Linear Algebra Notes

Lance Remigio

February 28, 2024

Contents

List of Theorems

Chapter 1

Vector Spaces

1.1 Vector Spaces

1.1.1 Basics

Definition 1.1.1 (Vector Spaces). A **vector space** (or **linear space**) over a field F consists of a set on which two operations (called **addition** and **scalar multiplication**, respectively) are defined so that for each pair of elements x, y, in V there is a unique element ax in V, such that the following conditions hold:

(VS 1) For all $x, y \in V$, $x + y = y + x$ (commutativity of addition).

(VS 2) For all $x, y, z \in V$, $(x + y) + z = x + (y + z)$ (associativity of addition).

(VS 3) There exists an element in V denoted by O such that $x + O = x$ for each $x \in V$

(VS 4) For each element $x \in V$, there exists an element $y \in V$ such that $x + y = O$.

(VS 5) For each element $x \in V$, we have $1x = x$.

(VS 6) For each $a, b \in F$ and each element $x \in V$, then $(ab)x = a(bx)$.

(VS 7) For each element $a \in F$ and each pair $x, y \in V$, we have $a(x + y) = ax + ay$.

(VS 8) For each pair $a, b \in F$ and each $x \in V$, we have $(a + b)x = ax + bx$.

The elements $x + y$ and ax are called the **sum** of x and y and the **product** of a and x, respectively.

- The elements of a field F are called **scalars** and the elements of a vector space V are called **vectors** (these should not be confused!).
- Every vector space will always be defined over a given field, mostly defined over the real numbers $\mathbb{R}$ or the complex numbers $\mathbb{C}$ unless otherwise noted.
- Every vector space should specify the operations of addition and scalar multiplication.

Definition 1.1.2 (n-tuples). An object of the form $(a_1, a_2, \ldots, a_n)$, where the entries $a_1, a_2, \ldots, a_n$ are elements of a field F, is called an **n-tuple** with entries from F. The elements $a_1, a_2, \ldots, a_n$ are called **entries** or **components** of the n-tuple.

Definition 1.1.3. We say that two n-tuples, $(a_1, a_2, \dots, a_n)$ and $(b_1, b_2, \dots, b_n)$, are **equal** if $a_i = b_i$ for $i = 1, 2, \dots, n$.

Example 1.1.1. The set of all n-tuples with entries from a field F denoted by F_n is a vector space. To see why, suppose $u, v \in F_n$ where $u = (a_1, a_2, \dots a_n)$ and $v = (b_1, b_2, \dots, b_n)$. If we take term-by-term addition of the entries in both u and v, then we end up with

$$u + v = (a_1 + b_1, a_2 + b_2, \dots, a_n + b_n)$$

and likewise,

$$cu = (ca_1, ca_2, \dots, ca_n).$$

These same set of operations define $\mathbb{R}^3$ as a vector space over $\mathbb{R}$ and likewise, $\mathbb{C}^2$ is a vector space over $\mathbb{C}$.

- Note that vectors in F^n can be written as **column vectors**

$$\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}$$

  rather than **row vectors** $(a_1, a_2, \dots, a_n)$.
- 1-tuples are are just scalars or an just an element from F.

Definition 1.1.4. An $m \times n$ **matrix** with entries from a field F is a rectangular array of the form

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

where each entry a_{ij} with $(1 \leq i \leq m, 1 \leq j \leq n)$ is an element of F. We call the entries a_{ij} with $i = j$ the **diagonal entries** of the matrix. The entries $a_{i1}, a_{i2}, \dots a_{in}$ compose the **ith row**of the matrix, and the entries $a_{1j}, a_{2j}, \dots a_{mj}$ compose the jth column of the matrix.

- The rows make a vector space which we denote F^n.
- Likewise, the columns make a vector space we denote F^m

Definition 1.1.5 (Zero Matrix). The $m \times n$ matrix in which each entry equals zero is called the **zero matrix** and is denoted by O.

Definition 1.1.6 (Square Matrix). A matrix is **square** if the number of rows and columns of a matrix are equal.

- Just like our tuple example, the set of all $m \times n$ matrices with entries from a field F form a vector space. Denote this vector space as $M_{m \times n}(F)$ endowed with two operations; that is,

matrix addition and **scalar multiplication**. Suppose for $A, B \in M_{m\times n}(F)$ and $c \in F$, we have

$$(A+B)_{ij} = A_{ij} + B_{ij}$$

and

$$(cA)_{ij} = cA_{ij}$$

for $1 \leq i \leq m$ and $1 \leq j \leq n$. In other words, the two operations can be performed entry-wise.

- The operations from our tuple case extends very naturally to $M_{m\times n}(F)$. In other words, if we add two matrices A_{ij} and B_{ij}, then we would expect to that $A_{ij} + B_{ij} \in M_{m\times n}(F)$ as well and likewise for the scalar multiplication case.

Definition 1.1.7 (Set of All Functions). Let S be any nonempty set and F be any field, and let $\mathcal{F}(S, F)$ denote the set of all functions from S to F.

Definition 1.1.8. Two functions $f, g \in \mathcal{F}(S, F)$ are called **equal** if $f(s) = g(s)$ for each $s \in S$.

The set of functions above forms a vector space with the operations of addition and scalar multiplication defined for every $f, g \in \mathcal{F}(S, F)$ and $c \in F$ with

$$(f+g)(s) = f(s) + g(s) \text{ and } (cf)(s) = cf(s)$$

Definition 1.1.9 (Set of All Polynomials). A **polynomial** with coefficients from a field F is an expression of the form

$$f(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0,$$

where n is a nonnegative integer and each $a_k \in F$ is called the **coefficient** of x_k.

Definition 1.1.10 (Zero Polynomial). We call $f(x) = 0$ the **zero polynomial** if $a_n = a_{n-1} = \cdots = a_0 = 0$.

Definition 1.1.11 (Degree). The **degree** of a given polynomial f is defined to be the largest exponent of x that appears in the representation

$$f(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0.$$

Definition 1.1.12 (Equality of Polynomials). We call two polynomials f, g, where

$$f(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$$

and

$$g(x) = b_m x^m + b_{m-1}x^{m-1} + \cdots + b_1 x + b_0,$$

are **equal** if $m = n$ and $a_i = b_i$ for all $i = 0, 1, \ldots, n$.

Suppose we have $c \in F$ and say we evaluated the polynomial $f \in F$ at c. Then we would have the following

$$f(c) = a_n c^n + a_{n-1}c^{n-1} + \cdots + a_1 c + a_0$$

where $f(c) \in F$.

Definition 1.1.13 (Basic Operations of Polynomials). Define polynomial **addition** $f+g$ as the following:

$$f(x)+g(x) = (a_n+b_n)x^n + (a_{n-1}+b_{n-1})x^{n-1} + \cdots + (a_1+b_1)x + (a_0+b_0).$$

Let $c \in F$. Let scalar **multiplication** be defined by

$$cf(x) = ca_nx^n + ca_{n-1}x^{n-1} + \cdots + ca_1x + ca_0.$$

The operations above form a vector space for $P(F)$ (the set of all polynomials).

Definition 1.1.14 (Sequences). A **sequence** in F is a function $\sigma : \mathbb{Z}^+ \to F$. A given sequence σ such that $\sigma(n) = a_n$ for $n = 1, 2, \ldots$ is denoted (a_n).

Let V be the set of all sequences $\sigma(n) \in F$. For every $(a_n), (b_n) \in V$ with $t \in F$, we have the following operations

$$(a_n) + (b_n) = (a_n + b_n) \text{ and } t(a_n) = (ta_n).$$

1.1.2 Non-examples

Example 1.1.2. Let $S = \{(a_1, a_2) : a_1, a_2 \in R\}$ where R is a field. For every $(a_1, a_2), (b_1, b_2) \in S$ and $c \in R$, define

$$(a_1, a_2) + (b_1, b_2) = (a_1 + b_1, a_2 - b_2) \text{ and } c(a_1, a_2) = (ca_1, ca_2).$$

Note that S is **NOT** a vector space since (VS 1), (VS 2), and (VS 8) fail.

Example 1.1.3. Let S be the same set as in the last example. For $(a_1, a_2), (b_1, b_2) \in S$ and $c \in R$, define

$$(a_1, a_2) + (b_1, b_2) = (a_1 + b_1, 0)$$

and

$$c(a_1, a_2) = (ca_1, 0).$$

Note that S is **NOT** a vector space with these operations since (VS 3), (VS 4), and (VS 5) fail.

1.1.3 Basic Extensions from Definition

Theorem 1.1.1 (Cancellation Law for Vector Addtion). If $x, y, z \in V$ such that $x + z = y + z$, then $x = y$.

Proof. There exists a vector $v \in V$ such that $z + v = O$ (VS 4). Thus, we have

$$\begin{aligned} x &= x + O \\ &= x + (z + v) \\ &= (x + z) + v \\ &= (y + z) + v \\ &= y + (z + v) \\ &= y + O \\ &= y. \end{aligned}$$

Hence, we have $x = y$. ∎

Corollary. The vector described O described in (VS 3) is unique.

Proof. Let $x \in V$. Suppose there exists two elements $O, O' \in V$ such that

$$x + O = x \tag{1}$$

and

$$x + O' = x. \tag{2}$$

Our objective is to show that $O = O'$. Equating (1) and (2), we find that

$$x + O = x + O' \tag{3}$$

Using the Cancellation Law for Vector Addition, (3) implies that $O = O'$. Hence, O is a unique vector in V. ∎

Corollary. The vector y described in (VS 4) is unique.

Proof. Let $x \in V$. Suppose there exists $y, y' \in V$ such that

$$x + y = O \text{ and } x + y' = O.$$

Equating the two equations above, we get that

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

Applying the Cancellation Law for Vector Addition once again, we get that $y = y'$. Hence, y is a unique vector in V. ∎

Theorem 1.1.2 (Properties of a Vector Space). In any vector space V, the following statements are true:

(a) $0x = O$ for any $x \in V$.

(b) $(-a)x = -(ax) = a(-x)$ for each $a \in F$ and each $x \in V$.

(c) $a0 = 0$ for any $a \in F$.

Proof. (a) Using (VS 8), (VS 3), and (VS 1), it follows that

$$\begin{aligned} 0x + 0x &= (0+0)x \\ &= 0x \\ &= 0x + O \\ &= O + 0x. \end{aligned}$$

Hence, $0x = O$ by Theorem 1.1.

(b) The vector $-(ax)$ is the unique element of V such that $ax + [-(ax)] = O$. If we know that $ax + (-a)x = O$, we can use Corollary 2 to Theorem 1.1 to state that $(-a)x = -(ax)$. But by (VS 8), we have

$$ax + (-ax) = [a + (-a)]x = 0x = O$$

by (a). Hence, we have that $(-a)x = -(ax)$. In particular, we have $(-1)x = -x$. Thus, (VS 6) implies that

$$a(-x) = a[(-1)x] = [a(-1)]x = (-a)x.$$

(c) The proof of (c) is similar to the proof of (a).

■

1.2 Subspaces

Definition 1.2.1 (Subspaces). A subset W of a vector space V over a field F is called a **subspace** of V if W is a vector space over F with the operations of addition and scalar multiplication.

The most simple examples of subspaces of V is V itself and the set containing just the zero vector. The latter is denoted as the **zero subspace** of V.

We don't have to check all the vector space properties to prove that a subset W of V is a subspace of V. This is because vectors in any subset of V already satisfy all the properties of a vector space. Hence, we need only check that W is closed under

1. **Addition**: $x + y \in W$ whenever $x \in W$ and $y \in W$.
2. **Scalar Multiplication**: $cx \in W$ whenever $c \in F$ and $x \in W$.
3. W contains a zero vector.
4. Each vector in W has an additive inverse in W.

Theorem 1.2.1 (Subspaces). Let V be a vector space and W a subset of V. Then W is a subspace of V if and only if the following three conditions hold for the operations defined in V.

(a) $O \in W$.

(b) $x + y \in W$ whenever $x \in W$ and $y \in W$.

(c) $cx \in W$ whenever $c \in F$ and $x \in W$.

Proof. Suppose W is a subspace of V. Since W is also a vector space with the operations of addition and scalar multiplication defined in V. Hence, we know that W is closed under addition and scalar multiplication. Note that W contains a zero vector $O' \in W$ such that for any $x \in W$, we have $x + 0' = x$. Since x is also in V (since $W \subseteq V$), we know that $x + 0 = x$. Using the cancellation , we can see that $x + 0' = x + 0$ implies $0' = 0$. Hence, properties (a), (b), and (c) are satisfied.

Conversely, suppose conditions (a), (b), and (c) hold. We need to show that W is a subspace. This tells us that W is closed under addition and scalar multiplication in addition to containing the zero vector. Let $x \in W$. Since W is closed under scalar multiplication, we know that $(-1)x \in W$. By part (b) of theorem 2, we know that $1(-x) = -x \in W$. Hence, W contains an additive inverse and we are done. ■

The theorem above provides a simpler way to determine whether a given subset of a vector space is a subspace. Instead of using the definition to show that a give subset is a subspace, it is more common to use the result above.

Definition 1.2.2 (Tranpose)**.** The **transpose** of A^t of an $m \times n$ matrix A is the $n \times m$ matrix obtained from A by interchanging the rows with the columns; that is, $(A^t)_{ij} = A_{ji}$.

Example 1.2.1.

$$\begin{pmatrix} 1 & -2 & 3 \\ 0 & 5 & -1 \end{pmatrix}^t = \begin{pmatrix} 1 & 0 \\ -2 & 5 \\ 3 & -1 \end{pmatrix}$$

and

$$\begin{pmatrix} 1 & 2 \\ 2 & 3 \end{pmatrix}^t = \begin{pmatrix} 1 & 2 \\ 2 & 3 \end{pmatrix}.$$

Definition 1.2.3 (Symmetric Matrices)**.** A **symmetric matrix** is a matrix A such that $A^t = A$.

- The easiest example of a symmetric matrix would be a square matrix where $i = j$.
- The set W of all symmetric matrices in $M_{n\times n}(F)$ is a subspace of $M_{n\times n}(F)$ since the conditions of Theorem 1.3 hold.

We can show that W is indeed a subspace.

1. The zero matrix is equal to its transpose and hence belongs to W.
2. Suppose $A \in W$ and $B \in W$. Hence, $A^t = A$ and $B^t = B$. Hence, we have

$$(A + B)^t = A^t + B^t = A + B$$

   which implies that $A + B \in W$.
3. We have $A \in W$ implies $A^t = A$. Now, let $a \in F$. Then we have that $(aA)^t = aA^t = aA$. Hence, $aA \in W$.

1.2.1 Examples of Subspaces

Example 1.2.2. Let n be non-negative integer, and let $P_n(F)$ consist of all polynomials in $P(F)$ having degree less than or equal to n. We get that $P_n(F)$ is a subspace because:

1. Zero polynomial has degree -1, it is in $P_n(F)$.
2. The sum of two polynomials with degrees less than or equal to n is another a polynomial less than or equal to n.
3. The product of a scalar and a polynomial of degree less than or equal to n is a polynomial less than or equal to n.

This tells us that $P_n(F)$ is a subspace of the space of all polynomials.

Example 1.2.3. Let $C(\mathbb{R})$ denote the set of all continuous real-valued functions defined on $\mathbb{R}$. We know that $C(\mathbb{R})$ is a subset of the vector space $\mathcal{F}(\mathbb{R}, \mathbb{R})$ defined in Example 3. We will show that $C(\mathbb{R})$ is a subspace of $\mathcal{F}(\mathbb{R}, \mathbb{R})$.

1. Note that the zero function $f(x) = 0$ for all $x \in \mathbb{R}$ of $\mathcal{F}(\mathbb{R}, \mathbb{R})$ is a continuous real-valued function.
2. Let $f, g \in C(\mathbb{R})$. Since the sum of f and g is also continuous for all $x \in \mathbb{R}$, we have that $f + g \in C(\mathbb{R})$.
3. Let $c \in \mathbb{R}$ and $f \in C(\mathbb{R})$. We have that for any $x \in \mathbb{R}$, $(cf)(x) = cf(x)$ is a continuous function. Hence, property (c) is satisfied.

Hence, $C(\mathbb{R})$ is a subspace.

Definition 1.2.4 (Upper Triangular). An $m \times n$ matrix A is called **upper triangular** if all its entries below the diagonal entries are zero; that is, if $A_{ij} = 0$ whenever $i > j$.

Example 1.2.4. Let B be an upper triangular 3×4 matrix

$$B = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 0 & 5 & 6 & 7 \\ 0 & 0 & 8 & 9 \end{pmatrix}$$

Definition 1.2.5 (Diagonal Matrix). An $n \times n$ matrix M is called a **diagonal matrix** if $M_{ij} = 0$ whenever $i \neq j$; that is, if all its non-diagonal entries are zero.

Example 1.2.5. Let A be diagonal 3×3 matrix.

$$A = \begin{pmatrix} 3 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 8 \end{pmatrix}$$

Example 1.2.6. The set of diagonal matrices is a subspace of $M_{n\times n}(F)$. To see why, we have

1. Let O be the zero matrix of $M_{n\times n}(F)$. Since O is also a diagonal matrix, we know that O must be in the set of diagonal matrices.
2. Let A, B in the set of diagonal matrices. Let $i \neq j$ such that

$$(A+B)_{ij} = A_{ij} + B_{ij} = 0 + 0 = 0.$$

Hence, $A + B$ is a diagonal matrix and so addition is closed.

3. Let A be a diagonal matrix as before and let $c \in F$. Let $i \neq j$ again, and observe that

$$(cA)_{ij} = cA_{ij} = c \cdot 0 = 0.$$

Hence, cA is a diagonal matrix and so scalar multiplication is closed.

Since the set of diagonal matrices satisfies all properties of theorem 3, we conclude that it is indeed a subspace of $M_{n\times n}(F)$.

Definition 1.2.6 (Trace). The **trace** of an $n \times n$ matrix M, denoted $\mathrm{tr}(M)$, is the sun of the diagonal entries of M; that is,

$$\mathrm{tr}(M) = M_{11} + M_{22} + \cdots + M_{nn}.$$

The set of all $n\times n$ matrices that have a trace equal to zero is a subspace of $M_{n\times n}(F)$ (proved in Exercise 6).

Example 1.2.7 (Non-example). Denote V as the set of matrices in $M_{m\times n}(\mathbb{R})$ having non-negative entries. The subset V is not a subspace because it is not closed under scalar multiplication. We can see this by multiplying any matrix in V be a negative number and observe that the entries of said matrix are no longer all non-negative.

We can create subspaces out of other subspaces.

1.2.2 Constructing Subspaces Out of Existing Subspaces

Theorem 1.2.2 (Intersection of Subspaces is a Subspace). Any intersection of subspaces of a vector space V is a subspace of V.

Proof. Let C be a collection of subspaces of V, and let W denote the intersection of the subspaces in C. Since every subspace contains the zero vector and the intersection $W \neq \emptyset$, the zero vector $0 \in W$. Let $a \in F$ and $x, y \in W$. Since each subspace of C is closed under addition and scalar multiplication, it follows that $x + y$ and ax are contained each subspace in C and hence W must be closed under addition and closed under scalar multiplication. Thus, W is a subspace of V by theorem 3. ∎

A natural question to ask is whether or not the union of subspaces of a vector space V is a subspace of V too. We can see that indeed the union of subspaces of V satisfy the first first and third property of theorem 3. However, property 2 need not be satisfied all the time. In fact, the union can only be a subspace of V if and only if one the subspaces is a subset of the other.

1.3 Linear Combinations and Systems of Linear Equations

1.3.1 Linear Combinations

Definition 1.3.1 (Linear Combinations). Let V be a vector space and S a nonempty subset of V. A vector $v \in V$ is called **linear combination** of vectors of S if there exist a finite number of vectors $u_1, u_2, \ldots, u_n \in S$ and scalars $a_1, a_2, \ldots, a_n \in F$ such that

$$v = a_1u_1 + a_2u_2 + \cdots + a_nu_n.$$

In this case, we say that v is a linear combination of $u_1, u_2, \ldots, u_n \in V$ and call $a_1, a_2, \ldots, a_n \in F$ the **coefficients** of the linear combination.

- An easy example of a vector expressed as a linear combination is the zero vector O where $0v = O$ for each $v \in V$.
- Determining whether a vector is a linear combination of other vectors often involves solving a system of linear equations.

Example 1.3.1. Suppose we wanted to express the vector $(2, 6, 8) \in \mathbb{R}^3$ as a linear combination of

$$u_1 = (1, 2, 1),\ u_2 = (-2, -4, -2),\ u_3 = (0, 2, 3),\ u_4 = (2, 0, -3), \text{ and } u_5 = (-3, 8, 16).$$

Our goal is to find scalars a_1, a_2, a_3, a_4 and a_5 such that

$$(2, 6, 8) = a_1u_1 + a_2u_2 + a_3u_3 + a_4u_4 + a_5u_5.$$

Doing a bit of algebra, we would need to solve the following system of linear equations

$$\begin{aligned} a_1 - 2a_2 + 2a_4 - 3a_5 &= 2 \\ 2a_1 - 4a_2 + 2a_3 + 8a_5 &= 6 \\ a_1 - 2a_2 + 3a_3 + 16a_5 &= 8 \end{aligned}$$

Solving the system of equations above involves three types of operations:

1. Interchanging the order of any two equations in the system;
2. multiplying any equation in the system by some non-zero constant;
3. adding a constant multiple of any equation to another equation in the system.

We will learn in a later section that the operations listed above do not change the set of solutions to the original system.

Example 1.3.2. We claim that

$$2x^3 - 2x^2 + 12x - 6$$

is a linear combination of

$$x^3 - 2x^2 - 5x - 3 \text{ and } 3x^3 - 5x^2 - 4x - 9$$

in $P_3(\mathbb{R})$, but that

$$3x^3 - 2x^2 + 7x + 8$$

is not. In the first case we wish to find scalars a and b such that

$$\begin{aligned} 2x^3 - 2x^2 + 12x - 6 &= a(x^3 - 2x^2 - 5x - 3) \\ &\quad + b(3x^3 - 5x^2 - 4x - 9) \\ &= (a + 3b)x^3 + (-2a - 5b)x^2 \\ &\quad + (-5a - 4b)x + (-3a - 9b). \end{aligned}$$

Thus, we have the following system of linear equations:

$$\begin{aligned} a + 3b &= 2 \\ -2a - 5b &= -2 \\ -5a - 4b &= 12 \\ -3a - 9b &= -6. \end{aligned}$$

Adding the appropriate multiples of the first equation to the others in order to eliminate a, we find that

$$\begin{aligned} a + 3b &= 2 \\ b &= 2 \\ 11b &= 22 \\ 0b &= 0. \end{aligned}$$

Then we get that $a = -4, b = 2, 0 = 0, 0 = 0$. Hence, we have

$$2x^3 - 2x^2 + 12x - 6 = -4(x^3 - 2x^2 - 5x - 3) + 2(3x^3 - 5x^2 - 4x - 9).$$

Looking at the second case now, we observe that using the preceding technique leads us to the following system of linear equations

$$\begin{aligned} a + 3b &= 3 \\ -2a - 5b &= -2 \\ -5a - 4b &= 7 \\ -3a - 9b &= 8. \end{aligned}$$

Eliminating a as before yields the following:

$$\begin{aligned} a + 3b &= 3 \\ b &= 4 \\ 11b &= 22 \\ 0 &= 17. \end{aligned}$$

The presence of the non-sensical result in the last equality tells us that there are no solutions to the system and that the second polynomial cannot be written as a linear combination of the two polynomials.

We can denote the set of all linear combinations of some set of vectors in the following:

1.3.2 The Span

Definition 1.3.2 (Span). Let S be a nonempty subset of a vector space V. The **span** of S, denoted span(S), is the set consisting of all linear combinations of the vectors in S. For convenience, we define span$(\emptyset) = \{0\}$.

Some immediate examples of spans are:

Example 1.3.3. In $\mathbb{R}^3$, the span of the set $S = \{(1,0,0),(0,1,0)\}$ consist of all vectors in $\mathbb{R}^3$ such that for some scalars $a, b \in \mathbb{R}$, we have

$$a(1,0,0) + b(0,1,0) = (a,b,0).$$

This tells us that S contains all the points in the xy-plane. One can show that S is a subspace of $\mathbb{R}^3$.

Theorem 1.3.1 (The Span of Subset is a Subspace). The span of any subset S of a vector space v is a subspace of V that contains S. Moreover, any subspace of V that contains S must also contain the span of S.

Proof. Suppose $S = \emptyset$. Then the span of S is just span$(\emptyset) = \{0\}$ which is a subspace in which S is contained in. Moreover, span$(\emptyset) = \{0\}$ is always contained in any subspace X of V because every subspace contains the zero vector 0. Suppose $S \neq \emptyset$, then S contains a vector z. We need to show that span(S) is a subspace of V.

(a) Since $S \neq \emptyset$, we know that S contains a vector v such that $0z = 0$. Hence, $0 \in$ span(S).

(b) Let $x, y \in$ span(S). We need to show that $x + y \in$ span(S). If $x \in$ span(S), then we can find $a_1, a_2, \dots, a_n \in F$ and $x_1, x_2, \dots, x_n \in S$ such that

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n.$$

Likewise, $y \in$ span(S) implies that we can find scalars $b_1, b_2, \dots, b_m$ and vectors $y_1, y_2, \dots, y_m$ such that

$$b_1y_1 + b_2y_2 + \cdots + b_my_m.$$

Then

$$x + y = a_1x_1 + a_2x_2 + \cdots + a_nx_n + b_1y_1 + b_2y_2 + \cdots + b_my_m$$

is a linear combination and thus $x + y \in$ span(S).

(c) Note that $c \in F$ implies that

$$\begin{aligned} cx &= c(a_1x_1 + a_2x_2 + \cdots + a_nx_n) \\ &= (ca_1)x_1 + (ca_2)x_2 + \cdots + (ca_n)x_n \end{aligned}$$

is a linear combination and thus $cx \in$ span(S).

Hence, span(S) is a subspace of V. Now we need to show that S is contained within span(S); that is, $S \subseteq$ span(S). Let $v \in S$. Then using (VS 5), we can see that $1 \cdot v = v$ is a linear combination; so we have $v \in$ span(S) and hence, the span of S contains S.

Now let W be any subspace of V that contains S. We need to show that $\text{span}(S) \subseteq W$. Let $v \in \text{span}(S)$. Then we can find scalars $a_1, a_2, \ldots, a_n \in F$ and $x_1, x_2, \ldots, x_n \in S$ such that

$$v = a_1x_1 + a_2x_2 + \cdots + a_nx_n.$$

Since $S \subseteq W$, we know that $x_1, x_2, \ldots, x_n \in W$. Using exercise 20 from section 1.3 and using the same set of scalars $a_1, a_2, \ldots, a_n \in F$, we have $v = a_1x_1 + a_2x_2 + \cdots + a_nx_n \in W$. Hence, $\text{span}(S) \subseteq W$. ■

1.3.3 Generating a Vector Space

Definition 1.3.3 (Generating/Spanning Sets). A subset S of a vector space V **generates** (or **spans**) V if $\text{span}(S) = V$. In this case, we also say that the vectors of S generates (or span) V.

Example 1.3.4 (Vectors in $\mathbb{R}^3$). The vectors $(1,1,0), (1,0,1), (0,1,1)$ generate $\mathbb{R}^3$ since any given vector $v \in \mathbb{R}^3$ is a linear combination of the three given vectors. Furthermore, there exists scalars $r, s, t \in \mathbb{R}$ such that

$$r(1,1,0) + s(1,0,1) + t(0,1,1) = (a_1, a_2, a_3) = v$$

where

$$r = \tfrac{1}{2}(a_1 + a_2 - a_3), s = \tfrac{1}{2}(a_1 - a_2 + a_3) \text{ and } t = \tfrac{1}{2}(-a_1 + a_2 + a_3).$$

Example 1.3.5 (Polynomials). The polynomials $x^2 + 3x - 2, 2x^2 + 5x - 3$, and $-x^2 - 4x + 4$ generate $P_2(\mathbb{R})$ because each of the three given polynomials belongs to $P_2(\mathbb{R})$ and each polynomial $ax^2 + bx + c \in P_2(\mathbb{R})$ is a linear combination of these three. It can be shown that we can find $a, b, c \in \mathbb{R}$ such that

$$a(x^2 + 3x - 2) + b(2x^2 + 5x - 3) + c(-x^2 - 4x + 4) = ax^2 + bx + c.$$

Example 1.3.6 (Matrices). Note that the matrices

$$\begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}, \text{ and } \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}$$

generate $M_{2\times2}(\mathbb{R})$ because an every $A \in M_{2\times2}(\mathbb{R})$ can be expressed as a linear combination of the four given matrices found below where there exists scalars $a, b, c, d \in \mathbb{R}$ such that

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = a\begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} + b\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} + c\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} + d\begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}$$

where it can be show that

$$\begin{aligned} a &= \frac{1}{3}a_{11} + \frac{1}{3}a_{12} + \frac{1}{3}a_{21} - \frac{2}{3}a_{22} \\ b &= \frac{1}{3}a_{11} + \frac{1}{3}a_{12} - \frac{2}{3}a_{21} + \frac{1}{3}a_{22} \\ c &= \frac{1}{3}a_{11} - \frac{2}{3}a_{12} + \frac{1}{3}a_{21} + \frac{1}{3}a_{22} \\ d &= -\frac{2}{3}a_{11} + \frac{1}{3}a_{12} + \frac{1}{3}a_{21} + \frac{1}{3}a_{22}. \end{aligned}$$

A non-example set of matrices are

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \text{ and } \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$$

do not generate $M_{2\times 2}(\mathbb{R})$ since all of the matrices have equal diagonal entries. Thus, every the set of matrices above cannot generate every $A \in M_{2\times 2}(\mathbb{R})$.

1.4 Linear Dependence and Linear Independence

1.4.1 Motivation

Suppose V is a vector space over a field F and that $W \subseteq V$ is a subspace of V.

- Suppose we have a set S that will generate every vector in V as a linear combination of vectors in S.
- It is desirable to find a subset of S that is as small as possible or rather just enough to generate each vector in V.
- The reason for this is to reduce the amount of computations done to represent a vector in V.
- Finding this subset is a matter of expressing one of the vectors in the spanning set as a linear combination of the other vectors in S.
- This will naturally lead to a system of linear equations for which we have to solve for the coefficients.
- The answer to the question is whether this system of equations leads to a solution or not.
- An easier way to answer this question is to find coefficients (not all zero) for which the zero vector can be written as a linear combination of vectors in S.
- Doing this, guarantees that we can write any one of the vectors in S as a linear combination of the others.

Definition 1.4.1 (Linearly Dependence). A subset S of a vector space V is called **linearly dependent** if there exists a finite number of distinct vectors in $u_1, u_2, \ldots, u_n \in S$ and scalars $a_1, a_2, \ldots, a_n$ not all zero, such that

$$a_1u_1 + a_2u_2 + \cdots + a_nu_n = 0.$$

In this case, we also say that the vectors of S are linearly dependent.

- The **trivial representation** of 0 is a linear combination of scalars $a_1, a_2, \ldots, a_n \in F$ and distinct vectors $v_1, v_2, \ldots, v_n \in S$ where for all $1 \leq i \leq n$, we have $a_i = 0$.

- This tells us that our definition of **linear dependence** implies that 0 is a non-trivial linear combination.

- Any subset of V that contains the zero vector is subsequently linearly dependent since 0 can be written as non-trivial representation; that is, $1 \cdot 0 = 0$.

1.4.2 Examples of Linearly Dependent Sets

Example 1.4.1. Consider a subset in $\mathbb{R}^4$ defined by

$$S = \{(1, 3, -4, 2), (2, 2, -4, 0), (1, -3, 2, -4), (-1, 0, 1, 0)\}.$$

To show that S is a linear dependent set, we need to find scalars $a_1, a_2, a_3, a_4 \in \mathbb{R}$ such that a_i for all $1 \leq i \leq 4$ not all zero such that

$$a_1(1, 3, -3, 2) + a_2(2, 2, -4, 0) + a_3(1, -3, 2, -4) + a_4(-1, 0, 1, 0) = 0.$$

We can represent this as a system of linear equations with each equation being set equal to zero and solving for the scalars. Solving this system of equations leads to the coefficients $a_1 = 4, a_2 = -3, a_3 = 2$, and $a_4 = 0$. Thus, we have that S is linearly dependent subset of $\mathbb{R}^4$ and hence we can write any vector in S as a linear combination of the other vectors contained in S.

Example 1.4.2. Define a subset of $M_{2\times 2}(\mathbb{R})$

$$S = \left\{ \begin{pmatrix} 1 & -3 & 2 \\ -4 & 0 & 5 \end{pmatrix} \}, \begin{pmatrix} -3 & 7 & 4 \\ 6 & -2 & -7 \end{pmatrix}, \begin{pmatrix} -2 & 3 & 11 \\ -1 & -3 & 2 \end{pmatrix} \right\}.$$

This set is **linearly dependent** because we can find coefficients a_1, a_2, a_3 such that $a_1 = 5, a_2 = 3$, and $a_3 = -2$ where

$$5 \begin{pmatrix} 1 & -3 & 2 \\ -4 & 0 & 5 \end{pmatrix} + 3 \begin{pmatrix} -3 & 7 & 4 \\ 6 & -2 & -7 \end{pmatrix} - 2 \begin{pmatrix} -2 & 3 & 11 \\ -1 & -3 & 2 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

Definition 1.4.2 (Linear Independence). A subset S of a vector space that is not linearly dependent is called **linearly independent**. As before, we also say that the vectors of S are linearly independent.

Whereas linear dependence requires solutions that are not all zero, linear independence **requires** all the solutions to be zero. This provides us an easy way to determine if a finite set is linearly independent.

1.4.3 Examples of Linear Independent Sets

Some facts about linearly independent sets include:

- The empty set is linearly independent since we don't have any vectors to take linear combinations of.

- A set consisting of only one non-zero vector; that is, $\{v\}$ is linearly independent.
- If $\{v\}$ is linearly dependent, then it the singleton has to be the zero vector 0. This is because $au = 0$
$$u = 1 \cdot u = (a^{-1}a)u = a^{-1}(au) = a^{-1} \cdot 0 = 0.$$

Example 1.4.3 (A Set of Vectors in $\mathbb{R}^4$)**.** It can be shown that the set

$$S = \{(1, 0, 0, -1), (0, 1, 0, -1), (0, 0, 1, -1), (0, 0, 0, 1)\}$$

is linearly independent by showing that all the coefficients of the linear combination of 0 represent the trivial representation; that is, we have scalars $a_1 = a_2 = a_3 = a_4 = 0$.

Example 1.4.4. For $k = 0, 1, \dots, n$, let $p_k = x^k + x^{k+1} + \cdots + x^n$. The set

$$\{p_0(x), p_1(x), \dots p_n(x)\}$$

is linearly independent in $P_n(F)$. It can be shown that for some scalars $a_0, a_1, \dots, a_n \in F$, the following equation

$$a_0 + (a_0 + a_1 + a_2)x^2 + \cdots + (a_0 + a_1 + \cdots + a_n)x^n = 0$$

has the trivial-representation; that is, $a_i = 0$ for all $1 \leq i \leq n$.

Theorem 1.4.1 (Linearly Dependent Subsets)**.** Let V be a vector space, and let $S_1 \subseteq S_2 \subseteq V$. If S_1 is linearly dependent, then S_2 is linearly dependent.

Proof. Let V be a vector space. Suppose S_1 is linearly dependent. Then there exists a finite number of distinct vectors $v_1, v_2, \dots, v_n \in S_1$ and scalars $a_1, a_2, \dots, a_n \in F$ not all zero such that
$$a_1v_1 + a_2v_2 + \cdots + a_nv_n = 0.$$
Since $S_1 \subseteq S_2$, we must have $v_1, v_2, \dots, v_n \in S_2$ as well. If these vectors are linearly dependent, then these vectors are also linearly dependent in S_2; that is, we have scalars $a_1, a_2, \dots, a_n \in F$ not all zero such that

$$a_1v_1 + a_2v_2 + \cdots + a_nv_n = 0.$$

Hence, S_2 is linearly dependent. ■

Corollary (Linearly Independent Subsets)**.** Let V be a vector space, and let $S_1 \subseteq S_2 \subseteq V$. If S_2 is linearly independent, then S_1 is linearly independent.

Proof. Note that this corollary is just the contrapositive of the theorem before it. Hence, S_1 is linearly independent. ■

- Determining whether there exists a minimal generating set for the span of S is related to solving the problem of finding whether some vector in S that can be written in terms of a linear combination of other vectors in S.
- We can see that in a given subset of $\mathbb{R}^3$ defined by
$$S = \{u_1, u_2, u_3, u_4\}$$

where $u_1 = (2, -1, 4), u_2 = (1, -1, 3), u_3 = (1, 1, -1)$, and $u_4 = (1, -2, 1)$. Note that this S is linearly dependent. This tells us that any vector in S can be written as a linear combination of the others. Suppose we pick u_3 and write as a linear combination of the vectors u_1, u_2, and u_4. We would find that the span of these three vectors (denote this set as S') generates the same spanning set S but with a cardinality one less than S; that is, $\text{span}(S) = \text{span}(S')$.

- If we find that there does not exist a proper subset that is equivalent to the span of S, then S must be a linearly independent set.

Theorem 1.4.2 (Adjoining Elements not in Linearly Independent Sets). Let S be a linearly independent subset of a vector space V, and let $v \in V$ but not in S. Then $S \cup \{v\}$ is linearly dependent if and only if $v \in \text{span}(S)$.

Proof. ($\Rightarrow$) Suppose $S \cup \{v\}$ is a linearly dependent set. Then there exists a finite number of scalars $a_1, a_2, \ldots, a_n \in F$ and vectors $u_1, u_2, \ldots, u_n \in S \cup \{v\}$ such that

$$a_1u_1 + a_2u_2 + \cdots + a_nu_n = 0.$$

Since S is a linearly independent set, we must have $u_i = v$ for some $1 \leq i \leq n$. Choose $i = 1$ (any choice of i will do) such that

$$a_1v + a_2u_2 + \cdots + a_nu_n = 0.$$

Solving for v by subtracting a_1v on both sides, multiplying by a_1^{-1} on both sides of the equation, and distributing by a_1^{-1} yields the following:

$$v = (-a_1^{-1}a_2)u_2 + (-a_1^{-1}a_3)u_3 + \cdots + (-a_1^{-1}a_n)u_n.$$

Since v is a linear combination of vectors $u_1, u_2, \ldots, u_n \in S$, we know that $v \in \text{span}(S)$.

($\Leftarrow$) Conversely, suppose $v \in \text{span}(S)$. This implies that there exists a finite amount of scalars $a_1, a_2, \ldots, a_n \in F$ and $u_1, u_2, \ldots, u_n \in S$ such that

$$v = a_1u_1 + a_2u_2 + \cdots + a_nu_n.$$

Subtracting v from both sides of the equation above gives us the following equation:

$$a_1u_1 + a_2u_2 + \cdots + a_nu_n - v = 0.$$

Note that v is not contained in S, so $v \neq u_i$ for all $1 \leq i \leq n$. Since S is a linearly independent set, we know that $a_i = 0$ for all $1 \leq i \leq n$. This implies that the only coefficient that is non-zero is with $-v = -1v$. Hence, the set of vectors $S' = \{u_1, u_2, \ldots, u_n, v\}$ is linearly dependent. Since $S' \subseteq S \cup \{v\}$, we know that $S \cup \{v\}$ is also linearly independent by Theorem 6. ■

1.5 Bases And Dimension

1.5.1 Bases

- Recall that S is a generating set for a subspace W and no proper subset of S is a generating set for W, then S must be linearly independent.

- Linearly independent sets possess the unique property that every vector that its spanning set generates is unique.
- This is property is what allows generating sets to be the building blocks of vector spaces.

Definition 1.5.1 (Basis). A **basis** β for a vector space V is linearly independent subset of V that generates V. If β is a basis for V, we also say that the vectors of β form a basis for V.

Example 1.5.1.
- Recall that the empty set $\emptyset$ is linearly independent and that $\text{span}(\emptyset) = \{0\}$. The empty set $\emptyset$ in this case is the basis for the zero vector space.
- Note that in F^n, the vectors $e_1 = (1, 0, \dots, 0)$, $e_2 = (0, 1, 0, \dots, 0) \dots, e_n = (0, 0, \dots, 0, 1)$ form a basis for F^n.
- The basis for $M_{m\times n}(F)$ is the set of matrices E^{ij} such that the only nonzero entry is a 1 in the ith and jth column.
- As we have seen in the last section, the set $\{1, x, x^2, \dots, x^n\}$ is a basis for $P_n(F)$.
- In $P(F)$, the set $\{1, x, x, x^2, \dots\}$ is a basis. *Bases are not limited to finite sets. They can be infinite.*

Theorem 1.5.1 (Unique Linear Combinations from Bases). Let V be a vector space and $u_1, u_2, \dots, u_n$ be distinct vectors in V. Then $\beta = \{u_1, u_2, \dots, u_n\}$ is a basis for V if and only if each $v \in V$ can be unique expressed as a linear combination of vectors in β, that is, expressed in the form

$$v = a_1v_1 + a_2v_2 + \cdots + a_nv_n$$

for unique scalars $a_1, a_2, \dots, a_n$.

Proof. ($\Rightarrow$) Suppose $\beta = \{u_1, u_2, \dots, u_n\}$ is a basis for V. Then $\text{span}(\beta) = V$. If $v \in V$, then $v \in \text{span}(\beta)$. Hence, we can write v as a linear combination of vectors in β such that choosing scalars $a_1, a_2, \dots, a_n \in F$ leads to

$$v = \sum_{i=1}^{n} a_iu_i.$$

Suppose there exists another representation of $v \in V$ such that

$$v = \sum_{i=1}^{n} b_ix_i$$

Hence, observe that

$$\sum_{i=1}^{n} a_ix_i = \sum_{i=1}^{n} b_iy_i$$
$$\Rightarrow \sum_{i=1}^{n} (a_i - b_i)x_i = 0.$$

Since β is linearly independent, we know that $a_i - b_i = 0$ which implies $a_i = b_i$ for all $1 \leq i \leq n$. Hence, v can be expressed as a unique linear combination of vectors in β.

($\Leftarrow$) Conversely, let $v \in V$ be expressed as a unique linear combination of vectors in β such that

$$v = a_1u_1 + a_2u_2 + \cdots + a_nu_n \tag{1}$$

for unique scalars $a_1, a_2, \ldots, a_n$. We want to show that β is a basis for V; that is, we want to show that β is a spanning set for V and β is linearly independent. To show that β is a spanning set for V, we need to show that $\text{span}(\beta) \subseteq V$ and $V \subseteq \text{span}(\beta)$. Note that $\beta \subseteq V$ and $\text{span}(\beta)$ is a subspace for V. Hence, $\text{span}(\beta) \subseteq V$. On the other hand, $V \subseteq \text{span}(\beta)$ follows immediately from (1). To show that β is linearly independent, we need to show that for scalars $\delta_1, \delta_2, \ldots, \delta_n$, we have

$$\delta_1u_1 + \delta_2u_2 + \cdots + \delta_nu_n = 0$$

such that $u_i = 0$ for all $1 \leq i \leq n$. Note that $v + 0 = v$. Hence, we can write

$$\begin{aligned}\sum_{i=1}^{n} a_iu_i + \sum_{i=1}^{n} \delta_iu_i &= \sum_{i=1}^{n} a_iu_i \\ \sum_{i=1}^{n}(a_i + \delta_i)u_i &= \sum_{i=1}^{n} a_iu_i\end{aligned}$$

Equating each term in the equation above yields the following equality:

$$a_i + \delta_i = a_i \Rightarrow \delta_i = 0$$

for all $1 \leq i \leq n$. But this tells us that β is linearly independent and we are done. ■

- Any vector $v \in V$ can be written as a linear combination of vectors from the basis containing $u_1, u_2, \ldots, u_n \in V$.
- This determines a unique n-tuple of scalars $(a_1, a_2, \ldots, a_n)$ and conversely, each n-tuple of scalars determines a unique vector $v \in V$ such that each coefficient from the linear combination of $u_1, u_2, \ldots, u_n$ is an entry from said tuple.
- For example, in our vector space F^n, n is the number of vectors that should be in the basis for F^n which is indeed the case.
- In this book, we are only concerned with finite bases.

Theorem 1.5.2 (Finite Spanning Set For a Vector Space). If a vector space V is generated by a finite set S, then some subset of S is a basis for V. Hence, V has a finite basis.

Proof. Suppose $S = \emptyset$ or $S = \{0\}$, then $V = \{0\}$ and $\emptyset$ is a subset of S that is a basis for V. If S neither of these choices, then S must contain at least one nonzero vector u_1; that is, $S = \{u_1\}$. Since u_1 is nonzero, it follows that S is a linearly independent set. We can continue this process of adding vectors $u_2, \ldots, u_k$ into S such that S is a linearly independent set of k vectors. Since S is a finite set, we must end with the linearly independent set $\beta = \{u_1, u_2, \ldots, u_n\}$. There are two cases for which this occurs, either $\beta = S$ or $\beta \subseteq S$:

(i) Suppose that $\beta = S$ (remember that β is a finite set by construction). Then we have S is a linearly independent set and spanning set for V (since S is a finite set that generates V). Hence, S is a finite basis for V.

(ii) Suppose $\beta \subseteq S$ is a linearly independent set such that adding $v \in S$ where $v \notin \beta$ makes a linearly dependent set. We claim that β is the desired subset of S that is a basis for V. Then we have two cases; that is, either $v \in \beta$ or $v \notin \beta$. Since $\beta \subseteq S$, we know by Theorem 1.5 that $\text{span}(\beta) \subseteq S$ (This applies for both cases). It suffices to show that $S \subseteq \text{span}(\beta)$. If $v \in \beta$, then surely $v \in \text{span}(\beta)$. Hence, we have $S \subseteq \text{span}(\beta)$. Suppose $v \notin \beta$. Since β is a linearly independent set, then by Theorem 1.7 , we have that $\beta \cup \{v\}$ being linearly dependent implies that $v \in \text{span}(\beta)$. Hence, $S \subseteq \text{span}(\beta)$. Thus, that both cases implies that β is a spanning set for V.

■

This theorem tells us that any spanning set of a vector space V can reduced to a finite basis for V. This is illustrated in the following examples.

Example 1.5.2. Define

$$S = \{(2, -3, 5), (8, -12, 20), (1, 0, -2), (0, 2, -1), (7, 2, 0)\}$$

which can be shown to generate $\mathbb{R}^3$. The idea is to create a proper subset of S such that none of the vectors in β are a multiples of each other. In other words, we want a set that is linearly independent; that is, our choices of vectors determine whether a set will be linearly dependent or independent (we want the latter to hold). Say, we pick $(2, -3, 5)$ as our first vector in our subset β. Right away, we can exclude $(8, -12, 20)$ since it is a multiple of $(2, -3, 5)$. Otherwise, including it would make β linearly dependent (see exercise 9 from section 1.5). Next, add the vectors $(1, 0, -2)$ and $(0, 2, -1)$ since they are not multiples of the other. Narrowing down our set to

$$\beta = \{(2, -3, 5), (1, 0, -2), (0, 2, -1)\}.$$

Note that adding the vector $(7, 2, 0)$ makes β linearly dependent, so $(7, 2, 0)$ is excluded from the list. Hence, we have arrived at a subset of S such that $\beta \subseteq S$ is both a linearly independent set and spanning set for $\mathbb{R}^3$.

The following theorem and its corollaries are the most important results in the Chapter 1.

1.5.2 Dimensions

Theorem 1.5.3 (Replacement Theorem)**.** Let V be a vector space that is generated by a set G containing exactly n vectors, and let L be a linearly independent subset of V containing exactly m vectors. Then $m \leq n$ and there exists a subset H of G containing exactly $n - m$ vectors such that $L \cup H$ generates V.

Proof. Let us proceed the proof via induction on m. Let $m = 0$ be our base case. Then we find that $L = \emptyset$ is linearly independent set with exactly 0 vectors. Letting $H = G$ gives us the desired result (since G contains exactly n vectors). Now suppose that the theorem is holds for some integer $m \geq 0$. We will show that the theorem holds for the $m+1$ case.Let $L = \{v_1, v_2, \ldots, v_{m+1}\}$ be a linearly independent subset of V consisting of exactly $m+1$ vectors. By the corollary to Theorem 1.6 , we find that $L' = \{v_1, v_2, \ldots, v_m\}$ is a linearly independent set (because $L' \subseteq L$ and L is linearly independent). Using our induction hypothesis, we can conclude that $m \leq n$ and that there exists a subset $H' = \{u_1, u_2, \dots, u_{n-m}\}$ of G such that $L' \cup H'$ generates V. Thus there exists scalars $a_1, a_2, \dots a_m, b_1, b_2, \dots, b_{n-m}$ such that

$$v_{m+1} = a_1v_1 + a_2v_2 + \cdots + a_mv_m + b_1u_1 + b_2u_2 + \cdots + b_{n-m}u_{n-m}. \tag{1}$$

Note that $n-m > 0$, **unless** v_{m+1} is a linear combination of $v_1, v_2, \dots, v_m$ which by Theorem 1.7 contradicts the assumption that L is a linearly independent set. Hence, $n > m$; that is, $n \geq m+1$. Furthermore, some b_i, say b_1 is nonzero, for otherwise we obtain the same contradiction. Solving (1) for u_1, we get

$$\begin{aligned} u_1 = (-b_1^{-1}a_1)v_1 + (-b_1^{-1}a_2)v_2 + \cdots + (-b_1^{-1})v_m + (b_1^{-1})v_{m+1} \\ + (-b_1^{-1}b_2)u_2 + \cdots + (-b_1^{-1}b_{n-m})u_{n-m}. \end{aligned}$$

Let $H = \{u_2, u_3, \dots, u_{n-m}\}$. Then $u_1 \in \text{span}(L \cup H)$ and because $v_1, v_2, \dots, v_m, u_2, \dots, u_{n-m}$ are clearly in $\text{span}(L \cup H)$, we have that

$$L' \cup H' \subseteq \text{span}(L \cup H).$$

Since $L' \cup H'$ generates V and the fact that $L' \cup H' \subseteq \text{span}(L \cup H)$ (note that $\text{span}(L \cup H)$ is also a subspace), we know by Theorem 1.5 that $\text{span}(L' \cup H') \subseteq \text{span}(L \cup H)$. Since $\text{span}(L' \cup H')$ generates V, we know that $V \subseteq \text{span}(L' \cup H') \subseteq \text{span}(L \cup H)$. Observe that $\text{span}(L \cup H) \subseteq V$ is true by default. Hence, $\text{span}(L \cup H)$ generates V and that H contains $(n-m)-1 = n-(m+1)$ vectors which concludes our induction proof. ■

Corollary. Let V be a vector space having a finite basis. Then all bases for V are finite, and every basis for V contains the same number of vectors.

Proof. Let β be a finite basis for V that contains exactly n vectors. Let γ be any other basis that contains more than n vectors. Suppose we pick a subset S of γ such that this subset contains exactly $n+1$ vectors. Since β is a finite basis for V, we know that β is a linearly independent set. By Theorem 1.6, we know that S is also a linearly independent set. By Replacement Theorem, we have that $n+1 \leq n$ which is a contradiction. Therefore, we must have γ is finite, and γ contains exactly m amount of vectors in γ which satisfies $m \leq n$. To show that $n \leq m$, we can reverse the roles of β and γ and use the same argument as above. Hence, $n \leq m$ and thus $m = n$. ■

The main takeaway from the corollary above is that the number of vectors in any basis for V is an inherent property of V.

Definition 1.5.2 (Finite-dimensional). A vector space is **finite-dimensional** if it has a basis consisting of a finite number of vectors.

Definition 1.5.3 (Dimension of a Vector Space). The unique integer n such that every basis for V contains exactly n elements is called the **dimension** of V and is denoted by $\dim(V)$.

Definition 1.5.4 (Infinite-dimensional). A vector space that is not finite-dimensional is called **infinite-dimensional**.

Example 1.5.3.

- The vector space $\{0\}$ has dimension zero. This is because $\{0\}$ is generated by the empty set $\emptyset$.

- The vector space F^n has dimension n.
- The vector space $M_{m\times n}(F)$ has dimension mn.
- The vector space $P_n(F)$ has dimension $n+1$.

It turns out that the dimension of a vector space depends on the choice of the field it's defined on.

Example 1.5.4.
- Over the field of complex numbers, the vector space of complex numbers has dimension 1. (A basis is $\{1\}$).
- On the other hand, if $\mathbb{C}$ is defined over the real numbers $\mathbb{R}$, then the vector space has dimension 2.

The first conclusion of the replacement theorem states that given a finite-dimensional vector space, a linearly independent subset of V cannot contain no more than $\dim(V)$ amount of vectors.

Example 1.5.5 (Example of an Infinite-Dimensional Vector Space). The vector space $P(F)$ is infinite-dimensional since it contains a linearly independent set $\{1, x, x^2, \dots\}$ that is infinite. This is peculiar because all of the results in this section thus far does not guarantee that an infinite-dimensional vector space containing a basis. In fact, it is proven in section 1.7 that infinite-dimensional vector space are guaranteed to have a basis.

The next corollary from the Replacement Theorem refers to the size of the generating set.

Corollary. Let V be a vector space with dimension n.

(a) Any finite generating set for V contains at least n vectors, and a generating set for V that contains exactly n vectors is a basis for V.

(b) Any linearly independent subset of V that contains exactly n vectors is a basis for V.

(c) Every linearly independent subset of V can be extended to a basis for V, that is, if L is a linearly independent subset of V, then there is a basis β of V such that $L \subseteq \beta$.

Proof. Let β be a basis for V.

(a) Let β be a finite generating set for V. By Theorem 1.9, β contains a subset S such that S is a finite basis for V. Since V is a vector space with dimension n, we know that S contains exactly n vectors. This means that β must contain at least n vectors since $S \subseteq \beta$.

(b) Let L be a linearly independent subset of V that contains exactly n vectors. Since β is a generating set for V, the Replacement Theorem states that we can find a subset of β, say L', such that L' contains exactly $n-n=0$ vectors implies that $L'=\emptyset$ and $L\cup L' = L\cup\emptyset = L$ generates V. Since L is also linearly independent, we conclude that L is a basis for V.

(c) Let L be a linearly independent subset of V containing m vectors. Then the Replacement Theorem implies that there exists $H \subseteq \beta$ containing $n-m$, where β is the generating set of V with exactly n amount of vectors, such that $L\cup H$ generates V and is finite. This implies that $L \cup H$ must contain at most n vectors, but part (a) also tells us that $L \cup H$ contains at least n vectors, so $L \cup H$ must contain exactly n vectors. Hence, $L \cup H$ is a basis for V.

■

Example 1.5.6.

- In Example 1.3.5 the set

$$\{x^2 + 3x - 2, 2x^2 + 5x - 3, -x^2 - 4x + 4\}$$

is a generating set for $P_2(\mathbb{R})$ that contains exactly 3 vectors. Hence, the set above is a basis for $P_2(\mathbb{R})$ by part (a).

- It follows from Example 1.3.6 and (a) of the second corollary to the Replacement Theorem that the set

$$\left\{ \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} \right\}$$

is a basis for $M_{2\times 2}(\mathbb{R})$.

- The set in Example 1.4.3 is a basis for $\mathbb{R}^4$.

- See Example 1.4.4. This set is a basis for $P_n(F)$.

1.5.3 An Overview of Dimension and Its Consequences

The main takeaways from our results are:

- We can reduce a finite generating set into a basis for a vector space V.
- A finite generating set contains at least n amount of vectors.
- A linearly independent set of vectors contain at most n amount of vectors.
- We can Extend a linearly independent set into a basis for V.
- A finite basis for V must contain exactly n vectors.
- Any basis for V contains the same number of vectors. In other words, if a basis for V, say β, contains n amount of vectors then any other basis must contain the same amount.
- The number of vectors in a basis is the dimension of a vector space.

1.5.4 The Dimension of Subspaces

We can relate the dimension of a vector space V to a subspace of V.

Theorem 1.5.4 (Dimension of Subspaces)**.** Let W be a subspace of a finite-dimensional vector space V. Then W is a finite-dimensional and $\dim(W) \leq \dim(V)$. Moreover, if $\dim(V) = \dim(W)$, then $V = W$.

Proof. Let $\dim(V) = n$ since V is finite-dimensional. Suppose $W = \{0\}$. Then W contains the empty set $\emptyset$ such that $\operatorname{span}(\emptyset) = W$ and $\emptyset$ is linearly independent. Hence, $\emptyset$ is a subset of W that is a basis for $W = \{0\}$ that contains 0 vectors. Hence, W is finite-dimensional. Otherwise, W contains a nonzero vector $= x_1$ such that $W' = \{x_1\}$ is a linearly independent set. Continue this process of adding vectors $x_2, x_3, \ldots, x_k$ into W' such that W' is linearly independent. Since no linearly independent subset of V cannot contain no more than n vectors, this process must stop at $k \leq n$. Note adding a vector from $w \in W$ into W' such that $w \notin W'$ will make a linearly dependent set. By Theorem 1.4.2, we get that $w \in \operatorname{span}(W')$. Hence, W' is a basis for W and that $\dim(W) = k \leq n = \dim(V)$. Suppose $\dim(W) = n$, then there exists a basis β for W such that β contains exactly n amount of vectors. Since β is also linearly independent and contains the same amount of vectors as a basis for V, then we must have that β must also be a basis for V by Corollary 2 of the Replacement Theorem. Hence, $W = \operatorname{span}(\beta) = V$. ∎

Example 1.5.7.

- Let

$$W = \{(a_1, a_2, a_3, a_4, a_5) \in F^5 : a_1 + a_3 + a_5 = 0, a_2 = a_4\}.$$

  It can be easily shown that W is a subspace of F^5 having

$$\{(-1, 0, 1, 0, 0), (-1, 0, 0, 0, 1), (0, 1, 0, 1, 0)\}$$

  as a basis. Hence, $\dim(W) = 3$

- The set of diagonal $n \times n$ matrices is a subspace W of $M_{n\times n}(F)$. A basis for W is the following set

$$\{E^{11}, E^{22}, \ldots, E^{nn}\},$$

  where E^{ij} is the matrix in which the only nonzero entry is a 1 in the ith row and jth column. Thus, we have $\dim(W) = n$.

Corollary. If W is a subspace of a finite-dimensional vector space V, then any basis for W can be extended to a basis for V.

Proof. Let β be a basis for the subspace W. Since $\beta \subseteq W \subseteq V$ is a linearly independent subset of V, we know that β can be extended as a basis for V by Corollary 2 of the Replacement Theorem. ∎

1.5.5 Geometrical View of Subspaces in Euclidean Space

- The subspaces of R^2 include $\{0\}$ of dimension 0, $\mathbb{R}$ of dimension 1, and R^2 itself.
  - The set $\{0\}$ of zero dimension can be visualized as the origin in R^2.
  - The set $\mathbb{R}$ is the field itself and can be visualized as the line $y = cx$ with $c \in \mathbb{R}$ where $x \in \mathbb{R}^2$.
  - The set $\mathbb{R}^2$ is the whole cartesian space itself.
- The subspaces of $\mathbb{R}^3$ must have subspaces with dimensions $0, 1, 2$, or 3. The same logic can be followed from the three bullet points above.

Chapter 2

Linear Transformations and Matrices

2.1 Linear Transformations, Null Spaces, and Ranges

Suppose we have a function T with domain V and codomain W denoted by $T : V \to W$.

Definition 2.1.1 (Linear Transformation). Let V and W be vector spaces (over F). We call a function $T : V \to W$ a **linear transformation from V to W**, for all $x, y \in V$ and $c \in F$, we have

(a) $T(x + y) = T(x) + T(y)$ and

(b) $T(cx) = cT(x)$.

Remark. If $F = \mathbb{Q}$, then (a) implies (b) in the definition above. Otherwise, (a) and (b) are logically independent statements.

The following are a list of properties for linear functions:

Proposition 2.1.1. Let V and W be vector spaces (over F). If $T : V \to W$ is a **linear transformation from V to W**, then the following properties hold:

(a) If T is linear, then $T(0) = 0$.

(b) T is linear if and only if $T(x + y) = T(x) + T(y)$ for all $x, y \in V$ and $c \in F$.

(c) If T is linear, then $T(x - y) = T(x) - T(y)$ for all $x, y \in V$.

(d) T is linear if and only if, for $x_1, x_2, \ldots, x_n \in V$ and $a_1, a_2, \ldots, a_n \in F$, we have

$$T\Big(\sum_{i=1}^{n} a_i x_i\Big) = \sum_{i=1}^{n} a_i T(x_i).$$

Proof. Let V and W be vector spaces such that $T : V \to W$.

(a) If T is linear, then $T(0) = 0$. Suppose T is linear, then $T(0 \cdot 0) = 0T(0) = 0$.

(b) We have T is linear if and only if $T(cx + y) = cT(x) + T(y)$ for all $x, y \in V$ and $c \in F$. Suppose T is linear. Let $x, y \in V$ and $c \in F$. Then

$$T(cx + y) = T(cx) + T(y) = cT(x) + T(y).$$

Conversely, if $c = 1$ then
$$T(x+y) = T(x) + T(y).$$
If $y = 0$, then
$$T(cx) = cT(x).$$
Hence, T is a linear transformation.

(c) If T is linear, then $T(x-y) = T(x) - T(y)$ for all $x, y \in V$. Let $x, y \in V$. Suppose T is linear, then
$$T(x-y) = T(x) + T(-y) = T(x) - T(y).$$

(d) T is linear if and only if, for $x_1, x_2, \ldots, x_n \in V$ and $a_1, a_2, \ldots, a_n \in F$, we have
$$T\Big(\sum_{i=1}^{n} a_i x_i\Big) = \sum_{i=1}^{n} a_i T(x_i).$$
Suppose T is linear. Let $x_1, x_2, \ldots, x_n \in V$ and $a_1, a_2, \ldots, a_n \in F$ such that
$$v = \sum_{i=1}^{n} a_i x_i.$$
Then observe that
$$T\Big(\sum_{i=1}^{n} a_i x_i\Big) = \sum_{i=1}^{n} T(a_i x_i) = \sum_{i=1}^{n} a_i T(x_i).$$
Conversely, for $i = 2$ and denote $a_i x_i = y_i$. Then
$$T\Big(\sum_{i=1}^{2} a_i x_i\Big) = T(a_1 x_1) + T(a_2 x_2) = T(y_1) + T(y_2)$$
and so property 1 is satisfied. If $i = 1$, then
$$T\Big(\sum_{i=1}^{1} a_i x_i\Big) = a_1 T(x_1)$$
and so property 2 is satisfied. Hence, T is a linear transformation.

■

Example 2.1.1. Define
$$T : \mathbb{R}^2 \to \mathbb{R}^2 \text{ by } T(a_1, a_2) = (2a_1 + a_2, a_1).$$
Show that T is linear. Let $c \in \mathbb{R}$ and $x, y \in \mathbb{R}^2$, where $x = (a_1, a_2)$ and $y = (b_1, b_2)$. Then
$$\begin{aligned} cx + y &= c(a_1, a_2) + (b_1, b_2) \\ &= (ca_1, ca_2) + (b_1, b_2) \\ &= (ca_1 + b_1, ca_2 + b_2). \end{aligned}$$

So, we have

$$\begin{aligned}
T(cx+y) &= T(ca_1+b_1, ca_2+b_2)\\
&= (2(ca_1+b_1)+ca_2+b_2, ca_1+b_1)\\
&= ((2ca_1+ca_2)+(2b_1+b_2), ca_1+b_1)\\
&= (2ca_1+ca_2, ca_1)+(2b_1+b_2, b_1)\\
&= c(2a_1+a_2, a_1)+(2b_1+b_2, b_1)\\
&= cT(a_1,a_2)+T(b_1,b_2)\\
&= cT(x)+T(y).
\end{aligned}$$

Hence, we have that $T(cx+y) = cT(x)+T(y)$ so $T : \mathbb{R}^2 \to \mathbb{R}^2$ is a linear.

2.1.1 Examples of Linear Transformations

The most common linear transformations come from geometry:

- Rotations: $T_\theta(a_1, a_2) = (a_1, a_2)$ by performing a counter-clockwise rotation by an angle θ if $(a_1, a_2) \neq (0, 0)$. This is called **rotation by** θ and $T_\theta(0, 0) = (0, 0)$ otherwise.
- Reflections about the $x-$axis: $T(a_1, a_2) = (a_1, -a_2)$.
- Projections on the $x-$axis: $T(a_1, a_2) = (a_1, 0)$.

Example 2.1.2 (Rotations). Define the rotation transformation above by $T_\theta : \mathbb{R}^2 \to \mathbb{R}^2$ by $T_\theta(a_1, a_2)$ where it is described as above. Our goal in this example is to define an explicit formula of this transformation. Let us fix a nonzero vector $(a_1, a_2) \in \mathbb{R}^2$. Let α be the angle such that (a_1, a_2) makes with the positive $x-$axis, and let $r = \sqrt{a_1^2 + a_2^2}$. Then using some trigonometry, we get that $a_1 = r\cos(\alpha)$ and $a_2 = r\sin(\alpha)$. Note that $T_\theta(a_1, a_2)$ has length r and makes an angle $\alpha + \theta$ with the positive $x-$axis. Using some trigonometric identities, we get that

$$\begin{aligned}
T_\theta(a_1, a_2) &= (r\cos(\alpha+\theta), r\sin(\alpha+\theta))\\
&= (r\cos(\alpha)\cos(\theta) - r\sin(\alpha)\sin(\theta), r\cos(\alpha)\sin(\theta) + r\sin(\alpha)\cos(\theta))\\
&= (a_1\cos(\theta) - a_2\sin(\theta), a_1\sin(\theta) + a_2\cos(\theta)).
\end{aligned}$$

Observe that this formula holds for $(a_1, a_2) = (0, 0)$. One can show that T_θ is linear.

Example 2.1.3 (Transpose). Define $T : M_{m\times n}(F) \to M_{n\times m}(F)$ by $T(A) = A^t$, where A^t is the transpose of A, defined in Section 1.3.

Example 2.1.4 (Derivatives of Polynomials). Define $T : P_n(\mathbb{R}) \to P_{n-1}(\mathbb{R})$ by $T(f(x)) = f'(x)$, where $f'(x)$ denotes the derivative of $f(x)$. We can show that T is linear. Let $g(x), h(x) \in P_n(\mathbb{R})$ and $a \in \mathbb{R}$.Assuming that the differentiation rules holds for any T, we have

$$\begin{aligned}
T(ag(x)+h(x)) &= (ag(x)+h(x))'\\
&= ag'(x)+h'(x)\\
&= aT(g(x))+T(h(x)).
\end{aligned}$$

By the second property of linearity , we have that T is linear.

Example 2.1.5 (Integrals). Let $V = C(\mathbb{R})$ be the vector space of continuous real-valued functions on $\mathbb{R}$. Let $a, b \in \mathbb{R}$, $a < b$. Define $T : V \to \mathbb{R}$ by

$$T(f) = \int_a^b f(t)\ dt$$

for all $f \in V$. We know that T is a linear transformation because the definite integral of a linear combination of functions is the same as the linear combination of the definite integrals of the functions; that is,

$$\int_a^b \sum_{i=1}^n \gamma_i f(t_i)\ dt = \sum_{i=1}^n \gamma_i \int_a^b f(t_i)\ dt = \sum_{i=1}^n \gamma_i T(f_i)$$

for scalars γ_i where $1 \leq i \leq n$ and

$$T(f_i) = \int_a^b f(t_i)\ dt.$$

2.1.2 Identity and Zero Transformations

Remark. Two very important examples of linear transformations are the **identity and zero transformations**.

Definition 2.1.2 (Identity Transformation). The **identity transformation** is denoted by $I_V : V \to V$ by $I_V(x) = x$ for all $x \in V$. From now on, we will denote this transformation as I instead of I_V.

Definition 2.1.3 (Zero Transformation). The **zero transformation** is denoted by $T_0 : V \to W$ by $T_0(x) = 0$ for all $x \in V$.

2.1.3 Range and Null Spaces

The *Range* and *Null* spaces give us important insights on the intrinsic properties of a linear transformation.

Definition 2.1.4 (Null Space). Let V and W be vector spaces, and let $T : V \to W$ be linear. The **null space** (or **kernel**), denoted by $N(T)$ of T, is the set of all vectors $x \in V$ such that $T(x) = 0$; that is, $N(T) = \{x \in V : T(x) = 0\}$.

Definition 2.1.5 (Range). Let V and W be vector spaces, and let $T : V \to W$ be linear. The **range** (or **image**) $R(T)$ of T is the subset of W consisting of all images (under T) of vectors in V; that is, $R(T) = \{T(x) : x \in V\}$.

Example 2.1.6. Let V and W be vector spaces, and let $I : V \to V$ and $T_0 : V \to W$ be the identity and zero transformations, respectively. Then $N(I) = \{0\}$, $R(I) = V$, $N(T_0) = V$, and $R(T_0) = \{0\}$.

Example 2.1.7. Let $T : \mathbb{R}^3 \to \mathbb{R}^2$ be the linear transformation defined by

$$T(a_1, a_2, a_3) = (a_1 - a_2, 2a_3).$$

One can verify that

$$N(T) = \{(a, a, 0) : a \in \mathbb{R}\} \text{ and } R(T) = \mathbb{R}^2.$$

Theorem 2.1.1 (The Null and Range are Both Subspaces). Let V and W be vector spaces and $T : V \to W$ be linear. Then $N(T)$ and $R(T)$ are subspaces of V and W, respectively.

Proof. Let $T : V \to W$ be linear. First, we show that $N(T)$ is a subspace of V.

(a) Note that $0_V \in V$ implies that $T(0_V) = 0_W$ since T is linear (property (1) of linearity). Hence, $0_V \in N(T)$.

(b) Let $x, y \in N(T)$. By definition, of the null space we have $T(x) = 0_W$ and $T(y) = 0_W$. Since T is linear, we get that

$$T(x + y) = T(x) + T(y) = 0_W + 0_W = 0_W \Rightarrow x + y \in N(T).$$

Hence, $N(T)$ is closed under addition.

(c) Let $x \in N(T)$ and let $c \in F$. Then

$$T(cx) = cT(x) = c0_V = 0_W \Rightarrow cx \in N(T).$$

Hence, Theorem 1.3 tells us that $N(T)$ is a subspace of V. Now, we will show that $R(T)$ is a subspace of W.

(a) Note that $0_V \in V$ implies that $T(0_V) = 0_W \in R(T)$ by property (1) of linearity. So, $0_W \in R(T)$.

(b) Now let $x, y \in R(T)$. Then there exists elements $z, w \in W$ such that $T(x) = z$ and $T(y) = w$. Then by linearity, we have $T(x + y) = T(x) + T(y) = z + y \in W$. Hence, we must have $x + y \in R(T)$ and thus $R(T)$ is closed under addition.

(c) Let $x \in R(T)$ and $c \in F$. Then we have $T(cx) = cT(x) \in W$ since $T(x) \in W$. Hence, $cx \in R(T)$.

Thus, Theorem 1.3 also tells us that $R(T)$ is a subspace of W. ■

The next theorem allows us to find a spanning set for the range of a linear transformation.

Theorem 2.1.2 (Spanning Set for a Linear Transformation). Let V and W be vector spaces, and let $T : V \to W$ be linear. If $\beta = \{v_1, v_2, \ldots v_n\}$ is a basis for V, then

$$R(T) = \operatorname{span}(T(\beta)) = \operatorname{span}(\{T(v_1), T(v_2), \ldots, T(v_n)\}).$$

Proof. Let $T : V \to W$ be linear where V and W are vector spaces. To show that $T(\beta)$ generates $R(T)$; we need to show that

$$\operatorname{span}(T(\beta)) \subseteq R(T) \text{ and } R(T) \subseteq \operatorname{span}(T(\beta)).$$

Observe that $T(v_i) \in R(T)$ for all i. Hence, $T(\beta) \subseteq R(T)$. By theorem 1.5, we know that $R(T)$ also contains the span of $T(\beta)$. Hence, $\text{span}(T(\beta)) \subseteq R(T)$.

Let $w \in R(T)$. Then for some $v \in W$, we have $T(v) = w$. Since β is a basis for V, we choose scalars $\alpha_1, \alpha_2, \ldots, \alpha_n$ such that

$$v = \sum_{i=1}^{n} \alpha_i v_i.$$

Since T is linear, we have that

$$w = T(v) = T\Big(\sum_{i=1}^{n} a_i v_i\Big) = \sum_{i=1}^{n} a_i T(v_i).$$

This tells us that $v \in \text{span}(T(\beta))$. Hence, $T(\beta)$ generates $R(T)$. ■

Example 2.1.8. Define the linear transformation $T : P_2(\mathbb{R}) \to M_{2\times 2}(\mathbb{R})$ by

$$T(f(x)) = \begin{pmatrix} f(1) - f(2) & 0 \\ 0 & f(0) \end{pmatrix}.$$

Note that

$$\beta = \{1, x, x^2\} \text{ is a basis for } P_2(\mathbb{R}) \text{ and } T(\beta) = \{T(1), T(x), T(x^2)\}.$$

Then observe that

$$\begin{aligned} R(T) &= \text{span}(T(\beta)) \\ &= \text{span}\left(\left\{ \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} -1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} -3 & 0 \\ 0 & 0 \end{pmatrix} \right\}\right). \end{aligned}$$

This tells us that $R(T)$ contains the basis found in the second equality above. Hence, we must have $\dim(R(T)) = 2$.

Example 2.1.9. Let $T : \mathbb{R}^3 \to \mathbb{R}^2$ be the linear transformation defined by

$$T(a_1, a_2, a_3) = (a_1 - a_2, 2a_3).$$

One can show that

$$N(T) = \{(a, a, 0)\} : a \in \mathbb{R} \text{ and } R(T) = \mathbb{R}^2.$$

Just like how the 'size' of a given subspace is denoted by its dimension, we can also determine the size of a null spaces and ranges. However, we will attach some special names associated with these sets.

Definition 2.1.6 (Nullity)**.** Let V and W be vector spaces, and let $T : V \to W$ be linear. If $N(T)$ is finite-dimensional, then we define **nullity** of T by $\text{nullity}(T)$ to be the dimension of $N(T)$.

Definition 2.1.7 (Rank)**.** Let V and W be vector spaces, and let $T : V \to W$. If $R(T)$ is finite-dimensional, then we denote the **rank** of T by $\text{rank}(T)$ to be the dimension of $R(T)$.

From these definitions, we can intuit the following relationships between Nullity and Rank of a linear transformation:

- The larger the nullity, the smaller the rank of a linear transformation.
- the larger the rank, the smaller the nullity.

This relationship between the two spaces is encompassed in the next theorem.

2.1.4 Dimension Theorem

Theorem 2.1.3 (Dimension Theorem). Let V and W be vector spaces, and let $T : V \to W$ be linear. If V is finite-dimensional, then

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

Proof. Suppose $\dim(V) = n$ and $\dim(N(T)) = k$, and $\{u_1, u_2, \dots, u_k\}$ is a basis for $N(T)$. By corollary to Theorem 1.11 β can be extended to be a basis for V. Denote this basis as

$$\{u_1, u_2, \dots, u_n\}.$$

We claim that that $S = \{T(u_{k+1}), T(u_{k+2}), \dots, T(u_n)\}$ is a basis for $R(T)$. First, we show that S generates $R(T)$. Using The Theorem 2.1 and the fact that $T(u_i) = 0$ for $1 \le i \le k$, we get that

$$\begin{aligned} R(T) &= \text{span}(\{T(u_1), T(u_2), \dots, T(u_n)\}) \\ &= \text{span}(\{T(u_{k+1}), T(u_{k+2}), \dots, T(u_n)\}) \\ &= \text{span}(T(S)). \end{aligned}$$

Hence, S generates $R(T)$.

Now, we want to show that S is linearly independent. Then choose scalars $\alpha_{k+1}, \alpha_{k+2}, \dots, \alpha_n$ such that

$$\sum_{i=k+1}^{n} \alpha_i T(u_i) = 0. \tag{1}$$

Since T is linear, we can re-write (1) into the following form

$$T\Big(\sum_{i=k+1}^{n} \alpha_i u_i\Big) = 0.$$

Hence, we find that

$$\sum_{i=k+1}^{n} \alpha_i u_i \in N(T). \tag{2}$$

Since $N(T)$ contains $\{u_1, u_2, \dots, u_k\}$ as a basis, we can express (2) as a linear combination of vectors in this set. Hence, we have

$$\sum_{i=k+1}^{n} \alpha_i u_i = \sum_{i=1}^{k} \beta_i u_i \tag{3}$$

for some scalars $\beta_1, \beta_2, \ldots, \beta_k$. Now, we have

$$\sum_{i=k+1}^{n} \alpha_i u_i - \sum_{i=1}^{k} \beta_i u_i = 0.$$

Since $\{u_1, u_2, \ldots, u_n\}$ is linearly independent, we find that both α_i's and β_i's are all zero. Hence, S is also linearly independent. Thus, S is a basis for $R(T)$ and that $\operatorname{rank}(T) = n-k$ and so we get our desired result

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

■

Applying the dimension theorem to example 9 allows us to conclude that $\operatorname{nullity}(T) + 2 = 3$ implies $\operatorname{nullity}(T) = 1$.

Before we move on, let us recall two key definitions needed for the next topic.

Definition 2.1.8 (Injective Functions). Let $f : A \to B$ and $x, y \in A$. We call f **one-to-one** if $f(x) = f(y)$ implies $x = y$ or, equivalently, if $x \neq y$ implies $f(x) \neq f(y)$ (this latter part is the contrapositive of the definition).

Definition 2.1.9 (Surjective Functions). Let $f : A \to B$. We call f **onto** if $f(A) = B$; that is, for any $y \in B$, there exists an $x \in A$ such that $f(x) = y$.

As we will see, these two definitions will give insights into the nullity and rank of linear transformations.

Theorem 2.1.4 (Injectivity $\Leftrightarrow$ Null Space is $\{0\}$). Let V and W be vector spaces, and let $T : V \to W$ be linear. Then T is **one-to-one** if and only if $N(T) = \{0\}$.

Proof. ($\Rightarrow$) Suppose T is injective. Let $x \in V$ be arbitrary. Then $T(x) = T(0_V)$ implies $x = 0_V$. This tells us that $N(T) = \{0\}$. ($\Leftarrow$) Conversely, let $N(T) = \{0\}$. Let $x, y \in V$ be arbitrary and assume $T(x) = T(y)$. By using linearity, we have

$$\begin{aligned} T(x) = T(y) &\Leftrightarrow T(x) - T(y) = 0_W \\ &\Leftrightarrow T(x-y) = 0_W \end{aligned}$$

Since $x - y \in N(T)$ and $N(T) = \{0\}$, we have $x - y = 0$ if and only if $x = y$. Hence, T is an injective transformation. ■

Referring back to example 9, we find that $\operatorname{nullity}(T) = 1$implies that example 9 is not injective.

Theorem 2.1.5 (Equal Finite Dimensions Between Vector Spaces). Let V and W be vector spaces of equal (finite) dimension, and let $T : V \to W$ be linear. Then the following are equivalent.

(a) T is injective.

(b) T is surjective.

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

Proof. Using the dimension theorem, we have

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

By theorem 2.5, We know that T is injective if and only if $N(T) = \{0\}$. By definition, we have $\text{nullity}(T) = 0$. By the dimension theorem, this is true if and only if

$$\text{rank}(T) = \dim(V) \Leftrightarrow \text{rank}(T) = \dim(W)$$

where V and W have equal dimensions. This is true if and only if $\dim(R(T)) = \dim(W)$. This is true if and only if $R(T) = W$ by theorem 1.11. By definition, we know that T is surjective. ■

Remark. If V is not finite-dimensional then we find that neither (a) nor (b) follows from each other.

Example 2.1.10. Let $T : P_2(\mathbb{R}) \to P_3(\mathbb{R})$ be the linear transformation defined by

$$T(f(x)) = 2f'(x) + \int_0^x 3f(t)\, dt.$$

Observe that

$$R(T) = \text{span}(T(1), T(x), T(x^2)) = \text{span}(\{3x, 2 + \tfrac{3}{2}x^2, 4x + x^3\}).$$

One can show that the set $\{3x, 2+\frac{3}{2}x^2, 4x+x^3\}$ is linearly independent and that $\text{rank}(T) = 3$. Since $\text{rank}(T) \neq \dim(P_3(\mathbb{R}))$, we have that T is not surjective. Since $\dim(P_2(\mathbb{R})) = 3$, we know that

$$\text{rank}(T) = \dim(P_2(\mathbb{R}))$$

by the Dimension Theorem. Thus, $\text{nullity}(T) = \{0\}$ and hence T is injective.

Example 2.1.11. Let $T : F^2 \to F^2$ be the linear transformation defined by

$$T(a_1, a_2) = (a_1 + a_2, a_1).$$

One can show that T is injective so that $N(T) = \{0\}$. The dimension theorem tells us that T must also be onto.

Example 2.1.12. Let $T : P_2(\mathbb{R}) \to \mathbb{R}^3$ be the linear transformation defined by

$$T(a_0 + a_1x + a_2x^2) = (a_0, a_1, a_2).$$

One can easily show that T is linear and one-to-one. We find that $S = \{2 - x + 3x^2, x + x^2, 1 - 2x^2\}$ is linearly independent in $P_2(\mathbb{R})$ since

$$T(S) = \{(2, -1, 3), (0, 1, 1), (1, 0, -2)\}$$

is linearly independent in $\mathbb{R}^3$.

Theorem 2.1.6 (Unique Linear Transformations). Let V and W be vector spaces over F, and suppose that $\{v_1, v_2, \ldots, v_n\}$ is a basis for V. For $w_1, w_2, \ldots, w_n \in W$, there exists exactly one linear transformation $T : V \to W$ such that $T(v_i) = w_i$ for all $1 \leq i \leq n$.

Proof. Let $x \in V$. Since $\{v_1, v_2, \ldots, v_n\}$ is a basis for V, we have

$$x = \sum_{i=1}^{n} \gamma_i v_i$$

for some scalars $\gamma_1, \gamma_2, \ldots, \gamma_n$. Define the map $T : V \to W$ by

$$T(x) = \sum_{i=1}^{n} \gamma_i w_i$$

or $w_i \in W$ and $x \in V$.

First, we show that T is linear. Let $x, y \in V$. Then

$$x = \sum_{i=1}^{n} \gamma_i v_i$$

and

$$y = \sum_{i=1}^{n} \delta_i v_i$$

for some scalars γ_i and δ_i for all $1 \leq i \leq n$. Then by definition of T, we get that

$$T(x) = \sum_{i=1}^{n} \gamma_i w_i$$

and

$$T(y) = \sum_{i=1}^{n} \delta_i w_i.$$

By definition of T, we get that

$$T(x+y) = \sum_{i=1}^{n} (\gamma_i + \delta_i) w_i = \sum_{i=1}^{n} \gamma_i w_i + \sum_{i=1}^{n} \delta_i w_i = T(x) + T(y).$$

Hence, the first property of linearity holds. For the second property, let $c \in F$ and observe that

$$T(cx) = \sum_{i=1}^{n} \gamma_i (c w_i) = c \sum_{i=1}^{n} \gamma_i w_i = cT(x).$$

Hence, the second property of linearity holds. We conclude that T is linear and that $T(u_i) = w_i$ for all $1 \leq i \leq n$.

Now suppose there exists another linear map $U : V \to W$ such that $U(v_i) = w_i$ for all $1 \leq i \leq n$ defined by.

$$U(x) = \sum_{i=1}^{n} \gamma_i U(v_i).$$

Then observe that

$$U(x) = \sum_{i=1}^{n} \gamma_i U(v_i) = \sum_{i=1}^{n} \gamma_i w_i = T(x).$$

Thus, $U = T$ for all $x \in V$ and so, we conclude that T is a unique linear map. ■

Corollary. Let V and W be vector spaces, and suppose that V has a finite basis $\{v_1, v_2, \dots, v_n\}$. If $U, T : V \to W$ are linear and $U(v_i) = T(v_i)$ for all $1 \le i \le n$, then $U = T$.

Example 2.1.13. Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be the linear transformation defined by

$$T(a_1, a_2) = (2a_2 - a_1, 3a_1)$$

and suppose that $U : \mathbb{R}^2 \to \mathbb{R}^2$ is linear. Let $\{(1,2), (1,1)\}$ be a basis for $\mathbb{R}^2$. If $U(1,2) = (3,3)$ and $U(1,1) = (1,3)$, then we find that $U(x_i) = T(x_i)$ $i = 1, 2$. Thus $U = T$ from the corollary to theorem 2.6.

2.2 The Matrix Representation of a Linear Transformation

In this section, we will develop a one-to-one correspondence between matrices and linear transformations so that we may study the properties of one utilizing the properties of another. Before we do this, we shall develop the concept of an *ordered basis* for a vector space V.

2.2.1 Ordered Bases

Definition 2.2.1 (Ordered Bases). Let V be a finite-dimensional vector space. An **ordered basis** for V is a basis for V endowed with a specific order; that is, an ordered basis for V is a finite sequence of linearly independent vectors in V that generates V.

Example 2.2.1. In F^3, let $\beta = \{e_1, e_2, e_3\}$ be an ordered basis for F^3. If we take another set, say, $\gamma = \{e_2, e_1, e_3\}$, be a basis for F^3, we will see that even though these two bases are equal in terms of the vectors within it, we still end up with different ORDERED bases.

- Note that e_i for all $1 \le i \le n$ are the **standard basis vectors** for F^n. The set $\{e_1, e_2, \dots, e_n\}$ in F^n is the **standard ordered basis** for F^n. Likewise, $\{1, x, \dots, x^n\}$ is the **standard ordered basis** for $P_n(F)$.
- We can now identify vectors in some finite-dimensional vector space of dimension n by using $n-$tuples.
- This is done through what is called *coordinate vectors*.

Definition 2.2.2 (Coordinate Vectors). Let $\beta = \{u_1, u_2, \dots, u_n\}$ be an ordered basis for a finite-dimensional vector space V. For $x \in V$, let $a_1, a_2, \dots, a_n$ be the unique scalars such that

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

We define the **coordinate vector of x relative to β**, denoted $[x]_\beta$, by

$$[x]_\beta = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}.$$

- In our definition of standard basis vectors, we see that $[u_i]_\beta = e_i$.
- It is quite easy to show that $x \to [x]_\beta$ provides us with a linear transformation from $T : V \to F^n$.

Example 2.2.2. Let $V = P_2(\mathbb{R})$ and let $\beta = \{1, x, x^2\}$ be the standard ordered basis for V. If $f(x) = 4 + 6x - 7x^2$, then

$$[f]_\beta = \begin{pmatrix} 4 \\ 6 \\ -7 \end{pmatrix}.$$

Suppose that V and W are finite-dimensional vector spaces with ordered bases $\beta = \{v_1, v_2, \dots, v_n\}$ and $\gamma = \{w_1, w_2, \dots, w_m\}$, respectively. Let $T : V \to W$ be linear. Then for each j, $1 \leq j \leq n$, there exists unique scalars (Theorem 1.5.1) $a_{ij} \in F$, $1 \leq i \leq m$ such that

$$T(v_j) = \sum_{i=1}^{m} a_{ij} w_i \quad \text{for} \quad 1 \leq j \leq n.$$

2.2.2 Matrix Representations

Definition 2.2.3 (Matrix Representation of T). Using the notation above, we call the $m \times n$ matrix A defined by $A_{ij} = a_{ij}$, the **matrix representation of T in the ordered bases β and γ** and write $A = [T]_\beta^\gamma$. If $V = W$ and $\beta = \gamma$, then we write $A = [T]_\beta$.

- The jth column of A is just $[T(v_j)]_\gamma$.
- If $U : V \to W$ is linear such that $[U]_\beta^\gamma = [T]_\beta^\gamma$, then $U = T$ by corollary to Theorem 2.6..

2.2.3 Examples of Computing Matrix Representations

Example 2.2.3. Let $T : \mathbb{R}^2 \to \mathbb{R}^3$ be the linear transformation defined by

$$T(a_1, a_2) = (a_1 + 3a_2, 0, 2a_1 - 4a_2)$$

Let β and γ be the standard ordered bases for $\mathbb{R}^2$ and $\mathbb{R}^3$, respectively. Now,

$$T(1, 0) = (1, 0, 2) = 1e_1 + 0e_2 + 2e_3$$

and

$$T(0, 1) = (3, 0, -4) = 3e_1 + 0e_2 - 4e_3.$$

Hence, we have

$$[T]_\beta^\gamma = \begin{pmatrix} 1 & 3 \\ 0 & 0 \\ 2 & -4 \end{pmatrix}.$$

Suppose we re-ordered our basis γ to be the basis $\gamma' = \{e_3, e_2, e_1\}$. Then we will see that

$$[T]_\beta^{\gamma'} = \begin{pmatrix} 2 & -4 \\ 0 & 0 \\ 1 & 3 \end{pmatrix}.$$

The example serves to demonstrate how different orderings of bases can lead to different matrix representations.

Example 2.2.4. Let $T : P_3(\mathbb{R}) \to P_2(\mathbb{R})$ be the linear transformation defined by $T(f(x)) = f'(x)$. Let β and γ be the standard ordered bases for $P_3(\mathbb{R})$ and $P_2(\mathbb{R})$, respectively. Since $\beta = \{1, x, x^2, x^3\}$ is a basis for $P_3(\mathbb{R})$ and $\gamma = \{1, x, x^2\}$ for $P_2(\mathbb{R})$, observe that

$$\begin{aligned} T(1) &= 0 \cdot 1 + 0 \cdot x + 0 \cdot x^2 \\ T(x) &= 1 \cdot 1 + 0 \cdot x + 0 \cdot x^2 \\ T(x^2) &= 0 \cdot 1 + 2 \cdot x + 0 \cdot x^2 \\ T(x^3) &= 0 \cdot 1 + 0 \cdot x + 2 \cdot x^2. \end{aligned}$$

So,

$$[T]_\beta^\gamma = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix}.$$

Observe that writing $T(x^j)$ is written as a linear combination of the vectors of γ, its coefficients give the entries of the jth column of $[T]_\beta^\gamma$.

Later on, we will prove a theorem about how associating matrices with linear transformations leads to preservation of addition and scalar multiplication.

2.2.4 Addition and Scalar Multiplication of Matrix Representations

Definition 2.2.4. Let $T, U : V \to W$ be arbitrary functions, where V and W are vector spaces over F, and let $a \in F$. We define $T + U : V \to W$ by $(T + U)(x) = T(x) + U(x)$ for all $x \in V$, and $aT : V \to W$ by $(aT)(x) = aT(x)$ for all $x \in V$.

These operations of addition and scalar multiplication lead to the preservation of linear transformations.

Theorem 2.2.1 (The Set of Linear Maps Over V is a Vector Space)**.** Let V and W be vector spaces over a field F, and let $T, U : V \to W$ be linear. Then we have

(a) For all $a \in F$, $aT + U$ is linear.

(b) Using the operations of addition and scalar multiplication in the preceding definition, the collection of all linear transformations from V to W is a vector space over F.

Proof. (a) To show linearity of $aT + U$, let $x, y \in V$ and $a \in F$. Then the linearity of T

and U implies that

$$\begin{aligned}(aT+U)(x+y) &= (aT)(x+y)+U(x+y)\\ &= (aT)(x)+(aT)(y)+U(x)+U(y)\\ &= (aT)(x)+U(x)+(aT)(y)+U(y)\\ &= (aT+U)(x)+(aT+U)(y).\end{aligned}$$

Let $c \in F$. Then

$$\begin{aligned}(c(aT+U))(x) &= (caT+cU)(x)\\ &= (caT)(x)+(cU)(x)\\ &= c((aT)(x)+cU(x)\\ &= c((aT)(x)+U(x))\\ &= c(aT+U)(x).\end{aligned}$$

Hence, we can see that T is linear.

(b) Let S denote the collection of all linear transformations from V to W. Note that the zero transformation T_0 is in S. The vector axioms follow from S being closed under addition and scalar multiplication; that is,

$$U+T \in S \text{ and } cT \in S.$$

Hence, S is a vector space over a field F.

■

Definition 2.2.5. Let V and W be vector spaces over F. We denote the vector space of all linear transformations from V into W by $\mathcal{L}(V,W)$. In the case that $V=W$, we write $\mathcal{L}(V)$ instead of $\mathcal{L}(V,W)$.

Theorem 2.2.2 (Addition and Scalar Mulitplication of Matrix Representations). Let V and W be finite-dimensional vector spaces with ordered bases β and γ, respectively, and let $T,U : V \to W$ be linear transformations. Then

(a) $[T+U]_\beta^\gamma = [T]_\beta^\gamma + [U]_\beta^\gamma$ and

(b) $[aT]_\beta^\gamma = a[T]_\beta^\gamma$ for all scalars a.

Proof. (a) Let $\beta = \{v_1, v_2, \dots, v_m\}$ and $\gamma = \{w_1, w_2, \dots, w_n\}$. There exists unique scalars a_{ij} and b_{ij} with $1 \le i \le m$ and $1 \le j \le n$ such that

$$T(v_j) = \sum_{i=1}^{m} a_{ij} w_i \quad \text{and} \quad U(v_j) = \sum_{i=1}^{m} b_{ij} w_i \quad \text{for } 1 \le j \le n.$$

Observe that

$$
\begin{aligned}
(T+U)(v_j) &= \sum_{i=1}^{m}(a_{ij}+b_{ij})w_i \\
&= \sum_{i=1}^{m} a_{ij}w_i + \sum_{i=1}^{m} b_{ij}w_i \\
&= T(v_j)+U(v_j).
\end{aligned}
$$

Hence, we can write that

$$([T+U]_\beta^\gamma)_{ij} = a_{ij}+b_{ij} = ([T]_\beta^\gamma + [U]_\beta^\gamma)_{ij}$$

(b) Using the same bases as above, the matrix representation of $[aT]_\beta^\gamma$ is written as

$$(aT)(v_j) = \sum_{i=1}^{n}(ca_{ij})w_i = c\sum_{i=1}^{n} a_{ij}w_i = cT(v_j) \quad \text{for} \quad 1 \le j \le n.$$

Hence, we can write

$$([aT]_\beta^\gamma)_{ij} = a([T]_\beta^\gamma)_{ij}.$$

■

Example 2.2.5. Let $T : \mathbb{R}^2 \to \mathbb{R}^3$ and $U : \mathbb{R}^2 \to \mathbb{R}^3$ be the linear transformations respectively defined by

$$T(a_1,a_2) = (a_1+3a_2, 2a_1-4a_2) \text{ and } U(a_1,a_2) = (a_1-a_2, 2a_1, 3a_1+2a_2).$$

Let β and γ be the standard ordered bases of $\mathbb{R}^2$ and $\mathbb{R}^3$, respectively. Computing $[T]_\beta^\gamma$, we get

$$[T]_\beta^\gamma = \begin{pmatrix} 1 & 3 \\ 0 & 0 \\ 2 & -4 \end{pmatrix},$$

and

$$[U]_\beta^\gamma = \begin{pmatrix} 1 & -1 \\ 2 & 0 \\ 3 & 2 \end{pmatrix}.$$

Since $T+U$ is linear, we can write

$$(T+U)(a_1,a_2) = (2a_1+2a_2, 2a_1, 5a_1-2a_2)$$

and so we can write

$$[T+U]_\beta^\gamma = \begin{pmatrix} 2 & 2 \\ 2 & 0 \\ 5 & -2 \end{pmatrix}$$

using Theorem 2.8.

2.3 Composition of Linear Transformations and Matrix Multiplication

2.3.1 Compositions

In the last section, we learned that

- Linear Transformations as well as their operations such as sums and scalar multiplications can represented in terms of matrices.
- The set of all linear transformations from V to W forms a vector space over some field F.

The main goal in this section is to describe how to represent the multiplication of two linear transformations and thus the multiplication of two matrices. The notation UT is used in replacement of $U \circ T$ for the composite of linear transformations of U and T.

Our first goal is to prove that the composite of linear transformations is linear.

Theorem 2.3.1 (Compositions are Linear). Let V, W, and Z be vector spaces over the same field F, and let $T : V \to W$ and $U : W \to Z$ be linear. Then $UT : V \to Z$ is linear.

Proof. Suppose $T : V \to W$ and $U : W \to Z$ are linear. Let $cx + y \in V$. Note that $T(cx + y) \in W$ and $U(T(cx + y)) \in Z$. So,

$$\begin{aligned}
(UT)(cx + y) &= U(T(cx + y)) && \text{(Definition of Composition)}\\
&= U(cT(x) + T(y)) && (T \text{ linear})\\
&= cU(T(x)) + U(T(y)) && (U \text{ linear})\\
&= c(UT)(x) + (UT)(y).
\end{aligned}$$

Hence, UT is linear. ■

The following are a list of properties of the composition of linear transformations.

Theorem 2.3.2 (Properties of Compositions in $\mathcal{L}(V)$). Let V be a vector space. Let $T, U_1, U_2 \in \mathcal{L}(V)$. Then

(a) $T(U_1 + U_2) = TU_1 + TU_2$ and $(U_1 + U_2)T = U_1T + U_2T$.

(b) $T(U_1U_2) = (TU_1)U_2$.

(c) $TI = IT = T$.

(d) $a(U_1U_2) = (aU_1)U_2 = U_1(aU_2)$ for all scalars a.

Proof. Let $T, U_1, U_2 \in \mathcal{V}$ with vector space V.

(a) Then for $x \in V$, we have

$$\begin{aligned}
T(U_1 + U_2)(x) &= T\Big((U_1 + U_2)(x)\Big) && \text{(Def of Composition)}\\
&= T\Big(U_1(x) + U_2(x)\Big) && (\mathcal{L}(V) \text{ V.S})\\
&= T\Big(U_1(x)\Big) + T\Big(U_2(x)\Big) && (T \text{ is linear})\\
&= (TU_1)(x) + (TU_2)(x) && \text{(Def of Composition)}
\end{aligned}$$

Hence, $T(U_1 + U_2) = TU_1 + TU_2$.

Let $x \in V$ again. Then

$$\begin{aligned} \Big((U_1 + U_2)T\Big)(x) &= (U_1 + U_2)(T(x)) && \text{(Def of Composition)} \\ &= U_1(T(x)) + U_2(T(x)) && (\mathcal{L}(V) \text{ V.S}) \\ &= (U_1T)(x) + (U_2T)(x). && \text{(Def of Composition)} \end{aligned}$$

Hence, $(U_1 + U_2)T = (U_1T) + U_2T$.

(b) Let $x \in V$. Then using the definition of composition, we have

$$\begin{aligned} T(U_1U_2)(x) &= T\Big((U_1U_2)(x)\Big) \\ &= T\Big(U_1(U_2(x))\Big) \\ &= (TU_1)(U_2(x)) \\ &= (TU_1)U_2(x). \end{aligned}$$

Hence, $T(U_1U_2) = TU_1$.

(c) Let $x \in V$. Using the definition of composition, we get

$$(TI)(x) = T(I(x)) = T(x) = I(T(x)) = (IT)(x).$$

Hence, $TI = IT = T$.

(d) Let $a \in F$ and $v \in V$. Then using the definition of composition and operations of $\mathcal{L}(V)$, we must have

$$\begin{aligned} a(U_1U_2)(x) &= aU_1(U_2(x)) \\ &= (aU_1)(U_2(x)) \\ &= (U_1a)(U_2(x)) \\ &= U_1(aU_2(x)) \\ &= U_1(aU_2)(x). \end{aligned}$$

Hence, we have $a(U_1U_2) = U_1(aU_2)$.

■

We can also prove a more general result when $T : V \to W$ where $\dim(V) \neq \dim(W)$.

2.3.2 Matrix Products

Definition 2.3.1 (Matrix Representation of the Composition). Let $T : V \to W$ and $U : W \to Z$ be linear transformations and let $A = [U]_\beta^\gamma$ and $B = [T]_\alpha^\beta$ where $\alpha = \{v_1, v_2, \ldots, v_n\}, \beta = \{w_1, w_2, \ldots, w_m\}$, and $\gamma = \{z_1, z_2, \ldots, z_p\}$ are ordered bases for V, W and Z, respectively. Define the product AB of two matrices so that $AB = [UT]_\alpha^\gamma$.

2.3.3 Summation Formula for Matrix Representation of Composition

For $1 \leq j \leq n$, we have

$$\begin{aligned}
(UT)(v_j) &= U(T(v_j)) && \text{(Definition of Composition)} \\
&= U\Big(\sum_{k=1}^{m} B_{kj} w_k\Big) && ([T]_\alpha^\beta \text{ for } 1 \leq j \leq n) \\
&= \sum_{k=1}^{m} B_{kj} U(w_k) && \text{(Linearity of } U\text{)} \\
&= \sum_{k=1}^{m} B_{kj} \Big(\sum_{i=1}^{p} A_{ik} z_i\Big) && ([U]_\beta^\gamma \text{ for } 1 \leq k \leq m) \\
&= \sum_{i=1}^{p} \Big(\sum_{k=1}^{m} A_{ik} B_{kj}\Big) z_i && \text{(Finite Sums are Interchangeable)} \\
&= \sum_{i=1}^{p} C_{ij} z_i
\end{aligned}$$

where

$$C_{ij} = \sum_{k=1}^{m} A_{ik} B_{kj}.$$

Definition 2.3.2 (Product of Two Matrices)**.** Let A be an $m \times n$ matrix and B be an $n \times p$ matrix. We define the **product** of A and B, denoted AB, to be the $m \times p$ matrix such that

$$(AB)_{ij} = \sum_{k=1}^{n} A_{ik} B_{kj} \quad \text{for} \quad 1 \leq i \leq m, \ 1 \leq j \leq p$$

- In order for the product AB to be defined, where A is an $m \times n$ matrix and B is a $n \times p$, the two inner dimensions must be equal (in this case n).
- Subsequently, the two outer dimensions (namely, m and p) determine the size of the resulting matrix.

Example 2.3.1. We have

$$\begin{pmatrix} 1 & 2 & 1 \\ 0 & 4 & -1 \end{pmatrix} \begin{pmatrix} 4 \\ 2 \\ 5 \end{pmatrix} = \begin{pmatrix} 1 \cdot 4 + 2 \cdot 2 + 1 \cdot 5 \\ 0 \cdot 4 + 4 \cdot 2 + (-1) \cdot 5 \end{pmatrix} = \begin{pmatrix} 13 \\ 3 \end{pmatrix}.$$

Notice how the number of columns of the first matrix matches the number of rows on the second.

Matrix multiplication is not commutative; that is, it is not always the case that $AB = BA$.If A and B have corresponding inner dimensions; that is, if A is an $m \times n$ matrix and B is an $n \times p$ matrix, then $(AB)^t = B^t A^t$since

$$(AB)^t_{ij} = (AB)_{ji} = \sum_{k=1}^{n} A_{jk} B_{ki}$$

and

$$(B^tA^t)_{ij} = \sum_{k=1}^{n} (B^t)_{ik} (A)^t{}_{kj} = \sum_{k=1}^{n} B_{ki} A_{jk}.$$

Theorem 2.3.3 (Compositions Represented by Product of Matrix Multiplication)**.** Let V, W, and Z be finite-dimensional vector spaces with ordered bases α, β, and γ respectively. Let $T : V \to W$ and $U : W \to Z$ be linear transformations. Then

$$[UT]_\alpha^\gamma = [U]_\beta^\gamma [T]_\alpha^\beta.$$

Proof. This fact follows immediately from our definition of matrix products. ■

Corollary. Let V be a finite-dimensional vector space with an ordered basis β. Let $T, U \in \mathcal{L}(V)$. Then $[UT]_\beta = [U]_\beta [T]_\beta$.

Example 2.3.2. Let $U : P_3(\mathbb{R}) \to P_2(\mathbb{R})$ and $T : P_2(\mathbb{R}) \to P_3(\mathbb{R})$ be the linear transformations respectively defined by

$$U(f(x)) = f'(x) \quad \text{and} \quad T(f(x)) = \int_0^x f(t)\, dt.$$

Let α and β be the standard ordered bases of $P_3(\mathbb{R})$ and $P_2(\mathbb{R})$, respectively. We claim that $UT = I$. To see why this is the case, observe that

$$[UT]_\beta = [U]_\alpha^\beta [T]_\beta^\alpha = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{3} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = [I]_\beta.$$

Definition 2.3.3 (Kronecker Delta)**.** We define the **Kronecker delta** δ_{ij} by $\delta_{ij} = 1$ if $i = j$ and $\delta_{ij} = 0$ if $i \neq j$.

Definition 2.3.4 (Identity Matrix)**.** The $n \times n$ **identity matrix** I_n is defined by $(I_n)_{ij} = \delta_{ij}$.

In other words, the identity matrix is made up of the Kronecker delta constants where entries along the diagonal are equal to 1 and 0 otherwise.

Theorem 2.3.4 (Properties of Matrices)**.** Let A be an $m \times n$ matrix, B and C be $n \times p$ matrices, and D and E be $q \times m$ matrices. Then

(a) $A(B + C) = AB + AC$ and $(D + E)A = DA + EA$.

(b) $a(AB) = (aA)B = A(aB)$ for any scalar a.

(c) $I_m A = A = AI_n$.

(d) If V is an $n-$dimensional vector space with an ordered basis β, then $[I_V]_\beta = I_n$.

Proof. Let A be an $m \times n$ matrix, B and C be $n \times p$ matrices, and D and E be $q \times m$

matrices. Then

(a) Let $1 \leq i \leq m$ and $1 \leq j \leq p$. By definition of the product of two matrices, we have

$$\begin{aligned}
(A(B+C))_{ij} &= \sum_{k=1}^{n} A_{ik}(B+C)_{kj} \\
&= \sum_{k=1}^{n} A_{ik}(B_{kj}+C_{kj}) \\
&= \sum_{k=1}^{n} A_{ik}B_{kj} + \sum_{k=1}^{n} A_{ik}C_{kj} \\
&= (AB)_{ij} + (AC)_{ij}.
\end{aligned}$$

Hence, $A(B+C) = AB + AC$.

Now, let $1 \leq i \leq q$ and $1 \leq j \leq n$. For the second formula, we can use the same definition to write

$$\begin{aligned}
((D+E)A)_{ij} &= \sum_{k=1}^{m} (D+E)_{ik}A_{kj} \\
&= \sum_{k=1}^{m} (D_{ik}+E_{ik})A_{kj} \\
&= \sum_{k=1}^{m} D_{ik}A_{kj} + \sum_{i=1}^{m} E_{ik}A_{kj} \\
&= (DA)_{ij} + (EA)_{ij}.
\end{aligned}$$

Hence, $(D+E)A = DA + EA$.

(b) Let $1 \leq i \leq m$ and $1 \leq j \leq p$. Let $a \in F$. Then using the definition of the product once again, we have

$$\begin{aligned}
a(AB)_{ij} &= a\sum_{k=1}^{n} A_{ik}B_{kj} \\
&= \sum_{k=1}^{n} a(A_{ik}B_{kj}) \\
&= \sum_{k=1}^{n} (aA_{ik})B_{kj} \\
&= \sum_{k=1}^{n} (aA)_{ik}B_{kj} \\
&= ((aA)B)_{ij}.
\end{aligned}$$

Then observe that

$$
\begin{aligned}
((aA)B)_{ij} &= \sum_{k=1}^{n} (aA_{ik})B_{kj} \\
&= \sum_{k=1}^{n} (A_{ik}a)B_{kj} \\
&= \sum_{k=1}^{n} A_{ik}(aB_{kj}) \\
&= \sum_{k=1}^{n} A_{ik}(aB)_{kj} \\
&= (A(aB))_{ij}.
\end{aligned}
$$

Hence, $(aA)B = A(aB)$. Thus, we conclude that

$$a(AB) = (aA)B = A(aB).$$

(c) Let $1 \leq i \leq m$ and $1 \leq j \leq n$. Since $\delta_{ik} = 1$ only when $i = k$ and 0 otherwise, we must have

$$(I_m A)_{ij} = \sum_{k=1}^{m} \delta_{ik} A_{kj} = A_{ij}.$$

Likewise,

$$(AI_n)_{ij} = \sum_{k=1}^{n} A_{ik}\delta_{kj} = A_{ij}$$

by the same reasoning. Hence, we have

$$I_m A = A = AI_n.$$

(d) Let V be an $n-$dimensional vector space with $\beta = \{v_1, v_2, \ldots, v_n\}$ as an ordered basis. Since $I_V(v_j) = v_j$ for all $1 \leq j \leq n$, we must have

$$v_j = I_V(v_j) = \sum_{i=1}^{n} a_{ij} v_i$$

which holds only if $a_{ij} = 1$ for all $i = j$ and 0 otherwise. But this means that $a_{ij} = \delta_{ij}$, so $[I_V]_\beta = I_n$.

■

Corollary. Let A be an $m \times n$ matrix, $B_1, B_2, \ldots, B_k$ be $n \times p$ matrices, $C_1, C_2, \ldots, C_k$ be $q \times m$ matrices, and $a_1, a_2, \ldots, a_k$ be scalars. Then

$$A\Big(\sum_{i=1}^{k} a_i B_i\Big) = \sum_{i=1}^{k} a_i AB_i$$

and

$$\Big(\sum_{i=1}^{k} a_i C_i\Big) A = \sum_{i=1}^{k} a_i C_i A.$$

Proof. Let A be an $m \times n$ matrix and $B_1, B_2, \ldots, B_k$ be $n \times p$ matrices. Let $1 \le \ell \le m$ and $1 \le s \le p$. Then

$$\begin{aligned}
\Big[A\Big(\sum_{i=1}^{k} a_i B_i\Big)\Big]_{\ell s} &= \sum_{\lambda=1}^{n} A_{\ell\lambda}\Big(\sum_{i=1}^{k} a_i B_i\Big)_{\lambda s} \\
&= \sum_{\lambda=1}^{n} A_{\ell\lambda}\Big(\sum_{i=1}^{k} (a_i B_i)_{\lambda s}\Big) \\
&= \sum_{\lambda=1}^{n} A_{\ell\lambda}\Big(\sum_{i=1}^{k} a_i (B_i)_{\lambda s}\Big) \\
&= \sum_{i=1}^{k} a_i\Big(\sum_{\lambda=1}^{n} A_{\ell\lambda}(B_i)_{\lambda s}\Big) && \text{(part (a) of Theorem 2.12)} \\
&= \sum_{i=1}^{k} a_i (AB_i)_{\ell s}.
\end{aligned}$$

Hence, we have

$$A\Big(\sum_{i=1}^{k} a_i B_i\Big) = \sum_{i=1}^{k} a_i A B_i.$$

To show the second formula, let $1 \le \ell \le q$ and $1 \le s \le n$. Then

$$\begin{aligned}
\Big[\Big(\sum_{i=1}^{k} a_i C_i\Big) A\Big]_{\ell s} &= \sum_{\lambda=1}^{n} \Big(\sum_{i=1}^{k} a_i C_i\Big)_{\ell\lambda} A_{\lambda s} \\
&= \sum_{\lambda=1}^{n} \Big(\sum_{i=1}^{k} (a_i C_i)_{\ell\lambda}\Big) A_{\lambda s} \\
&= \sum_{\lambda=1}^{n} \Big(\sum_{i=1}^{k} a_i (C_i)_{\ell\lambda}\Big) A_{\lambda s} \\
&= \sum_{i=1}^{k} a_i\Big(\sum_{i=1}^{k} (C_i)_{\ell\lambda} A_{\lambda s}\Big) && \text{(part (a) of Theorem 2.12)} \\
&= \sum_{i=1}^{k} a_i (C_i A)_{\ell s}.
\end{aligned}$$

Hence, we have

$$\Big(\sum_{i=1}^{k} a_i C_i\Big) A = \sum_{i=1}^{k} a_i C_i A.$$

■

Theorem 2.3.5 (Marix Products Represented by Column Vectors)**.** Let A be an $m \times n$ matrix and B be an $n \times p$ matrix. For each j $(1 \le j \le p)$ let u_j and v_j denote the jth columns of

AB and B respectively. Then

(a) $u_j = Av_j$

(b) $v_j = Be_j$, where e_j is the jth standard vector of F^p.

Proof. Let $1 \leq j \leq p$. Then we have

$$u_j = \begin{pmatrix} (AB)_{1j} \\ (AB)_{2j} \\ \vdots \\ (AB)_{mj} \end{pmatrix} = \begin{pmatrix} \sum_{k=1}^{n} A_{1k}B_{kj} \\ \sum_{k=1}^{n} A_{2k}B_{kj} \\ \vdots \\ \sum_{k=1}^{n} A_{mk}B_{kj} \end{pmatrix} = A \begin{pmatrix} B_{1j} \\ B_{2j} \\ \vdots \\ B_{nj} \end{pmatrix} = Av_j.$$

To show the other equation, we apply part (c) of Theorem 2.12, to write

$$v_j = \begin{pmatrix} B_{1j} \\ B_{2j} \\ \vdots \\ B_{nj} \end{pmatrix} = \begin{pmatrix} (BI_p)_{1j} \\ (BI_p)_{2j} \\ \vdots \\ (BI_p)_{nj} \end{pmatrix} = \begin{pmatrix} \sum_{k=1}^{n} B_{1k}\delta_{kj} \\ \sum_{k=1}^{n} B_{2k}\delta_{kj} \\ \vdots \\ \sum_{k=1}^{n} B_{nk}\delta_{kj} \end{pmatrix} = B \begin{pmatrix} \delta_{1j} \\ \delta_{2j} \\ \vdots \\ \delta_{pj} \end{pmatrix} = Be_j$$

where δ_{1j} are the Kronecker delta constants. ■

- This result gives us the ability to write columns of AB (that is, the jth column of AB) as a linear combination of the columns of A with the coefficients being entries of column j of B.

- In the case for rows, the ith row of AB can be written as a linear combination of the rows of B with coefficients being entries from ith row in A.

Theorem 2.3.6 (Linear Maps and Matrix Products)**.** Let V and W be finite-dimensional vector spaces having ordered bases β and γ, respectively, and let $T : V \to W$ be linear. Then, for each $u \in V$, we have

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

Proof. Let β and γ be ordered bases for V and W respectively. Fix $u \in V$, and define the linear transformations

$$f : F \to V \text{ by } f(a) = au \text{ and } g : F \to W \text{ by } g(a) = aT(u) \text{ for all } a \in F.$$

Let $\alpha = \{1\}$ be the standard ordered basis for F. Note that

$$T(u) = 1 \cdot T(u) = g(1) = g = Tf$$

since $f = f(1) = 1 \cdot u = u$. Using Theorem 2.11, we can write

$$\begin{aligned} [T(u)]_\gamma = [g(1)]_\gamma &= [g]_\alpha^\gamma \\ &= [Tf]_\alpha^\gamma \\ &= [T]_\beta^\gamma [f]_\alpha^\beta \\ &= [T]_\beta^\gamma [f(1)]_\beta \\ &= [T]_\beta^\gamma [u]_\beta. \end{aligned}$$

Hence, we have

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

■

Example 2.3.3. Let $T : P_3(\mathbb{R}) \to P_2(\mathbb{R})$ be the linear transformation defined by

$$T(f(x)) = f'(x),$$

and let β and γ be the standard ordered bases for $P_3(\mathbb{R})$ and $P_2(\mathbb{R})$, respectively. Let $A = [T]_\beta^\gamma$, then we have

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix}$$

from Example 4 of Section 2.2. Utilizing Theorem 2.3.6, we can verify that

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

where $p(x) \in P_3(\mathbb{R})$ is the polynomial defined by

$$p(x) = 2 - 4x + x^2 + 3x^3.$$

Hence, we have

$$[T(p(x))]_\gamma = [q(x)]_\gamma = \begin{pmatrix} -4 \\ 2 \\ 9 \end{pmatrix},$$

and

$$[T]_\beta^\gamma [p(x)]_\beta = A[p(x)]_\beta = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix} \begin{pmatrix} 2 \\ -4 \\ 1 \\ 3 \end{pmatrix} = \begin{pmatrix} -4 \\ 2 \\ 9 \end{pmatrix}.$$

Definition 2.3.5 (Left Multiplication Transformation)**.** Let A be an $m \times n$ matrix with entries from a field F. We denote by L_A the mapping $L_A : F^n \to F^m$ defined by $L_A(x) = Ax$ (the matrix product of A and x) for each column vector $x \in F^n$. We call L_A a **left-multiplication transformation**.

Example 2.3.4. Let

$$A = \begin{pmatrix} 1 & 2 & 1 \\ 0 & 1 & 2 \end{pmatrix}.$$

Then $A \in M_{2\times 3}(\mathbb{R})$ and $L_A : \mathbb{R}^3 \to \mathbb{R}^2$. If

$$x = \begin{pmatrix} 1 \\ 3 \\ -1 \end{pmatrix},$$

then

$$L_A(x) = Ax = \begin{pmatrix} 1 & 2 & 1 \\ 0 & 1 & 2 \end{pmatrix} \begin{pmatrix} 1 \\ 3 \\ -1 \end{pmatrix} = \begin{pmatrix} 6 \\ 1 \end{pmatrix}.$$

Remark. It is relatively straight forward to show that L_A is linear.

Theorem 2.3.7 (Properties of Left-Multiplication Transformations)**.** Let A be an $m \times n$ matrix with entries from F. Then the left-multiplication transformation $L_A : F^n \to F^m$ is linear. Furthermore, if B is any other $m \times n$ matrix (with entries from F) and β and γ are the standard ordered bases for F^n and F^m, respectively, then we have the following properties.

(a) $[L_A]_\beta^\gamma = A$.

(b) $L_A = L_B$ if and only if $A = B$.

(c) $L_{A+B} = L_A + L_B$ and $L_{aA} = aL_A$ for all $a \in F$.

(d) If $T : F^n \to F^m$ is linear, then there exists a unique $m \times n$ matrix C such that $T = L_C$. In fact, $C = [T]_\beta^\gamma$.

(e) If E is an $n \times p$ matrix, then $L_{AE} = L_A L_E$.

(f) If $m = n$, then $L_{I_n} = I_{F^n}$Transformation.

Proof. First, we show that L_A is linear. Let A be an $m \times n$ matrix and $L_A : F^n \to F^m$. Given an arbitrary $c \in F$ and $cx + y \in F^n$, we can write

$$\begin{aligned}
L_A(cx+y) &= A(cx+y) \\
&= A(cx) + A(y) && \text{(Part (a) of Property of Matrices)} \\
&= c(A(x)) + A(y) && \text{(Part (b) of Property of Matrices)} \\
&= cL_A(x) + L_A(y).
\end{aligned}$$

Hence, L_A is a linear transformation. Now, let's show properties (a)-(f)

(a) Let $1 \leq j \leq n$ and A be an $m \times n$ matrix. Observe that

$$([L_A]_\beta^\gamma)_j = L_A(e_j) = Ae_j = u_j$$

where u_j is the jth column of A. Hence, we have

$$[L_A]_\beta^\gamma = A.$$

(b) Suppose $L_A = L_B$. By part (a), we get that

$$A = [L_A]_\beta^\gamma = [L_B]_\beta^\gamma = B.$$

Conversely, suppose $A = B$. Let $x \in F^n$. Then performing left-multiplication, we have

$$L_A(x) = Ax = Bx = L_B(x).$$

Hence, we have $L_A = L_B$.

(c) Let $x \in F^n$. Using the Part (a) of Property of Matrices, we must have

$$L_{A+B}(x) = (A+B)(x) = A(x) + B(x) = L_A(x) + L_B(x).$$

Hence, $L_{A+B} = L_A + L_B$. Now, let $a \in F$. Using the same reasoning, we have

$$L_{aA}(x) = (aA)(x) = a(A(x)) = aL_A(x).$$

Hence, $L_{aA} = aL_A$.

(d) Let $c = [T]_\beta^\gamma$. Since $T : F^n \to F^m$, we see that

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

or

$$T(x) = C(x) = L_C(x)$$

where $x \in F^n$. Hence, $T = L_C$.

(e) Let $1 \leq j \leq p$. Applying Theorem 2.13 several times so we may note that $(AE)e_j$ is the jth column of AE and that the jth column of AE is also equal to $A(Ee_j)$; that is,

$$(AE)e_j = A(Ee_j).$$

Hence, we have

$$\begin{aligned} L_{AE}(e_j) = (AE)e_j &= A(E(e_j)) \\ &= L_A(E(e_j)) \\ &= L_A(L_E(e_j)) \\ &= (L_A L_E)(e_j). \end{aligned}$$

Hence, we have $L_{AE} = L_A L_E$.

(f) Let $1 \leq j \leq n$. Then

$$\begin{aligned} L_{I_n}(e_j) = I_n(e_j) &= e_j \\ &= I_{F^n}(e_j). && \text{(Part (d) of Theorem 2.3.4)} \end{aligned}$$

Hence, $L_{I_n} = I_{F^n}$.

■

Theorem 2.3.8 (Associativity of Left-Multiplication Transformations). Let A, B and C be matrices such that $A(BC)$ is defined. Then $(AB)C$ is also defined and $A(BC) = (AB)C$; that is, matrix multiplication is associative.

Proof. Let A be an $m \times n$ matrix, B an $n \times p$ matrix, and C $p \times \ell$ matrix so $(AB)C$ be defined with dimension $m \times \ell$. So, let $L_{(AB)C} : F^\ell \to F^m$ be the left multiplication transformation. Our goal is to show that

$$L_{(AB)C} = L_{A(BC)}.$$

Using the properties of compositions and left-multiplication transformations, we write

$$\begin{aligned} L_{(AB)C} &= (L_{AB})L_C && \text{(Def of Left-Multiplication)} \\ &= (L_A L_B)L_C && \text{(Prop of Left-Multiplication)} \\ &= L_A(L_B L_C) && \text{(Prop of Compositions)} \\ &= L_A(L_{BC}) && \text{(Prop of Left-Multiplication)} \\ &= L_{A(BC)}. && \text{(Def of Left-Multiplication)} \end{aligned}$$

Hence, we conclude that $L_{(AB)C} = L_{A(BC)}$. ■

2.4 Invertibility and Isomorphisms

2.4.1 Inverses of Linear Maps

- We will see in this section that inverses of linear maps are linear.
- Inverses of linear maps allow us to gain insight on properties of inverted matrices.
- Many of the results found in this section are applied to concept of *isomorphisms*.

Definition 2.4.1 (Invertibility of Linear Maps). Let V and W be vector spaces, and let $T : V \to W$ be linear. A function $U : W \to V$ is said to be an **inverse** of T if $TU = I_W$ and $UT = I_V$. If T has an inverse, then T is said to be **invertible**. As noted in Appendix B, if T is invertible, then the inverse of T is unique and is denoted by T^{-1}.

The following facts hold for invertible functions T and U.

1. $(TU)^{-1} = U^{-1}T^{-1}$.
2. $(T^{-1})^{-1} = T$; in particular, T^{-1} is invertible.

- We often refer to functions being invertible if and only if they are both injective and surjective. This allows us to restate Dimension Theorem when applied to two equal finite-dimensional vector spaces.

Theorem 2.4.1 (Dimension Theorem In Terms of Invertibility). Let $T : V \to W$ be a linear transformation, where V and W are finite-dimensional spaces of equal dimension. Then T is invertible if and only if $\text{rank}(T) = \dim(V)$.

Example 2.4.1. Let $T : P_1(\mathbb{R}) \to \mathbb{R}^2$ be the linear transformation defined by

$$T(a + bx) = (a, a + b).$$

One can verify directly that $T^{-1} : \mathbb{R}^2 \to P_1(\mathbb{R})$ is defined by

$$T^{-1}(c, d) = c + (d - c)x$$

and that T^{-1} is also linear.

Theorem 2.4.2 (Inverses Of Linear Maps Are Linear). Let V and W be vector spaces, and let $T : V \to W$ be linear and invertible. Then $T^{-1} : W \to V$ is linear.

Proof. Let $y_1, y_2 \in W$ and $c \in F$. Since T is surjective and injective, there exists unique vectors x_1 and x_2 such that $T(x_1) = y_1$ and $T(x_2) = y_2$. Thus, we have $x_1 = T^{-1}(y_1)$ and $x_2 = T^{-1}(y_2)$. Since T is linear, we must have that

$$\begin{aligned} T^{-1}(cy_1 + y_2) &= T^{-1}(cT(x_1) + T(x_2)) \\ &= T^{-1}(T(cx_1) + T(x_2)) \\ &= T^{-1}(T(cx_1 + x_2)) \\ &= T^{-1}T(cx_1 + x_2) \\ &= I_V(cx_1 + x_2) \\ &= cx_1 + x_2 \\ &= cT^{-1}(y_1) + T^{-1}(y_2). \end{aligned}$$

Hence, T^{-1} is a linear map. ■

- This tells us that the notions of invertibility, injectivity, and surjectivity in Theorem 2.5 are all equivalent.

2.4.2 Inverse of a Matrix

Definition 2.4.2 (Invertibility of Matrices). Let A be an $n \times n$ matrix. Then A is **invertible** if there exists an $n \times n$ matrix B such that $AB = BA = I$.

- When A is invertible, the matrix B and its product AB can be written in the following way:
$$AB = BA = I.$$
- If C was another such matrix such that the above is true (replace B with C), then we can write
$$C = CI = C(AB) = (CA)B = IB = B.$$
- The matrix B, in this case, is called the **inverse** of A and is denoted by A^{-1}.

Lemma 2.4.1. Let T be an invertible linear transformation from V to W. Then V is finite-dimensional if and only if W is finite-dimensional. In this case, $\dim(V) = \dim(W)$.

Proof. Suppose that V is finite-dimensional. Let $\beta = \{x_1, x_2, \dots, x_n\}$ be a basis for V. By Theorem 2.2, we get that $\operatorname{span}(T(\beta)) = R(T)$. Since T is invertible, we know that T is also surjective. Hence, $\operatorname{span}(T(\beta)) = R(T) = W$ and so W is finite-dimensional by Theorem 1.9. Conversely, suppose that W is finite-dimensional. So, let $\gamma = \{y_1, y_2, \dots, y_n\}$ be a basis for W. Since T is surjective, we know that for each y_i in γ that $y_i = T(x_i)$ for each i. Hence, we have that $T^{-1}(y_i) = x_i$ for all i. By Theorem 2.2 again, we know that $T^{-1}(\gamma)$ spans $R(T^{-1})$ where $R(T^{-1}) = V$ since T^{-1} is surjective. Hence, V is finite-dimensional.

Now, suppose that both V and W are finite-dimensional. Since T is injective and surjective, we must have that

$$\operatorname{nullity}(T) = 0 \text{ and } \operatorname{rank}(T) = \dim(R(T)) = \dim(W).$$

By the Dimension Theorem, we have that $\text{rank}(T) = \dim(V)$ which implies that

$$\dim(V) = \dim(W).$$

■

Theorem 2.4.3 (Invertibility of Linear Maps ⇔ Invertibility of Matrices). Let V and W be finite-dimensional vector spaces with ordered bases β and γ, respectively. Let $T : V \to W$ be linear. Then T is invertible if and only if $[T]_\beta^\gamma$ is invertible. Furthermore, $[T^{-1}]_\gamma^\beta = ([T]_\beta^\gamma)^{-1}$.

Proof. Let V and W be vector spaces with ordered bases β and γ respectively. Let T be linear. Suppose T is invertible. Then there exists a unique linear transformation $T^{-1} : W \to V$ by definition. Hence, observe that

$$[T^{-1}]_\gamma^\beta [T]_\beta^\gamma = [T^{-1}T]_\beta = [I_V]_\beta = I_n$$

and similarly

$$[T]_\beta^\gamma [T^{-1}]_\gamma^\beta = [TT^{-1}]_\beta = [I_V]_\beta = I_n.$$

Hence, $[T]_\beta^\gamma$ is invertible and that

$$[T^{-1}]_\gamma^\beta = ([T]_\beta^\gamma)^{-1}.$$

Conversely, suppose that $[T]_\beta^\gamma$ is invertible. Since V and W are finite-dimensional vector spaces where $\beta = \{v_1, v_2, \dots, v_n\}$ and $\gamma = \{w_1, w_2, \dots, w_n\}$ are ordered bases for V and W, respectively. Hence, we know by Theorem 2.6 there exists a unique linear transformation $U : W \to V$ such that

$$U(w_j) = \sum_{i=1}^{n} B_{ij} v_i$$

where $v_i \in \beta$. Since $[T]_\beta^\gamma = A$ is invertible, we know that

$$AB = BA = I_n.$$

We want to show that $UT = I_V$ and $TU = I_W$. Define the matrix representation $[U]_\gamma^\beta$. We need to show that $U = T^{-1}$. Hence, we have

$$[UT]_\beta = [U]_\gamma^\beta [T]_\beta^\gamma = BA = I_n = [I_V]_\beta$$

and

$$[TU]_\gamma = [T]_\beta^\gamma [U]_\gamma^\beta = AB = I_n = [I_W]_\gamma$$

by Theorem 2.11. Hence, we have $UT = I_V$ and $TU = I_W$. Thus, T is invertible. ■

Example 2.4.2. Let β and γ be the standard ordered bases of $P_1(\mathbb{R})$ and $\mathbb{R}^2$, respectively. For T as in Example 1, we have

$$[T]_\beta^\gamma = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \quad \text{and} \quad [T^{-1}]_\gamma^\beta = \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix}.$$

We can verify via matrix multiplication that both matrices are inverses of each other.

Corollary. Let V be a finite-dimensional vector space with an ordered basis β, and let $T : V \to V$ be linear. Then T is invertible if and only if $[T]_\beta$ is invertible. Furthermore, $[T^{-1}]_\beta = ([T]_\beta)^{-1}$.

Proof. Let $T : V \to V$ be linear and let $\beta = \{v_1, v_2, \dots, v_n\}$ be an ordered basis for V. Suppose T is invertible. Then there exists a unique linear transformation denoted by $T^{-1} : V \to V$ such that

$$TT^{-1} = T^{-1}T = I_V.$$

By using the Corollary to Theorem 2.11 and part (d) of Theorem 2.12, we must have that

$$[T]_\beta[T^{-1}]_\beta = [TT^{-1}]_\beta = [I_V]_\beta = I_n$$

and similarly,

$$[T^{-1}]_\beta[T]_\beta = [T^{-1}T]_\beta = [I_V]_\beta = I_n.$$

This tells us that $[T]_\beta$ is invertible and that

$$[T^{-1}]_\beta = ([T]_\beta)^{-1}.$$

For the backwards direction, the ordered basis β defined earlier implies that there exists a unique linear transformation $U : V \to V$ defined by

$$U(v_j) = \sum_{i=1}^{n} A_{ij}v_i \text{ for } 1 \le j \le n.$$

We need to show that $U = T^{-1}$. Using the fact that $[T]_\beta$ is invertible, we can write

$$[TU]_\beta = [T]_\beta[U]_\beta = I_n = [I_V]_\beta$$

and similarly

$$[UT]_\beta = [U]_\beta[T]_\beta = I_n = [I_V]_\beta.$$

But this tells us that $UT = TU = I_V$. So, $U = T^{-1}$ and that T is invertible. ■

Corollary. Let A be an $n \times n$ matrix. Then A is invertible if and only if L_A is invertible. Furthermore, $(L_A)^{-1} = L_{A^{-1}}$.

Proof. Let $L_A : F^n \to F^n$. Let $\beta = \{v_1, v_2, \dots, v_n\}$ be a basis for F^n. By part (a) of Theorem 2.15, we have that $[L_A]_\beta = A$. Since A is invertible, we know that L_A must also be invertible by Corollary to Theorem 2.18. Furthermore, we have that $L_A L_{A^{-1}} = I$ implies that

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

Conversely, L_A invertible implies that $[L_A]_\beta$ is invertible by Corollary to Theorem 2.18. By Theorem 2.15, we must have that $[L_A]_\beta = A$. But this means that A is invertible. ■

2.4.3 Isomorphisms

Definition 2.4.3 (Isomorphisms). Let V and W be vector spaces. We say that V is **isomorphic** to W if there exists a linear transformation $T : V \to W$ that is invertible. Such a linear transformation is called an **isomorphism** from V to W.

Theorem 2.4.4 (Isomorphisms $\Leftrightarrow$ Equal Finite-Dimensions). Let V and W be finite-dimensional vector spaces (over the same field). Then V is isomorphic to W if and only if $\dim(V) = \dim(W)$.

Proof. Suppose that V is isomorphic to W and $T : V \to W$ is an isomorphism from V to W. By the lemma preceding Theorem 2.18, we have that $\dim(V) = \dim(W)$.

Now, suppose that $\dim(V) = \dim(W)$, and let $\beta = \{v_1, v_2, \dots, v_n\}$ and $\gamma = \{w_1, w_2, \dots, w_n\}$. By Theorem 2.6, there exists a unique linear transformation $T : V \to W$ such that $T(v_i) = w_i$ for all i. Then by Theorem 2.2, we must have that

$$R(T) = \operatorname{span}(T(\beta)) = \operatorname{span}(\gamma) = W.$$

Hence, T is onto. By Theorem 2.5, we must also have that T is one-to-one. Hence, T is an isomorphism. ∎

Remark. Using the lemma to Theorem 2.18, we find that V and W are finite-dimensional or infinite-dimensional when they are isomorphic to each other.

Corollary. Let V be a vector space over F. Then V is isomorphic to F^n if and only if $\dim(V) = n$.

We are now in the position to show that the collection of all linear transformations between two vector spaces, say V and W with different dimensions m and n, can be identified with the appropriate vector space of $m \times n$ matrices.

Theorem 2.4.5 (Isomorphism Between Linear Transformations and Matrices). Let V and W be finite-dimensional vector spaces F of dimensions n and m, respectively, and let β and γ be ordered bases for V and W respectively. Then the function $\Phi : \mathcal{L}(V, W) \to M_{m\times n}(F)$, defined by

$$\Phi(T) = [T]_\beta^\gamma \text{ for } T \in \mathcal{L}(V, W)$$

is an isomorphism.

Proof. First, we show that Φ is linear. Using Theorem 2.8, we find that

$$\begin{aligned}\Phi(aT + U) &= [aT + U]_\beta^\gamma \\ &= a[T]_\beta^\gamma + [U]_\beta^\gamma \\ &= a\Phi(T) + \Phi(U).\end{aligned}$$

Hence, Φ is linear.

In order to show that Φ is isomorphic, it suffices to show that Φ is both injective and surjective. Clearly, Φ is injective since for every $T, U \in \mathcal{L}(V, W)$, we have $\Phi(T) = \Phi(U)$ implies $[T]_\beta^\gamma = [U]_\beta^\gamma \Rightarrow T = U$. For surjectivity, let $\beta = \{v_1, v_2, \dots, v_n\}$ and $\gamma = \{w_1, w_2, \dots, w_m\}$ be ordered bases for V and W respectively, and let A be a given $m \times n$ matrix. By Theorem 2.6, there exists a unique linear transformation $T : V \to W$ such that

$$T(v_j) = \sum_{i=1}^{m} A_{ij} w_i, \text{ for } 1 \le j \le n.$$

But then we have $[T]_\beta^\gamma = A$ or $\Phi(T) = A$. Hence, Φ is an isomorphism. ■

Corollary. Let V and W be finite-dimensional vector spaces of dimensions n and m, respectively. Then $\mathcal{L}(V, W)$ is finite-dimensional of dimension mn.

Proof. By Theorem 2.20, $\Phi : \mathcal{L}(V, W) \to M_{m\times n}(F)$ is an isomorphism. Then we have $\mathcal{L}(V, W)$ finite-dimensional with $\dim(\mathcal{L}(V, W)) = \dim(M_{m\times n}(F)) = mn$ by Theorem 2.19. ■

2.4.4 Standard Representation of Vector Spaces

Definition 2.4.4 (Standard Representation of V). Let β be an ordered basis for an $n-$dimensional vector space V over the field F. The **standard representation of V with respect to β** is the function $\phi_\beta : V \to F^n$ defined by $\phi_\beta(x) = [x]_\beta$ for each $x \in V$.

Example 2.4.3. Let $\beta = \{(1, 0), (0, 1)\}$ and $\gamma = \{(1, 2), (3, 4)\}$. It is easily observed that β and γ are ordered bases for $\mathbb{R}^2$. For $x = (1, -2)$, we have

$$\phi_\beta(x) = [x]_\beta = \begin{pmatrix} 1 \\ -2 \end{pmatrix} \text{ and } \phi_\gamma(x) = [x]_\gamma = \begin{pmatrix} -5 \\ 2 \end{pmatrix}.$$

In the last two sections, we showed that ϕ_β is a linear transformation.

Theorem 2.4.6. For any finite-dimensional vector space V with ordered basis β, ϕ_β is an isomorphism.

Proof. Let $\dim(V) = n$. Let $\beta = \{v_1, v_2, \ldots, v_n\}$ be an ordered basis for V. We can show that $\phi_\beta(y)$ is an isomorphism by showing that ϕ_β is a surjective and injective linear map. Suppose $\phi_\beta(x) = \phi_\beta$. Then by definition of ϕ_β, we must have that $[x]_\beta = [y]_\beta$ which further implies that $x = y$. Now, let $y \in V$. Since β is an ordered basis for V, we can find scalars $\delta_1, \delta_2, \ldots, \delta_n$ such that

$$y = \sum_{i=1}^{n} \delta_i v_i.$$

This implies that we have constructed a coordinate vector such that $[y]_\beta = \phi_\beta(y)$. Hence, ϕ_β is surjective. Thus, we find that ϕ_β is an isomorphism. ■

- Given finite-dimensional vector spaces V and W with corresponding bases β and γ respectively, we can use ϕ_β and ϕ_γ to map V to F^n and W to F^m respectively by the Theorem above.
- Let $T : V \to W$ be a linear transformation. We can map V into W using T and use ϕ_γ to map W onto F^m.
- Likewise, let $A = [T]_\beta^\gamma$. Then mapping V onto F^n using ϕ_β allows us to map the result onto F^m using $L_A : F^n \to F^m$ which produces the same vector in F^m as the transformation above.
- We can conclude, using Theorem 2.14, that

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

- Since V and W are associated with F^n and F^m respectively, we can now 'identify' T with L_A.

Example 2.4.4. Suppose we have the linear transformation $T : P_3(\mathbb{R}) \to P_2(\mathbb{R})$ defined by

$$T(f(x)) = f'(x).$$

Let β and γ be the standard ordered bases for $P_3(\mathbb{R})$ and $P_2(\mathbb{R})$, respectively. Let $\phi_\beta : P_3(\mathbb{R}) \to \mathbb{R}^4$ and $\phi_\gamma : P_2(\mathbb{R}) \to R^3$ be the corresponding standard representations of $P_3(\mathbb{R})$ and $P_2(\mathbb{R})$. If $A = [T]_\beta^\gamma$, then

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix}.$$

Consider the polynomial $p(x) = 2 + x - 3x^2 + 5x^3$. We will show that

$$L_A\phi_\beta(p(x)) = \phi_\gamma T(p(x)).$$

Using standard matrix operations, we get

$$L_A\phi_\beta(p(x)) = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix} \begin{pmatrix} 2 \\ 1 \\ -3 \\ 5 \end{pmatrix} = \begin{pmatrix} 1 \\ -6 \\ 15 \end{pmatrix}.$$

But notice that by the same operations, we have

$$\phi_\gamma T(p(x)) = \begin{pmatrix} 1 \\ -6 \\ 15 \end{pmatrix}.$$

So $L_A\phi_\beta(p(x)) = \phi_\gamma T(p(x))$.

2.5 The Change of Coordinate Matrix

2.5.1 The Change of Coordinate Matrix

Definition 2.5.1 (Change of Coordinate Matrix). Let β and β' be two ordered bases for a finite-dimensional vector space V. The matrix $Q = [I_V]_{\beta'}^{\beta}$ is the **Change of Coordinate Matrix**.

Theorem 2.5.1 (Properties of Change of Coordinates Matrix). Let β and β' be two ordered bases for a finite-dimensional vector space V, and let $Q = [I_V]_{\beta'}^{\beta}$. Then

(a) Q is invertible.

(b) For any $v \in V$, $[v]_\beta = Q[v]_{\beta'}$.

Proof. (a) Since I_V is invertible, Q is also invertible by Theorem 2.18.

(b) Let $v \in V$. Since $I_V(v) = v$ and that β and β' are ordered bases for V, we can write that

$$[v]_\beta = [I_V(v)]_\beta = [I_V]_{\beta'}^{\beta}[v]_{\beta'} = Q[v]_{\beta'}$$

by using Theorem 2.14 .

■

- The matrix $Q = [I_V]_{\beta'}^{\beta}$ is the **Change of Coordinate Matrix**.
- In the theorem above, part (b) states Q is what allows us to change from one set of coordinates to another.
- Letting $\beta = \{x_1, x_2, \dots, x_n\}$ and $\beta' = \{x_1', x_2', \dots, x_n'\}$, then

  $$x_j' = \sum_{i=1}^{n} Q_{ij} x_i \quad \text{for} \quad 1 \le j \le n$$

  where the jth column of Q is $[x_j']_\beta$.
- Since Q is invertible, we can say that Q^{-1} is its inverse and that Q^{-1} changes β coordinates back into β' coordinates.

Example 2.5.1. In $\mathbb{R}^2$, let $\beta = \{(1,1),(1,-1)\}$ and $\beta' = \{(2,4),(3,1)\}$. We can express each basis vector of β' as a linear combination of vectors from β; that is,

$$(2,4) = 3(1,1) - 1(1,-1) \text{ and } (3,1) = 2(1,1) + 1(1,-1),$$

the matrix that changes $\beta'-$coordinates into $\beta-$coordinates is

$$Q = \begin{pmatrix} 3 & 2 \\ -1 & 1 \end{pmatrix}.$$

Now, using part (b) of the above the theorem before this, we get that

$$[(2,4)]_\beta = Q[(2,4)]_{\beta'} = Q\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 3 & 2 \\ -1 & 1 \end{pmatrix}\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 3 \\ -1 \end{pmatrix}.$$

For the remainder of this section, we consider only linear transformations that map a vector space V into itself.

2.5.2 Properties of Change of Coordinate Matrices

Definition 2.5.2 (Linear Operators). Let V be a vector space. A linear transformation T that maps V into V is called a **Linear Operator** on V.

If V is finite-dimensional such that β and β' are ordered bases for V, then V can be represented in terms of the matrix representations $[T]_\beta$ and $[T]_{\beta'}$. Our goal for the rest of section is to find the relationship between these two matrices. Indeed, the next theorem answers this question using the help of coordinate matrices.

Theorem 2.5.2. Let T be a linear operator on a finite-dimensional vector space V, and let β and β' be ordered bases for V. Suppose that Q is the change of coordinate matrix that changes β'−coordinates into β−coordinates. Then

$$[T]_{\beta'} = Q^{-1}[T]_\beta Q.$$

Proof. Let I_V be the identity transformation on V and note that $Q = [I_V]_{\beta'}^{\beta}$ is invertible. Then $T = I_V T = T I_V$ and thus Theorem 2.11 implies that

$$\begin{aligned}
[T]_{\beta'} &= [TI_V]_{\beta'} \\
&= [T]_\beta [I_V]_{\beta'}^{\beta} \\
&= [T]_\beta Q \\
&= [I_V T]_\beta Q \\
&= [I_V]_{\beta}^{\beta'} [T]_\beta Q \\
&= Q^{-1}[T]_\beta Q.
\end{aligned}$$

Thus, we have $[T]_{\beta'} = Q^{-1}[T]_\beta Q$. ■

Example 2.5.2. Let T be the linear operator on $\mathbb{R}^2$ defined by

$$T\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} 3a-b \\ a+3b \end{pmatrix}$$

and let β and β' be the ordered bases as defined in Example 1. One can easily verify that

$$[T]_\beta = \begin{pmatrix} 3 & 1 \\ -1 & 3 \end{pmatrix}.$$

Using the change of coordinate matrix that changes β'−coordinates into β−coordinates found in Example 1

$$Q = \begin{pmatrix} 3 & 2 \\ -1 & 1 \end{pmatrix}$$

and

$$Q^{-1} = \frac{1}{5}\begin{pmatrix} 1 & -2 \\ 1 & 3 \end{pmatrix}.$$

By using Theorem 2.23, we have

$$[T]_{\beta'} = Q^{-1}[T]_\beta Q = \begin{pmatrix} 4 & 1 \\ -2 & 2 \end{pmatrix}.$$

To show that this is, indeed, the correct matrix, we can verify that the image under T of each vector of β' can be expressed as the linear combination of the vectors of β' with the entries of the corresponding column as its coefficients. That is, the image of the second vector in β' is

$$T\begin{pmatrix} 3 \\ 1 \end{pmatrix} = \begin{pmatrix} 8 \\ 6 \end{pmatrix} = 1\begin{pmatrix} 2 \\ 4 \end{pmatrix} + 2\begin{pmatrix} 3 \\ 1 \end{pmatrix}$$

and likewise the first vector of β' can be written as

$$T\begin{pmatrix} 2 \\ 4 \end{pmatrix} = \begin{pmatrix} 2 \\ 14 \end{pmatrix} = 4\begin{pmatrix} 2 \\ 4 \end{pmatrix} - 2\begin{pmatrix} 3 \\ 1 \end{pmatrix}.$$

Example 2.5.3. In this example, we wish to find an expression for $T(a,b)$ for any $(a,b) \in \mathbb{R}^2$. Note that T being linear implies that its values can be determined on a basis for $\mathbb{R}^2$. Observe from Figure 2.5 that $T(1,2) = (1,2)$ and that $T(-2,1) = -(-2,1) = (2,-1)$. If we let

$$\beta' = \left\{ \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} -2 \\ 1 \end{pmatrix} \right\},$$

then β' is an ordered basis for $\mathbb{R}^2$ and that

$$[T]_{\beta'} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Let β be the standard basis for $\mathbb{R}^2$, and let Q be the matrix that changes $\beta'-$coordinates into $\beta-$coordinates. Then

$$Q = \begin{pmatrix} 1 & -2 \\ 2 & 1 \end{pmatrix}$$

and $Q^{-1}[T]_\beta Q = [T]_{\beta'}$. We can solve this equation for $[T]_\beta$ to obtain that $[T]_\beta = Q[T]_{\beta'}Q^{-1}$. Since

$$Q^{-1} = \frac{1}{5}\begin{pmatrix} 1 & 2 \\ -2 & 1 \end{pmatrix},$$

we can easily verify that

$$[T]_\beta = \frac{1}{5}\begin{pmatrix} -3 & 4 \\ 4 & 3 \end{pmatrix}.$$

Since β is the standard ordered basis, it follows that T is left-multiplication by $[T]_\beta$. Hence, for any $(a,b) \in \mathbb{R}^2$, we have

$$T\begin{pmatrix} a \\ b \end{pmatrix} = \frac{1}{5}\begin{pmatrix} -3 & 4 \\ 4 & 3 \end{pmatrix}\begin{pmatrix} a \\ b \end{pmatrix} = \frac{1}{5}\begin{pmatrix} -3a+4b \\ 4a+3b \end{pmatrix}.$$

Corollary. Let $A \in M_{n\times n}(F)$, and let γ be an ordered basis for F^n. Then $[L_A]_\gamma = Q^{-1}AQ$ where Q is the $n \times n$ matrix whose jth column is the jth column is the jth vector of γ.

Proof. Observe that part (a) of Theorem 2.15 implies that $[L_A]_\gamma = A$. Note that $Q = [I_{F^n}]_\gamma$ and that $I_{F^n}L_A = L_A I_{F^n}$ so we write

$$\begin{aligned} Q[L_A]_\gamma &= [I_{F^n}]_\gamma [L_A]_\gamma \\ &= [I_{F^n}L_A]_\gamma \\ &= [L_A I_{F^n}]_\gamma \\ &= [L_A]_\gamma [I_{F^n}]_\gamma \\ &= AQ. \end{aligned}$$

Hence, we have

$$Q[L_A]_\gamma = AQ \tag{1}$$

. Since Q is invertible, we can do left-multiplication of Q^{-1} on both sides of (1) to get our desired result

$$[L_A]_\gamma = Q^{-1}AQ.$$

■

Example 2.5.4. Let

$$A = \begin{pmatrix} 2 & 1 & 0 \\ 2 & 1 & 3 \\ 0 & -1 & 0 \end{pmatrix}$$

and let

$$\gamma = \left\{ \begin{pmatrix} -1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right\},$$

which is an ordered basis for $\mathbb{R}^3$. Let Q be the 3×3 matrix whose jth column is the jth vector of γ. Then

$$Q = \begin{pmatrix} -1 & 2 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix} \quad \text{and} \quad Q^{-1} = \begin{pmatrix} -1 & 2 & -1 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{pmatrix}.$$

Using the Corollary to Theorem 2.23, we can see that

$$[L_A]_\gamma = Q^{-1}AQ = \begin{pmatrix} 0 & 2 & 8 \\ -1 & 4 & 6 \\ 0 & -1 & -1 \end{pmatrix}.$$

2.5.3 Similar Matrices

Definition 2.5.3 (Similar Matrices)**.** Let A and B be matrices in $M_{n\times n}(F)$. We say that B is **similar** to A if there exists an invertible matrix Q such that $B = Q^{-1}AQ$.

- Similarity, just as the notion of the isomorphism seen in the last section, is an equivalence relation.
- In the context of Theorem 2.23, if T is a linear operator on a finite-dimensional vector space V, and if β and β' are any ordered bases for V, then $[T]_{\beta'}$ is similar to $[T]_\beta$.
- We can also extend Theorem 2.23 to allow linear maps such as $T : V \to W$ with distinct vector spaces V and W.

Chapter 3

Elementary Matrix Operations and Systems of Linear Equations

3.1 Matrix Operations and Matrices

Definition 3.1.1 (Elementary Row (Column) Operations). Let A be an $m \times n$ matrix. Any one of the following three operations on the rows [columns] of A is called an **elementary row [column] operation**:

(1) Interchanging any two rows [columns] of A;

(2) Multiplying any row [column] of A by a nonzero scalar;

(3) Adding any scalar multiple of a row [column] of A to another row [column].

Any of these operations is called an **elementary operation**. Elementary operations of **type 1, type 2,** or **type 3** depending on whether they are obtained by (1), (2), (3).

If a matrix Q can be obtained from a matrix P via an elementary row operation, then P can be obtained from Q via an elementary row operation of the same type.

Definition 3.1.2 (Elementary Matrix). An $n \times n$ **elementary matrix** is a matrix obtained by performing an elementary operation on I_n. The elementary matrix is said to be of **type 1, 2,** or **3** according to whether the elementary operation performed on I_n is a type 1,2, or 3 operation, respectively.

The next theorem shows that performing an elementary row operation on a matrix is equivalent to multiplying the matrix by an elementary matrix.

Theorem 3.1.1. Let $A \in M_{m\times n}(F)$, and suppose that B is obtained from A by performing an elementary row [column] operation. Then there exists an $m \times m$ $[n \times n]$ elementary matrix E such that $B = EA[B = AE]$. In fact, E is obtained from I_m $[I_n]$ by performing the same elementary row [column] operation as that which was performed on A to obtain B. Conversely, if E is an elementary $m \times m$ $[n \times n]$ matrix, then $EA[AE]$ is the matrix obtained from A by performing the same elementary row [column] operation as that which produces E from $I_m[I_n]$.

Theorem 3.1.2 (Elementary Matrices are Invertible). Elementary matrices are invertible, and the inverse of an elementary matrix is an elementary matrix of the same type.

Proof. Let E be an elementary $n \times n$ matrix. Then E can be obtained by an elementary row operation on I_n. By reversing the steps used to transform I_n into E, we can transform E back into I_n. ■

3.2 The Rank Of A Matrix And Matrix Inverses

3.2.1 Defining the Rank of a Matrix

In this section, we will

- Define the *rank* of a matrix.
- Compute the rank of a matrix and a linear transformation.
- Create a procedure for computing the inverse of an invertible matrix.

Definition 3.2.1 (Rank). If $A \in M_{m\times n}(F)$, we define the **rank** of A, denoted $\text{rank}(A)$ to be the rank of the linear transformation $L_A : F^n \to F^m$.

- An important result derives from fact 3 and Corollary 2 to Theorem 2.18 is that *an $n \times n$ matrix is invertible if and only if its rank is n.*
- The rank of a matrix is equivalent to the rank of the linear transformation L_A.

Theorem 3.2.1. Let $T : V \to W$ be a linear transformation between finite-dimensional vector spaces, and let β and γ be ordered bases for V and W, respectively. Then $\text{rank}(T) = \text{rank}([T]_\beta^\gamma)$.

Proof. This is a restatement of Exercise 20 of Section 2.4. ■

- We can see now that finding the rank of a linear transformation can be done by finding the rank of a matrix.
- The next theorem will show us a method on how to do this.
- This allows us to do these operations in a rank-preserving way on matrices.

Theorem 3.2.2. Let A be an $m \times n$ matrix. If P and Q are invertible $m \times m$ and $n \times n$ matrices, respectively, then

(a) $\text{rank}(AQ) = \text{rank}(A)$,

(b) $\text{rank}(PA) = \text{rank}(A)$,

(c) $\text{rank}(PAQ) = \text{rank}(A)$.

Proof. (a) Let $A \in M_{m\times n}(F)$. Observe that

$$R(L_{AQ}) = R(L_A L_Q) = L_A L_Q(F^n) = L_A(L_Q(F^n)) = L_A(F^n) = R(L_A)$$

where L_Q is surjective. Hence, we have

$$\text{rank}(AQ) = \dim(R(L_{AQ})) = \dim(R(L_A)) = \text{rank}(A)$$

and so, $\text{rank}(AQ) = \text{rank}(A)$.

(b) Note that $L_A(F^n)$ is a subspace of F^m. So, we must have $\dim(L_A(F^n)) = \dim(L_P(L_A(F^n)))$ by Exercise 17. We can apply this result to write that

$$R(L_{PA}) = R(L_P L_A) = L_P L_A(F^n) = L_P(L_A(F^n)) = L_A(F^n) = R(L_A)$$

which implies that

$$\text{rank}(PA) = \dim(R(L_{PA})) = \dim(R(L_A)) = \text{rank}(A)$$

and so, we have $\text{rank}(PA) = \text{rank}(A)$.

(c) Applying part (a) and (b), we can see that

$$\text{rank}(PAQ) = \text{rank}(PA) = \text{rank}(A).$$

■

Corollary. Elementary row and column operations on a matrix are rank-preserving.

Proof. If B is obtained from a matrix A by an elementary row operation, then there exists an elementary matrix E such that $B = EA$ by Theorem 3.1. Since E is an invertible matrix, we use Theorem 3.4 to conclude that $\text{rank}(A) = \text{rank}(B)$. Similarly, if B is obtained from a matrix A by an elementary column operation, then there exists an elementary matrix E such that $B = AE$ by Theorem 3.1. Then we have $\text{rank}(B) = \text{rank}(AE) = \text{rank}(A)$ by Theorem 3.4 and we are done. ■

This result allows us to preserve the rank of matrices while using elementary operations on them. The next theorem allows us to find out the rank of a transformed matrix.

Theorem 3.2.3. The rank of any matrix equals the maximum number of its linearly independent columns; that is, the rank of a matrix is the dimension of the subspace generated by its columns.

Proof. Let $A \in M_{m\times n}(F)$. Then observe that

$$\text{rank}(A) = \text{rank}(L_A) = \dim(R(L_A)).$$

Let $\beta = \{e_1, e_2, \dots, e_n\}$ be the standard ordered basis of F^n. Since $L_A : F^n \to F^m$ is linear, we know by Theorem 2.2 that

$$R(L_A) = \text{span}(L_A(\beta)) = \text{span}(\{L_A(e_1), L_A(e_2), \dots, L_A(e_n)\}).$$

Observe that for each $1 \leq j \leq n$ that

$$L_A(e_j) = Ae_j = a_j$$

where a_j is the jth column of A. So, we have that

$$R(L_A) = \text{span}(\{a_1, a_2, \ldots, a_n\}).$$

■

Example 3.2.1. Suppose we have the matrix

$$A = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix}.$$

In this example, we can see quite easily that the first two rows are linearly independent and that the third column is a linear combination of the first two. Thus, we can see that the rank of A is just 2.

Remark. In much less trivial matrices, it is often much easier to perform appropriate elementary row and column operations to convert the matrix in question to a suitable one so that one can easily see the linearly independent columns of said matrix through the accumulation of more zero entries. Remember that the Corollary to Theorem 3.4 guarantees that our simplifying operations will preserve the rank of our matrix.

Example 3.2.2. Let

$$A = \begin{pmatrix} 1 & 2 & 1 \\ 1 & 0 & 3 \\ 1 & 1 & 2 \end{pmatrix}.$$

Observe that

$$A = \begin{pmatrix} 1 & 2 & 1 \\ 1 & 0 & 3 \\ 1 & 1 & 2 \end{pmatrix} \xrightarrow{R_2 - R_1 \text{ and } R_3 - R_1} \begin{pmatrix} 1 & 2 & 1 \\ 0 & -2 & 2 \\ 0 & -1 & 1 \end{pmatrix}.$$

Then we have

$$\begin{pmatrix} 1 & 2 & 1 \\ 0 & -2 & 2 \\ 0 & -1 & 1 \end{pmatrix} \xrightarrow{R_2 - 2R_1 \text{ and } R_3 - 2R_1} \begin{pmatrix} 1 & 0 & 0 \\ 0 & -2 & 2 \\ 0 & -1 & 1 \end{pmatrix}$$

for which we can see that the maximum number of linearly independent columns of A is 2 (again this fact is possible since our operations are rank preserving). Hence, the rank of A is 2.

These set of operations to convert a given matrix into a simpler form is used to prove the following fact.

Theorem 3.2.4 (Echelon Row (Column) Reduction)**.** Let A be an $m \times n$ matrix of rank r. Then $r \leq m, r \leq n$, and, y means of a finite number of elementary row and column operations, A can be transformed into the matrix

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

where $O_1, O_2, \ldots, O_3$ are zero matrices. Thus, $D_{ii} = 1$ for $i \leq r$ and $D_{ij} = 0$ otherwise.

Proof. If A is the zero matrix, then we have $r = 0$ by Exercise 3. Hence, we have $D = A$. Otherwise, suppose $A \neq O$ and $r = \text{rank}(A)$. So, $r > 0$. Let us proceed with induction on m, the number of rows of A. Let $m = 1$ be our base case. Using at most one type 1 column operation and at most one type 2 column operation, A can be transformed into a matrix with a 1 in the $1,1$ position. By means of at most $n-1$ type 3 column operations, A can be turned into the following matrix

$$\begin{pmatrix} 1 & 0 & \cdots & 0 \end{pmatrix}.$$

Thus, we can see that the first column is a (the only) linearly independent column in D. Since the rank of A is preserved, we get that $\text{rank}(D) = \text{rank}(A) = 1$ by Theorem 3.4. Hence, this ends our base case.

Assume that this theorem holds for any matrix with at most $m-1$ rows (that is, for some $m > 0$). Our goal is to show that the theorem holds for any matrix with m rows.

Suppose that A is any $m \times n$ matrix. Suppose $n = 1$. Then Theorem 3.6 can be shown in similarly as we have done wen we fixed $m = 1$. Otherwise, let $n > 1$. Since $A \neq O$, we must have that $A_{ij} \neq 0$ for some i, j.

Utilizing both at most one elementary row and at most elementary column operation (both of each are type 1 operations), we can move the nonzero entry to the $1,1$ position. Now, an additional type 2 operation can ensure a 1 in the $1,1$ position. This has an effect of creating the following matrix

$$B = \begin{pmatrix} I_r & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & B' & \\ 0 & & & \end{pmatrix}$$

where B' has a rank of one less than B. Since $\text{rank}(A) = \text{rank}(B)$, we mus have $\text{rank}(B') = r - 1$. Hence, $r - 1 \leq m - 1$ and $r - 1 \leq n - 1$ by the induction hypothesis. Thus, we have $r \leq m$ and $r \leq n$.

Using the induction hypothesis again, B' can be transformed into

$$D' = \begin{pmatrix} I_{r-1} & O_4 \\ O_5 & O_6 \end{pmatrix}$$

(an $(m-1) \times (n-1)$ matrix) via a finite number of elementary row and column operations. Note that O_4, O_5, and O_6 are zero matrices and that D' contains $r-1$ diagonal entries containing 1 and everywhere else containing 0. Thus, we have

$$D = \begin{pmatrix} I_r & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & D' & \\ 0 & & & \end{pmatrix}.$$

We can show through a repeated set of applications involving elementary row and column operations that D can be obtained from B (Exercise 12). Since A can be transformed into B and B can be transformed into D through a finite number of elementary operations, A can be transformed into D by a finite number of elementary operations.

Since D' contains ones in its first $r-1$ diagonal entries, we can see that D contains ones in its first r diagonal entires and zeros elsewhere. This establishes the theorem. ■

Corollary. Let A be an $m \times n$ matrix of $\text{rank}(A) = r$. Then there exist invertible matrices B and C of sizes $m \times m$ and $n \times n$, respectively, such that $D = BAC$, where

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

is the $m \times n$ matrix in which O_1, O_2, and O_3 are zero matrices.

Proof. By Theorem 3.6, A can be transformed into D via a finite number of elementary row and elementary column operations. We can use Theorem 3.1 to state that there exist elementary $m \times m$ matrices $E_1, E_2, \ldots, E_p$ and elementary $n \times n$ matrices $G_1, G_2, \ldots, G_q$ such that

$$D = E_p E_{p-1} \cdots E_2 E_1 A G_1 G_2 \cdots G_q$$

for each elementary row and column operation done on A to convert it into D. By Theorem 3.2, we can see that each E_j and G_j is invertible. So, let $B = E_p E_{p-1} \cdots E_2 E_1$ and let $C = G_1 G_2 \cdots G_q$. Using exercise 4 of section 2.4, we can see that B and C are invertible and thus $D = BAC$. ■

Corollary. Let A be an $m \times n$ matrix. Then

(a) $\text{rank}(A^t) = \text{rank}(A)$.

(b) The rank of any matrix equals the maximum number of its linearly independent rows; that is, the rank of a matrix is the dimension of the subspace generated by its rows.

(c) The rows and columns of any matrix generate subspaces of the same dimension, numerically equal to the rank of the matrix.

Proof. (a) Using Corollary 1, there exists invertible matrices B and C such that $D = BAC$, where

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

is an $m \times n$ matrix with the conditions satisfied in the Corollary. Taking the transpose of both sides of $D = BAC$, we get that

$$D^t = (BAC)^t = C^t A^t B^t.$$

Note that B^t and C^t are invertible by Exercise 5 of Section 2.4. So, Theorem 3.4 implies that

$$\text{rank}(A^t) = \text{rank}(C^t A^t B^t) = \text{rank}(D^t).$$

Since $\text{rank}(A) = r$, we know that D^t is an $n \times n$ matrix with the form found as in D in Corollary 1, and thus $\text{rank}(D^t)$ must also have the same rank by Theorem 3.5. Hence, we have

$$\text{rank}(A^t) = \text{rank}(D^t) = r = \text{rank}(A).$$

(b) Left as an exercise. *Similar process used to prove Theorem 3.5*

(c) Left as an exercise.

■

Corollary. Every invertible matrix is a product of elementary matrices.

Proof. Let $A \in M_{m\times n}(F)$ be an invertible matrix. Thus, $\text{rank}(A) = n$ and then by Corollary 1, there exists $m \times m$ and $n \times n$ invertible matrices B and C, respectively, such that $I_n = BAC$ where $B = E_p E_{p-1} \cdots E_2 E_1$ and $C = G_q G_{q-1} \cdots G_2 G_1$. Thus, we have that

$$A = B^{-1} I_n C^{-1} = B^{-1} C^{-1}$$

where $B^{-1} = E_1 E_2 \cdots E_{p-1} E_p$ and $C^{-1} = G_1 G_2 \cdots G_{q-1} G_q$. Note that that each E_j and G_j making up the product of B^{-1} and C^{-1} are elementary matrices by Theorem 3.2. ■

Theorem 3.2.5. Let $T : V \to W$ and $U : W \to Z$ be linear transformations on finite-dimensional vector spaces V, W, and Z, and let A and B matrices such that the product AB is defined. Then

(a) $\text{rank}(UT) \leq \text{rank}(U)$.

(b) $\text{rank}(UT) \leq \text{rank}(T)$.

(c) $\text{rank}(AB) \leq \text{rank}(A)$.

(d) $\text{rank}(AB) \leq \text{rank}(B)$.

Proof. (a) Note that $R(T) \subseteq W$. So, we must have

$$R(UT) = UT(V) = U(T(V)) = U(R(T)) \subseteq U(W) = R(U).$$

We can see by Theorem 1.11 that $R(UT) \subseteq R(U)$ implies that $\text{rank}(UT) \leq \text{rank}(U)$.

(c) Observe that

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

Thus, $\text{rank}(AB) \leq \text{rank}(A)$.

(d) Observe that $(AB)^t = B^t A^t$. So, we have

$$\text{rank}((AB)^t) = \text{rank}(B^t A^t) \leq \text{rank}(B^t).$$

By Corollary 3 to Theorem 3.6, we must have $\text{rank}(B^t) = \text{rank}(B)$ and that $\text{rank}((AB)^t) = \text{rank}(AB)$. Thus, we have that $\text{rank}(AB) \leq \text{rank}(B)$.

(b) Let α, β, and γ be ordered bases for V, W and Z, respectively. Thus, $[T]_\alpha^\beta$, $[U]_\beta^\gamma$ and

$[UT]_\alpha^\gamma$ are defined. Since these are just matrices, we can see that

$$\begin{aligned}
\operatorname{rank}(UT) &= \operatorname{rank}([UT]_\alpha^\gamma) \\
&= \operatorname{rank}([U]_\beta^\gamma [T]_\alpha^\beta) && \text{(Theorem 2.11)} \\
&\leq \operatorname{rank}([T]_\alpha^\beta) && \text{(part (d))} \\
&= \operatorname{rank}(T).
\end{aligned}$$

Hence, we have $\operatorname{rank}(UT) \leq \operatorname{rank}(T)$.

■

3.2.2 The Inverse of a Matrix

Definition 3.2.2 (Augmented Matrices). Let A and B be $m \times n$ and $m \times p$ matrices, respectively. By the **augmented matrix** $(A|B)$, we mean the $m \times (n+p)$ matrix $(A|B)$, that is, the matrix who first n columns are the columns of A, and whose last p columns are the columns of B.

We can see that performing a left-multiplication by an elementary matrix creates a matrix via an elementary row operation (See Theorem 3.1). This tells us the following important fact:

Proposition 3.2.1. If A is an invertible $n \times n$ matrix, then it is possible to transform the matrix $(A|I_n)$ into the matrix $(I_n|A^{-1})$ by means of a finite number of elementary row operations. Conversely, if for some $n \times n$ matrix B, the matrix $(A|I_n)$ can be transformed into the matrix $(I_n|B)$ by a finite number of elementary row operations, then $B = A^{-1}$.

Proof. Suppose A is an invertible $n \times n$ matrix. Consider the augmented matrix $C = (A|I_n)$. Using Exercise 15, we can see that

$$A^{-1}C = (A^{-1}A|A^{-1}I_n) = (I_n|A^{-1}).$$

Using Corollary 3 to Theorem 3.6, we can see that A^{-1} is the product of elementary matrices, say $A^{-1} = E_pE_{p-1}\cdots E_2E_1$. Thus, we have that

$$E_pE_{p-1}\cdots E_1(A|I_n) = A^{-1}C = (I_n|A^{-1})$$

and we are done. Conversely, suppose that A is invertible and that, for some $n \times n$ matrix B, the matrix $(A|I_n)$ can be transformed into the matrix $(I_n|B)$ by a finite number of elementary row operations. Let $E_1, E_2, \ldots, E_p$ be elementary matrices associated with elementary row operations found in Theorem 3.1. Then we see that

$$E_pE_{p-1}\ldots E_1(A|I_n) = (I_n|B).$$

Letting $M = E_pE_{p-1}\cdots E_1$, we have from the equation above that

$$(MA|M) = M(A|I_n) = (I_n|B).$$

Thus, we can see that $MA = I_n$ and $M = B$. Since A is invertible, this must mean that $M = A^{-1}$ and that $B = A^{-1}$. ■

- If A is $n \times n$ matrix not invertible, then it is not possible to transform $(A|I_n)$ into $(I_n|B)$.
- We also have that $\operatorname{rank}(A) < n$.

- An attempt to transform $(A|I_n)$ into $(B|I_n)$ will produce a row that contains zeros in the first n entries.

Chapter 4

Determinants

4.1 Determinants of Order 2

This section will go over the definition of the determinant of 2×2 matrices and its geometric significance in terms of area and orientation.

4.1.1 Basic Properties of Determinants

Definition 4.1.1 (Determinants). If

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

is a 2×2 matrix with entries from a field F, then we define the **determinant** of A, denoted $\det(A)$ or $|A|$, to be the scalar $ad - bc$.

Example 4.1.1. For the matrices

$$A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} 3 & 2 \\ 6 & 4 \end{pmatrix}$$

in $M_{2\times 2}(\mathbb{R})$, the determinant of these two matrices are

$$\det(A) = 1 \cdot 4 - 2 \cdot 3 = -2 \text{ and } \det(B) = 3 \cdot 4 - 2 \cdot 6 = 0$$

A common intuition that we have about determinants is that we think that

$$\det(A + B) = \det(A) + \det(B)$$

which is NOT generally true. Keep in mind that the mapping $\det : M_{2\times 2}(F) \to F$ is NOT a linear transformation, but it does posses an important linearity property which will presented below.

Theorem 4.1.1. The function $\det : M_{2\times 2}(F) \to F$ is a linear function of each row of a 2×2 matrix when the other row is held fixed. That is, if u, v and w are in F^2 and k is some scalar, then

$$\det \begin{pmatrix} u + kv \\ w \end{pmatrix} = \det \begin{pmatrix} u \\ w \end{pmatrix} + k \begin{pmatrix} u \\ w \end{pmatrix}$$

and

$$\det\begin{pmatrix} w \\ u+kv \end{pmatrix} = \det\begin{pmatrix} w \\ u \end{pmatrix} + k\det\begin{pmatrix} w \\ v \end{pmatrix}.$$

Proof. Let $u, v, w \in F^2$ be defined by $u = (a_1, a_2)$, $v = (b_1, b_2)$, and $w = (c_1, c_2)$. Using the definition of determinant, we have that

$$\begin{aligned}
\det\begin{pmatrix} u \\ w \end{pmatrix} + k\det\begin{pmatrix} v \\ w \end{pmatrix} &= \det\begin{pmatrix} a_1 & a_2 \\ c_1 & c_2 \end{pmatrix} + k\det\begin{pmatrix} b_1 & b_2 \\ c_1 & c_2 \end{pmatrix} \\
&= (a_1c_2 - a_2c_1) + k(b_1c_2 - b_2c_1) \\
&= (a_1 + kb_1)c_2 - (a_2 + kb_2)c_1 \\
&= \det\begin{pmatrix} a_1 + kb_1 & a_2 + kb_2 \\ c_1 & c_2 \end{pmatrix} \\
&= \det\begin{pmatrix} (a_1, a_2) + k(b_1, b_2) \\ (c_1, c_2) \end{pmatrix} \\
&= \det\begin{pmatrix} u + kv \\ w \end{pmatrix}.
\end{aligned}$$

The proof for the other equation is similar. ■

- In Example 1, notice how A is invertible and B is not.
- Also, A contains a nonzero determinant and B has a zero determinant.
- It turns our the invertibility is tied to nonzero determinants which will be presented in the next theorem.

Theorem 4.1.2. Let $A \in M_{2\times 2}(F)$. Then the determinant of A is nonzero if and only if A is invertible. Moreover, if A is invertible, then

$$A^{-1} = \frac{1}{\det(A)}\begin{pmatrix} A_{22} & -A_{12} \\ -A_{21} & A_{11} \end{pmatrix}.$$

Proof. Suppose $\det(A) \neq 0$, then we can define a matrix

$$M = \frac{1}{\det(A)}\begin{pmatrix} A_{22} & -A_{12} \\ -A_{21} & A_{11} \end{pmatrix}.$$

Thus, all we need to show is that $MA = AM = I$. Observe that

$$\begin{aligned}
AM &= \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}\frac{1}{\det(A)}\begin{pmatrix} A_{22} & -A_{12} \\ -A_{21} & A_{11} \end{pmatrix} \\
&= \frac{1}{\det(A)}\begin{pmatrix} A_{11}A_{22} - A_{12} & -A_{11}A_{12} + A_{12}A_{11} \\ A_{22}A_{21} - A_{21}A_{22} & A_{11}A_{22} - A_{12}A_{21} \end{pmatrix} \\
&= \frac{1}{\det(A)}\begin{pmatrix} \det(A) & 0 \\ 0 & \det(A) \end{pmatrix} \\
&= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \\
&= I.
\end{aligned}$$

Hence, $AM = I$ and a similar computation proves that $MA = I$. Hence, A is invertible and that $M = A^{-1}$.

Conversely, suppose that A is invertible. Using the remark found in page 152, we have that the rank of

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$$

must be 2. Hence, $A_{11} \neq 0$ or $A_{21} \neq 0$. If $A_{11} \neq 0$, then add $-A_{21}/A_{11}$ times row 1 of A to row 2 to obtain the matrix

$$\begin{pmatrix} A_{11} & A_{12} \\ 0 & A_{22} - \frac{A_{12}A_{21}}{A_{11}} \end{pmatrix}.$$

Since elementary row operations are rank-preserving by the corollary to Theorem 3.4, it follows that

$$A_{22} - \frac{A_{12}A_{21}}{A_{11}} \neq 0 \Rightarrow A_{11}A_{22} - A_{12}A_{21} \neq 0.$$

Now, suppose $A_{21} \neq 0$, we can see that $\det(A) \neq 0$ by adding $-A_{11}/A_{21}$ times row 2 of A to row 1 and applying a similar argument. Thus, in either case, $\det(A) \neq 0$.

■

In the upcoming sections, we will generalize the definition of the determinant to $n \times n$ matrices and show that the theorem above holds for these more general matrices.

4.1.2 The Area of a Parallelogram

This section will cover the geometric significance of the determinant of a 2×2 matrix. First, we start off by defining the notion of an angle in $\mathbb{R}^2$.

Definition 4.1.2 (Angle). Define **angle** between two vectors $u, v \in \mathbb{R}^2$ as the measure θ $(0 \leq \theta \leq \pi)$ that is formed by the vectors having the same magnitude and direction as the given vectors but emanating from the origin.

Definition 4.1.3 (Orientation). If $\beta = \{u, v\}$ is an ordered basis for $\mathbb{R}^2$, we define the **orientation** of β to be the real number

$$O\begin{pmatrix} u \\ v \end{pmatrix} = \frac{\det\begin{pmatrix} u \\ v \end{pmatrix}}{\left|\det\begin{pmatrix} u \\ v \end{pmatrix}\right|}.$$

4.2 Determinants of Order n

4.2.1 Cofactor Expansions

Before introducing the definition of the determinant for $n \times n$ matrices for $n \geq 3$, we will give a definition of deleting rows in a given matrix A.

Suppose we have

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} \in M_{3\times 3}(\mathbb{R}).$$

Denote the $(n-1) \times (n-1)$ matrix obtained from A by deleting row i and column j by $\tilde{A}_{ij}$.

Then we have

$$\tilde{A}_{11} = \begin{pmatrix} 5 & 6 \\ 8 & 9 \end{pmatrix}, \quad \tilde{A}_{13} = \begin{pmatrix} 4 & 5 \\ 7 & 8 \end{pmatrix}, \quad \text{and} \quad \tilde{A}_{32} = \begin{pmatrix} 1 & 3 \\ 4 & 6 \end{pmatrix}.$$

Definition 4.2.1. Let $A \in M_{n\times n}(F)$. If $n = 1$, so that $A = (A_{11})$, we define $\det(A) = A_{11}$. For $n \geq 2$, we define $\det(A)$ recursively as

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

The scalar $\det(A)$ is called the **determinant** of A and is also denoted by $|A|$. The scalar

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

is called the **cofactor** of the entry of A in row i, column j.

We can re-write the cofactor of the row i, column j entry of A as

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

so that we can express the formula given in the definition above as

$$\det(A) = A_{11}c_{11} + A_{12}c_{12} + \cdots + A_{1n}c_{1n}.$$

The formula above is denoted as the **cofactor expansion along the first row** of A. If $n = 2$, then this formula corresponds with the definition given in section 4.1.

Example 4.2.1. We will show that the $n \times n$ identity matrix is 1. We prove this assertion by using induction on n. This result holds for our base case of $n = 1$ by definition of the determinant. Assume that the determinant of the $(n-1) \times (n-1)$ identity matrix is 1 for some $n \geq 2$, and let I denote the $n \times n$. Using cofactor expansion along the first row of I, we can write

$$\begin{aligned} \det(I) &= (-1)^2(1) \cdot \det(\tilde{I}_{11} + (-1)^3(0) \cdot \det(\tilde{I}_{12})) + \cdots \\ &\quad + (-1)^{1+n}(0) \cdot \det(\tilde{I}_{1n}) \\ &= 1(1) + 0 + \cdots + 0 \\ &= 1 \end{aligned}$$

since $\tilde{I}_{11}$ is the $(n-1) \times (n-1)$ identity matrix. This shows that the determinant of the $n \times n$ matrix is 1, and so the determinant of any identity matrix is 1 by the principle of mathematical induction.

- As one can see, the operations involved when computing determinants using the recursive definition can be quite tedious.
- From here on, we will try to develop a more efficient method for computing them using the linearity property of determinants (not to mistaken linearity that we have been going over in this book).

4.2.2 The Determinant is a Linear Function

Theorem 4.2.1. The determinant of an $n \times n$ matrix is a linear function of each row when the remaining rows are held fixed. That is, for $1 \leq r \leq n$, we have

$$\det \begin{pmatrix} a_1 \\ \vdots \\ a_{r-1} \\ u + kv \\ a_{r+1} \\ \vdots \\ a_n \end{pmatrix} = \det \begin{pmatrix} a_1 \\ \vdots \\ a_{r-1} \\ u \\ a_{r+1} \\ \vdots \\ a_n \end{pmatrix} + k\det \begin{pmatrix} a_1 \\ \vdots \\ a_{r-1} \\ v \\ a_{r+1} \\ \vdots \\ a_n \end{pmatrix}$$

whenever k is a scalar and u, v, and each a_i are row vectors in F^n.

Proof. We proceed by mathematical induction on n. The result is immediate if $n = 1$. Assume that for some integer $n \geq 2$ the determinant of any $(n-1) \times (n-1)$ matrix is a linear function of each row when the remaining rows are held fixed. Let A be an $n \times n$ matrix with rows $a_1, a_2, \ldots, a_n$, respectively, and suppose that for some $r (1 \leq r \leq n)$, we have $a_r = u + kv$ for some $u, v \in F^n$ and some scalar k. Let $u = (b_1, b_2, \ldots, b_n)$ and $v = (c_1, c_2, \ldots, c_n)$, and let B and C be the matrices obtained from A by replacing row r of A by u and v, respectively.

For $r > 1$ and $1 \leq j \leq n$, the rows of the deleted matrices $\tilde{A}_{1j}, \tilde{B}_{1j}$, and $\tilde{C}_{1j}$ are the same except for row $r-1$. Moreover, row $r-1$ of $\tilde{A}_{1j}$ is

$$(b_1 + kc_1, \ldots, b_{j-1} + kc_{j-1}, b_{j+1} + kc_{j+1}, \ldots, b_n + kc_n),$$

which is the sum of row $r-1$ of $\tilde{B}_{1j}$ and k times row $r-1$ of $\tilde{C}_{1j}$. Since $\tilde{B}_{1j}$ and $\tilde{C}_{1j}$ are $(n-1) \times (n-1)$ matrices, we have

$$\det(\tilde{A}_{1j}) = \det(\tilde{B}_{1j}) + k\det(\tilde{C}_{1j})$$

by induction hypothesis. Using the recursive definition for the determinant and the fact that $A_{1j} = B_{1j} = C_{1j}$, we write

$$\begin{aligned}
\det(A) &= \sum_{j=1}^{n} (-1)^{1+j} A_{1j} \cdot \det(\tilde{A}_{1j}) \\
&= \sum_{j=1}^{n} (-1)^{1+j} A_{1j} \cdot \left[\det(\tilde{B}_{1j}) + k\det(\tilde{C}_{1j})\right] \\
&= \sum_{j=1}^{n} (-1)^{1+j} A_{1j} \cdot \det(\tilde{B}_{1j}) + k \sum_{j=1}^{n} (-1)^{1+j} A_{1j} \cdot \det(\tilde{C}_{1j}) \\
&= \sum_{j=1}^{n} (-1)^{1+j} B_{1j} \cdot \det(\tilde{B}_{1j}) + k \sum_{j=1}^{n} (-1)^{1+j} C_{1j} \cdot \det(\tilde{C}_{1j}) \\
&= \det(B) + k\det(C).
\end{aligned}$$

Hence, this shows that the theorem is true for $n \times n$ matrices, and so the theorem is true for all square matrices by mathematical induction. ■

Remark. For $r = 1$, notice that $a_1 = u + kv$ with $u, v \in F^n$ as defined as before where k

is a non-zero scalar. Furthermore, we have that the rows of the deleted matrices $\tilde{A}_{1j}, \tilde{B}_{1j}$, and $\tilde{C}_{1j}$ are all equal to each other except for $r = 1$. Using the recursive definition of the determinant of A and that $A_{1j} = b_j + kv_j$, we get that

$$\begin{aligned}
\det(A) &= \sum_{j=1}^{n}(-1)^{1+j}A_{1j} \cdot \det(\tilde{A}_{1j}) \\
&= \sum_{j=1}^{n}(-1)^{1+j}(b_j + kc_j)\det(\tilde{A}_{1j}) \\
&= \sum_{j=1}^{n}(-1)^{1+j}b_j \cdot \det(\tilde{A}_{1j}) + k\sum_{j=1}^{n}(-1)^{1+j}c_j \cdot \det(\tilde{A}_{1j}) \\
&= \sum_{j=1}^{n}(-1)^{1+j}B_{1j} \cdot \det(\tilde{B}_{1j}) + k\sum_{j=1}^{n}(-1)^{1+j}C_{1j} \cdot \det(\tilde{C}_{1j}) \\
&= \det(A) + k\det(C).
\end{aligned}$$

Corollary. If $A \in M_{n\times n}(F)$ has a row consisting entirely of zeros, then $\det(A) = 0$.

Proof. Left to the reader as an exercise (of course). ■

4.2.3 Cofactor Expansion Along Any Row

The next theorem will show that the determinant of any square matrix can be computed using cofactor expansion along any row. Before we proceed with the proof of this fact, we need to prove a preliminary result.

Lemma 4.2.1. Let $B \in M_{n\times n}(F)$, where $n \geq 2$. If row i of B equals e_k for some $k(1 \leq k \leq n)$, then $\det(B) = (-1)^{i+k}\det(\tilde{B}_{1k})$.

Proof. We proceed by performing mathematical induction on n. The lemma can be easily proved for $n = 2$. Assume that for some integer $n \geq 3$, the lemma is true for $(n-1)\times(n-1)$ matrices, and let B be an $n\times n$ matrix in which row i of B equals e_k for some k $(1 \leq k \leq n)$. The result follows immediately from the definition of the determinant if $i = 1$.

Suppose therefore that $1 < i \leq n$. For each $j \neq k$ $(1 \leq j \leq n)$, let C_{ij} denote the $(n-2)\times(n-2)$ matrix obtained from B by deleting rows 1 and i and columns j and k. For each j, row $i-1$ of $\tilde{B}_{1j}$ is the following vector in F^{n-1}:

$$\begin{cases} e_{k-1} & \text{if } j < k \\ 0 & \text{if } j = k \\ e_k & \text{if } j > k. \end{cases}$$

Using the induction hypothesis and the corollary to Theorem 4.3, we can write

$$\det(\tilde{B}_{1j}) = \begin{cases} (-1)^{(i-1)+(k-1)}\det(C_{ij}) & \text{if } j < k \\ 0 & \text{if } j = k \\ (-1)^{(i-1)+k}\det(C_{ij}) & \text{if } j > k. \end{cases}$$

Then we can see that

$$\begin{aligned}
\det(B) &= \sum_{j=1}^{n}(-1)^{1+j}B_{1j}\cdot\det(\tilde{B}_{1j}) \\
&= \sum_{j<k}(-1)^{1+j}B_{1j}\cdot\det(\tilde{B}_{1j}) + \sum_{j>k}(-1)^{1+j}B_{1j}\cdot\det(\tilde{B}_{1j}) \\
&= \sum_{j<k}(-1)^{1+j}B_{1j}\cdot\left[(-1)^{(i-1)+(k-1)}\det(C_{ij})\right] \\
&+ \sum_{j>k}(-1)^{1+j}B_{1j}\cdot\left[(-1)^{(i-1)+k}\det(C_{ij})\right] \\
&= (-1)^{i+k}\left[\sum_{j<k}(-1)^{1+j}B_{1j}\cdot\det(C_{ij})\right. \\
&\left. + \sum_{j>k}(-1)^{1+(j-1)}B_{1j}\cdot\det(C_{ij})\right].
\end{aligned}$$

Note that the expression on the inside of the last equality is just the cofactor expansion of $\tilde{B}_{ik}$ along the first row, it follows that

$$\det(B) = (-1)^{i+k}\det(\tilde{B}_{ik})$$

which ends our induction argument on why this lemma holds for all $n\times n$ matrices. ■

Now, this gives us the machinery to be able to use cofactor expansion along any row which is the basis for our next theorem.

Theorem 4.2.2. The determinant of a square matrix can be evaluated by cofactor expansion along any row. That is, if $A\in M_{n\times n}(F)$, then for any integer i $(1\le i\le n)$,

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

Proof. Let $i=1$. Using cofactor along the $i=1$ row gives us the determinant of A by definition. Fix $i>1$. Note that the ith row of A can be written as $\sum_{j=1}^{n}A_{ij}e_j$. For $1\le j\le n$, replace the ith row of A by e_j to obtain B_j. Using Theorem 4.3 and lemma, we have

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

■

Corollary. If $A\in M_{n\times n}(F)$ has two identical rows, then $\det(A)=0$.

Proof. We proceed the proof via mathematical induction on n. The proof for the case that $n=2$ is proved in Exercise 6 of Section 4.1. Now, assume that for some integer $n\ge 3$, it is true for $(n-1)\times(n-1)$ matrices, and let rows r and s of $A\in M_{n\times n}(F)$ be identical for $r\ne s$. Since $n\ge 3$, we can choose an integer i $(1\le i\le n)$ other than r and s. Using

Theorem 4.4, we can see that

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

Since $\tilde{A}_{ij}$ is an $(n-1)\times(n-1)$ matrix with two identical rows, our induction hypothesis tells us that for each $\det(\tilde{A}_{ij}) = 0$, and thus we have $\det(A) = 0$. This completes the induction proof for $n \times n$ matrices, and so the lemma must be true for all square matrices. ■

4.2.4 Effect of Elementary Operations on Determinant

Now, we will investigate what the behavior of the determinant will be if we were to perform type 1 operations on a matrix; that is, exchanging rows in a matrix.

Theorem 4.2.3. If $A \in M_{n\times n}(F)$ and B is a matrix obtained from A by interchanging any two rows of A, then $\det(B) = -\det(A)$.

Proof. Let the rows of $A \in M_{n\times n}(F)$ be $a_1, a_2, \dots, a_n$, and let B be the matrix obtained from A by interchanging rows r and s, where $r < s$. Thus, we have

$$A = \begin{pmatrix} a_1 \\ \vdots \\ a_r \\ \vdots \\ a_s \\ \vdots \\ a_n \end{pmatrix} \text{ and } B = \begin{pmatrix} a_1 \\ \vdots \\ a_s \\ \vdots \\ a_r \\ \vdots \\ a_n \end{pmatrix}.$$

Suppose we replaced rows r and s of A by $a_r + a_s$. Then by Corollary to Theorem 4.4 and Theorem 4.3, we have

$$\begin{aligned} 0 = \det \begin{pmatrix} a_1 \\ \vdots \\ a_r + a_s \\ \vdots \\ a_r + a_s \\ \vdots \\ a_n \end{pmatrix} &= \det \begin{pmatrix} a_1 \\ \vdots \\ a_r \\ \vdots \\ a_r + a_s \\ \vdots \\ a_n \end{pmatrix} + \det \begin{pmatrix} a_1 \\ \vdots \\ a_s \\ \vdots \\ a_r + a_s \\ \vdots \\ a_n \end{pmatrix} \\ &= \det \begin{pmatrix} a_1 \\ \vdots \\ a_r \\ \vdots \\ a_r \\ \vdots \\ a_n \end{pmatrix} + \det \begin{pmatrix} a_1 \\ \vdots \\ a_r \\ \vdots \\ a_s \\ \vdots \\ a_n \end{pmatrix} + \det \begin{pmatrix} a_1 \\ \vdots \\ a_s \\ \vdots \\ a_r \\ \vdots \\ a_n \end{pmatrix} + det \begin{pmatrix} a_1 \\ \vdots \\ a_s \\ \vdots \\ a_s \\ \vdots \\ a_n \end{pmatrix} \\ &= 0 + \det(A) + \det(B) + 0. \end{aligned}$$

Solving for $\det(B)$ now gives us the following result

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

■

Next, we show that performing a type 3 elementary row operation does not change the determinant of a matrix.

Theorem 4.2.4. Let $A \in M_{n\times n}(F)$, and let B be a matrix obtained by adding a multiple of one row of A. Then $\det(B) = \det(A)$.

Proof. Suppose that B is the $n \times n$ matrix obtained from A by adding k times row r to row s, where $r \neq s$. Let the rows of A be $a_1, a_2, \ldots, a_n$, and the rows of B $b_1, b_2, \ldots, b_n$. Then $a_i = b_i$ except for when $i = s$ and $b_s = a_s + ka_r$. Suppose we replace row s with a_r to create a matrix C. Then using Theorem 4.3 leads to

$$\begin{aligned}
\det(B) = \det\begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ a_s + ka_r \\ \vdots \\ b_n \end{pmatrix} &= \det\begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ a_s \\ \vdots \\ b_n \end{pmatrix} + k\det\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_r \\ \vdots \\ a_n \end{pmatrix} \\
&= \det(A) + k\det(C) \\
&= \det(A) + 0 \qquad \text{(Corollary to Theorem 4.4)}
\end{aligned}$$

which implies that $\det(B) = \det(A)$. ■

Corollary. If $A \in M_{n\times n}(F)$ has rank less than n, then $\det(A) = 0$.

Proof. If $\text{rank}(A) < n$, then the rows $a_1, a_2, \ldots, a_n$ of A are linearly dependent. Using Exercise 14 of Section 1.5, we know that some row of A, say row r, is a linear combination of the other rows. Hence, there must exist scalars c_i such that

$$a_r = c_1a_1 + \cdots + c_{r-1}a_{r-1} + c_{r+1}a_{r+1} + \cdots + c_na_n.$$

Let B be the matrix obtained from A by adding $-c_i$ times row i to row r for each $i \neq r$. Then row r of B consists entirely of zeros, and so $\det(B) = 0$. Now, Theorem 4.6 tells us that $\det(B) = \det(A)$ and that $\det(A) = 0$. ■

Now, we summarize the effect that specific elementary operations have on the determinant of a matrix.

(a) Interchanging any two rows of A implies $\det(B) = -\det(A)$.

(b) Scaling any row of A by some non-zero scalar implies that $\det(B) = k\det(A)$.

(c) Adding a multiple of one row of A to another row of A leaves the determinant of a matrix unchanged; that is, $\det(B) = \det(A)$.

Let

$$A = \begin{pmatrix} 1 & 3 & -3 \\ -3 & -5 & 2 \\ -4 & 4 & -6 \end{pmatrix}.$$

Adding 3 times row 1 of A to row 2 and 4 times row 1 to row 3, we obtain

$$M = \begin{pmatrix} 1 & 4 & -3 \\ 0 & 4 & -7 \\ 0 & 16 & -18 \end{pmatrix}.$$

Since we obtained M via type 3 elementary row operations, we see that $\det(A) = \det(M)$. We can use cofactor expansion along the first row to give us

$$\begin{aligned} \det(M) = (-1)^{1+1} \cdot \det(\tilde{M}_{11}) + (-1)^{1+2} \cdot \det(\tilde{M}_{12}) \\ + (-1)^{1+3}(-3) \cdot \det(\tilde{M}_{13}). \end{aligned}$$

Notice how $\tilde{M}_{12}$ and $\tilde{M}_{13}$ contain a column that consists entirely of zeros, and so we must have $\det(\tilde{M}_{12}) = \det(\tilde{M}_{13}) = 0$ by corollary to Theorem 4.6. Hence, we can write

$$\begin{aligned} \det(M) &= (-1)^{1+1}(1) \cdot \det(\tilde{M}_{11}) \\ &= (-1)^{1+1}(1) \cdot \det \begin{pmatrix} 4 & -7 \\ 16 & -18 \end{pmatrix} \\ &= 1[4(-18) - (-7)(16)] = 40. \end{aligned}$$

We can make this even easier by turning M into an upper triangular matrix so that computing the determinant is a matter of just taking the product of the diagonal entries of M. Adding -4 times row 2 of M to row 3 (type 3 elementary row operation), we obtain

$$P = \begin{pmatrix} 1 & 4 & -3 \\ 0 & 4 & -7 \\ 0 & 0 & 10 \end{pmatrix}.$$

Taking the cofactor expansion of the first row, we find that

$$\begin{aligned} \det(P) &= (-1)^{1+1}(1) \cdot \det(\tilde{P}_{11}) \\ &= (-1)^{1+1}(1) \cdot \det \begin{pmatrix} 4 & -7 \\ 0 & 10 \end{pmatrix} \\ &= 1 \cdot 4 \cdot 10 = 40. \end{aligned}$$

So, we have $\det(A) = \det(M) = \det(P) = 40$ and thus $\det(A) = 40$