Linear Algebra

Preliminary Lecture Notes

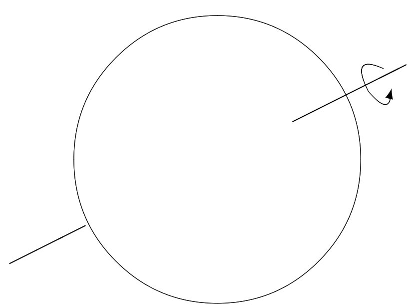
Adolfo J. Rumbos

Contents

Chapter 1

Motivation for the course

Imagine a ball whose center is at a fixed location in three–dimensional space, but is free to rotate about its center around any axis through the center. The center of the ball is not allowed to move away from its fixed location. Imagine that we perform several rotations about various axes, one after the other. We claim that there are two antipodal points on the surface of the ball which are exactly at the same locations they were at the beginning of the process. Furthermore, the combination of all the rotations that we perform has the same affect on the ball as that of a single rotation performed about the axis going through the fixed antipodal points. This result is know in the literature as Euler's Theorem



on the Axis of Rotation (see [PPR09]).

One of the goals of this course will be the proof if this fact. We will require all of the machinery of Linear Algebra to prove this result. The machinery of Linear Algebra consists of a new language we need to learn, new concepts we need to master and several theorems that we need to understand. The language and concepts of Linear Algebra will help us find convenient ways to represent rotations in space. Rotations, we will see, are special kinds of *linear* *transformations*, which are functions that map points in space to points in space and which satisfy some special properties.

We have studied functions in Calculus already. In Calculus I and II we dealt with real valued functions defined on intervals of the real line, $\mathbb{R}$; in Calculus III, we learned about functions which may be defined in regions of the plane, $\mathbb{R}^2$, or three dimensional space, $\mathbb{R}^3$, and which may be real valued or *vector* valued (also known as vector fields). In Linear Algebra we focus on a class of functions which are defined in all of the space (one–, two–, or three–dimensional space, or higher dimensional space) and can take on values in a one–dimensional or higher–dimensional space. The functions we will deal with have the property known as *linearity*. Loosely speaking, linearity means that the functions interact nicely with the algebraic structure that the spaces on which the functions act have: the structure of a *linear space* or a *vector space*.

The study of vector spaces will be one of the major topics of this course. We begin our discussion of vector spaces by introducing the example of Euclidean n–dimensional space. The main concepts of Linear Algebra will first be defined in the context of Euclidean space and then will be presented in more general context later on in the course.

Chapter 2

Euclidean n–dimensional Space

2.1 Definition of n–Dimensional Euclidean Space

Euclidean space of dimension n, denoted by $\mathbb{R}^n$ in this course, will consist of columns of real numbers of the form

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}.$$

These are called **column vectors**. In many textbooks elements of $\mathbb{R}^n$ are denoted by row–vectors; in the lectures and homework assignments, we will use column vectors to represents the elements in $\mathbb{R}^n$. Vectors in $\mathbb{R}^n$ can be used to locate points in n–dimensional space. They can also be used to indicate displacements in a certain direction and through certain distance.

Example 2.1.1. Consider two–dimensional space, $\mathbb{R}^2$. This can be represented by the familiar xy–plane pictured in Figure 2.1.1.

The vectors $\begin{pmatrix} 1 \\ 2 \end{pmatrix}$ and $\begin{pmatrix} 2 \\ -1 \end{pmatrix}$ are represented in the figure as arrows, or directed line segments, emanating from the origin of the xy–plane.

In the previous example, the vector $\begin{pmatrix} 1 \\ 2 \end{pmatrix}$ can be used to locate a point in the xy–plane with coordinates $(1, 2)$. However, it can also indicate a displacement from the origin to the point $(1, 2)$ through the straight line segment joining them.

Notation (Vector Notation and Conventions)**.** In the lectures and in these notes we will use the symbols u, v, w, etc. to denote vectors. In several linear algebra

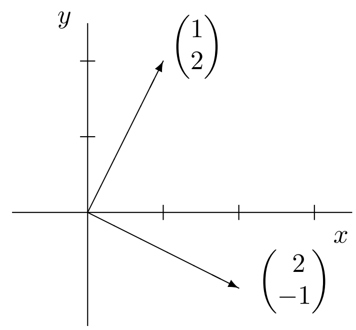


Figure 2.1.1: Two–dimensional Euclidean Space

texts, though, these symbols are usually written in boldface, $\mathbf{u}, \mathbf{v}, \mathbf{w}$, etc., or with an arrow on top of the letter, $\overrightarrow{u}, \overrightarrow{v}, \overrightarrow{w}$, etc. In these notes, real numbers will usually be denoted by the letters $a, b, c, d, t, s, x, y, x$, etc. and will be called **scalars** in order to distinguish them from vectors. I will also try to follow my own convention that if we are interested in locating a point in space, we will use the row vector made up of the Cartesian coordinates of the point; for instance, a point P in $\mathbb{R}^n$ will be indicated by $P(x_1, x_2, \ldots, x_n)$, where $x_1, x_2, \ldots, x_n$ are the coordinates of the point.

As mentioned earlier, vectors in $\mathbb{R}^n$ can also be used to indicate displacement along a straight line segment. For instance, the point $P(x_1, x_2, \ldots, x_n)$ is located by the vector

$$v = \overrightarrow{OP} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix},$$

where O denotes the origin, or **zero vector**, in n–dimensional Euclidean space. The arrow over the symbols OP emphasizes the "displacement" nature of the vector v.

Example 2.1.2. Denote the vectors $\begin{pmatrix} 1 \\ 2 \end{pmatrix}$ and $\begin{pmatrix} 2 \\ -1 \end{pmatrix}$ in Figure 2.1.1 by v_1 and v_2, respectively. Then, v_1 and v_2 locate the point $P_1(1, 2)$ and $P_2(2, -1)$, respectively. See Figure 2.1.2. Note, however, that the arrow representing the vector v_2 in Figure 2.1.2 does not have to be drawn with its starting point at the origin. It can be drawn anywhere as long as its length and direction are the same (see Figure 2.1.2). We will still call it the vector v_2. Only when the base of the arrow representing v_2 is located at the origin will it be locating the point $P_2(2, -1)$. In all other instances, the vector v_2 represents a displacement parallel to that from the origin to the point $(2, -1)$.

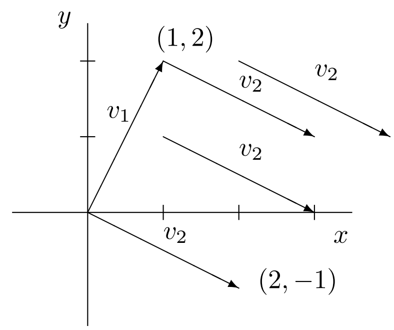


Figure 2.1.2: Dual Nature of Vectors in Euclidean Space

2.2 Algebraic Structure on Euclidean Space

What makes $\mathbb{R}^n$ into a vector space are the algebraic operations that we will define in this section. We begin with **vector addition**.

1. **Vector Addition**

   Given $v = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$ and $w = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$, the vector sum $v + w$ or v and w is

   $$v + w = \begin{pmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_n + y_n \end{pmatrix}$$

   Example 2.2.1. Let $v_1 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$ and $v_2 = \begin{pmatrix} 2 \\ -1 \end{pmatrix}$. Then, the vector sum of v_1 and v_2 is

   $$v_1 + v_2 = \begin{pmatrix} 1+2 \\ 2-1 \end{pmatrix} = \begin{pmatrix} 3 \\ 1 \end{pmatrix}.$$

   Figure 2.2.3 shows a geometric interpretation of the vector sum of the vectors v_1 and v_2 in the previous example. It is known as the *parallelogram rule:* the arrow representing the vector v_2 is drawn with its base at the tip of the arrow representing the vector v_1. The vector sum $v_1 + v_2$ is then represented by the arrow going from the base of the arrow representing v_1 to the tip of the translated arrow representing v_2.

   Notice that we could have obtained the same vector sum, $v_1 + v_2$, if, instead of translating the arrow representing v_2, we would have translated

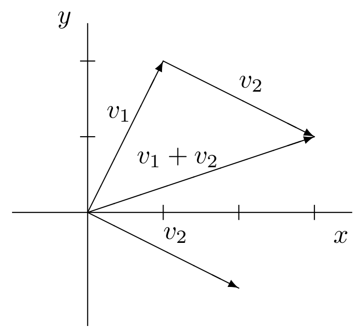


Figure 2.2.3: Parallelogram Rule

the arrow representing v_1 to the tip of the arrow representing v_2; see Figure 2.2.4

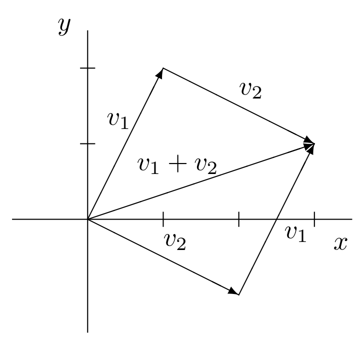


Figure 2.2.4: Commutative Property for Vector Addition

The picture in Figure 2.2.4 illustrates the fact that

$$v_1 + v_2 = v_2 + v_1.$$

This is known as the **commutative property** of vector addition, which can be derived algebraically from the definition and the fact that addition of real numbers is commutative: for any vectors $v = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$ and $w =$

$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$ in $\mathbb{R}^n$,

$$w + v = \begin{pmatrix} y_1 + x_1 \\ y_2 + x_2 \\ \vdots \\ y_n + x_n \end{pmatrix} = \begin{pmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_n + y_n \end{pmatrix} = v + w.$$

Properties of Vector Addition

Let u, v, w denote vectors in $\mathbb{R}^n$. Then,

(a) **Commutativity of Vector Addition**

$$v + w = w + v$$

(b) **Associativity of Vector Addition**

$$(u + v) + w = u + (v + w)$$

Like commutativity, this property follows from the definition and the fact that addition of real numbers is associative:

Write $v = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$, $\quad w = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$ and $u = \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{pmatrix}$. Then,

$$(u + w) + v = \begin{pmatrix} z_1 + x_1 \\ z_2 + x_2 \\ \vdots \\ z_n + x_n \end{pmatrix} + \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} (z_1 + x_1) + y_1 \\ (z_2 + x_2) + y_2 \\ \vdots \\ (z_n + x_n) + y_n \end{pmatrix}.$$

Thus, since $(z_i + x_i) + y_i = x_i + (x_i + y_i)$, for each $i = 1, 2, \ldots, n$, by associativity of addition of real numbers, it follows that

$$(u+w)+v = \begin{pmatrix} (z_1 + x_1) + y_1 \\ (z_2 + x_2) + y_2 \\ \vdots \\ (z_n + x_n) + y_n \end{pmatrix} = \begin{pmatrix} z_1 + (x_1 + y_1) \\ z_2 + (x_2 + y_2) \\ \vdots \\ z_n + (x_n + y_n) \end{pmatrix} = u+(v+w).$$

(c) **Existence of an Additive Identity**

The vector $\mathbf{0} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$ in $\mathbb{R}^n$ has the property that

$$v + \mathbf{0} = \mathbf{0} + v = v \quad \text{for all } v \text{ in } \mathbb{R}^n.$$

This follows from the fact that $x + 0 = x$ for all real numbers x.

(d) **Existence of an Additive Inverse**
Given $v = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$ in $\mathbb{R}^n$, the vector w defined by $v = \begin{pmatrix} -x_1 \\ -x_2 \\ \vdots \\ -x_n \end{pmatrix}$ has the property that

$$v + w = \mathbf{0}.$$

The vector w is called an additive inverse of v.

2. **Scalar Multiplication**

Given a real number t, also called a *scalar*, and a vector $v = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$, the scaling of v by t, denoted by tv, is given by

$$tv = \begin{pmatrix} tx_1 \\ tx_2 \\ \vdots \\ tx_n \end{pmatrix}$$

Example 2.2.2. Given the vector $v_1 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$ in $\mathbb{R}^2$, the scalar products $\left(-\frac{1}{2}\right) v_1$ and $\frac{3}{2}v_1$ are given by

$$\left(-\frac{1}{2}\right) v_1 = \begin{pmatrix} -1/2 \\ -1 \end{pmatrix} \text{ and } \frac{3}{2}v_1 = \begin{pmatrix} 3/2 \\ 3 \end{pmatrix},$$

respectively. The arrows representing these vectors are shown in Figure 2.2.5. Observe that the arrows representing the scalar products of v_1 lie on the same line as the arrow representing v_1.

Properties of Scalar Multiplication

(a) **Associativity of Scalar Multiplication**
Given scalars t and s and a vector v in $\mathbb{R}^n$,

$$t(sv) = (ts)v.$$

This property follows from the definition of scalar multiplication and the fact that $s(tx) = (st)x$ for all real numbers x; that is, associativity of multiplication of real numbers.

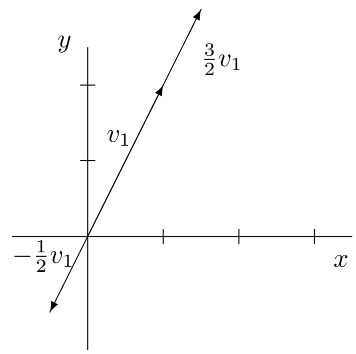


Figure 2.2.5: Scalar Multiplication

(b) **Identity in Scalar Multiplication**
The scalar 1 has the property that

$$1\ v = v \quad \text{for all } v \in \mathbb{R}^n.$$

3. **Distributive Properties**
Given vectors v and w in $\mathbb{R}^n$, and scalars t and s,

(a) $t(v + w) = tv + tw$
(b) $(t + s)v = tv + sv$.

These properties follow from the distributive properties for addition and multiplication in the set of real numbers; namely

$$t(x + y) = tx + ty \quad \text{for all} \quad t, x, y \in \mathbb{R},$$

and

$$(t + s)x = tx + sx \quad \text{for all} \quad t, s, x \in \mathbb{R},$$

respectively.

2.3 Linear Combinations and Spans

Given a vector v in $\mathbb{R}^n$, the set of all scalar multiples of v is called the **span** of the set $\{v\}$. We denote the span of $\{v\}$ by span($\{v\}$). In symbols, we write

$$\text{span}(\{v\}) = \{tv \mid t \in \mathbb{R}\}.$$

Geometrically, if v is not the zero vector in $\mathbb{R}^n$, span$\{v\}$ is the line through the origin on $\mathbb{R}^n$ in the direction of the vector v.

Example 2.3.1 (In $\mathbb{R}^3$)**.** Let $v = \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}$. Then,

$$\text{span}\{v\} = \left\{ t \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} \;\Big|\; t \in \mathbb{R} \right\}$$

We can also write this set as

$$\text{span}\{v\} = \left\{ \begin{pmatrix} x \\ y \\ z \end{pmatrix} \in \mathbb{R}^n \;\Big|\; \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} t \\ 2t \\ t \end{pmatrix}, \; t \in \mathbb{R} \right\}$$

Figure 2.3.6 shows a sketch of the line in $\mathbb{R}^3$ representing span$\{v\}$.

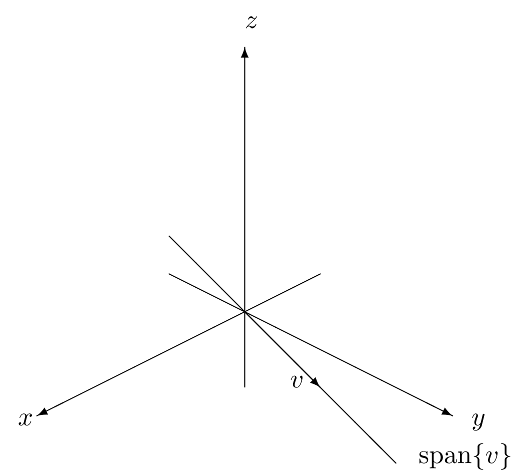


Figure 2.3.6: Line in $\mathbb{R}^3$

Note that $\begin{pmatrix} x \\ y \\ z \end{pmatrix}$ is a vector on the line, span$\{v\}$, if and only if

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} t \\ 2t \\ t \end{pmatrix}$$

for some scalar t. In other words, $\begin{pmatrix} x \\ y \\ z \end{pmatrix}$ is on the line if and only if the coordinates x, y and z satisfy the equations

$$\begin{cases} x = t \\ y = 2t \\ z = t. \end{cases}$$

These are known as the **parametric equations** of the line and t is called a **parameter**.

Definition 2.3.2 (Linear Combinations)**.** Given vectors $v_1, v_2, \ldots, v_k$ in $\mathbb{R}^n$, the expression

$$c_1v_1 + c_2v_2 + \cdots + c_kv_k,$$

where $c_1, c_2, \ldots, c_k$ are scalars, is called a **linear combination** of the vectors $v_1, v_2, \ldots, v_k$.

Definition 2.3.3 (Span)**.** Given vectors $v_1, v_2, \ldots, v_k$ in $\mathbb{R}^n$, the collection of all linear combinations of the vectors $v_1, v_2, \ldots, v_k$ is called the **span** of the set of vectors $\{v_1, v_2, \ldots, v_k\}$. We denote the span of $\{v_1, v_2, \ldots, v_k\}$ by

$$\text{span}\{v_1, v_2, \ldots, v_k\}.$$

We then have that

$$\text{span}\{v_1, v_2, \ldots, v_k\} = \{t_1v_1 + t_2v_2 + \cdots + t_kv_k \mid t_1, t_2, \ldots, t_k \in \mathbb{R}\}.$$

Example 2.3.4. Consider the vectors v_1 and v_2 in $\mathbb{R}^3$ given by

$$v_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \quad \text{and} \quad v_2 = \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix}.$$

Let's compute $\text{span}\{v_1, v_2\}$.

Solution: Write

$$\begin{aligned} \text{span}\{v_1, v_2\} &= \{c_1v_1 + c_2v_2 \mid c_1, c_2 \in \mathbb{R}\} \\ &= \left\{ c_1 \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + c_2 \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix} \;\middle|\; c_1, c_2 \in \mathbb{R} \right\} \\ &= \left\{ \begin{pmatrix} c_1 \\ c_1 \\ c_1 \end{pmatrix} + \begin{pmatrix} c_2 \\ 0 \\ 2c_2 \end{pmatrix} \;\middle|\; c_1, c_2 \in \mathbb{R} \right\} \\ &= \left\{ \begin{pmatrix} c_1 + c_2 \\ c_1 \\ c_1 + 2c_2 \end{pmatrix} \;\middle|\; c_1, c_2 \in \mathbb{R} \right\}. \end{aligned}$$

We then have that a vector $\begin{pmatrix} x \\ y \\ z \end{pmatrix}$ is in $\text{span}\{v_1, v_2\}$ if and only if

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} c_1 + c_2 \\ c_1 \\ c_1 + 2c_2 \end{pmatrix}$$

for some scalars c_1 and c_2; that is, if

$$\begin{cases} c_1 + c_2 & = & x \\ c_1 & = & y \\ c_1 + 2c_2 & = & z. \end{cases}$$

Substituting the second equation, $c_1 = y$, into the first and third equation leads to the two equation

$$\begin{cases} y + c_2 & = & x \\ y + 2c_2 & = & z. \end{cases}$$

Solving for c_2 in the first equation and substituting into the second yields the single equation

$$2x - y - z = 0.$$

This is the equation of a plane through the origin in $\mathbb{R}^3$ and containing the points with coordinates $(1, 1, 1)$ and $(1, 0, 2)$. □

In the previous example we showed that if a vector $\begin{pmatrix} x \\ y \\ z \end{pmatrix}$ is in the span,

$$W = \text{span}\left\{ \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix} \right\},$$

of the vectors $v_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$ and $v_2 = \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix}$ in $\mathbb{R}^3$, then it determines a point with coordinates (x, y, z) in $\mathbb{R}^3$ lying in the plane with equation $2x - y - z = 0$. Denote the plane by Q; that is,

$$Q = \left\{ \begin{pmatrix} x \\ y \\ z \end{pmatrix} \in \mathbb{R}^3 \,\middle|\, 2x - y - z = 0 \right\}.$$

Then, the previous example shows that W is a **subset** of Q. We write

$$W \subseteq Q,$$

meaning that every element in W is also an element in Q. We will presently show that Q is also a subset of W; that is, every point in the plane Q must also be in the span of the vectors $v_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$ and $v_2 = \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix}$.

Example 2.3.5. Let

$$W = \text{span}\left\{ \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix} \right\}$$

and

$$Q = \left\{ \begin{pmatrix} x \\ y \\ z \end{pmatrix} \in \mathbb{R}^3 \;\middle|\; 2x - y - z = 0 \right\}.$$

Show that $Q \subseteq W$.

Solution: To show that Q is a subset of W, we need to show that every point in the plane Q is a linear combination of the vectors $v_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$ and $v_2 = \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix}$.

Thus, let $\begin{pmatrix} x \\ y \\ z \end{pmatrix} \in Q$. Then,

$$2x - y - z = 0.$$

Solving for z in terms of x and y in the previous equation yields

$$z = 2x - y.$$

Thus, z depends on both x and y, which can be thought of as parameters. We therefore set $x = t$ and $y = s$, where t and s are parameters. We then have that, if $\begin{pmatrix} x \\ y \\ z \end{pmatrix} \in Q$, then

$$\left\{ \begin{array}{rcl} x & = & t \\ y & = & s \\ z & = & 2t - s. \end{array} \right.$$

In vector notation, we then have that, if $\begin{pmatrix} x \\ y \\ z \end{pmatrix} \in Q$, then

$$\begin{aligned} \begin{pmatrix} x \\ y \\ z \end{pmatrix} &= \begin{pmatrix} t \\ s \\ 2t - s \end{pmatrix} \\ &= \begin{pmatrix} t \\ 0 \\ 2t \end{pmatrix} + \begin{pmatrix} 0 \\ s \\ -s \end{pmatrix}, \end{aligned}$$

where we have used the definition of vector addition in $\mathbb{R}^3$. Thus, using now the definition of scalar multiplication, we get that

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = t \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix} + s \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix},$$

which shows that, if $\begin{pmatrix} x \\ y \\ z \end{pmatrix} \in Q$, then

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} \in \text{span}\left\{ \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix} \right\} = \text{span}\left\{ v_2, \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix} \right\}.$$

In order to complete the proof that $\begin{pmatrix} x \\ y \\ z \end{pmatrix} \in \text{span}\{v_1, v_2\}$, we will need to show that the vector $\begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}$ is in the span of the vectors v_1 and v_2; that is, we need to find scalars c_1 and c_2 such that

$$c_1 v_1 + c_2 v_2 = \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix},$$

or

$$\begin{pmatrix} c_1 + c_2 \\ c_1 \\ c_1 + 2c_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}.$$

This leads to the system of equations

$$\left\{ \begin{array}{rcl} c_1 + c_2 & = & 0 \\ c_1 & = & 1 \\ c_1 + 2c_2 & = & -1, \end{array} \right.$$

which has solution: $c_1 = 1$, $c_2 = -1$. Thus,

$$\begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix} = v_1 - v_2.$$

Consequently, if $\begin{pmatrix} x \\ y \\ z \end{pmatrix} \in Q$, then

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = c_1 v_2 + c_2 \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}$$

for some scalars c_1 and c_2, by what we have seen in the first part of this proof. Hence, since $\begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix} = v_1 - v_2$, it follows that

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = c_1 v_2 + c_2(v_1 - v_2) = c_2 v_1 + (c_2 - c_1) v_2,$$

which is a linear combination of v_1 and v_2. Hence,

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} \in Q \Rightarrow \begin{pmatrix} x \\ y \\ z \end{pmatrix} \in \text{span}\{v_1, v_2\} = W.$$

We have therefore shown that $Q \subseteq W$. □

The previous two examples show that the span of v_1 and v_2 is the same set as the plane in $\mathbb{R}^3$ with equation $2x - y - z = 0$. In other words, the combination of the statements

$$W \subseteq Q \quad \text{and} \quad Q \subseteq W$$

is equalivalent to the statement

$$W = Q.$$

2.4 Linear Independence

In the previous example we showed that the vector $v_3 = \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}$ is in the span of the set $\{v_1, v_2\}$, where

$$v_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \quad \text{and} \quad v_2 = \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix}.$$

When this happens (i.e., when one vector in the set is in the span of the other vectors) we say that the set $\{v_1, v_2, v_3\}$ is **linearly dependent**. In general, we have the following definition:

Definition 2.4.1 (Linear Dependence in $\mathbb{R}^n$)**.** A set of vectors, S, in $\mathbb{R}^n$ is said to be **linearly dependent** if at least one of the vectors in S is a finite linear combination of other vectors in S.

Example 2.4.2. We have already seen that the set $S = \{v_1, v_2, v_3\}$, where

$$v_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \quad v_2 = \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix}, \text{and} \quad v_3 = \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix},$$

is a linearly dependent subset of $\mathbb{R}^3$ since

$$v_3 = v_1 - v_2;$$

that is, v_3 is in the span of the other vectors in S.

Example 2.4.3. Let $v_1, v_2, \ldots, v_k$ be any vectors in $\mathbb{R}^n$. Then, the set

$$S = \{\mathbf{0}, v_1, v_2, \ldots, v_k\},$$

where $\mathbf{0}$ denotes the zero vector in $\mathbb{R}^n$, is linearly dependent since

$$\mathbf{0} = 0 \cdot v_1 + 0 \cdot v_2 + \cdots + 0 \cdot v_k;$$

that is, $\mathbf{0}$ is in the span of the vectors $v_1, v_2, \ldots, v_k$.

If a subset, S, of $\mathbb{R}^n$ is **not** linear dependent, we say that it is **linearly independent**.

Definition 2.4.4 (Linear Independence in $\mathbb{R}^n$)**.** A set of vectors, S, in $\mathbb{R}^n$ is said to be **linearly independent** if it is not linearly dependent; that is, no vector in S can be expressed as a linear combination of other vectors in S.

The following proposition gives an alternate characterization of linear independence for a finite subset of $\mathbb{R}^n$.

Proposition 2.4.5. The set $S = \{v_1, v_2, \ldots, v_k\}$ of vectors in $\mathbb{R}^n$ is linearly independent if and only if

$$c_1 = 0,\ c_2 = 0,\ \ldots,\ c_k = 0$$

is the only solution to the vector equation

$$c_1 v_1 + c_2 v_2 + \cdots + c_k v_k = \mathbf{0}.$$

Remark 2.4.6. Note that it is not hard to see that $c_1 = 0, c_2 = 0, \ldots, c_k = 0$ is a solution to the equation

$$c_1 v_1 + c_2 v_2 + \cdots + c_k v_k = \mathbf{0}. \tag{2.1}$$

The solution $c_1 = 0, c_2 = 0, \ldots, c_k = 0$ is usually referred to as the **trivial solution**. Thus, linear independence is equivalent to the statement that the trivial solution is the only solution to the equation in (2.1). Thus, linear dependence of the set $\{v_1, v_2, \ldots, v_k\}$ is equivalent to the statement that the equation

$$c_1 v_1 + c_2 v_2 + \cdots + c_k v_k = \mathbf{0}$$

has solutions in addition to the trivial solution.

Remark 2.4.7. The statement of Proposition 2.4.5 is a **bi–conditional**; that is, it is the combination of the two implications:

1. If the set $S = \{v_1, v_2, \ldots, v_k\}$ is linearly independent, then

   $$c_1 = 0,\ c_2 = 0,\ \ldots,\ c_k = 0$$

   is the only solution to the vector equation

   $$c_1 v_1 + c_2 v_2 + \cdots + c_k v_k = \mathbf{0};$$

2. Conversely, if
$$c_1 = 0,\ c_2 = 0,\ \ldots,\ c_k = 0$$
is the only solution to the vector equation
$$c_1 v_1 + c_2 v_2 + \cdots + c_k v_k = \mathbf{0},$$
then $S = \{v_1, v_2, \ldots, v_k\}$ is linearly independent.

Thus, in order to prove Proposition 2.4.5, the two implications need to be established.

We will now prove Proposition 2.4.5. This is the first formal proof that we present in the course and will therefore be presented with lots of details in order to illustrate how a mathematical argument is presented. Subsequent arguments in these notes will not be as detailed as this one.

Proof of Proposition 2.4.5. We first prove that if the set $S = \{v_1, v_2, \ldots, v_k\}$ is linearly independent, then
$$c_1 = 0,\ c_2 = 0,\ \ldots,\ c_k = 0$$
is the only solution to the vector equation
$$c_1 v_1 + c_2 v_2 + \cdots + c_k v_k = \mathbf{0}.$$
Suppose therefore that S is linearly independent. This means that no vector in S is in the span of the other vectors in S.

We wish to prove that the equation
$$c_1 v_1 + c_2 v_2 + \cdots + c_k v_k = \mathbf{0}$$
has only the trivial solution
$$c_1 = 0,\ c_2 = 0,\ \ldots,\ c_k = 0.$$
If this is not the case, then there exist scalars $c_1, c_2, \ldots, c_k$, such not all of them are zero and
$$c_1 v_1 + c_2 v_2 + \cdots + c_k v_k = \mathbf{0}.$$
Suppose the non–zero scalar is c_j, for some j in $\{1, 2, \ldots, k\}$, and write
$$c_1 v_1 + c_2 v_2 + \cdots + c_{j-1} v_{j-1} + c_j v_j + c_{j+1} v_{j+1} + \cdots + c_k v_k = \mathbf{0}. \tag{2.2}$$
We can solve for $c_j v_j$ in equation (2.2) by adding on both sides the additive inverses of the other vectors. Using the properties of vector addition we then get that
$$c_j v_j = -c_1 v_1 - c_2 v_2 - \cdots - c_{j-1} v_{j-1} - c_{j+1} v_{j+1} - \cdots - c_k v_k,$$

which, using now the properties of scalar multiplication can now be re–written as

$$c_jv_j = (-c_1)v_1+(-c_2)v_2+\cdots+(-c_{j-1})v_{j-1}+(-c_{j+1})v_{j+1}+\cdots+(-c_k)v_k. \quad (2.3)$$

Now, since $c_j \neq 0$, $1/c_j$ exists. We can then multiply both sides of equation (2.3) by $1/c_j$, and using now the distributive properties and the associative property for addition and scalar multiplication we obtain that

$$v_j = \left(\frac{-c_1}{c_j}\right)v_1 + \cdots + \left(\frac{-c_{j-1}}{c_j}\right)v_{j-1} + \left(-\frac{-c_{j+1}}{c_j}\right)v_{j+1} + \cdots + \left(\frac{-c_k}{c_j}\right)v_k. \quad (2.4)$$

Equation (2.4) displays v_j as a linear combination of $v_1, \ldots, v_{j-1}, v_{j+1}, \ldots, v_k$. However, this is impossible since we are assuming that S is linearly independent and therefore no no vector in S is in the span of the other vectors in S. This contradiction then implies that the equation

$$c_1v_1 + c_2v_2 + \cdots + c_kv_k = \mathbf{0}$$

has only the trivial solution

$$c_1 = 0,\ c_2 = 0,\ \ldots,\ c_k = 0,$$

which we had set out to prove.

Next, we prove the converse statement: if

$$c_1 = 0,\ c_2 = 0,\ \ldots,\ c_k = 0$$

is the only solution to the vector equation

$$c_1v_1 + c_2v_2 + \cdots + c_kv_k = \mathbf{0}, \quad (2.5)$$

then $S = \{v_1, v_2, \ldots, v_k\}$ is linearly independent.

Suppose that

$$c_1v_1 + c_2v_2 + \cdots + c_kv_k = \mathbf{0}$$

has only the trivial solution

$$c_1 = 0,\ c_2 = 0,\ \ldots,\ c_k = 0.$$

Arguing by contradiction again, assume that S is not linearly independent. Then, one of the vectors in S, say v_j, is in the span of the other vectors in S; that is, there exist scalars $c_1, c_2, \ldots, c_{j-1}, c_{j+1}, \ldots, c_k$ such that

$$v_j = c_1v_1 + c_2v_2 + \cdots + c_{j-1}v_{j-1} + c_{j+1}v_{j+1} + \cdots + c_kv_k. \quad (2.6)$$

Adding the additive inverse to both sides of equation (2.6) we obtain that

$$c_1v_1 + c_2v_2 + \cdots + c_{j-1}v_{j-1} - v_j + c_{j+1}v_{j+1} + \cdots + c_kv_k = \mathbf{0},$$

which may be re–written as

$$c_1v_1 + c_2v_2 + \cdots + c_{j-1}v_{j-1} + (-1)v_j + c_{j+1}v_{j+1} + \cdots + c_kv_k = \mathbf{0}. \tag{2.7}$$

Since $-1 \neq 0$, equation (2.7) shows that there is a non–trivial solution to the equation

$$c_1v_1 + c_2v_2 + \cdots + c_kv_k = \mathbf{0}.$$

This contradicts the assumption that the only solution to the equation in (2.5) is the trivial one. Therefore, it is not the case that S is linearly dependent and hence it must be linearly independent. □

Proposition 2.4.5 is very useful in determining whether a given set of vectors, $\{v_1, v_2, \ldots, v_k\}$, in $\mathbb{R}^n$ is linearly independent or not. According to Proposition 2.4.5, all we have to do is to solve the equation

$$c_1v_1 + c_2v_2 + \cdots + c_kv_k = \mathbf{0}$$

and determine whether it has one solution or more than one solution. In the first case (only the trivial solution) we can conclude by virtue of Proposition 2.4.5 that the set is linearly independent. In the second case (more than one solution), the set is linearly dependent.

Example 2.4.8. Determine whether the set $\left\{ \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix} \right\}$ is linearly independent in $\mathbb{R}^3$ or not.

***Solution**:* Consider the equation

$$c_1 \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + c_2 \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix} + c_3 \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}. \tag{2.8}$$

This equation leads to the system of linear equations

$$\left\{ \begin{array}{rcl} c_1 + c_2 + c_3 & = & 0 \\ c_1 + 2c_3 & = & 0 \\ c_1 + 2c_2 & = & 0, \end{array} \right. \tag{2.9}$$

Solving for c_3 in the first equation and substituting into the second equation leads to the system of two equations

$$\left\{ \begin{array}{rcl} -c_1 - 2c_2 & = & 0 \\ c_1 + 2c_2 & = & 0. \end{array} \right. \tag{2.10}$$

Observe that the system of equations in (2.10) is really a single equation in two unknowns

$$c_1 + 2c_2 = 0. \tag{2.11}$$

We can solve for c_1 in equation (2.11) and c_3 in the first equation 1n (2.9) to obtain that

$$\begin{aligned} c_1 &= -2c_2 \\ c_3 &= c_2, \end{aligned} \tag{2.12}$$

which shows that the unknown scalars c_1 and c_3 depend on c_2, which could taken on arbitrarily any value. To stress the arbitrary nature of c_2, let's rename it t, an arbitrary parameter. We then get from (2.12) that

$$\begin{aligned} c_1 &= -2t \\ c_2 &= t \\ c_3 &= t. \end{aligned} \tag{2.13}$$

Since the parameter t in (2.13) is arbitrary, we see that the system in (2.9) has infinitely many solutions. In particular, the vector equation (2.21) has non–trivial solutions. It then follows by virtue of Proposition 2.4.5 that the set $\left\{ \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix} \right\}$ is linearly dependent. □

Example 2.4.9. Determine whether the set $\left\{ \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\}$ is linearly independent in $\mathbb{R}^3$ or not.

Solution: Consider the equation

$$c_1 \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + c_2 \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix} + c_3 \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}. \tag{2.14}$$

This equation leads to the system of linear equations

$$\left\{ \begin{aligned} c_1 + c_2 \quad &= 0 \\ c_1 + 2c_3 \quad &= 0 \\ c_1 + 2c_2 + c_3 &= 0, \end{aligned} \right. \tag{2.15}$$

Solving for c_1 and c_2 in the first two equations in (2.15) leads to

$$\begin{aligned} c_1 &= 0 \\ c_2 &= 0. \end{aligned}$$

Substituting for these in the third equation in (2.15) then leads to

$$c_3 = 0.$$

We have therefore shown that the vector equation in (2.14) has only the trivial solution. Consequently, by virtue of Proposition 2.4.5 that the set $\left\{ \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\}$ is linearly independent. □

Remark 2.4.10. In the previous two examples we have seen that the question of whether a finite set of vectors in $\mathbb{R}^n$ is linearly independent or not leads to the question of whether a system of equations, like those in (2.9) and (2.15), has only the trivial solution or not. The systems in (2.9) and (2.15) are examples of **homogeneous** systems. In general, a homogenous system of linear of m equations in n unknowns is of the form

$$\left\{\begin{array}{rcl} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n & = & 0 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n & = & 0 \\ \vdots & = & \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n & = & 0, \end{array}\right. \tag{2.16}$$

where the $x_1, x_2, \ldots, x_n$ are the unknowns, and a_{ij}, for $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, n$, are known **coefficients**. We will study systems more systematically later in the course and we will see that what is illustrated in the previous two examples is what happens in general: either the linear homogenous system has only the trivial solution, or it has infinitely many solutions.

2.5 Subspaces of Euclidean Space

In this section we study some special subsets of Euclidean space, $\mathbb{R}^n$. These are called **subspaces** and are defined as follows

Definition 2.5.1 (Subspaces of $\mathbb{R}^n$)**.** A non–empty subset, W, of Euclidean space, $\mathbb{R}^n$, is said to be a **subspace** of $\mathbb{R}^n$ iff

(i) $v, w \in W$ implies that $v + w \in W$; and

(ii) $t \in \mathbb{R}$ and $v \in W$ implies that $tv \in W$.

If (i) and (ii) in Definition 2.5.1 hold, we say that the set W is **closed** under the vector space operations in $\mathbb{R}^n$. For this reason, properties (i) and (ii) are usually referred to as **closure properties**.

There are many examples of subspaces of $\mathbb{R}^n$; but there are also many examples of subsets of $\mathbb{R}^n$ which are not subspaces. We shall begin by presenting a few examples of subsets which are not subspaces.

Example 2.5.2 (Subsets which are not subspaces)**.**

1. The empty set, denoted by the symbol $\emptyset$, is not a subspace of any Euclidean space by definition.

2. Consider the subset, S, of $\mathbb{R}^2$ given by the first quadrant in the xy–plane:

$$S = \left\{ \begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R}^2 \;\middle|\; x \geqslant 0, y \geqslant 0 \right\}$$

S is not a subspace since $\begin{pmatrix} 1 \\ 1 \end{pmatrix} \in S$, but

$$(-1)\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} -1 \\ -1 \end{pmatrix}$$

is not in S because $-1 < 0$. That is, S is not closed under scalar multiplication.

3. Let $S \subseteq \mathbb{R}^2$ this time be given by

$$S = \left\{ \begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R}^2 \;\middle|\; xy \geqslant 0 \right\}.$$

In this case, S is closed under scalar multiplication, but it is not closed under vector addition. To see why this is so, let $\begin{pmatrix} x \\ y \end{pmatrix} \in S$. Then, $xy \geqslant 0$. Then, for any scalar t, note that

$$(tx)(ty) = t^2 xy \geqslant 0$$

since $t^2 \geqslant 0$ for any real number t. Thus, S is closed under scalar multiplication. However, S is not closed under vector addition; to see this, consider the vectors

$$v = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \text{and} \quad w = \begin{pmatrix} 0 \\ -1 \end{pmatrix}.$$

Then, v and w are both in S since

$$1 \cdot 0 = 0 \cdot (-1) = 0.$$

However,

$$v + w = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

is not in S since $1 \cdot (-1) = -1 < 0$.

Example 2.5.3 (Subsets which are subspaces)**.**

1. Let $W = \{\mathbf{0}\}$; that is, W consists solely of the additive identity, $\mathbf{0}$, in $\mathbb{R}^n$. W is a subspace of $\mathbb{R}^n$ because

$$\mathbf{0} + \mathbf{0} = \mathbf{0} \in W,$$

so that W is closed under vector addition; and

$$t \cdot \mathbf{0} = \mathbf{0} \in W \quad \text{for all } t \in \mathbb{R};$$

that is, W is closed under scalar multiplication.

2. $W = \mathbb{R}^n$, the entire Euclidean space, is also a subspace of $\mathbb{R}^n$.

3. Let $W = \left\{ \begin{pmatrix} x \\ y \\ z \end{pmatrix} \;\middle|\; ax + by + cz = 0 \right\}$, where a, b and c are real numbers, is a subspace of $\mathbb{R}^3$.

   Proof: Let $v = \begin{pmatrix} x_1 \\ y_1 \\ z_1 \end{pmatrix}$ and $w = \begin{pmatrix} x_2 \\ y_2 \\ z_2 \end{pmatrix}$ be in W. Then,

$$\begin{array}{rcl} ax_1 + by_1 + cz_1 & = & 0 \\ ax_2 + by_2 + cz_2 & = & 0. \end{array}$$

   Adding both equations yields

$$a(x_1 + x_2) + b(y_1 + y_2) + c(z_1 + z_2) = 0,$$

   where we have used the distributive property for real numbers. It then follows that

$$v + w = \begin{pmatrix} x_1 + x_2 \\ y_1 + y_2 \\ z_1 + z_2 \end{pmatrix} \in W,$$

   and so W is closed under vector addition in $\mathbb{R}^3$.

   Next, multiply $ax_1 + by_1 + cz_1 = 0$ on both sides by a scalar t and apply the distributive and associative properties for real numbers to get that

$$a(tx_1) + b(ty_1) + c(tz_1) = 0,$$

   which show that

$$tv = \begin{pmatrix} tx_1 \\ ty_1 \\ tz_1 \end{pmatrix} \in W,$$

   and therefore W is also closed with respect to scalar multiplication.

   Hence, W is closed with respect to the vector space operations in $\mathbb{R}^3$; that is, W is a subspace of $\mathbb{R}^n$.

   Let $S = \{v_1, v_2, \ldots, v_k\}$ be a subset of $\mathbb{R}^n$ and put $W = \text{span}(S)$. Then, W is a subspace of $\mathbb{R}^n$.

□

Proposition 2.5.4. Given a non–empty subset, S, of $\mathbb{R}^n$, $\text{span}(S)$ is a subspace of $\mathbb{R}^n$.

Proof: Since $S \neq \emptyset$, there is a vector $v \in S$. Observe that $v = 1 \cdot v$ is a linear combination of a vector from S; therefore, $v \in \text{span}(S)$ and therefore $\text{span}(S)$ is non–empty.

Next, we show that $\text{span}(S)$ is closed under the vector space operations of $\mathbb{R}^n$. Let $v \in \text{span}(S)$; then, there exist vectors $v_1, v_2, \ldots, v_k$ in S such that

$$v = c_1 v_1 + c_2 v_2 + \cdots + c_k v_k$$

for some scalars $c_1, c_2, \ldots, c_k$. Thus, for any scalar t,

$$\begin{aligned} tv &= t(c_1 v_1 + c_2 v_2 + \cdots + c_k v_k) \\ &= t(c_1 v_1) + t(c_2 v_2) + \cdots + t(c_k v_k) \\ &= (tc_1) v_1 + (tc_2) v_2 + \cdots + (tc_k) v_k, \end{aligned}$$

which shows that tv is a linear combination of elements in S; that is, $tv \in \text{span}(S)$. Consequently, $\text{span}(S)$ is closed under scalar multiplication.

To show that $\text{span}(S)$ is closed under vector addition, let v and w be in $\text{span}(S)$. Then, there exist vectors $v_1, v_2, \ldots, v_k$ and $w_1, w_2, \ldots, w_m$ in S such that

$$v = c_1 v_1 + c_2 v_2 + \cdots + c_k v_k$$

and

$$w = d_1 w_1 + d_2 w_2 + \cdots + d_m w_m,$$

for for some scalars $c_1, c_2, \ldots, c_k$ and $d_1, d_2, \ldots, d_m$. Thus,

$$v + w = c_1 v_1 + c_2 v_2 + \cdots + c_k v_k + d_1 w_1 + d_2 w_2 + \cdots + d_m w_m,$$

which is a linear combination of vectors in S. Therefore, $v + w \in \text{span}(S)$.

We have therefore that $\text{span}(S)$ is a non–empty subset of $\mathbb{R}^n$ which is closed under the vector space operations in $\mathbb{R}^n$; that is, $\text{span}(S)$ is a subspace of $\mathbb{R}^n$. □

Proposition 2.5.5. Given a non–empty subset, S, of $\mathbb{R}^n$, $\text{span}(S)$ is the smallest subspace of $\mathbb{R}^n$ which contains S; that is, is W is any subspace of $\mathbb{R}^n$ such that $S \subseteq W$, then $\text{span}(S) \subseteq W$.

Proof: Let V denote the smallest subspace of $\mathbb{R}^n$ that contains S; that is,

(i) V is a subspace of $\mathbb{R}^n$;

(ii) $S \subseteq V$; and

(iii) for any subspace, W, of $\mathbb{R}^n$ such that $S \subseteq W$, $V \subseteq W$.

We show that

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

By Proposition 2.5.4, $\text{span}(S)$ is a subspace of $\mathbb{R}^n$. Observe also that

$$S \subseteq \text{span}(S),$$

since $v \in S$ implies that $v = 1 \cdot v \in \text{span}(S)$. It then follows that

$$V \subseteq \text{span}(S), \tag{2.17}$$

since V is the smallest subset of $\mathbb{R}^n$ which contains S. It remains to show then that

$$\text{span}(S) \subseteq V.$$

Let $v \in \text{span}(S)$; then, there exist vectors $v_1, v_2, \ldots, v_k$ in S such that

$$v = c_1 v_1 + c_2 v_2 + \cdots + c_k v_k$$

for some scalars $c_1, c_2, \ldots, c_k$. Now, since $S \subseteq V$, $v_i \in V$ for all $i = 1, 2, \ldots, k$. It then follows from the closure of V with respect to scalar multiplication that

$$c_i v_i \in V \quad \text{for all } i - 1, 2, \ldots, k.$$

Applying the closure of V with respect to vector addition we then get that

$$c_1 v_1 + c_2 v_2 + \cdots + c_k v_k \in V;$$

that is $v \in V$. We have then shown that

$$v \in \text{span}(S) \Rightarrow v \in V;$$

that is,

$$\text{span}(S) \subseteq V.$$

Combining this with (2.17), we conclude that $\text{span}(S) = V$; that is, $\text{span}(S)$ is the smallest subspace of $\mathbb{R}^n$ which contains S. □

Remark 2.5.6 (The Span of the Empty Set)**.** In view of Proposition 2.5.5, it makes sense to define

$$\text{span}(\emptyset) = \{\mathbf{0}\}.$$

Indeed, $\{\mathbf{0}\}$ is the smallest subset of $\mathbb{R}^n$ and $\emptyset \subseteq \{\mathbf{0}\}$.

2.6 Finitely Generated Subspaces

We have seen that for any subset, S, of $\mathbb{R}^n$, $\text{span}(S)$ is a subspace of $\mathbb{R}^n$. If the set S is finite, we will say that $\text{span}(S)$ is a **finitely generate** subspace of $\mathbb{R}^n$.

Definition 2.6.1 (Finitely Generated Subspaces)**.** A subspace, W, of $\mathbb{R}^n$ is said to be finitely generate iff $W = \text{span}(S)$ for some finite subset S of $\mathbb{R}^n$.

Example 2.6.2. Since $\{\mathbf{0}\} = \text{span}(\emptyset)$, by definition, it follows that $\{\mathbf{0}\}$ is finitely generated because $\emptyset$ is finite.

Example 2.6.3. Let $e_1, e_2, \ldots, e_n$ be vectors in $\mathbb{R}^n$ given by

$$e_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \; e_2 = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \; \ldots, \; e_n = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}.$$

We show that

$$\mathbb{R}^n = \text{span}\{e_1, e_2, \ldots, e_n\}. \tag{2.18}$$

This will prove that $\mathbb{R}^n$ is finitely generated. To see why (2.18) is true, first observe that

$$\text{span}\{e_1, e_2, \ldots, e_n\} \subseteq \mathbb{R}^n. \tag{2.19}$$

Next, let $\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$ denote any vector in $\mathbb{R}^n$. We then have that

$$\begin{aligned} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} &= \begin{pmatrix} x_1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ x_2 \\ \vdots \\ 0 \end{pmatrix} + \cdots + \begin{pmatrix} 0 \\ 0 \\ \vdots \\ x_n \end{pmatrix} \\ &= x_1 \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + x_2 \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix} + \cdots + x_n \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix} \\ &= x_1 e_1 + x_2 e_2 + \cdots + x_n e_n, \end{aligned}$$

which shows that $\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$ is in the span of $\{e_1, e_2, \cdots, e_n\}$. Thus,

$$\mathbb{R}^n \subseteq \text{span}\{e_1, e_2, \ldots, e_n\}.$$

Combining this with (2.19) yields (2.18), which shows that $\mathbb{R}^n$ is finitely generated.

We will eventually show that all subspaces of $\mathbb{R}^n$ are finitely generated. Before we do so, however, we need to make a short incursion into the theory of systems of liner equations.

2.7 Connections with the Theory of Systems Linear Equations

We have seen that the questions of whether a given set of vectors in $\mathbb{R}^m$ is linearly independent can be translated into question of whether a homogeneous system of the form

$$\left\{\begin{array}{rcl} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n & = & 0 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n & = & 0 \\ \vdots & = & \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n & = & 0, \end{array}\right. \tag{2.20}$$

has only the trivial solution or many solutions. In this section we study these systems in more detail. In particular, we will see that in the case $m < n$, then the system (2.20) has infinitely many solutions. This result will imply that any set of n vectors in $\mathbb{R}^m$, where $n > m$, is linearly dependent. We will illustrate this with an example in $\mathbb{R}^2$.

Example 2.7.1. Let $v_1 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$, $v_2 = \begin{pmatrix} 2 \\ -1 \end{pmatrix}$ and $v_3 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$. Show that the set $\{v_1, v_2, v_3\}$ is linearly dependent.

Solution: Consider the equation

$$c_1v_1 + c_2v_2 + c_3v_3 = \mathbf{0}, \tag{2.21}$$

where $\mathbf{0}$ denotes the zero–vector in $\mathbb{R}^2$ and c_1, c_2 and c_3 are scalars. This vector equation leads to the system of equations

$$\left\{\begin{array}{rcl} c_1 + 2c_2 + c_3 & = & 0 \\ 2c_1 - c_2 + c_3 & = & 0. \end{array}\right. \tag{2.22}$$

Solving for c_1 in the first equation and substituting into the second equation leads to the system

$$\left\{\begin{array}{lcl} c_1 + 2c_2 + c_3 & = & 0 \\ -5c_2 - c_3 & = & 0. \end{array}\right. \tag{2.23}$$

Observe that systems (2.22) and (2.23) have the same solutions since we simply solved for one of the variables in one equation and substituted into the other. Similarly, we can now solve for c_2 in the second equation in (2.23) and substitute for it in the first equation of the same system to get

$$\left\{\begin{array}{lcl} c_1 + \frac{3}{5}c_3 & = & 0 \\ -5c_2 - c_3 & = & 0. \end{array}\right. \tag{2.24}$$

We can then solve for c_1 and c_2 in system (2.24) to get

$$\left\{\begin{array}{lcl} c_1 & = & -\frac{3}{5}c_3 \\ c_2 & = & -\frac{1}{5}c_3. \end{array}\right. \tag{2.25}$$

The variables c_1 and c_2 in system (2.24) are usually called the **leading variables** of the system; thus, the process of going from (2.24) to (2.25) is usually referred to as **solving for the leading variables**.

System (2.25) shows that the leading variables, c_1 and c_2, depend on c_3, which is arbitrary. We may therefore define $c_3 = -5t$, where t is an arbitrary parameter to get the solutions

$$\left\{\begin{array}{rcl} c_1 & = & 3t \\ c_2 & = & t \\ c_3 & = & -5t, \end{array}\right. \tag{2.26}$$

so that the solution spaces of system (2.22) is

$$W = \operatorname{span}\left\{\begin{pmatrix} 3 \\ 2 \\ -5 \end{pmatrix}\right\}.$$

We therefore conclude that the vector equation (2.21) has non–trivial solutions and therefore $\{v_1, v_2, v_3\}$ is linearly dependent. □

2.7.1 Elementary Row Operations

The process of going from system (2.22) to the system in (2.24) can also be achieved by a procedure that uses elimination of variables instead of substitution. For instances, we can multiply the first equation in (2.22) by the scalar -2, adding to the second equation and replacing the second equation by the result leads to the system:

$$\left\{\begin{array}{rcl} c_1 + 2c_2 + c_3 & = & 0 \\ -5c_2 - c_3 & = & 0, \end{array}\right. \tag{2.27}$$

which is the same system that we got in (2.24). This procedure does not change the solution space of the original system. In general, the solution space for the pair of equations

$$\left\{\begin{array}{rcl} a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n & = & 0 \\ a_{k1}x_1 + a_{k2}x_2 + \cdots + a_{kn}x_n & = & 0 \end{array}\right. \tag{2.28}$$

is the same as that of the pair

$$\left\{\begin{array}{lcl} a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n & = & 0 \\ (ca_{i1} + a_{k1})x_1 + (ca_{i2} + a_{k2})x_2 + \cdots + (ca_{in} + a_{kn})x_n & = & 0, \end{array}\right. \tag{2.29}$$

where c is any scalar. To see why this is so, let $\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$ be a solution of system (2.29); thus, from the second equation in the system,

$$(ca_{i1} + a_{k1})x_1 + (ca_{i2} + a_{k2})x_2 + \cdots + (ca_{in} + a_{kn})x_n \quad = \quad 0.$$

It then follows, using the distributive properties, that

$$ca_{i1}x_1 + a_{k1}x_1 + ca_{i2}x_2 + a_{k2}x_2 + \cdots + ca_{in}x_n + a_{kn}x_n = 0.$$

Thus, by the associative properties and the distributive property again,

$$c(a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n) + a_{k1}x_1 + a_{k2}x_2 + \cdots + a_{kn}x_n = 0.$$

Consequently, since $\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$ also satisfies the first equation in (2.29), we get that

$$a_{k1}x_1 + a_{k2}x_2 + \cdots + a_{kn}x_n = 0,$$

which is the second equation in (2.28). Hence, $\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$ is also a solution of system (2.28). A similar argument shows that if $\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$ is also a solution of system (2.28), then it is also a solution of system (2.29).

Adding a scalar multiple of one equation to another equation and replacing the second equation by the resulting equation is an example of an **elementary row operation**. Other elementary row operations are: (1) multiply an equation by a no–zero scalar and replace the equation by the result of the scalar multiple, and (2) swap two equations. It is clear that the later operation does not change the solution space of the system; in the former operation, since the scalar is non-zero, the solution space does not change either. To see why this is the case, note that if $\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$ is a solution of

$$c(a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n) = 0,$$

then, since $c \neq 0$, we see that

$$a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n = 0.$$

We illustrate this by multiplying the second equation in (2.27) by $-1/5$ to get.

$$\left\{ \begin{aligned} c_1 + 2c_2 + c_3 &= 0 \\ c_2 + \tfrac{1}{5}c_3 &= 0, \end{aligned} \right. \tag{2.30}$$

The system in (2.30) is in what is known as **row echelon form**, in which the leading variables c_1 and c_2 have 1 as coefficient.

We can perform a final row operation on the system in (2.30) by multiplying the second equation in the system by the scalar -2 and adding to the first equation to get

$$\left\{ \begin{array}{rcl} c_1 + \frac{3}{5}c_3 & = & 0 \\ c_2 + \frac{1}{5}c_3 & = & 0. \end{array} \right. \tag{2.31}$$

The system in (2.31) is said to be in **reduced row echelon form**. It can be solved for the leading variables to yield the system in (2.25).

2.7.2 Gaussian Elimination

Observe that in going from system (2.22) to system (2.31) by performing elementary row operations in the equations, as outlined in the previous section, the operations only affected the coefficients; the variables c_1, c_2 and c_3 acted as place–holders. It makes sense, therefore, to consider the coefficients only in order to optimize calculations. The coefficients in each equation in system (2.22) can be represented as rows in an array of numbers shown in equation

$$\begin{pmatrix} 1 & 2 & 1 & | & 0 \\ 2 & -1 & 1 & | & 0 \end{pmatrix}. \tag{2.32}$$

The two–dimensional array in (2.32) is known as the **augmented matrix** for the system (2.22). The elementary operations can then be performed on the rows of the augmented matrix in (2.32) (hence the name, elementary **row** operations). If we denote the rows first and second row in the matrix in (2.32) by R_1 and R_2, respectively, we can denote and keep track of the row operations as follows:

$$-2R_1 + R_2 \to R_2 : \quad \begin{pmatrix} 1 & 2 & 1 & | & 0 \\ 0 & -5 & -1 & | & 0 \end{pmatrix}. \tag{2.33}$$

$-2R_1 + R_2 \to R_2$ in (2.33) indicates that we have multiplied the first row in (2.32) by -2, added the scalar product to the second, and replaced the second row by the result. The rest of the operations can be indicated as follows:

$$(-1/5)R_2 \to R_2 : \quad \begin{pmatrix} 1 & 2 & 1 & | & 0 \\ 0 & 1 & 1/5 & | & 0 \end{pmatrix}, \tag{2.34}$$

and

$$-2R_2 + R_1 \to R_1 : \quad \begin{pmatrix} 1 & 0 & 3/5 & | & 0 \\ 0 & 1 & 1/5 & | & 0 \end{pmatrix}. \tag{2.35}$$

The matrix in (2.34) is in row echelon form, and that in (2.35) is in reduced row echelon form.

The process of going from an augmented matrix for a system to any of its row echelon forms by performing elementary row operations is known as **Gaussian Elimination** or **Gauss–Jordan reduction**. We will present here two more examples in the context of determining whether a given set of vectors is linearly independent or not.

Example 2.7.2. Determine whether the set of vectors $\{v_1, v_2, v_3\}$ in $\mathbb{R}^3$, where

$$v_1 = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \quad v_2 = \begin{pmatrix} 2 \\ 5 \\ 1 \end{pmatrix} \quad \text{and} \quad v_3 = \begin{pmatrix} 0 \\ -4 \\ 3 \end{pmatrix},$$

is linearly independent or not.

Solution: Consider the equation

$$c_1 \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} + c_2 \begin{pmatrix} 2 \\ 5 \\ 1 \end{pmatrix} + c_3 \begin{pmatrix} 0 \\ -4 \\ 3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}. \tag{2.36}$$

This leads to the system

$$\left\{ \begin{aligned} c_1 + 2c_2 \qquad &= 0 \\ 5c_2 - 4c_3 &= 0 \\ -c_1 + c_2 + 3c_3 &= 0. \end{aligned} \right. \tag{2.37}$$

Starting with the augmented matrix

$$\left(\begin{array}{ccc|c} 1 & 2 & 0 & 0 \\ 0 & 5 & -4 & 0 \\ -1 & 1 & 3 & 0 \end{array} \right), \tag{2.38}$$

we perform the following elementary row operations on the matrix in (2.38):

$$R_1 + R_3 \to R_3 \qquad \left(\begin{array}{ccc|c} 1 & 2 & 0 & 0 \\ 0 & 5 & -4 & 0 \\ 0 & 3 & 3 & 0 \end{array} \right),$$

$$(1/5)R_2 \to R_2 \qquad \left(\begin{array}{ccc|c} 1 & 2 & 0 & 0 \\ 0 & 1 & -4/5 & 0 \\ 0 & 3 & 3 & 0 \end{array} \right),$$

$$-3R_2 + R_3 \to R_3 \qquad \left(\begin{array}{ccc|c} 1 & 2 & 0 & 0 \\ 0 & 1 & -4/5 & 0 \\ 0 & 0 & 27/5 & 0 \end{array} \right),$$

and

$$(5/27)R_3 \to R_3 \qquad \left(\begin{array}{ccc|c} 1 & 2 & 0 & 0 \\ 0 & 1 & -4/5 & 0 \\ 0 & 0 & 1 & 0 \end{array} \right),$$

where we have indicated the row operation by the row on which the operation was performed. It then follows that the system in (2.37) is equivalent to the system

$$\left\{ \begin{aligned} c_1 + 2c_2 \qquad &= 0 \\ c_2 - (4/5)c_3 &= 0 \\ c_3 &= 0. \end{aligned} \right. \tag{2.39}$$

System (2.39) is in row echelon form and can be solved to yield

$$c_3 = c_2 = c_1 = 0.$$

Consequently, the vector equation (2.36) has only the trivial solution, and therefore the set $\{v_1, v_2, v_3\}$ is linearly independent. □

Example 2.7.3. Determine whether the set of vectors $\{v_1, v_2, v_3\}$ in $\mathbb{R}^3$, where

$$v_1 = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \quad v_2 = \begin{pmatrix} 2 \\ 5 \\ 1 \end{pmatrix} \quad \text{and} \quad v_3 = \begin{pmatrix} 0 \\ 5 \\ 3 \end{pmatrix},$$

is linearly independent or not.

Solution: Consider the equation

$$c_1 \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} + c_2 \begin{pmatrix} 2 \\ 5 \\ 1 \end{pmatrix} + c_3 \begin{pmatrix} 0 \\ 5 \\ 3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}. \tag{2.40}$$

This leads to the system

$$\begin{cases} c_1 + 2c_2 & = & 0 \\ \quad 5c_2 + 5c_3 & = & 0 \\ -c_1 + c_2 + 3c_3 & = & 0. \end{cases} \tag{2.41}$$

Starting with the augmented matrix

$$\left(\begin{array}{ccc|c} 1 & 2 & 0 & 0 \\ 0 & 5 & 5 & 0 \\ -1 & 1 & 3 & 0 \end{array}\right), \tag{2.42}$$

we perform the following elementary row operations on the matrix in (2.42):

$$\begin{array}{c} (1/5)R_2 \to R_2 \\ R_1 + R_3 \to R_3 \end{array} \quad \left(\begin{array}{ccc|c} 1 & 2 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 3 & 3 & 0 \end{array}\right),$$

$$-3R_2 + R_3 \to R_3 \quad \left(\begin{array}{ccc|c} 1 & 2 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right),$$

and

$$-2R_2 + R_1 \to R_1 \quad \left(\begin{array}{ccc|c} 1 & 0 & -2 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right).$$

We then conclude that the system (2.41) is equivalent to the system

$$\begin{cases} c_1 \quad - \quad 2c_3 & = & 0 \\ \qquad c_2 + c_3 & = & 0, \end{cases} \tag{2.43}$$

which is in reduced row echelon form and can be solved for the leading variables c_1 and c_2 to yield

$$\begin{aligned} c_1 &= 2t \\ c_2 &= -t \\ c_3 &= t, \end{aligned}$$

where t is an arbitrary parameter. It then follows that the system in (2.41) has infinitely many solutions; consequently, the vector equation in (2.40) has nontrivial solutions and therefore the set $\{v_1, v_2, v_3\}$ is linearly dependent. □

Examples 2.7.2 and 2.7.3 illustrate what can happen in general when we are solving homogeneous linear systems: either (1) the system has only the trivial solution; or (2) the system has infinitely many solutions. In the case in which there are infinitely many solutions, the system is equivalent to one in which there are more unknowns than equations, as was the case in system (2.43) Example 2.7.3. In the following section we prove that this is always the case: a homogenous systems with more unknowns than equations has infinitely many solutions.

2.7.3 The Fundamental Theorem of Homogenous Systems of Linear Equations

The fundamental theorem for homogenous linear systems states that a homogenous system of more unknowns than equations has a nontrivial solutions. This is Theorem 1.5E in our text [TT70, pg. 16]. We present here slight variation of that theorem:

Theorem 2.7.4 (Fundamental Theorem of Homogeneous Linear Systems)**.** *A homogeneous system of m linear equations in n unknowns,*

$$\left\{\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= 0 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= 0 \\ \vdots \qquad\qquad & \ \vdots \ \ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= 0, \end{aligned}\right. \tag{2.44}$$

with $n > m$ has infinitely many solutions.

Proof: Perform Gauss–Jordan reduction of the augmented matrix of the system in (2.44) to obtain the equivalent augmented matrix

$$\left(\begin{array}{cccccccc|c} 1 & b_{12} & b_{13} & \cdots & b_{1k} & b_{1,k+1} & \cdots & b_{1n} & 0 \\ 0 & 1 & b_{23} & \cdots & b_{2k} & b_{2,k+1} & \cdots & b_{2n} & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & b_{k,k+1} & \cdots & b_{kn} & 0 \end{array}\right)$$

in row–echelon form, where $k \leqslant m$. The previous system can be further reduced to

$$\begin{pmatrix} 1 & 0 & 0 & \cdots & 0 & b'_{1,k+1} & \cdots & b'_{1n} & | & 0 \\ 0 & 1 & 0 & \cdots & 0 & b'_{2,k+1} & \cdots & b'_{2n} & | & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & | & \vdots \\ 0 & 0 & 0 & \cdots & 1 & b'_{k,k+1} & \cdots & b'_{kn} & | & 0 \end{pmatrix},$$

which leads to the system

$$\left\{ \begin{array}{rcl} x_1 + b'_{1,k+1}x_{k+1} + b'_{1,k+2}x_3 + \cdots + b'_{1n}x_n & = & 0 \\ x_2 + b_{2,k+1}x_{k+1} + \cdots + b_{2n}x_n & = & 0 \\ \vdots & \vdots & \vdots \\ x_k + b'_{k,k+1}x_{k+1} \cdots + b'_{k,n}x_n & = & 0, \end{array} \right. \tag{2.45}$$

where $k \leqslant m$, equivalent to (2.44). We can solve for the leading variables, $x_1, x_2, \ldots, x_k$ in (2.45) in terms of $x_{k+1}, \ldots, x_n$, which can be set to equal arbitrary parameters. Since $n > m$ and $k \leqslant m$, there are $n-k \geqslant 1$ such parameters, It follows that system (2.45) has infinitely many solutions. Consequently, (2.44) has infinitely many solutions. □

A consequence of the Fundamental Theorem 2.7.4 is the following Proposition which will play a crucial role in the study of subspaces of $\mathbb{R}^n$ in the next section.

Proposition 2.7.5. *Any set of vectors $\{v_1, v_2, \ldots, v_k\}$ in $\mathbb{R}^n$ with $k > n$ must be linearly dependent.*

Proof: Consider the vector equation

$$c_1v_1 + c_2v_2 + \cdots + c_kv_k, = \mathbf{0}. \tag{2.46}$$

Since the set $\{v_1, v_2, \ldots, v_k\}$ is a subset of $\mathbb{R}^n$, we can write

$$v_1 = \begin{pmatrix} a_{11} \\ a_{21} \\ a_{31} \\ \vdots \\ a_{n1} \end{pmatrix}, \quad v_2 = \begin{pmatrix} a_{12} \\ a_{22} \\ a_{32} \\ \vdots \\ a_{n2} \end{pmatrix}, \quad \cdots, \quad v_k = \begin{pmatrix} a_{1k} \\ a_{2k} \\ a_{3k} \\ \vdots \\ a_{nk} \end{pmatrix}.$$

Hence, the vector equation in (2.46) translate into the homogeneous system

$$\left\{ \begin{array}{rcl} a_{11}c_1 + a_{12}c_2 + \cdots + a_{1k}c_k & = & 0 \\ a_{21}c_1 + a_{22}c_2 + \cdots + a_{2k}c_k & = & 0 \\ \vdots & \vdots & \vdots \\ a_{n1}c_1 + a_{n2}n_2 + \cdots + a_{nk}c_k & = & 0, \end{array} \right. \tag{2.47}$$

of n linear equations in k unknowns. Since $k > n$, the homogenous system in (2.47) has more unknowns than equations. It then follows from the Fundamental Theorem 2.7.4 that system (2.47) has infinitely many solutions. It then follows that the vector equation in (2.46) has a nontrivial solution, and therefore, by Proposition 2.4.5, the set $\{v_1, v_2, \ldots, v_k\}$ is linearly dependent. □

Example 2.7.6. By Proposition 2.7.5, the set

$$S = \left\{ \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}, \begin{pmatrix} 0 \\ 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 5 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 5 \\ 4 \end{pmatrix} \right\}$$

is a linearly dependent subset of $\mathbb{R}^3$. We will now show how to find a subset of S which is linearly independent and which also spans span(S).

Solution: Denote the elements of S by v_1, v_2, v_3 and v_4, respectively, and consider the vector equation

$$c_1 v_1 + c_2 v_2 + c_3 v_3 + c_4 v_4 = \mathbf{0}. \tag{2.48}$$

Since S is a linearly dependent, equation (2.48) has nontrivial solutions. Our goal now is to find those nontrivial solutions to obtain nontrivial linear relations between the elements of S which will allow us to express some of the vectors as linear combinations of the other ones. Those vectors in S which can be expressed as linear combinations of the others can be discarded. We perform this procedure until we find a linearly independent subset of S which which also spans span(S).

Equation (2.48) leads to the system

$$\left\{ \begin{aligned} c_1 + c_3 - c_4 &= 0 \\ c_1 + 2c_2 + 5c_3 + 5c_4 &= 0 \\ -c_1 + c_2 + c_3 + 4c_4 &= 0, \end{aligned} \right. \tag{2.49}$$

which has the augmented matrix

$$\begin{matrix} R_1 \\ R_2 \\ R_3 \end{matrix} \quad \left(\begin{array}{cccc|c} 1 & 0 & 1 & -1 & 0 \\ 1 & 2 & 5 & 5 & 0 \\ -1 & 1 & 1 & 4 & 0 \end{array} \right).$$

Performing the elementary row operations $-R_1 + R_2 \to R_2$ and $R_1 + R_3 \to R_3$, we obtain the augmented matrix:

$$\left(\begin{array}{cccc|c} 1 & 0 & 1 & -1 & 0 \\ 0 & 2 & 4 & 6 & 0 \\ 0 & 1 & 2 & 3 & 0 \end{array} \right).$$

Next, perform $\frac{1}{2}R_2 \to R_2$ and $-R_2 + R_3 \to R_3$ in succession to obtain

$$\left(\begin{array}{cccc|c} 1 & 0 & 1 & -1 & 0 \\ 0 & 1 & 2 & 3 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array} \right).$$

Hence, the system in (2.57) is equivalent to the system

$$\left\{ \begin{aligned} c_1 + c_3 - c_4 &= 0 \\ c_2 + 2c_3 + 3c_4 &= 0. \end{aligned} \right. \tag{2.50}$$

Solving for the leading variables c_1 and c_2 in (2.50) then yields the solutions

$$\begin{cases} c_1 &= t+s \\ c_2 &= 2t-3s \\ c_3 &= -t \\ c_4 &= s, \end{cases} \tag{2.51}$$

where t and s are arbitrary parameters.

Taking $t = 1$ and $s = 0$ in (2.51) yields the nontrivial linear relation

$$v_1 + 2v_2 - v_3 = \mathbf{0},$$

by virtue of the vector equation (2.48), which shows that $v_3 = v_1 + 2v_2$ and therefore

$$v_3 \in \text{span}\{v_1, v_2\}. \tag{2.52}$$

Similarly, taking $t = 0$ and $s = 1$ in (2.51) yields the nontrivial linear relation

$$v_1 - 3v_2 + v_4 = \mathbf{0},$$

from which we get that $v_4 = -v_1 + 3v_2$, and therefore

$$v_4 \in \text{span}\{v_1, v_2\}. \tag{2.53}$$

It follows from (2.52) and (2.53) that

$$\{v_1, v_2, v_3, v_4\} \subseteq \text{span}\{v_1, v_2\}.$$

Consequently, since $\text{span}\{v_1, v_2, v_3, v_4\}$ is the smallest subspace of $\mathbb{R}^3$ which contains $\{v_1, v_2, v_3, v_4\}$, by Proposition 2.5.5,

$$\text{span}\{v_1, v_2, v_3, v_4\} \subseteq \text{span}\{v_1, v_2\}.$$

Combining this with

$$\text{span}\{v_1, v_2\} \subseteq \text{span}\{v_1, v_2, v_3, v_4\},$$

we obtain that

$$\text{span}\{v_1, v_2\} = \text{span}(S).$$

It remains to check that $\{v_1, v_2\}$ is linearly independent. However, this follows from the fact that v_1 and v_2 are not scalar multiples of each other. □

2.7.4 Nonhomogeneous Systems

Asking whether a vector $v \in \mathbb{R}^n$ is in the span of the set $\{v_1, v_2, \ldots, v_k\}$ in $\mathbb{R}^n$ leads to the system of n linear equations in k unknowns

$$\begin{cases} a_{11}c_1 + a_{12}c_2 + \cdots + a_{1k}c_k &= b_1 \\ a_{21}c_1 + a_{22}c_2 + \cdots + a_{2k}c_k &= b_2 \\ \qquad\qquad \vdots & \vdots\ \ \vdots \\ a_{n1}c_1 + a_{n2}c_2 + \cdots + a_{nk}c_k &= b_n, \end{cases} \tag{2.54}$$

where

$$v_1 = \begin{pmatrix} a_{11} \\ a_{21} \\ a_{31} \\ \vdots \\ a_{n1} \end{pmatrix}, \quad v_2 = \begin{pmatrix} a_{12} \\ a_{22} \\ a_{32} \\ \vdots \\ a_{n2} \end{pmatrix}, \quad \cdots, \quad v_k = \begin{pmatrix} a_{1k} \\ a_{2k} \\ a_{3k} \\ \vdots \\ a_{nk} \end{pmatrix},$$

and

$$v = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \\ \vdots \\ b_n \end{pmatrix}.$$

If v is not the zero–vector in $\mathbb{R}^n$, then the system in (2.54) is a nonhomogeneous. In general, nonhomogeneous system might or might not have solutions. If they do have a solution, they either have exactly one solution or infinitely many solutions.

We can analyze the system in (2.54) by considering the augmented matrix

$$\left(\begin{array}{cccc|c} a_{11} & a_{12} & \cdots & a_{1k} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2k} & b_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nk} & b_n \end{array}\right) \tag{2.55}$$

and performing elementary row operations on the rows of the matrix in (2.55).

Example 2.7.7. Determine whether or not the vector $\begin{pmatrix} 1 \\ 2 \\ -3 \end{pmatrix}$, is in the span of the set

$$S = \left\{ \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}, \begin{pmatrix} 0 \\ 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 5 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 5 \\ 4 \end{pmatrix} \right\}.$$

Solution: Denote the elements of S by v_1, v_2, v_3 and v_4, respectively, and consider the vector equation

$$c_1 v_1 + c_2 v_2 + c_3 v_3 + c_4 v_4 = v, \tag{2.56}$$

where

$$v = \begin{pmatrix} 1 \\ 2 \\ -3 \end{pmatrix}.$$

Equation (2.56) leads to the system

$$\left\{ \begin{array}{rcl} c_1 + c_3 - c_4 & = & 1 \\ c_1 + 2c_2 + 5c_3 + 5c_4 & = & 2 \\ -c_1 + c_2 + c_3 + 4c_4 & = & -3, \end{array} \right. \tag{2.57}$$

which has the augmented matrix

$$\begin{matrix} R_1 \\ R_2 \\ R_3 \end{matrix} \quad \left(\begin{array}{cccccc} 1 & 0 & 1 & -1 & | & 1 \\ 1 & 2 & 5 & 5 & | & 2 \\ -1 & 1 & 1 & 4 & | & -3 \end{array}\right).$$

Performing the elementary row operations $-R_1 + R_2 \to R_2$ and $R_1 + R_3 \to R_3$, we obtain the augmented matrix:

$$\left(\begin{array}{cccccc} 1 & 0 & 1 & -1 & | & 1 \\ 0 & 2 & 4 & 6 & | & 1 \\ 0 & 1 & 2 & 3 & | & -2 \end{array}\right).$$

Next, perform $\frac{1}{2}R_2 \to R_2$ and $-R_2 + R_3 \to R_3$ in succession to obtain

$$\left(\begin{array}{cccccc} 1 & 0 & 1 & -1 & | & 1 \\ 0 & 1 & 2 & 3 & | & 1/2 \\ 0 & 0 & 0 & 0 & | & -5/2 \end{array}\right).$$

The third row in the previous matrix yields $0 = -5/2$, which is impossible. Therefore, the vector equation in (2.56) is not solvable. Hence, $\begin{pmatrix} 1 \\ 2 \\ -3 \end{pmatrix}$ is not in the span of the set S. □

2.8 Maximal Linearly Independent Subsets

The goal of this section is to prove that every subspace, W, of $\mathbb{R}^n$ is the span of a linearly independent subset, S. In other words,

Theorem 2.8.1. *Let W be a subspace of $\mathbb{R}^n$. There exists a subset, S, of W such that*

(i) S is linearly independent, and

(ii) $W = span(S)$.

In the proof of Theorem 2.8.1 we will use Proposition 2.7.5, which says that any set of vectors $\{v_1, v_2, \ldots, v_k\}$ in $\mathbb{R}^n$ with $k > n$ must be linearly dependent, and the following

Lemma 2.8.2. *Let $S = \{v_1, v_2, \ldots, v_k\}$ be a linearly independent subset of $\mathbb{R}^n$. If $v \notin span(S)$, then the set*

$$S \cup \{v\} = \{v_1, v_2, \ldots, v_k, v\}$$

is linearly independent.

Remark 2.8.3. The set $S \cup \{v\}$ is called the **union** of the sets S and $\{v\}$.

Proof of Lemma 2.8.2: Suppose that S is linearly independent and that $v \notin \text{span}(S)$. Consider the vector equation

$$c_1v_1 + c_2v_2 + \cdots + c_kv_k + cv = \mathbf{0}. \tag{2.58}$$

We first show that $c = 0$. For, if $c \neq 0$, then we can solve for v in the vector equation (2.58) to get

$$v = \left(-\frac{c_1}{c}\right)v_1 + \left(-\frac{c_2}{c}\right)v_2 + \cdots + \left(-\frac{c_k}{c}\right)v_k, \tag{2.59}$$

where we have used the additive inverse, additive identity, associative and distributive properties of the vector space operations in $\mathbb{R}^n$. Equation (2.59) displays v as a linear combination of the vectors in S; that is, v is an element of the span of S. However, this contradicts the assumption that $v \notin \text{span}(S)$. It then follows that $c = 0$, and therefore, using (2.58),

$$c_1v_1 + c_2v_2 + \cdots + c_kv_k = \mathbf{0}. \tag{2.60}$$

It then follows from (2.60) and the linear independence of S that

$$c_1 = c_2 = \cdots = c_k = 0.$$

Hence, $c_1 = c_2 = \cdots = c_k = c = 0$ is the only solution of (2.58) and, therefore, $S \cup \{v\}$ is linearly independent by Proposition 2.4.5. □

We are now in a position to prove Theorem 2.8.1.

Proof of Theorem 2.8.1: Let W be a subspace of $\mathbb{R}^n$. If $W = \{\mathbf{0}\}$, then

$$W = \text{span}(\emptyset);$$

therefore, $S = \emptyset$ in this case, and the proof is done.

On the other hand, if $W \neq \{\mathbf{0}\}$, there exists v_1 in W such that $v_1 \neq \mathbf{0}$. Thus, $\{v_1\}$ is linearly independent. If $W = \text{span}\{v_1\}$, set $S = \{v_1\}$ and the proof is done. Otherwise, there exists v_2 in W such that $v_2 \notin \text{span}\{v_1\}$. Then, by Lemma 2.8.2, the set $\{v_1, v_2\}$ is linearly independent.

We may now proceed by induction to obtain a linearly independent subset $S = \{v_1, v_2, \ldots, v_k\}$ of W as follows: having found a linearly independent subset $\{v_1, v_2, \ldots, v_{k-1}\}$ of W such that $\text{span}\{v_1, v_2, \ldots, v_{k-1}\} \neq W$, pick $v_k \in W$ such that $v_k \notin \text{span}\{v_1, v_2, \ldots, v_{k-1}\}$. Then, by Lemma 2.8.2, the set $\{v_1, v_2, \ldots, v_{k-1}, v_k\}$ is linearly independent.

We claim that this process has to stop for some value of $k \leqslant n$. The reason for this is that, by Proposition 2.7.5, if $k > n$, then S is linearly dependent. Furthermore, $S = \{v_1, v_2, \ldots, v_k\}$ has the property that, every vector, v, in W, the set $S \cup \{v\}$ is linearly dependent. We therefore obtain a subset, S, of W with the properties

(i) S is linearly independent, and

(ii) for every $v \in W$, the set $S \cup \{v\}$ is linearly dependent.

We claim that S must span W. To see why this is so, first observe that, since $S \subseteq W$, and W is a subspace of $\mathbb{R}^n$, we get that

$$\text{span}(S) \subseteq W, \tag{2.61}$$

since span(S) is the smallest subspace of $\mathbb{R}^n$ which contains the set S. Thus, it remains to show that

$$W \subseteq \text{span}(S). \tag{2.62}$$

If (2.62) does not hold true, then there exists $v \in W$ such that $v \notin \text{span}(S)$. It then follows by Lemma 2.8.2 that the set $S \cup \{v\}$ is linearly independent, but this contradicts (ii) above. Consequently, every v in W must in span(S) and (2.62) follows.

Combining (2.61) and (2.62) yields

$$\text{span}(S) = W,$$

which is (ii) in the statement of Theorem 2.8.1. Since S was constructed to be linearly independent, we also get that (i) in Theorem 2.8.1 also holds and we have therefore completed the proof of Theorem 2.8.1. □

Remark 2.8.4. The subset S of W which we constructed in the proof of Theorem 2.8.1 has the properties that: (i) S is linearly independent, and (ii) for every vector $v \in W$, the set $S \cup \{v\}$ is linearly dependent. A set with these two properties is called a **maximal linearly independent subset** subset of W. Thus, we have proved that every subspace of $\mathbb{R}^n$ has a maximal linearly independent subset.

2.9 Bases

A maximal linearly independent subset for a subspace, W, of $\mathbb{R}^n$ is also called a **basis** for W.

Definition 2.9.1 (Basis of a Subspace)**.** Let W be a subspace of $\mathbb{R}^n$. A subset B of W is said to be a basis for W if and only if

(i) B is linearly independent, and

(ii) $W = \text{span}(B)$.

Example 2.9.2. Let $W = \mathbb{R}^n$ and B consist of the vectors $e_1, e_2, \ldots, e_n$ in $\mathbb{R}^n$ given by

$$e_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \; e_2 = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \; \ldots, \; e_n = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}.$$

We show that B is a basis for $\mathbb{R}^n$; in other words B is linearly independent and it spans $\mathbb{R}^n$.

We first show that B is linearly independent.

Consider the vector equation

$$c_1e_1 + c_2e_2 + \cdots + c_ne_n = \mathbf{0}, \tag{2.63}$$

or

$$c_1 \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + c_2 \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix} + \cdots + c_n \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix},$$

which leads to

$$\begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix},$$

from which we get that

$$c_1 = c_2 = \ldots = c_n = 0$$

is the only solution of the vector equation in (2.63). Hence, B is linearly independent.

Next, we show that $\mathbb{R}^n = \mathrm{span}(B)$. To see why this is so, observe that for any vector, $\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$, in $\mathbb{R}^n$,

$$\begin{aligned} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} &= \begin{pmatrix} x_1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ x_2 \\ \vdots \\ 0 \end{pmatrix} + \cdots + \begin{pmatrix} 0 \\ 0 \\ \vdots \\ x_n \end{pmatrix} \\ &= x_1 \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + x_2 \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix} + \cdots + x_n \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix} \\ &= x_1e_1 + x_2e_2 + \cdots + x_ne_n, \end{aligned}$$

which shows that $\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$ is in the span of $\{e_1, e_2, \cdots, e_n\}$. Thus,

$$\mathbb{R}^n \subseteq \mathrm{span}(B).$$

On the other hand, since $B \subseteq \mathbb{R}^n$, we get that

$$\text{span}(B) \subseteq \mathbb{R}^n.$$

Thus,

$$\mathbb{R}^n = \text{span}(B).$$

Definition 2.9.3 (Standard Basis for $\mathbb{R}^n$)**.** The set $\{e_1, e_2, \cdots, e_n\}$, denoted by $\mathcal{E}_n$, is called the **standard basis** for $\mathbb{R}^n$.

Example 2.9.4. Let $W = \left\{ \begin{pmatrix} x \\ y \\ z \end{pmatrix} \in \mathbb{R}^3 \;\middle|\; x + 2y - z = 0 \right\}$. We have seen that W is a subspace of $\mathbb{R}^3$. Find a basis for W.

Solution: W is the solution space of the homogeneous linear equation

$$x + 2y - z = 0.$$

Solving for x in tees of y and z, and setting these to be arbitrary parameters $-t$ and s, respectively, we get the solutions

$$\begin{aligned} x &= 2t + s \\ y &= -t \\ z &= s, \end{aligned}$$

from which we get that

$$W = \left\{ \begin{pmatrix} x \\ y \\ z \end{pmatrix} \in \mathbb{R}^3 \;\middle|\; \begin{pmatrix} x \\ y \\ z \end{pmatrix} = t \begin{pmatrix} 2 \\ -1 \\ 0 \end{pmatrix} + s \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \right\}.$$

In other words,

$$W = \text{span} \left\{ \begin{pmatrix} 2 \\ -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \right\}.$$

Thus, the set

$$B = \left\{ \begin{pmatrix} 2 \\ -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \right\}$$

is a candidate for a basis for W. To show that B is a basis, it remains to show that it is linearly independent. So, consider the vector equation

$$c_1 \begin{pmatrix} 2 \\ -1 \\ 0 \end{pmatrix} + c_2 \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix},$$

which is equivalent to the system

$$\begin{cases} 2c_1 - c_2 &= 0 \\ -c_1 &= 0 \\ \quad c_2 &= 0, \end{cases}$$

from which we read that $c_1 = c_2 = 0$ is the only solution. Consequently, B is linearly independent.

We therefore conclude that B is a basis for W. □

2.10 Dimension

A remarkable fact about bases for a subspace, W, of $\mathbb{R}^n$ is that any two bases of W must have the same number of vectors. For example, in Example 2.9.4 we saw that

$$B = \left\{ \begin{pmatrix} 2 \\ -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \right\}$$

is a basis for the plane in $\mathbb{R}^3$ given by $W = \left\{ \begin{pmatrix} x \\ y \\ z \end{pmatrix} \in \mathbb{R}^3 \,\middle|\, x + 2y - z = 0 \right\}$.

We did this by solving the equation

$$x + 2y - z = 0$$

for x in terms of y and z and setting the last two variables to be arbitrary parameters. However, we could have instead solved for z in terms of x and y. This would have yielded the basis

$$B_1 = \left\{ \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix} \right\}.$$

Another basis for W is provided by the set

$$B_2 = \left\{ \begin{pmatrix} 1 \\ -1 \\ -1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 3 \end{pmatrix} \right\}.$$

Notice that, in all three cases, the bases consist of two vectors; i.e., the three bases for W displayed above have the same number of elements. The goal of this section is to prove that this result holds true in general:

Theorem 2.10.1 (Invariance of number of elements in bases)**.** Let W be a subspace of $\mathbb{R}^n$. If B_1 and B_2 are two bases of W, then B_1 and B_2 have the same number of elements.

Theorem 2.10.1 is the basis for the following definition:

Definition 2.10.2 (Definition of Dimension)**.** Let W be a subspace of $\mathbb{R}^n$. The **dimension** of W, denoted $\dim(W)$, is the number of elements in a basis of W.

Example 2.10.3.

- If $W = \mathbb{R}^n$, then $\dim(W) = n$ since the standard basis, $\mathcal{E}_n$, for $\mathbb{R}^n$ has n vectors (see Example 2.9.2 on page 44 in these notes).
- If $W = \{\mathbf{0}\}$, then $\dim(W) = 0$ since $\{\mathbf{0}\} = \text{span}(\emptyset)$ and $\emptyset$ has no vectors.
- If $W = \left\{ \begin{pmatrix} x \\ y \\ z \end{pmatrix} \in \mathbb{R}^3 \;\middle|\; x + 2y - z = 0 \right\}$, then $\dim(W) = 2$, since

$$B = \left\{ \begin{pmatrix} 2 \\ -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \right\}$$

  is a basis for W.

Remark 2.10.4. Note that for any subspace W of $\mathbb{R}^n$, $\dim(W) \leqslant n$. This last statement follows from Theorem 2.8.1 and Proposition 2.7.5.

In order to prove the Invariance Theorem 2.10.1, we will need the following lemma, which can be thought of as an extension of Proposition 2.7.5:

Lemma 2.10.5. *Let W be a subspace of $\mathbb{R}^n$ with a basis $B = \{w_1, w_2, \ldots, w_k\}$. Any set of vectors $\{v_1, v_2, \ldots, v_m\}$ in W, with $m > k$, must be linearly dependent.*

Proof: Consider the vector equation

$$c_1 v_1 + c_2 v_2 + \cdots + c_m v_m, = \mathbf{0}. \tag{2.64}$$

Since the set $B = \{w_1, w_2, \ldots, w_k\}$ is a basis for W, we can write each v_j, $j = 1, 2, \ldots, m$, as liner combination of the vectors in B:

$$\begin{aligned}
v_1 &= a_{11} w_1 + a_{21} w_2 + \cdots + a_{k1} w_k, \\
v_2 &= a_{12} w_1 + a_{22} w_2 + \cdots + a_{k2} w_k, \\
v_3 &= a_{13} w_1 + a_{23} w_2 + \cdots + a_{k3} w_k, \\
&\vdots \\
v_m &= a_{1m} w_1 + a_{2m} w_2 + \cdots + a_{km} w_k.
\end{aligned}$$

Substituting for v_j, $j = 1, \ldots, m$, in the vector equation in (2.64) and applying the distributive and associative properties yields the vector equation

$$\begin{aligned}
&(a_{11} c_1 + a_{12} c_2 + \cdots + a_{1m} c_m) w_1 \\
&\quad + (a_{21} c_1 + a_{22} c_2 + \cdots + a_{2m} c_m) w_2 \\
&\quad\quad + \cdots \\
&\quad\quad + (a_{k1} c_1 + a_{k2} c_2 + \cdots + a_{km} c_m) w_k = \mathbf{0}.
\end{aligned} \tag{2.65}$$

Next, since the set $B = \{w_1, w_2, \ldots, w_k\}$ is linearly independent, it follows from (2.65) that

$$\begin{cases} a_{11}c_1 + a_{12}c_2 + \cdots + a_{1m}c_m &= 0 \\ a_{21}c_1 + a_{22}c_2 + \cdots + a_{2m}c_m &= 0 \\ \vdots & \vdots \quad \vdots \\ a_{k1}c_1 + a_{k2}n_2 + \cdots + a_{km}c_m &= 0, \end{cases} \tag{2.66}$$

which is a homogeneous system of of k linear equations in m unknowns. Since $m > k$, the homogenous system in (2.66) has more unknowns than equations. It then follows from the Fundamental Theorem 2.7.4 that system (2.66) has infinitely many solutions. Consequently, the vector equation in (2.64) has a nontrivial solution, and therefore, by Proposition 2.4.5, the set $\{v_1, v_2, \ldots, v_m\}$ is linearly dependent. □

Proof of the Invariance Theorem 2.10.1. Let B_1 and B_2 be two bases for the subspace, W, of $\mathbb{R}^n$. Let k denote the number of vectors in B_1 and m the number of vectors in B_2. We show that

$$k = m. \tag{2.67}$$

If $m > k$, it follows from Lemma 2.10.5 the B_2 is linearly dependent; but this impossible since B_2 is a basis for W and is, therefore, linearly independent. Thus,

$$m \leqslant k. \tag{2.68}$$

The same argument applied to B_1 and B_2 interchanged implies that

$$k \leqslant m. \tag{2.69}$$

Equation (2.67) follows by combining (2.68) and (2.69), and the Theorem is proved. □

2.11 Coordinates

Another remarkable fact about bases for subspaces of $\mathbb{R}^n$ is the following

Theorem 2.11.1 (Coordinates Theorem)**.** *Let W be a subspace of $\mathbb{R}^n$ and*

$$B = \{w_1, w_2, \ldots, w_k\}$$

be a basis for W. Given any vector, v, in W, there exists a unique set of scalars $c_1, c_2, \ldots, c_k$ such that

$$v = c_1w_1 + c_2w_2 + \cdots + c_kw_k.$$

Proof: Since B spans W, there exist scalars $c_1, c_2, \ldots, c_k$ such that

$$v = c_1w_1 + c_2w_2 + \cdots + c_kw_k. \tag{2.70}$$

It remains to show that $c_1, c_2, \ldots, c_k$ are the only scalars for which (2.70) works.

Suppose that there is another set of scalars $d_1, d_2, \ldots, d_k$ such that

$$v = d_1 w_1 + d_2 w_2 + \cdots + d_k w_k. \tag{2.71}$$

Combining (2.70) and (2.71), we then obtain that

$$c_1 w_1 + c_2 w_2 + \cdots + c_k w_k = d_1 w_1 + d_2 w_2 + \cdots + d_k w_k. \tag{2.72}$$

Adding $(-d_1)w_1 + (-d_2)w_2 + \cdots + (-d_k)w_k$ on both sides of equation (2.72) and applying the associative and distributive properties we obtain that

$$(c_1 - d_1)w_1 + (c_2 - d_2)w_2 + \cdots + (c_k - d_k)w_k = \mathbf{0}. \tag{2.73}$$

It then follows from (2.73) and the linear independence of the basis $B = \{w_1, w_2, \ldots, w_k\}$ that

$$c_1 - d_1 = c_2 - d_2 = \cdots = c_k - d_k = 0,$$

from which we get

$$d_1 = c_1, \; d_2 = c_2, \; \ldots, \; d_k = c_k.$$

This proves the uniqueness of the coefficients $c_1, c_2, \ldots, c_k$ for the expansion of v given in (2.70) in terms of the vectors in the basis B. □

Definition 2.11.2 (Ordered Basis)**.** Let W be a subspace of $\mathbb{R}^n$ of dimension k and let B denote a basis for W. If the elements in B are listed in a specified order: $B = \{w_1, w_2, \ldots, w_k\}$, then B is called an **ordered basis**. In this sense, the basis $B_1 = \{w_2, w_1, \ldots, w_k\}$ is different from B even though, as sets, B and B_1 are the same; that is, the contain the same elements. However, as ordered bases, B and B_1 are not the same.

Definition 2.11.3 (Coordinates Relative to a Basis)**.** Let W be a subspace of $\mathbb{R}^n$ and

$$B = \{w_1, w_2, \ldots, w_k\}$$

be an ordered basis for W. Given any vector, v, in W, the **coordinates of** v **relative to the basis** B, are the unique set of scalars $c_1, c_2, \ldots, c_k$ such that

$$v = c_1 w_1 + c_2 w_2 + \cdots + c_k w_k.$$

We denote the coordinates of v relative to the basis B by the symbol $[v]_B$ and write $[v]_B = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_k \end{pmatrix}$. The vector $[v]_B$ in $\mathbb{R}^k$ is also called the **coordinates vector for** v **with respect to the basis** B.

Example 2.11.4. Let $W = \left\{ \begin{pmatrix} x \\ y \\ z \end{pmatrix} \in \mathbb{R}^3 \;\middle|\; x + 2y - z = 0 \right\}$. We have seen that the set

$$B = \left\{ \begin{pmatrix} 2 \\ -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \right\}$$

is a basis for W. Observe that the vector $v = \begin{pmatrix} 1 \\ 1 \\ 3 \end{pmatrix}$ is in W. To find the coordinates of v relative to B, we need to solve the vector equation

$$c_1 \begin{pmatrix} 2 \\ -1 \\ 0 \end{pmatrix} + c_2 \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 3 \end{pmatrix}$$

for c_1 and c_2. We see that $c_1 = -1$ and $c_2 = 3$, so that

$$[v]_B = \begin{pmatrix} -1 \\ 3 \end{pmatrix}.$$

Observe that the coordinate vector $[v]_B$ is a vector in $\mathbb{R}^2$ since W is a two–dimensional subspace of $\mathbb{R}^3$.

2.12 Euclidean Inner Product and Norm

The reason $\mathbb{R}^n$ is called Euclidean space is that, in addition to the vector space structure that we have discussed so far, there is also defined on $\mathbb{R}^n$ a product between vectors in $\mathbb{R}^n$ which produces a scalar. We shall denote the new product by the symbol $\langle v, w\rangle$ for vectors v and w in $\mathbb{R}^n$. We will call $\langle v, w\rangle$ the **Euclidean inner product** of v and w, or simply, the inner product of v and w.

2.12.1 Definition of Euclidean Inner Product

Before we give a formal definition of the inner product, let us show how we can multiply a row–vector and a column–vector.

Definition 2.12.1 (Row–Column Product)**.** Given a row–vector, R, of dimension n and a column–vector, C, also of the same dimension n, we define the product RC as follows:

Write $R = [x_1 \quad x_2 \quad \cdots \quad x_n]$ and $C = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$; then,

$$RC = [x_1 \quad x_2 \quad \cdots \quad x_n] \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = x_1y_1 + x_2y_2 + \cdots + x_ny_n.$$

Definition 2.12.2 (Transpose of a vector)**.** Given a vector $v = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$ in $\mathbb{R}^n$, the **transpose** of v, denoted by v^T, is the row vector

$$v^T = \begin{pmatrix} x_1 & x_2 & \cdots & x_n \end{pmatrix}.$$

Definition 2.12.3. *Given vectors* $v = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$ *and* $w = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$, *the* **inner product** *of v and w is the real number (or scalar), denoted by $\langle v, w\rangle$, obtained as follows*

$$\langle v, w\rangle = v^T w = \begin{pmatrix} x_1 & x_2 & \cdots & x_n \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = x_1y_1 + x_2y_2 + \cdots + x_ny_n.$$

The inner product defined above satisfies the following properties:
Given vectors v, w, v_1, v_2, w_1 and w_2 in $\mathbb{R}^n$,

(i) Symmetry: $\langle v, w\rangle = \langle w, v\rangle$;

(ii) Bi-Linearity: $\langle c_1v_1 + c_2v_2, w\rangle = c_1\langle v_1, w\rangle + c_2\langle v_2, w\rangle$, for scalars c_1 and c_2, and $\langle v, d_1w_1 + d_2w_2\rangle = d_1\langle v, w_1\rangle + d_2\langle v, w_2\rangle$, for scalars d_1 and d_2; and

(iii) Positive Definiteness: $\langle v, v\rangle \geqslant 0$ for all $v \in \mathbb{R}^n$ and $\langle v, v\rangle = 0$ if and only if v is the zero vector.

These properties follow from the definition can be easily checked; for instance, to verify (i), write $v = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$ and $w = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$. Then,

$$\langle w, v \rangle = [y_1 \quad y_2 \quad \cdots \quad y_n] \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = y_1x_1 + y_2x_2 + \cdots + y_nx_n.$$

Thus, since multiplication of real numbers is commutative,

$$\langle w, v \rangle = x_1y_1 + x_2y_2 + \cdots + x_ny_n = \langle v, w \rangle,$$

which shows the symmetry of the Euclidean inner product.

To verify the second part of the bi–linearity property, write

$$v = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \ w_1 = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \text{ and } \ w_2 = \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{pmatrix}.$$

Then, for scalars d_1 and d_2,

$$\begin{aligned} \langle v, d_1w_1 + d_2w_2 \rangle &= [x_1 \quad x_2 \quad \cdots \quad x_n] \begin{pmatrix} d_1y_1 + d_2z_1 \\ d_1y_2 + d_2z_2 \\ \vdots \\ d_1y_n + d_2z_n \end{pmatrix} \\ &= x_1(d_1y_1 + d_2z_1) + x_2(d_1y_2 + d_2z_2) + \cdots + x_n(d_1y_n + d_2z_n). \end{aligned}$$

Next, use the distributive and associative properties to get

$$\begin{aligned} \langle v, d_1w_1 + d_2w_2 \rangle &= d_1(x_1y_1x_2y_2 + \cdots + x_ny_n) + d_2(x_1z_1 + x_2z_2 + \cdots + x_nz_n) \\ &= d_1\langle v, w_1 \rangle + d_2\langle v, w_2 \rangle. \end{aligned}$$

Finally, the positive–definiteness property of the Euclidean inner product follows from the observation that, if $v = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$, then

$$\langle v, v \rangle = x_1^2 + x_2^2 + \cdots + x_n^2$$

is a sum of non–negative terms; and this sum is zero if and only if all the terms are zero.

Given an inner product in a vector space, we can define a norm as follows.

Definition 2.12.4 (Euclidean Norm in $\mathbb{R}^n$)**.** *For any vector $v \in \mathbb{R}^n$, its Euclidean norm, denoted $\|v\|$, is defined by*

$$\|v\| = \sqrt{\langle v, v\rangle}.$$

Observe that, by the positive definiteness of the inner product, this definition makes sense. Note also that we have defined the norm of a vector to be the *positive* square root of the the inner product of the vector with itself. Thus, the norm of any vector is always non–negative.

If P is a point in $\mathbb{R}^n$ with coordinates $(x_1, x_2, \ldots, x_n)$, the norm of the vector $\overrightarrow{OP}$ that goes from the origin to P is the distance from P to the origin; that is,

$$\text{dist}(O, P) = \|\overrightarrow{OP}\| = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}.$$

If $P_1(x_1, x_2, \ldots, x_n)$ and $P_2(y_1, y_2, \ldots, y_n)$ are any two points in $\mathbb{R}^n$, then the distance from P_1 to P_2 is given by

$$\text{dist}(P_1, P_2) = \|\overrightarrow{OP_2} - \overrightarrow{OP_2}\| = \sqrt{(y_1 - x_1)^2 + (y_2 - x_2)^2 + \cdots + (y_n - x_n)^2}.$$

2.12.2 Euclidean Norm

As a consequence of the properties of the inner product, we obtain the following properties of the norm:

Proposition 2.12.5 (Properties of the Norm)**.** *Let v denote a vector in $\mathbb{R}^n$ and c a scalar. Then,*

(i) $\|v\| \geqslant 0$ *and* $\|v\| = 0$ *if and only if v is the zero vector.*

(ii) $\|cv\| = |c|\|v\|$.

We also have the following very important inequality

Theorem 2.12.6 (The Cauchy–Schwarz Inequality)**.** *Let v and w denote vectors in $\mathbb{R}^n$; then,*

$$|\langle v, w\rangle| \leqslant \|v\|\|w\|.$$

Proof. Consider the function $f\colon \mathbb{R} \to \mathbb{R}$ given by

$$f(t) = \|v + tw\|^2 \quad \text{for all} \quad t \in \mathbb{R}.$$

Using the definition of the norm, we can write

$$f(t) = \langle v + tw, v + tw\rangle.$$

We can now use the properties of the inner product to expand this expression and get

$$f(t) = \|v\|^2 + 2t\langle v, w\rangle + t^2\|w\|^2.$$

Thus, $f(t)$ is a quadratic polynomial in t which is always non–negative. Therefore, it can have at most one real root. It then follows that

$$(2\langle v, w\rangle)^2 - 4\|w\|^2\|v\|^2 \leqslant 0,$$

from which we get

$$(\langle v, w\rangle)^2 \leqslant \|w\|^2\|v\|^2.$$

Taking square roots on both sides yields the inequality. □

The Cauchy–Schwarz inequality, together with the properties of the inner product and the definition of the norm, yields the following inequality known as the *Triangle Inequality.*

Proposition 2.12.7 (The Triangle Inequality)**.** *For any v and w in $\mathbb{R}^n$,*

$$\|w + w\| \leqslant \|v\| + \|w\|.$$

Proof. This is an Exercise. □

Definition 2.12.8 (Unit vectors)**.** A vector $u \in \mathbb{R}^n$ is said to be a **unit vector** if $\|u\| = 1$.

Remark 2.12.9 (Normalization)**.** Given a non–zero vector v in $\mathbb{R}^n$, we can define a unit vector in the direction of v as follows:

$$u = \frac{1}{\|v\|}v.$$

Then,

$$\|u\| = \left\|\frac{1}{\|v\|}v\right\| = \left|\frac{1}{\|v\|}\right|\|v\| = \frac{1}{\|v\|}\|v\| = 1.$$

We call $\dfrac{1}{\|v\|}v$ the **normalization** of v and usually denotes it by $\widehat{v}$.

2.12.3 Orthogonality

Definition 2.12.10 (Orthogonality)**.** *Two vectors v and w in $\mathbb{R}^n$ are said to be **orthogonal**, or perpendicular, if*

$$\langle v, w\rangle = 0.$$

Example 2.12.11. Let $v \in \mathbb{R}^n$ and define $W = \{w \in \mathbb{R}^n \mid \langle w, v\rangle = 0\}$; that is, W is the set of all vectors in $\mathbb{R}^n$ which are orthogonal to v.

(a) Prove that W is a subspace of $\mathbb{R}^n$.

   Solution: First, observe that $W \neq \emptyset$ because $\langle \mathbf{0}, v\rangle = 0$ and therefore $\mathbf{0} \in W$ and so W is nonempty.

Next, we show that W is closed under addition and scalar multiplication.

To see that W is closed under scalar multiplication, observe that, by the bi–linearity property of the inner product, if $w \in W$, then

$$\langle\langle v, tw\rangle = t\langle v, w\rangle = t \cdot 0 = 0$$

for all $t \in \mathbb{R}$.

To show that W is closed under vector addition, let w_1 and w_2 be two vectors in W. Then, applying the bi–linearity property of the inner product again,

$$\langle w_1 + w_2, v\rangle = \langle w_1, v\rangle + \langle w_2, v\rangle = 0 + 0 = 0;$$

hence, $w_1 + w_2 \in W$. □

(b) Suppose that $v \neq \mathbf{0}$ and compute $\dim(W)$.

Solution: Let $B = \{w_1, w_2, \ldots w_k\}$ be a basis for W. Then, $\dim(W) = k$ and we would like to determine what k is.

First note that $v \notin \text{span}(B)$. For, suppose that $v \in \text{span}(B) = W$, then

$$\langle v, v\rangle = 0.$$

Thus, by the positive definiteness of the Euclidean inner product, it follows that $v = \mathbf{0}$, but we are assuming that $v \neq \mathbf{0}$. Consequently, the set

$$B \cup \{v\} = \{w_1, w_2, \ldots w_k, v\}$$

is linearly independent. We claim that $B \cup \{v\}$ also spans $\mathbb{R}^n$. To see why this is so, let $u \in \mathbb{R}^n$ be any vector in $\mathbb{R}^n$, and let

$$t = \frac{\langle u, v\rangle}{\|v\|^2}.$$

Write

$$u = tv + (u - tv),$$

and observe that $u - tv \in W$. To see why this is so, compute

$$\begin{aligned}
\langle u - tv, v\rangle &= \langle u, v\rangle - t\langle v, v\rangle \\
&= \langle u, v\rangle - t\|v\|^2 \\
&= \langle u, v\rangle - \frac{\langle u, v\rangle}{\|v\|^2}\|v\|^2 \\
&= \langle u, v\rangle - \langle u, v\rangle \\
&= 0.
\end{aligned}$$

Thus, $u - tv \in W$. It then follows that there exist scalars $c_1, c_2, \ldots, c_k$ such that

$$u - tv = c_1 w_1 + c_2 w_2 + \cdots + c_k w_k.$$

Thus,

$$u = c_1 w_1 + c_2 w_2 + \cdots + c_k w_k + tv,$$

which shows that $u \in \text{span}(B \cup \{v\})$. Consequently, $B \cup \{v\}$ spans $\mathbb{R}^n$. Therefore, since $B \cup \{v\}$ is also linearly independent, it forms a basis for $\mathbb{R}^n$. We then have that $B \cup \{v\}$ must have n vectors in it, since $\dim(\mathbb{R}^n) = n$; that is,

$$k + 1 = n,$$

from which we get that

$$\dim(W) = n - 1.$$

□

Chapter 3

Spaces of Matrices

Matrices are rectangular arrays of numbers. More precisely, an $m \times n$ matrix is an array of numbers made up of n columns, with each column consisting of m scalar entries:

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \tag{3.1}$$

The columns of the matrix in (3.1) are the vectors

$$v_1 = \begin{pmatrix} a_{11} \\ a_{21} \\ a_{31} \\ \vdots \\ a_{m1} \end{pmatrix}, \quad v_2 = \begin{pmatrix} a_{12} \\ a_{22} \\ a_{32} \\ \vdots \\ a_{m2} \end{pmatrix}, \quad \cdots, \quad v_n = \begin{pmatrix} a_{1n} \\ a_{2n} \\ a_{3n} \\ \vdots \\ a_{mn} \end{pmatrix}$$

in $\mathbb{R}^m$.

We have already encountered matrices in this course, in connection with systems of linear equations, when we discussed elementary row operations in the augmented matrix corresponding to a system. We will see later in this course that the connection between linear systems and matrices is a very important in the theory of linear equations.

We will denote by $\mathbb{M}(m, n)$ the collection of all $m \times n$ matrices with real entries. We will see that $\mathbb{M}(m, n)$ has the structure of a vector space with addition and scalar multiplication defined in a manner analogous to those for vectors in Euclidean space. In addition to the vector space structure, there is a way to define a **matrix product** between a matrix in $\mathbb{M}(m, n)$ and a matrix in $\mathbb{M}(n, k)$, in that order, to yield a matrix in $\mathbb{M}(m, k)$. This gives rise to a **matrix algebra** in the space of square matrices (i.e., matrices in $\mathbb{M}(n, n)$), which we will also discuss in this chapter.

3.1 Vector Space Structure in $\mathbb{M}(m, n)$

Given matrices

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

and

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

in $\mathbb{M}(m, n)$, we will use the shorthand notation

$$A = [a_{ij}], \qquad 1 \leqslant i \leqslant m,\ 1 \leqslant j \leqslant n;$$

and

$$B = [b_{ij}], \qquad 1 \leqslant i \leqslant m,\ 1 \leqslant j \leqslant n.$$

We define the vector sum of A and B, denoted by $A + B$, by

$$A + B = [a_{ij} + b_{ij}], \qquad 1 \leqslant i \leqslant m,\ 1 \leqslant j \leqslant n;$$

that is, we add corresponding components to obtain the matrix sum of A and B.

Example 3.1.1. Let A and B be the 2×3 matrices given by

$$A = \begin{pmatrix} 4 & 0 & 7 \\ -7 & 4 & 0 \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} 7 & -4 & 0 \\ 4 & -7 & -4 \end{pmatrix}.$$

Then,

$$A + B = \begin{pmatrix} 11 & -4 & 7 \\ -3 & -3 & -4 \end{pmatrix}.$$

Note that if $A, B \in \mathbb{M}(m, n)$, then $A + B \in \mathbb{M}(m, n)$.

Similarly, we can define the scalar product of a scalar, c, with a matrix $A = [a_{ij}]$ in $\mathbb{M}(m, n)$ by

$$cA = [ca_{ij}], \qquad 1 \leqslant i \leqslant m,\ 1 \leqslant j \leqslant n.$$

Example 3.1.2. Let A and B be as in Example 3.1.1. Then,

$$2A = \begin{pmatrix} 8 & 0 & 14 \\ -14 & 8 & 0 \end{pmatrix} \quad \text{and} \quad (-3)B = \begin{pmatrix} -21 & 12 & 0 \\ -12 & 21 & 12 \end{pmatrix}.$$

We can therefore form the linear combination

$$2A + (-3)B = \begin{pmatrix} -13 & 12 & 14 \\ -26 & 29 & 12 \end{pmatrix}.$$

Definition 3.1.3 (Equality of Matrices)**.** We say that two matrices are equal iff corresponding entries are the same. In symbols, write $A = [a_{ij}]$ and $B = [b_{ij}]$; we say that $A = B$ iff

$$a_{ij} = b_{ij}, \quad \text{for } 1 \leqslant i \leqslant m \text{ and } 1 \leqslant j \leqslant n.$$

The operations of matrix addition and scalar multiplication can be shown to satisfy the following properties:

1. **Properties of Matrix Addition**

   Let A, B and C denote matrices in $\mathbb{M}(m, n)$. Then,

   (i) **Commutativity of Matrix Addition**

   $$A + B = B + A$$

   (ii) **Associativity of Matrix Addition**

   $$(A + B) + C = A + (B + C)$$

   (iii) **Existence of an Additive Identity**

   The matrix $O = [o_{ij}] \in \mathbb{M}(m, n)$ given by $o_{i,j} = 0$, for all $1 \leqslant i \leqslant m$ and $1 \leqslant j \leqslant n$, has the property that

   $$A + O = O + A = A \quad \text{for all } A \text{ in } \mathbb{M}(m, n).$$

   (iv) **Existence of an Additive Inverse**

   Given $A = [a_{ij}]$ in $\mathbb{M}(m, n)$, the matrix $W = [w_{ij}] \in \mathbb{M}(m, n)$ defined by $w_{ij} = -a_{ij}$ for $1 \leqslant i \leqslant m$ and $1 \leqslant j \leqslant n$ has the property that

   $$A + W = W + A = O.$$

   The matrix W is called an additive inverse of A and is denoted by $-A$.

2. **Properties of Scalar Multiplication**

   (i) **Associativity of Scalar Multiplication**

   Given scalars t and s and a matrix A in $\mathbb{M}(m, n)$,

   $$t(sA) = (ts)A.$$

   (ii) **Identity in Scalar Multiplication**

   The scalar 1 has the property that

   $$1 \cdot A = A \quad \text{for all } A \in \mathbb{M}(m, n).$$

3. **Distributive Properties**

Given matrices A and B in $\mathbb{M}(m,n)$, and scalars t and s,

(i) $t(A+B) = tA + tB$

(ii) $(t+s)A = tA + sA$.

All these properties can be easily verified using the definitions. For instance, to establish the distributive property (i) $t(A+B) = tA + tA$, write $A = [a_{ij}]$ and $B = [b_{ij}]$, for $1 \leqslant i \leqslant m$ and $1 \leqslant j \leqslant n$; then,

$$\begin{aligned} t(A+B) &= [t(a_{ij}+b_{ij})] \\ &= [ta_{ij}+tb_{ij}] \\ &= [ta_{ij}] + [tb_{ij}] \\ &= tA + tB. \end{aligned}$$

The properties of matrix addition and scalar multiplication are analogous to those for vector addition and scalar multiplication in Euclidean space, and they make $\mathbb{M}(m,n)$ into a vector space or linear space. Thus, we can talk about spans of sets of matrices and whether a given set of matrices is linearly independent or not.

Example 3.1.4. Consider the 2×2 matrices

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

Denote them by A_1, A_2, A_4 and A_4, respectively.

We first show that the set $\{A_1, A_2, A_3, A_4\}$ spans $\mathbb{M}(2,2)$. To see why this is the case, note that for any matrix 2×2 matrix, $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$,

$$\begin{aligned} \begin{pmatrix} a & b \\ c & d \end{pmatrix} &= \begin{pmatrix} a & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & b \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ c & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & d \end{pmatrix} \\ &= a\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + b\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} + c\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} + d\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \end{aligned}$$

so that

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \text{span}\{A_1, A_2, A_3, A_4\}.$$

It then follows that

$$\mathbb{M}(2,2) = \text{span}\{A_1, A_2, A_3, A_4\}.$$

Next, we see that $\{A_1, A_2, A_3, A_4\}$ is linearly independent.

Consider the matrix equation

$$c_1\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + c_2\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} + c_3\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} + c_4\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \tag{3.2}$$

or

$$\begin{pmatrix} c_1 & c_2 \\ c_3 & c_4 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix},$$

which implies that

$$c_1 = c_2 = c_3 = c_4 = 0.$$

Hence, the matrix equation in (3.2) has only the trivial solution. Consequently, the set $\{A_1, A_2, A_3, A_4\}$ is linearly independent.

We therefore have that $\{A_1, A_2, A_3, A_4\}$ is a basis for $\mathbb{M}(2,2)$. Consequently, $\dim(\mathbb{M}(2,2)) = 4$. Furthermore, the coordinate vector of the matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ relative to the basis $B = \{A_1, A_2, A_3, A_4\}$ is

$$\left[\begin{pmatrix} a & b \\ c & d \end{pmatrix}\right]_B = \begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix}.$$

3.2 Matrix Algebra

There is a way to define the product of a matrix $A \in \mathbb{M}(m,n)$ and a matrix $B \in \mathbb{M}(n,k)$ to obtain an $m \times k$ matrix AB. In this section we show how to obtain that product and derive its properties.

3.2.1 The row–column product

We begin with the row–column product, which we have already defined in connection with the Euclidean inner product in Section 2.12.1. Given $R \in \mathbb{M}(1,n)$ and $C \in \mathbb{M}(n,1)$, the product RC is the scalar obtained as follows:

Write $R = [x_1 \quad x_2 \quad \cdots \quad x_n]$ and $C = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$; then,

$$RC = [x_1 \quad x_2 \quad \cdots \quad x_n] \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = x_1y_1 + x_2y_2 + \cdots + x_ny_n,$$

or

$$RC = \sum_{j=1}^{n} x_j y_j.$$

We also saw in Section 2.12.1 that the row–column product satisfies the distributive properties:

(i) $(R_1 + R_2)C = R_1C + R_2C$ for $R_1, R_1 \in \mathbb{M}(1, n)$ and $C \in \mathbb{M}(n, 1)$;

(ii) $R(C_1 + C_2) = RC_1 + RC_2$ for $R \in \mathbb{M}(1, n)$ and $C_1, C_2 \in \mathbb{M}(n, 1)$.

3.2.2 The product of a matrix and a vector

We will now see how to use the row–column product to define the product of a matrix $A \in \mathbb{M}(m, n)$ and a (column) vector, $x \in \mathbb{R}^n$, given by $x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$:

Write

$$A = \begin{pmatrix} R_1 \\ R_2 \\ \vdots \\ R_m \end{pmatrix},$$

where

$$\begin{array}{rcl} R_1 & = & \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \end{pmatrix}, \\ R_2 & = & \begin{pmatrix} a_{21} & a_{22} & \cdots & a_{2n} \end{pmatrix}, \\ & \vdots & \\ R_m & = & \begin{pmatrix} a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}. \end{array}$$

Then, the product Ax is given by

$$Ax = \begin{pmatrix} R_1x \\ R_2x \\ \vdots \\ R_mx \end{pmatrix},$$

where, for each $1 \leqslant i \leqslant m$, R_ix is the row–column product

$$R_ix = \sum_{j=1}^{n} a_{ij}x_j.$$

Thus, the product, Ax, of an $m \times n$ matrix, A, and a (column) vector, x, in $\mathbb{M}(n, 1) = \mathbb{R}^n$ is a (column) vector in $\mathbb{M}(m, 1) = \mathbb{R}^m$.

Example 3.2.1. Let $A = \begin{pmatrix} -1 & 0 & 1 \\ 2 & -1 & 0 \end{pmatrix}$ and $x = \begin{pmatrix} 1 \\ -3 \\ 2 \end{pmatrix}$. Then,

$$Ax = \begin{pmatrix} -1 & 0 & 1 \\ 2 & -1 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ -3 \\ 2 \end{pmatrix} = \begin{pmatrix} 1 \\ 5 \end{pmatrix}.$$

Note that in this example $A \in \mathbb{M}(2, 3)$, $x \in \mathbb{M}(3, 1) = \mathbb{R}^3$ and $Ax \in \mathbb{M}(2, 1) = \mathbb{R}^2$.

3.2.3 Interpretations of the matrix product Ax

Observe that, using the definition of the matrix product Ax, the system of linear equations

$$\left\{\begin{array}{rcl} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1k}x_n & = & b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2k}x_n & = & b_2 \\ \vdots & \vdots & \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mk}x_n & = & b_m, \end{array}\right. \tag{3.3}$$

may be written in matrix form

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

or

$$Ax = b, \tag{3.4}$$

where $A = [a_{ij}] \in \mathbb{M}(m,n)$, $x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n$ and $b = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix} \in \mathbb{R}^m$. We therefore see that there exists a very close connection between matrix algebra and the theory of systems of linear equations. In particular, the system in (3.3) is solvable if and only if the matrix equation in (3.4) has a solution $x \in \mathbb{R}^n$ for the given vector $b \in \mathbb{R}^m$.

Another interpretation of the matrix product Ax is provided by the following observation: Note that the product

$$Ax = \begin{pmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \end{pmatrix},$$

may be re–written as

$$\begin{aligned} Ax &= \begin{pmatrix} a_{11}x_1 \\ a_{21}x_1 \\ \vdots \\ a_{m1}x_1 \end{pmatrix} + \begin{pmatrix} a_{12}x_2 \\ a_{22}x_2 \\ \vdots \\ a_{m2}x_2 \end{pmatrix} + \cdots + \begin{pmatrix} a_{1n}x_n \\ a_{2n}x_n \\ \vdots \\ a_{mn}x_n \end{pmatrix} \\ &= x_1 \begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{pmatrix} + x_2 \begin{pmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{pmatrix} + \cdots + x_n \begin{pmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{pmatrix} \\ &= x_1v_1 + x_2v_2 + \cdots + x_nv_n, \end{aligned}$$

where we have set

$$v_1 = \begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{pmatrix}, \; v_2 = \begin{pmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{pmatrix}, \; \cdots, v_n = \begin{pmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{pmatrix},$$

the columns of the matrix A. Hence, Ax is a linear combination of the columns, $v_1, v_2, \ldots, v_n$, of the matrix A where the coefficients are the coordinates of x relative to the standard basis $\mathcal{E} = \{e_1, e_2, \ldots, e_n\}$ in $\mathbb{R}^n$. We may therefore write

$$\begin{aligned} Ax &= [v_1 \quad v_2 \quad \cdots \quad v_n] \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \\ &= x_1 v_1 + x_2 v_2 + \cdots + x_n v_n. \end{aligned}$$

These observations can be used to derive the following facts about the matrix equation in (3.4).

Proposition 3.2.2 (Connections between matrix algebra and the theory of linear equations)**.** *Write the $m \times n$ matrix A in terms of its columns $v_1, v_2, \ldots, v_n$ in $\mathbb{R}^m$; that is,*

$$A = [v_1 \quad v_2 \quad \cdots \quad v_n].$$

1. *Given $b \in \mathbb{R}^m$, the matrix equation*

   $$Ax = b$$

   has a solution if and only if $b \in span\{v_1, v_2, \ldots, v_n\}$; that is, the matrix equation in (3.4) is solvable if and only if b is in the span of the columns of A.

2. *The homogenous equation*

   $$Ax = \mathbf{0}$$

   has only the trivial solution if and only if the columns of A (namely, $v_1, v_2, \ldots, v_n$) are linearly independent.

3. *If the columns of A are linearly independent and span $\mathbb{R}^m$, then $n = m$; that is, A must be a square matrix.*

3.2.4 The Matrix Product

Given matrices $A \in \mathbb{M}(m, n)$ and $B \in \mathbb{M}(n, k)$, write B it terms of its columns,

$$B = [v_1 \quad v_2 \quad \cdots \quad v_k],$$

where $v_1, v_2, \ldots, v_k$ are (column) vectors in $\mathbb{R}^n$. We define the product AB by

$$AB = A[v_1 \quad v_2 \quad \cdots \quad v_k] = [Av_1 \quad Av_2 \quad \cdots \quad Av_k],$$

where, for each $j \in \{1, 2, \ldots, k\}$,

$$Av_j = \begin{pmatrix} R_1 v_j \\ R_2 v_j \\ \vdots \\ R_m v_j \end{pmatrix},$$

where $R_1, R_2, \ldots, R_m$ are the rows of the matrix A. We therefore have that

$$\begin{aligned} AB &= \begin{pmatrix} R_1 \\ R_2 \\ \vdots \\ R_m \end{pmatrix} [v_1 \quad v_2 \quad \cdots \quad v_k] \\ &= \begin{pmatrix} R_1 v_1 & R_1 v_2 & \cdots & R_1 v_k \\ R_2 v_1 & R_2 v_2 & \cdots & R_2 v_k \\ \vdots & \vdots & \cdots & \vdots \\ R_m v_1 & R_m v_2 & \cdots & R_m v_k \end{pmatrix}. \end{aligned}$$

Thus, if $A \in \mathbb{M}(m, n)$ and $B \in \mathbb{M}(n, k)$, the product AB is the $m \times k$ matrix given by

$$AB = [R_i v_j] \quad 1 \leqslant i \leqslant m, \ 1 \leqslant j \leqslant k,$$

where $R_1, R_2, \ldots, R_m$ are the rows of A and $v_1, v_2, \ldots, v_k$ are the columns of B.

Example 3.2.3. Let $A = \begin{pmatrix} -1 & 0 & 1 \\ 2 & -1 & 0 \end{pmatrix}$ and $B = \begin{pmatrix} 2 & -1 \\ -1 & 2 \\ 0 & 1 \end{pmatrix}$. Then,

$$AB = \begin{pmatrix} -1 & 0 & 1 \\ 2 & -1 & 0 \end{pmatrix} \begin{pmatrix} 2 & -1 \\ -1 & 2 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} -2 & 2 \\ 4 & -4 \end{pmatrix}.$$

Thus, $A \in \mathbb{M}(2, 3)$, $B \in \mathbb{M}(3, 2)$ and $AB \in \mathbb{M}(2, 2)$.

Observe that we can also compute BA to obtain the 3×3 matrix:

$$BA = \begin{pmatrix} 2 & -1 \\ -1 & 2 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} -1 & 0 & 1 \\ 2 & -1 & 0 \end{pmatrix} = \begin{pmatrix} -4 & 1 & 2 \\ 5 & -2 & -1 \\ 2 & -1 & 0 \end{pmatrix}.$$

Thus, in this example, $AB \neq BA$.

The previous example shows that matrix multiplication is not commutative. Even when AB and BA have the same dimensions (e.g., when A and B are square matrices of the same dimension), there is no guarantee that AB and BA will be equal to each other.

Example 3.2.4. Let $A = \begin{pmatrix} -1 & 1 \\ 2 & 3 \end{pmatrix}$ and $B = \begin{pmatrix} 0 & -1 \\ 1 & -2 \end{pmatrix}$. Then,

$$AB = \begin{pmatrix} -1 & 1 \\ 2 & 3 \end{pmatrix} \begin{pmatrix} 0 & -1 \\ 1 & -2 \end{pmatrix} = \begin{pmatrix} 1 & -1 \\ 3 & -8 \end{pmatrix},$$

and

$$BA = \begin{pmatrix} 0 & -1 \\ 1 & -2 \end{pmatrix} \begin{pmatrix} -1 & 1 \\ 2 & 3 \end{pmatrix} = \begin{pmatrix} -2 & -3 \\ -5 & -5 \end{pmatrix}.$$

Hence, $AB \neq BA$.

3.2.5 Properties of Matrix Multiplication

We have already seen that matrix multiplication, when it is defined, is not commutative. It is, however, associative and it distributes with respect to matrix addition, as we will show in this section.

Proposition 3.2.5 (Distributive Properties)**.**

(i) For $A \in \mathbb{M}(m,n)$ and $B, C \in \mathbb{M}(n,k)$,

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

(ii) For $A, B \in \mathbb{M}(m,n)$ and $C \in \mathbb{M}(n,k)$,

$$(A + B)C = AC + BC.$$

Proof of (i): Write

$$A = \begin{pmatrix} R_1 \\ R_2 \\ \vdots \\ R_m \end{pmatrix}, \ B = [v_1 \quad v_2 \quad \cdots \quad v_k], \text{ and } C = [w_1 \quad w_2 \quad \cdots \quad w_k],$$

where $R_1, R_2, \ldots, R_m \in \mathbb{M}(1,n)$ are the rows of A, $v_1, v_2, \ldots, v_k \in \mathbb{R}^n$ are the columns of B, and $w_1, w_2, \ldots, w_k \in \mathbb{R}^n$ are the columns of C. Then, using the distributive property for the row–column product,

$$\begin{aligned} A(B + C) &= [R_i(v_j + w_j)], \quad 1 \leqslant i \leqslant m, \ 1 \leqslant j \leqslant k, \\ &= [R_i v_j + R_i w_j] \quad 1 \leqslant i \leqslant m, \ 1 \leqslant j \leqslant k, \\ &= [R_i v_j] + [R_i w_j] \quad 1 \leqslant i \leqslant m, \ 1 \leqslant j \leqslant k, \\ &= AB + AC, \end{aligned}$$

which was to be shown. □

Given a matrix $A = [a_{ij}] \in \mathbb{M}(m,n)$ and a matrix $B = [b_{j\ell}] \in \mathbb{M}(n,k)$, where $1 \leqslant i \leqslant m$, $1 \leqslant j \leqslant n$ and $1 \leqslant \ell \leqslant k$, we have seen that

$$AB = [R_i v_\ell] \quad 1 \leqslant i \leqslant m,\ 1 \leqslant \ell \leqslant k,$$

where $R_1, R_2, \ldots, R_m$ are the rows of A and $v_1, v_2, \ldots, v_k$ are the columns of B. Note that, for each i in $\{1, 2, \ldots, m\}$ and each ℓ in $\{1, 2, \ldots, k\}$,

$$R_i = \begin{pmatrix} a_{i1} & a_{i2} & \cdots & a_{in} \end{pmatrix},$$

and

$$v_\ell = \begin{pmatrix} b_{1\ell} \\ b_{2\ell} \\ \vdots \\ a_{n\ell} \end{pmatrix},$$

so that

$$R_i v_\ell = \sum_{j=1}^{n} a_{ij} b_{j\ell}.$$

We can therefore write

$$AB = [d_{i\ell}],$$

where

$$d_{i\ell} = \sum_{j=1}^{n} a_{ij} b_{j\ell}$$

for $1 \leqslant i \leqslant m$ and $1 \leqslant \ell \leqslant k$. We will use this short–hand notation for the matrix product in the proof of the associative property below.

Proposition 3.2.6 (Associative Property). *Let $A \in \mathbb{M}(m,n)$, $B \in \mathbb{M}(n,k)$ and $C \in \mathbb{M}(k,p)$. Then,*

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

Proof: Write $A = [a_{ij}]$, $B = [b_{j\ell}]$ and $C = [c_{\ell r}]$, where $1 \leqslant i \leqslant m$, $1 \leqslant j \leqslant n$, $1 \leqslant \ell \leqslant k$ and $1 \leqslant r \leqslant p$. Then,

$$AB = [d_{i\ell}], \tag{3.5}$$

where

$$d_{i\ell} = \sum_{j=1}^{n} a_{ij} b_{j\ell} \tag{3.6}$$

for $1 \leqslant i \leqslant m$ and $1 \leqslant \ell \leqslant k$, and

$$BC = [e_{jr}], \tag{3.7}$$

where

$$e_{jr} = \sum_{\ell=1}^{k} b_{j\ell} c_{\ell r} \tag{3.8}$$

for $1 \leqslant j \leqslant n$ and $1 \leqslant r \leqslant p$. We then have that

$$A(BC) = [f_{ir}]$$

where

$$f_{ir} = \sum_{j=1}^{n} a_{ij} e_{jr}$$

for $1 \leqslant 1 \leqslant m$ and $1 \leqslant r \leqslant p$, where we have used (3.7).

Thus, using (3.8) and the distributive property for real numbers,

$$\begin{aligned} f_{ir} &= \sum_{j=1}^{n} a_{ij} \left(\sum_{\ell=1}^{k} b_{j\ell} c_{\ell r} \right) \\ &= \sum_{j=1}^{n} \sum_{\ell=1}^{k} a_{ij} b_{j\ell} c_{\ell r}, \end{aligned}$$

where we have distributed a_{ij} in the the second sum. Thus, since interchanging the order of summation does not alter the sum, we get that

$$\begin{aligned} f_{ir} &= \sum_{\ell=1}^{k} \sum_{j=1}^{n} a_{ij} b_{j\ell} c_{\ell r} \\ &= \sum_{\ell=1}^{k} \left(\sum_{j=1}^{n} a_{ij} b_{j\ell} \right) c_{\ell r}, \end{aligned}$$

where we have used the distributive property for real numbers to factor out $c_{\ell r}$ from the second sum. Using (3.6), we then have that

$$f_{ir} = \sum_{\ell=1}^{k} d_{i\ell} c_{\ell r},$$

so

$$A(BC) = [f_{ir}] = \left[\sum_{\ell=1}^{k} d_{i\ell} c_{\ell r} \right] = (AB)C,$$

since

$$AB = [d_{i\ell}] \quad 1 \leqslant i \leqslant m, \ 1 \leqslant \ell \leqslant k,$$

by (3.5). This completes the proof of the associative property for matrix multiplication. □

As a consequence of the associative property of matrix multiplication, we can define the powers, A^n, for $n = 1, 2, 3 \ldots$, of a square matrix $A \in \mathbb{M}(n, n)$, by computing

$$\begin{aligned} A^2 &= AA \\ A^3 &= AAA = A^2A \\ A^4 &= AAAA = A^3A \\ &\vdots \\ A^m &= A^{m-1}A \\ &\vdots \end{aligned}$$

We define the power A^0 to be the $n \times n$ identity matrix $I = [\delta_{ij}]$ defined by

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

for $1 \leqslant i, j \leqslant n$.

We note that two powers, A^m and A^k, of the matrix A commute with each other; that is,

$$A^m A^k = A^k A^m.$$

To see why this is the case, use the associative property of matrix multiplication to show that

$$A^m A^k = A^{m+k},$$

so that

$$A^m A^k = A^{k+m} = A^k A^m.$$

Example 3.2.7. A square matrix, $A = [a_{ij}] \in \mathbb{M}(n, n)$, is said to be a **diagonal** matrix if $a_{ij} = 0$ for all $i \neq j$. Writing $d_i = a_{ii}$ for $i = 1, 2, \ldots, n$, we have that

$$A = \begin{pmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & d_n \end{pmatrix}.$$

Then,

$$A^2 = \begin{pmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & d_n \end{pmatrix} \begin{pmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & d_n \end{pmatrix} = \begin{pmatrix} d_1^2 & 0 & \cdots & 0 \\ 0 & d_2^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & d_n^2 \end{pmatrix}.$$

By induction on m, we then see that

$$A^m = \begin{pmatrix} d_1^m & 0 & \cdots & 0 \\ 0 & d_2^m & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & d_n^m \end{pmatrix} \quad \text{for } m = 1, 2, 3, \ldots$$

3.3 Invertibility

In Section 3.2.3 on page 65 we saw how to use the matrix product to turn the system

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1k}x_n & = & b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2k}x_n & = & b_2 \\ \vdots & \vdots & \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mk}x_n & = & b_m, \end{cases} \tag{3.9}$$

into the matrix equation

$$Ax = b, \tag{3.10}$$

where A is the $m \times n$ matrix

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

and x and b are the vectors

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \quad \text{and} \quad b = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}$$

in $\mathbb{R}^n$ and $\mathbb{R}^m$, respectively. We will see in this section how matrix algebra and the vector space theory that we developed in the study of Euclidean spaces can be used to answer questions regarding the solvability of the system in (3.9), which is equivalent to the matrix equation in (3.10). For instance, suppose we can find a matrix $C \in \mathbb{M}(n, m)$ with the property that

$$AC = I, \tag{3.11}$$

where I denotes the identity matrix in $\mathbb{M}(m, m)$. Then, using the associativity of the matrix product, which we proved in Proposition 3.2.6, we see that

$$A(Cb) = (AC)b = Ib = b,$$

so that $x = Cb$ is a solution to the matrix equation in (3.10). A matrix $C \in \mathbb{M}(n, m)$ with the property that $AC = I$ is called a right–inverse for A.

3.3.1 Right and Left Inverses

Definition 3.3.1 (Right–Inverse)**.** A matrix $A \in \mathbb{M}(m, n)$ is said to have a **right–inverse** if there exists a matrix $C \in \mathbb{M}(n, m)$ with the property that

$$AC = I,$$

where I denotes the identity matrix in $\mathbb{M}(m, m)$.

We have just proved the following

Proposition 3.3.2. Suppose that $A \in \mathbb{M}(m,n)$ has a right–inverse. Then, for any vector $b \in \mathbb{R}^m$, the matrix equation

$$Ax = b$$

has at least one solution.

Example 3.3.3. Let $A = \begin{pmatrix} 2 & -1 & -3 \\ 1 & 1 & 1 \\ 1 & 2 & 3 \end{pmatrix}$. Then, the matrix

$$C = \begin{pmatrix} 1 & -3 & 2 \\ -2 & 9 & -5 \\ 1 & -5 & 3 \end{pmatrix}$$

is a right–inverse for A since $AC = I$, where I is the 3×3 identity matrix. Then, for any $b = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} \in \mathbb{R}^3$,

$$x = Cb = \begin{pmatrix} 1 & -3 & 2 \\ -2 & 9 & -5 \\ 1 & -5 & 3 \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} = \begin{pmatrix} b_1 - 3b_2 + 2b_3 \\ -2b_1 + 9b_2 - 5b_3 \\ b_1 - 5b_2 + 3b_3 \end{pmatrix}$$

is a solution to the equation

$$Ax = b$$

and, therefore, it is a solution to the system

$$\left\{ \begin{array}{rcl} 2x_1 - x_2 - 3x_3 & = & b_1 \\ x_1 + x_2 + x_3 & = & b_2 \\ x_1 + 2x_2 + 3x_3 & = & b_3, \end{array} \right.$$

for any scalars b_1, b_2 and b_3.

We now turn to the question: When does the equation $Ax = b$ have only one solution?

Definition 3.3.4 (Left–Inverse)**.** A matrix $A \in \mathbb{M}(m,n)$ is said to have a **left–inverse** if there exists a matrix $B \in \mathbb{M}(n,m)$ with the property that

$$BA = I,$$

where I denotes the identity matrix in $\mathbb{M}(n,n)$.

Proposition 3.3.5. Suppose that $A \in \mathbb{M}(m,n)$ has a left–inverse. Then, for any vector $b \in \mathbb{R}^m$, the matrix equation

$$Ax = b$$

can have at most one solution

Proof: Assume that A has a left–inverse, B, then $BA = I$.

Suppose that $v, w \in \mathbb{R}^n$ are two solutions to the equation $Ax = b$. It then follows that

$$Av = b \quad \text{and} \quad Aw = b.$$

Consequently,

$$Av = Aw.$$

Thus,

$$Av - Aw = \mathbf{0}.$$

Using the distributive property for matrix multiplication proved in Proposition 3.2.5 we then obtain that

$$A(v - w) = \mathbf{0}.$$

Multiply on both sides by B we obtain that

$$B[A(v - w)] = B\mathbf{0},$$

so that, by the associative property of the matrix product,

$$(BA)(v - w) = \mathbf{0},$$

or

$$I(v - w) = \mathbf{0}.$$

We therefore get that $v - w = \mathbf{0}$, or $v = w$. Hence, $Ax = b$ can have at most one solution. □

Corollary 3.3.6. Suppose that $A \in \mathbb{M}(m, n)$ has a left–inverse. Then, the columns of A are linearly independent.

Proof: Assume that A has a left–inverse and write $A = [\begin{pmatrix} v_1 & v_2 & \cdots & v_n \end{pmatrix}]$, where $v_1, v_2, \ldots, v_n \in \mathbb{R}^m$ are the columns of A, and suppose that $\begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix}$ is a solution to the vector equation

$$c_1 v_1 + c_2 v_2 + \cdots + c_n v_n = \mathbf{0},$$

which can be written in matrix form as

$$A \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix} = \mathbf{0}.$$

Thus, $\begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix}$ is a solution to

$$Ax = \mathbf{0}. \tag{3.12}$$

Since, A has a left–inverse, it follows from Proposition 3.3.5 that the equation in (3.12) has at most one solution. Observe that the $x = \mathbf{0}$ is already a solution of (3.12). Consequently,

$$c_1 = c_2 = \cdots = c_n = 0,$$

and therefore the set $\{v_1, v_2, \ldots, v_n\}$ is linearly independent. □

Theorem 3.3.7. Let $A \in \mathbb{M}(m, n)$ have a left–inverse, B, and a right–inverse, C. Then, $m = n$ and $B = C$.

Proof: Assume that $A \in \mathbb{M}(m, n)$ has a left–inverse, B, and a right–inverse, C. By Corollary 3.3.6, the columns of A are linearly independent. Denote the columns of A by $v_1, v_2, \ldots, v_n$. We show that $\{v_1, v_2, \ldots, v_n\}$ spans $\mathbb{R}^m$. To see why this is so, let $b \in \mathbb{R}^m$ and consider the equation

$$Ax = b. \tag{3.13}$$

Since A has a right inverse, it follows from Proposition 3.3.2 that equation (3.13) has a solution. Thus, there exist scalars $x_1, x_2, \ldots, x_n$ such that

$$x_1 v_1 + x_2 v_2 + \cdots + x_n v_n = b,$$

so that $b \in \text{span}\{v_1, v_2, \ldots, v_n\}$.

We have shown that $\{v_1, v_2, \ldots, v_n\}$ is linearly independent and spans $\mathbb{R}^m$. Hence, it is a basis for $\mathbb{R}^m$ and therefore $n = m$, since $\dim(\mathbb{R}^m) = m$.

Next, multiply $BA = I$ by C on the left to get

$$(BA)C = IC$$

or, by the associative property,

$$B(AC) = C,$$

which implies that $BI = C$ or $B = C$. □

3.3.2 Definition of Inverse

Theorem 3.3.7 is the basis for the following definition of invertibility for a square matrix.

Definition 3.3.8. A square matrix, $A \in \mathbb{M}(n, n)$, is said to be invertible is there exists a matrix $B \in \mathbb{M}(n, n)$ such that

$$BA = AB = I,$$

where I denotes the $n \times n$ identity matrix.

As a consequence of Theorem 3.3.7 we get the following

Proposition 3.3.9. Let $A \in \mathbb{M}(n,n)$ and suppose that there exists a matrix $B \in \mathbb{M}(n,n)$ such that

$$BA = AB = I,$$

where I denotes the $n \times n$ identity matrix. Then, if $C \in \mathbb{M}(n,n)$ is such that

$$CA = AC = I,$$

then $C = B$.

Hence, if $A \in \mathbb{M}(n,n)$ is invertible, then there exists a unique matrix $B \in \mathbb{M}(n,n)$ such that

$$BA = AB = I.$$

Definition 3.3.10. If $A \in \mathbb{M}(n,n)$ is invertible, then the unique matrix $B \in \mathbb{M}(n,n)$ such that

$$BA = AB = I$$

is called the **inverse** of A and is denoted by A^{-1}.

Example 3.3.11. Suppose that $A \in \mathbb{M}(n,n)$ is invertible. Then, A^{-1} is also invertible and

$$\left(A^{-1}\right)^{-1} = A.$$

To see why this is so, simply observe that, from

$$A^{-1}A = AA^{-1} = I,$$

A is both a right–inverse and a left–inverse of A^{-1}.

3.3.3 Constructing Inverses

In Example 3.3.3 we saw that $C = \begin{pmatrix} 1 & -3 & 2 \\ -2 & 9 & -5 \\ 1 & -5 & 3 \end{pmatrix}$ is a right–inverse of the matrix $A = \begin{pmatrix} 2 & -1 & -3 \\ 1 & 1 & 1 \\ 1 & 2 & 3 \end{pmatrix}$. We can also compute $CA = I$, so that C is also a left–inverse of A and therefore A is invertible with inverse $A^{-1} = C$. In this section we present an algorithm based on elementary row operations which can be used to determine whether a given square matrix is invertible or not and to compute its inverse, if it is invertible.

Before we proceed any further, let's establish the following lemma which is very useful when looking for inverses.

Lemma 3.3.12. If $A \in \mathbb{M}(n,n)$ has a left inverse B, then A is invertible and $A^{-1} = B$.

Proof: Assume that $A \in \mathbb{M}(n,n)$ has a left inverse B. By Corollary 3.3.6, the columns of A form a linearly independent subset, $\{v_1, v_2, \ldots, v_n\}$, of $\mathbb{R}^n$. Hence, since $\dim(\mathbb{R}^n) = n$, it follows that $\{v_1, v_2, \ldots, v_n\}$ is a basis for $\mathbb{R}^n$ and, therefore, $\{v_1, v_2, \ldots, v_n\}$ spans $\mathbb{R}^n$. Consequently, any vector in $\mathbb{R}^n$ is a linear combination of the vectors in $\{v_1, v_2, \ldots, v_n\}$. In particular, there exist c_{ij}, for $1 \leqslant i, j \leqslant n$, such that

$$\begin{array}{rcl} c_{11}v_1 + c_{21}v_2 + \cdots + c_{n1}v_n & = & e_1 \\ c_{12}v_1 + c_{22}v_2 + \cdots + c_{n2}v_n & = & e_2 \\ \vdots & \vdots & \vdots \\ c_{1n}v_1 + c_{2n}v_2 + \cdots + c_{nn}v_n & = & e_1, \end{array}$$

where $\{e_1, e_2, \cdots, e_n\}$ is the standard basis is $\mathbb{R}^n$. We then get that

$$A \begin{pmatrix} c_{1,j} \\ c_{2j} \\ \vdots \\ c_{nj} \end{pmatrix} = e_j$$

for $j = 1, 2, \ldots, n$. Consequently, if we set $C = [c_{ij}]$ for $1 \leqslant i, j \leqslant n$, we see that

$$AC_j = e_j,$$

where C_j is the j^{th} column of C; in other words

$$AC = [AC_1 \quad AC_2 \quad \cdots \quad AC_n] = [e_1 \quad e_2 \quad \cdots \quad e_n] = I.$$

We have therefore shown that A has right–inverse, C. Thus, A has both a right and a left inverse, which shows that A is invertible and therefore $A^{-1} = B$. □

It is also possible to prove that, if A has a right–inverse, then A is invertible.

Proposition 3.3.13. If $A \in \mathbb{M}(n,n)$ has a right–inverse, C, then A is invertible and $A^{-1} = C$.

Proof: Assume $A \in \mathbb{M}(n,n)$ has a right–inverse, $C \in \mathbb{M}(n,n)$; then

$$AC = I. \tag{3.14}$$

Taking transpose on both sides of (3.14) yields

$$C^T A^T = I, \tag{3.15}$$

where we have used the result of Problem 3 in Assignment #15. It follows from (3.15) that A^T has a left–inverse. Thus, applying Lemma 3.3.12, A^T is invertible with inverse $(A^T)^{-1} = C^T$. Finally, applying the result of Problem 5 in Assignment #16, we obtain that $A = (A^T)^T$ is invertible with

$$A^{-1} = [(A^T)^{-1}]^T = (C^T)^T = C,$$

which was to be shown. □

Corollary 3.3.14. Let $A \in \mathbb{M}(n,n)$. If the columns of A are linearly independent, then A is invertible.

Proof: Write $A = [v_1 \quad v_2 \quad \cdots \quad v_n]$, where $v_1, v_2, \ldots, v_n \in \mathbb{R}^n$ are the columns of A. Assume that the set $\{v_1, v_2, \ldots, v_n\}$ is linearly independent; then, since $\dim(\mathbb{R}^n) = n$, $\{v_1, v_2, \ldots, v_n\}$ forms a basis for $\mathbb{R}^n$. In particular, $\{v_1, v_2, \ldots, v_n\}$ spans $\mathbb{R}^n$ so that, for any $b \in \mathbb{R}^n$, the equation

$$Ax = b$$

has a solution in $\mathbb{R}^n$. Applying this result to the equations

$$Ax = e_j, \qquad \text{for } j = 1, 2, \ldots, n,$$

where $\{e_1, e_2, \ldots, e_n\}$ is the standard basis in $\mathbb{R}^n$, we obtain vectors $w_1, w_2, \ldots, w_n \in \mathbb{R}^n$ such that

$$Aw_j = e_j, \qquad \text{for } j = 1, 2, \ldots, n. \tag{3.16}$$

Set $C = [w_1 \quad w_2 \quad \cdots \quad w_n]$; then

$$\begin{aligned} AC &= [Aw_1 \quad Aw_2 \quad \cdots \quad Aw_n] \\ &= [e_1 \quad e_2 \quad \cdots \quad e_n] \\ &= I, \end{aligned}$$

where we have used (3.16). It follows that A has a right–inverse. Consequently, by Proposition 3.3.13, A is invertible. □

Next, we introducing the concept of an **elementary matrix**.

Definition 3.3.15 (Elementary Matrix)**.** A matrix, $E \in \mathbb{M}(n,n)$, which is obtained from the $n \times n$ identity matrix, I, by performing a single elementary row operation on I is called an **elementary matrix**.

Example 3.3.16. Start with the 3×3 identity matrix $I = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ and perform the elementary row operation $cR_1 + R_3 \to R_3$ to obtain

$$E = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ c & 0 & 1 \end{pmatrix}.$$

Observe that if we multiply any 3×3 matrix A on the left by the matrix E

in Example 3.3.16 we obtain

$$\begin{aligned} EA &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ c & 0 & 1 \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \\ &= \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ ca_{11} + a_{31} & ca_{12} + a_{32} & ca_{13} + a_{33} \end{pmatrix} \\ &= \begin{pmatrix} R_1 \\ R_2 \\ cR_1 + R_3 \end{pmatrix}, \end{aligned}$$

where R_1, R_2 and R_3 denote the rows of A. Hence, the effect of multiplying A by E on the left it to perform the same elementary row operation on A that was used on I to obtain E. This is true of all elementary matrices.

Note that we can revert from E to the identity by performing the elementary row operation $-cR_1+R_3$. This is equivalent to multiplying E by the elementary matrix

$$F = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -c & 0 & 1 \end{pmatrix}.$$

We then get that

$$FE = I,$$

and therefore, by Lemma 3.3.12, E is invertible with $E^{-1} = F$. This is also true for all elementary matrices; that is, any elementary matrix is invertible and its inverse is an elementary matrix.

We summarize the previous two observations about elementary matrices in the following

Proposition 3.3.17. Let $E \in \mathbb{M}(m,m)$ denote an elementary matrix.

(i) For any matrix $A \in \mathbb{M}(m,n)$, EA yields a matrix resulting from A by performing on A the same elementary row operation which led from $I \in \mathbb{M}(m,m)$ to E.

(ii) E is invertible and its inverse is also an elementary matrix.

Definition 3.3.18 (Row Equivalence)**.** A matrix $A \in \mathbb{M}(m,n)$ is said to be **row equivalent** to a matrix $B \in \mathbb{M}(m,n)$ if there exist elementary matrices, $E_1, E_2, \ldots, E_k \in \mathbb{M}(m,m)$ such that

$$E_k E_{k-1} \cdots E_2 E_1 A = B.$$

The most important example of row equivalence for this section is the case in which and $n \times n$ matrix, A, is row equivalent to the identity $I \in \mathbb{M}(n,n)$.

Proposition 3.3.19. If $A \in \mathbb{M}(n,n)$ is row equivalent to the identity, then A is invertible and

$$A^{-1} = E_k E_{k-1} \cdots E_2 E_1,$$

where $E_1, E_2, \ldots, E_k$ are $n \times n$ elementary matrices.

Proof: Assume that $A \in \mathbb{M}(n,n)$ is row equivalent to the identity $I \in \mathbb{M}(n,n)$. Then, there exist elementary matrices, $E_1, E_2, \ldots, E_k \in \mathbb{M}(n,n)$ such that

$$E_k E_{k-1} \cdots E_2 E_1 A = I,$$

or

$$(E_k E_{k-1} \cdots E_2 E_1) A = I.$$

It then follows from Lemma 3.3.12 that A is invertible and

$$A^{-1} = E_k E_{k-1} \cdots E_2 E_1.$$

□

Thus, if A is invertible, to find its inverse, all we need to do is find a sequence of elementary matrices $E_1, E_2, \ldots, E_k \in \mathbb{M}(n,n)$ such that

$$E_k E_{k-1} \cdots E_2 E_1 A = I.$$

Since multiplying by an elementary matrix on the left is equivalent to performing an elementary row operation on the matrix, $E_k E_{k-1} \cdots E_2 E_1 A$ is the result of performing k successive elementary row operations on the matrix A. The product $E_k E_{k-1} \cdots E_2 E_1$ keeps track of those operations. This can also be done by performing elementary row operations on the augmented matrix

$$[\ A \quad | \quad I\ ]. \tag{3.17}$$

Performing the first elementary row operation on the matrix in (3.17) yields

$$[\ E_1 A \quad | \quad E_1 I\ ],$$

or

$$[\ E_1 A \quad | \quad E_1\ ].$$

Performing the second elementary row operation on the augmented matrix in (3.17) then yields

$$[\ E_1 E_1 A \quad | \quad E_2 E_1\ ].$$

Continuing in this fashion we obtain

$$[\ E_k E_{k-1} \cdots E_1 E_1 A \quad | \quad E_k E_{k-1} \cdots E_2 E_1\ ],$$

or

$$[\ I \quad | \quad A^{-1}\ ]. \tag{3.18}$$

Hence, if after performing elementary row operations on the augmented matrix in (3.17) we obtain the augmented matrix in (3.18), we can conclude that A is invertible and and its inverse is the matrix obtained in the right–hand side of the augmented matrix in (3.18).

Example 3.3.20. Use Gaussian elimination to compute the inverse of the matrix

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

Solution: Begin with the augmented matrix

$$\left(\begin{array}{ccc|ccc} 2 & -1 & -3 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 2 & 3 & 0 & 0 & 1 \end{array}\right). \tag{3.19}$$

Then, perform the elementary row operations $R_1 \leftrightarrow R_2$, $-2R_1 + R_2 \to R_2$ and $-R_1 + R_3 \to R_3$ in succession to turn the matrix in (3.19) into

$$\left(\begin{array}{ccc|ccc} 1 & 1 & 1 & 0 & 1 & 0 \\ 0 & -3 & -5 & 1 & -2 & 0 \\ 0 & 1 & 2 & 0 & -1 & 1 \end{array}\right). \tag{3.20}$$

Next, perform on the augmented matrix in (3.20) the elementary row operations $R_2 \leftrightarrow R_3$ and $3R_2 + R_3 \to R_3$ in succession to get

$$\left(\begin{array}{ccc|ccc} 1 & 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 2 & 0 & -1 & 1 \\ 0 & 0 & 1 & 1 & -5 & 3 \end{array}\right). \tag{3.21}$$

Finally, perform the elementary row operations $-2R_3 + R_2 \to R_2$, $-R_3 + R_1 \to R_1$ and $-R_2 + R_1 \to R_1$ in succession to obtain from (3.21) the augmented matrix

$$\left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & -3 & 2 \\ 0 & 1 & 0 & -2 & 9 & -5 \\ 0 & 0 & 1 & 1 & -5 & 3 \end{array}\right). \tag{3.22}$$

We then read from (3.22) that

$$A^{-1} = \begin{pmatrix} 1 & -3 & 2 \\ -2 & 9 & -5 \\ 1 & -5 & 3 \end{pmatrix}$$

□

It follows from Proposition 3.3.19 and the fact that the inverse of an elementary matrix is also an elementary matrix that every invertible matrix is the product of elementary matrices. Indeed, if A is an invertible $n \times n$ matrix, then, by virtue of Proposition 3.3.19,

$$A^{-1} = E_k E_{k-1} \cdots E_2 E_1, \tag{3.23}$$

where $E_1, E_2, \ldots, E_k$ are $n \times n$ elementary matrices. Thus, taking inverses on both sides of (3.23),

$$A = E_1^{-1} E_2^{-1} \cdots E_k^{-1}.$$

We have therefore proved the following proposition.

Proposition 3.3.21. Every invertible $n \times n$ matrix is a product of elementary matrices.

3.4 Nullity and Rank

Given an $m \times n$ matrix, A, we can define the following subspaces associated with A:

1. **The column cpace of** A, denoted by $\mathcal{C}_A$ is the subspace of $\mathbb{R}^m$ defined as the span of the columns of A; that is, if $A = [v_1 \quad v_2 \quad \cdots \quad v_n]$, then

$$\mathcal{C}_A = \text{span}\{v_1, v_2, \ldots, v_n\}.$$

Example 3.4.1. Let A denote the matrix

$$\begin{pmatrix} 1 & 3 & -1 & 0 \\ 2 & 2 & 2 & 4 \\ 1 & 0 & 2 & 3 \end{pmatrix}. \tag{3.24}$$

Then, $\mathcal{C}_A$ is the subspace of $\mathbb{R}^3$ given by

$$\mathcal{C}_A = \text{span}\left\{ \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 3 \\ 2 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ 2 \\ 2 \end{pmatrix}, \begin{pmatrix} 0 \\ 4 \\ 3 \end{pmatrix} \right\}.$$

We saw in Problem 2 of Assignment #9 that the set

$$\left\{ \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 3 \\ 2 \\ 0 \end{pmatrix} \right\}$$

is a basis for $\mathcal{C}_A$. Hence, $\dim(\mathcal{C}_A) = 2$.

Definition 3.4.2 (Column Rank)**.** Given an $m \times n$ matrix, A, the dimension of $\mathcal{C}$ is called the **column rank** of the matrix A. In these notes, we will denote the row rank of A by $c(A)$; thus,

$$c(A) = \dim(\mathcal{C}_A).$$

Observe that, since $\mathcal{C}_A$ is a subspace of $\mathbb{R}^m$,

$$c(A) \leqslant m.$$

2. **The row space of** A, denoted by $\mathcal{R}_A$, is the subspace of $\mathbb{M}(1, n)$ spanned by the rows of A. If we let $R_1, R_2, \ldots, R_m$ denote the rows of A, then

$$\mathcal{R}_A = \text{span}\{R_1, R_2, \ldots, R_m\}.$$

The dimension of the row space of A is called the **row rank** of A and we will denote it by $r(A)$. We then have that

$$r(A) \leqslant n,$$

since $\mathcal{R}_A$ is a subspace of $\mathbb{M}(1,n)$ and $\dim(\mathbb{M}(1,n)) = n$.

Example 3.4.3. Let A denote the matrix in Example 3.4.1 given in (3.24). We would like to compute the row rank of A. In order to do this we need to find a basis for the span of the rows of A. Denote the rows of A by R_1, R_2 and R_3. We can find a linearly independent subset of $\{R_1, R_2, R_3\}$ which also spans $\mathcal{R}_A$ by performing elementary row operations on the matrix A and keeping track of them as follows: Start with the matrix

$$\begin{pmatrix} R_1 \\ R_2 \\ R_3 \end{pmatrix} \qquad \begin{pmatrix} 1 & 3 & -1 & 0 \\ 2 & 2 & 2 & 4 \\ 1 & 0 & 2 & 3 \end{pmatrix}, \tag{3.25}$$

where we are keeping track of the operations on the left–hand side of (3.25). Performing $-2R_1 + R_2 \to R_2$ and $-R_1 + R_3 \to R_3$ in succession on the matrix in (3.25) and keeping track of the results of the operations on the left of the matrix in (3.25) yields

$$\begin{pmatrix} R_1 \\ -2R_1 + R_2 \\ -R_1 + R_3 \end{pmatrix} \qquad \begin{pmatrix} 1 & 3 & -1 & 0 \\ 0 & -4 & 4 & 4 \\ 0 & -3 & 3 & 3 \end{pmatrix}. \tag{3.26}$$

Next, perform the operations $-\frac{1}{4}R_2 \to R_2$ and $3R_2 + R_3 \to R_3$ in succession to the matrices in (3.26) to get

$$\begin{pmatrix} R_1 \\ \frac{1}{2}R_1 - \frac{1}{4}R_2 \\ -\frac{1}{2}R_1 - \frac{3}{4}R_2 + R_3 \end{pmatrix} \qquad \begin{pmatrix} 1 & 3 & -1 & 0 \\ 0 & 1 & -1 & -1 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \tag{3.27}$$

We then get from the matrices in (3.27) that

$$-\frac{1}{2}R_1 - \frac{3}{4}R_2 + R_3 = O,$$

where O denotes the zero matrix in $\mathbb{M}(1,4)$. Hence,

$$R_3 = \frac{1}{2}R_1 + \frac{3}{4}R_2,$$

which shows that $R_3 \in \operatorname{span}\{R_1, R_2\}$ and therefore

$$\operatorname{span}\{R_1, R_2, R_3\} = \operatorname{span}\{R_1, R_2\}.$$

Since R_1 and R_2 are clearly not multiple of each other, it follows that $\{R_1, R_2\}$ is linearly independent and therefore it is a basis for $\mathcal{R}_A$. It then follows that $r(A) = \dim(\mathcal{R}_A) = 2$.

3. **The null space of** A, denoted by $\mathcal{N}_A$ is the subset of $\mathbb{R}^n$ defined by

$$\mathcal{N}_A = \{v \in \mathbb{R}^n \mid Av = 0\}.$$

$\mathcal{N}_A$ is a subspace of $\mathbb{R}^n$. In order to see why this is so, first observe that $\mathcal{N}_A \neq \emptyset$ since $\mathbf{0} \in \mathcal{N}_A$ because $A\mathbf{0} = \mathbf{0}$. Next, suppose that $v, w \in \mathcal{N}_A$; then

$$Av = \mathbf{0} \quad \text{and} \quad Aw = \mathbf{0}.$$

It then follows from the distributive property for matrix multiplication that

$$A(v + w) = Av + Aw = \mathbf{0} + \mathbf{0} = \mathbf{0},$$

and so $v + w \in \mathcal{N}_A$; thus, $\mathcal{N}_A$ is closed under vector addition. Finally, note that for any $v \in \mathbb{R}^n$ and $c \in \mathbb{R}$,

$$\begin{aligned} A(cv) &= \begin{pmatrix} R_1 \\ R_2 \\ \vdots \\ R_m \end{pmatrix} (cv) \\ &= \begin{pmatrix} R_1(cv) \\ R_2(cv) \\ \vdots \\ R_m(cv) \end{pmatrix} \\ &= \begin{pmatrix} \langle R_1^T, cv\rangle \\ \langle R_2^T, cv\rangle \\ \vdots \\ \langle R_m^T, cv\rangle \end{pmatrix}, \end{aligned}$$

Where $\langle \cdot, \cdot \rangle$ denotes the Euclidean inner product in $\mathbb{R}^n$. It then follows from the bilinearity of the inner product that

$$\begin{aligned} A(cv) &= \begin{pmatrix} c\langle R_1^T, v\rangle \\ c\langle R_2^T, v\rangle \\ \vdots \\ c\langle R_m^T, v\rangle \end{pmatrix} \\ &= c \begin{pmatrix} \langle R_1^T, v\rangle \\ \langle R_2^T, v\rangle \\ \vdots \\ \langle R_m^T, v\rangle \end{pmatrix} \\ &= cAv. \end{aligned}$$

Thus, if $v \in \mathcal{N}_A$, then

$$A(cv) = cA(v) = c\mathbf{0} = \mathbf{0},$$

which shows that $cv \in \mathcal{N}_A$ and therefore $\mathcal{N}_A$ is closed under scalar multiplication.

Example 3.4.4. Let A denote the matrix in Example 3.4.1 given in (3.24). To compute the null space of A, we find the solution space of the system

$$Ax = \mathbf{0},$$

or

$$\begin{cases} x_1 + 3x_2 - x_3 & = & 0 \\ 2x_1 + 2x_2 + 2x_3 + 4x_4 & = & 0 \\ x_1 + 2x_3 + 3x_4 & = & 0. \end{cases} \tag{3.28}$$

We can use Gauss–Jordan reduction to turn the system in (3.28) into the equivalent system

$$\begin{cases} x_1 + 2x_3 + 3x_4 & = & 0 \\ \quad x_2 - x_3 - x_4 & = & 0, \end{cases} \tag{3.29}$$

which can be solved to yield

$$\mathcal{N}_A = \operatorname{span}\left\{ \begin{pmatrix} 2 \\ -1 \\ -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 3 \\ -1 \\ 0 \\ -1 \end{pmatrix} \right\}.$$

Thus, the set

$$\left\{ \begin{pmatrix} 2 \\ -1 \\ -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 3 \\ -1 \\ 0 \\ -1 \end{pmatrix} \right\}.$$

is a basis for $\mathcal{N}_A$, and therefore $\dim(\mathcal{N}_A) = 2$.

Given $A \in \mathbb{M}(m,n)$, the dimension of the null space, $\mathcal{N}_A$, of A is called the **nullity** of A and we will denote it by $n(A)$. We then have that

$$n(A) = \dim(\mathcal{N}_A).$$

Observe that, an $m \times n$ matrix A, since $\mathcal{N}_A$ is a subspace of $\mathbb{R}^n$, it follows that

$$n(A) \leqslant n.$$

In the previous example we showed that for the 3×4 matrix A given in (3.24), the nullity of A is $n(A) = 2$.

The main goal of this section is to prove the following facts about the row rank, the column rank and the nullity of an $m \times n$ matrix A:

Theorem 3.4.5. Let $A \in \mathbb{M}(m,n)$. Then,

(i) *(Equality of row rank and column rank)*

$$r(A) = c(A);$$

and

(ii) *(Dimension Theorem for Matrices)*

$$n(A) + r(A) = n.$$

We will therefore call the dimension of the column space of A simply the **rank** of A and denote it by $r(A)$.

We will present here a proof of the equality of the row rank and the column rank based on an argument given by Mackiw in [Mac95, pp. 285–286]). We first prove the following

Lemma 3.4.6. Let $A \in \mathbb{M}(m,n)$ and denote the row space of A by $\mathcal{R}_A$. Define

$$\mathcal{R}_A^\perp = \{w \in \mathbb{R}^n \mid R_i w = 0 \text{ for } i = 1, 2, \ldots, m\},$$

where $R_1, R_2, \ldots, R_m$ denote the rows of the matrix A; i.e., $\mathcal{R}_A^\perp$ is the set of vectors in $\mathbb{R}^n$ which are orthogonal to the vectors $R_1^T, R_2^T, \ldots, R_m^T$ in $\mathbb{R}^n$. Then,

(i) $\mathcal{R}_A^\perp = \mathcal{N}_A$, and

(ii) if $w \in \mathcal{N}_A$ and $w^T \in \mathcal{R}_A$, then $w = \mathbf{0}$.

Proof of (i): Observe that $w \in \mathcal{N}_A$ if and only if $Aw = \mathbf{0}$, or

$$\begin{pmatrix} R_1 w \\ R_2 w \\ \vdots \\ R_m w \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Hence, $w \in \mathcal{N}_A$ if and only if $R_i w = 0$ for $i = 1, 2, \ldots, m$. This is equivalent to $\mathcal{N}_A = \mathcal{R}_A^\perp$. □

Proof of (ii): Assume that $w \in \mathcal{N}_A$ and $w^T \in \mathcal{R}_A^\perp$. Then, by the result of part (1), $w \in \mathcal{R}_A^\perp$, which implies that $v^T w = 0$ for all $v^T \in \mathcal{R}_A$. Thus, in particular, $w^T w = 0$, or $\langle w, w \rangle = 0$, which implies that $w = \mathbf{0}$, by the positive definiteness of the Euclidean inner product. □

Proof of the equality of the row and columns ranks: Let $r(A) = k$. Then, there exist $w_1, w_2, \ldots, w_k$ in $\mathbb{R}^n$ such that $\{w_1^T, w_2^T, \ldots, w_k^T\}$ is a basis for $\mathcal{R}_A$.

Consider the set $\{Aw_1, Aw_2, \ldots, Aw_k\}$, which is a subset of $\mathbb{R}^m$. We first observe that

$$\{Aw_1, Aw_2, \ldots, Aw_k\} \subseteq \mathcal{C}_A. \tag{3.30}$$

To see why this is the case, write w_j, for $j = 1, 2, \ldots, k$, in terms of the standard basis $\{e_1, e_2, \ldots, e_n\}$ for $\mathbb{R}^n$:

$$w_j = c_{1j}e_1 + c_{2j}e_2 + \cdots + c_{nj}e_n,$$

and apply A to get

$$\begin{aligned} Aw_j &= c_{1j}e_1 + c_{2j}e_2 + \cdots + c_{nj}e_n \\ &= A(c_{1j}e_1 + c_{2j}e_2 + \cdots + c_{nj}e_n) \\ &= A(c_{1j}e_1) + A(c_{2j}e_2) + \cdots + A(c_{nj}e_n) \\ &= c_{1j}Ae_1 + c_{2j}Ae_2 + \cdots + c_{nj}Ae_n, \end{aligned}$$

where we have used the distributive property of matrix multiplication and the fact that $A(cv) = cAv$ for all scalars c and all vectors $v \in \mathbb{R}^n$. Noting that $Ae_1, Ae_2, \ldots, Ae_n$ are the columns of A, we see that (3.30) follows.

Next, we show that $\{Aw_1, Aw_2, \ldots, Aw_k\}$ is linearly independent. To prove this, suppose that $c_1, c_2, \ldots, c_k$ is a solution of the vector equation

$$c_1Aw_1 + c_1Aw_2 + \cdots + Aw_k = \mathbf{0}. \tag{3.31}$$

Then, using the distributive property of the matrix product and the fact that $A(cv) = cAv$ for all scalars c and all vectors $v \in \mathbb{R}^n$, we get from (3.31) that

$$A(c_1w_1 + c_1w_2 + \cdots + w_k) = \mathbf{0},$$

which shows that the vector $w = c_1w_1 + c_1w_2 + \cdots + w_k$ is in the null space, $\mathcal{N}_A$, of the matrix A. On the other hand,

$$w^T = c_1w_1^T + c_1w_2^T + \cdots + w_k^T, \tag{3.32}$$

is in $\mathcal{R}_A$, since $\{w_1^T, w_2^T, \ldots, w_k^T\}$ is a basis for $\mathcal{R}_A$. It then follows from part (ii) in Lemma 3.4.6 that $w = \mathbf{0}$. We then get from (3.32) that

$$c_1w_1^T + c_1w_2^T + \cdots + w_k^T = \mathbf{0},$$

which implies that

$$c_1 = c_2 = \cdots = c_k = 0,$$

since the set $\{w_1^T, w_2^T, \ldots, w_k^T\}$ is linearly independent. We have therefore shown that the only solution to the vector equation in (3.31) is the trivial solution, and hence the set $\{Aw_1, Aw_2, \ldots, Aw_k\}$ is linearly independent. It then follows from Lemma 2.10.5 that

$$k \leqslant c(A),$$

or

$$r(A) \leqslant c(A). \tag{3.33}$$

Applying the previous argument to A^T we see that

$$r(A^T) \leqslant c(A^T),$$

which is equivalent to

$$c(A) \leqslant r(A). \tag{3.34}$$

Combining (3.33) and (3.34) proves the equality of the row and column ranks. □

Next, we preset a proof of part (ii) of Theorem 3.4.5, the Dimension Theorem for Matrices. In the proof we will use the following Lemma, which is Theorem 3.13 (the Expansion Theorem) in Messer, [Mes94, pg. 119].

Lemma 3.4.7. Let $\{w_1, w_2, \ldots, w_k\}$ denote a linearly independent subset of $\mathbb{R}^n$. If $k < n$, there exist vectors $v_1, v_2, \ldots, v_\ell$ in $\mathbb{R}^n$ such that

$$\{w_1, w_2, \ldots, w_k, v_1, v_2, \ldots, v_\ell\}$$

is a basis for $\mathbb{R}^n$, where $k + \ell = n$.

Proof: Since $k < n$, $\text{span}\{w_1, w_2, \ldots, w_k\} \neq \mathbb{R}^n$ because $\dim(\mathbb{R}^n) = n$. Hence, there exists $v_1 \in \mathbb{R}^n$ such that $v_1 \notin \text{span}\{w_1, w_2, \ldots, w_k\}$. Consequently, by Lemma 2.8.2 on page 42 in these notes, $\{w_1, w_2, \ldots, w_k, v_1\}$ is linearly independent. If

$$\{w_1, w_2, \ldots, w_k, v_1\}$$

spans $\mathbb{R}^n$, it is a basis for $\mathbb{R}^n$ and the Lemma is proved in this case. If not, there exists $v_2 \in \mathbb{R}^n$ such that $v_2 \notin \text{span}\{w_1, w_2, \ldots, w_k, v_1\}$. Thus, invoking Lemma 2.8.2 again, the set $\{w_1, w_2, \ldots, w_k, v_1, v_2\}$ is linearly independent. If $\{w_1, w_2, \ldots, w_k, v_1, v_2\}$ also spans $\mathbb{R}^n$, the Lemma is proved. If not, we continue as before. We therefore conclude that there exist $v_1, v_2, \ldots, v_\ell$ in $\mathbb{R}^n$ such that

$$\{w_1, w_2, \ldots, w_k, v_1, v_2, \ldots, v_\ell\}$$

is a linearly independent subset such that

$$\text{span}\{w_1, w_2, \ldots, w_k, v_1, v_2, \ldots, v_\ell\} = \mathbb{R}^n.$$

This proves the Lemma. □

Proof of the Dimension Theorem for Matrices: We show that for any $m \times n$ matrix, A,

$$n(A) + r(A) = n, \tag{3.35}$$

where $n(A)$ is the nullity of A and $r(A)$ is the rank of A, which we know to be the same as the dimension of the columns space of A, $\mathcal{C}_A$.

If $n(A) = 0$, then $\mathcal{N}_A = \{\mathbf{0}\}$ and therefore the equation

$$Ax = 0$$

has only the trivial solution and, therefore, the columns of A are linearly independent. Thus, they form a basis for the column space of A and therefore $\dim(\mathbf{C}_A) = n$; that is, $r(A) = n$ which implies (3.35) for the case $n(A) = 0$.

Thus, assume that $n(A) = k > 0$. Then, since $\mathcal{N}_A$ is a subspace of $\mathbb{R}^n$, $1 \leqslant k \leqslant n$. Let $\{w_1, w_2, \ldots, w_k\}$ denote a basis for $\mathcal{N}_A$. If $k = n$, then $\mathcal{N}_A = \mathbb{R}^n$, since $\dim(\mathbb{R}^n) = n$ and therefore $Ax = 0$ for all $x \in \mathbb{R}^n$ and therefore all the columns of A are the zero vector in $\mathbb{R}^n$, which implies that $\mathcal{C}_A = \{\mathbf{0}\}$; therefore, $\dim(\mathcal{C}_A) = 0$, which shows that $r(A) = 0$ and therefore 3.35) holds true for the case $n(A) = n$.

Next, consider the case $1 \leqslant k < n$. Then, by Lemma 3.4.7, we can find vectors $v_1, v_2, \ldots, v_\ell$ in $\mathbb{R}^n$ such that $\{w_1, w_2, \ldots, w_k, v_1, v_2, \ldots, v_\ell\}$ is a basis for $\mathbb{R}^n$, where

$$k + \ell = n. \tag{3.36}$$

It remains to prove that

$$\ell = \dim(\mathcal{C}_A); \tag{3.37}$$

for, if (3.37) is true, then equation (3.36) implies (3.35) and the Dimension Theorem for Matrices is proved.

In order to prove (3.37), consider the set

$$\mathcal{B} = \{Av_1, Av_2, \ldots, Av_\ell\}.$$

First note that $\mathcal{B}$ is a subset of $\mathcal{C}_A$ since each Av_j, for $j = 1, 2, \ldots, \ell$, is a linear combinations of the columns of A.

We first see that $\mathcal{B}$ spans $\mathcal{C}_A$. To show this, let $w \in \mathcal{C}_A$. Then, w is a linear combination of the columns of A, which implies that $w = Av$ for some $v \in \mathbb{R}^n$. Since the set $\{w_1, w_2, \ldots, w_k, v_1, v_2, \ldots, v_\ell\}$ is basis for $\mathbb{R}^n$, there exist scalars $d_1, d_2, \ldots, d_k, c_1, c_2, \ldots, c_\ell$ such that

$$v = d_1 w_1 + d_2 w_2 + \cdots + d_k w_k + c_1 v_1 + c_2 v_2 + \cdots + c_\ell v_\ell.$$

Then

$$\begin{aligned} w &= Av \\ &= A(d_1 w_1 + d_2 w_2 + \cdots + d_k w_k + c_1 v_1 + c_2 v_2 + \cdots + c_\ell v_\ell \\ &= d_1 A w_1 + d_2 A w_2 + \cdots + d_k A w_k + c_1 A v_1 + c_2 A v_2 + \cdots + c_\ell A v_\ell, \end{aligned}$$

where we have used the distributive property of matrix multiplication and the fact that $A(cv) = cAv$ for all scalars c and all vectors $v \in \mathbb{R}^n$. It then follows, since $w_1, w_2, \ldots, w_k \in \mathcal{N}_A$, that

$$w = c_1 A v_1 + c_2 A v_2 + \cdots + c_\ell A v_\ell,$$

which shows that $w \in \operatorname{span}(\mathcal{B})$ and therefore $\mathcal{C}_A = \operatorname{span}(\mathcal{B})$.

Next, we prove that $\mathcal{B}$ is linearly independent. To see why why this is the case, suppose that $c_1, c_2, \ldots, c_k$ is a solution of the vector equation

$$c_1 A v_1 + c_1 A v_2 + \cdots + A v_\ell = \mathbf{0}. \tag{3.38}$$

Then, using the distributive property of the matrix product and the fact that $A(cv) = cAv$ for all scalars c and all vectors $v \in \mathbb{R}^n$, we get from (3.38) that

$$A(c_1 v_1 + c_1 v_2 + \cdots + v_\ell) = \mathbf{0},$$

which shows that the vector $w = c_1 v_1 + c_1 v_2 + \cdots + v_\ell$ is in the null space, $\mathcal{N}_A$, of the matrix A. Thus, since $\{w_1, w_2, \ldots, w_k\}$ is a basis for $\mathcal{N}_A$, there exist scalars $d_1, d_2, \ldots, d_k$ such that

$$w = d_1 w_1 + d_2 w_2 + \cdots + d_k w_k.$$

It then follows that

$$c_1 v_1 + c_1 v_2 + \cdots + v_\ell = d_1 w_1 + d_2 w_2 + \cdots + d_k w_k,$$

from which we get that

$$(-d_1) w_1 + (-d_2) w_2 + \cdots + (-d_k) w_k + c_1 v_1 + c_1 v_2 + \cdots + v_\ell = \mathbf{0}.$$

We now use the fact that $\{w_1, w_2, \ldots, w_k, v_1, v_2, \ldots, v_\ell\}$ is basis for $\mathbb{R}^n$ to conclude that

$$c_1 = c_2 = \cdots = c_\ell = 0.$$

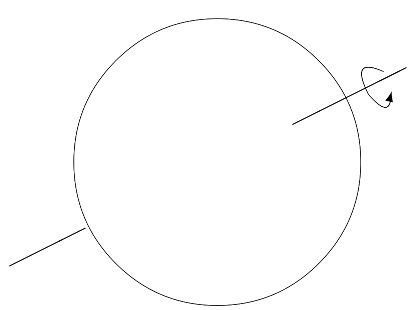
We have therefore shown that the only solution to the vector equation in (3.38) is the trivial solution, and hence the set $\{Av_1, Av_2, \ldots, Av_\ell\}$ is linearly independent. This proves (3.37) and the proof of the Dimension Theorem for Matrices is now complete. □

Chapter 4

Linear Transformations

The main goal of this chapter and the next is solve the problem stated in Chapter 1, which has served as the motivation for theory of vector spaces and matrix algebra that we have developed so far. The problem is simple to state:

Problem 4.0.8 (Euler's Theorem on the Axis of Rotation (see [PPR09])). *Imagine a ball whose center is at a fixed location in three–dimensional space, but is free to rotate about its center around any axis through the center. The center of the ball is not allowed to move away from its fixed location. Imagine that we perform several rotations about various axes, one after the other. We claim that there are two antipodal points on the surface of the ball which are exactly at the same locations they were at the beginning of the process. Furthermore, the combination of all the rotations that we perform has the same affect on the ball as that of a single rotation performed about the axis going through the fixed antipodal points.*



In order to prove the claims stated in Problem 4.0.8, we will first model a rotation in $\mathbb{R}^3$ by a function,

$$R\colon \mathbb{R}^3 \to \mathbb{R}^3,$$

which takes a point v in the ball and yields a point $R(v)$, also in the ball, which locates the point v after the rotation has been performed. Note that (i) every point v in $\mathbb{R}^3$ is mapped to a point $R(v)$ by the rotation ($R(v)$ could be the same point as v; for example, if v is on the axis of rotation, then $R(v) = v$); (ii) no point in $\mathbb{R}^3$ gets mapped to more than one point by the rotation R. Hence, R does indeed defines a function. It is an example of a **vector valued function** defined on an Euclidean space.

4.1 Vector Valued Functions on Euclidean Space

A vector valued function,

$$f\colon \mathbb{R}^n \to \mathbb{R}^m, \tag{4.1}$$

assigns to each vector, v, in $\mathbb{R}^n$ one vector, $f(v)$, in $\mathbb{R}^m$. We have already seen examples of these functions in this course. For instance, the function $f\colon \mathbb{R}^n \to \mathbb{R}$ given by

$$f(v) = \|v\| \quad \text{for all} \quad v \in \mathbb{R}^n,$$

where $\|\cdot\|$ denotes the Euclidean norm in $\mathbb{R}^n$. In this case $m = 1$. Also, for a fixed $w \in \mathbb{R}^n$, define

$$f(v) = \langle w, v\rangle \quad \text{for all} \quad v \in \mathbb{R}^n,$$

where $\langle\cdot,\cdot\rangle$ denotes the Euclidean inner product in $\mathbb{R}^n$; then, f is also a map from $\mathbb{R}^n$ to R.

The set $\mathbb{R}^n$ in (4.1) is called the **domain** of the function f, while $\mathbb{R}^m$ is called the **co-domain** of f.

Definition 4.1.1 (Image)**.** Given a function $f\colon \mathbb{R}^n \to \mathbb{R}^m$ and a subset, S, of $\mathbb{R}^n$, the **image** of S under f is the subset of $\mathbb{R}^m$, denoted by $f(S)$, and defined as follows

$$f(S) = \{w \in \mathbb{R}^m \mid w = f(v) \text{ for some } v \in S\}.$$

In other words, $f(S)$ is the set to which the vectors in S get mapped by the function f.

Example 4.1.2 (Rotations in $\mathbb{R}^2$)**.** Let $R_\theta\colon \mathbb{R}^2 \to \mathbb{R}^2$ denote the function that takes every line through the origin in $\mathbb{R}^2$ and rotates it through an angle of θ is the counterclockwise sense. Figure 4.1.1 shows a typical line through the origin, L, and its image, $R_\theta(L)$ under the rotation R_θ.

Suppose that the line L is generated by a vector $v \neq \mathbf{0}$; that is, $L = \text{span}\{v\}$. The image of v under R_θ is the vector $R_\theta(v)$ in $\mathbb{R}_\theta(L)$. Since a rotation does change the length of vectors, it follows that $\|R_\theta(v)\| = \|v\| \neq 0$. Thus, the vector $R_\theta(v)$ can be used to generate $R_\theta(L)$; that is, $R_\theta(L) = \text{span}\{R_\theta(v)\}$. We

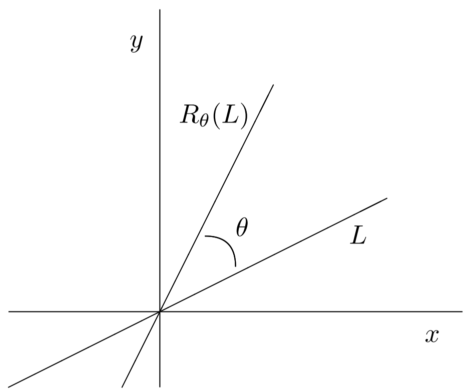


Figure 4.1.1: Image of a line under rotation

then get that for any vector $w \in L$, $w = tv$ for some scalar t, and $R_\theta(tv) \in R_\theta(L)$ so that

$$R_\theta(tv) = sR_\theta(v), \tag{4.2}$$

for some scalar s. Again, since R_θ does not change lengths of vectors, it follows from (4.2) that

$$|t|\|v\| = |s|\|v\|,$$

from which w get that $|t| = |s|$, since $v \neq \mathbf{0}$. Observe also that, for $0 < \theta < \pi$, R_θ does not reverse the orientation the vector v, so that t and s must have the same sign. We therefore conclude that $t = s$ and therefore (4.2) turns into

$$R_\theta(tv) = tR_\theta(v); \tag{4.3}$$

that is R_θ takes a scalar multiple of v to a scalar multiple of $R_\theta(v)$ with the same scaling factor.

Next, consider two linearly independent vectors, v and w, in $\mathbb{R}^2$. The vectors v and w generate a parallelogram defined by

$$P(v, w) = \{tv + sw \mid 0 \leqslant t \leqslant 1, 0 \leqslant s \leqslant 1\}$$

and pictured in Figure 4.1.2

Observe from the picture in Figure 4.1.2 that the diagonal of $P(v, w)$ going from the origin to the point determined by $v + w$ gets mapped by R_θ by the corresponding diagonal in the parallelogram $P(R_\theta(v), R_\theta(w))$; namely, the one determined by $R_\theta(v) + R_\theta(w)$. It then follows that

$$R_\theta(v + w) = R_\theta(v) + R_\theta(w); \tag{4.4}$$

that is, the rotation R_θ maps the sum of two vectors to the sum of the images of the two vectors.

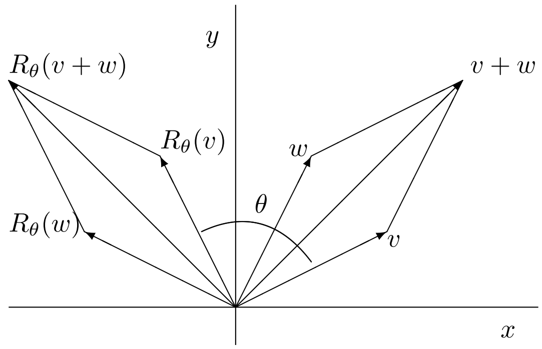


Figure 4.1.2: Image of a parallelogram under rotation

In Example 4.1.2 we have shown that the function R_θ maps the scalar multiple of a vector to the scalar multiple of its image with the same scaling factor (this is (4.3)), and it maps the sum of two vectors to the sum of their images (see Equation (4.4)); in other words, R_θ preserves the vector space operations in $\mathbb{R}^2$. A function satisfying the properties in (4.3) and (4.4) is said to be a **linear** function. We will spend a large portion of this chapter studying linear functions and learning about their properties. We will then see how the theory of linear functions can be used to solve Problem 4.0.8.

4.2 Linear Functions

Definition 4.2.1 (Linear Function)**.** A function $T\colon \mathbb{R}^n \to \mathbb{R}^n$ is said to be a **linear function**, or a **linear transformation**, if T satisfies the properties

(i) $T(cv) = cT(v)$ for all scalars c and all $v \in \mathbb{R}^n$, and

(ii) $T(u+v) = T(u) + T(v)$ for all $u, v \in \mathbb{R}^n$.

Example 4.2.2. Let $A \in \mathbb{M}(m, n)$ and define $T\colon \mathbb{R}^n \to \mathbb{R}^m$ by

$$T(v) = Av \quad \text{for all } v \in \mathbb{R}^n;$$

that is, $T(v$ is obtained by multiplying the column vector v by the $m \times n$ matrix on the left. Then, T is a linear function.

To see why T is linear, use the fact that $A(cv) = cAv$ for all scalars c and vectors v. This proves that (i) in Definition 4.2.1. Next, use the distributive property in matrix algebra to see that

$$A(v+w) = Av + Aw \quad \text{for all } v, w \in \mathbb{R}^n.$$

This proves that (ii) in Definition 4.2.1 holds true.

We therefore conclude that $T(v) = Av$, where A is an $m \times n$ matrix, defines linear function from $\mathbb{R}^n$ to $\mathbb{R}^n$.

Example 4.2.3 (Reflection on the x–axis)**.** Let $R\colon \mathbb{R}^2 \to \mathbb{R}^2$ denote refection of the x–axis; that is, for each $v \in \mathbb{R}^2$, $R(v)$ determines a point in $\mathbb{R}^2$ lying on a line through the point determined by v and perpendicular to the x–axis. The point determined by $R(v)$ lies on one of the two half–planes determined by the x–axis, which is opposite to that of where the point determined by v is located, and the distance from v to the x–axis is the same as the distance from $R(v)$ to the x–axis (see Figure 4.2.3).

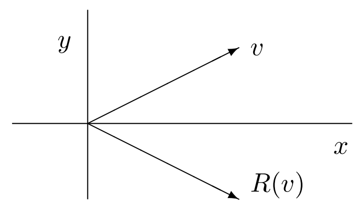


Figure 4.2.3: Reflection on the x-axis

Observe that if the coordinates of v are $\begin{pmatrix} x \\ y \end{pmatrix}$, then the coordinates of $R(v)$ are $\begin{pmatrix} x \\ -y \end{pmatrix}$. It then follows that

$$R\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x \\ -y \end{pmatrix},$$

which we can write as

$$R\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}.$$

Thus, R is of the form $R(v)Av$, where A is the 2×2 matrix given by

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

Consequently, by the result presented in Example 4.2.2, R is a linear function from $\mathbb{R}^2$ to $\mathbb{R}^2$.

Remark 4.2.4. Linear transformations form a very specialized class of vector valued functions. It is important to bear in mind that not all functions between Euclidean spaces are linear. For example, we have already encountered in this course the function $f\colon \mathbb{R}^n \to \mathbb{R}$ given by

$$f(v) = \|v\| \quad \text{for all } v \in \mathbb{R}^n,$$

where $\|\cdot\|$ denotes the Euclidean norm in $\mathbb{R}^n$. To see why f is not linear, simply consider the case of the vectors

$$v = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \text{and} \quad w = \begin{pmatrix} -1 \\ 0 \end{pmatrix}$$

in $\mathbb{R}^2$. Observe that $f(v) = 1$ and $f(w) = 1$; however, $f(v + w) = f(\mathbf{0}) = 0$. This, $f(v+w) \neq f(v) + f(w)$, and therefore condition (ii) in Definition 4.2.1 is not fulfilled.

Most functions dealt with in a single variable Calculus course are not linear. For instance, the quadratic function $f(x) = x^2$ for all $x \in \mathbb{R}$ is not linear since

$$f(x + y) = x^2 + y^2 + 2xy,$$

so that, if x and y are not 0, $f(x + y) \neq f(x) + f(y)$. Another example is provided by the sine function. Recall that

$$\sin(x + y) = \cos(y)\sin(x) + \cos(x)\sin(y).$$

In fact, the only linear function, $f \colon \mathbb{R} \to \mathbb{R}$, according to Definition 4.2.1, is

$$f(x) = ax \quad \text{for all } x \in \mathbb{R},$$

where a is a real constant. This is essentially the one–dimensional version of Example 4.2.2.

Functions that are not linear are usually referred in the literature as **nonlinear** functions, even though they actually form the bulk of functions arising in the applications of mathematics to the sciences and engineering. So, why do we spend a whole semester–course studying linear functions? Why not study the class of **all** functions, linear and nonlinear? There are two reasons for the in–depth study of linear functions. First, there is a rich, beautiful, complete and well known theory of linear functions, a glimpse of which is provided in this Linear Algebra course. Secondly, understanding linear functions provides a very powerful and simple tool for studying nonlinear functions. A very common approach in applications is to use linear functions, when possible, to approximate nonlinear functions. In a lot of cases, the behavior of the linear approximation near a point in $\mathbb{R}^n$ yields a lot of information about the nonlinear function around that point.

We will see in the next section that the function $T(v) = Av$, where A is an $m \times n$ matrix, given in Example 4.2.2 is essentially the only example of a linear transformation form $\mathbb{R}^n$ to $\mathbb{R}^m$.

We end this section by presenting an important class of linear transformations in $\mathbb{R}^n$.

Example 4.2.5 (Orthogonal Projections)**.** Let u denote a unit vector in $\mathbb{R}^n$ and let $L = \operatorname{span}\{u\}$; that is, L is the line through the origin in $\mathbb{R}^3$ in the direction of u. For each v in $\mathbb{R}^n$, we denote by $P_u(v)$ the point in L that is the closest to v. For instance, if $v = tu$, for some scalar t, then $P_u(v) = P_u(tu) = tu$. Thus, P_u defines a mapping from R_n to R_n whose image, $\mathcal{I}_{P_u}$, is the line L. We prove that

$$P_u \colon \mathbb{R}^n \to \mathbb{R}^n$$

is a linear function.

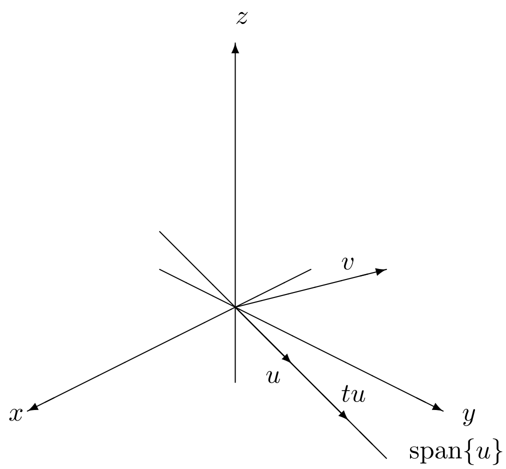


Figure 4.2.4: Orthogonal Projection

Before we prove the linearity of P_u, we first get a formula for computing $P_u(v)$ for any $v \in \mathbb{R}^n$. In order to do this, we define the function

$$f(t) = \|v - tu\|^2 \quad \text{for all } t \in \mathbb{R};$$

that is, $f(t)$ gives the square of the distance from v to the point tu on the line L. Figure 4.2.4 shows the situation we are discussion in $\mathbb{R}^3$. Using the Euclidean inner product, we can write $f(t)$ as

$$\begin{aligned} f(t) &= \langle v - tu, v - tu \rangle \\ &= \langle v, v \rangle + \langle v, -tu \rangle + \langle -tu, v \rangle + \langle -tu, -tu \rangle \\ &= \|v\|^2 - 2t\langle v, u \rangle + t^2\|u\|^2, \end{aligned}$$

where we have uses the bi–linearity of the Euclidean inner product. We therefore get that

$$f(t) = \|v\|^2 - 2t\langle v, u \rangle + t^2,$$

since u is a unit vector. Thus, $f(t)$ is a quadratic polynomial in t which can be shown to have an absolute minimum when

$$t = \langle v, u \rangle.$$

Hence,

$$P_u(v) = \langle v, u \rangle u.$$

The linearity of P_u then follows from the bi–linearity of the inner–product.

4.3 Matrix Representation of Linear Functions

In this section we show that every linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^n$ can be expressed as multiplication by an $m \times n$ matrix. In order to show this, observe that any vector, v, in $\mathbb{R}^n$ can be expressed as a linear combination of the standard basis, $\mathcal{E}_n = \{e_1, e_2, \ldots, e_n\}$, in $\mathbb{R}^n$; that is,

$$v = x_1 e_1 + x_2 e_2 + \cdots + x_n e_n,$$

where $x_1, x_2, \ldots, x_n$ are the coordinates of v relative to the basis $\mathcal{E}$,

$$[v]_{\mathcal{E}_n} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}.$$

Thus, if T is a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$, then

$$\begin{aligned} T(v) &= T(x_1 e_1 + x_2 e_2 + \cdots + x_n e_n) \\ &= T(x_1 e_1) + T(x_2 e_2) + \cdots + T(x_n e_n) \\ &= x_1 T(e_1) + x_2 T(e_2) + \cdots + x_n T(e_n), \end{aligned}$$

where we have used properties (i) and (ii) defining a linear transformation in Definition 4.2.1. We have therefore shown that a linear transformation, $T\colon \mathbb{R}^n \to \mathbb{R}^m$, is completely determined by what T does to the standard basis in $\mathbb{R}^n$. Writing $T(v)$ in terms of its coordinates relative to the standard basis $\mathcal{E}_m$ in $\mathbb{R}^m$, we get that

$$[T(v)]_{\mathcal{E}_m} = x_1 [T(e_1)]_{\mathcal{E}_m} + x_2 [T(e_2)]_{\mathcal{E}_m} + \cdots + x_n [T(e_n)]_{\mathcal{E}_m}; \tag{4.5}$$

in other words, the coordinate vector of $T(v)$ relative the standard basis, $\mathcal{E}_m$ is a linear combination of the coordinate vectors of $T(e_1), T(e_2), \ldots, T(e_n)$ relative to $\mathcal{E}_m$.

The expression in (4.5) can be written in terms of the matrix product as follows

$$[T(v)]_{\mathcal{E}_m} = \begin{bmatrix} [T(e_1)]_{\mathcal{E}_m} & [T(e_2)]_{\mathcal{E}_m} & \cdots & [T(e_n)]_{\mathcal{E}_m} \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix},$$

or

$$[T(v)]_{\mathcal{E}_m} = \begin{bmatrix} [T(e_1)]_{\mathcal{E}_m} & [T(e_2)]_{\mathcal{E}_m} & \cdots & [T(e_n)]_{\mathcal{E}_m} \end{bmatrix} [v]_{\mathcal{E}_m} \tag{4.6}$$

We denote the matrix $\begin{bmatrix} [T(e_1)]_{\mathcal{E}_m} & [T(e_2)]_{\mathcal{E}_m} & \cdots & [T(e_n)]_{\mathcal{E}_m} \end{bmatrix}$ in (4.6) by M_T and call it the **matrix representation** of T relative to the standard bases, $\mathcal{E}_n$ and $\mathcal{E}_m$, in $\mathbb{R}^n$ and $\mathbb{R}^m$, respectively, and denote it by M_T. We then have that

$$[T(v)]_{\mathcal{E}_m} = M_T [v]_{\mathcal{E}_m} \tag{4.7}$$

and usually write

$$T(v) = M_T v \tag{4.8}$$

with the understanding that $T(v)$ and v are expressed in terms of their coordinates relative to the standard bases in $\mathbb{R}^m$ and $\mathbb{R}^n$, respectively. The matrix representation of T, M_T, is obtained by computing the vectors $T(e_1), T(e_2), \ldots, T(e_n)$ and putting them as columns in the matrix M_T, in that order; that is,

$$M_T = \begin{bmatrix} T(e_1) & T(e_2) & \cdots & T(e_n) \end{bmatrix}. \tag{4.9}$$

The value of $T(v)$ is then computed by using the equation in (4.8).

Example 4.3.1 (Rotations in $\mathbb{R}^2$ (continued))**.** Let $R_\theta \colon \mathbb{R}^2 \to \mathbb{R}^2$ denote rotation in $\mathbb{R}^2$ through an angle of θ is the counterclockwise sense. We saw in Example 4.1.2 that R_θ is linear. In this example we compute the matrix representation for R_θ. In order to do this we compute $R_\theta(e_1)$ and $R_\theta(e_2)$ and use these as the columns of M_{R_θ}. Inspection of the sketch in Figure 4.3.5 reveals

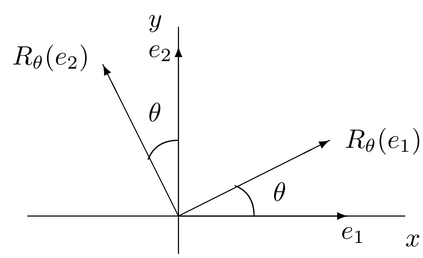


Figure 4.3.5: $R_\theta(e_1)$ and $R_\theta(e_2)$

that

$$R_\theta(e_1) = \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix} \quad \text{and} \quad R_\theta(e_2) = \begin{pmatrix} -\sin\theta \\ \cos\theta \end{pmatrix}.$$

It then follows that

$$M_{R_\theta} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

Thus, for any vector $v = \begin{pmatrix} x \\ y \end{pmatrix}$ in $\mathbb{R}^2$, the rotated image of v is given by

$$R_\theta(v) = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x\cos\theta - y\sin\theta \\ x\sin\theta + y\cos\theta \end{pmatrix}.$$

Example 4.3.2 (Rotations in $\mathbb{R}^3$)**.** Give the linear transformation,

$$R_{z,\theta} \colon \mathbb{R}^3 \to \mathbb{R}^3,$$

which rotates a vector around the z–axis through an angle of θ in the counterclockwise sense on the xy–plane.

Solution: In this case we want

$$R_{z,\theta}(e_1) = \begin{pmatrix} \cos\theta \\ \sin\theta \\ 0 \end{pmatrix}, \quad R_{z,\theta}(e_2) = \begin{pmatrix} -\sin\theta \\ \cos\theta \\ 0 \end{pmatrix}, \quad \text{and} \quad R_{z,\theta}(e_3) = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

We then have that the matrix representation for $R_{z,\theta}$ is

$$M_{R_{z,\theta}} = \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

□

Example 4.3.3. Find a linear transformation, $T\colon \mathbb{R}^2 \to \mathbb{R}^2$, which maps the square determined by the vectors e_1 and e_2 to the parallelogram determined by the vectors v_1 and v_2 in $\mathbb{R}^2$, and given by

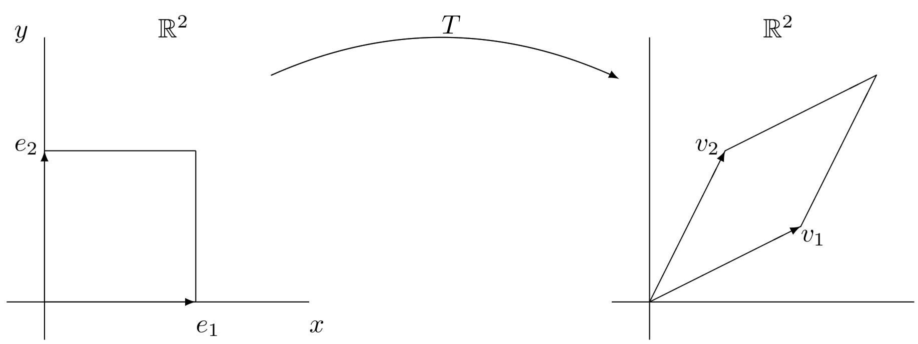


Figure 4.3.6: Picture for Example 4.3.3

$$v_1 = \begin{pmatrix} 2 \\ 1 \end{pmatrix} \quad \text{and} \quad v_2 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$$

and which are pictured in Figure 4.3.6.

Solution: We define T so that it maps e_1 to v_1 and e_2 to v_2. We then have that

$$T(e_1) = \begin{pmatrix} 2 \\ 1 \end{pmatrix} \quad \text{and} \quad T(e_2) = \begin{pmatrix} 1 \\ 2 \end{pmatrix}.$$

Thus, since we want T to be linear, its matrix representation relative to the standard basis in $\mathbb{R}^2$ is, according to (4.9),

$$M_T = \big[T(e_1) \quad T(e_2)\big] = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}.$$

It then follows that

$$T\begin{pmatrix} x \\ y \end{pmatrix} = M_T\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix},$$

or

$$T\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 2x + 1 \\ x + 2y \end{pmatrix}$$

for all $\begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R}^2$. Notice that this function does indeed maps the parallelogram $P(e_1, e_2)$ to the parallelogram $P(v_1, v_2)$ because the point determined by $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ on the upper right corner of the square gets mapped to the point determined by $\begin{pmatrix} 3 \\ 3 \end{pmatrix}$ and, since T is linear, lines get mapped to lines. □

Example 4.3.4. Find a linear transformation, $T\colon \mathbb{R}^2 \to \mathbb{R}^2$, which maps the parallelogram determined by the vectors

$$v_1 = \begin{pmatrix} 2 \\ 1 \end{pmatrix} \quad \text{and} \quad v_2 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$$

to the parallelogram determined by the vectors w_1 and w_2 in $\mathbb{R}^2$,

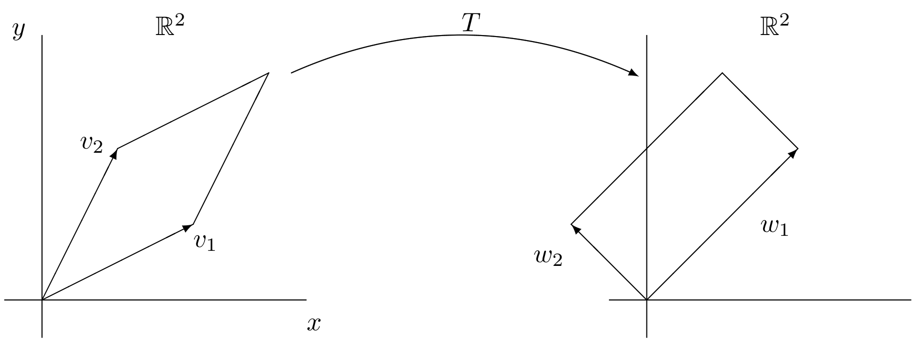


Figure 4.3.7: Picture for Example 4.3.4

$$w_1 = \begin{pmatrix} 2 \\ 2 \end{pmatrix} \quad \text{and} \quad w_2 = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

which are pictured in Figure 4.3.7.

***Solution**:* We define T so that it maps v_1 to w_1 and v_2 to v_2; that is, T is linear from $\mathbb{R}^2$ to $\mathbb{R}^2$ and

$$T(v_1) = w_1 \quad \text{and} \quad T(v_2) = w_2.$$

Thus, since we want T to be linear, its matrix representation relative to the standard basis in $\mathbb{R}^2$ is, according to (4.9),

$$M_T = \begin{bmatrix} T(e_1) & T(e_2) \end{bmatrix}.$$

Thus, we need to find $T(e_1)$ and $T(e_2)$.

Observe that $v_1 = 2e_1 + e_2$ and $v_2 = e_1 + 2e_2$. Thus, by the assumed linearity of T,

$$T(v_1) = 2T(e_1) + T(e_2)$$

and

$$T(v_2) = T(e_1) + 2T(e_2)$$

We therefore get the system

$$\begin{cases} 2T(e_1) + T(e_2) = w_1 \\ T(e_1) + 2T(e_2) = w_2, \end{cases}$$

which can be solved for $T(e_1)$ and $T(e_2)$ to yield that

$$T(e_1) = \begin{pmatrix} 5/3 \\ 1 \end{pmatrix}$$

and

$$T(e_2) = \begin{pmatrix} -4/3 \\ 0 \end{pmatrix}.$$

It then follows that

$$M_T = \begin{pmatrix} 5/3 & -4/3 \\ 1 & 0 \end{pmatrix}.$$

It then follows that

$$T\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x/3 - 4y/3 \\ x \end{pmatrix}$$

for all $\begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R}^2$. □

In addition to providing a way for computing the action of linear transformations on vectors in their domains, the matrix representation of a linear transformation can be used to answer questions about the linear transformation. For instance, the null space of a linear transformation $T \colon \mathbb{R}^n \to \mathbb{R}^n$ is the set

$$\mathcal{N}_T = \{v \in \mathbb{R}^n \mid T(v) = \mathbf{0}\}.$$

The linearity of T implies that $\mathcal{N}_T$ is a subspace of $\mathbb{R}^n$. Observe that

$$v \in \mathcal{N}_T \text{ if and only if } T(v) = \mathbf{0}$$

or

$$v \in \mathcal{N}_T \text{ if and only if } M_T v = \mathbf{0}.$$

It then follows that the null space of T is the same as the null space of the matrix representation, M_T, of T. Similarly, we can show that the image of T,

$$\mathcal{I}_T = \{w \in \mathbb{R}^m \mid w = T(v) \text{ for some } v \in \mathbb{R}^n\}$$

is the span of the columns of the matrix representation, M_T, of T.

4.4 Compositions

Given vector–valued functions $T\colon \mathbb{R}^n \to \mathbb{R}^m$ and $R\colon \mathbb{R}^m \to \mathbb{R}^k$, we can define a new function from $\mathbb{R}^n$ to $\mathbb{R}^k$, which we denote by $R \circ T$, as follows

$$R \circ T(v) = R(T(v)) \quad \text{for all } v \in \mathbb{R}^n. \tag{4.10}$$

Notice that, since T maps $\mathbb{R}^n$ to $\mathbb{R}^m$, and $\mathbb{R}^m$ is the domain of R, the definition of $R \circ T$ in (4.10) makes sense and yields a vector in $\mathbb{R}^k$. We call the function

$$R \circ T\colon \mathbb{R}^n \to \mathbb{R}^k$$

define in (4.10) the **composition of R and T**. Intuitively, the conposition of R and T is the successive application of T and R, in that order.

It is not hard to show that, if both T and R are linear functions, then the composition $R \circ T$ is a linear function as well. In fact, for $v, w \in \mathbb{R}^n$ we have that

$$R \circ T(v + w) = R(T(v + w)) = R(T(v) + T(w)),$$

since T is linear (here we used property (ii) in Definition 4.2.1). Applying next the linearity of R, we then get that

$$R \circ T(v + w) = R(T(v)) + R(T(w)) = R \circ T(v) + R \circ T(w).$$

This verifies condition (ii) in Definition 4.2.1.

We verify condition (i) in Definition 4.2.1 in a similar way:

$$R \circ T(cv) = R(T(cv)) = R(cT(v)) = cR(T(v)) = cR \circ T(v).$$

We next see how the matrix representation for $R \circ T$ relates to the matrix representations for R and T. We have the following proposition:

Proposition 4.4.1. Let $T\colon \mathbb{R}^n \to \mathbb{R}^m$ and $R\colon \mathbb{R}^m \to \mathbb{R}^k$ denote linear functions with corresponding matrix representations $M_T \in \mathbb{M}(m, n)$ and $M_R \in \mathbb{M}(k, m)$, respectively, with respect to the standard basis in $\mathbb{R}^n$, $\mathbb{R}^m$ and $\mathbb{R}^k$. Then, the matrix representation of the composition $R \circ T\colon \mathbb{R}^n \to \mathbb{R}^k$, with respect to the standard bases in $\mathbb{R}^n$and $\mathbb{R}^k$, is given by

$$M_{R\circ T} = M_R M_T;$$

that is, the matrix representation of a composition of linear functions is the matrix product of their matrix representations.

Proof: Compute $R \circ T(e_j)$ for $j = 1, 2, \ldots, n$ to get

$$R \circ T(e_j) = R(T(e_j) = R(M_T e_j),$$

since M_T is the matrix representation of T relative to the standard basis in $\mathbb{R}^n$ and $\mathbb{R}^m$. Using the same result for R we get

$$R \circ T(e_j) = R(T(e_j) = M_R M_T e_j \quad \text{for } j = 1, 2, \ldots, n.$$

Thus, the columns of $M_{R\circ T}$ are the columns of the matrix product $M_R M_T$ and the result follows. □

Example 4.4.2 (Rotations in $\mathbb{R}^3$ continued)**.** We saw in Example 4.3.2 that

$$M_{R_{z,\theta}} = \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

is the matrix representation for a rotation around the z–axis through an angel of θ in a direction that moves the positive x–axis towards the positive y–axis (see Figure 4.4.8).

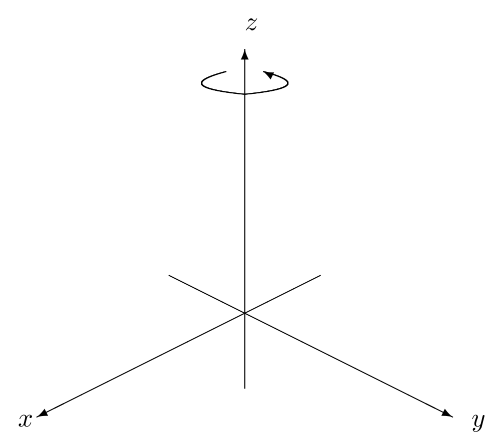


Figure 4.4.8: Positive rotation in $\mathbb{R}^3$ around the z–axis through an angle θ

A similar calculation to that used to obtain $M_{R_{z,\theta}}$ shows that the matrix representation of for a rotation, $R_{y,\varphi}\colon \mathbb{R}^3 \to \mathbb{R}^3$, around the y–axis through an angel of φ in a direction that moves the positive x–axis towards the positive z–axis is given by

$$M_{R_{y,\varphi}} = \begin{pmatrix} \cos\varphi & 0 & -\sin\varphi \\ 0 & 1 & 0 \\ \sin\varphi & 0 & \cos\varphi \end{pmatrix}$$

Suppose we perform a positive rotation around the z–axis through an angle θ followed by a positive rotation around the y–axis through an angle φ. Let $R\colon \mathbb{R}^3 \to \mathbb{R}^3$ denote the linear transformation which which performs the two rotations in succession; then,

$$R = R_{y,\varphi} \circ R_{z,\theta}$$

and, therefore, by the result of Proposition 4.4.1,

$$M_R = M_{R_{y,\varphi}} M_{R_{z,\theta}}.$$

we then have the matrix for the transformation that combines the two rotations in succession is

$$M_R = \begin{pmatrix} \cos\varphi & 0 & -\sin\varphi \\ 0 & 1 & 0 \\ \sin\varphi & 0 & \cos\varphi \end{pmatrix} \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

or

$$M_R = \begin{pmatrix} \cos\varphi\cos\theta & -\cos\varphi\sin\theta & -\sin\varphi \\ \sin\theta & \cos\theta & 0 \\ \sin\varphi\cos\theta & -\sin\varphi\sin\theta & \cos\varphi \end{pmatrix}. \tag{4.11}$$

Our solution to Problem 4.0.8 will show that R corresponds to a single rotation about some axis through the origin. We will eventually learn how to determine the axis and the angle of rotation.

Remark 4.4.3. Note that, like matrix multiplication, composition of functions is associative. In fact, let $T\colon \mathbb{R}^n \to \mathbb{R}^m$, $R\colon \mathbb{R}^m \to \mathbb{R}^k$ and $S\colon \mathbb{R}^k \to \mathbb{R}^p$ be functions. Then,

$$\begin{aligned} (T\circ R)\circ S(v) &= T\circ R(S(v)) \\ &= T(R(S(v))) \\ &= T(R\circ S(v)) \\ &= T\circ(R\circ S)(v) \end{aligned}$$

for all $v \in \mathbb{R}^n$. It then follows that

$$(T\circ R)\circ S = T\circ(R\circ S).$$

Function composition also distributes with the sum of functions. Let $T\colon \mathbb{R}^m \to \mathbb{R}^m$, $R\colon \mathbb{R}^m \to \mathbb{R}^k$ and $S\colon \mathbb{R}^m \to \mathbb{R}^k$. We can then define the sum of R and S as follows:

$$(R+S)(w) = R(w) + S(w) \quad \text{for all } w \in \mathbb{R}^m.$$

Note that this definition is possible because there is a vector addition defined in $\mathbb{R}^k$. We can then prove that

$$(R+S)\circ T = R\circ T + S\circ T.$$

To see why this is the case, observe that, for every $v \in \mathbb{R}^n$

$$\begin{aligned} (R+S)\circ T(v) &= (R+S)(T(v) \\ &= R(T(v)) + S(T(v)) \\ &= R\circ T(v) + S\circ T(v) \\ &= (R\circ T + S\circ T)(v). \end{aligned}$$

Similarly, if $T\colon \mathbb{R}^n \to \mathbb{R}^n$, $R\colon \mathbb{R}^n \to \mathbb{R}^m$ and $S\colon \mathbb{R}^m \to \mathbb{R}^k$,

$$S\circ(T+R) = S\circ T + S\circ R.$$

Given a linear transformation $T\colon \mathbb{R}^n \to \mathbb{R}^n$, if M_T is an invertible matrix, then we can define the transformation $R\colon \mathbb{R}^n \to \mathbb{R}^n$ by

$$R(w) = M_T^{-1}w \quad \text{for all } \ w \in \mathbb{R}^n.$$

It then follows that

$$R \circ T(v) = R(M_T v) = M_T^{-1}M_T v = Iv = v.$$

That is, $R \circ T$ maps every vector, v, in $\mathbb{R}^n$ to itself. This transformation is called the **identity transformation** and we denote it by I. We then have that

$$R \circ T = I.$$

Similarly,

$$T \circ R = I.$$

Definition 4.4.4 (Invertible Transformations)**.** A function $f\colon \mathbb{R}^n \to \mathbb{R}^n$ is said to be **invertible** if there exists a function $g\colon \mathbb{R}^n \to \mathbb{R}^n$ such that

$$f \circ g = g \circ f = I,$$

where $I\colon \mathbb{R}^n \to \mathbb{R}^n$ denotes the identity function; that is,

$$I(v) = v \quad \text{for all } \ v \in \mathbb{R}^n.$$

The function g is called the **inverse** of f, and f is the inverse of g. We usually denote g by f^{-1}.

We have just seen that if $T\colon \mathbb{R}^n \to \mathbb{R}^n$ is linear and its matrix representation, M_T, is invertible, then T is invertible and the inverse of T is given by

$$T^{-1}(v) = M_T^{-1}v \quad \text{for all } \ v \in \mathbb{R}^n.$$

4.5 Orthogonal Transformations

The matrix representation, M_R, given in (4.11) for the linear transformation R given in Example 4.6.15 has the following interesting property: If we write M_R in terms if its columns, u_1, u_2 and u_3, then it is not hard to check that

$$\|u_1\| = 1, \ \|u_2\| = 1, \ \|u_3\| = 1,$$

and

$$\langle u_i, u_j\rangle = 0 \quad \text{for } i \neq j.$$

It then follows that

$$\begin{aligned} M_R^T M_R &= \begin{pmatrix} u_1^T \\ u_2^T \\ u_3^T \end{pmatrix} \begin{bmatrix} u_1 & u_2 & u_3 \end{bmatrix} \\ &= \begin{pmatrix} u_1^T u_1 & u_1^T u_2 & u_1^T u_3 \\ u_2^T u_1 & u_2^T u_2 & u_2^T u_3 \\ u_3^T u_1 & u_3^T u_2 & u_3^T u_3 \end{pmatrix} \\ &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \end{aligned}$$

Thus,

$$M_R^T M_R = I.$$

Definition 4.5.1 (Orthogonal Matrix)**.** An $n \times n$ matrix, A, is said to be **orthogonal** if

$$A^T A = I,$$

where I denotes the identity matrix in $\mathbb{M}(n, n)$.

Thus, an $n \times n$ orthogonal matrix is invertible and its inverse is its transpose.

Definition 4.5.2 (Orthogonal Transformations)**.** A linear transformation, $R\colon \mathbb{R}^n \to \mathbb{R}^n$, is said to be **orthogonal** if its matrix representation M_R is orthogonal.

Proposition 4.5.3 (Properties of Orthogonal Transformations (Part I))**.** Let $R\colon \mathbb{R}^n \to \mathbb{R}^n$ denote an orthogonal transformation. Then,

(i) $\langle R(v), R(w)\rangle = \langle v, w\rangle$ for all $v, w \in \mathbb{R}^n$.

That is, an orthogonal transformation preserve the Euclidean inner product.

(ii) $\|R(v)\| = \|v\|$

That is, an orthogonal transformation preserve the Euclidean norm, or length, of vectors.

Proof of (i): Assume $R\colon \mathbb{R}^n \to \mathbb{R}^n$ is orthogonal. Then, its matrix representation, M_R, satisfies

$$M_R^T M_R = I,$$

where I is the $n \times n$ identity matrix. Thus, for v and w in $\mathbb{R}^n$,

$$\begin{aligned} \langle Rv, Rw \rangle &= \langle M_R v, M_R w \rangle \\ &= (M_R v)^T M_R w \\ &= v^T M_R^T M_R w \\ &= v^T I w \\ &= v^T w \\ &= \langle v, w \rangle. \end{aligned}$$

□

The second part of Proposition 4.5.3 is a straightforward consequence of the first part.

The first part of Proposition 4.5.3 can be interpreted geometrically as saying that orthogonal transformations preserve angles between vectors.

Example 4.5.4. In this example we see the connection of Euclidean inner product of two vectors and the angle between the vectors. We consider the situation in the xy–plane. Let u denote a unit vector in $\mathbb{R}^2$ and suppose that u makes an angle of φ with the positive x–axis; that is φ is the angle between u and e_1 (see Figure 4.5.9). We then have that

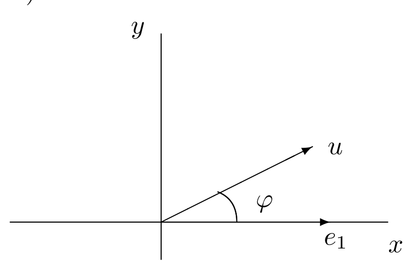


Figure 4.5.9: Angle between u and e_1

$$u = \begin{pmatrix} \cos\varphi \\ \sin\varphi \end{pmatrix},$$

since $\|u\| = 1$. Consequently,

$$\langle e_1, u \rangle = \cos\varphi.$$

That is, the inner product of the unit vectors e_1 and u is the cosine of the angle between them.

Next, consider two unit vectors, u_1 and u_2, whose angle is φ pictured in Figure 4.5.10.

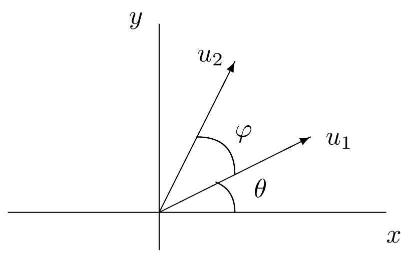


Figure 4.5.10: Angle between u_1 and u_2

Let θ denote the angle that u_1 in Figure 4.5.10 makes with the positive $x-$ axis. Apply a rotation around the origin through an angle θ in the clockwise sense. This is the linear function $R_{-\theta}\colon \mathbb{R}^2 \to \mathbb{R}^2$ whose matrix representation is

$$M_{R_{-\theta}} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}.$$

Observe that $M_{R_{-\theta}}$ is an orthogonal matrix. The result of applying the rotation $R_{-\theta}$ then yields situation like the one picture in Figure 4.5.11. Observe that

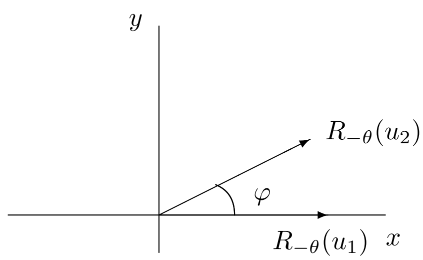


Figure 4.5.11: Angle between $R_{-\theta}(u_1)$ and $R_{-\theta}(u_1)$

$R_{-\theta}(u_1) = e_1$. Thus, since $R_{-\theta}$ is orthogonal,

$$\begin{aligned} \langle u_1, u_2\rangle &= \langle R_{-\theta}(u_1), R_{-\theta}(u_2)\rangle \\ &= \langle e_1, R_{-\theta}(u_2)\rangle \\ &= \cos\varphi. \end{aligned}$$

To see why the last equality it true, assume that the vectors u_1 and u_2 and the angles θ and φ are as pictured in Figure 4.5.10. Then, it is the case that

$$u_2 = \begin{pmatrix} \cos(\theta+\varphi) \\ \sin(\theta+\varphi) \end{pmatrix} = \begin{pmatrix} \cos\theta\cos\varphi - \sin\theta\sin\varphi \\ \sin\theta\cos\varphi + \cos\theta\sin\varphi \end{pmatrix},$$

which we can write in matrix form as

$$u_2 = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} \cos\varphi \\ \sin\varphi \end{pmatrix} = R_\theta \begin{pmatrix} \cos\varphi \\ \sin\varphi \end{pmatrix},$$

from which we get that

$$R_{-\theta}u_2 = \begin{pmatrix}\cos\varphi\\ \sin\varphi\end{pmatrix}.$$

It then follows that

$$\langle u_1, u_2\rangle = \cos\varphi;$$

that is, the Euclidean inner product of the unit vectors u_1 and u_2 is the cosine of the angle between them.

The second part of Proposition 4.5.3 says that orthogonal transformations preserve lengths. Thus, orthogonal transformations preserve angles and lengths. It is reasonable, therefore, to expect that orthogonal transformations preserve areas and volumes of parallelograms and parallelepipeds, respectively. We will see why this is the case in the next section.

4.6 Areas, Volumes and Orientation

4.6.1 Areas of Parallelograms

Two linearly independent vectors, v and w, in $\mathbb{R}^n$ determine a parallelogram

$$P(v, w) = \{tv + sw \mid 0 \leqslant t \leqslant 1,\ 0 \leqslant s \leqslant 1\}.$$

We would like to compute the area of $P(v, w)$. Figure 4.6.12 shows $P(v, w)$ for the special situation in which v and w lie in the first quadrant in the xy–plane. $\mathbb{R}^2$ We can see from the picture in Figure 4.6.12 that the area of $P(v, w)$ is given

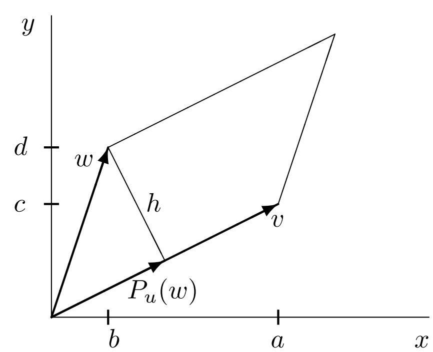


Figure 4.6.12: Parallelogram $P(v, w)$ in the xy–plane

by

$$\text{area}(P(v, w)) = \|v\| h, \tag{4.12}$$

where h is the distance from the point determined by w to the line spanned by v. In order to compute h, let u denote a unit vector in the direction of v; that is,

$$u = \frac{1}{\|v\|} v. \tag{4.13}$$

Recall that the orthogonal projection of w onto the direction of u,

$$P_u(w) = \langle w, u \rangle u, \tag{4.14}$$

gives the point on the line spanned by u which is closest to w. We then see that the norm of the vector $w - P_u(w)$ is the shortest distance from w to the line spanned by v. Consequently, $h = \|w - P_u(w)\|$. Substituting this expression for h into the expression for area$(P(v, w))$ in Equation (4.12) and squaring both sides of the equation then yields

$$\begin{aligned}
(\text{area}(P(v,w)))^2 &= \|v\|^2 \|w - P_u(w)\|^2 \\
&= \|v\|^2 \langle w - P_u(w), w - P_u(w) \rangle \\
&= \|v\|^2 (\|w\|^2 - 2\langle w, P_u(w)\rangle + \|P_u(w)\|^2) \\
&= \|v\|^2 \left(\|w\|^2 - 2 \left\langle w, \frac{\langle v, w\rangle}{\|v\|^2} v \right\rangle + \frac{\langle v, w\rangle^2}{\|v\|^2} \right) \\
&= \|v\|^2 \left(\|w\|^2 - 2 \frac{\langle v, w\rangle}{\|v\|^2} \langle w, v\rangle + \frac{\langle v, w\rangle^2}{\|v\|^2} \right) \\
&= \|v\|^2 \left(\|w\|^2 - 2 \frac{\langle v, w\rangle^2}{\|v\|^2} + \frac{\langle v, w\rangle^2}{\|v\|^2} \right) \\
&= \|v\|^2 \|w\|^2 - \langle v, w\rangle^2,
\end{aligned}$$

where we have used the properties of the Euclidean inner product, the definition of $P_u(w)$ in (4.14), and the fact that u is the unit vector given in (4.13). We have therefore shown that

$$(\text{area}(P(v,w)))^2 = \|v\|^2 \|w\|^2 - \langle v, w\rangle^2. \tag{4.15}$$

4.6.2 Determinant of 2×2 matrices

Applying formula (4.15) to the case in which the vectors v and w lie in $\mathbb{R}^2$ and have coordinates

$$\begin{pmatrix} a \\ c \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} b \\ d \end{pmatrix},$$

respectively, we can write (4.15) as

$$\begin{aligned}
(\text{area}(P(v,w)))^2 &= \|v\|^2\|w\|^2 - (v\cdot w)^2 \\
&= (a^2+c^2)(b^2+d^2) - (ab+cd)^2 \\
&= a^2b^2 + a^2d^2 + c^2b^2 + c^2d^2 - (a^2b^2 + 2abcd + c^2d^2) \\
&= a^2d^2 + c^2b^2 - 2adbc \\
&= (ad)^2 - 2(ad)(bc) + (bc)^2 \\
&= (ad-bc)^2.
\end{aligned}$$

Taking square roots on both sides we then have that

$$\text{area}(P(v,w)) = |ad-bc|. \tag{4.16}$$

Definition 4.6.1 (Determinant of a 2×2 matrix)**.** The expression $ad-bc$ in (4.16) is called the **determinant** of the 2×2 matrix

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

We denote the determinant of A by $\det(A)$ or $|A|$. We then have that

$$\det(A) = ad - bc,$$

or

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc.$$

Thus, the expression in (4.16) for the area of the parallelogram, $P(v,w)$, determined by the vectors v and W in $\mathbb{R}^2$ can be written as

$$\text{area}(P(v,w)) = |\det([\ v \quad w\ ])|; \tag{4.17}$$

that is, the area of $P(v,w)$ is the absolute value of the determinant of the 2×2 matrix, $[\ v \quad w\ ]$, whose columns are the vectors v and w.

The following properties of the determinant for a 2×2 matrices can be easily verified.

Proposition 4.6.2 (Properties of determinants of 2×2 matrices)**.** Let A denote a 2×2 matrix, $v, v_1, v_2, w \in \mathbb{R}^2$ and c denote a scalar. Then,

(i) $\det(I) = 1$, where I denotes the 2×2 identity matrix.

(ii) $\det(A^T) = \det(A)$, where A^T denotes the transpose of A.

(iii) $\det([\ v \quad w\ ]) = -\det([\ w \quad v\ ])$; that is, switch the columns of A once changes the sign of the determinant of A.

(iv) If the columns of A are linearly dependent, then $\det(A) = 0$. Conversely, if $\det(A) = 0$, then the columns of A are linearly dependent.

(v) $\det([\ cv \quad w\ ]) = c\det([\ v \quad w\ ])$.

(vi) $\det([\ v_1 + v_2 \quad w\ ]) = \det([\ v_1 \quad w\ ]) + \det([\ v_2 \quad w\ ])$.

(vii) $\det([\ v \quad cv + w\ ]) = \det([\ v \quad w\ ])$.

(viii) $\det(A) \neq 0$ if an only if A is invertible.

(ix) $\det(A) = 0$ if and only if A is singular; that is, $\det(A) = 0$ if and only if the equation $Ax = \mathbf{0}$ has nontrivial solutions.

Definition 4.6.3 (Determinant of a linear function in $\mathbb{R}^2$)**.** The determinant of a linear function, $T\colon \mathbb{R}^2 \to \mathbb{R}^2$, is the determinant of its matrix representation relative to the standard basis in $\mathbb{R}^2$; that is,

$$\det(T) = \det(M_T).$$

Example 4.6.4. The determinant of the rotation, $R_\theta\colon \mathbb{R}^2 \to \mathbb{R}^2$, is

$$\det(R_\theta) = \begin{vmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{vmatrix} = \cos^2\theta + \sin^2\theta = 1.$$

Example 4.6.5. Let $T\colon \mathbb{R}^2 \to \mathbb{R}^2$ denote reflection across the y–axis. Then,

$$T\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} -x \\ y \end{pmatrix},$$

so that

$$M_T = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$$

Thus,

$$\det(T) = -1.$$

Observe that the transformations R_θ and T in the previous two examples are orthogonal; therefore, it is not surprising that they they preserve areas of parallelogram. In fact, given an orthogonal transformation, $R\colon \mathbb{R}^2 \to \mathbb{R}^2$, the area of the transformed parallelogram $P(R(v), R(w))$ can be computed using (4.15) as follows

$$\begin{aligned} (\text{area}(P(R(v), R(w))))^2 &= \|R(v)\|^2\|R(w)\|^2 - \langle R(v), R(w)\rangle^2 \\ &= \|v\|^2\|w\|^2 - \langle v, w\rangle^2 \\ &= (\text{area}(P(v, w)))^2, \end{aligned}$$

where we have used Proposition 4.5.3. It then follows that

$$\text{area}(P(R(v), R(w))) = \text{area}(P(v, w)); \tag{4.18}$$

that is, orthogonal transformations preserve areas of parallelograms.

We can use (4.17) to write (4.18) in terms of the determinant of an orthogonal transformation from $\mathbb{R}^2$ to $\mathbb{R}^2$; in fact, applying (4.18) to the unit parallelogram $P(e_1, e_2)$ in $\mathbb{R}^2$ we obtain that

$$|\det([\ R(e_1) \quad R(e_2)\ ])| = |\det(I)| = 1.$$

It then follows that, for any orthogonal transformation, $R\colon \mathbb{R}^2 \to \mathbb{R}^2$,

$$|\det(R)| = 1.$$

Thus, there are two possibilities for the determinant of an orthogonal transformation, either 1 or -1. Examples 4.6.4 and 4.6.5 show these two possibilities for the case of a rotation and a reflection, respectively. It turns out the sign of the determinant is what distinguishes rotations from reflections. The determinant of a rotation is 1, while that of a reflection is -1. We will see that a positive determinant implies that the transformation preserves "orientation," while a negative determinant implies that it reverses "orientation." In order to see this, we first need to define the term "orientation." This will be done after we have defined the determinant of an $n \times n$ matrix for $n \geqslant 3$. However, before we do that, we will first define a special products of vectors in $\mathbb{R}^3$ known as the **cross product** and the **triple scalar product** in the next section. In the next section we deal with the simpler task of defining orientation in $\mathbb{R}^2$.

4.6.3 Orientation in $\mathbb{R}^2$

Given an ordered basis, $\mathcal{B} = \{v_1, v_2\}$, we say that $\mathcal{B}$ has a positive orientation if

$$\det([\ v_1 \quad v_2\ ]) > 0.$$

If $\det([\ v_1 \quad v_2\ ]) < 0$, we say that $\mathcal{B}$ has a negative orientation. For example, the standard, ordered basis, $\mathcal{E}_2 = \{e_1, e_2\}$, in $\mathbb{R}^2$ has a positive orientation since

$$\det([\ e_1 \quad e_2\ ]) = \det(I) = 1 > 0.$$

On the other hand, the ordered basis $\mathcal{B} = \{e_2, e_1\}$ has a negative orientation.

Definition 4.6.6 (Orientation Preserving Transformation in $\mathbb{R}^2$)**.** A linear transformation $T\colon \mathbb{R}^2 \to \mathbb{R}^2$ is said to be orientation preserving if

$$\det(T) > 0;$$

that is, if

$$\det([\ T(e_1) \quad T(e_2)\ ]) > 0.$$

Example 4.6.7 (Orientation Preserving Orthogonal, Transformations in $\mathbb{R}^2$)**.** In this example we see that an orthogonal transformation, $R\colon \mathbb{R}^2 \to \mathbb{R}^2$, which preserves orientation must be a rotation around the origin.

Let $M_R = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ denote the matrix representation for for R relative to the standard basis in $\mathbb{R}^2$. Then, since R is orthogonal, $M_R^{-1} = M_R^T$, where

$$M_R^{-1} = \frac{1}{\det(R)} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix},$$

since $\det(R) = 1$, and

$$M_R^T = \begin{pmatrix} a & c \\ b & d \end{pmatrix}.$$

It then follows that $a = d$ and $b = -c$. We then have that the matrix representation of R relative to the standard basis in $\mathbb{R}^2$ must be of the form

$$M_R = \begin{pmatrix} a & -c \\ c & a \end{pmatrix},$$

where

$$a^2 + c^2 = 1.$$

Setting $\sin\theta = c$ and $\cos\theta = a$, we then see that $R = R_\theta$; that is, R is rotation around the origin by θ. If $c > 0$, we set $\theta = \arccos(a)$, which is an angle between 0 and π, and so R is a rotation in the counterclockwise sense. On the other hand, if $c < 0$, we set $\theta = -\arccos(a)$, and so R is a rotation in the clockwise sense. If $c = 0$, R is the identity for $a = 1$, or R is rotation by π for $a = -1$.

4.6.4 The Cross–Product

Given two linearly independent vectors, v and w, in $\mathbb{R}^3$, we would like to associate to them a vector, denoted $v \times w$ and called the *cross product* of v and w, satisfying the following properties:

- $v \times w$ is orthogonal to the plane spanned by v and w.

- There are two choices for a perpendicular direction to the span of v and w. The direction for $v \times w$ is determined according to the so called "right–hand rule":

  > *With the fingers of your right hand, follow the direction of v while curling them towards the direction of w. The thumb will point in the direction of $v \times w$.*

- The norm of $v \times w$ is the area of the parallelogram, $P(v, w)$, determined by the vectors v and w.

Example 4.6.8. Suppose that v and w lie in the xy–plane and write

$$v = \begin{pmatrix} a \\ c \\ 0 \end{pmatrix} \quad \text{and} \quad w = \begin{pmatrix} b \\ d \\ 0 \end{pmatrix}.$$

Then,

$$v \times w = \begin{vmatrix} a & b \\ c & d \end{vmatrix} e_3. \tag{4.19}$$

Observe that e_3 is orthogonal to the xy–plane and therefore $v \times w$ is orthogonal to the plane spanned by v and w. Furthermore, for $v \times w$ given by (4.19),

$$\|v \times w\| = |ad - bc| = \text{area}(P(v, w)),$$

by the calculations leading to (4.16). Finally, to check that (4.19) gives the correct direction for $v \times w$, according to the right–hand rule, observe that, for $v = e_1$ and $w = e_2$, the formula in (4.19) yields

$$e_1 \times e_2 = e_3, \tag{4.20}$$

which is in agreement with the right–hand rule as shown in Figure 4.6.13

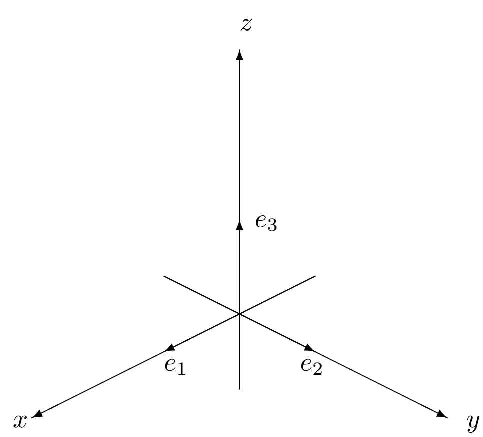


Figure 4.6.13: Right–hand Rule

Using the illustration in Figure 4.6.13 we also get that cross–product relations for the vectors in the standard basis in $\mathbb{R}^3$:

$$e_1 \times e_3 = -e_2, \tag{4.21}$$

and

$$e_2 \times e_3 = e_1. \tag{4.22}$$

Observe that, from the definition in (4.19),

$$w \times v = \begin{vmatrix} b & a \\ d & c \end{vmatrix} e_3 = -\begin{vmatrix} a & b \\ c & d \end{vmatrix} e_3 = -v \times w.$$

Thus, the **anti–symmetry** relation,

$$w \times v = -v \times w, \tag{4.23}$$

is inherent in the definition of $v \times w$ given so far. Observe that (4.23) implies that

$$v \times v = \mathbf{0} \quad \text{for all} \quad v \in \mathbb{R}^3. \tag{4.24}$$

To complete the definition of the cross product in $\mathbb{R}^3$, we require that it be bi–linear; that is, $v \times w$ is linear in both variables v and w; that is,

$$(c_1 v_1 + c_2 v_2) \times w = c_1 v_1 \times w + c_2 v_2 \times w, \tag{4.25}$$

and

$$v \times (d_1 w_1 + d_2 w_2) = d_1 v \times w_1 + d_2 v \times w_2, \tag{4.26}$$

for all vectors v, v_2, v_3, w, w_1 and w_2 in $\mathbb{R}^3$ and all scalars c_1, c_2, d_1 and d_2.

The relations in (4.20), (4.21 and (4.22) for the cross products of the vectors in the standard basis in $\mathbb{R}^3$, the anti–symmetry relation in (4.23) and the bi–linearity relations in (4.25) and (4.26) can be used to define the cross product in $\mathbb{R}^3$ as follows: Given vectors

$$v = \begin{pmatrix} a_1 \\ a_1 \\ a_1 \end{pmatrix} \quad \text{and} \quad w = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix},$$

write then in terms of the standard basis in $\mathbb{R}^3$,

$$\begin{aligned} v &= a_1 e_1 + a_2 e_2 + a_3 e_3, \\ w &= b_1 e_1 + b_2 e_2 + b_3 e_3. \end{aligned}$$

Then,

$$\begin{aligned} v \times w &= (a_1 e_1 + a_2 e_2 + a_3 e_3) \times (b_1 e_1 + b_2 e_2 + b_3 e_3) \\ &= a_1 b_2 \ e_1 \times \ e_2 + a_1 b_3 \ e_1 \times \ e_3 + a_2 b_1 \ e_2 \times \ e_1 \\ &\qquad + a_2 b_3 \ e_2 \times \ e_3 + a_3 b_1 \ e_3 \times \ e_1 + a_3 b_2 \ e_3 \times \ e_2, \end{aligned}$$

where we have used the bi-linearity relations and (4.24). Thus, using the relations in (4.20), (4.21 and (4.22), we get that

$$v \times w = a_1 b_2 \ e_3 - a_1 b_3 \ e_2 - a_2 b_1 \ e_3 + a_2 b_3 \ e_1 + a_3 b_1 \ e_2 - a_3 b_2 \ e_1,$$

which we could re–arrange as

$$v \times w = (a_2 b_3 - a_3 b_2) \ e_1 - (a_1 b_3 - a_3 b_1) \ e_2 + (a_1 b_2 - a_2 b_1) \ e_3.$$

We can write this vector product in terms of the determinants of the 2×2 matrices

$$\begin{pmatrix} a_2 & b_2 \\ a_3 & b_3 \end{pmatrix}, \quad \begin{pmatrix} a_1 & b_1 \\ a_3 & b_3 \end{pmatrix}, \quad \text{and} \quad \begin{pmatrix} a_1 & b_1 \\ a_2 & b_2 \end{pmatrix}$$

as follows

$$v \times w = \begin{vmatrix} a_2 & b_2 \\ a_3 & b_3 \end{vmatrix} e_1 - \begin{vmatrix} a_1 & b_1 \\ a_3 & b_3 \end{vmatrix} e_2 + \begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix} e_3. \tag{4.27}$$

We take (4.27) as our definition of the cross product of the vectors

$$v = \begin{pmatrix} a_1 \\ a_1 \\ a_1 \end{pmatrix} \quad \text{and} \quad w = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}$$

in $\mathbb{R}^3$.

We presently verify that the cross product, $v \times w$, satisfies the required properties stated at the beginning of this section. Specifically, we verify that

- $v \times w$ is orthogonal to the plane spanned by v and w;

  and

- the norm of $v \times w$ is the area of the parallelogram, $P(v, w)$, determined by the vectors v and w.

First, we verify that $v \times w$ is orthogonal to v by computing

$$\begin{aligned} \langle v, v \times w \rangle &= a_1 \begin{vmatrix} a_2 & b_2 \\ a_3 & b_3 \end{vmatrix} - a_2 \begin{vmatrix} a_1 & b_1 \\ a_3 & b_3 \end{vmatrix} + a_3 \begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix} \\ &= a_1(a_2b_3 - a_3b_2) - a_2(a_1b_3 - a_3b_1) + a_3(a_1b_2 - a_2b_1) \\ &= 0. \end{aligned}$$

Similarly, we can compute $\langle w, v \times w \rangle = 0$. Therefore, $v \times w$ is orthogonal to both v and w.

Calculations involving the definition of the Euclidean inner product and norm can be used to show that, if $v \times w$ is given by (4.27), then

$$\|v \times w\|^2 = \|v\|^2 \|w\|^2 - \langle v, w \rangle^2.$$

which, by virtue of (4.15) shows that

$$\|v \times w\| = \text{area}(P(v, w)).$$

Thus, the norm of $v \times w$ is the area of the parallelogram, $P(v, w)$, determined by the vectors v and w.

4.6.5 The Triple–Scalar Product

Given vectors u, v and w in $\mathbb{R}^3$, whose coordinates relative to the standard basis in $\mathbb{R}^3$ are

$$u = \begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix}, \quad v = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} \quad \text{and} \quad w = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix},$$

respectively, we define the **triple scalar product** of u, v and w to be

$$\langle u, v \times w \rangle = c_1 \begin{vmatrix} a_2 & b_2 \\ a_3 & b_3 \end{vmatrix} - c_2 \begin{vmatrix} a_1 & b_1 \\ a_3 & b_3 \end{vmatrix} + c_3 \begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}. \tag{4.28}$$

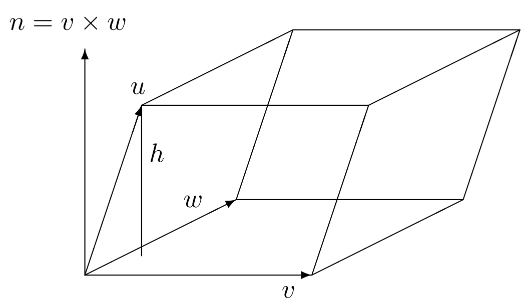


Figure 4.6.14: Volume of Parallelepiped

Geometrically, the absolute value of the triple scalar product $\langle u, v \times w \rangle$ is the volume of the parallelepiped generated by the vectors u, v and w. To see why this is so, denote by $P(v, w, u)$ the parallelepiped spanned by v, w, and u, and by $P(v, w)$ the parallelogram spanned by v and w. Observe that the volume of the parallelepiped drawn in Figure 4.6.14 is the area of the parallelogram spanned by v and w times the height, h, of the parallelepiped:

$$\text{volume}(P(v, w, u)) = \text{area}(P(v, w)) \cdot h, \tag{4.29}$$

where h can be obtained by projecting u onto the cross–product, $v \times w$, of v and w; that is

$$h = \|P_n(u)\| = \left\| \frac{\langle u, n \rangle}{\|n\|^2} \, n \right\|,$$

where

$$n = v \times w.$$

We then have that

$$h = \frac{|\langle u, v \times w \rangle|}{\|v \times w\|}.$$

Consequently, since $\text{area}(P(v, w)) = \|v \times w\|$, we get from (4.29) that

$$\text{volume}(P(v, w, u)) = |\langle u, v \times w \rangle|. \tag{4.30}$$

4.6.6 Determinant of 3×3 matrices

We can use the triple scalar product of vectors in $\mathbb{R}^3$ to define the determinant of a 3×3 matrix, A, as follows:

Definition 4.6.9 (Determinant of a 3×3 matrix)**.** Write the matrix A in terms of its columns,

$$A = [\ v_1 \quad v_2 \quad v_3\ ],$$

where v_1, v_2 and v_3 are vectors in $\mathbb{R}^3$. We define $\det(A)$ to be the triple scalar product of v_1, v_2 and v_3, in that order; that is,

$$\det(A) = \langle v_1, v_2 \times v_3 \rangle. \tag{4.31}$$

Thus, for A given by

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix},$$

using (4.31) and the definition of the triple scalar product in (4.28), we obtain the formula

$$\det(A) = a_{11} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{21} \begin{vmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{vmatrix} + a_{31} \begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix}. \tag{4.32}$$

Using the expression in (4.30) for the volume of a parallelepiped and (4.31) we then obtain that

$$\text{volume}(P(v_1, v_2, v_3)) = |\det([\ v_1 \quad v_2 \quad v_3\ ])|; \tag{4.33}$$

that is, the volume of $P(v_1, v_2, v_3)$ is the absolute value of the determinant of the 3×3 matrix, $[\ v_1 \quad v_2 \quad v_3\ ]$, whose columns are the vectors v_1, v_2 and v_3.

Using the definition of the determinant of a 3×3 matrix as a triple scalar product, or that given in (4.32), we can can derive the following properties for the determinant of 3×3 matrices,

Proposition 4.6.10 (Properties of determinant of 3×3 matrices)**.** Let A denote a 3×3 matrix, $u, u_1, u_2, v, w \in \mathbb{R}^3$ and c denote a scalar. Then,

(i) $\det(I) = 1$, where I denotes the 3×3 identity matrix.

(ii) $\det(A^T) = \det(A)$, where A^T denotes the transpose of A.

(iii) $\det([\ v \quad u \quad w\ ]) = -\det([\ u \quad v \quad w\ ])$, $\det([\ w \quad v \quad u\ ]) = -\det([\ u \quad v \quad w\ ])$ and $\det([\ u \quad w \quad v\ ]) = -\det([\ u \quad v \quad w\ ])$; that is, swapping two columns of A once changes the sign of the determinant of A.

(iv) If the columns of A are linearly dependent, then $\det(A) = 0$. Conversely, if $\det(A) = 0$, then the columns of A are linearly dependent.

(v) $\det([\ cu \quad v \quad w\ ]) = c{\cdot}\det([\ u \quad v \quad w\ ])$; $\det([\ u \quad cv \quad w\ ]) = c{\cdot}\det([\ u \quad v \quad w\ ])$; and $\det([\ u \quad v \quad cw\ ]) = c \cdot \det([\ u \quad v \quad w\ ])$.

(vi) $\det([\ u_1 + u_2 \quad v \quad w\ ]) = \det([\ u_1 \quad v \quad w\ ]) + \det([\ u_2 \quad v \quad w\ ])$;
$\det([\ u \quad v_1 + v_2 \quad w\ ]) = \det([\ u \quad v_1 \quad w\ ]) + \det([\ u \quad v_2 \quad w\ ])$; and
$\det([\ u \quad v \quad w_1 + w_2\ ]) = \det([\ u \quad v \quad w_1\ ]) + \det([\ u \quad v \quad w_2\ ])$.

(vii) $\det([\ u \quad cu + v \quad w\ ]) = \det([\ u \quad v \quad w\ ])$ and
$\det([\ u \quad v \quad cu + w\ ]) = \det([\ u \quad v \quad w\ ])$.

(viii) $\det(A) \neq 0$ if an only if A is invertible.

(ix) $\det(A) = 0$ if and only if A is singular; that is, $\det(A) = 0$ if and only if the equation $Ax = \mathbf{0}$ has nontrivial solutions.

Remark 4.6.11. These properties can be derived from the definition of the determinant of $A \in \mathbb{M}(3,3)$ as the triple–scalar product of the columns of A (see the formulas in (4.31) and (4.28)), or the formula for $\det(A)$ in (4.32), and the interpretation of $|\det(A)|$ as the volume of the parallelepiped generated by the columns of A (see (4.30)). For instance, to prove part (ii) of Proposition 4.6.10, write

$$A = \begin{pmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{pmatrix}.$$

Then, using the definition of $\det(A)$ in (4.31) and (4.28), we have that

$$\begin{aligned}
\det(A) &= a_1 \begin{vmatrix} b_2 & c_2 \\ b_3 & c_3 \end{vmatrix} - a_2 \begin{vmatrix} b_1 & c_1 \\ b_3 & c_3 \end{vmatrix} + a_3 \begin{vmatrix} b_1 & c_1 \\ b_2 & c_2 \end{vmatrix} \\
&= a_1 \begin{vmatrix} b_2 & b_3 \\ c_2 & c_3 \end{vmatrix} - a_2(b_1c_3 - b_3c_1) + a_3(b_1c_2 - b_2c_1) \\
&= a_1 \begin{vmatrix} b_2 & b_3 \\ c_2 & c_3 \end{vmatrix} - a_2b_1c_3 + a_2b_3c_1 + a_3b_1c_2 - a_3b_2c_1 \\
&= a_1 \begin{vmatrix} b_2 & b_3 \\ c_2 & c_3 \end{vmatrix} - a_2b_1c_3 + a_3b_1c_2 + a_2b_3c_1 - a_3b_2c_1 \\
&= a_1 \begin{vmatrix} b_2 & b_3 \\ c_2 & c_3 \end{vmatrix} - b_1(a_2c_3 - a_3c_2) + c_1(a_2b_3 - a_3b_2) \\
&= a_1 \begin{vmatrix} b_2 & b_3 \\ c_2 & c_3 \end{vmatrix} - b_1 \begin{vmatrix} a_2 & a_3 \\ c_2 & c_3 \end{vmatrix} + c_1 \begin{vmatrix} a_2 & a_3 \\ b_2 & b_3 \end{vmatrix} \\
&= \det \begin{pmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{pmatrix},
\end{aligned}$$

where we have used again the definition of the determinant in (4.32). Observe that the matrix

$$\begin{pmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{pmatrix}$$

is the transpose of A, and therefore part (ii) of Proposition 4.6.10 follows

Definition 4.6.12 (Triangular Matrices)**.** A 3×3 matrix, $A = [a_{ij}]$, is said to be **upper triangular** if $a_{ij} = 0$ for $i > j$. $A = [a_{ij}]$, is said to be **lower triangular** if $a_{ij} = 0$ for $i < j$.

Proposition 4.6.13 (Determinants of 3×3 triangular matrices)**.** Let $A = [a_{ij}]$ be 3×3 upper triangular or lower triangular matrix. Then,

$$\det(A) = a_{11} \cdot a_{22} \cdot a_{33}.$$

Proof: Assume that A is upper triangular; so that

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ 0 & a_{22} & a_{23} \\ 0 & 0 & a_{33} \end{pmatrix}.$$

Then, using the definition of $\det(A)$ in (4.32),

$$\det(A) = a_{11} \begin{vmatrix} a_{22} & a_{23} \\ 0 & a_{33} \end{vmatrix} - 0 \cdot \begin{vmatrix} a_{12} & a_{13} \\ 0 & a_{33} \end{vmatrix} + 0 \cdot \begin{vmatrix} a_{12} & a_{13} \\ a_{22} & 0 \end{vmatrix} = a_{11} \cdot a_{22} \cdot a_{33},$$

which was to be shown.

If A is upper triangular, then A^T is lower triangular; then, the result just proved and part (ii) of Proposition 4.6.10 imply that $\det(A) = \det(A^T) = a_{11} \cdot a_{22} \cdot a_{33}$. □

Definition 4.6.14 (Determinant of a linear function in $\mathbb{R}^3$)**.** The determinant of a linear function, $T\colon \mathbb{R}^3 \to \mathbb{R}^3$, is the determinant of its matrix representation relative to the standard basis in $\mathbb{R}^2$; that is,

$$\det(T) = \det(M_T).$$

Example 4.6.15. Let $R\colon \mathbb{R}^3 \to \mathbb{R}^3$ denote the transformation obtained in Example 4.6.15 as the composition of two rotations: $R = R_{y,\varphi} \circ R_{z,\theta}$. We saw in Example 4.6.15 that

$$M_R = \begin{pmatrix} \cos\varphi\cos\theta & -\cos\varphi\sin\theta & -\sin\varphi \\ \sin\theta & \cos\theta & 0 \\ \sin\varphi\cos\theta & -\sin\varphi\sin\theta & \cos\varphi \end{pmatrix}. \tag{4.34}$$

We compute $\det(M_R)$ using the formula for the triple scalar product in (4.28) to get

$$\begin{aligned}
\det(M_R) &= \cos\varphi\cos\theta\begin{vmatrix}\cos\theta & 0\\ -\sin\varphi\sin\theta & \cos\varphi\end{vmatrix}\\
&\quad +\cos\varphi\sin\theta\begin{vmatrix}\sin\theta & 0\\ \sin\varphi\cos\theta & \cos\varphi\end{vmatrix}\\
&\quad -\sin\varphi\begin{vmatrix}\sin\theta & \cos\theta\\ \sin\varphi\cos\theta & -\sin\varphi\sin\theta\end{vmatrix}\\
&= \cos^2\varphi\cos^2\theta+\cos^2\varphi\sin^2\theta+\sin\varphi(\sin\varphi\sin^2\theta+\sin\varphi\cos^2\theta)\\
&= \cos^2\varphi+\sin^2\varphi\\
&= 1.
\end{aligned}$$

It the follows that $\det(R) = 1$.

In what remains of this section, we will prove the following important property of the determinant function:

Proposition 4.6.16. Let A and B denote 3×3 matrices. Then,

$$\det(AB) = \det(A)\det(B). \tag{4.35}$$

As an application of Proposition 4.6.16, we prove the following

Proposition 4.6.17. For any scalar c and any 3×3 matrix B

$$\det(cB) = c^3 \det(B).$$

Proof: We first prove the result for the 3×3 identity matrix; namely,

$$\det(cI) = c^3,$$

which follows from Proposition 4.6.13 because

$$cI = \begin{pmatrix} c & 0 & 0\\ 0 & c & 0\\ 0 & 0 & c\end{pmatrix}$$

is a triangular matrix.

Next, apply Proposition 4.6.16 with $A = cI$ to get

$$\det(cB) = \det[(cI)B] = \det(cI)\det(B) = c^3\det(B).$$

□

The proof of Proposition 4.6.16 will proceed by stages. We will prove first the case in which A is singular and then prove the case in which A is nonsingular. The case in which A is nonsingular will also proceed by stages by first considering the case in which A is an elementary matrix.

Proposition 4.6.18. Let A and B be 3×3 matrices. If A is singular, then

$$\det(AB) = 0, \tag{4.36}$$

for any 3×3 matrix B.

Proof: Assume that A is a 3×3 singular matrix. The proof of (4.36) will follow from part (ix) of Proposition 4.6.10 once we establish the fact that A is singular implies that AB is singular for any 3×3 matrix B. Assume, by way of contradiction that AB is nonsingular; it then follows that $(AB)^T = B^T A^T$ is nonsingular. Since we are assuming that A is singular, we obtain from parts (ii) and (ix) of Proposition 4.6.10 that A^T is singular; so, there exists $v \in \mathbb{R}^3$, $v \neq \mathbf{0}$, such that

$$A^T v = \mathbf{0};$$

thus,

$$B^T A^T v = \mathbf{0}, \quad \text{for } v \neq \mathbf{0},$$

which shows that $B^T A^T$ is singular. This is a contradiction; hence, AB is singular if A is singular, and (4.36) follows. □

Lemma 4.6.19. Let B be a 3×3 matrix and E an elementary 3×3 matrix. Then

$$\det(EB) = \det(E)\det(B). \tag{4.37}$$

Proof: There are three kinds of elementary matrices: (i) those obtained from the 3×3 identity matrix by interchanging two rows; for example,

$$E_1 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}; \tag{4.38}$$

(ii) those obtain from the 3×3 identity matrix by multiplying a row by a constant c; for example,

$$E_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & c & 0 \\ 0 & 0 & 1 \end{pmatrix}; \tag{4.39}$$

and (iii) those obtained from the 3×3 identity matrix by adding a multiple of one row to another row and putting the result in the latter row; for example,

$$E_3 = \begin{pmatrix} 1 & 0 & 0 \\ c & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{4.40}$$

Next, we compute the determinants of the matrices E_1, E_2 and E_3 in (4.38), (4.39) and (4.40), respectively.

Note that $E_1 = [e_2 \quad e_1 \quad e_3]$; so that, by part (iii) of Proposition 4.6.10,

$$\det(E_1) = -\det([e_1 \quad e_2 \quad e_3] = -1. \tag{4.41}$$

Since matrices E_2 and E_3 are triangular matrices, we can use Proposition 4.6.13 to compute

$$\det(E_2) = c, \tag{4.42}$$

and

$$\det(E_3) = 1. \tag{4.43}$$

Write $B = \begin{pmatrix} R_1 \\ R_2 \\ R_3 \end{pmatrix}$, where $R_i \in \mathbb{M}(1,3)$, for $i = 1, 2, 3$, are the rows of B. Then,

$$E_1 B = \begin{pmatrix} R_2 \\ R_1 \\ R_3 \end{pmatrix};$$

so that

$$(E_1 B)^T = [R_2^T \quad R_1^T \quad R_3^T];$$

Thus,

$$\det((E_1 B)^T) = -\det([R_1^T \quad R_2^T \quad R_3^T]), \tag{4.44}$$

where we have used part (iii) of Proposition 4.6.10. It follows from (4.44) and part (ii) of Proposition 4.6.10 that

$$\det(E_1 B) = -\det(B). \tag{4.45}$$

Combining (4.45) and (4.41) then yields

$$\det(E_1 B) = \det(E_1)\det(B). \tag{4.46}$$

Next, note that

$$E_2 B = \begin{pmatrix} R_1 \\ cR_2 \\ R_3 \end{pmatrix};$$

thus, $(E_2 B)^T = [R_1^T \quad cR_2^T \quad R_3^T]$ and, using part (v) of Proposition 4.6.10,

$$\det[(E_2 B)^T] = c\det[R_1^T \quad R_2^T \quad R_3^T] = c\det(B^T).$$

Hence, by virtue of part (ii) of Proposition 4.6.10,

$$\det(E_2 B) = c\det(B). \tag{4.47}$$

Combining (4.42) and (4.47) we get

$$\det(E_2 B) = \det(E_2)\det(B). \tag{4.48}$$

Next, observe that

$$E_3B = \begin{pmatrix} R_1 \\ cR_1 + R_2 \\ R_3 \end{pmatrix};$$

so that

$$(E_3B)^T = [R_1^T \quad cR_1^T + R_2^T \quad R_3^T]. \tag{4.49}$$

Applying part (vii) of Proposition 4.6.10 to (4.49) we have

$$\begin{aligned} \det[(E_3B)^T] &= \det[R_1^T \quad cR_1^T + R_2^T \quad R_3^T] \\ &= \det[R_1^T \quad R_2^T \quad R_3^T]; \end{aligned}$$

so that

$$\det[(E_3B)^T] = \det(B^T);$$

thus, by virtue of part (ii) of Proposition 4.6.10,

$$\det(E_3B) = \det(B). \tag{4.50}$$

In view of (4.43) and (4.50) we see that

$$\det(E_3B) = \det(E_3)\det(B). \tag{4.51}$$

Finally, note that (4.46), (4.48) and (4.51) are instances of (4.37) for the three classes of elementary 3×3 matrices. We have therefore established Lemma 4.6.19. □

Proposition 4.6.20. Let B be a 3×3 matrix and A an invertible 3×3 matrix. Then

$$\det(AB) = \det(A)\det(B). \tag{4.52}$$

Proof: Let A and B denote 3×3 matrices and assume that A is invertible. It then follows from Proposition 3.3.21 that

$$A = E_1E_2\cdots E_k, \tag{4.53}$$

for elementary 3×3 matrices $E_1, E_2 \ldots, E_k$.

Applying Lemma 4.6.19 to (4.54) successively we obtain

$$\det(A) = \det(E_1)\det(E_2)\cdots\det(E_k). \tag{4.54}$$

Next, write

$$AB = E_1E_2\cdots E_kB, \tag{4.55}$$

and apply Lemma 4.6.19 to (4.55) successively we obtain

$$\det(AB) = \det(E_1)\det(E_2)\cdots\det(E_k)\det(B). \tag{4.56}$$

Finally, combine (4.54) and (4.56) to obtain (4.52). □

We end this section with the proof of Proposition 4.6.16.

Proof of Proposition 4.6.16: Let A and B be 3×3 matrices. Assume that A is singular. It then follows from part (ix) of Proposition 4.6.10 that

$$\det(A) = 0,$$

and from Proposition 4.6.18 that

$$\det(AB) = 0.$$

Consequently,

$$\det(AB) = \det(A)\det(B),$$

and (4.35) is established in this case.

On the other hand, if A is nonsingular, (4.35) follows from Proposition 4.6.20. The proof of Proposition 4.6.16 is now complete. □

4.6.7 Orientation in $\mathbb{R}^3$

It is not surprising that $|\det(R)| = 1$ in the Example 4.6.15, since R is an orthogonal transformation and therefore it preserves angles between vectors and lengths. The fact that $\det(R) > 0$ will then imply that R also preserves orientation. Given an ordered basis $\mathcal{B} = \{v_1, v_2, v_3\}$ of $\mathbb{R}^3$, we say that $\mathcal{B}$ has a positive orientation if

$$\langle v_1, v_2 \times v_3 \rangle > 0.$$

If $\langle v_1, v_2 \times v_3 \rangle < 0$, we say that $\mathcal{B}$ has a negative orientation. We say that a transformation $T\colon \mathbb{R}^3 \to \mathbb{R}^3$ preserves orientation if $\det(T) > 0$. If $\det(T) < 0$, we say that T reverses orientation.

Example 4.6.21. Let $T\colon \mathbb{R}^3 \to \mathbb{R}^3$ denote reflection on the xy–plane; that is,

$$T\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x \\ y \\ -z \end{pmatrix},$$

or

$$T\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix}.$$

Thus,

$$M_T = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}$$

and, therefore, $\det(T) = \det(M_T) = -1 < 0$. Hence, T reverses orientation.

In the next chapter we will prove that any orthogonal transformation from $\mathbb{R}^3$ to $\mathbb{R}^3$ which preserves orientation must be a rotation. This will complete the solution to the problem that we stated at the beginning of these notes.

Chapter 5

The Eigenvalue Problem

We have seen in the previous chapter that a rotation in $\mathbb{R}^3$ can be modeled by an orthogonal transformation from $\mathbb{R}^3$ to $\mathbb{R}^3$ which also preserves orientation. It is not hard to see that compositions of orthogonal, orientation preserving transformations are also orthogonal and orientation preserving. Thus, a partial solution to the motivating problem stated at the start of these notes, and re–stated as Problem 4.0.8 on page 91, will be attained if we can show that for any orientation preserving, orthogonal transformation,

$$R\colon \mathbb{R}^3 \to \mathbb{R}^3,$$

there exists a nonzero vector $v \in \mathbb{R}^3$ such that

$$R(v) = v;$$

that is, R fixes the line spanned by v. This would correspond to the axis of rotation of the transformation.

Given a linear transformation, $T\colon \mathbb{R}^n \to \mathbb{R}^n$, a scalar, λ, for which the equation

$$T(v) = \lambda v$$

has a nontrivial solution, v, is called an **eigenvalue** of the transformation T, and a nontrivial solution of $T(v) = \lambda v$ is called an **eigenvector** corresponding to the eigenvalue λ. Thus, in order to solve Problem 4.0.8, we will have to show that any orientation preserving, orthogonal transformation from $\mathbb{R}^3$ to $\mathbb{R}^3$ must have the scalar 1 as an eigenvalue.

We begin our discussion of the eigenvalue problem by presenting the example of characterizing all orthogonal, orientation reversing transformations in $\mathbb{R}^2$.

5.1 Orientation reversing, orthogonal transformations in $\mathbb{R}^2$

This section is a follow–up to Example 4.6.6. In that example, we proved that any orientation preserving, orthogonal transformation in $\mathbb{R}^2$ must be a rotation. In what follows we will prove that any orientation reversing, orthogonal transformation in $\mathbb{R}^2$ must be a reflection.

Assume that $R\colon \mathbb{R}^2 \to \mathbb{R}^2$ is an orthogonal transformation with

$$\det(R) = -1. \tag{5.1}$$

Let $M_R = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ denote the matrix representation for for R relative to the standard basis in $\mathbb{R}^2$. Then, since R is orthogonal,

$$M_R^{-1} = M_R^T, \tag{5.2}$$

where

$$M_R^{-1} = \frac{1}{\det(R)}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix} = \begin{pmatrix} -d & b \\ c & -a \end{pmatrix}, \tag{5.3}$$

since $\det(R) = -1$. We also note that

$$M_R^T = \begin{pmatrix} a & c \\ b & d \end{pmatrix}. \tag{5.4}$$

It then follows from (5.2)–(5.4) that $d = -a$ and $b = c$. We then have that the matrix representation of R relative to the standard basis in $\mathbb{R}^2$ must be of the form

$$M_R = \begin{pmatrix} a & b \\ b & -a \end{pmatrix}, \tag{5.5}$$

where

$$a^2 + b^2 = 1. \tag{5.6}$$

We claim that there exist nonzero vectors, v_1 and v_2, in $\mathbb{R}^2$ such that

$$M_R v_1 = v_1$$

and

$$M_R v_2 = -v_2$$

Definition 5.1.1 (Eigenvalues and Eigenvectors)**.** Let $T\colon \mathbb{R}^n \to \mathbb{R}^n$ be a linear transformation. A scalar, λ, is said to be an **eigenvalue** of T if and only if the equation

$$T(v) = \lambda v \tag{5.7}$$

has a nontrivial solution.

A nontrivial solution, v, of the equation $T(v) = \lambda v$ is called an **eigenvector** corresponding to the eigenvalue λ.

Observe that the equation in (5.7) can also be written as

$$(T - \lambda I)v = \mathbf{0}, \tag{5.8}$$

where $I\colon \mathbb{R}^n \to \mathbb{R}^n$ denotes the identity transformation in $\mathbb{R}^n$. Thus, λ is an eigenvalue of T if and only if the null space of the linear transformation $T - \lambda I$ is nontrivial; that is $\mathcal{N}_{T-\lambda I} \neq \{\mathbf{0}\}$. The null space of $T - \lambda I$ is called the **eigenspace** of T corresponding to λ and is denoted by $E_T(\lambda)$.

Thus, according to Definition 5.1.1, we wish to prove that the linear function $R\colon \mathbb{R}^2 \to \mathbb{R}^2$, whose matrix representation, M_R, is given by (5.5) has eigenvalues $\lambda_1 = 1$ and $\lambda_2 = -1$. This will prove that R is a reflection on the line given by $E_R(1)$. To see why this is the case, we will show that eigenspace corresponding to $\lambda_2 = -1$ is a line orthogonal to $E_R(1)$ which gets reflected across the line $E_R(1)$ (see the picture in Figure 5.1.1).

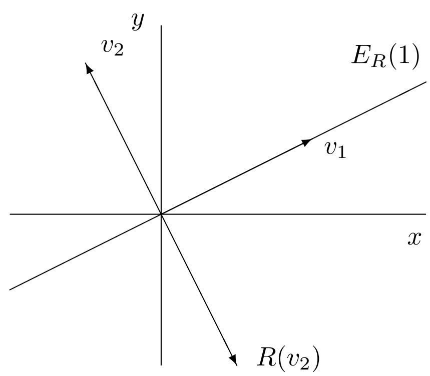


Figure 5.1.1: Reflection in $\mathbb{R}^2$

In order to find eigenvalues of R, we look for values of λ for which the system

$$(M_R - \lambda I)v = \mathbf{0} \tag{5.9}$$

has nontrivial solutions, where M_R is the matrix given in (5.5) and I is the 2×2 identity matrix. Now the system in (5.9) has nontrivial solutions when the columns of the matrix

$$M_R - \lambda I = \begin{pmatrix} a - \lambda & b \\ b & -a - \lambda \end{pmatrix}$$

are linearly dependent, which occurs if and only if the determinant of $M_R - \lambda I$ is 0; that is, λ is an eigenvalue of R if and only if

$$(\lambda + a)(\lambda - a) - b^2 = 0$$

or

$$\lambda^2 - 1 = 0,$$

since $a^2 + b^2 = 1$. We then get that $\lambda_1 = 1$ and $\lambda_2 = -1$ are eigenvalues of R, which was to be shown.

In order to find the eigenspace corresponding to $\lambda_1 = 1$, we solve the homogeneous system

$$\begin{pmatrix} a - 1 & b \\ b & -a - 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \tag{5.10}$$

In order to solve (5.10), we first consider the case $b = 0$. Then, from $a^2+b^2 = 1$, we get that $a^2 = 1$, so that $a = 1$ or $a = -1$. If $a = 1$, the system in (5.10) is equivalent to the system

$$\begin{pmatrix} 0 & 0 \\ 0 & -2 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

which is equivalent to the equation $y = 0$. Thus, setting $x = t$, where t is arbitrary we get that the solution space of (5.10) for the case $b = 0$ and $a = 1$ is given by

$$\begin{pmatrix} x \\ y \end{pmatrix} = t\begin{pmatrix} 1 \\ 0 \end{pmatrix},$$

so that

$$E_R(1) = \text{span}\left\{\begin{pmatrix} 1 \\ 0 \end{pmatrix}\right\},$$

or the x–axis. Thus, in this case R is reflection across the x–axis. Similarly, if $b = 0$ and $a = -1$, we get from the system in (5.10) that

$$E_R(1) = \text{span}\left\{\begin{pmatrix} 0 \\ 1 \end{pmatrix}\right\},$$

so that, in this case, R is reflection across the y–axis.

Next, assume that $b \neq 0$ and perform Gaussian elimination on the system in (5.10) to the get the system

$$\begin{pmatrix} 1 & -(a+1)/b \\ 0 & 0 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \qquad (5.11)$$

where we have used $a^2 + b^2 = 1$.

Observe that the system in (5.11) is equivalent to the equation

$$x - \frac{a+1}{b}y = 0,$$

which has solutions space given by

$$\begin{pmatrix} x \\ y \end{pmatrix} = t\begin{pmatrix} a+1 \\ b \end{pmatrix},$$

where t is arbitrary. We therefore get that the eigenspace of R corresponding to $\lambda = 1$ is

$$E_R(1) = \text{span}\left\{\begin{pmatrix} a+1 \\ b \end{pmatrix}\right\}. \qquad (5.12)$$

Next, we solve the system in (5.9) for $\lambda = -1$, which is the same as

$$\begin{pmatrix} a+1 & b \\ b & -a+1 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \qquad (5.13)$$

A calculation similar to the one leading to (5.12) can be used to show that, for the case $b \neq 0$, the eigenspace corresponding to $\lambda = -1$ is

$$E_R(-1) = \text{span}\left\{\begin{pmatrix} a-1 \\ b \end{pmatrix}\right\}. \tag{5.14}$$

Thus, we have shown that

$$v_1 = \begin{pmatrix} a+1 \\ b \end{pmatrix}$$

is an eigenvector corresponding to $\lambda_1 = 1$ and

$$v_2 = \begin{pmatrix} a-1 \\ b \end{pmatrix}$$

is an eigenvector corresponding to $\lambda_2 = -1$. That is,

$$R(v_1) = v_1$$

and

$$R(v_2) = -v_2.$$

Furthermore, v_1 and v_2 are orthogonal; to see why this is so, compute

$$\langle v_1, v_2\rangle = (a+1)(a-1) + b^2 = a^2 - 1 + b^2 = 0,$$

since $a^2+b^2 = 1$. Thus, R is indeed a reflection across the line $E_R(1)$. Note that R fixes the line $E_R(1)$; that is, $R(v) = v$ for all $v \in E_R(1)$; for, if $v \in E_R(1)$, then

$$v = cv_1,$$

for some scalar c, so that, by the linearity of R,

$$R(v) = R(cv_1) = cR(v_1) = cv_1 = v.$$

Note that R does not fix $E_R(-1)$, given in (5.14). However, it maps $E_R(-1)$ to itself; that is, $R(v) \in E_R(-1)$ for all $v \in E_R(-1)$. To see this, let $v \in E_R(-1)$; then, $v = cv_2$ for some scalar, c. Then,

$$R(v) = R(cv_2) = cR(v_2) = -cv_2 \in \text{span}\{v_2\} = E_R(-1).$$

Definition 5.1.2 (Invariant Subspaces)**.** Let $T\colon \mathbb{R}^n \to \mathbb{R}^n$ denote a linear transformation. A subspace, W, of $\mathbb{R}^n$ is said to be **invariant under** T if and only if

$$T(w) \in W \quad \text{for all } w \in W;$$

in other words, W is invariant under T iff

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

We have seen in this section that, if $R\colon \mathbb{R}^2 \to \mathbb{R}^2$ is an orthogonal, orientation reversing transformation, then R has invariant subspaces $E_R(1)$ and $E_R(-1)$. The invariant subspace $E_R(1)$ is the line of reflection of R. The line $E_R(-1)$ is orthogonal to $E_R(1)$ and is reflected across $E_R(1)$ by the transformation R.

5.2 Orientation preserving, orthogonal transformations in $\mathbb{R}^3$

In this section we solve Problem 4.0.8 on page 91. We will first re-formulate the problem in the language of linear transformation and the eigenvalue problem.

Theorem 5.2.1. *Let $R\colon \mathbb{R}^3 \to \mathbb{R}^3$ denote an orthogonal transformation which preserves orientation. We claim that $\lambda = 1$ is an eigenvalue of R. Furthermore, if u is a eigenvector corresponding to $\lambda = 1$ of norm 1, then R is a rotation around the span of u.*

Proof: We first prove that if $R\colon \mathbb{R}^3 \to \mathbb{R}^3$ is orthogonal and preserves orientation, then $\lambda = 1$ is an eigenvalue of R. To show this, let M_R denote the matrix representation for R relative to the standard basis in $\mathbb{R}^3$ and assume that

$$M_R^T M_R = M_R M_R^T = I, \tag{5.15}$$

where I denotes the 3×3 identity matrix, and

$$\det(M_R) = 1. \tag{5.16}$$

We prove that the equation

$$M_R v = v$$

has a nontrivial solution in $\mathbb{R}^3$, or equivalently, the homogenous system

$$(M_R - I)v = \mathbf{0} \tag{5.17}$$

has nontrivial solutions. This occurs if and only if

$$\det(M_R - I) = 0. \tag{5.18}$$

Observe that

$$\begin{aligned} \det(M_R - I) &= \det(M_R - M_R M_R^T) \\ &= \det(M_R(I - M_R^T)) \\ &= \det(M_R)\det((I - M_R^T)) \\ &= \det((I - M_R^T)), \end{aligned}$$

where we have used (5.15), the distributive property of matrix multiplication, Proposition 4.6.16 and (5.16). Thus, using the fact that, for any matrices A and B of the same dimension, $(A + B)^T = A^T + B^T$, we get that

$$\begin{aligned} \det(M_R - I) &= \det((I - M_R)^T) \\ &= \det(I - M_R), \end{aligned}$$

by part (ii) of Proposition 4.6.10. It then follows that

$$\begin{aligned} \det(M_R - I) &= \det(-(M_R - I)) \\ &= (-1)^3 \det(M_R - I), \end{aligned}$$

by Proposition 4.6.17. Consequently,

$$\det(M_R - I) = -\det(M_R - I),$$

from which (5.18) follows, and therefore the homogeneous system in (5.18) has nontrivial solutions. Hence, $\lambda = 1$ is an eigenvalue of R.

Next, let u denote an eigenvector of R corresponding to the eigenvalue $\lambda = 1$; assume also that $\|u\| = 1$. Define $P_u\colon \mathbb{R}^3 \to \mathbb{R}^3$ by

$$P_u(v) = \langle u, v\rangle u \quad \text{for all } v \in \mathbb{R}^3;$$

that is, P_u is orthogonal projection onto the direction of u. Then, the image of P_u is the span of the vector u,

$$\mathcal{I}_{P_u} = \text{span}\{u\};$$

so

$$\dim(\mathcal{I}_{P_u}) = 1. \tag{5.19}$$

Let W denote the null space of P_u. We then have that

$$W = \{w \in \mathbb{R}^3 \mid \langle u, w\rangle = 0\};$$

that is, W is the space of vectors in $\mathbb{R}^3$ which are orthogonal to u. By the Dimension Theorem we then get, in view of (5.19), that

$$\dim(W) = 2.$$

Thus, W is a two–dimensional subspace in $\mathbb{R}^3$; in other words, W is a plane through the origin in $\mathbb{R}^3$ which is perpendicular to u.

Since, W is two–dimensional, it has a basis, $\{w_1, w_2\}$, which we may assume consists of unit vectors. We may further assume that w_1 and w_2 are orthogonal to each other. To see why this is the case, let $\{v_1, v_2\}$ denote any basis for W. By multiplying by the reciprocal of their norms, if necessary, we may assume that $\|v_1\| = \|v_2\| = 1$. Set $w_1 = v_1$ and find a scalar c such that $v_2 + cw_1$ is orthogonal to w_1; in other words,

$$\langle v_2 + cw_1, w_1\rangle = 0,$$

which yields

$$c = -\langle v_2, w_1\rangle.$$

Finally, set

$$w_2 = \frac{1}{\|v_2 - \langle v_2, w_1\rangle w_1\|}(v_2 - \langle v_2, w_1\rangle w_1).$$

Then, w_2 is a unit vector which is orthogonal to w_1.

We may also choose w_1 and w_2 so that

$$\det([\ w_1 \quad w_2 \quad u\ ]) = 1. \tag{5.20}$$

To see why we can do this, observe that, since w_1, w_2 and u are mutually orthogonal and have length 1,

$$\text{volume}(P(w_1, w_2, u)) = 1,$$

so that, by (4.33), $|\det([\ w_1 \quad w_2 \quad u\ ])| = 1$. We therefore have two possibilities for $\det([\ w_1 \quad w_2 \quad u\ ])$: 1 or -1. If the determinant of $[\ w_1 \quad w_2 \quad u\ ] = -1$, we may switch the order of w_1 and w_2, and rename them w_2 and w_1, respectively to get (5.20).

Next, we show that W is an invariant subspace of R; that is, we show that

$$R(W) \subseteq W,$$

or equivalently

$$R(w) \in W \quad \text{for all } w \in W. \tag{5.21}$$

To show (5.21), let $w \in W$. Then, $\langle u, w\rangle = 0$ and, using the fact that $R(u) = u$,

$$\begin{aligned} \langle R(w), u\rangle &= \langle R(w), R(u)\rangle \\ &= \langle w, u\rangle \end{aligned}$$

since R is orthogonal. Consequently, $\langle R(w), u\rangle = 0$, which shows that $R(w)$ is in W, and (5.21) is established. It then follows that

$$R(w_1) = aw_1 + cw_2 \tag{5.22}$$

and

$$R(w_2) = bw_1 + dw_2, \tag{5.23}$$

for some scalars a, b, c and d, since $W = \text{span}\{w_1, w_2\}$.

In what remains of this section we will show that the effect of R on W is that of rotating it by some angle θ. To see why this is the case, set $\mathcal{B} = \{w_1, w_2, u\}$. We see by (5.20) that $\mathcal{B}$ is a basis for $\mathbb{R}^3$; this can also be seen from the observation that $\mathcal{B}$ forms an orthonormal set of three vectors in $\mathbb{R}^3$. Thus, any vector, v, in $\mathbb{R}^3$ can be expressed as

$$v = y_1 w_1 + y_2 w_2 + y_3 u, \tag{5.24}$$

where y_1, y_1 and y_3 are the coordinates of v relative to $\mathcal{B}$. Thus,

$$[v]_{\mathcal{B}} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}$$

is the coordinates vector of v relative to $\mathcal{B}$. Applying the linear transformation, R, on v given in (5.24) we obtain

$$\begin{aligned} R(v) &= y_1 R(w_1) + y_2 R(w_2) + y_3 R(u) \\ &= y_1(aw_1 + cw_2) + y_2(bw_1 + dw_2) + y_3 u \\ &= (ay_1 + by_2)w_1 + (cy_1 + dy_2)w_2 + y_3 u, \end{aligned}$$

where we have used (5.22) and (5.23) and the fact that u is an eigenvector for R corresponding to the eigenvalue $\lambda = 1$. We then have that the coordinates of $R(v)$ relative to $\mathcal{B}$ are given by

$$[R(v)]_{\mathcal{B}} = \begin{pmatrix} ay_1 + by_2 \\ cy_1 + dy_2 \\ y_3 \end{pmatrix},$$

which may be written as

$$[R(v)]_{\mathcal{B}} = \begin{pmatrix} a & b & 0 \\ c & d & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix},$$

or

$$[R(v)]_{\mathcal{B}} = \begin{pmatrix} a & b & 0 \\ c & d & 0 \\ 0 & 0 & 1 \end{pmatrix} [v]_{\mathcal{B}}. \tag{5.25}$$

We claim that the entries a, b, c and d in the matrix in (5.25) satisfy the relations

$$\begin{cases} d = a \\ b = -c \\ a^2 + c^2 = 1. \end{cases} \tag{5.26}$$

These relations will imply that (5.25) may be further re-written as

$$[R(v)]_{\mathcal{B}} = \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix} [v]_{\mathcal{B}}. \tag{5.27}$$

The expression in (5.27) shows that, when viewed from the frame of reference provided by the basis $\mathcal{B} = \{w_1, w_2, u\}$, R is a rotation around the axis determined by the eigenvector u through an angle of θ, where θ is determined by $\sin\theta = c$ and $\cos\theta = a$.

In order to prove the relations in (5.26) for the entries a, b, c and d in the 3×3 matrix in (5.25), denote it by A; that is, let

$$A = \begin{pmatrix} a & b & 0 \\ c & d & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

where a, b, c and d are determined by (5.22) and (5.23). We claim that

(i) A is orthogonal, and

(ii) $\det(A) = 1$.

To see why (i) is true, compute

$$\begin{aligned} \langle R(w_i), R(w_j)\rangle &= \langle w_i, w_j\rangle \\ &= \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j, \end{cases} \end{aligned}$$

for $i, j = 1, 2$, where we have used the assumption that R is orthogonal and the fact that w_1 and w_2 are mutually orthogonal with norm 1. On the other hand, using (5.22) and (5.23), we obtain that

$$\begin{aligned} \langle R(w_1), R(w_1)\rangle &= \langle aw_1 + cw_2, aw_1 + cw_2\rangle \\ &= a^2\langle w_1, w_1\rangle + ac\langle w_1, w_2\rangle + ca\langle w_2, w_2\rangle + c^2\langle w_2, w_2\rangle \\ &= a^2 + c^2, \end{aligned}$$

again by the orthonormality of the basis $\{w_1, w_2\}$. It then follows that

$$a^2 + c^2 = 1. \tag{5.28}$$

Similar calculations show that

$$b^2 + d^2 = 1 \tag{5.29}$$

and

$$ab + cd = 0. \tag{5.30}$$

The relations in (5.28), (5.29) and (5.30) imply that A is orthogonal; in fact,

$$\begin{aligned} A^T A &= \begin{pmatrix} a & c & 0 \\ b & d & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} a & b & 0 \\ c & d & 0 \\ 0 & 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} a^2 + c^2 & ab + cd & 0 \\ ab + cd & b^2 + d^2 & 0 \\ 0 & 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \\ &= I. \end{aligned}$$

Next, to see that $\det(A) = 1$, let Q denote the matrix $[\, w_1 \quad w_2 \quad u \,]$. Then, by (5.20),

$$\det(Q) = 1.$$

It then follows that Q is invertible and that

$$Q^{-1}w_1 = e_1, \quad Q^{-1}w_2 = e_2 \quad \text{and} \quad Q^{-1}u = e_3,$$

since

$$Qe_1 = w_1, \quad Qe_2 = w_2 \quad \text{and} \quad Qe_3 = u.$$

Consider the matrix $Q^{-1}M_RQ$. Observe that the first column of this matrix is

$$\begin{aligned} Q^{-1}M_RQe_1 &= Q^{-1}M_Rw_1 \\ &= Q^{-1}R(w_1) \\ &= Q^{-1}(aw_1 + cw_2) \\ &= aQ^{-1}w_1 + cQ^{-1}w_2 \\ &= ae_1 + ce_2 \\ &= \begin{pmatrix} a \\ c \\ 0 \end{pmatrix}. \end{aligned}$$

Similarly, the second and third column of $Q^{-1}M_RQ$ are

$$Q^{-1}M_RQe_2 \begin{pmatrix} b \\ d \\ 0 \end{pmatrix}$$

and

$$Q^{-1}M_RQe_3 \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix},$$

respectively. We then conclude that

$$Q^{-1}M_RQ = \begin{pmatrix} a & b & 0 \\ c & d & 0 \\ 0 & 0 & 1 \end{pmatrix} = A.$$

We then get that

$$\begin{aligned}
\det(A) &= \det(Q^{-1}M_RQ) \\
&= \det(Q^{-1})\det(M_R)\det(Q) \\
&= \frac{1}{\det(Q)}\det(R)\det(Q) \\
&= \det(R) \\
&= 1.
\end{aligned}$$

Observe that

$$\det(A) = a\begin{vmatrix} d & 0 \\ 0 & 1 \end{vmatrix} - c\begin{vmatrix} b & 0 \\ 0 & 1 \end{vmatrix} = ad - bc.$$

Consequently, $ad - bc = 1$. Observe that this implies that

$$A^{-1} = \begin{pmatrix} d & -b & 0 \\ -c & a & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

On the other hand,

$$A^T = \begin{pmatrix} a & c & 0 \\ b & d & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

This, since A is orthogonal, $A^{-1} = A^T$ and, therefore, the relations in (5.26) follow, which we wanted to prove. □

Example 5.2.2. Let $R\colon \mathbb{R}^3 \to \mathbb{R}^3$ by a linear transformation obtained by composing a rotation around the x–axis by $-\frac{\pi}{2}$ and rotation around the y–axis by $-\frac{\pi}{2}$; that is,

$$R = R_{x,-\frac{\pi}{2}} \circ R_{y,\frac{\pi}{2}}; \tag{5.31}$$

The rotations $R_{x,-\frac{\pi}{2}}$ and $R_{y,\frac{\pi}{2}}$ are shown pictorially in Figure 5.2.2.

The matrix representation for $R_{x,-\frac{\pi}{2}}$ can be obtained from Figure 5.2.2 to be

$$M_{R_{x,-\frac{\pi}{2}}} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix}, \tag{5.32}$$

since $R_{x,-\frac{\pi}{2}}$ maps e_1, e_2 and e_3 to e_1, $-e_3$ and e_2, respectively. Similarly,

$$M_{R_{y,\frac{\pi}{2}}} = \begin{pmatrix} 0 & 0 & -1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}. \tag{5.33}$$

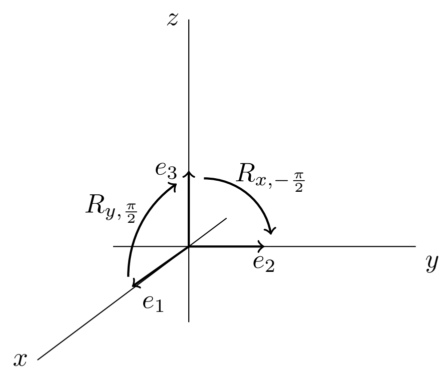


Figure 5.2.2: $R = R_{x,-\frac{\pi}{2}} \circ R_{y,\frac{\pi}{2}}$.

It follows from (5.31), (5.32) and (5.33) that the matrix representation for R relative to the standard basis is

$$\begin{aligned} M_R &= M_{R_{x,-\frac{\pi}{2}}} M_{R_{y,\frac{\pi}{2}}} \\ &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 & -1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} \\ &= \begin{pmatrix} 0 & 0 & -1 \\ 1 & 0 & 0 \\ 0 & -1 & 0 \end{pmatrix} \end{aligned} \tag{5.34}$$

Since M_R is orthogonal (see Problem 1 in Assignment #22) and

$$\det(M_R) = \det(M_{R_{x,-\frac{\pi}{2}}}) \det(M_{R_{y,\frac{\pi}{2}}}) = 1,$$

it follows from Theorem 5.2.1 that $\lambda = 1$ is an eigenvalue of R. In order to find an eigenvector for R corresponding to the eigenvalue $\lambda = 1$, we solve the system

$$(M_R - I)v = \mathbf{0}, \tag{5.35}$$

where I denotes the 3×3 identity matrix and M_R is the matrix in (5.34). In order to solve the equation in (5.35) we perform elementary row operations to the augmented matrix

$$\left(\begin{array}{ccc|c} -1 & 0 & -1 & 0 \\ 1 & -1 & 0 & 0 \\ 0 & -1 & -1 & 0 \end{array}\right)$$

to obtain

$$\left(\begin{array}{ccc|c} 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right) \tag{5.36}$$

It follows from the matrix in (5.36) that the equation in (5.35) is equivalent to the system

$$\begin{cases} x_1 \quad + x_3 &= 0 \\ \quad x_2 + x_3 &= 0, \end{cases}$$

which has solutions

$$\begin{cases} x_1 &= t; \\ x_2 &= t; \\ x_3 &= -t, \end{cases}$$

for $t \in \mathbb{R}$. It then follows that $v = \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}$ is an eigenvector for R corresponding to the eigenvalue $\lambda = 1$. According to Theorem 5.2.1, the line span$\{v\}$ is the axis of rotation of the orthogonal transformation R. Next, we see how to determine the angle of rotation around that axis.

Set

$$u = \begin{pmatrix} 1/\sqrt{3} \\ 1/\sqrt{3} \\ -1/\sqrt{3} \end{pmatrix}, \tag{5.37}$$

so that u is a unit vector in the direction of v and, therefore, u is also an eigenvector for R corresponding to the eigenvalue $\lambda = 1$.

Let Γ denote the plane through the origin in $\mathbb{R}^3$ that is orthogonal to u; so that

$$\Gamma = \{w \in \mathbb{R}^3 \mid \langle u, w\rangle = 0\}, \tag{5.38}$$

or

$$\Gamma = \left\{ \begin{pmatrix} x \\ y \\ z \end{pmatrix} \in \mathbb{R}^3 \;\middle|\; x + y - z = 0 \right\}. \tag{5.39}$$

Then, Γ is a 2–dimensional subspace of $\mathbb{R}^3$ that is invariant under the transformation R; that is,

$$R(\Gamma) \subseteq \Gamma. \tag{5.40}$$

The assertion in (5.40) follows from the fact that R is orthogonal. Indeed, if $w \in \Gamma$, it follows from (5.38) that

$$\langle u, w\rangle = 0, \tag{5.41}$$

where u is given in (5.37). Now, since u is an eigenvector for R corresponding to $\lambda = 1$, we have that

$$R(u) = u. \tag{5.42}$$

It follows from (5.41), (5.42) and the fact that R is orthogonal that

$$\langle u, R(w)\rangle = \langle R(u), R(w)\rangle = \langle u, w\rangle = 0,$$

which shows that $R(w) \in \Gamma$. Thus, we have shown that

$$w \in \Gamma \Rightarrow R(w) \in \Gamma,$$

which establishes (5.40).

Next, we construct a set $\{w_1, w_2\}$ of unit vectors in Γ that are also orthogonal to each other. First, we find a vector $v_1 = \begin{pmatrix} x \\ y \\ z \end{pmatrix} \in \Gamma$ with $x = 1$ and $y = 0$; thus, in view of (5.39), $z = 1$; so that

$$v_1 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}. \tag{5.43}$$

We then take

$$w_1 = \begin{pmatrix} 1/\sqrt{2} \\ 0 \\ 1/\sqrt{2} \end{pmatrix}; \tag{5.44}$$

that is, w_1 is a unit vector in the direction of v_1 in (5.43).

Next, we look for a vector $v_2 = \begin{pmatrix} x \\ y \\ z \end{pmatrix} \in \Gamma$ that is orthogonal to v_1 in (5.43). It then follows from (5.39) and (5.43) that

$$\left\{ \begin{array}{rcl} x + y - z & = & 0 \\ x \quad + \quad z & = & 0. \end{array} \right. \tag{5.45}$$

The system in (5.45) can be solved by reducing the augmented matrix

$$\left(\begin{array}{rrrcr} 1 & 1 & -1 & | & 0 \\ 1 & 0 & 1 & | & 0 \end{array} \right) \tag{5.46}$$

to

$$\left(\begin{array}{rrrcr} 1 & 0 & 1 & | & 0 \\ 0 & 1 & -2 & | & 0. \end{array} \right) \tag{5.47}$$

From the equivalence of the matrices in (5.46) and (5.47) it follows that the system in (5.45) is equivalent to the system

$$\left\{ \begin{array}{rcrcl} x & + & z & = & 0 \\ y & - & 2z & = & 0. \end{array} \right. \tag{5.48}$$

Solving the system in (5.48) yields a solution

$$v_2 = \begin{pmatrix} 1 \\ -2 \\ -1 \end{pmatrix}. \tag{5.49}$$

Thus, we can take

$$w_2 = \begin{pmatrix} 1/\sqrt{6} \\ -2/\sqrt{6} \\ -1/\sqrt{6} \end{pmatrix}, \tag{5.50}$$

the unit vector in the direction of v_2 in (5.49).

The vectors u, w_1 and w_2 in (5.37), (5.44) and (5.50), respectively, are mutually orthogonal unit vectors in $\mathbb{R}^3$; hence, the set

$$\mathcal{B} = \{u, w_1, w_2\} \tag{5.51}$$

forms a basis for $\mathbb{R}^3$ (see Problem 3 in Assignment #12).

Set

$$Q = [u \quad w_1 \quad w_2]; \tag{5.52}$$

that is, Q is the matrix whose columns are the vectors in the ordered basis $\mathcal{B}$ in (5.51).

Computing the determinant of Q in (5.52) we obtain

$$\begin{aligned}
\det(Q) &= \frac{1}{6}\begin{vmatrix} 1 & 1 & 1 \\ 1 & 0 & -2 \\ -1 & 1 & -1 \end{vmatrix} \\
&= \frac{1}{6}\left[-\begin{vmatrix} 1 & 1 \\ 1 & -1 \end{vmatrix} + 2\begin{vmatrix} 1 & 1 \\ -1 & 1 \end{vmatrix}\right] \\
&= \frac{1}{6}\left[-(-2) + 2(2)\right],
\end{aligned}$$

so that

$$\det(Q) = 1 > 0,$$

and therefore the basis $\mathcal{B}$ in (5.51) has a positive orientation.

Next, we find the matrix representation of R relative to the ordered basis $\mathcal{B}$ in (5.51).

We have already noted that $R(u) = u$, so that

$$R(u) = 1 \cdot u + 0 \cdot w_1 + 0 \cdot w_2,$$

and therefore, the coordinates of $R(u)$ relative to $\mathcal{B}$ are

$$[R(u)]_{\mathcal{B}} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}. \tag{5.53}$$

Next, we compute the coordinates of $R(w_1)$ and $R(w_2)$ relative to $\mathcal{B}$. First, note that, by virtue of (5.40), we can write

$$R(w_1) = aw_1 + cw_2 \tag{5.54}$$

and

$$R(w_2) = bw_1 + dw_2, \tag{5.55}$$

for some scalars a, b, c and d, where

$$\begin{aligned} R(w_1) &= M_R\, w_1 \\ &= \begin{pmatrix} 0 & 0 & -1 \\ 1 & 0 & 0 \\ 0 & -1 & 0 \end{pmatrix} \begin{pmatrix} 1/\sqrt{2} \\ 0 \\ 1/\sqrt{2} \end{pmatrix} \\ &= \begin{pmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \end{pmatrix}, \end{aligned} \tag{5.56}$$

and

$$\begin{aligned} R(w_2) &= M_R\, w_2 \\ &= \begin{pmatrix} 0 & 0 & -1 \\ 1 & 0 & 0 \\ 0 & -1 & 0 \end{pmatrix} \begin{pmatrix} 1/\sqrt{6} \\ -2/\sqrt{6} \\ -1/\sqrt{6} \end{pmatrix} \\ &= \begin{pmatrix} 1/\sqrt{6} \\ 1/\sqrt{6} \\ 2/\sqrt{6} \end{pmatrix}. \end{aligned} \tag{5.57}$$

Since w_1 and w_2 are unit vectors that are orthogonal to each other, we can use the result of Problem 3 in Assignment #12 to compute the scalars a, b, c and d in (5.54) and (5.55) to obtain

$$\begin{aligned} a &= \langle R(w_1), w_1\rangle \\ &= R(w_1)^T w_1 \\ &= \begin{pmatrix} -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \end{pmatrix} \begin{pmatrix} 1/\sqrt{2} \\ 0 \\ 1/\sqrt{2} \end{pmatrix}, \end{aligned}$$

so that

$$a = -\frac{1}{2}, \tag{5.58}$$

where we have used the result in (5.56). Similarly,

$$\begin{aligned} c &= \langle R(w_1), w_2\rangle \\ &= R(w_1)^T w_2 \\ &= \begin{pmatrix} -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \end{pmatrix} \begin{pmatrix} 1/\sqrt{6} \\ -2/\sqrt{6} \\ -1/\sqrt{6} \end{pmatrix}, \end{aligned}$$

so that

$$c = -\frac{\sqrt{3}}{2}, \tag{5.59}$$

where we have also used the definition of w_2 in (5.50). Putting (5.54), (5.58) and (5.59) together we obtain the coordinates of $R(w_1)$ relative to $\mathcal{B}$ to be

$$[R(w_1)]_{\mathcal{B}} = \begin{pmatrix} 0 \\ -1/2 \\ -\sqrt{3}/2 \end{pmatrix}. \tag{5.60}$$

Calculations similar to those leading to (5.60), using the results of (5.57) and (5.55) can be used to obtain

$$[R(w_2)]_{\mathcal{B}} = \begin{pmatrix} 0 \\ \sqrt{3}/2 \\ -1/2 \end{pmatrix}. \tag{5.61}$$

Combining (5.53), (5.60) and (5.61), we get that the matrix representation for R relative to the basis $\mathcal{B}$ is

$$[R]_{\mathcal{B}} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1/2 & \sqrt{3}/2 \\ 0 & -\sqrt{3}/2 & -1/2 \end{pmatrix} \tag{5.62}$$

Thus, the matrix representation for R relative to $\mathcal{B}$ is of the form

$$[R]_{\mathcal{B}} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{pmatrix}, \tag{5.63}$$

where, comparing (5.62) and (5.63), we see that R corresponds to a rotation around the line spanned by u through an angle θ such that

$$\cos\theta = -\frac{1}{2} \quad \text{and} \quad \sin\theta = -\frac{\sqrt{3}}{2}.$$

Thus, viewed from the frame of reference provided by the vectors u, w_1 and w_2 in $\mathcal{B}$, R is a rotation around the axis generated by the unit vector u through and angle $\theta = -\dfrac{2\pi}{3}$ or -120°.

Bibliography

[Mac95] George Mackiw. A note on the equality of the column and row rank of a matrix. *Mathematics Magazine*, 68(4):285–286, October 1995.

[Mes94] Robert Messer. *Linear Algebra: Gateway to Mathematics*. Harper-Collins College Publishers, 1994.

[PPR09] Bob Palais, Richard Palais, and Stephen Rodi. A disorienting look at Euler's theorem on the axis of a rotation. *American Mathematical Monthly*, 116(10):892–909, August 2009.

[TT70] Robert Thrall and Leonard Thornheim. *Vector Spaces and Matrices*. Dover, 1970.