1) and Part (2) follow immediately from the analogous parts of the definition of an inner product, together with the Symmetry Law.

For Part (3), observe that $\langle x, 0\rangle = \langle x, 0 + 0\rangle = \langle x, 0\rangle + \langle x, 0\rangle$, and then use cancelation.

It follows from Part (3) that $\langle 0, 0\rangle = 0$. If $x \neq 0$, then we know that $\langle x, x\rangle > 0$. Those two observations imply Part (4).

For Part (5), suppose that $\langle w, y\rangle = \langle w, z\rangle$ for all $w \in V$. Then $\langle w, y + (-z)\rangle = 0$ for all $w \in V$. In particular, we deduce that $\langle y + (-z), y + (-z)\rangle = 0$. By Part (4) it follows that $y + (-z) = 0$, and that implies that $y = z$. □

Definition 9.1.4. Let V be an inner product space. Let $x \in V$. The **norm** of x, denoted as $\|x\|$, is defined by $\|x\| = \sqrt{\langle x, x\rangle}$. △

Remark 9.1.5. Let V be an inner product space. Let $x \in V$. Then $\|x\|^2 = \langle x, x\rangle$. ◊

Lemma 9.1.6. *Let V be an inner product space over $\mathbb{R}$, let $x, y \in V$, and let $c \in \mathbb{R}$.*

1. $\|cx\| = |c| \cdot \|x\|$.

2. $\|x\| \geq 0$.

3. $\|x\| = 0$ *if and only if* $x = 0$.

4. $|\langle x, y\rangle| \leq \|x\| \cdot \|y\|$ *(Cauchy Schwarz Inequality).*

5. $\|x + y\| \leq \|x\| + \|y\|$ *(Triangle Inequality).*

Proof. Observe that for any real number $a \in \mathbb{R}$, we have $a^2 = |a|^2$ and $\sqrt{a^2} = |a|$.

(1), **(2)**, **(3).** These three part are straightforward, and we omit the details.

(4). There are two cases. First, suppose that that $y = 0$. In that case $\|y\| = 0$, and $\langle x, y\rangle = 0$, so clearly $|\langle x, y\rangle| = 0 = \|x\| \cdot \|y\|$.

Second, suppose that $y \neq 0$. Then $\langle y, y\rangle \neq 0$. Let $b \in \mathbb{R}$. Then $\|x - by\| \geq 0$, and hence

$$0 \leq \|x - by\|^2 = \langle x + (-b)y, x + (-b)y\rangle = \langle x, x\rangle + 2(-b)\langle x, y\rangle + b^2\langle y, y\rangle.$$

That holds for any value of b, and in particular it holds for $b = \frac{\langle x,y\rangle}{\langle y,y\rangle}$, which is defined because $\langle y, y\rangle \neq 0$. We then have

$$0 \leq \langle x, x\rangle - 2\frac{\langle x, y\rangle}{\langle y, y\rangle}\langle x, y\rangle + \frac{\langle x, y\rangle^2}{\langle y, y\rangle^2}\langle y, y\rangle = \|x\|^2 - \frac{|\langle x, y\rangle|^2}{\|y\|^2}.$$

The desired result follows.

(5). Using Part (4), we compute

$$\begin{aligned}\|x + y\|^2 &= \langle x + y, x + y\rangle = \langle x, x\rangle + 2\langle x, y\rangle + \langle y, y\rangle \\ &= \|x\|^2 + 2|\langle x, y\rangle| + \|y\|^2 \\ &\leq \|x\|^2 + 2\|x\| \cdot \|y\| + \|y\|^2 = (\|x\| + \|y\|)^2.\end{aligned}$$

The desired result follows. □

Definition 9.1.7. Let V be an inner product space. Let $x, y \in V$, and let $S \subseteq V$.

1. The vectors x, y are **orthogonal** if $\langle x, y\rangle = 0$.

2. The vector x is a **unit vector** if $\|x\| = 1$.

3. The set S is an **orthogonal set** if $v, w \in S$ and $v \neq w$ implies v, w are orthogonal.

4. The set S is an **orthonormal set** if S is orthogonal and every vector in S is a unit vector. △

Remark 9.1.8. Let V be an inner product space. Let $S = \{v_1, \dots, v_k\}$ be a subset of V. Then S is an orthonormal set if and only if

$$\langle v_i, v_j \rangle = \delta_{ij} = \begin{cases} 1, & \text{if } i = j \\ 0, & \text{if } i \neq j \end{cases}$$

for all $i, j \in \{1, \dots, k\}$. ◊

Remark 9.1.9. Let V be an inner product space. Let $x \in V$. Suppose $x \neq 0$. Then $\frac{x}{\|x\|}$ is a unit vector. Hence, there is a unit vector that is a scalar multiple of the vector x. ◊

Exercises

Exercise 9.1.1. Let $\mathsf{C}([-\pi, \pi])$ denote the set of all continuous functions $[-\pi, \pi] \to \mathbb{R}$. We define an inner product on $\mathsf{C}([-\pi, \pi])$ as follows. Let $\langle f, g \rangle = \int_{-\pi}^{\pi} f(t)g(t)\,dt$ for all $f, g \in \mathsf{C}([-\pi, \pi])$. It can be verified that this definition is indeed an inner product.

Let $a(t) = \sin t$, let $b(t) = t$ and let $c(t) = t$ for all $t \in [-\pi, \pi]$.

(1) Which pairs of a, b and c are orthogonal?

(2) Find $\|b\|$.

(3) Find a unit vector that is a scalar multiple of b.

Exercise 9.1.2. Let V be an inner product space over $\mathbb{R}$, let β be a basis for V, and let $x, y \in V$. Suppose that V is finite-dimensional.

(a) Prove that if $\langle x, b \rangle = 0$ for all $b \in \beta$, then $x = 0$.

(b) Prove that if $\langle x, b \rangle = \langle y, b \rangle$ for all $b \in \beta$, then $x = y$.

Exercise 9.1.3. Let V be an inner product space over $\mathbb{R}$, and let $f\colon V \to V$ be a linear map. Suppose that f preserves the norm on V, that is, suppose that $\|f(x)\| = \|x\|$ for all $x \in V$. Prove that f is injective.

Exercise 9.1.4. Let V be an inner product space over $\mathbb{R}$, and let $x, y \in V$. Prove that

$$\langle x, y \rangle = \frac{1}{4}\|x + y\|^2 - \frac{1}{4}\|x - y\|^2,$$

which is called the **polar identity**.

9.2 Orthonormal Bases

Friedberg-Insel-Spence, 4th ed. – Section 6.2

Definition 9.2.1. Let V be a vector space over a field F, and let $B \subseteq V$. The set B is a **orthogonal basis**, respectively **orthonormal basis**, for V if B is a basis for V and if it is orthogonal, respectively orthonormal, set. △

Lemma 9.2.2. *Let V be an inner product space over $\mathbb{R}$, and let $S = \{v_1, \ldots, v_k\} \subseteq V$. Suppose that $v_i \neq 0$ for all $i \in \{1, \ldots, k\}$. Let $y \in \operatorname{span}(S)$.*

1. *If S is orthogonal, then*
$$y = \sum_{i=1}^{k} \frac{\langle y, v_i\rangle}{\|v_i\|^2} v_i.$$
2. *If S is orthonormal, then*
$$y = \sum_{i=1}^{k} \langle y, v_i\rangle v_i.$$

Proof.

(1). Suppose that S is orthogonal. Because $y \in \operatorname{span}(S)$, there are $c_1, \ldots, c_k \in F$ such that $y = c_1v_1 + \cdots + c_kv_k$.

Let $i \in \{1, \ldots, k\}$. Because S is orthogonal, then $\langle v_i, v_s\rangle = 0$ for all $s \in \{1, \ldots, k\}$ such that $s \neq i$. Then

$$\begin{aligned}\langle y, v_i\rangle &= \langle c_1v_1 + \cdots + c_kv_k, v_i\rangle \\ &= c_1\langle v_1, v_i\rangle + \cdots + c_k\langle v_k, v_i\rangle = c_i\langle v_i, v_i\rangle = c_i\|v_i\|^2.\end{aligned}$$

Hence $c_i = \frac{\langle y, v_i\rangle}{\|v_i\|^2}$, and that completes the proof of this part of the lemma.

(2). Suppose that S is orthonormal. Then $\|v_i\|^2 = 1$ for all $i \in \{1, \ldots, k\}$. This part of the lemma now follows immediately from Part (1). □

Corollary 9.2.3. *Let V be an inner product space over $\mathbb{R}$, and let $f: V \to V$ be a linear map. Suppose that V is finite-dimensional. Let $\beta = \{v_1, \ldots, v_n\}$ be an ordered orthonormal basis for V. Let $\left[a_{ij}\right] = [f]_\beta$. Then $a_{ij} = \langle f(v_j), v_i\rangle$ for all $i, j \in \{1, \ldots, n\}$.*

Proof. Let $j \in \{1, \ldots, n\}$. By Remark 5.5.2, we know that $f(v_j) = \sum_{i=1}^{n} a_{ij}v_i$.

By Lemma 9.2.2 (2), we also know that $f(v_j) = \sum_{i=1}^{n}\langle f(v_j), v_i\rangle v_i$. Equating these two expressions for $f(v_j)$ and using Theorem 3.6.2 (2) implies that $a_{ij} = \langle f(v_j), v_i\rangle$ for all $i \in \{1, \ldots, n\}$. □

Lemma 9.2.4. *Let V be an inner product space over $\mathbb{R}$, and let $S = \{v_1, \ldots, v_k\} \subseteq V$.*

1. *If S is orthogonal and if $v_i \neq 0$ for all $i \in \{1, \ldots, k\}$, then S is linearly independent.*
2. *If S is orthonormal, then S is linearly independent.*

Proof.

(1). Suppose S is orthogonal. Let $a_1, \ldots, a_n \in \mathbb{R}$. Suppose $a_1 v_1 + \ldots + a_k v_k = 0$. Then $0 \in \operatorname{span}(S)$. It then follows from Lemma 9.2.2 (1), using $y = 0$, that $a_i = \frac{\langle 0, v_i \rangle}{\|v\|_i^2} = 0$ for all $i \in \{1, \ldots, k\}$. Hence S is linearly independent.

(2). This part follows immediately from Part (1), together with the fact that a vector with norm 1 cannot be 0. □

Corollary 9.2.5. *Let V be an inner product space over $\mathbb{R}$. Suppose that V is finite-dimensional. Let $n = \dim(V)$. Let $S = \{v_1, \ldots, v_n\} \subseteq V$.*

1. *If S is orthogonal and if $v_i \neq 0$ for all $i \in \{1, \ldots, n\}$, then S is a basis for V.*
2. *If S is orthonormal, then $S$is a basis for V.*

Proof. Combine Lemma 9.2.4 (1) and Corollary 3.6.9 (4). □

Theorem 9.2.6 (Gram-Schmidt). *Let V be an inner product space over $\mathbb{R}$, and let $S = \{w_1, \ldots, w_n\} \subseteq V$. Suppose that S is linearly independent. Let $S' = \{v_1, \ldots, v_n\} \subseteq V$ be defined recursively as follows. Let $v_1 = w_1$, and let*

$$v_k = w_k - \sum_{i=1}^{k-1} \frac{\langle w_k, v_i \rangle}{\|v_i\|^2} v_i \tag{1}$$

for all $k \in \{2, \ldots, n\}$.

1. *S' is orthogonal.*
2. *None of the vectors in S' is 0.*
3. $\operatorname{span}(S') = \operatorname{span}(S)$.

Proof. We prove all three parts of the theorem by induction on n, which is the number of elements of S.

For each $k \in \{1, \ldots, n\}$, let $S_k = \{w_1, \ldots w_k\}$. Then $(S_k)' = \{v_1, \ldots v_k\}$ for all $k \in \{1, \ldots, n\}$.

Base Case: Let $k = 1$. Observe that $(S_1)' = S_1$, because $v_1 = w_1$. Clearly all three parts of the theorem hold for S_1.

Inductive Step: Second, let $k \in \{2, \ldots, n\}$. Suppose that all three parts of the theorem hold for S_{k-1}. We will show that all three parts of the theorem hold for S_k, which will complete the proof.

By the inductive hypothesis we know that $(S_{k-1})'$ is orthogonal. Let $r \in \{1, \ldots, k-1\}$. Then

$$\begin{aligned}\langle v_k, v_r\rangle &= \langle w_k - \sum_{i=1}^{k-1} \frac{\langle w_k, v_i\rangle}{\|v_i\|^2} v_i, v_r\rangle = \langle w_k, v_r\rangle - \sum_{i=1}^{k-1} \frac{\langle w_k, v_i\rangle}{\|v_i\|^2}\langle v_i, v_r\rangle \\ &= \langle w_k, v_r\rangle - \frac{\langle w_k, v_r\rangle}{\|v_r\|^2}\langle v_r, v_r\rangle = \langle w_k, v_r\rangle - \frac{\langle w_k, v_r\rangle}{\|v_r\|^2}\|v_r\|^2 = 0.\end{aligned}$$

Hence Part (1) holds for S_k.

By the inductive hypothesis we know that none of the vectors in $(S_{k-1})'$ is 0. Suppose that $v_k = 0$. Then Equation (1) implies that $w_k = \sum_{i=1}^{k-1} \frac{\langle w_k, v_i\rangle}{\|v_i\|^2} v_i$. Hence $w_k \in \operatorname{span}((S_{k-1})')$. But by the inductive hypothesis, we know that $\operatorname{span}((S_{k-1})') = \operatorname{span}(S_{k-1})$, and hence $w_k \in \operatorname{span}(S_{k-1})$. By Lemma 3.5.3 we deduce that S_k is linearly dependent, which by Lemma 3.5.7 (1) implies that S is linearly dependent, which is a contradiction. We conclude that $v_k \neq 0$. Hence Part (2) holds for S_k.

By the inductive hypothesis we know that $\operatorname{span}((S_{k-1})') = \operatorname{span}(S_{k-1})$. Hence $\operatorname{span}((S_{k-1})') \subseteq \operatorname{span}(S_k)$. Clearly $\sum_{i=1}^{k-1} \frac{\langle w_k, v_i\rangle}{\|v_i\|^2} v_i \in \operatorname{span}((S_{k-1})')$, and hence $\sum_{i=1}^{k-1} \frac{\langle w_k, v_i\rangle}{\|v_i\|^2} v_i \in \operatorname{span}(S_{k-1})$. It therefore follows from Equation (1) that $v_k \in \operatorname{span}(S_k)$. Putting all that together we deduce that $\operatorname{span}((S_k)') \subseteq \operatorname{span}(S_k)$. We know S_k is linearly independent. Because we have already proved Part (1) and Part (2) for S_k, it follows from Lemma 9.2.4 (1) that $(S_k)'$ is linearly independent. Hence $(S_k)'$ and S_k are bases for $\operatorname{span}((S_k)')$ and $\operatorname{span}(S_k)$, respectively. Hence $\dim(\operatorname{span}((S_k)')) = |(S_k)'| = k = |S_k| = \dim(\operatorname{span}(S_k))$. It then follows from Theorem 3.6.10 (3) that $\operatorname{span}((S_k)') = \operatorname{span}(S_k)$. Hence Part (3) holds for S_k. ◻

Corollary 9.2.7. *Let V be an inner product space over $\mathbb{R}$. Suppose that V is finite-dimensional. Then V has an orthonormal basis.*

Proof. Let B be any finite basis for V. Applying the Gram-Schmidt process (Theorem 9.2.6) to B yields an orthogonal basis S for V. Dividing each element of S by its norm yields an orthonormal basis for V. ◻

Corollary 9.2.8. *Let V be an inner product space over $\mathbb{R}$. Suppose that V is finite-dimensional. Let $S = \{v_1, \ldots, v_k\}$ be an orthonormal set.*

1. *Let S' be the result of doing the Gram-Schmidt process to S. Then $S' = S$.*

2. *S can be extended to an orthonormal basis for V.*

Proof.

(1). Left to the reader in Exercise 9.2.3.

(2). By Lemma 9.2.4 (2) we know that S is linearly independent, and then by Corollary 3.6.9 (5) we see that S can be extended to a basis B of V. Applying the Gram-Schmidt process (Theorem 9.2.6) to B yields an orthogonal basis T for V. By Part (1) of this corollary, we see that when the Gram-Schmidt process was applied to B, it did not change S. Hence $S \subseteq T$. Finally, divide every element of T by its norm to obtain an orthonormal basis for V that contains S. $\square$

Exercises

Exercise 9.2.1. Let $B = \left\{ \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 3 \\ 0 \\ 1 \end{bmatrix} \right\}$. It can be verified that S is a basis for $\mathbb{R}$, but not an orthogonal basis; there is no need to do that verification.

(1) Apply the Gram-Schmidt process to B, to obtain an orthogonal basis S for $\mathbb{R}^3$.

(2) Use S to make an orthonormal basis T for $\mathbb{R}^3$.

Exercise 9.2.2. Let V be an inner product space over $\mathbb{R}$. Suppose that V is finite-dimensional. Let $\{v_1, \ldots, v_n\}$ be an orthonormal basis for V, and let $x, y \in V$. Prove that

$$\langle x, y \rangle = \sum_{i=1}^{n} \langle x, v_i \rangle \langle y, v_i \rangle,$$

which is called the **Parseval's Identity**.

Exercise 9.2.3. Prove Corollary 9.2.8 (1).

9.3 Orthogonal Complement

Friedberg-Insel-Spence, 4th ed. – Section 6.2

Definition 9.3.1. Let V be an inner product space.

1. Let $S \subseteq V$. Suppose that $S \neq \emptyset$. The **orthogonal complement** of S, denoted $S^\perp$, is the set
$$S^\perp = \{x \in V \mid \langle x, y \rangle = 0 \text{ for all } y \in S\}.$$
2. Let $\emptyset^\perp = V$. △

Lemma 9.3.2. *Let V be an inner product space over $\mathbb{R}$, and let $A, B \subseteq V$.*

1. $\{0\}^\perp = V$.
2. *$A^\perp$ is a subspace of V.*
3. *If $A \subseteq B$, then $B^\perp \subseteq A^\perp$.*
4. $A \subseteq A^{\perp\perp}$.
5. *If $A \neq \emptyset$, then $A \cap A^\perp = \{0\}$.*

Proof. Left to the reader in Exercise 9.3.1. □

Lemma 9.3.3. *Let V be an inner product space over $\mathbb{R}$, and let $W \subseteq V$ be a subspace. Suppose that V is finite-dimensional.*

1. $\dim(W) + \dim(W^\perp) = \dim(V)$.
2. $W^{\perp\perp} = W$.
3. $W \oplus W^\perp = V$.

Proof. We know from Theorem 3.6.10 that W is finite-dimensional and $\dim(W) \leq \dim(V)$.

(1). By Corollary 9.2.7 we know that W has an orthonormal basis. Let $S = \{v_1, \ldots, v_k\}$ be such a basis of W. By Corollary 9.2.8 (2) we know that S can be extended to an orthonormal basis $B = \{v_1, \ldots, v_n\}$ of V, where $n \geq k$. Then $\dim(W) = k$ and $\dim(V) = n$. Let $T = \{v_{k+1}, \ldots, v_n\}$. We will show that T is a basis for $W^\perp$, and that will prove Part (1) of the lemma.

First, we note that T is orthonormal, so by Lemma 9.2.4 (2) we know that T is linearly independent. Because B is orthonormal, then $T \subseteq S^\perp$. By Exercise 9.3.3 we know that $S^\perp = (\mathrm{span}(S))^\perp = W^\perp$. Hence $T \subseteq W^\perp$. By Lemma 9.3.2 (2) and Lemma 3.4.3 (3) we see that $\mathrm{span}(T) \subseteq W^\perp$.

Let $z \in W^\perp$. By Lemma 9.2.2 (2) we know that $z = \sum_{i=1}^n \langle z, v_i \rangle v_i$. But $z \in W^\perp$ implies that $\langle z, v_i \rangle = 0$ for all $i \in \{1, \ldots, k\}$. Hence $z = \sum_{i=k+1}^n \langle z, v_i \rangle v_i \in \mathrm{span}(T)$. Therefore $W^\perp \subseteq \mathrm{span}(T)$. We conclude that $\mathrm{span}(T) = W^\perp$. Hence T is a basis for $W^\perp$.

(2). We know by Lemma 9.3.2 (4) that $W \subseteq W^{\perp\perp}$. By Part (1) we know that lemAHT1 $\dim(W) + \dim(W^{\perp}) = \dim(V)$, and by applying that part of the lemma to $W^{\perp}$ we see that $\dim(W^{\perp}) + \dim(W^{\perp\perp}) = \dim(V)$. It follows that $\dim(W) = \dim(W^{\perp\perp})$. By Lemma 3.6.10 (3) we deduce that $W = W^{\perp\perp}$

(3). We use the sets S and T from the proof of Part (1) of the lemma. Recall from that part of the proof that $\mathrm{span}(S) = W$ and $\mathrm{span}(T) = W^{\perp}$. The desired result now follows from Exercise 3.6.3. □

Exercises

Exercise 9.3.1. Prove Lemma 9.3.2.

Exercise 9.3.2. Let V be an inner product space over $\mathbb{R}$, let $W \subseteq V$ be a subspace, let β be a basis for W, and let $z \in V$. Prove that

$$W^{\perp} = \{x \in V \mid \langle x, b \rangle = 0 \text{ for all } b \in \beta\}.$$

Exercise 9.3.3. Let V be an inner product space over $\mathbb{R}$, and let $S \subseteq V$. Prove that $S^{\perp} = (\mathrm{span}(S))^{\perp}$.

Exercise 9.3.4. Let V be an inner product space over $\mathbb{R}$, and let $X, Y \subseteq V$ be subspaces. Suppose that V is finite-dimensional. Recall the definition of $X+Y$ given in Definition 3.3.8.

(a) Prove that $(X + Y)^{\perp} = X^{\perp} \cap Y^{\perp}$.

(b) Prove that $(X \cap Y)^{\perp} = X^{\perp} + Y^{\perp}$.

9.4 Adjoint of a Linear Map

Friedberg-Insel-Spence, 4th ed. – Section 6.3

Lemma 9.4.1. *Let V be an inner product space over $\mathbb{R}$, and let $h\colon V \to \mathbb{R}$ be a linear map. Suppose that V is finite-dimensional. Then there exists a unique $y \in V$ such that $h(x) = \langle x, y\rangle$ for all $x \in V$.*

Proof. By Corollary 9.2.7, we know that V has an orthonormal basis. Let $\beta = \{v_1, \ldots, v_n\}$ be such a basis for V. Let $y = \sum_{i=1}^{n} h(v_i)v_i$. Let $p\colon V \to \mathbb{R}$ be defined by $p(x) = \langle x, y\rangle$ for all $x \in V$. We know that p is a linear map, using the definition of an inner product.

Let $i = \{1, \ldots, n\}$. Then

$$p(v_k) = \langle v_k, y\rangle = \langle v_k, \sum_{i=1}^{n} h(v_i)v_i\rangle = \sum_{i=1}^{n} h(v_i)\langle v_k, v_i\rangle = h(v_k)\cdot 1 = h(v_k).$$

Hence h and p agree on the basis for V, and therefore by Corollary 4.1.7 we see that $h = p$.

To show that y is unique, suppose there is some $y' \in V$ such that $h(x) = \langle x, y'\rangle$ for all $x \in V$. Then $\langle x, y\rangle = \langle x, y'\rangle$ for all $x \in V$. It follows from Lemma 9.1.3 (5) that $y = y'$. □

Theorem 9.4.2. *Let V be an inner product space over $\mathbb{R}$, and let $f\colon V \to V$ be a linear map. Suppose that V is finite-dimensional. Then there exists a unique function $f^*\colon V \to V$ such that $\langle f(x), y\rangle = \langle x, f^*(y)\rangle$ for all $x, y \in V$.*

Proof. We define the function $f^*\colon V \to V$ as follows. Let $v \in V$. Let $g_v\colon V \to \mathbb{R}$ be defined by $g_v(x) = \langle f(x), v\rangle$ for all $x \in V$. Then g_v is a linear map by the definition of inner products. By Lemma 9.4.1 there is a unique $w_v \in V$ such that $g_v(x) = \langle x, w_v\rangle$ for all $x \in V$. Let $f^*(v) = w_v$. We have now defined the function f^*.

Let $x, y \in V$. Then $\langle f(x), y\rangle = g_y(x) = \langle x, w_y\rangle = \langle x, f^*(y)\rangle$.

To show that f^* is unique, suppose there is some linear map $q\colon V \to V$ such that $\langle f(x), y\rangle = \langle x, q(y)\rangle$ for all $x, y \in V$. Then $\langle x, f^*(y)\rangle = \langle x, q(y)\rangle$ for all $x, y \in V$. It follows from Lemma 9.1.3 (5) that $f^*(y) = q(y)$ for all $y \in V$. Hence $f^* = q$. □

Definition 9.4.3. Let V be an inner product space over $\mathbb{R}$, and let $f\colon V \to V$ be a linear map. An **adjoint** of f is a function $f^*\colon V \to V$ such that $\langle f(x), y\rangle = \langle x, f^*(y)\rangle$ for all $x, y \in V$. △

Remark 9.4.4. Let V be an inner product space over $\mathbb{R}$, and let $f\colon V \to V$ be a linear map. Suppose that V is finite-dimensional. Then Lemma 9.4.2 says that f has a unique adjoint. ◊

Remark 9.4.5. Let V be an inner product space over $\mathbb{R}$, and let $f\colon V \to V$ be a linear map. Suppose that V is finite-dimensional. It is straightforward to verify that $\langle x, f(y)\rangle = \langle f^*(x), y\rangle$ for all $x, y \in V$. ◊

Lemma 9.4.6. *Let V be an inner product space over $\mathbb{R}$, and let $f\colon V \to V$ be a linear map. Suppose that f has an adjoint.*

1. *The adjoint of f is unique.*
2. *The adjoint of f is a linear map.*

Proof.

(1). The uniqueness part of the proof of Theorem 9.4.2 holds whether or not V is finite-dimensional.

(2). Let $x, y \in V$ and let $c \in \mathbb{R}$. If $w \in V$, then

$$\begin{aligned}\langle w, f^*(x+y)\rangle &= \langle f(w), x+y\rangle = \langle f(w), x\rangle + \langle f(w), y\rangle \\ &= \langle w, f^*(x)\rangle + \langle w, f^*(y)\rangle = \langle w, f^*(x) + f^*(y)\rangle.\end{aligned}$$

It follows from Lemma 9.1.3 (5) that $f^*(x+y) = f^*(x) + f^*(y)$.

A similar argument shows that $f^*(cx) = cf^*(x)$, and we omit the details. Hence f^* is a linear map. □

Lemma 9.4.7. *Let V be an inner product space over $\mathbb{R}$, let $f, g\colon V \to V$ be linear maps, and let $c \in \mathbb{R}$. Suppose that f and g have adjoints.*

1. *$f + g$ has an adjoint, and $(f+g)^* = f^* + g^*$.*
2. *cf has an adjoint, and $(cf)^* = cf^*$.*
3. *$g \circ f$ has an adjoint, and $(g \circ f)^* = f^* \circ g^*$.*
4. *f^* has an adjoint, and $f^{**} = f$.*
5. *1_V has an adjoint, and $(1_V)^* = 1_V$.*

Proof. We prove Part (1); the remaining parts of this lemma are left to the reader in Exercise 9.4.1.

(1). Let $x, y \in V$. Then

$$\begin{aligned}\langle (f+g)(x), y\rangle &= \langle f(x) + g(x), y\rangle = \langle f(x), y\rangle + \langle g(x), y\rangle \\ &= \langle x, f^*(y)\rangle + \langle x, g^*(y)\rangle = \langle x, f^*(y) + g^*(y)\rangle \\ &= \langle x, (f^* + g^*)(y)\rangle.\end{aligned}$$

We therefore see that the function $f^* + g^*$ satisfies Definition 9.4.3 with respect to the function $f + g$. Hence $f + g$ has an adjoint, which is $f^* + g^*$. □

Definition 9.4.8. Let V be a vector space over F, let $W \subset V$ be a subspace, and let $f: V \to V$ be a linear map. The subspace W is **invariant** under f if $f(W) \subseteq W$. △

Lemma 9.4.9. *Let V be an inner product space over $\mathbb{R}$, let $f: V \to V$ be a linear map, and let $W \subseteq V$ be a subspace. Suppose that f has an adjoint.*

1. *W is invariant under f if and only if $W^\perp$ is invariant under f^*.*

2. *If W is invariant under f and f^*, then $f|W$ has an adjoint, and $(f|W)^* = f^*|W$.*

Proof.

(1). Suppose that W is invariant under f. That means that $f(W) \subseteq W$. Let $y \in W^\perp$. For each $x \in W$, we have $\langle f^*(y), x\rangle = \langle x, f^*(y)\rangle = \langle f(x), y\rangle = 0$, because $f(x) \in W$ and $y \in W^\perp$. It follows that $f^*(y) \in W^\perp$. We deduce that $f^*(W^\perp) \subseteq W^\perp$, which means that $W^\perp$ is invariant under f^*.

Now suppose $W^\perp$ is invariant under f^*. Hence $f^*(W^\perp) \subseteq W^\perp$. A similar argument as before shows that $f^{**}(W^{\perp\perp}) \subseteq W^{\perp\perp}$. However, by Lemma 9.4.7 (4) we know that $f^{**} = f$, and by Lemma 9.3.3 (2) we know that $W^{\perp\perp} = W$. Hence $f(W) \subseteq W$, which means that W is invariant under f.

(2). Suppose that W is invariant under f and f^*. Let $x, y \in W$. Then $\langle (f|W)(x), y\rangle = \langle f(x), y\rangle = \langle x, f^*(y)\rangle = \langle x, (f^*|W)(y)\rangle$. We therefore see that the function $f^*|W$ satisfies Definition 9.4.3 with respect to the function $f|W$. Hence $f|W$ has an adjoint, which is $f^*|W$. □

Theorem 9.4.10. *Let V be an inner product space over $\mathbb{R}$, and let $f: V \to V$ be a linear map. Suppose that V is finite-dimensional. Let β be an ordered orthonormal basis for V. Then $[f^*]_\beta = \left([f]_\beta\right)^t$.*

Proof. Let $\beta = \{v_1, \ldots, v_n\}$. Let $\left[a_{ij}\right] = [f]_\beta$ and $\left[c_{ij}\right] = [f^*]_\beta$. Let $i, j \in \{1, \ldots, n\}$. By Corollary 9.2.3, we know that $a_{ij} = \langle f(v_j), v_i\rangle$, and that $c_{ij} = \langle f^*(v_j), v_i\rangle = \langle v_i, f^*(v_j)\rangle = \langle f(v_i), v_j\rangle = a_{ji}$. Hence $\left[c_{ij}\right] = \left[a_{ij}\right]^t$, which means $[f^*]_\beta = \left([f]_\beta\right)^t$. □

Exercises

Exercise 9.4.1. Prove Lemma 9.4.7 (2), (3), (4) and (5).

Exercise 9.4.2. Let V be an inner product space over $\mathbb{R}$, and let $f: V \to V$ be a linear map.

(1) Let $g = f + f^*$. Prove that g has an adjoint, and that $g^* = g$.

(2) Let $h = f \circ f^*$. Prove that h has an adjoint, and that $h^* = h$.

Exercise 9.4.3. Let V be an inner product space over $\mathbb{R}$, and let $f : V \to V$ be a linear map. Suppose that V is finite-dimensional. Suppose that f is an isomorphism. Prove that f^* is an isomorphism, and that $(f^*)^{-1} = (f^{-1})^*$.

Exercise 9.4.4. Let V be an inner product space over $\mathbb{R}$, and let $f : V \to V$ be a linear map. Suppose that f has an adjoint.

(1) Prove that $(\operatorname{im} f^*)^\perp = \ker f$.

(2) Suppose that V is finite dimensional. Prove that $\operatorname{im} f^* = (\ker f)^\perp$.

9.5 Self-Adjoint Linear Maps

Friedberg-Insel-Spence, 4th ed. – Section 6.4

Definition 9.5.1. Let V be an inner product space over $\mathbb{R}$, and let $f : V \to V$ be a linear map. The function f is **self-adjoint** if $f^* = f$. △

Remark 9.5.2. Let V be an inner product space over $\mathbb{R}$, and let $f : V \to V$ be a linear map. The function f is self-adjoint if and only if $\langle f(x), y \rangle = \langle x, f(y) \rangle$ for all $x, y \in V$. ◊

Lemma 9.5.3.

1. *Let V be an inner product space over $\mathbb{R}$, and let $f : V \to V$ be a linear map. Suppose that V is finite-dimensional. Let β be an ordered orthonormal basis for V. Then f is self-adjoint if and only if $[f]_\beta$ is symmetric.*

2. *Let $A \in \mathsf{M}_{n\times n}(\mathbb{R})$. Then A is symmetric if and only if L_A is self-adjoint.*

Proof.

(1). By Theorem 9.4.10 we know that $[f^*]_\beta = \big([f]_\beta\big)^t$.

Suppose that f is self-adjoint. Then $f^* = f$. Because $[f^*]_\beta = \big([f]_\beta\big)^t$, it follows that $[f]_\beta = \big([f]_\beta\big)^t$, which means that $[f]_\beta$ is symmetric.

Suppose that $[f]_\beta$ is symmetric. Then $[f]_\beta = \big([f]_\beta\big)^t$. Because $[f^*]_\beta = \big([f]_\beta\big)^t$, it follows that $[f]_\beta = [f^*]_\beta$. By Lemma 5.5.3 (1) we deduce that $f = f^*$, which means that f is self-adjoint.

(2). Let γ be the standard ordered basis for $\mathbb{R}^n$. Observe that γ is an orthonormal basis.

By Part (1) of this lemma, we see that L_A is self-adjoint if and only if $[\mathsf{L}_A]_\gamma$ is symmetric. By Lemma 5.6.3 (1) we know that $[\mathsf{L}_A]_\gamma = A$, which implies that L_A is self-adjoint if and only if A is symmetric. □

Definition 9.5.4. Let $n \in \mathbb{N}$. The **$(n-1)$-sphere** in $\mathbb{R}^n$, denoted S^{n-1}, is the set

$$S^{n-1} = \{v \in \mathbb{R}^n \mid \|v\| = 1\}.$$ △

Theorem 9.5.5. *Let $A \in \mathsf{M}_{n\times n}(\mathbb{R})$. If A is symmetric, then A has an eigenvector.*

Proof. This proof is from [Lan66, p. 192]. Suppose that A is symmetric. Let $f : S^{n-1} \to \mathbb{R}$ be defined by $f(x) = \langle Ax, x\rangle$ for all $x \in S^{n-1}$. It can be shown (using the methods of Real Analysis) that f is differentiable, and hence continuous.

Let $v \in S^{n-1}$ be such that f achieves its maximum value at v.

Let $w \in S^{n-1}$. Suppose that $\langle v, w \rangle = 0$. We will show that $\langle Av, w \rangle = 0$. It will then follow that Av is orthogonal to all the vectors in $S^{n-1} \cap \{v\}^{\perp}$; we deduce that Av is orthogonal to $\{v\}^{\perp}$, and hence Av is orthogonal to $\operatorname{span}(\{v\})^{\perp}$. Therefore $Av \in \operatorname{span}(\{v\})^{\perp\perp}$. By Lemma 9.3.3 (2) we deduce that $Av \in \operatorname{span}(\{v\})$. It follows that Av is a multiple of v, which means that v is an eigenvector of A, which is what we are trying to prove.

We now show that $\langle Av, w \rangle = 0$. Let $c\colon (-\pi/2, \pi/2) \to S^{n-1}$ be defined by $c(t) = (\cos t)v + (\sin t)w$ for all $t \in (-\pi/2, \pi/2)$. Let $t \in (-\pi/2, \pi/2)$. Recalling that v and w are unit vectors, and that $\langle v, w \rangle = 0$, we see that $\|c(t)\|^2 = \langle c(t), c(t) \rangle = (\cos t)^2 \langle v, v \rangle + 2(\cos t)(\sin t)\langle v, w \rangle + (\sin t)^2 \langle w, w \rangle = (\cos t)^2 + (\sin t)^2 = 1$. Hence $c(t)$ is a unit vector, and so $c(t) \in S^{n-1}$, which makes the function c validly defined.

Clearly c is differentiable. It is straightforward to see that $c(0) = v$ and $c'(0) = w$. We can form the function $f \circ c$. Because each of f and c are differentiable, so is $f \circ c$. Because $f(v)$ is the maximal value of f, then certainly $(f \circ c)(0)$ is the largest value of $f \circ c$. Hence $(f \circ c)'(0) = 0$.

By hypothesis A is symmetric, and hence by Lemma 9.5.3 (2) we know

It is then seen that L_A is self-adjoint, which means $\mathsf{L}_A^* = \mathsf{L}_A$. We then use the Product Rule to compute

$$\begin{aligned}
(f \circ c)'(t) &= \frac{d}{dt}\langle Ac(t), c(t) \rangle = \langle Ac'(t), c(t) \rangle + \langle Ac(t), c'(t) \rangle \\
&= \langle \mathsf{L}_A(c'(t)), c(t) \rangle + \langle Ac(t), c'(t) \rangle = \langle c'(t), \mathsf{L}_A^*(c(t)) \rangle + \langle Ac(t), c'(t) \rangle \\
&= \langle c'(t), \mathsf{L}_A(c(t)) \rangle + \langle Ac(t), c'(t) \rangle = \langle c'(t), Ac(t) \rangle + \langle Ac(t), c'(t) \rangle \\
&= 2\langle Ac(t), c'(t) \rangle.
\end{aligned}$$

We deduce that

$$0 = (f \circ c)'(0) = 2\langle Ac(0), c'(0) \rangle = 2\langle Av, w \rangle,$$

which is what we needed to show. □

Corollary 9.5.6. *Let V be an inner product space over $\mathbb{R}$, and let $f\colon V \to V$ be a linear map. Suppose that V is finite-dimensional. If f is self-adjoint, then f has an eigenvector.*

Proof. Suppose that f is self-adjoint. By Corollary 9.2.7, there is an ordered orthonormal basis β for V. By Lemma 9.5.3 (1) the matrix $[f]_\beta$ is symmetric. We now use Theorem 9.5.5 to deduce that the matrix $[f]_\beta$ has an eigenvector. It now follows from Corollary 8.1.13 that f has an eigenvector. □

Theorem 9.5.7 (Spectral Theorem). *Let V be an inner product space over $\mathbb{R}$, and let $f\colon V \to V$ be a linear map. Suppose that V is finite-dimensional. Then f is self-adjoint if and only if V has an orthonormal basis of eigenvectors of f.*

Proof. First, suppose f is self-adjoint. That means $f^* = f$. Let $n = \dim(V)$. The proof is by induction on n.

Base Case: Suppose that $n = 1$. Then $V = \mathbb{R}$ (thought of as a vector space over itself), and f is a linear map $\mathbb{R} \to \mathbb{R}$. Clearly $\{1\}$ is an orthonormal basis for V, and $f(1)$ is some multiple of 1, so 1 is an eigenvector of f.

Inductive Step: Let $n \in \mathbb{N}$. Suppose $n \geq 2$, and suppose that the result is true for $n-1$. By Corollary 9.5.6, we know that f has an eigenvector; let w be an eigenvector of f. By definition $w \neq 0$. Let $v = \frac{w}{\|w\|}$. Then v is an eigenvector and a unit vector. Let $W = \operatorname{span}(\{v\})$. By Lemma 3.4.3 (2) we see that W is a subspace of V. Clearly $\{v\}$ is a basis for W, and hence $\dim(W) = 1$. Because v is an eigenvector of v, then $f(v) \in W$, and it follows that $f(W) \subseteq W$, which means that W is invariant under f. By Lemma 9.4.9 (1) we know that $W^\perp$ is invariant under f^*. Because $f^* = f$, it follows that $W^\perp$ is invariant under f. By Lemma 9.4.9 (2) applied to $W^\perp$, we see that $(f|W^\perp)^* = f^*|W^\perp = f|W^\perp$. Hence $f|W^\perp$ is self-adjoint. By Lemma 9.3.3 (1) we know that $\dim(W^\perp) = n-1$. We can then apply the inductive hypothesis to $f|W^\perp$, to find an orthonormal basis $\{v_2, \ldots, v_n\}$ for $W^\perp$. Clearly $\{v, v_2, \ldots, v_n\}$ is orthonormal, and by Corollary 9.2.5 (2) we deduce that it is a basis for V.

Second, suppose that V has an orthonormal basis of eigenvectors of f. Let $\beta = \{v_1, \ldots, v_n\}$ be such a basis, with corresponding eigenvalues $\{\lambda_1, \ldots, \lambda_n\}$. Let $x, y \in V$. Then $x = a_1v_1 + \cdots + a_nv_n$ and $y = b_1v_1 + \cdots + b_nv_n$ for unique $a_1, \ldots, a_n, b_1, \ldots, b_n \in \mathbb{R}$. It is then straightforward to see that both $\langle f(x), y\rangle$ and $\langle x, f(y)\rangle$ are equal to $\lambda_1a_1b_1 + \cdots + \lambda_na_nb_n$. Hence f is self-adjoint by Remark 9.5.2. □

Corollary 9.5.8. *Let F be a field. Let $A \in \mathsf{M}_{n\times n}(F)$. Then A is symmetric if and only if there is an invertible matrix P, which has orthonormal columns, such that $P^{-1}AP$ is a diagonal matrix.*

Proof. Left to the reader in Exercise 9.5.2. □

Exercises

Exercise 9.5.1. Let V be an inner product space over $\mathbb{R}$, and let $f, g\colon V \to V$ be self-adjoint linear maps. Prove that $g \circ f$ is self-adjoint if and only if $g \circ f = f \circ g$.

Exercise 9.5.2. Prove Corollary 9.5.8.

Exercise 9.5.3. Let F be a field. Let $A \in \mathsf{M}_{n\times n}(\mathbb{R})$. We say that A is **Gramian** if there exists $B \in \mathsf{M}_{n\times n}(\mathbb{R})$ such that $A = B^tB$.

Prove that A is Gramian if and only if A is symmetric and all of its eigenvalues are non-negative.

Bibliography

[Ber92] Sterling Berberian, *Linear Algebra*, Oxford University Press, Oxford, 1992.

[Cur74] Charles Curtis, *Linear Algebra: An Introductory Approach*, 3rd ed., Allyn & Bacon, Boston, 1974.

[Lan66] Serge Lang, *Linear Algebra*, Addison-Wesley, Reading, MA, 1966.