Linear Algebra

Fourth Edition

Kunquan Lan
Ryerson University

Contents

Preface

This textbook is intended as an aid for students who are attending a course on Linear Algebra in universities and colleges. It is suitable for students who are tackling Linear Algebra for the first time and even those who have little prior knowledge in mathematics.

This is a self-contained textbook that contains full solutions to the questions in the exercises at the end of the book. The aim is to present the core of Linear Algebra in a clear way and provide many new and simple proofs for some fundamental results. For example, we give a very simple proof for the well-known Cauchy-Schwarz inequality. The goal is to enhance students' abilities of analyzing, solving, and generalizing problems.

The material covered in this textbook is well organized and can be easily understood by students. The main topics include vectors, matrices, determinants, linear systems, linear transformations in the Euclidean spaces, bases and dimensions of spanning spaces in the Euclidean spaces, eigenvalues and diagonalizability, subspaces of the Euclidean spaces, and general vector spaces. When we discuss linear transformations, bases, and dimensions, we restrict our attention to the Euclidean spaces, but the ideas and methods involved can be extrapolated to general vector spaces.

This textbook is based on the third edition of Linear Algebra. Some material has been rewritten for greater clarity and some sections and chapters have been combined for better understanding. New examples and exercises have been added. One new chapter introducing vector spaces has also been added. In particular, subspaces of the Euclidean spaces are discussed in more detail.

I would like to thank Tim Robinson, Melody Vincent, Jared Steuernol, and Corina Wilshire from Pearson for their support during production of this textbook.

Chapter 1 Euclidean spaces

1.1 Euclidean spaces

We denote by $\mathbb{R}$ the set of real numbers and by $\mathbb{N}$ the set of positive integers, that is, $\mathbb{N} = \{1, 2, \cdots\}.$ If *x* is an element in $\mathbb{R},$ we write $x \in \mathbb{R}$ where the symbol $\in$ means "belongs to." For example, $2 \in \mathbb{R}$ means that 2 is an element in $\mathbb{R},$ that is, 2 is a real number. Let $n \in \mathbb{N}$ and $I_n = \left\{1, 2, \cdots, n\right\}.$

Definition 1.1.1.

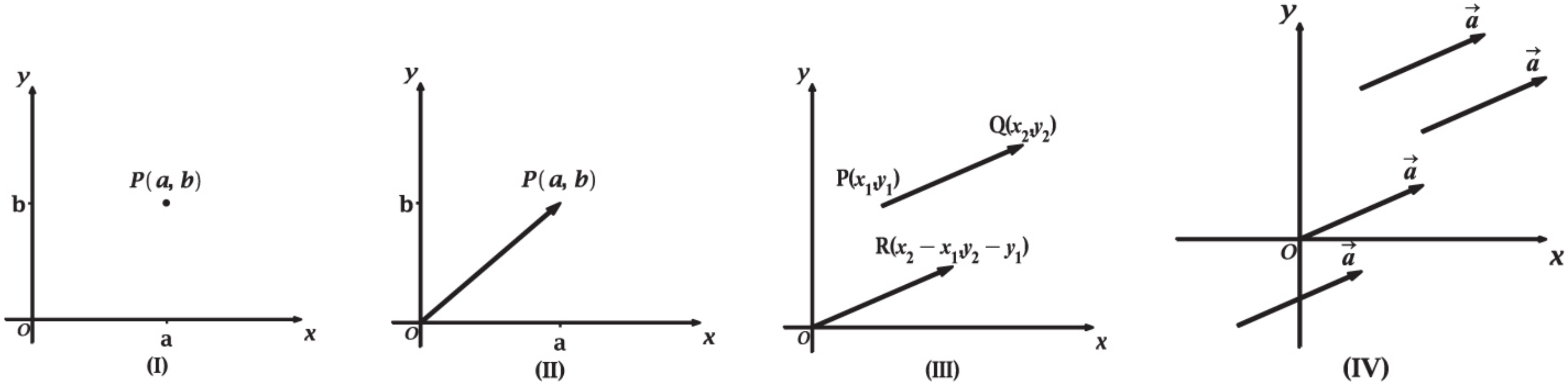
The set of all elements with *n*-ordered numbers is called $\mathbb{R}^n$-space ($\mathbb{R}^n$ for short) or an n-dimensional Euclidean space and denoted by

$$\mathbb{R}^n = \left\{(x_1, x_2, \cdots, x_n) : x_i \in \mathbb{R} \quad \text{and } i \in \mathrm{I}_n\right\}.$$

In geometry, $\mathbb{R}^1$ represents the *x*-axis and we write $\mathbb{R}^1 = \mathbb{R};$ $\mathbb{R}^2$ represents the *xy*-plane and $\mathbb{R}^3$ is the *xyz*-space. There is no geometry to show $\mathbb{R}^n$ for $n \geq 4,$ but these spaces exist. For example, we can denote a particle moving in *xyz*-space by a 4-dimensional space $\{(x, y, x, t) : x, y, t \in \mathbb{R}\},$ where (*x*, *y*, *z*, *t*) denotes the position of the particle at time *t*.

The element $(x_1, x_2, \cdots, x_n)$ is called a point in $\mathbb{R}^n$ and $x_1, x_2, \cdots, x_n$are called the coordinates of the point in $\mathbb{R}^n.$ We use symbols like $P(x_1, x_2, \cdots, x_n)$ or *P* to denote the point $(x_1, x_2, \cdots, x_n)$ (see **Figure 1.1** (I)). In particular, we use the symbol $O(0,\ 0, \cdots, 0)$ or O to denote the origin $(0,\ 0,\ \cdots,\ 0)$ of $\mathbb{R}^n$.

Figure 1.1: (I), (II), (III), (IV)



Definition 1.1.2.

Let $P(x_1,\ x_2,\ \cdots,\ x_n)$ be point in $\mathbb{R}^n$. The directed line segment starting at the origin *O* and ending at the point *P* is called a vector or *n*-vector, and is denoted by $\overrightarrow{OP}$ (see **Figure 1.1** (II)). The distance between the points *O* and *P* is said to be the length of the vector. If *P* is the origin, then $\overrightarrow{OP}$ is called the zero vector and if *P* is not the origin, then $\overrightarrow{OP}$ is called a nonzero vector.

We denote a vector by $\vec{a},\ \vec{u},$ and so on. If $\vec{a} = \left(x_1,\ x_2,\ \cdots,\ x_n\right),$ then x_i is called the *i*th component (*i*th entry) of the vector for each $i\ \in\ I_n$.

A vector $\vec{a}$ in $\mathbb{R}^n$ can be written into either a column form, called a column vector (or *n*-column vector), or a row form, called a column vector (or *n*-column vector), that is

$$\vec{a} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \quad \text{or } \vec{a} = \left(x_1, x_2, \cdots, x_n\right).$$

For example, $(0, 0, 0)^T$ is a column vector (or 3-vector), $\vec{a} = (5)$ is a row (or column) vector (or 1-vector); $\vec{a} = (1, 2)$ is a row vector (or 2-vector). We can write

$$\begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix} = (1, 2, 3, 4).$$

Note that two vectors with the same components written in different orders may not be the same. For example, the column vectors $(1, 2, 3)^T$ and $(3, 2, 1)^T$ are different vectors. A vector whose components are all zero is a zero vector while a nonzero vector has at least one nonzero component. For example, (0, 0, 0) is a zero vector in $\mathbb{R}^3$ while (1, 0, 0) and (1, 2, 3) are nonzero vectors in

Let

(1.1.1)

$$\vec{e_1} = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \vec{e_2} = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \cdots, \vec{e_n} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}$$

be vectors in $\mathbb{R}^n$. Then each vector has *n* components. These vectors $\vec{e_1}, \vec{e_2}, \cdots, \vec{e_n}$ are called the standard vectors in $\mathbb{R}^n$. In particular, $\vec{e_1} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $\vec{e_2} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ are the standard vectors in $\mathbb{R}^2$ and $\vec{e_1} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \vec{e_2} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$ and $\vec{e_3} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$ are the standard vectors in $\mathbb{R}^3$.

The difference between a point $P(x_1, x_2, \cdots, x_n)$ and a vector $\vec{a} = (x_1, x_2, \cdots, x_n)^T$ is that a point *P* has neither a direction nor a length while a vector $\vec{a}$ has both a direction, which starts at the origin and ends at the point *P*, and a length, which is the distance between *O* and *P*.

In **Definition 1.1.2** , a vector is defined as a directed line segment starting at the origin. The following definition introduces the notion of a vector whose initial point is not necessarily at the origin.

Definition 1.1.3.

Let $P(x_1, x_2, \cdots, x_n)$ and $Q(y_1, y_2, \cdots, y_n)$ be points in $\mathbb{R}^n$. Let *R* denote the point $(y_1 - x_1, y_2 - x_2, \cdots, y_n - x_n)$. The directed line segment starting at *P* and ending at *Q* is called a vector and is defined by

(1.1.2)

$$\overrightarrow{PQ} = \overrightarrow{OR} = (y_1 - x_1,\ y_2 - x_2,\ \cdots,\ y_n - x_n),$$

see **Figure 1.1** (III).

By **Definition 1.1.3**, we see that the vector $\overrightarrow{PQ}$ has different initial and terminating points than the vector $\overrightarrow{OR}$, but its mathematical expression is the same as $\overrightarrow{OR}$; both are $(y_1 - x_1,\ y_2 - x_2,\ \cdots,\ y_n - x_n)$.

Hence, if you move a nonzero vector $\vec{a}$ anywhere in the *xy*-plane without changing its direction, the resulting vector is equal to $\vec{a}$ (see **Figure 1.1** (IV)). The same vector may have different initial and terminal points.

Mathematically, **Definition 1.1.3** does not introduce any new vectors. This establishes a one-to-one relation between points and vectors in $\mathbb{R}^n$ that is, for each point $P(x_1,\ x_2,\ \cdots,\ x_n)$ in $\mathbb{R}^n$, there is a unique vector $\overrightarrow{OP}$ corresponding to the point $P(x_1,\ x_2,\ \cdots,\ x_n)$. Conversely, for each vector $\vec{a} = (x_1,\ x_2,\ \cdots,\ x_n)$, there is a unique point $(x_1,\ x_2,\ \cdots,\ x_n)$ corresponding to the vector $\vec{a}$. Hence, the space $\mathbb{R}^n$ defined in **(1.1.1)** can be treated as the set of all vectors in $\mathbb{R}^n$.

Example 1.1.1.

1. Let $P(1,\ 2)$, $Q(3,\ 4)$, $P_1(0,\ -2,\ 3)$, and $Q_1(1,\ 0,\ 2)$. Find $\overrightarrow{PQ}$ and $\overrightarrow{P_1Q_1}$.

Solution

By **(1.1.2)**, $\overrightarrow{PQ} = (3-1,\ 4-2) = (2,\ 2)$ and $\overrightarrow{P_1Q_1} = (1-0,\ 0-(-2),\ 2-3) = (1,\ 2,\ 1)$.

Operations on vectors

For real numbers, there are four common operations: addition, subtraction, multiplication, and division. In this section, we generalize the first two operations: addition and subtraction to $\mathbb{R}^n$ in a natural way, and introduce scalar multiplication and the dot product of two vectors in $\mathbb{R}^n$. The dot product is a real number–not a vector. The multiplication and division of two vectors introduced in a natural way are not useful, so we do not introduce such multiplication and division operations for two vectors. However, in **Section 6.1**, we introduce a cross product of two vectors in $\mathbb{R}^3$, which is a vector in $\mathbb{R}^3$.

We start with equal vectors in $\mathbb{R}^n$.

> **Definition 1.1.4.**
>
> Two vectors are said to be equal if the following two conditions are satisfied:
>
> i. The number of components of the two vectors are the same.
> ii. The corresponding components of the two vectors are equal.

If $\vec{a}, \vec{b}$ are equal, we write $\vec{a} = \vec{b}$. Otherwise, we write $\vec{a} \neq \vec{b}$.

In **Definition 1.1.3**, we see that the two vectors $\overrightarrow{PQ}$ and $\overrightarrow{OR}$ in **(1.1.2)** are equal although they have different initial and terminating points.

Let $m, n \in \mathbb{N}$ and let

$$\vec{a} = (a_1, a_2, \cdots, a_m) \quad \text{and} \quad \vec{b} = (b_1, b_2, \cdots, b_n).$$

According to **Definition 1.1.4**, $\vec{a} = \vec{b}$ if and only if $m = n$, that is, $\vec{a}$ and $\vec{b}$ must be in the same space, and $a_i = b_i$ for each $i \in I_n$. Moreover, $\vec{a} \neq \vec{b}$ if and only if either $m \neq n$ or there exists some $i \in I_n$ such that $a_i \neq b_i$

Example 1.1.2.

1. Let $\vec{a} = \begin{pmatrix} x+y \\ x-y \\ z \end{pmatrix}$ and $\vec{b} = \begin{pmatrix} 1 \\ 3 \\ 1 \end{pmatrix}$. Find all $x, \ y, \ z \in \mathbb{R}$ such that $\vec{a} = \vec{b}$.

2. Let $\vec{a} = \begin{pmatrix} 1 \\ x^2 \end{pmatrix}$ and $\vec{b} = \begin{pmatrix} 1 \\ 4 \end{pmatrix}$. Find all $x \in \mathbb{R}$ such that $\vec{a} \neq \vec{b}$.

3. $\vec{a} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$ and $\vec{b} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$. Identify whether $\vec{a}$ and $\vec{b}$ are equal.

4. Let $P(1, \ 1), \ Q(3, \ 4), \ P_1(-2, \ 3),$ and $Q_1(0, 6)$ be points in $\mathbb{R}^2$. Show that $\overrightarrow{PQ} = \overrightarrow{P_1Q_1}$.

Solution

1. Because the number of components of $\vec{a}$ and $\vec{b}$ are equal, by **Definition 1.1.4** , in order to make $\vec{a} = \vec{b},$ we need the corresponding components of $\vec{a}$ and $\vec{b}$ to be equal, that is, *x*, *y*, *z* satisfy the following system of equations:

$$\begin{cases} x+y=1 \\ x-y=3 \\ z=1. \end{cases}$$

Solving the above system, we get $x=2, \ y=-1$ and $z=1$. Hence, when $x=2, \ y=-1,$ and $z=1, \ \vec{a}=\vec{b}$.

2. Because the number of components of $\vec{a}$ and $\vec{b}$ are equal, by **Definition 1.1.4** , if $x^2 \neq 4,$ than $\vec{a} \neq \vec{b}$. Solving $x^2 \neq 4,$ we get $x \neq 2$ and $x \neq -2$. Hence, when $x \neq 2$ and $x \neq -2, \ \vec{a} \neq \vec{b}$.

3. Because the number of components of $\vec{a}$ and $\vec{b}$ are 3 and 2, respectively, it follows from **Definition 1.1.4** that $\vec{a} \neq \vec{b}$.

4. By **(1.1.2)**, we have $\overrightarrow{PQ} = (3-1,\ 4-2) = (2,\ 3)$ and $\overrightarrow{P_1Q_1} = (0-(-2),\ 6-3) = (2, 3)$. Hence, $\overrightarrow{PQ} = \overrightarrow{P_1Q_1}$.

Definition 1.1.5.

Let $\vec{a} = (a_1,\ a_2,\ \cdots,\ a_n)$ and $\vec{b} = (b_1,\ b_2,\ \cdots,\ b_n)$ We define the following operations in a natural way.

Addition: $\vec{a} + \vec{b} = (a_1 + b_1,\ a_2 + b_3,\ \cdots,\ a_n + b_n)$.

Subtraction: $\vec{a} - \vec{b} = (a_1 - b_1,\ a_2 - b_3,\ \cdots,\ a_n - b_n)$.

Scalar multiplication: $k\vec{a} = (kx_1,\ kx_2,\ \cdots,\ kx_n)$

Remark 1.1.1.

The addition and subtraction of vectors are introduced in the same space $\mathbb{R}^n$. If two vectors are in different spaces, say, one is in $\mathbb{R}^m$ and another in $\mathbb{R}^n$, where $m \neq n$, then the addition and subtraction of the two vectors are not defined. For example, $\begin{pmatrix} 1 \\ 2 \end{pmatrix} + \begin{pmatrix} 3 \\ 3 \\ 5 \end{pmatrix}$ is not defined.

When $n = 2$, the addition of two vectors is illustrated in **Figure 1.2** . If $\vec{a} = (x_1, y_1)$ and $\vec{b} = (x_2, y_2)$ then **Figure 1.2** (I) shows that

$$\vec{a} + \vec{b} = (x_1 + x_2, y_1 + y_2).$$

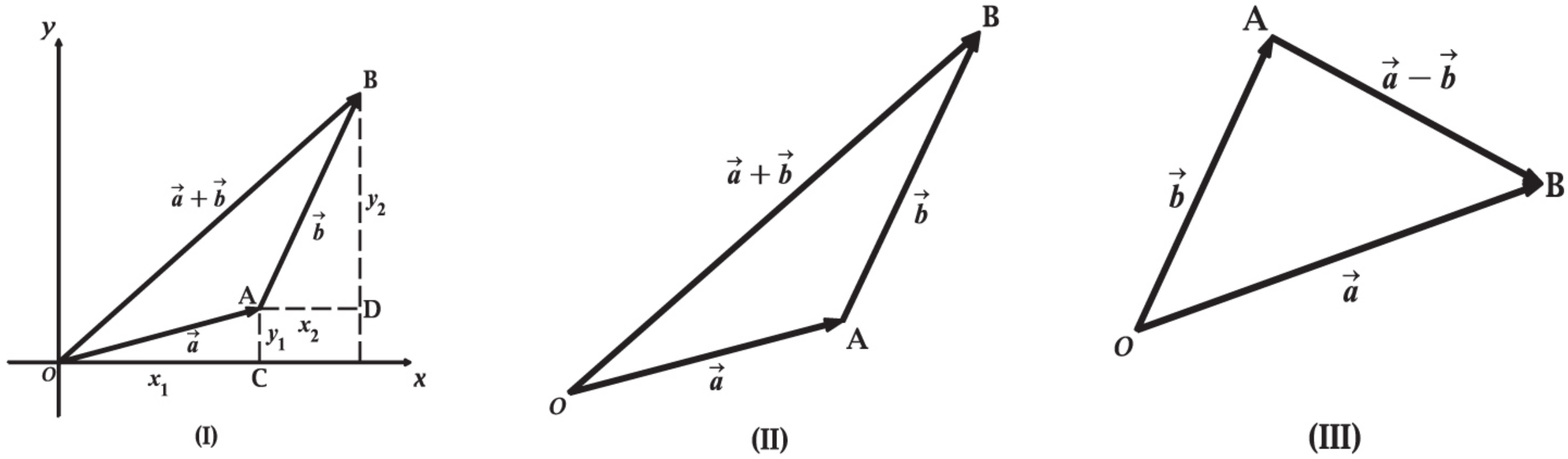
Figure 1.2 (II) shows that if we place the initial point of the vector $\vec{b}$ on the terminating point of the vector $\vec{a}$, then the vector $\vec{a} + \vec{b}$ is the vector starting at the initial point of the vector $\vec{a}$ and ending at the terminating point of vector $\vec{b}$. In other words, the vector $\overrightarrow{OA}$ plus the vector $\overrightarrow{AB}$ is equal to the vector $\overrightarrow{OB}$.

(1.1.3)

$$\overrightarrow{OA} + \overrightarrow{AB} = \overrightarrow{OB}.$$

The subtraction of two vectors is illustrated in **Figure 1.2** (III), which can be derived from **Figure 1.2** (II) by the relation **(1.1.3)** .

Figure 1.2: (I), (II), (III)



By **Definition 1.1.5**, the following results on the vector operations can be easily proved, so we leave the proofs to the reader.

Theorem 1.1.1.

Let $\vec{a},\ \vec{b},\ \vec{c} \in \mathbb{R}^n$ *and let* $\alpha,\ \beta \in \mathbb{R}$. *Then*

1. $\vec{a} + \vec{0} = \vec{a}$.
2. $0\vec{a} = \vec{0}$.
3. $\vec{a} + \vec{b} = \vec{b} + \vec{a}$.
4. $(\vec{a} + \vec{b}) + \vec{c} = \vec{a} + \overrightarrow{(b + \vec{a})}$.
5. $\alpha(\vec{a} + \vec{b}) = \alpha\vec{a} + \alpha\vec{b}$.
6. $(\alpha + \beta)\vec{a} = \alpha\vec{a} + \beta\vec{a}$.
7. $(\alpha\beta)\vec{a} = \alpha(\beta\vec{a})$.

Example 1.1.3.

Let $\vec{a} = \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}$ and $\vec{b} = \begin{pmatrix} 3 \\ 6 \\ 5 \end{pmatrix}$. Compute

i. $\vec{a} + \vec{b}$;
ii. $\vec{a} - \vec{b}$;
iii. $5\vec{a}$;
iv. $-\vec{a}$;
v. $2\vec{a} - 4\vec{b}$.

Solution

By **Definition 1.1.5**, we obtain

i. $\vec{a}+\vec{b}=\begin{pmatrix}1\\2\\0\end{pmatrix}+\begin{pmatrix}3\\6\\5\end{pmatrix}=\begin{pmatrix}1+3\\2+6\\0+5\end{pmatrix}=\begin{pmatrix}4\\8\\5\end{pmatrix}.$

ii. $\vec{a}-\vec{b}=\begin{pmatrix}1\\2\\0\end{pmatrix}-\begin{pmatrix}3\\6\\5\end{pmatrix}=\begin{pmatrix}1-3\\2-6\\0-5\end{pmatrix}=\begin{pmatrix}-2\\-4\\-5\end{pmatrix}.$

iii. $5\vec{a}=5\begin{pmatrix}1\\2\\0\end{pmatrix}=\begin{pmatrix}(5)(1)\\(5)(2)\\(5)(0)\end{pmatrix}=\begin{pmatrix}5\\10\\0\end{pmatrix}.$

iv. $-\vec{a}=(-1)\begin{pmatrix}1\\2\\0\end{pmatrix}=\begin{pmatrix}(-1)(1)\\(-1)(2)\\(-1)(0)\end{pmatrix}=\begin{pmatrix}-1\\-2\\0\end{pmatrix}.$

v. $2\vec{a}-4\vec{b}=2\begin{pmatrix}1\\2\\0\end{pmatrix}-4\begin{pmatrix}3\\6\\5\end{pmatrix}=\begin{pmatrix}2\\4\\0\end{pmatrix}-\begin{pmatrix}12\\24\\20\end{pmatrix}=\begin{pmatrix}-10\\-20\\-20\end{pmatrix}.$

Example 1.1.4.

Let $\vec{a}=\begin{pmatrix}1\\2\\0\end{pmatrix}$, $\vec{b}=\begin{pmatrix}3\\6\\5\end{pmatrix}$ and $\vec{c}=\begin{pmatrix}4\\6\\7\end{pmatrix}$. Compute $\vec{a}-\vec{b}+2\vec{c}$.

Solution

By **Theorem 1.1.1**, we have

$$\vec{a} - \vec{b} + 2\vec{c} = (\vec{a} - \vec{b}) + 2\vec{c} = \left[\begin{pmatrix}1\\2\\0\end{pmatrix} - \begin{pmatrix}3\\6\\5\end{pmatrix}\right] + 2\begin{pmatrix}4\\6\\7\end{pmatrix} = \begin{pmatrix}6\\8\\9\end{pmatrix}.$$

Example 1.1.5.

Find $\vec{a}$ if $\frac{1}{5}\left[4\vec{a} - \begin{pmatrix}9\\3\end{pmatrix}\right] = \begin{pmatrix}1\\5\end{pmatrix} - 2\vec{a}$.

Solution

From the given equation, we see that $\vec{a} \in \mathbb{R}^2$ and

$$4\vec{a} - \begin{pmatrix}9\\3\end{pmatrix} = 5\left[\begin{pmatrix}1\\5\end{pmatrix} - 2\vec{a}\right] = \begin{pmatrix}5\\25\end{pmatrix} - 10\vec{a}.$$

Hence,

$4\vec{a} + 10\vec{a} = \begin{pmatrix}5\\25\end{pmatrix} + \begin{pmatrix}9\\3\end{pmatrix} = \begin{pmatrix}14\\28\end{pmatrix}$ and $14\vec{a} = \begin{pmatrix}14\\28\end{pmatrix}$. Hence, $\vec{a} = \frac{1}{14}\begin{pmatrix}14\\28\end{pmatrix} = \begin{pmatrix}1\\2\end{pmatrix}$.

Parallel vectors

Definition 1.1.6.

Two nonzero vectors $\vec{a}$ and $\vec{b}$ in $\mathbb{R}^n$ are said to be parallel if there exists a number $k \in \mathbb{R}$ such that $\vec{a} = k\vec{b}$. When $k > 0$, $\vec{a}$ and $\vec{b}$ have the same direction and when $k < 0$, their directions are opposite.

For example, because $(2, 4, 6) = 2(1, 2, 3)$, the two vectors (2, 4, 6) and (1, 2, 3) are parallel and have the same direction. Similarly, because $(-2, -4, -6) = -2(1, 2, 3)$, the two vectors $(-2, -4, -6)$ and (1, 2, 3) are parallel and have the opposite direction.

If $\vec{a}$ and $\vec{b}$ in $\mathbb{R}^n$ are parallel, we write

$$\vec{a} \parallel \vec{b}.$$

Example 1.1.6.

Let $P(1, -2, -3)$, $Q(2, 0, -1)$, $P_1(-2, -7, -4)$, and $Q_1(1, -1, 2)$. Show that $\overrightarrow{PQ} \parallel \overrightarrow{QP}$.

Solution

By **(1.1.2)** , $\overrightarrow{PQ} = (2-1, 0-(-2), -1-(-3)) = (1, 2, 2)$ and

$$\overrightarrow{P_1Q_1} = (1-(-2), -1-(-7), 2-(-4)) = (3, 6, 6).$$

It follows that $\overrightarrow{P_1Q_1} = (3, 6, 6) = 3(1, 2, 2) = 3\overrightarrow{PQ}$. By **Definition 1.1.6** , $\overrightarrow{PQ} \parallel \overrightarrow{QP}$.

Dot product

The scalar multiplication, addition, and subtraction of two vectors in $\mathbb{R}^n$ result in vectors in $\mathbb{R}^n$. We don't define the multiplication and division of $\vec{a}$ and $\vec{b}$ in the natural way as $(a_1b_1, a_2b_2, \cdots, a_nb_n)$ and

$\left(\frac{a_1}{b_1}, \frac{a_2}{b_2}, \cdots, \frac{a_n}{b_n}\right)$ because they are not useful. But we introduce the notion of the product of two vectors, called the dot product, which can be seen to be useful later. The dot product of two vectors is a real number–not a vector in $\mathbb{R}^n$. In **Section 6.1** , we shall introduce another product of two vectors in $\mathbb{R}^3$ called the cross product of two vectors, which is a vector in $\mathbb{R}^3$.

Definition 1.1.7.

Let $\vec{a} = (a_1, a_2, \cdots, a_n)$ and $\vec{b} = (b_1, b_2, \cdots, b_n)$. The dot product of $\vec{a}$ and $\vec{b}$ denoted by $\vec{a} \cdot \vec{b}$, is defined by

(1.1.4)

$$\vec{a} \cdot \vec{b} = a_1b_1 + a_2b_2 + \cdots + a_nb_n.$$

It is convenient to write the dot product of $\vec{a}$ and $\vec{b}$ in the following row times column form:

(1.1.5)

$$\vec{a} \cdot \vec{b} = (a_1, a_2, \cdots, a_n)\begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} = a_1b_1 + a_2b_2 + \cdots + a_nb_n.$$

Hence, $\vec{a} \cdot \vec{b}$ is the sum of the products of the corresponding components of $\vec{a}$ and $\vec{b}$. Sometimes, we use sigma notation to write $\vec{a} \cdot \vec{b}$, that is,

(1.1.6)

$$\vec{a} \cdot \vec{b} = \sum_{i=1}^{n} a_i b_i.$$

Example 1.1.7.

Let $\vec{a} = (-1,\ 3,\ 5,\ -7)$ and $\vec{b} = (5,\ -4,\ 7,\ 0)$. Find $\vec{a} \cdot \vec{b}$.

Solution

$$\vec{a} \cdot \vec{b} = (-1)(5) + (3)(-4) + (5)(7) + (-7)(0) = 18.$$

The following result follows directly from **Definition 1.1.7** .

Theorem 1.1.2.

Let $\vec{a},\ \vec{b},\ \vec{c} \in \mathbb{R}^n$ *and let* $\alpha \in \mathbb{R}$. *Then*

1. $\vec{a} \cdot \vec{0} = 0.$
2. $\vec{a} \cdot \vec{b} = \vec{b} \cdot \vec{a}.$ *(Commutative law for dot product)*
3. $\vec{a} \cdot (\vec{b} + \vec{c}) = \vec{a} \cdot \vec{b} + \vec{a} \cdot \vec{c}.$ *(Distributive law for dot product)*
4. $(\alpha\vec{a}) \cdot \vec{b} = \vec{a} \cdot (\alpha\vec{b}) = \alpha(\vec{a} \cdot \vec{b}).$

Example 1.1.8.

Let $\vec{a} = \begin{pmatrix} 1 \\ 0 \\ 4 \\ 2 \end{pmatrix}$, $\vec{b} = \begin{pmatrix} 2 \\ 5 \\ 0 \\ 1 \end{pmatrix}$, and $\vec{c} = \begin{pmatrix} 1 \\ 2 \\ 0 \\ 1 \end{pmatrix}$. Calculate $\vec{a} \cdot (\vec{b} + \vec{c})$.

Solution

$$\begin{aligned} \vec{a} \cdot (\vec{b} + \vec{c}) &= \vec{a} \cdot \vec{b} + \vec{a} \cdot \vec{c} = (1,\ 0,\ 4,\ 2)\begin{pmatrix} 2 \\ 5 \\ 0 \\ 1 \end{pmatrix} + (1,\ 0,\ 4,\ 2)\begin{pmatrix} 1 \\ 2 \\ 0 \\ 1 \end{pmatrix} \\ &= 4 + 3 = 7. \end{aligned}$$

Exercises

1. Write a 3-column zero vector and a 3-column nonzero vector, and a 5-column zero vector and a 5-column nonzero vector.
2. Suppose that the buyer for a manufacturing plant must order different quantities of oil, paper, steel, and plastics. He will order 40 units of oil, 50 units of paper, 80 units of steel, and 20 units of plastics. Write the quantities in a single vector.
3. Suppose that a student's course marks for quiz 1, quiz 2, test 1, test 2, and the final exam are 70, 85, 80, 75, and 90, respectively. Write his marks as a column vector.
4. Let $\vec{a} = \begin{pmatrix} x - 2y \\ 2x - y \\ 2z \end{pmatrix}$ and $\vec{b} = \begin{pmatrix} 2 \\ -2 \\ 1 \end{pmatrix}$. Find all $x,\ y,\ z \in \mathbb{R}$ that $\vec{a} = \vec{b}$.

5. $\vec{a} = \begin{pmatrix} |x| \\ y^2 \end{pmatrix}$ and $\vec{b} = \begin{pmatrix} 1 \\ 4 \end{pmatrix}$. Find all $x, \ y \in \mathbb{R}$ such that $\vec{a} = \vec{b}$.

6. $\vec{a} = \begin{pmatrix} x - y \\ 4 \end{pmatrix}$ and $\vec{b} = \begin{pmatrix} 2 \\ x + y \end{pmatrix}$. Find all $x, \ y \in \mathbb{R}$ such that $\vec{a} - \vec{b}$ is a nonzero vector.

7. Let $P(1, \ x^2), \ Q(3, \ 4), \ P_1(4, \ 5),$ and $Q_1(6, \ 1).$ Find all possible values of $x \in \mathbb{R}$ such that $\overrightarrow{PQ} = \overrightarrow{P_1Q_1}$.

8. A company with 553 employees lists each employee's salary as a component of a vector $\vec{a}$ in $\mathbb{R}^{553}$. If a 6% salary increase has been approved, find the vector involving $\vec{a}$ that gives all the new salaries.

9. Let $\vec{a} = \begin{pmatrix} 110 \\ 88 \\ 40 \end{pmatrix}$ denote the current prices of three items at a store. Suppose that the store announces a sale so that the price of each item is reduced by 20%.
   a. Find a 3-vector that gives the price changes for the three items.
   b. Find a 3-vector that gives the new prices of the three items.

10. Let $\vec{a} = \begin{pmatrix} -1 \\ 2 \\ 3 \end{pmatrix}, \ \vec{b} = \begin{pmatrix} 2 \\ -2 \\ 0 \end{pmatrix}, \ \vec{c} = \begin{pmatrix} 3 \\ 0 \\ -1 \end{pmatrix}, \ \alpha \in \mathbb{R}.$ Compute
    i. $2\vec{a} - \vec{b} + 5\vec{c};$
    ii. $4\vec{a} + \overrightarrow{\alpha b} - 2\vec{c}$

11. Find *x*, *y*,and *z* such that $\begin{pmatrix} 9 \\ 4y \\ 2z \end{pmatrix} + \begin{pmatrix} 3x \\ 8 \\ -6 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$.

12. Let $\vec{u} = (1,\ -1,\ 0,\ 2),\ \vec{v} = (-2,\ -1,\ 1,\ -4),$ and $\vec{w} = (3,\ 2,\ -1,\ 0).$

    a. Find a vector $\vec{a} \in \mathbb{R}^4$ such that $2\vec{u} - 3\vec{v} - \vec{a} = \vec{w}.$

    b. Find a vector $\vec{a}$

$$\frac{1}{2}[2\vec{u} - 3\vec{v} + \vec{a}] = 2\vec{w} + \vec{u} - 2\vec{a}.$$

13. Determine whether $\vec{a} \parallel \vec{b},$

    1. $\vec{a} = (1,\ 2,\ 3)$ and $\vec{b} = (-2,\ -4,\ -6).$
    2. $\vec{a} = (-1,\ 0,\ 1,\ 2)$ and $\vec{b} = (-3,\ 0,\ 3,\ 6).$
    3. $\vec{a} = (1,\ -1,\ 1,\ 2)$ and $\vec{b} = (-2,\ 2,\ 2,\ 3).$

14. Let $x,\ y,\ a,\ b \in \mathbb{R}$ with $x \neq y.$ Let $P(x,\ 2x),\ Q(y,\ 2y),\ P_1(a,\ 2a),$ and $Q_1(b,\ 2b)$ be points in $\mathbb{R}^2.$ Show that $\overrightarrow{PQ} \parallel \overrightarrow{P_1Q_1}.$

15. Let $P(x,\ 0),\ Q(3,\ x^2),\ P_1(2x,\ 1),$ and $Q_1(6,\ x^2).$ Find all possible values of $x \in \mathbb{R}$ such that $\overrightarrow{PQ} \parallel \overrightarrow{P_1Q_1}.$

16. Let $\vec{a} = \begin{pmatrix} -1 \\ 2 \\ 3 \end{pmatrix},\ \vec{b} = \begin{pmatrix} 2 \\ -2 \\ 0 \end{pmatrix},\ \vec{c} = \begin{pmatrix} 3 \\ 0 \\ -1 \end{pmatrix},$ and $\alpha \in \mathbb{R}.$

    Compute

    a. $\vec{a} \cdot \vec{b};$

    b. $\vec{a} \cdot \vec{c};$

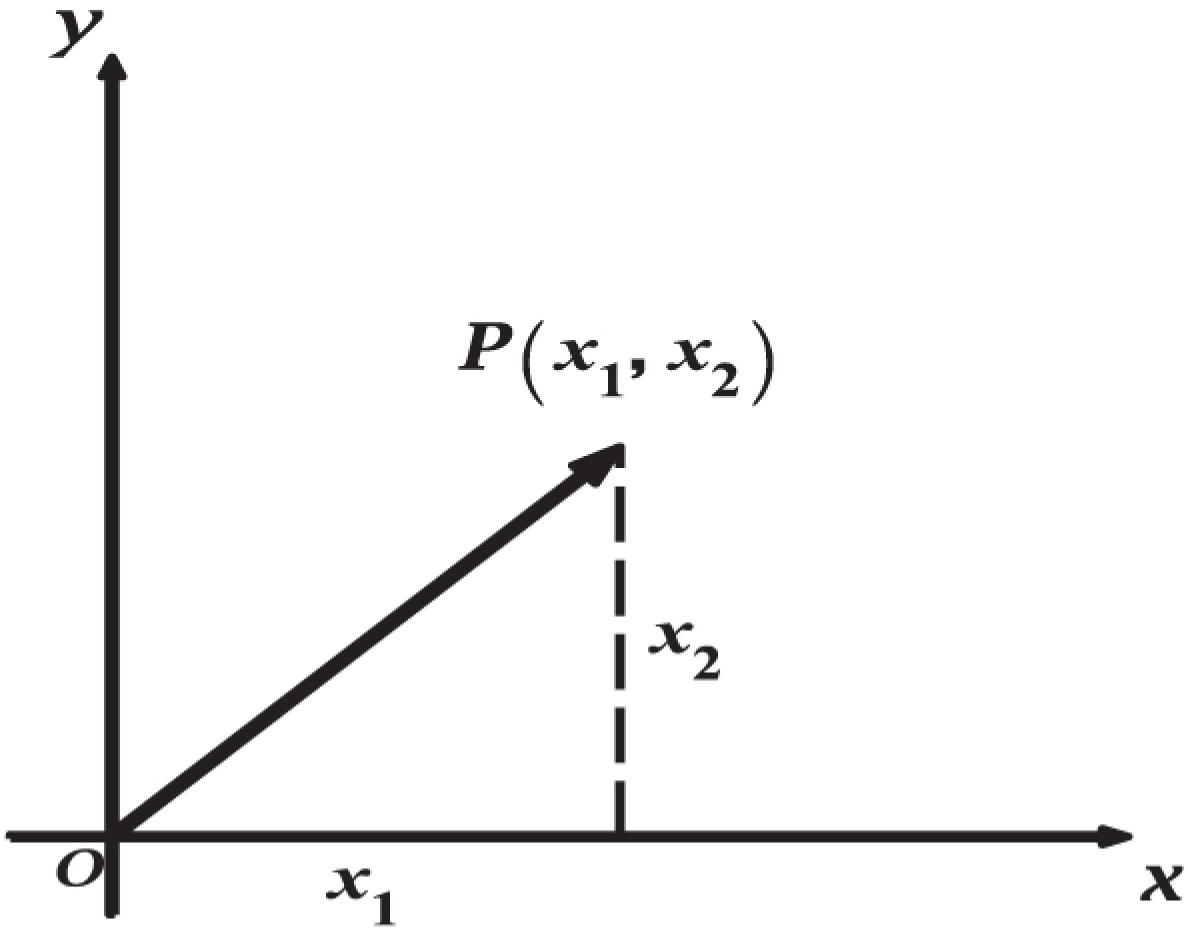
    c. $\vec{b} \cdot \vec{c};$

    d. $\vec{a} \cdot (\vec{b} + \vec{c}).$

17. Let $\vec{u} = (1,\ -1,\ 0,\ 2),\ \vec{v} = (-2,\ -1,\ 1,\ -4),$ and $\vec{w} = (3,\ 2,\ -1,\ 0).$
    Compute $\vec{u} \cdot \vec{v},\ \ \vec{u} \cdot \vec{w},$ and $\vec{w} \cdot \vec{v}.$

18. Assume that the percentages for homework, test 1, test 2, and the final exam for a course are 10%, 25%, 25%, and 40%, respectively. The total marks for homework, test 1, test 2 and the final exam are 10, 50, 50, and 90, respectively. A student's corresponding marks are 8, 46, 48, and 81, respectively. What is the student's final mark out of 100?
19. Assume that the percentages for homework, test 1, test 2, and the final exam for a course are 14%, 20%, 20%, and 46%, respectively. The total marks for homework, test 1, test 2, and the final exam are 14, 60, 60, and 80, respectively. A student's corresponding marks are 12, 45, 51, and 60, respectively. What is the student's final mark out of 100?
20. A manufacturer produces three different types of products. The demand for the products is denoted by the vector $\vec{a} = (10,\ 20,\ 30)$. The price per unit for the products is given by the vector $\vec{b} = (\$200,\ \$150,\ \$100)$. If the demand is met, how much money will the manufacturer receive?
21. A company pays four groups of employees a salary. The numbers of the employees for the four groups are expressed by a vector $\vec{a} = (5,\ 20,\ 40)$. The payments for the groups are expressed by a vector $\vec{b} = (\$100,\ 000,\ \$80,\ 000,\ \$60,\ 000)$. Use the dot product to calculate the total amount of money the company paid its employees.

Figure 1.3:



22. There are three students who may buy a Calculus or Algebra book. Use the dot product to find the total number of students who buy both Calculus and Algebra books.
23. There are *n* students who may buy a Calculus or Algebra book $(n \geq 2)$. Use the dot product to find the total number of students who buy both Calculus and Algebra books.
24. Assume that a person *A* has contracted a contagious disease and has direct contacts with four people: *P*1, *P*2, *P*3, and *P*4. We denote the contacts by a vector $\vec{a} := (a_{11}, a_{12}, a_{13}, a_{14})$, where if the person *A* has made contact with the person *Pj*, then $a_{1j} = 1$, and if the person *A* has made no contact with the person *Pj*, then $a_{1j} = 0$. Now we suppose that the four people then have had a variety of direct contacts with another individual *B*, which we denote by a vector $\vec{b} := (b_{11}, b_{21}, b_{31}, b_{41})$,

where if the person *Pj* has made contact with the person *B*, then $b_{j1} = 1,$ and if the person *Pj* has made no contact with the person *B*, then $b_{j1} = 0.$

i. Find the total number of indirect contacts between *A* and *B*.

ii. If $\vec{a} = (1, \ 0, \ 1, \ 1)$ and $\vec{a} = (1, \ 1, \ 1, \ 1),$ find the total number of indirect contacts between *A* and *B*.

1.2 Norm and angle

In this section, we shall give definitions of the lengths of vectors and derive formulas used to compute the distance of two points. Some important equalities and inequalities are studied. In particular, the Cauchy-Schwarz Inequality is obtained and applied to derive the angle between two vectors.

In $\mathbb{R}^1$, it is known that the distance between a real number *x* and zero 0 is $|x|$, the absolute value of *x*. In $\mathbb{R}^2$ by the Pythagorean theorem, the distance formula between a point $P(x_1, x_2)$ and the origin *O*(0, 0) is given by

(1.2.1)

$$\sqrt{x_1^2 + x_2^2},$$

(see **Figure 1.2**), which is the length of the vector $\overrightarrow{OP}$. Now, we generalize the formula **(1.2.1)** of the length of a vector from $\mathbb{R}^2$ to $\mathbb{R}^n$, where the length of a vector is called a norm of the vector.

Definition 1.2.1.

Let $\vec{a} = (x_1, x_2, \cdots, x_n) \in \mathbb{R}^n$. The norm of the vector $\vec{a}$ denoted by $\|\vec{a}\|$, is defined by

(1.2.2)

$$\|\vec{a}\| = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}.$$

Geometrically, the norm $\|\vec{a}\|$ represents the length of the vector $\vec{a}$. It is obvious that $\vec{a}$ is a zero vector if and only if $\|\vec{a}\| = 0$.

Example 1.2.1.

Let $\vec{a} = (1, \ 3, \ -2)$. Find $\|\vec{a}\|$.

Solution

$\|\vec{a}\| = \sqrt{1^2 + 3^2 + (-2)} = \sqrt{14}.$

The following result gives the basic properties of norms of vectors.

Theorem 1.2.1.

Let $\vec{a}, \ \vec{b} \in \mathbb{R}^n$. *Then*

i. $\|\vec{a}\|^2 = \vec{a} \cdot \vec{a}$.
ii. $\|k\vec{a}\| = \left| k \right| \|\vec{a}\|$ *for* $k \in \mathbb{R}$, *and* $\|k\vec{a}\| = k\|\vec{a}\|$ *for* $k \geq 0$.
iii. $\|\vec{a} + \vec{b}\|^2 = \|\vec{a}\|^2 + 2(\vec{a} \cdot \vec{b}) + \|\vec{b}\|^2$.
iv. $\|\vec{a} - \vec{b}\|^2 = \|\vec{a}\|^2 - 2(\vec{a} \cdot \vec{b}) + \|\vec{b}\|^2$.
v. $(\vec{a} - \vec{b}) \cdot (\vec{a} + \vec{b}) = \|\vec{a}\|^2 - \|\vec{b}\|^2$.

Proof

We only prove (*iii*). By the result (*i*) and **Theorem 1.1.2**, we obtain

$$\begin{aligned} \|\vec{a} + \vec{b}\|^2 &= (\vec{a} + \vec{b}) \cdot (\vec{a} + \vec{b}) = \vec{a} \cdot \vec{a} + \vec{a} \cdot \vec{b} + \vec{b} \cdot \vec{a} + \vec{b} \cdot \vec{b} \\ &= \|\vec{a}\|^2 + 2(\vec{a} \cdot \vec{b}) + \|\vec{b}\|^2. \end{aligned}$$

Example 1.2.2.

Let $\vec{a} \in \mathbb{R}^n$ be a nonzero vector. Find the norm of the following vector

(1.2.3)

$$\frac{1}{\|\vec{a}\|}\vec{a}.$$

Solution

Because $\vec{a} \in \mathbb{R}^n$ is a nonzero vector, $\|\vec{a}\| \neq 0$. Let $k = \frac{1}{\|\vec{a}\|}$. Then $k > 0$. By **Theorem 1.2.1** (*ii*) we have

$$\|\frac{1}{\|\vec{a}\|}\vec{a}\| = \|k\vec{a}\| = k\|\vec{a}\| = \frac{1}{\|\vec{a}\|}\|\vec{a}\| = 1.$$

Definition 1.2.2.

A vector with norm 1 is called a unit vector.

By **Example 1.2.2**, we see that every nonzero vector can be normalized to a unit vector by **(1.2.3)**.

Example 1.2.3.

Let $\vec{a} = (1,\ 2,\ 3,\ \sqrt{2})$. Normalize $\vec{a}$ to a unit vector.

Solution

Because $\|\vec{a}\| = \sqrt{1^2 + 2^2 + 3^3 + \sqrt{2}^2} = \sqrt{16} = 4,$ the unit vector is

$$\frac{1}{\|\vec{a}\|}\vec{a} = \frac{1}{4}(1,\ 2,\ 3,\ \sqrt{2}) = \left(\frac{1}{2},\ \frac{1}{2},\ \frac{3}{4},\ \frac{\sqrt{2}}{4}\right).$$

Gram determinant

Gram determinant can be used to obtain the Cauchy-Schwarz Inequality and the formula for the area of a parallelogram in $\mathbb{R}^n$.

A 2×2 matrix A is a rectangular array of four numbers $a,\ b,\ c,\ d \in \mathbb{R}$ arranged in two rows and two columns:

(1.2.4)

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

The determinant of A denoted by $|A|$ is defined by

(1.2.5)

$$\left| A \right| = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc.$$

Example 1.2.4.

Compute the following determinants.

i. $\begin{vmatrix} 2 & 3 \\ 4 & 5 \end{vmatrix}$

ii. $\begin{vmatrix} -1 & 2 \\ -3 & 6 \end{vmatrix}$

iii. $\begin{vmatrix} 1 & 0 \\ -2 & 5 \end{vmatrix}$

Solution

i. $\begin{vmatrix} 2 & 3 \\ 4 & 5 \end{vmatrix} = (2)(5) - (3)(4) = 10 - 12 = -2.$

ii. $\begin{vmatrix} -1 & 2 \\ -3 & 6 \end{vmatrix} = (-1)(6) - (2)(-3) = -6 + 6 = 0.$

iii. $\begin{vmatrix} 1 & 0 \\ -2 & 5 \end{vmatrix} = (1)(5) - (0)(-2) = 5 - 0 = 5.$

From **Example 1.2.4** , we see that determinants can be negative, zero, or positive.

Definition 1.2.3.

Let $\vec{a}, \vec{b} \in \mathbb{R}^n$. The number

(1.2.6)

$$G(\vec{a}, \vec{b}) = \begin{vmatrix} a \cdot a & a \cdot \vec{b} \\ \vec{b} \cdot \vec{a} & \vec{b} \end{vmatrix} = \|\vec{a}\|^2 \|\vec{b}\|^2 - (\vec{a} \cdot \vec{b})^2$$

is called the Gram determinant of $\vec{a}$ and $\vec{b}$

Lemma 1.2.1.

$\in \mathbb{R}^n$. *Then*

(1.2.7)

$$G(\vec{a},\ \vec{b})\|\vec{a}\|^2 = \left\|\|\vec{a}\|^2\vec{b} - (\vec{a}\cdot\vec{b})\vec{a}\right\|^2$$

Proof

Let

(1.2.8)

$$\vec{g} = \|\vec{a}\|^2\vec{b} - (\vec{a}\cdot\vec{b})\overrightarrow{a}.$$

By **Theorem 1.1.2** (2), (4) and **Theorem 1.2.1** (*i*), (*iv*), we have

$$\begin{aligned}
\|\vec{g}\|^2 &= \left\|\|\vec{a}\|^2\vec{b}\right\|^2 - 2(\|\vec{a}\|^2\vec{b})\cdot\Big((\vec{a}\cdot\vec{b})\vec{a}\Big) + \|(\vec{a}\cdot\vec{b})\vec{a}\|^2 \\
&= \|\vec{a}\|^4\|\vec{b}\|^2 - 2\|\vec{a}\|^2(\vec{a}\cdot\vec{b})^2 + (\vec{a}\cdot\vec{b})^2\|\vec{a}\|^2 \\
&= \|\vec{a}\|^4\|\vec{b}\|^2 - \|\vec{a}\|^2(\vec{a}\cdot\vec{b})^2 = \|\vec{a}\|^2\Big[\|\vec{a}\|^2\|\vec{b}\|^2 - (\vec{a}\cdot\vec{b})^2\Big].
\end{aligned}$$

This implies **(1.2.7)** .

Using **Lemma 1.2.1** , we prove the Cauchy-Schwarz inequality, which provides the relation between the dot product and the norm of the two vectors.

Theorem 1.2.2 (Cauchy-Schwarz Inequality)

Let $\vec{a},\ \vec{b} \in \mathbb{R}^n$. *Then*

$$\left|\vec{a}\cdot\vec{b}\right| \le \|\vec{a}\|\|\vec{b}\|.$$

Proof

If $\vec{a} = \vec{0}$, then we see that the Cauchy-Schwarz Inequality holds. If $\vec{a} \neq \vec{0}$, then because $\left\| \|\vec{a}\|^2 \vec{b} - (\vec{a} \cdot \vec{b})\vec{a} \right\|^2 \geq 0$, by **(1.2.7)** , $G(\vec{a}, \vec{b}) \geq 0$. It follows that

$$(\vec{a} \cdot \vec{b})^2 \leq \|\vec{a}\|^2 \|\vec{b}\|^2.$$

and

$$|\vec{a} \cdot \vec{b}| = \sqrt{(\vec{a} \cdot \vec{b})^2} \leq \sqrt{\|\vec{a}\|^2 \|\vec{b}\|^2} = \|\vec{a}\| \|\vec{b}\|.$$

In terms of components of $\vec{a} = (x_1, x_2, \cdots, x_n)$ and $\vec{b} = (y_1, y_2, \cdots, y_n)$, the Cauchy-Schwarz inequality is equivalent to the following inequality.

$$|x_1 y_1 + x_2 y_2 + \cdots + x_n y_n| \leq \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2} \sqrt{y_1^2 + y_2^2 + \cdots + y_n^2}. \tag{1.2.9}$$

If $\vec{a} = 0$ or $\vec{b} = 0$, then the equality of the Cauchy-Schwarz inequality holds. The following result shows that if $\vec{a} \neq 0$ and $\vec{b} \neq 0$, then the equality of the Cauchy-Schwarz inequality holds if and only if the two vectors are parallel.

Theorem 1.2.3.

Let $\vec{a}, \vec{b} \in \mathbb{R}^n$ be nonzero vectors. Then the following assertions are equivalent.

i. $\left| \vec{a} \cdot \vec{b} \right| = \|\vec{a}\| \|\vec{b}\|$.

ii. $\vec{b} = \frac{\vec{a} \cdot \vec{b}}{\|\vec{a}\|^2} \vec{a}$.

iii. $\vec{a} \parallel \vec{b}$.

Proof

i. implies (*ii*). Assume that (*i*) holds. Let $\vec{g}$ be the same as in **(1.2.8)** . By **(1.2.7)** , we have

$$\begin{aligned}\|\vec{g}\|^2 &= \|\vec{a}\|^2\left[\|\vec{a}\|^2\|\vec{b}\|^2 - (\vec{a}\cdot\vec{b})^2\right] \\ &= \|\vec{a}\|^2\left[\|\vec{a}\|^2\|\vec{b}\|^2 - (\|\vec{a}\|\|\vec{b}\|)^2\right] = 0.\end{aligned}$$

This implies that

$$\vec{g} = \|\vec{a}\|^2\vec{b} - (\vec{a}\cdot\vec{b})\vec{a} = \vec{0}.$$

Hence,

$$\vec{b} = \frac{\vec{a}\cdot\vec{b}}{\|\vec{a}\|^2}\vec{a}.$$

ii. implies (*iii*). Assume that (*ii*) holds. By **Definition 1.1.6** , $\vec{a} \parallel \vec{b}$.

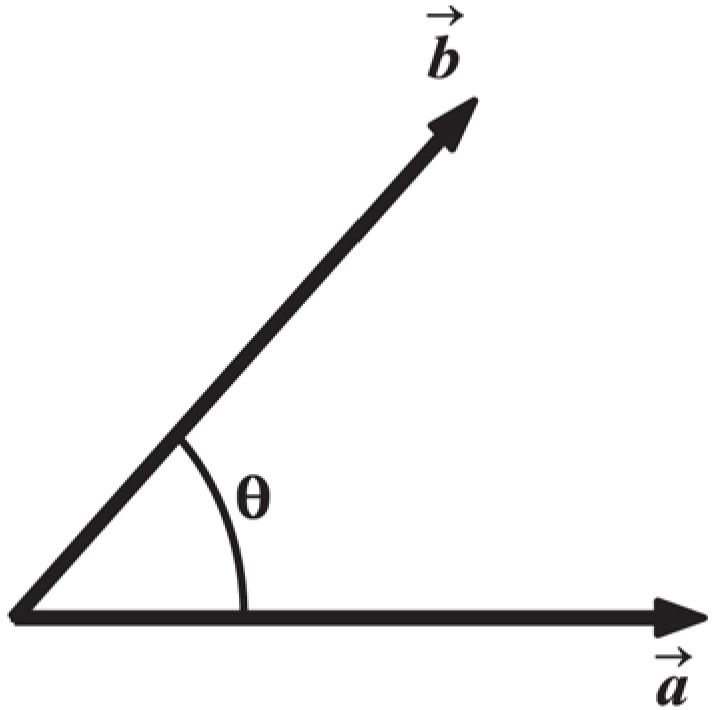
iii. implies (*i*). Assume that (*iii*) holds. By **Definition 1.1.6** , there exists $k \in \mathbb{R}$ such that $\vec{a} = k\vec{b}$. Then

$$\left|\vec{a}\cdot\vec{b}\right| = \left|(k\vec{b})\cdot\vec{b}\right| = \left|k\right|\left|\vec{b}\right\|^2 \quad \text{and } \|\vec{a}\|\|\vec{b}\| = \|k\vec{b}\|\|\vec{b}\| = \left|k\right|\left|\vec{b}\right\|^2.$$

Hence, (*i*) holds

By **Theorem 1.2.2** , we prove the triangle inequality, which is illustrated by **Figure 1.5** (I).

Figure 1.5: (I)



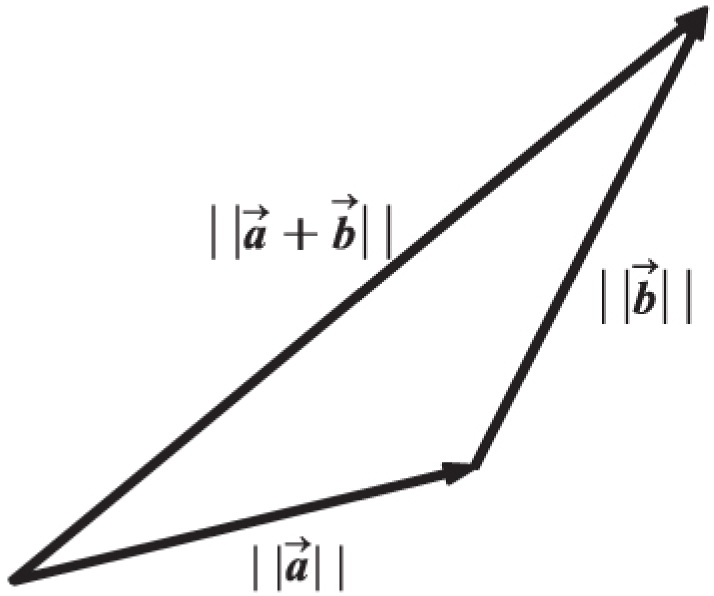
Theorem 1.2.4.

Let $\vec{a}, \vec{b} \in \mathbb{R}^n$. *Then*

$$\|\vec{a} + \vec{b}\| \leq \|\vec{a}\| + \|\vec{b}\|.$$

Moreover, equality holds if $\vec{a} = \vec{0}$ *or* $\vec{b} = \vec{0}$ *or* $\vec{a} \parallel \vec{b}$.

Figure 1.4: (I)



Proof

By **Theorem 1.2.1** (*iii*) and **Theorem 1.2.2**, we have

$$\begin{aligned}\|\vec{a}+\vec{b}\|^2 &= \|\vec{a}\|^2+2\vec{a}\cdot\vec{b}+\|\vec{b}\|^2 \le \|\vec{a}\|^2+2\|\vec{a}\|\|\vec{b}\|+\|\vec{b}\|^2\\ &= (\|\vec{a}\|+\|\vec{b}\|)^2.\end{aligned}$$

This implies that $\|\vec{a}+\vec{b}\| \le \|\vec{a}\|+\|\vec{b}\|$. If $\vec{a} \parallel \vec{b}$, by **Theorem 1.2.3**,

$$\|\vec{a}+\vec{b}\|^2 = (\|\vec{a}\|+\|\vec{b}\|)^2$$

and equality holds. It is obvious that equality holds if $\vec{a}=\vec{0}$ or $\vec{b}=\vec{0}$.

Let $P(x_1, x_2, \cdots, x_n)$ and $Q(y_1, y_2, \cdots, y_n)$ be two points in $\mathbb{R}^n$. By **(1.1.2)** and **Definition 1.2.1**, we obtain the formula of the distance between *P* and *Q* :

$$\|\overrightarrow{PQ}\| = \sqrt{(y_1 - x_1)^2 + (y_2 - x_2)^2 + \cdots + (y_n - x_n)^2}.$$

Example 1.2.5.

Find the distance between $P_1(2, \ -1, \ 5)$ and $P_2(4, \ -3, \ 1)$.

Solution

$$\|\overrightarrow{P_1P_2}\| = \sqrt{(4-2)^2 + (-3+1)^2 + (1-5)^2} = 2\sqrt{6}.$$

Now, we derive the formula for points between two points in $\mathbb{R}^2$. In particular, we obtain the midpoint formula.

Theorem 1.2.5.

Let $A(x_1, \ y_1)$ and $B(x_2, \ y_2)$ be two different points in $\mathbb{R}^2$. Then the following assertions hold.

1. *If C(x, y) is a point on the line segment joining A and B, then there exists $t \in [0, \ 1]$ such that* (1.2.11)

$$\begin{pmatrix} x \\ y \end{pmatrix} = (1-t)\begin{pmatrix} x_1 \\ y_1 \end{pmatrix} + t\begin{pmatrix} x_2 \\ y_2 \end{pmatrix}.$$

2. *For each $t \in [0, \ 1]$, $C(x, y)$ given by* **(1.2.11)** *is a point on the line segment joining A and B.*

Proof

ecause *C* is a point on the line segment joining *A* and *B*, then $\overrightarrow{AC}$ is parallel to $\overrightarrow{AB}$. By **Definition 1.1.6**, there exists $t \in \mathbb{R}$ such that

(1.2.12)

$$\overrightarrow{AC} = t\overrightarrow{AB}.$$

It follows that

$$\begin{pmatrix} x \\ y \end{pmatrix} - \begin{pmatrix} x_1 \\ y_1 \end{pmatrix} = t\left[\begin{pmatrix} x_2 \\ y_2 \end{pmatrix} - \begin{pmatrix} x_1 \\ y_1 \end{pmatrix}\right]$$

and

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x_1 \\ y_1 \end{pmatrix} + t\left[\begin{pmatrix} x_2 \\ y_2 \end{pmatrix} - \begin{pmatrix} x_1 \\ y_1 \end{pmatrix}\right] = (1-t)\begin{pmatrix} x_1 \\ y_1 \end{pmatrix} + t\begin{pmatrix} x_2 \\ y_2 \end{pmatrix}.$$

By **(1.2.12)** , we have

(1.2.13)

$$\|\overrightarrow{AC}\| = \|t\overrightarrow{AB}\| = t\|\overrightarrow{AB}\|.$$

Because *C* is between *A* and *B*, $\|\overrightarrow{AC}\| \le \|\overrightarrow{AB}\|$. This, together with **(1.2.13)** , implies $t \in [0, 1]$.

2. By **(1.2.11)** , we have

$$\overrightarrow{AC} = \begin{pmatrix} x \\ y \end{pmatrix} - \begin{pmatrix} x_1 \\ y_1 \end{pmatrix} = (1-t)\left[\begin{pmatrix} x_2 \\ y_2 \end{pmatrix} - \begin{pmatrix} x_1 \\ y_1 \end{pmatrix}\right].$$

This implies that $\overrightarrow{AC}$ is parallel to $\overrightarrow{AB}$. Because the starting point of $\overrightarrow{AC}$ and $\overrightarrow{AB}$ is *A*, *C* is a point on the line segment joining *A* and *B*.

Note that **Theorem 1.2.5** holds in $\mathbb{R}^n$ for $n \geq 3$.

By **Theorem 1.2.5** (2) with $t = \frac{1}{2}$, we obtain the following midpoint formula of the line segment joining *A* and *B*

(1.2.14)

$$\left(\frac{x_1 + x_2}{2}, \frac{y_1 + y_2}{2}\right).$$

Example 1.2.6.

Let $A(1, 2)$, $B(-2, 1)$, and $C(-1, 1)$ be three different points in $\mathbb{R}^2$. Find the distance between the midpoints of the segments *AB* and *AC*.

Solution

By **(1.2.14)**, the two midpoints of the segments *AB* and *AC* are

$$\left(\frac{1-2}{2}, \frac{2+1}{2}\right) = \left(-\frac{1}{2}, \frac{3}{2}\right) \quad \text{and} \quad \left(\frac{1-2}{2}, \frac{2+1}{2}\right) = \left(0, \frac{3}{2}\right).$$

By **(1.2.10)** with $n = 2$, we obtain the distance between the two midpoints

$$\sqrt{\left(-\frac{1}{2}-0\right)^2+\left(\frac{3}{2}-\frac{3}{2}\right)^2}=\frac{1}{2}.$$

Angle between two vectors

Let $\vec{a},\ \vec{b} \in \mathbb{R}^n$ be nonzero vectors. Then $\|\vec{a}\| \neq 0$ and $\|\vec{b}\| \neq 0$. By the Cauchy-Schwarz inequality given in **Theorem 1.2.2** , we have

$$-1 \leq \frac{\vec{a} \cdot \vec{b}}{\|\vec{a}\|\|\vec{b}\|} \leq 1.$$

Let θ be an angle whose radian measure varies from 0 to π. Then $\cos\theta$ is strictly decreasing on $[0,\ \pi]$ and

$$-1 \ \leq \cos\theta \leq 1 \ \text{ for } \theta \in [0,\ \pi].$$

Hence, for any two nonzero vectors $\vec{a},\ \vec{b} \in \mathbb{R}^n,$ there exists a unique angle $\theta \in [0,\ \pi]$ such that

(1.2.15)

$$\cos\theta = \frac{\vec{a} \cdot \vec{b}}{\|\vec{a}\|\|\vec{b}\|}.$$

If $n = 2$ or $n = 3,$ then we can position the two vectors such that they have the same initial point. We denote by θ the angle between the two vectors $\vec{a}$ and $\vec{b},$ which satisfies $0 \leq \theta \leq \pi.$ By the cosine law, we have

(1.2.16)

$$\|\vec{a}-\vec{b}\|^2 = \|\vec{a}\|^2 + \|\vec{b}\|^2 - 2\|\vec{a}\|\|\vec{b}\|\cos\theta.$$

Thus, together with **Theorem 1.2.1** (*iv*), we obtain

$$\|\vec{a}\|\|\vec{b}\| \cos\theta = \vec{a}\cdot\vec{b}$$

and **(1.2.15)** holds. Hence, it is natural to define the angle given in **(1.2.15)** as the angle between any two nonzero vectors $\vec{a},\ \vec{b}$ in $\mathbb{R}^n$.

Definition 1.2.4.

Let $\vec{a},\ \vec{b}$ be nonzero vectors in $\mathbb{R}^n$. We define the angle, denoted by θ, between $\vec{a}$ and $\vec{b}$ by **(1.2.15)** .

By the graph of $y = \cos\theta$ for $\theta \in [0,\ \pi]$, we see that $\theta \in \left[0,\ \frac{\pi}{2}\right]$ if and only if $\vec{a}\cdot\vec{b} \geq 0$ and $\theta \in \left(\frac{\pi}{2},\ \pi\right]$ if and only if $\vec{a}\cdot\vec{b} < 0$.

Definition 1.2.5.

Two vectors $\vec{a},\ \vec{b} \in \mathbb{R}^n$ are said to be orthogonal (or perpendicular) if $\theta = \frac{\pi}{2}$. We write

$$\vec{a} \perp \vec{b}.$$

By **(1.2.15)** , we obtain

Theorem 1.2.6.

Let $\vec{a},\ \vec{b} \in \mathbb{R}^n$. Then $\vec{a} \perp \vec{b}$ if and only if $\vec{a}\cdot\vec{b} = 0$.

If the two vectors $\vec{a}$ and $\vec{b}$ in **Theorem 1.2.4** are orthogonal, then we have the following Pythagorean Theorem.

ı 1.2.7 (Pythagorean Theorem)

Let $\vec{a},\ \vec{b} \in \mathbb{R}^n$. Then $\vec{a} \perp \vec{b}$ if and only if

(1.2.17)

$$\|\vec{a} + \vec{b}\|^2 = \|\vec{a}\|^2 + \|\vec{b}\|^2.$$

Proof

By **Theorem 1.2.1** (*iii*), we have

(1.2.18)

$$\|\vec{a} + \vec{b}\|^2 = \|\vec{a}\|^2 + 2(\vec{a} \cdot \vec{b}) + \|\vec{b}\|^2.$$

If $\vec{a} \perp \overrightarrow{b}$, then $\vec{a} \cdot \vec{b} = 0$. It follows from **(1.2.18)** that **(1.2.17)** holds. Conversely, if **(1.2.17)** holds, then by **(1.2.18)**, $\vec{a} \cdot \vec{b} = 0$ and $\vec{a} \perp \overrightarrow{b}$.

By **Theorem 1.2.3** and **(1.2.15)**, we have the following result.

Theorem 1.2.8.

Let $\vec{a},\ \vec{b} \in \mathbb{R}^n$ be two nonzero vectors. Then $\vec{a} \parallel \vec{b}$ if and only if $\theta = 0$ or $\theta = \pi$.

Proof

By **Theorem 1.2.3**, $\vec{a} \parallel \vec{b}$ if and only if $\left|\vec{a} \cdot \vec{b}\right| = \|\vec{a}\|\|\vec{b}\|$. Because $\vec{a},\ \vec{b} \in \mathbb{R}^n$ are two nonzero vectors, $\|\vec{a}\| \neq 0$ and $\|\vec{b}\| \neq 0$. This implies that $\left|\vec{a} \cdot \vec{b}\right| \neq 0$ By **(1.2.15)**,

$$\cos\theta = \frac{\vec{a}\cdot\vec{b}}{\|\vec{a}\|\|\vec{b}\|} = \frac{\vec{a}\cdot\vec{b}}{\left|\vec{a}\cdot\vec{b}\right|}.$$

If $\vec{a}\cdot\vec{b} > 0,$ then $\left|\vec{a}\cdot\vec{b}\right| = \vec{a}\cdot\vec{b}$ and $\cos\theta = 1$ and $\theta = 0.$ If $\vec{a}\cdot\vec{b} < 0,$ then $\left|\vec{a}\cdot\vec{b}\right| = -\vec{a}\cdot\vec{b},$ $\cos\theta = -1$ and $\theta = \pi.$

Example 1.2.7.

Let $\vec{a} = (-2, 3, 1)$ and $\vec{b} = (1, 2, -4).$ Show that $\vec{a} \perp \overrightarrow{b}.$

Solution

Because $\vec{a}\cdot\vec{b} = -2+6-4 = 0,$ by **(1.2.6)** , we have $\vec{a} \perp \vec{b}.$

Example 1.2.8.

Let $\vec{a} = \begin{pmatrix} x^2 \\ 1 \end{pmatrix}$ and $\vec{b} = \begin{pmatrix} 2 \\ -8 \end{pmatrix}.$ Find $x \in \mathbb{R}$ such that $\vec{a} \perp \vec{b}.$

Solution

$\vec{a}\cdot\vec{b} = (x^2, 1)\begin{pmatrix} 2 \\ -8 \end{pmatrix} = 2x^2 - 8 = 0$ if and only if $x = 2$ or $x = -2.$ Hence, when $x = 2$ or $x = -2,$ $\vec{a}\cdot\vec{b} = 0$ and $\vec{a} \perp \vec{b}.$

Example 1.2.9.

Find the angle θ between $\vec{a}$ and $\vec{b}.$

i. $\vec{a} = (2, -1, 1)$ and $\vec{b} = (1, 1, 2).$

ii. $\vec{a} = (1, \ -1, \ 0, \ -1)$ and $\vec{b} = (1, \ -1, \ 1, \ -1)$.

iii. $\vec{a} = (a, \ a, \ a)$, where $a > 0$ is given and $\vec{b} = (a, \ 0, \ 0)$.

iv. $\vec{a} = (1, \ -1, \ 2)$ and $\vec{b} = (0, \ 1, \ -1)$.

Solution

i. Because $\vec{a} \cdot \vec{b} = 3$, $\|\vec{a}\| = \sqrt{6}$ and $\|\vec{b}\| = \sqrt{6}$, by **(1.2.15)** ,

$$\cos\theta = \frac{\vec{a} \cdot \vec{b}}{\|\vec{a}\|\|\vec{b}\|} = \frac{3}{\sqrt{6}\sqrt{6}} = \frac{1}{2} \quad \text{and } \theta = \frac{\pi}{3}.$$

ii. Because $\vec{a} \cdot \vec{b} = 3$, $\|\vec{a}\| = \sqrt{3}$ and $\|\vec{b}\| = 2$, by **(1.2.15)** ,

$$\cos\theta = \frac{\vec{a} \cdot \vec{b}}{\|\vec{a}\|\|\vec{b}\|} = \frac{3}{2\sqrt{3}} = \frac{\sqrt{3}}{2} \quad \text{and } \ \theta = \frac{\pi}{6}.$$

iii. Because $\vec{a} \cdot \vec{b} = a^2$, $\|\vec{a}\| = \sqrt{3}a$ and $\|\vec{b}\| = a$, by **(1.2.15)** ,

$$\cos\theta = \frac{\vec{a} \cdot \vec{b}}{\|\vec{a}\|\|\vec{b}\|} = \frac{a^2}{\sqrt{3}a^2} = \frac{\sqrt{3}}{3} \quad \text{and } \theta = \arccos\frac{\sqrt{3}}{3} \approx 54.74^\circ.$$

iv. Because $\vec{a} \cdot \vec{b} = -3$, $\|\vec{a}\| = \sqrt{6}$ and $\|\vec{b}\| = \sqrt{2}$, by **(1.2.15)** ,

$$\cos\theta = \frac{\vec{a} \cdot \vec{b}}{\|\vec{a}\|\|\vec{b}\|} = -\frac{3}{\sqrt{6}\sqrt{2}} = -\frac{\sqrt{3}}{2} \quad \text{and } \theta = \frac{5\pi}{6}.$$

Exercises

1. For each of the following vectors, find its norm and normalize it to a unit vector.
   i. $\vec{a} = (1,\ 0,\ -2)$;
   ii. $\vec{b} = (1,\ 1,\ -2)$;
   iii. $\vec{c} = (2,\ 2,\ -2)$.

2. Let $\vec{a},\ \vec{b} \in \mathbb{R}^n$. Show that the following identities hold.
   i. $\|\vec{a} + \vec{b}\|^2 + \|\vec{a} - \vec{b}\|^2 = 2\|\vec{a}\|^2 + 2\|\vec{b}\|^2$.
   ii. $\|\vec{a} + \vec{b}\|^2 - \|\vec{a} - \vec{b}\|^2 = 4(\vec{a} \cdot \vec{b})$.

3. Evaluate the determinant of each of the following matrices.

$$A = \begin{pmatrix} 1 & 1 \\ 4 & -2 \end{pmatrix} B = \begin{pmatrix} -2 & 2 \\ 1 & -3 \end{pmatrix} C = \begin{pmatrix} 2 & -1 \\ 3 & 0 \end{pmatrix} D = \begin{pmatrix} -2 & 1 \\ -6 & 3 \end{pmatrix}$$

4. For each pair of vectors, verify that Cauchy-Schwarz and triangle inequalities hold.
   i. $\vec{a} = (1,\ 2),\ \vec{b} = (2,\ -1)$
   ii. $\vec{a} = (1,\ 2,\ 1),\ \vec{b} = (0,\ 2,\ -2)$
   iii. $\vec{a} = (1,\ -1,\ 0,\ 1),\ \vec{b} = (0,\ -1,\ 2,\ 1)$

5. Let $\vec{a} = (2,\ 1)$ and $\vec{b} = (1,\ x)$. Use **Theorem 1.2.3** to find $x \in \mathbb{R}$ such that $\vec{a} \parallel \vec{b}$.
6. For each pair of points P_1 and P_2, find the distance between the two points.
   1. $P_1(1,\ -2),\ P_2(-3,\ -1)$
   2. $P_1(3,\ -1),\ P_2(-2,\ 1)$
   3. $P_1(1,\ -1,\ 5),\ P_2(1,\ -3,\ 1)$
   4. $P_1(0,\ -1,\ 2),\ P_2(1,\ -3,\ 1)$
   5. $P_1(-1,\ -1,\ 5,\ 3),\ P_2(0,\ -2,\ -3,\ -1)$

7. Let $A(-1, -2)$, $B(2, 3)$, $C(-1, 1)$, and *D*(0, 1) be four different points in $\mathbb{R}^2$. Find the distance between the midpoints of the segments *AB* and *CD*.
8. For each pair of vectors, determine whether they are orthogonal.
   i. $\vec{a} = (1, 2)$, $\vec{b} = (2, -1)$
   ii. $\vec{a} = (1, 1, 1)$, $\vec{b} = (0, 2, -2)$
   iii. $\vec{a} = (1, -1, -1)$, $\vec{b} = (1, 2, -2)$
   iv. $\vec{a} = (1, -1, 1)$, $\vec{b} = (0, 2, -1)$
   v. $\vec{a} = (2, -1, 0, 1)$, $\vec{b} = (0, -1, 3, -1)$

9. For each pair of vectors, find all values of $x \in \mathbb{R}$ such that they are orthogonal.
   i. $\vec{a} = (-1, 2)$, $\vec{b} = (2, x^2)$
   ii. $\vec{a} = (1, 1, 1)$, $\vec{b} = (x, 2, -2)$
   iii. $\vec{a} = (5, x, 0, 1)$, $\vec{b} = (0, -1, 3, -1)$

10. For each pair of vectors, find $\cos\theta$, where *θ* is the angle between them.
    1. $\vec{a} = (1, -1, 1)$ and $\vec{b} = (1, 1, -2)$;
    2. $\vec{a} = (1, 0, 1)$ and $\vec{b} = (-1, -1, -1)$.

11. Let $\vec{a} = (6, 8)$ and $\vec{b} = (1, x)$. Find $x \in \mathbb{R}$ such that
    i. $\vec{a} \perp \vec{b}$
    ii. $\vec{a} \parallel \vec{b}$
    iii. the angle between $\vec{a}$ and $\vec{b}$ is $\frac{\pi}{4}$.

12. Three vertices of a triangle are

$$P_1(6, 6, 5, 8),\ P_2(6, 8, 6, 5),\ \text{and}\ P_3(5, 7, 4, 6).$$

    a. Show that the triangle is a right triangle.

b. Find the area of the triangle.

13. Suppose three vertices of a triangle in $\mathbb{R}^5$ are

$$P(4,\ 5,\ 5,\ 6,\ 7),\ O(3,\ 6,\ 4,\ 8,\ 4),\ Q(5,\ 7,\ 7,\ 13,\ 6).$$

a. Show that the triangle is a right triangle.
b. Find the area of the triangle.

14. Four vertices of a quadrilateral are

$$A(4,\ 5,\ 5,\ 7),\ B(6,\ 3,\ 3,\ 9),\ C(5,\ 2,\ 2,\ 8),\ D(3,\ 4,\ 4,\ 6).$$

a. Show that the quadrilateral is a rectangle.
b. Find the area of the quadrilateral.

1.3 Projection and areas of parallelograms

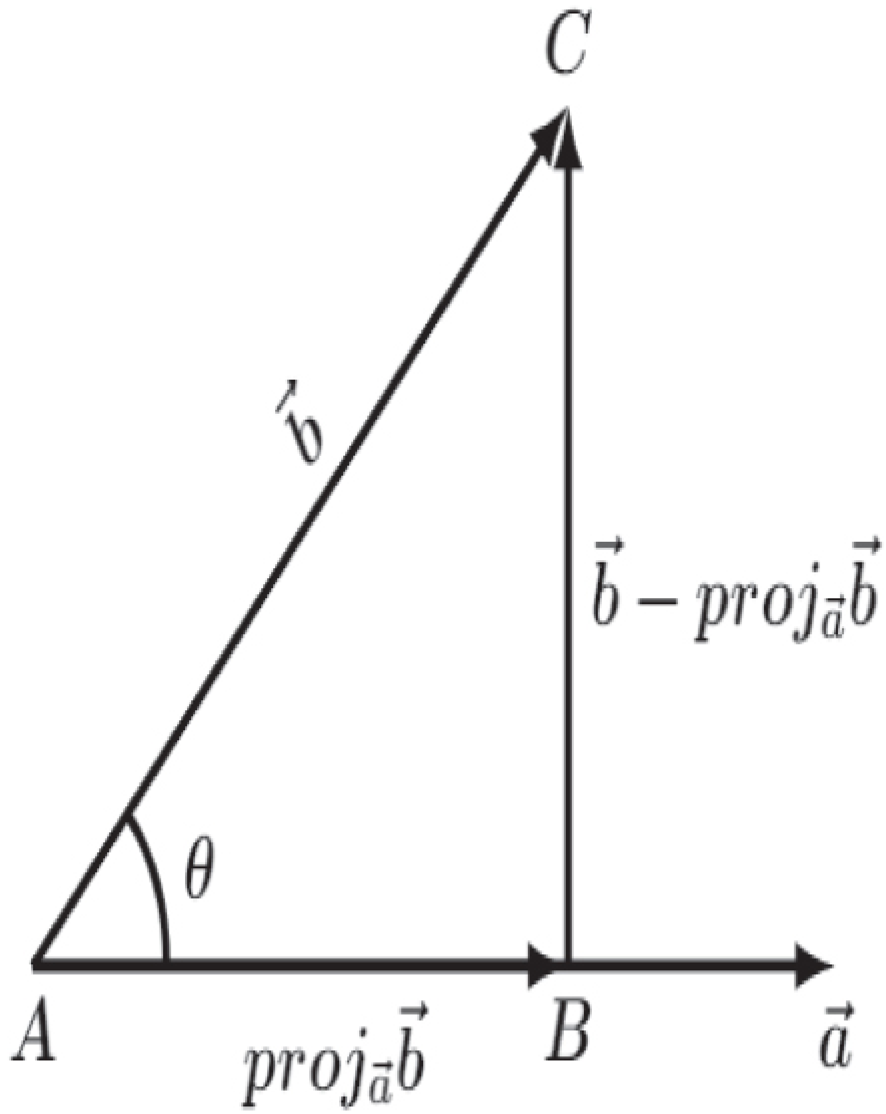
In this section, we study the projection of a vector $\vec{b}$ on a vector $\vec{a}$ in $\mathbb{R}^n$. The projection has many applications. For example, it can be applied to find the height of a parallelogram in $\mathbb{R}^n$ later in this section, the volume of a parallelpiped in $\mathbb{R}^3$ (**Section 6.1**), and the distance from a point in $\mathbb{R}^3$ to a plane in $\mathbb{R}^3$ (**Section 6.3**).

Definition 1.3.1.

Let $\vec{a}, \vec{b}$ be nonzero vectors in $\mathbb{R}^n$. A vector $\vec{p}$ in $\mathbb{R}^n$ is said to be a projection of $\vec{b}$ on $\vec{a}$ if $\vec{p}$ satisfies the following two conditions.

1. $\vec{p}$ is parallel to $\vec{a}$.
2. $\vec{b} - \vec{p}$ is perpendicular to $\vec{a}$.

Figure 1.6:

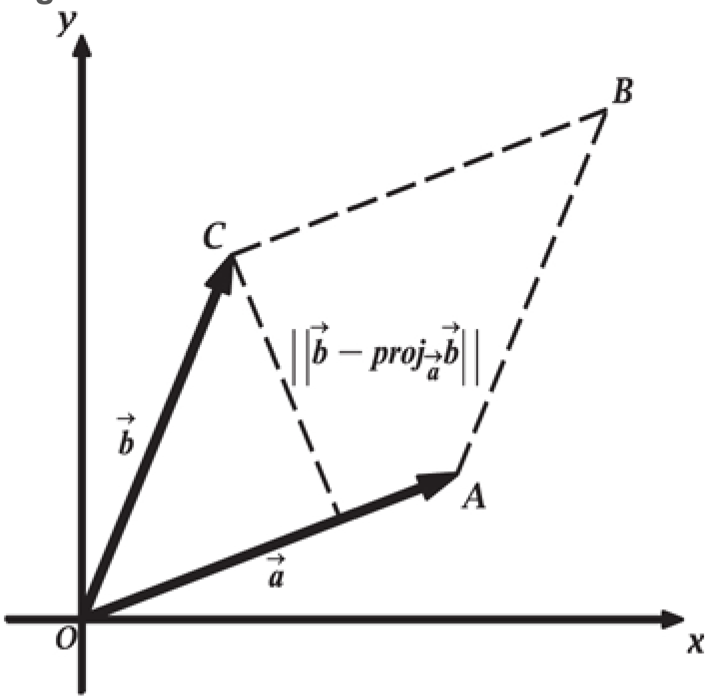


If a vector $\vec{p}$ is a projection of $\vec{b}$ on $\vec{a}$, we write

$$\vec{p} = \mathrm{Proj}_{\vec{a}}\,\vec{b}.$$

We refer to **Figure 1.7** for the relations among the vectors $\vec{a}$, $\vec{b}$, $\mathrm{proj}_{\vec{a}}\vec{b}$, and $\vec{b} - \mathrm{Proj}_{\vec{a}}\vec{b}$. Moreover, if $\vec{a}\cdot\vec{b} > 0$, then the angle between $\vec{a}$ and $\vec{b}$ belongs to $\left(0, \frac{\pi}{2}\right)$ and $\mathrm{Proj}_{\vec{a}}$ and $\vec{a}$ have the same direction. If $\vec{a}\cdot\vec{b} < 0$, then the angle between $\vec{a}$ and $\vec{b}$ belongs to $\left(\frac{\pi}{2}, \pi\right)$ and $\mathrm{Proj}_{\vec{a}}$ and $\vec{a}$ have the opposite direction. If $\vec{a}\cdot\vec{b} = 0$, then $\vec{a} \perp \vec{b}$ and $\mathrm{Proj}_{\vec{a}}\vec{b} = \vec{0}$.

Figure 1.7:



The following result gives the formula for the projection of $\vec{b}$ on $\vec{a}$.

Theorem 1.3.1.

Let $\vec{a}$, $\vec{b}$ be nonzero vectors in $\mathbb{R}^n$. Then

(1.3.1)

$$\text{Proj}_{\vec{a}}\,\vec{b} = \left(\frac{\vec{a}\cdot\vec{b}}{\|\vec{a}\|^2}\right)\vec{a}.$$

Proof

By **Definition 1.3.1** , $\text{Proj}_{\vec{a}}\,\vec{b} \parallel \vec{a}$ and $(\vec{b} - \text{Proj}_{\vec{a}}\,\vec{b}) \perp \vec{a}$. By **Definition 1.1.6** and **Theorem 1.2.6** , there exists $k \in \mathbb{R}$ such that

(1.3.2)

$$\text{Proj}_{\vec{a}}\,\vec{b} = k\vec{a}$$

and

(1.3.3)

$$\left(\vec{b} - \text{Proj}_{\vec{a}}\,\vec{b}\right)\cdot\vec{a} = 0.$$

Hence, by **(1.3.3)** , **Theorem 1.1.2** (2) and (3), **(1.3.2)** , and **Theorem 1.2.1** (*i*),

$$\vec{a}\cdot\vec{b} = \vec{a}\cdot\text{Proj}_{\vec{a}}\,\vec{b} = \vec{a}\cdot(k\vec{a}) = k(\vec{a}\cdot\vec{a}) = k\|\vec{a}\|^2.$$

This imples $k = \frac{\vec{a}\cdot\vec{b}}{\|\vec{a}\|^2}$. This, together with **(1.3.2)** , implies **(1.3.1)** .

Example 1.3.1.

Let $\vec{a} = (4, \ -1, \ 2)$ and $\vec{b} = (2, \ -1, \ 3)$. Find $\text{Proj}_{\vec{a}}\,\vec{b}$.

Solution

Because

$$\vec{a} \cdot \vec{b} = 4(2) + (-1)(-1) + 2(3) = 15 \quad \text{and } \|\vec{a}\|^2 = 16 + 1 + 4 = 21,$$

by **(1.3.1)** we obtain

$$\mathrm{Proj}_{\vec{a}}\, \vec{b} = \frac{15}{21}(4,\ -1,\ 2) = \frac{5}{7}(4,\ -1,\ 2) = \left(\frac{20}{7},\ -\frac{5}{7},\ \frac{10}{7}\right).$$

Theorem 1.3.2.

Let $\vec{a},\ \vec{b} \in \mathbb{R}^n$ be nonzero vectors. Then the following assertions hold.

i. $\|\mathrm{Proj}_{\vec{a}}\, \vec{b}\| = \dfrac{\left|\vec{a} \cdot \vec{b}\right|}{\|\vec{a}\|}$.
ii. $|\cos\theta| = \dfrac{\|\mathrm{Proj}_{\vec{a}}\, \vec{b}\|}{\|\vec{b}\|}$, *where θ is the angle between $\vec{a}$ and $\vec{b}$.*

Proof

i. Because $\mathrm{Proj}_{\vec{a}}\, \vec{b} = \left(\dfrac{\vec{a} \cdot \vec{b}}{\|\vec{a}\|^2}\right)\vec{a}$, we have

$$\|\mathrm{Proj}_{\vec{a}}\, \vec{b}\| = \left\|\left(\frac{\vec{a} \cdot \vec{b}}{\|\vec{a}\|^2}\right)\vec{a}\right\| = \frac{\left|\vec{a} \cdot \vec{b}\right|}{\|\vec{a}\|^2}\|\vec{a}\| = \frac{\left|\vec{a} \cdot \vec{b}\right|}{\|\vec{a}\|}.$$

ii. By **(1.2.15)** and (*i*), we have

$$|\cos\theta| = \frac{\left|\vec{a}\cdot\vec{b}\right|}{\|\vec{a}\|\|\vec{b}\|} = \frac{\|\mathrm{Proj}_{\vec{a}}\,\vec{b}\|}{\|\vec{b}\|}$$

and the result (*ii*)holds.

By **Theorem 1.3.2** , we see that the following assertions hold.

1. If $\theta \in \left(0,\ \frac{\pi}{2}\right)$, then $\cos\theta = \frac{\|\mathrm{Proj}_{\vec{a}}\,\vec{b}\|}{\|\vec{b}\|}$.
2. If $\theta \in \left(\frac{\pi}{2},\ \pi\right)$, then $\cos\theta = -\frac{\|\mathrm{Proj}_{\vec{a}}\,\vec{b}\|}{\|\vec{b}\|}$.

Example 1.3.2.

Let $\vec{a} = (1,\ -1,\ 2)$ and $\vec{b} = (0,\ 1,\ -1)$.

1. Use **Theorem 1.3.2** (*i*) to find $\|\mathrm{Proj}_{\vec{a}}\,\vec{b}\|$
2. Use **Theorem 1.3.2** (*ii*) to find the angle θ between $\vec{a}$ and $\vec{b}$.

Solution

1. Because
   $\vec{a}\cdot\vec{b} = (1)(0) + (-1)(1) + (2)(-1) = -3$ and $\|\vec{a}\| = \sqrt{1^2 + (-1)^2 + 2^2} = \sqrt{6}$,
   by **Theorem 1.3.2** (*i*),

$$\|\mathrm{Proj}_{\vec{a}}\,\vec{b}\| = \frac{\left|\vec{a}\cdot\vec{b}\right|}{\|\vec{a}\|} = \frac{|1-3|}{\sqrt{6}} = \frac{3}{\sqrt{6}} = \frac{\sqrt{6}}{2}.$$

2. Because $\|\vec{b}\| = \sqrt{0^2 + 1^2 + (-1)^2} = \sqrt{2}$, by **Theorem 1.3.2** (*ii*),

$$|\cos\theta| = \frac{\|\mathrm{Proj}_{\vec{a}}\vec{b}\|}{\|\vec{b}\|} = \frac{\frac{\sqrt{6}}{2}}{\sqrt{2}} = \frac{\sqrt{3}}{2}.$$

Because $\vec{a}\cdot\vec{b} = -3 < 0$, we have cos θ < 0. Hence,

$$\cos\theta = -|\cos\theta| = -\frac{\sqrt{3}}{2} \text{ and } \theta = \pi - \frac{\pi}{6} = \frac{5\pi}{6}.$$

Areas of parallelograms

Let $\vec{a} = (x_1, x_2, \cdots, x_n)$ and $\vec{b} = (y_1, y_2, \cdots, y_n)$ be nonzero vectors in $\mathbb{R}^n$. By the condition (2) of **Definition 1.3.1**, we see that the vector $(\vec{b} - \mathrm{Proj}_{\vec{a}}\vec{b})$ is always orthogonal to $\vec{a}$ whenever the angle θ between $\vec{a}$ and $\vec{b}$ belongs to $\left(0, \frac{\pi}{2}\right]$ or $\left(\frac{\pi}{2}, \pi\right)$. In $\mathbb{R}^2$, it is easy to see that the norm $\|\vec{b} - \mathrm{Proj}_{\vec{a}}\vec{b}\|$ is the the height of the parallelogram determined by $\vec{a}$ and $\vec{b}$ whenever $\theta \in \left(0, \frac{\pi}{2}\right]$ or $\theta \in \left(\frac{\pi}{2}, \pi\right)$.

Hence, we define the norm of $(\vec{b} - \mathrm{Proj}_{\vec{a}}\vec{b})$ as the height of the parallelogram determined by $\vec{a}$ and $\vec{b}$ in $\mathbb{R}^n$ (see **Figure 1.7**). We denote by $A(\vec{a}, \vec{b})$ the area of the parallelogram determined by $\vec{a}$ and $\vec{b}$. Then

(1.3.4)

$$A(\vec{a}, \vec{b}) = \|\vec{a}\|\|\vec{b} - \mathrm{Proj}_{\vec{a}}\vec{b}\|.$$

The formula for the area $A(\vec{a},\ \vec{b})$ given in Definition 1.3.4 depends heavily on the projection $\mathrm{Proj}_{\vec{a}}\vec{b}$. By Pythagorean **Theorem 1.2.7** and **Theorem 1.3.2**, we derive a formula for $A(\vec{a},\ \vec{b})$, which depends only on the Gram determinant of $\vec{a}$ and $\vec{b}$ so it is easier to use it to compute $A(\vec{a},\ \vec{b})$.

Theorem 1.3.3.

Let $\vec{a}$ and $\vec{b}$ be nonzero vectors in $\mathbb{R}^n$. Then

(1.3.5)

$$A(\vec{a},\ \vec{b}) = \sqrt{G(\vec{a},\ \vec{b})} = \sqrt{\|\vec{a}\|^2\|\vec{b}\|^2 - (\vec{a}\cdot\vec{b})^2}.$$

Proof

Because $(\vec{b} - \mathrm{Proj}_{\vec{a}}\vec{b}) \perp \vec{a}$ and $\vec{a}\,||\,\mathrm{Proj}_{\vec{a}}\vec{b}$, we have

$$(\vec{b} - \mathrm{Proj}_{\vec{a}}\vec{b}) \perp \mathrm{Proj}_{\vec{a}}\vec{b}.$$

By **Theorems 1.2.7** and **(1.3.2)**,

$$\|\vec{b} - \mathrm{Proj}_{\vec{a}}\vec{b}\|^2 = \|\vec{b}\|^2 - \|\mathrm{Proj}_{\vec{a}}\vec{b}\|^2 = \|\vec{b}\|^2 - \frac{|\vec{a}\cdot\vec{b}|^2}{\|\vec{a}\|^2}$$

and

$$A(\vec{a},\ \vec{b})^2 = \|\vec{a}\|^2\|\vec{b} - \mathrm{Proj}_{\vec{a}}\vec{b}\|^2 = \|\vec{a}\|^2\|\vec{b}\|^2 - |\vec{a}\cdot\vec{b}|^2 = G(\vec{a},\ \vec{b}).$$

This implies $A(\vec{a},\ \vec{b}) = \sqrt{G(\vec{a},\ \vec{b})}$.

If $\vec{a} \perp \vec{b}$, then the parallelogram determined by $\vec{a}$ and $\vec{b}$ is called a rectangle. By **Theorem 1.3.3** , we see that if $\vec{a} \perp \vec{b}$, then the area of the rectangle equals $\|\vec{a}\|\|\vec{b}\|$ because $\vec{a} \cdot \vec{b} = 0$.

To give a formula of $A(\vec{a}, \vec{b})$ in terms of components of $\vec{a}, \vec{b} \in \mathbb{R}^n$, we prove the Lagrange's Identity on the Gram determinant.

Lemma 1.3.1 (Lagrange's Identity)

Let $\vec{a} = (x_1, x_2, \cdots, x_n)$ and $\vec{b} = (y_1, y_2, \cdots, y_n)$ *be two vectors in* $\mathbb{R}^n$. *Then*

(1.3.6)

$$G(\vec{a}, \vec{b}) = \sum_{i=1}^{n=1} \sum_{j=i+1}^{n} \begin{vmatrix} x_i & x_j \\ y_i & y_j \end{vmatrix}^2 .$$

Proof

By computation, we have

$$
\begin{aligned}
\|\vec{a}\|^2 \|\vec{b}\|^2 &= \left(\sum_{i=1}^{n} x_i^2\right)\left(\sum_{j=1}^{n} y_j^2\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} x_i^2 y_j^2 \\
&= \sum_{i=1}^{n} (x_i y_i)^2 + \sum_{i=1}^{n-1}\sum_{j=i+1}^{n} x_i^2 y_j^2 + \sum_{j=1}^{n-1}\sum_{i=j+1}^{n} x_i^2 y_j^2 \\
&= \sum_{i=1}^{n} (x_i y_i)^2 + \sum_{i=1}^{n-1}\sum_{j=i+1}^{n} x_i^2 y_j^2 + \sum_{i=1}^{n-1}\sum_{j=i+1}^{n} x_j^2 y_i^2 \\
&= \sum_{i=1}^{n} (x_i y_i)^2 + \sum_{i=1}^{n-1}\sum_{j=i+1}^{n} (x_i^2 y_j^2 + x_j^2 y_i^2)
\end{aligned}
$$

and

$$
\begin{aligned}
(\vec{a}\cdot\vec{b})^2 &= \left(\sum_{i=1}^{n} x_i y_i\right)^2 = \sum_{i=1}^{n} x_i y_i \left(\sum_{j=1}^{n} x_j y_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} (x_i y_i)(x_j y_j) \\
&= \sum_{i=1}^{n} (x_i y_i)^2 + \sum_{i=1}^{n-1}\sum_{j=i+1}^{n} (x_i y_i)(x_j y_j) + \sum_{j=1}^{n-1}\sum_{i=j+1}^{n} (x_i y_i)(x_j y_j) \\
&= \sum_{i=1}^{n} (x_i y_i)^2 + \sum_{i=1}^{n-1}\sum_{j=i+1}^{n} (x_i y_i)(x_j y_j) + \sum_{i=1}^{n-1}\sum_{j=i+1}^{n} (x_j y_j)(x_i y_i) \\
&= \sum_{i=1}^{n} (x_i y_i)^2 + 2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} (x_i y_i)(x_j y_j).
\end{aligned}
$$

$$G(\vec{a},\ \vec{b}) = \left[\sum_{i=1}^{n}(x_iy_i)^2 + \sum_{i=1}^{n-1}\sum_{j=i+1}^{n}(x_i^2y_j^2 + x_j^2y_i^2)\right]$$

$$-\left[\sum_{i=1}^{n}(x_iy_i)^2 + 2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}(x_iy_i)(x_jy_j)\right]$$

$$= \sum_{i=1}^{n}\sum_{j=i+1}^{n}(x_i^2y_j^2 + x_j^2y_i^2) - 2\sum_{i=1}^{n}\sum_{j=i+1}^{n}(x_iy_i)(x_jy_j)$$

$$= \sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\left[(x_i^2y_j^2 + x_j^2y_i^2) - 2(x_iy_i)(x_jy_j)\right] = \sum_{i=1}^{n-1}\sum_{j=i+1}^{n}(x_iy_j - x_jy_i)^2.$$

Hence, **(1.3.6)** holds.

By **Theorem 1.3.3** and **Lemma 1.3.1**, we obtain the following formula of $A(\vec{a}, \vec{b})$ in terms of components of $\vec{a},\ \vec{b} \in \mathbb{R}^n$.

(1.3.7)

$$A(\vec{a}, \vec{b}) = \sqrt{\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\begin{vmatrix} x_i & x_j \\ y_i & y_j \end{vmatrix}^2}.$$

The formula for $A(\vec{a}, \vec{b})$ given in **(1.3.7)** only depends on the components of $\vec{a}$ and $\vec{b}$, where norms and dot products are not involved, but when $n \geq 4$, it contains many terms of 2×2 determinants. So we only state the simple cases when $n = 2$ or $n = 3$ as a corollary. For $n \geq 4$, it would be simpler to use **(1.3.5)** to compute $A(\vec{a},\ \vec{b})$.

In Corollary 1.3.1

1. *Let* $\vec{a} = (x_{1,} x_2)$ *and* $\vec{b} = (y_1, y_2)$. *Then*

$$G(\vec{a}, \vec{b}) = \begin{vmatrix} x_1 & x_2 \\ y_1 & y_2 \end{vmatrix}^2$$

and

$$A(\vec{a},\ \vec{b}) = \left\| \begin{matrix} x_1 & x_2 \\ y_1 & y_2 \end{matrix} \right\|.$$

2. *Let* a→=(x1,x2,x3) *and* b→=(y1,y2,y3). *Then*

G(a→,b→)=| x1x2y1y2 |2+| x1x3y1y3 |2+| x2x3y2y3 |2

and

A(a→,b→)=| x1x2y1y2 |2+| x1x3y1y3 |2+| x2x3y2y3 |2.

In Corollary 1.3.1 (2), the three determinants are obtained by deleting the third, second, and first columns of the following matrix, respectively:

(x1x2x3y1y2y3).

Example 1.3.3.

1. Find A(a→,b→), where a→=(1,2) and b→=(0,−1).

2. Find $A(\vec{a}, \vec{b})$ and $\|\vec{b}-\operatorname{proj}_{\vec{a}} \vec{b}\|$, where $\vec{a}=(1, -2, 0)$ and $\vec{b}=(-5,0,2)$.
3. Find $A(\vec{a},\vec{b})$ and $\|\vec{b}-\operatorname{proj}_{\vec{a}} \vec{b}\|$, where $\vec{a}=(2, -1, 1, -1)$ and $\vec{b}=(0, 1, -1, 2)$.

Solution

1. By Corollary 1.3.1 (1), $A(\vec{a},\vec{b})=\left|\,\begin{vmatrix} 1 & 2 \\ 0 & -1 \end{vmatrix}\,\right|=|-1|=1$.
2. Because
$$G(\vec{a},\vec{b})=\begin{vmatrix} 1 & -2 \\ -5 & 0 \end{vmatrix}^2+\begin{vmatrix} 1 & 0 \\ -5 & 2 \end{vmatrix}^2+\begin{vmatrix} -2 & 0 \\ 0 & 2 \end{vmatrix}^2=120,$$

   by Corollary 1.3.1 (2), $A(\vec{a},\vec{b})=\sqrt{G(\vec{a},\vec{b})}=\sqrt{120}=2\sqrt{30}$.
   By **(1.3.4)**, $\|\vec{b}-\operatorname{proj}_{\vec{a}} \vec{b}\|=\frac{A(\vec{a},\vec{b})}{\|\vec{a}\|}=\frac{2\sqrt{30}}{\sqrt{5}}=2\sqrt{6}$.
3. Because
$$G(\vec{a}, \vec{b})=\|\vec{a}\|^2\|\vec{b}\|^2-(\vec{a}\cdot\vec{b})^2=(7)(6)-(-4)^2=26,$$

   by **Theorem 1.3.3**, $A(\vec{a},\vec{b})=\sqrt{G(\vec{a},\vec{b})}=\sqrt{26}$.

Exercises

1. For each pair of vectors $\vec{a}$ and $\vec{b}$, find $\operatorname{Proj}_{\vec{a}} \vec{b}$ and verify that $(\vec{b}-\operatorname{proj}_{\vec{a}} \vec{b})\perp\vec{a}$.
   1. $\vec{a}=(-1, 2)$, $\vec{b}=(1, 1)$;
   2. $\vec{a}=(2, -1, 1)$, $\vec{b}=(1, -1, 2)$;
   3. $\vec{a}=(-1, -2, 2)$, $\vec{b}=(1, 0, -1)$;
   4. $\vec{a}=(0, -1, 1, -1)$, $\vec{b}=(1, 1, -1, 2)$.

2. Let $\vec{a}=(1,1,1)$ and $\vec{b}=(1,-1,1)$.
   1. Use **Theorem 1.3.2** (*i*) to find $\|\operatorname{Proj}_{\vec{a}} \vec{b}\|$.
   2. Use **Theorem 1.3.2** (*ii*) to find $\cos\theta$, where θ is the angle between $\vec{a}$ and $\vec{b}$.

3. Let $\vec{a}=(1,0,1,2)$ and $\vec{b}=(-\frac{1}{2},\frac{1}{2},\frac{1}{2},-\frac{3}{2})$.

1. Use **Theorem 1.3.2** (*i*) to find ∥ Proj$\vec{a}$ $\vec{b}$ ∥.
2. Use **Theorem 1.3.2** (*ii*) to find θ, where θ is the angle between $\vec{a}$ and $\vec{b}$.

4. Find the area of the parallelogram determined by $\vec{a}$ and $\vec{b}$.
    1. $\vec{a}=(1,0)$, $\vec{b}=(0,-1)$;
    2. $\vec{a}=(1,0)$, $\vec{b}=(-1, 1)$;
    3. $\vec{a}=(0,2)$, $\vec{b}=(1, 1)$;
    4. $\vec{a}=(1, -1)$, $\vec{b}=(-1, 2)$.

5. Find the area of the parallelogram determined by $\vec{a}$ and $\vec{b}$.
    1. $\vec{a}=(1, 0, 1)$, $\vec{b}=(-1, 0, 1)$;
    2. $\vec{a}=(1, 0, 0)$, $\vec{b}=(-1, 1, -2)$;
    3. $\vec{a}=(0, 2, -1)$, $\vec{b}=(-5, 1, 1)$;
    4. $\vec{a}=(1, -1, -1)$, $\vec{b}=(-1, 1, 2)$;
    5. $\vec{a}=(1, 0, 2, 1)$, $\vec{b}=(2, 1, 0, 3)$;
    6. $\vec{a}=(-2, -1, 0, 3)$, $\vec{b}=(3, 1, -1, 0)$.

6. Find $G(\vec{a}, \vec{b})$ and ∥$\vec{b}$ −Porj$\vec{a}$ $\vec{b}$∥ for each pair of vectors.
    a. $\vec{a}=(-1, 2)$, $\vec{b}=(1, -1)$;
    b. $\vec{a}=(2, -1, 1)$, $\vec{b}=(1, -1, 2)$;
    c. $\vec{a}=(1, 2, 0, 1)$, $\vec{b}=(1, -1, 2, 3)$;
    d. $\vec{a} = (2, -1, 1, 0)$, $\vec{b}=(3, 0, 1, 3)$.

1.4 Linear combinations and spanning spaces

In this section, we present the notions of linear combinations and spanning spaces, which will be used in Chapters 4–8. These notions employ the operations given in **Definition 1.1.5** .

Let $m, \ n \in \mathbb{N}$ and let

(1.4.1)

$$\vec{a}_1 = \begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{pmatrix}, \ \vec{a}_2 = \begin{pmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{pmatrix}, \ \cdots, \ \vec{a}_n = \begin{pmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{pmatrix}$$

be vectors in $\mathbb{R}^m$. Let $\vec{b} = (b_1, \ b_2 \cdots, \ b_n)^T$ and

(1.4.2)

$$S = \left\{ \vec{a}_1, \ \vec{a}_2, \ \cdots, \ \vec{a}_n \right\}.$$

Definition 1.4.1.

If there exist real numbers $x_1, \ x_2, \ \cdots, \ x_n \in \mathbb{R}$ such that

(1.4.3)

$$\vec{b} = x_1\vec{a_1} + x_2\vec{a_2} + \cdots + x_n\vec{a_n},$$

then $\vec{b}$ is said to be a linear combination of $\vec{a_1}, \vec{a_2}, \cdots, \vec{a_n}$or *S*.

Example 1.4.1.

Let

$$\vec{a_1} = \begin{pmatrix} 1 \\ 0 \\ 2 \\ 4 \end{pmatrix}, \vec{a_2} = \begin{pmatrix} 2 \\ 3 \\ 1 \\ 2 \end{pmatrix}, \vec{a_3} = \begin{pmatrix} 0 \\ 1 \\ 3 \\ 1 \end{pmatrix}, \vec{a_4} = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \vec{a_5} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

i. Let $\vec{b} = (-4, \ -9, \ 1, \ 2)^T$. Verify that $2\vec{a_1} - 3\vec{a_2} = \vec{b}$. Hence, $\vec{b}$ is a linear combination of $\vec{a_1}, \vec{a_2}$.

ii. Let $\vec{b} = \big(5, \ 7, \ 7, \ 9)^T$. Verify that $\vec{a_1} + 2\vec{a_2} + \vec{a_3} = \vec{b}$. Hence, $\vec{b}$ is a linear combination of $\vec{a_1}, \vec{a_2}, \vec{a_3}$.

iii. Let $\vec{b} = (-1, \ -7, \ 6, \ 7)^T$. Verify that $3\vec{a_1} - 2\vec{a_2} - \vec{a_3} + 5\vec{a_4} = \vec{b}$. Hence, $\vec{b}$ is a linear combination of $\vec{a_1}, \vec{a_2}, \vec{a_3}, \vec{a_4}$.

iv. Let $\vec{b} = (-5,\ 1,\ 2,\ 4)^T$. Verify that $\vec{a_1} + \vec{a_2} - 2\vec{a_3} + 5\vec{a_4} - 8\vec{a_5} = \vec{b}$. Hence, $\vec{b}$ is a linear combination of $\vec{a_1},\ \vec{a_2},\ \vec{a_3},\ \vec{a_4},\ \vec{a_5}$.

Solution

i. $2\vec{a_1} - 3\vec{a_2} = 2\begin{pmatrix}1\\0\\2\\4\end{pmatrix} - 3\begin{pmatrix}2\\3\\1\\2\end{pmatrix} = \begin{pmatrix}2-6\\0-9\\4-3\\8-6\end{pmatrix} = \begin{pmatrix}-4\\-9\\1\\2\end{pmatrix}.$

ii. $\vec{a_1} + 2\vec{a_2} + \vec{a_3} = \begin{pmatrix}1\\0\\2\\4\end{pmatrix} + 2\begin{pmatrix}2\\3\\1\\2\end{pmatrix} + \begin{pmatrix}0\\1\\3\\1\end{pmatrix} = \begin{pmatrix}1+4+0\\0+6+1\\2+2+3\\4+4+1\end{pmatrix} = \begin{pmatrix}5\\7\\7\\9\end{pmatrix}.$

iii. Let $\vec{c} = 3\vec{a_1} - 2\vec{a_2} - \vec{a_3} + 5\vec{a_4}$. Then

$$\vec{c} - = 3\begin{pmatrix}1\\0\\2\\4\end{pmatrix} - 2\begin{pmatrix}2\\3\\1\\2\end{pmatrix} - \begin{pmatrix}0\\1\\3\\1\end{pmatrix} + 5\begin{pmatrix}0\\0\\1\\0\end{pmatrix} = \begin{pmatrix}-1\\-7\\6\\7\end{pmatrix}.$$

iv. Let $\vec{c} = \vec{a_1} + \vec{a_2} - 2\vec{a_3} + 5\vec{a_4} - 8\vec{a_5}$. Then

$$\vec{c} = \begin{pmatrix} 1 \\ 0 \\ 2 \\ 4 \end{pmatrix} + \begin{pmatrix} 2 \\ 3 \\ 1 \\ 2 \end{pmatrix} - 2\begin{pmatrix} 0 \\ 1 \\ 3 \\ 1 \end{pmatrix} + 5\begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} - 8\begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} -5 \\ 1 \\ 2 \\ 4 \end{pmatrix}.$$

In **Example 1.4.1** , we verify that the vector $\vec{b}$ is a linear combination of some vectors from $\vec{a_1}, \vec{a_2}, \vec{a_3}, \vec{a_4}, \vec{a_5}$. For example, in **Example 1.4.1** (1), we verify that

$$\vec{b} = 2\vec{a_1} - 3\vec{a_2}.$$

A question is how to find the coefficients 2 and −3 of $\vec{a_1}$ and $\vec{a_2}$ if $\vec{a_1}, \vec{a_2},$ and $\vec{b}$ are given? In general, given vectors $\vec{a_1}, \vec{a_2}, \cdots, \vec{a_n}$ and a vector $\vec{b}$, how to show that $\vec{b}$ is a linear combination of $\vec{a_1}, \vec{a_2}, \cdots, \vec{a_n}$? Or equivalently, by **Definition 1.4.1** , how to find the numbers $x_1, x_2, \cdots, x_n \in \mathbb{R}$ such that **(1.4.3)** holds?

In the following, we only give a simple example to show how to find these numbers, and give more detailed discussions in **Section 4.6** after we study the approaches of solving systems of linear equations in Sections 4.2–4.4.

Example 1.4.2.

Let $\vec{a_1} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$ and $\vec{a_2} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$. Show that $\vec{b} = \begin{pmatrix} 8 \\ 7 \end{pmatrix}$ is a linear combination of $\vec{a_1}, \vec{a_2}$.

Solution

Let $x_1\overrightarrow{a_1} + x_2\overrightarrow{a_2} = \vec{b}$. Then

$$x_1\begin{pmatrix}1\\2\end{pmatrix} + x_2\begin{pmatrix}2\\1\end{pmatrix} = \begin{pmatrix}8\\7\end{pmatrix}.$$

This implies

$$\begin{pmatrix}x_1 + 2x_2\\2x_1 + x_2\end{pmatrix} = \begin{pmatrix}8\\7\end{pmatrix}.$$

By **Definition 1.1.4** , we have

$$\begin{cases}x_1 + 2x_2 & = & 8\\2x_1 + x_2 & = & 7.\end{cases}$$

Solving the system, we get $x_1 = 2$ and $x_2 = 3$. Hence, $\vec{b} = 2\overrightarrow{a_1} + 3\overrightarrow{a_2}$ and $\vec{b}$ is a linear combination of $\overrightarrow{a_1}$ and $\overrightarrow{a_2}$.

Now, we discuss some special cases of **Definition 1.4.1** . In **Definition 1.4.1** , if *n* = 1, then $\vec{b}$ is a linear combination of $\overrightarrow{a_1}$ if and only if there exists $x_1 \in \mathbb{R}$ such that

$$\vec{b} = x_1\overrightarrow{a_1}.$$

If $\vec{b} \neq \vec{0}$ and $\overrightarrow{a_1} \neq \vec{0}$, then by **Definition 1.1.6**, $\vec{b}$ is parallel to $\overrightarrow{a_1}$. Hence, If $\vec{b} \neq \vec{0}$ and $\overrightarrow{a_1} \neq \vec{0}$, then $\vec{b}$ is a linear combination of $\overrightarrow{a_1}$. if and only if $\vec{b}$ is parallel to $\overrightarrow{a_1}$.

Theorem 1.4.1.

The zero vector $\vec{0} \in \mathbb{R}^m$ is a linear combination of S.

Proof

Let $x_1 = x_2 = \cdots = x_n = 0$. Then

$$\vec{0} = 0\overrightarrow{a_1} + 0\overrightarrow{a_2} + \cdots + 0\overrightarrow{a_n} = x_1\overrightarrow{a_1} + x_2\overrightarrow{a_2} + \cdots + x_n\overrightarrow{a_n}.$$

The result follows.

The following result shows that if $\vec{b}$ is a linear combination of *S* and we add finitely many vectors to *S*, then $\vec{b}$ is a linear combination of all these vectors.

Theorem 1.4.2.

Assume that $\vec{b}$ is a linear combination of S. Let $\vec{b}_1, \overrightarrow{b_2}, \cdots, \overrightarrow{b_k}$ be vectors in $\mathbb{R}^m$. Then $\vec{b}$ is a linear combination of T, where

$$T = \left\{\overrightarrow{a_1}, \overrightarrow{a_2}, \cdots, \overrightarrow{a_n}, \overrightarrow{b_1}, \overrightarrow{b_2}, \cdots, \overrightarrow{b_k}\right\}.$$

Proof

To show that $\vec{b}$ is a linear combination of *T*, we need to find $n + k$ numbers, say $y_1, \cdots, y_n, y_{n+1}, \cdots, y_{n+k}$ such that

(1.4.4)

$$\vec{b} = y_1 \overrightarrow{a_1} + y_2 \overrightarrow{a_2} + \cdots + y_n \overrightarrow{a_n} + y_{n+1} \overrightarrow{b_1} + \cdots + y_{n+k} \overrightarrow{b_k}.$$

Because $\vec{b}$ is a linear combination of *S*, by **Definition 1.4.1**, there exist $\left\{x_i \in \mathbb{R} : i \in I_n\right\}$ such that

(1.4.5)

$$\vec{b} = x_1 \overrightarrow{a_1} + x_2 \overrightarrow{a_2} + \cdots + x_n \overrightarrow{a_n}.$$

Comparing **(1.4.4)** with **(1.4.5)**, we can choose

$$y_1 = x_1, \cdots, y_n = x_n, y_{n+1} = y_{n+2} = \cdots = y_{n+k} = 0.$$

Hence, we have

$$\begin{aligned} \vec{b} &= x_1 \overrightarrow{a_1} + x_2 \overrightarrow{a_2} + \cdots + x_n \overrightarrow{a_n} \\ &= x_1 \overrightarrow{a_1} + x_2 \overrightarrow{a_2} + \cdots + x_n \overrightarrow{a_n} + 0\overrightarrow{b_1} + \cdots + 0\overrightarrow{b_k}. \end{aligned}$$

By **Definition 1.4.1**, $\vec{b}$ is a linear combination of *T*.

Example 1.4.3.

Let $\overrightarrow{a_1} = (1,\ 2,\ 3)^T$. Show that the following assertions hold.

1. $\vec{0} = (0,\ 0,\ 0)^T$ is a linear combination of $\overrightarrow{a_1}$.
2. Let $\vec{b} = (2,\ 4,\ 6)^T$. Then $\vec{b}$ is a linear combination of $\overrightarrow{a_1}$.
3. Let $\overrightarrow{a_2} = (1,\ 0,\ 1)^T$ and $\overrightarrow{a_3} = (1,\ 1,\ 1)^T$. Then $\vec{b} = (2,\ 4,\ 6)^T$ is a linear combination of $\overrightarrow{a_1},\ \overrightarrow{a_2},$ and $\overrightarrow{a_3}$.

Solution

The result (1) follows from **Theorem 1.4.1** . Because $\vec{b} = 2\overrightarrow{a_1},$ the result (2) follows from **Definition 1.4.1** . Because $\vec{b}$ is a linear combination of $\overrightarrow{a_1},$ the result (3) follows from **Theorem 1.4.2** .

By **(1.4.3)** , and **Definition 1.1.5** , we obtain

(1.4.6)

$$
\begin{aligned}
\vec{b} &= x_1\overrightarrow{a_1} + x_2\overrightarrow{a_2} + \cdots + x_n\overrightarrow{a_n} \\
&= x_1\begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{pmatrix} + x_2\begin{pmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{pmatrix} + \cdots + x_n\begin{pmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{pmatrix} \\
&= \begin{pmatrix} a_{11}x_1 \\ a_{21}x_1 \\ \vdots \\ a_{m1}x_1 \end{pmatrix} + \begin{pmatrix} a_{12}x_2 \\ a_{22}x_2 \\ \vdots \\ a_{m2}x_2 \end{pmatrix} + \cdots + \begin{pmatrix} a_{1m}x_n \\ a_{2m}x_n \\ \vdots \\ a_{mn}x_n \end{pmatrix} \\
&= \begin{pmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ a_{12}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \end{pmatrix}.
\end{aligned}
$$

By **Definition 1.1.4**, we have

(1.4.7)

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

Now we introduce spanning spaces, which will be used in **Section 4.6** and Chapters 5,7, and 8. By **Definition 1.4.1**, the vector

$$x_1\overrightarrow{a_1} + x_2\overrightarrow{a_2} + \cdots + x_n\overrightarrow{a_n}$$

is a linear combination of *S* for each set of $x_1, \cdots, x_n \in \mathbb{R}$. We collect all these linear combinations together as a set called a spanning space of *S*.

Definition 1.4.2.

The spanning space of *S*, denoted by span *S*, is the set of all linear combinations of the vectors in *S*, that is,

(1.4.8)

$$\text{span } S = \left\{x_1\overrightarrow{a_1} + x_2\overrightarrow{a_2} + \cdots + x_n\overrightarrow{a_n} : x_i \in \mathbb{R},\ i \in I_n\right\}.$$

By **Definitions 1.4.1** and **1.4.2**, we have

Theorem 1.4.3.

The following assertions are equivalent.

1. *A vector* $\vec{b} \in$ span *S.*
2. $\vec{b}$ *is a linear combination of* $\overrightarrow{a_1}, \overrightarrow{a_2}, \cdots, \overrightarrow{a_n}$.

Let $\overrightarrow{a_1}, \overrightarrow{a_2}, \overrightarrow{b}$ be the same as in **Example 1.4.2**. By **Example 1.4.2**, $\vec{b}$ is a linear combination of $\overrightarrow{a_1}$ and $\overrightarrow{a_1}$. By **Theorem 1.4.3**, $\vec{b} \in \operatorname{span}\{\overrightarrow{a_1}, \overrightarrow{a_2}\}$.

Example 1.4.4.

Let $\overrightarrow{a_1} = \begin{pmatrix} 6 \\ 3 \end{pmatrix}$, $\overrightarrow{a_2} = \begin{pmatrix} -4 \\ -2 \end{pmatrix}$, and $\vec{b} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$. Show that $\vec{b} \in \operatorname{span}\{\overrightarrow{a_1}, \overrightarrow{a_2}\}$.

Solution

Let $x_1\overrightarrow{a_1} + x_2\overrightarrow{a_2} = \vec{b}$. Then

$$x_1\begin{pmatrix} 6 \\ 3 \end{pmatrix} + x_2\begin{pmatrix} -4 \\ -2 \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}.$$

This implies

$$\begin{cases} 6x_1 - 4x_2 & = & 2 \\ 3x_1 - 2x_2 & = & 1. \end{cases}$$

The above two equations are the same, so the system is equivalent to the single equation $3x_1 - 2x_2 = 1$. Let $x_2 = 1$. Then $3x_1 - 2x_2 = 1$ implies

$$x_1 = \frac{1}{3}(1 + 2x_2) = \frac{1}{3}(1 + 2) = 1.$$

Hence, $\vec{b} = \vec{a_1} + \vec{a_2}$ and $\vec{b} \in \operatorname{span}\{\vec{a_1}, \vec{a_2}\}$.

Example 1.4.5.

Let $\vec{e_1} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, $\vec{e_2} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$. Show $\mathbb{R}^2 = \operatorname{span}\{\vec{e_1}, \vec{e_2}\}$.

Solution

$$\operatorname{span}\left\{\vec{e_1}, \vec{e_2}\right\} = \left\{x_1\vec{e_1}, x_2\vec{e_2} : x_1, x_2 \in \mathbb{R}\right\} = \left\{(x_1, x_2) : x_1, x_2 \in \mathbb{R}\right\} = \mathbb{R}^2.$$

Similarly, we can show

(1.4.9)

$$\mathbb{R}^n = \operatorname{span}\left\{\vec{e_1}, \vec{e_2}, \cdots, \vec{e_n}\right\},$$

where $\vec{e}$, $\vec{e_2}$, $\cdots$, $\vec{e_n}$ are the standard vectors in $\mathbb{R}^n$ given in (1.1.1).

Example 1.4.6.

Let $\overrightarrow{e_1} = (1,\ 0,\ 0)$ and $\overrightarrow{e_2} = (0,\ 1,\ 0)$. Find $\operatorname{span}\left\{\overrightarrow{e_1},\ \overrightarrow{e_2}\right\}$.

Solution

$$\operatorname{span}\left\{\overrightarrow{e_1},\ \overrightarrow{e_2}\right\} = \left\{x\overrightarrow{e_1} + y\overrightarrow{e_2} : x,\ y \in \mathbb{R}\right\} = \left\{(x,\ y,\ 0) : x,\ y \in \mathbb{R}\right\}.$$

The above example shows that the spanning space $\operatorname{span}\left\{\overrightarrow{e_1},\ \overrightarrow{e_2}\right\}$ is the *xy*-plane in $\mathbb{R}^3$.

Exercises

1. Let $\overrightarrow{a_1} = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$, $\overrightarrow{a_2} = \begin{pmatrix} 1 \\ 3 \\ 2 \end{pmatrix}$, and $\overrightarrow{a_3} = \begin{pmatrix} -1 \\ 1 \\ 3 \end{pmatrix}$.

   i. Let $\vec{b} = (-1,\ -9,\ -4)^T$. Verify whether $2\overrightarrow{a_1} - 3\overrightarrow{a_2} = \vec{b}$.

   ii. Let $\vec{b} = (3,\ 6,\ 6)^T$. Verify whether $\overrightarrow{a_1} - 2\overrightarrow{a_2} = \vec{b}$.

   iii. Let $\vec{b} = (2,\ 7,\ 7)^T$. Verify whether $\overrightarrow{a_1} - 2\overrightarrow{a_2} + \overrightarrow{a_3} = \vec{b}$.

2. Let $\overrightarrow{\ }_1 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$, $\overrightarrow{a_2} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$, and $\vec{b} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$. Determine whether $\vec{b}$ is a linear combination of $\overrightarrow{a_1},\ \overrightarrow{a_2}$.

3. Let $\overrightarrow{a_1} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ and $\overrightarrow{a_2} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$. Show whether $\vec{b} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ is a linear combination of $\overrightarrow{a_1}, \overrightarrow{a_2}$.

4. Let $\overrightarrow{a_1} = (1,\ -1,\ 1)^T$, $\overrightarrow{a_2} = (0,\ 1,\ 1)^T$, $\overrightarrow{a_3} = (1,\ 3,\ 2)^T$, and $\overrightarrow{a_4} = (0,\ 0,\ 1)^T$.

   1. Is $\vec{0} = (0,\ 0,\ 1)^T$ a linear combination of $\overrightarrow{a_1}, \overrightarrow{a_2}, \overrightarrow{a_3}, \overrightarrow{a_4}$?
   2. Is $\vec{b} = (1,\ 0,\ 2)^T$ a linear combination of $\overrightarrow{a_1}, \overrightarrow{a_2}, \overrightarrow{a_3}, \overrightarrow{a_4}$?

5. Let $\overrightarrow{a_1} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$, $\overrightarrow{a_2} = \begin{pmatrix} 2 \\ -1 \end{pmatrix}$, and $\vec{b} = \begin{pmatrix} 4 \\ 3 \end{pmatrix}$. Show that $\vec{b} \in \text{span}\left\{\overrightarrow{a_1}, \overrightarrow{a_2}\right\}$.

Chapter **2** Matrices

2.1 Matrices

Definition of a matrix

Definition 2.1.1.

An $m \times n$ matrix *A* is a rectangular array of *mn* numbers arranged in *m* rows and *n* columns:

(2.1.1)

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

The *ij*th component of *A*, denoted a_{ij}, is the number appearing in the *i*th row and *j*th column of *A*. a_{ij}, is called an (*i*, *j*)-entry of *A*, or entry or element of *A* if there is no confusion. Sometimes, it is useful to write $A = (a\ )$. Capital letters are usually applied to denote matrices. $m \times n$ is called the size of *A*.

Let

(2.1.2)

$$\vec{a}_1 = \begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{pmatrix}, \ \vec{a} = \begin{pmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{pmatrix}, \ \cdots, \ \vec{a}_n = \begin{pmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{pmatrix}.$$

The vectors $\vec{a}_1, \ \vec{a}_2, \ \cdots, \ \vec{a}_n$ are called the column vectors of the matrix *A*. We can rewrite the matrix *A* as

(2.1.3)

$$A = (\vec{a}_1, \ \vec{a}_2, \ \cdots, \ \vec{a}_n).$$

Let

(2.1.4)

$$\begin{aligned} \vec{r}_1 &= (a_{11}, \ a_{12}, \ \cdots, \ a_{1n}), \\ \vec{r}_2 &= (a_{21}, \ a_{22}, \ \cdots, \ a_{2n}), \\ &\vdots \\ \vec{r}_m &= (a_{m1}, \ a_{m2}, \ \cdots, \ a_{mn}). \end{aligned}$$

The vectors $\overrightarrow{r_1}, \overrightarrow{r_2}, \cdots, \overrightarrow{r_m}$ are called the row vectors of *A*. We can rewrite the matrix *A* as

(2.1.5)

$$A = \begin{pmatrix} \overrightarrow{r_1} \\ \overrightarrow{r_2} \\ \vdots \\ \overrightarrow{r_m} \end{pmatrix}.$$

Note that $\overrightarrow{a_1}, \overrightarrow{a_2}, \cdots, \overrightarrow{a_n}$ are in $\mathbb{R}^m$ while $\overrightarrow{r_1}, \overrightarrow{r_2}, \cdots, \overrightarrow{r_m}$ are in $\mathbb{R}^m$.

Example 2.1.1.

The size of $A = (2)$ is $1 \times 1,$ the size of $\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$ is $2 \times 2,$ the size of $\begin{pmatrix} 1 & 2 & 0 \\ 3 & 4 & -1 \end{pmatrix}$ is $2 \times 3,$ and the size of $\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$ is 4 × 3.

Example 2.1.2.

Let $A = \begin{pmatrix} 1 & 0 & 0 & 1 \\ 2 & 2 & 1 & 1 \\ 0 & 1 & 2 & 4 \end{pmatrix}$. Use the column vectors and the row vectors of *A* to rewrite *A*.

Solution

Let

$$\overrightarrow{a_1} = \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}, \overrightarrow{a_2} = \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix}, \overrightarrow{a_3} = \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix}, \text{ and } \overrightarrow{a_4} = \begin{pmatrix} 1 \\ 1 \\ 4 \end{pmatrix}.$$

Then *A* can be rewritten as $A = (\overrightarrow{a_1}\overrightarrow{a_2}\overrightarrow{a_3}\overrightarrow{a_4})$.

Let $\overrightarrow{r_1} = (1,\ 0,\ 0,\ 1), \overrightarrow{r_2} = (2,\ 2,\ 1,\ 1)$, and $\overrightarrow{r_3} = (0,\ 1,\ 2,\ 4)$. Then *A* can be rewritten as $A = \begin{pmatrix} \overrightarrow{r_1} \\ \overrightarrow{r_2} \\ \overrightarrow{r_3} \end{pmatrix}$.

Definition 2.1.2.

1. An $m \times n$ matrix is called a zero matrix if all entries of *A* are zero.
2. A $1 \times n$ matrix is called a row matrix.
3. An $m \times 1$ matrix is called a column matrix.

Sometimes, we denote by 0 a zero matrix if there is no confusion. A $1 \times n$ matrix and an $m \times 1$ matrix can be treated as a row vector and a column vector, respectively.

Example 2.1.3.

1. The following are zero matrices.

$$A = (0), \quad B = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \text{and} \quad C = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

2. (0 0 0) is a 1×3 row matrix and (1 3 5 7) is a 1×4 row matrix.

3. $\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$ is a 4×1 column matrix and $\begin{pmatrix} 2 \\ 1 \\ 3 \end{pmatrix}$ is a 3×1 column matrix.

Transpose of a matrix

Let $A = (a_{ij})$ bean $m \times n$ matrix defined in **(2.1.1)** . The transpose of *A*, denoted by A^T, is the $n \times m$ matrix obtained by interchanging the rows and columns of *A*. Hence, $A^T = (a_{ji})$ or

(2.1.6)

: 100%

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

From **(2.1.6)** , we see that the *i*th column of A^T and the *i*th row of *A* are the same. It is obvious that $\left(A^T\right)^T = A$.

Example 2.1.4.

Find the transposes of the following matrices:

$$A = (8)\ B = (1\ 3\ 5) \qquad C = \begin{pmatrix} 7 & 9 \\ 18 & 31 \\ 52 & 68 \end{pmatrix}.$$

Solution

$$A^T = (8),\ B^T = \begin{pmatrix} 1 \\ 3 \\ 5 \end{pmatrix},\ C^T = \begin{pmatrix} 7 & 18 & 52 \\ 9 & 31 & 68 \end{pmatrix}.$$

Operations on matrices

Definition 2.1.3.

Two matrices *A* and *B* are said to be equal if the following conditions hold:

1. *A* and *B* have the same size.
2. All the corresponding entries are the same.

If *A* and *B* are equal, then we write $A = B$. Let $A = (a_{ij})$ and $B = (b_{ij})$ be two $m \times n$ matrices. Then $A = B$ if and only if

$$a_{ij} = b_{ij} \quad \text{for all } i \in I_m \text{ and } j \in I_n.$$

Example 2.1.5.

Let $A = \begin{pmatrix} 2 & 1 \\ 3 & x^2 \end{pmatrix}$ and $B = \begin{pmatrix} 2 & 1 \\ 3 & 4 \end{pmatrix}$. Find all $x \in \mathbb{R}$ such that $A = B$.

Solution

Note that *A* and *B* have the same size. Hence, if $x^2 = 4$, then $A = B$. This implies that when $x = 2$ or $x = -2$, $A = B$.

Example 2.1.6.

Let $A = \begin{pmatrix} a & 2x + y \\ b & 4x + 3y \end{pmatrix}$ and $B = \begin{pmatrix} 1 & 1 \\ 3 & 6 \end{pmatrix}$. Find all $a, b, x, y \in \mathbb{R}$ such that $A = B$

Solution

Note that *A* and *B* have the same size. Hence, $A = B$ if $a = 1$, $b = 3$, and *x*, *y* satisfy the following system.

$$\begin{cases} 2x + y = 1, \\ 4x + 3y = 6. \end{cases}$$

Solving the above system, we obtain $x = -3/2$ and $y = 4$. Hence, when $a = 1,\ b = 3,\ x = -3/2,$ and $y = 4,\ A = B$.

Example 2.1.7.

Let $A = (1\ \ 2)$ and $B = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$ be two matrices. Determine if $A = B$.

Solution

The sizes of *A* and *B* are 1×2 and 2×1 respectively. Because *A* and *B* have different sizes, $A \neq B$.

Note that in **Example 2.1.7** , if we treat *A* and *B* as vectors, then they are equal vectors.

Definition 2.1.4.

Let $A = (a_{ij}),\ B = (b_{ij})$ be $m \times n$ matrices and *k* a real number. We define

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

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

Scalar multiplication: $kA = \begin{pmatrix} ka_{11} & ka_{12} & \cdots & ka_{1n} \\ ka_{21} & ka_{22} & \cdots & ka_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ ka_{m1} & ka_{m2} & \cdots & ka_{mn} \end{pmatrix}$.

Example 2.1.8.

Let

$$A = \begin{pmatrix} 3 & 4 & 2 \\ 2 & -3 & 0 \end{pmatrix} \text{ and } B = \begin{pmatrix} -4 & 1 & 0 \\ 5 & -6 & 1 \end{pmatrix}.$$

Compute

i. $A + B$;
ii. $A - B$;
iii. $-A$
iv. 3*A*.

Solution

i. $A + B = \begin{pmatrix} 3-4 & 4+1 & 2+0 \\ 2+5 & -3-6 & 0+1 \end{pmatrix} = \begin{pmatrix} -1 & 5 & 2 \\ 7 & -9 & 1 \end{pmatrix}$.

ii. $A - B = \begin{pmatrix} 3-(-4) & 4-1 & 2-0 \\ 2-5 & -3-(-6) & 0-1 \end{pmatrix} = \begin{pmatrix} 7 & 3 & 2 \\ -3 & 3 & -1 \end{pmatrix}$.

iii. $-A = \begin{pmatrix} (-1)(3) & (-1)(4) & (-1)(2) \\ (-1)(2) & (-1)(-3) & (-1)(0) \end{pmatrix} = \begin{pmatrix} -3 & -4 & -2 \\ -2 & 3 & 0 \end{pmatrix}.$

iv. $3A = \begin{pmatrix} (3)(3) & (3)(4) & (3)(2) \\ (3)(2) & (3)(-3) & (3)(0) \end{pmatrix} = \begin{pmatrix} 9 & 12 & 6 \\ 6 & -9 & 0 \end{pmatrix}.$

The following result can be easily proved and its proof is left to the reader.

Theorem 2.1.1.

Let A, B, and C be $m \times n$ matrices and let $\alpha, \beta \in \mathbb{R}$. Then

1. $A + 0 = A$,
2. $0A = 0$,
3. $A + B = B + A$,
4. $(A + B) + C = A + (B + C)$,
5. $\alpha(A + B) = \alpha A + \alpha B$,
6. $(\alpha + \beta)A = \alpha A + \beta A$,
7. $(\alpha\beta)A = \alpha(\beta A)$,
8. $1A = A$,
9. $(A + B)^T = A^T + B^T$,
10. $(A - B)^T = A^T - B^T$,
11. $(\alpha A)^T = \alpha A^T$.

Example 2.1.9.

Let

$$A = \begin{pmatrix} 1 & 0 & 1 \\ 2 & 1 & 1 \\ 0 & 0 & 2 \end{pmatrix} \quad B = \begin{pmatrix} 1 & 2 & 3 \\ 0 & 0 & 0 \\ 1 & 4 & 2 \end{pmatrix} \quad C = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 2 \end{pmatrix}$$

Compute $2A - B + C$ and $[2(A+B)]^T$.

Solution

$2A - B + C = (2A - B) + C$

$$= \left[2\begin{pmatrix} 1 & 0 & 1 \\ 2 & 1 & 1 \\ 0 & 0 & 2 \end{pmatrix} - \begin{pmatrix} 1 & 2 & 3 \\ 0 & 0 & 0 \\ 1 & 4 & 2 \end{pmatrix}\right] + \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 2 \end{pmatrix}$$

$$= \left[\begin{pmatrix} 2 & 0 & 2 \\ 4 & 2 & 2 \\ 0 & 0 & 4 \end{pmatrix} - \begin{pmatrix} 1 & 2 & 3 \\ 0 & 0 & 0 \\ 1 & 4 & 2 \end{pmatrix}\right] + \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 2 \end{pmatrix}$$

$$= \begin{pmatrix} 1 & -2 & -1 \\ 4 & 2 & 2 \\ -1 & -4 & 2 \end{pmatrix} + \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 2 \end{pmatrix} = \begin{pmatrix} 2 & -2 & -1 \\ 4 & 3 & 3 \\ -1 & -4 & 4 \end{pmatrix}.$$

$[2(A+B)]^T = 2(A+B)^T = 2(A^T + B^T)$

$$= 2\left[\begin{pmatrix} 1 & 2 & 0 \\ 0 & 1 & 0 \\ 1 & 1 & 2 \end{pmatrix} + \begin{pmatrix} 1 & 0 & 1 \\ 2 & 0 & 4 \\ 3 & 0 & 2 \end{pmatrix}\right] = 2\begin{pmatrix} 2 & 2 & 1 \\ 2 & 1 & 4 \\ 4 & 1 & 4 \end{pmatrix} = \begin{pmatrix} 4 & 4 & 2 \\ 4 & 2 & 8 \\ 8 & 2 & 8 \end{pmatrix}.$$

Example 2.1.10.

Find the matrix A if

$$\left[2A^T - \begin{pmatrix} 1 & 0 \\ 2 & -1 \end{pmatrix}^T\right]^T = \begin{pmatrix} 0 & 1 \\ 0 & -1 \end{pmatrix}.$$

Solution

Because

$$\begin{aligned} \left[2A^T - \begin{pmatrix} 1 & 0 \\ 2 & -1 \end{pmatrix}^T\right]^T &= (2A^T)^T - \left[\begin{pmatrix} 1 & 0 \\ 2 & -1 \end{pmatrix}^T\right]^T \\ &= 2(A^T)^T - \begin{pmatrix} 1 & 0 \\ 2 & -1 \end{pmatrix} = 2A - \begin{pmatrix} 1 & 0 \\ 2 & -1 \end{pmatrix}, \end{aligned}$$

we obtain

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

and

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

Hence, $A = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ 2 & -2 \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ 1 & -1 \end{pmatrix}$.

Exercises

1. Find the sizes of the following matrices:

$$A = (2)\ B = \begin{pmatrix} 5 & 8 \\ 1 & 4 \end{pmatrix}\ C = \begin{pmatrix} 10 & 3 & 8 \\ 10 & 3 & 4 \end{pmatrix} D = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 10 & 2 \\ 0 & 0 & 1 \\ 6 & 9 & 10 \end{pmatrix}$$

2. Use column vectors and row vectors to rewrite each of the following matrices.

$$A = \begin{pmatrix} 3 & 3 & 2 \\ 1 & 8 & 1 \\ 5 & 4 & 10 \end{pmatrix}\ B = \begin{pmatrix} 1 & 9 & 7 & 2 \\ 3 & 10 & 8 & 7 \\ 4 & 3 & 10 & 4 \end{pmatrix}\ C = \begin{pmatrix} 9 & 8 & 7 \\ 6 & 5 & 4 \\ 3 & 2 & 1 \\ 0 & -1 & -2 \end{pmatrix}$$

3. Find the transposes of the following matrices:

$$C = (4)\ B = (\ 3\ \ 11\ \ 2\ )\quad C = \begin{pmatrix} 33 \\ 8 \\ 12 \end{pmatrix}\ D = \begin{pmatrix} 3 & 9 & -19 \\ -2 & 8 & -7 \\ 5 & 3 & -9 \end{pmatrix}$$

4. Let $A = \begin{pmatrix} 9 & 3 \\ 6 & x^2 \end{pmatrix}$ and $B = \begin{pmatrix} 9 & 3 \\ 6 & 4 \end{pmatrix}$. Find all $x \in \mathbb{R}$ such that $A = B$.

5. Let $C = \begin{pmatrix} 12 & 5 & 25 \\ 19 & 4 & 6 \end{pmatrix}$ and $D = \begin{pmatrix} 12 & 5 & x^2 \\ 19 & 4 & 6 \end{pmatrix}$. Find all $x \in \mathbb{R}$ such that $C = D$.

6. Let $E = \begin{pmatrix} 120 & 25 & 122 \\ 123 & 124 & 125 \\ 126 & 127 & 128 \end{pmatrix}$ and $F = \begin{pmatrix} 120 & x^2 & 122 \\ 123 & 124 & x^3 \\ 26x-4 & 127 & 128 \end{pmatrix}$. Find all $x \in \mathbb{R}$ such that $E = F$.

7. Let $A = \begin{pmatrix} a & b \\ 3x+2y & -x+y \end{pmatrix}$ and $B = \begin{pmatrix} 1 & 1 \\ 3 & 6 \end{pmatrix}$. Find all $a,\ b,\ x,\ y \in \mathbb{R}$ such that $A = B$.

8. Let

$$C = \begin{pmatrix} a+b & 2b-a \\ x-2y & 5x+3y \end{pmatrix} \text{ and } D = \begin{pmatrix} 6 & 0 \\ 8 & 14 \end{pmatrix}.$$

Find all $a,\ b,\ x,\ y \in \mathbb{R}$ such that $C = D$.

9. Let $A = \begin{pmatrix} -2 & 3 & 4 \\ 6 & -1 & -8 \end{pmatrix}$ and $B = \begin{pmatrix} 7 & -8 & 9 \\ 0 & -1 & 0 \end{pmatrix}$. Compute

i. $A + B$;
ii. $-A$;
iii. $4A - 2B$;
iv. $100A + B$.

10. Let $A = \begin{pmatrix} 9 & 5 & 1 \\ 8 & 0 & 0 \\ 0 & 3 & 2 \end{pmatrix}$, $B = \begin{pmatrix} 2 & 2 & 2 \\ 0 & 0 & 0 \\ 4 & 6 & 8 \end{pmatrix}$, and $C = \begin{pmatrix} 4 & 7 & 0 \\ 0 & 3 & 1 \\ 0 & 0 & 2 \end{pmatrix}$.

Compute

1. $3A - 2B + C$,
2. $[3(A+B)]^T$,
3. $\left(4A + \frac{1}{2}B - C\right)^T$.

11. Find the matrix A if $\left[\left(3A^T\right) - \begin{pmatrix} -7 & -2 \\ -6 & 9 \end{pmatrix}^T\right]^T = \begin{pmatrix} -5 & -10 \\ 33 & 12 \end{pmatrix}$.

12. Find the matrix B if

$$\left[\frac{1}{2}B + \begin{pmatrix} 6 & 3 \\ 8 & 3 \\ 1 & 4 \end{pmatrix}\right]^T - 3\begin{pmatrix} -5 & 6 \\ 8 & -9 \\ -4 & 2 \end{pmatrix}^T = \begin{pmatrix} 23 & -16 & 17 \\ -16 & 26.5 & 2 \end{pmatrix}.$$

: 100%

2.2 Product of two matrices

Product of a matrix and a vector

Definition 2.2.1.

Let A be the same as in **(2.1.1)** and $\vec{X} = (x_1, \cdots, x_n)^T$. The product $A\vec{X}$ of A and $\vec{X}$ is defined by

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

Note that $A\vec{X}$ is a vector in $\mathbb{R}^m$, where m is the number of rows of A while $\vec{X}$ is a vector in $\mathbb{R}^n$, where n is the number of columns of A. $A\vec{X}$ is called the image of $\vec{X}$ under A.

We denote by $A(\mathbb{R}^n)$ the set of all the vectors $A\vec{X}$ as $\vec{X}$ varies in $\mathbb{R}^n$, that is,

(2.2.1)

$$A(\mathbb{R}^n) = \left\{A\vec{X} \in \mathbb{R}^m : \vec{X} \in \mathbb{R}^n\right\}.$$

By **Definition 2.2.1**, we see

(2.2.2)

$$A\vec{X} = \begin{pmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}a_x \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \end{pmatrix} = \begin{pmatrix} \overrightarrow{r_1} \cdot \vec{X} \\ \overrightarrow{r_2} \cdot \vec{X} \\ \vdots \\ \overrightarrow{r_m} \cdot \vec{X} \end{pmatrix},$$

where $\overrightarrow{r_1}, \overrightarrow{r_2}, \cdots, \overrightarrow{r_m}$ are the row vectors of *A* given in **(2.1.4)** .

Example 2.2.1.

Let $A = \begin{pmatrix} 1 & 2 & 0 \\ 3 & -1 & 4 \end{pmatrix}$, $\vec{X} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$, $\overrightarrow{X_1} = \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}$, and $\overrightarrow{X_2} = \begin{pmatrix} 0 \\ 2 \\ 2 \end{pmatrix}$. Compute $A\vec{X}$, $A\overrightarrow{X_1}$, and $A\overrightarrow{X_2}$.

Solution

Let $\overrightarrow{r_1} = (1, 2, 0)$ and $\overrightarrow{r_2} = (1, -3, 4)$ be the row vectors of *A*. Then

$$\vec{r}_1 \cdot \vec{X} = (1,\ 2,\ 0)\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = (1)(x_1) + (2)(x_2) + (0)(x_3) = x_1 + 2x_2$$

and

$$\vec{r}_2 \cdot \vec{X} = (1,\ -3,\ 4)\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = (3)(x_1) + (-1)(x_2) + (4)(x_3) = 3x_1 - x_2 + 4x_3.$$

Hence,

$$A\vec{X} = \begin{pmatrix} 1 & 2 & 0 \\ 3 & -1 & 4 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} \vec{r}_1 \cdot \vec{X} \\ \vec{r}_2 \cdot \vec{X} \end{pmatrix} = \begin{pmatrix} x_1 + 2x_2 \\ 3x_1 - x_2 + 4x_3 \end{pmatrix}.$$

We can rewrite $A\vec{X}$ in a simple way.

$$
\begin{aligned}
A\vec{X} &= \begin{pmatrix} 1 & 2 & 0 \\ 3 & -1 & 4 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} (1)(x_1) + (2)(x_2) + (0)(x_3) \\ (3)(x_1) + (-1)(x_2) + (4)(x_3) \end{pmatrix} \\
&= \begin{pmatrix} x_1 + 2x_2 \\ 3x_1 - x_2 + 4x_3 \end{pmatrix}.
\end{aligned}
$$

$$
\overrightarrow{AX_1} = \begin{pmatrix} 1 & 2 & 0 \\ 3 & -1 & 4 \end{pmatrix}\begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} = \begin{pmatrix} (1)(1) + (2)(2) + (0)(1) \\ (3)(1) + (-1)(2) + (4)(1) \end{pmatrix} = \begin{pmatrix} 5 \\ 5 \end{pmatrix}.
$$

$$
\overrightarrow{AX_2} = \begin{pmatrix} 1 & 2 & 0 \\ 3 & -1 & 4 \end{pmatrix}\begin{pmatrix} 0 \\ 2 \\ 2 \end{pmatrix} = \begin{pmatrix} (1)(0) + (2)(2) + (0)(2) \\ (3)(0) + (-1)(2) + (4)(2) \end{pmatrix} = \begin{pmatrix} 4 \\ 6 \end{pmatrix}.
$$

Example 2.2.2.

Let

$$
A = \begin{pmatrix} -4 & 1 \\ 5 & -1 \\ 3 & 0 \\ 2 & 4 \end{pmatrix}, \quad \vec{X} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \text{ and } \overrightarrow{X_1} = \begin{pmatrix} 0 \\ 4 \end{pmatrix}.
$$

Compute $A\vec{X}$ and $\overrightarrow{AX_1}$.

Solution

$$A\vec{X} = \begin{pmatrix} -4 & 1 \\ 5 & -1 \\ 3 & 0 \\ 2 & 4 \end{pmatrix}\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} (-4)(1) + (1)(1) \\ (5)(1) + (-1)(1) \\ (3)(1) + (0)(1) \\ (2)(1) + (4)(1) \end{pmatrix} = \begin{pmatrix} -3 \\ 4 \\ 3 \\ 6 \end{pmatrix}.$$

$$\overrightarrow{AX_1} = \begin{pmatrix} -4 & 1 \\ 5 & -1 \\ 3 & 0 \\ 2 & 4 \end{pmatrix}\begin{pmatrix} 0 \\ 4 \end{pmatrix} = \begin{pmatrix} -4(0) + (1)(4) \\ (5)(0) + (-1)(4) \\ (3)(0) + (0)(4) \\ (2)(0) + (4)(4) \end{pmatrix} = \begin{pmatrix} 4 \\ -4 \\ 0 \\ 16 \end{pmatrix}.$$

From **(2.2.2)** , a matrix and a vector can be multiplied together only when the number of columns of the matrix equals the number of components of the vector. Hence, if the vectors $\vec{r}_i$ and $\vec{X}$ have a different number of components, then the scalar product $\overrightarrow{r_i} \cdot \vec{X}$ is not defined.

Example 2.2.3.

Let

$$A = \begin{pmatrix} -1 & 1 \\ 2 & -1 \\ 3 & 1 \end{pmatrix}, \ \overrightarrow{X_1} = \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}, \ \overrightarrow{X_2} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \text{ and } \overrightarrow{X_3} = \begin{pmatrix} 1 \\ 1 \\ 2 \\ 4 \end{pmatrix}.$$

Determine whether $\overrightarrow{AX_i}$ is defined for $i = 1, 2, 3$.

Solution

$\overrightarrow{AX_2}$ is defined, but $\overrightarrow{AX_1}$ and $\overrightarrow{AX_3}$ are not defined.

Theorem 2.2.1.

Let A be the same as in **(2.1.1)** $, \vec{X}, \vec{Y} \in \mathbb{R}^n$ *and* $\alpha, \beta \in \mathbb{R}$.

Then

$A(\alpha\vec{X} + \beta\vec{Y}) = \alpha(A\bar{X}) + \beta(A\vec{Y})$.

Example 2.2.4.

Let

$$A = \begin{pmatrix} 1 & 1 & 0 \\ 2 & 0 & -1 \\ 1 & 0 & 1 \end{pmatrix}, \ \vec{X} = \begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix}, \ \text{and } \vec{Y} = \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}.$$

Compute $A(2\vec{X} + 3\vec{Y})$.

Solution

$$A(2\vec{X}+3\vec{Y}) = 2(A\vec{X}) + 3(A\vec{Y})$$

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

$$= 2\begin{pmatrix} 3 \\ 3 \\ 0 \end{pmatrix} + 3\begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix} = \begin{pmatrix} 6 \\ 6 \\ 0 \end{pmatrix} + \begin{pmatrix} 3 \\ 3 \\ -3 \end{pmatrix} = \begin{pmatrix} 9 \\ 9 \\ -3 \end{pmatrix}.$$

Let $\overrightarrow{a_1}, \overrightarrow{a_2}, \cdots, \overrightarrow{a_n}$ be the column vectors of *A* given in **(2.1.2)** . By **(1.4.6)** and **(2.2.2)** , $A\vec{X}$ can be expressed by a linear combination of these column vectors:

(2.2.3)

$$A\vec{X} = x_1\overrightarrow{a_1} + x_2\overrightarrow{a_2} + \cdots + x_n\overrightarrow{a_n}.$$

Example 2.2.5.

Let $A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}$ and $\vec{X} = \begin{pmatrix} 2 \\ 3 \\ 4 \end{pmatrix}$. Write $A\vec{X}$ as a linear combination of the column vectors of *A*.

Solution

$$A\vec{X} = 2\begin{pmatrix} 1 \\ 4 \\ 7 \end{pmatrix} + 3\begin{pmatrix} 2 \\ 5 \\ 8 \end{pmatrix} + 4\begin{pmatrix} 3 \\ 6 \\ 9 \end{pmatrix}.$$

The following result gives the relation between $A(\mathbb{R}^n)$ and the spanning space.

Theorem 2.2.2.

Let $S = \left\{\overrightarrow{a_1},\ \overrightarrow{a_2},\ \cdots,\ \overrightarrow{a_n}\right\}$ *be the set of the column vectors of A given in* **(2.1.2)** *. Then*

(2.2.4)

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

Proof

Let $\vec{b} = (b_1,\ b_2,\ \cdots,\ b_m)^T \in A(\mathbb{R}^n)$. By **(2.2.1)** , there exists a vector $\vec{X} \in \mathbb{R}^n$ such that $T\vec{X} = \vec{b}$. By **(2.2.3)** and **(1.4.8)** ,

$$\vec{b} = A\vec{X} = x_1\overrightarrow{a_1} + x_2\overrightarrow{a_2} + \cdots + x_n\overrightarrow{a_n} \in \text{span } S.$$

This shows that $A(\mathbb{R}^n)$ is a subset of span *S*. Conversely, if $\vec{b} \in \text{span } S,$ then by **Theorem 1.4.3** , there exists a vector $\vec{X} \in \mathbb{R}^n$ such that

$$\vec{b} = x_1\overrightarrow{a_1} + x_2\overrightarrow{a_2} + \cdots + x_n\overrightarrow{a_n}.$$

This, together with **(2.2.3)** and **(2.2.1)**, implies $\vec{b} = A\vec{X} \in A(\mathbb{R}^n)$. This shows that span *S* is a subset of $A(\mathbb{R}^n)$. We have shown that $A(\mathbb{R}^n)$. is a subset of span *S* and span *S* is a subset of $A(\mathbb{R}^n)$. Hence, **(2.2.4)** holds.

Product of two matrices

Let *A* be the same as in **(2.1.1)** and $\overrightarrow{r_1}, \overrightarrow{r_2}, \cdots, \overrightarrow{r_m}$ be the row vectors of *A* given in **(2.1.4)**. Let

$$B_{n\times r} = \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1r} \\ b_{21} & b_{22} & \cdots & b_{2r} \\ \cdots & \cdots & \ddots & \cdots \\ b_{n1} & b_{n2} & \cdots & b_{nr} \end{pmatrix}$$

and let

$$\overrightarrow{b_1} = \begin{pmatrix} b_{11} \\ b_{21} \\ \vdots \\ b_{n1} \end{pmatrix}, \overrightarrow{b_2} = \begin{pmatrix} b_{12} \\ b_{22} \\ \vdots \\ b_{n2} \end{pmatrix}, \cdots, \overrightarrow{b_r} = \begin{pmatrix} b_{1r} \\ b_{2r} \\ \vdots \\ c_{nr} \end{pmatrix}$$

be the column vectors of *B*.

Definition 2.2.2.

The product *AB* of *A* and *B* is defined by

(2.2.5)

$$AB = \begin{pmatrix} \vec{r_1} \cdot \vec{b_1} & \vec{r_1} \cdot \vec{b_2} & \cdots & \vec{r_1} \cdot \vec{br} \\ \vec{r_2} \cdot \vec{b_1} & \vec{r_2} \cdot \vec{b_2} & \cdots & \vec{r_2} \cdot \vec{b_r} \\ \vdots & \vdots & \ddots & \vdots \\ \vec{r_m} \cdot \vec{b_1} & \vec{r_m} \cdot \vec{b_2} & \cdots & \vec{r_m} \cdot \vec{b_r} \end{pmatrix}.$$

From Definition 2.2.5, we see that

(2.2.6)

$$AB = \begin{pmatrix} A\vec{b_1} \; A\vec{b_2} \cdots A\vec{b_r} \end{pmatrix}$$

and the size of *AB* is $m \times r$

The order for computing *AB* is to compute the first row

$$\vec{r_1} \cdot \vec{b_1}, \; \vec{r_1} \cdot \vec{b_2}, \; \cdots, \; \vec{r_1} \cdot \vec{b_r}$$

then the second row

$$\vec{r_2} \cdot \vec{b_1},\ \vec{r_2} \cdot \vec{b_2},\ \cdots,\ \vec{r_2} \cdot \vec{b_r}$$

and so on until the last row

$$\vec{r_m} \cdot \vec{b_1},\ \vec{r_m} \cdot \vec{b_2},\ \cdots,\ \vec{r_m} \cdot \vec{b_r}.$$

Example 2.2.6.

Let

$$A = \begin{pmatrix} 1 & 3 \\ -2 & 4 \end{pmatrix} \text{ and } B = \begin{pmatrix} 3 & -2 & 0 \\ 5 & 6 & 0 \end{pmatrix}.$$

Find *AB* and the sizes of *A*, *B*, and *AB*.

The order of computing *AB* is to compute the first row of *AB*,

$$\begin{aligned} (1 \ \ 3)\begin{pmatrix} 3 & -2 & 0 \\ 5 & 6 & 0 \end{pmatrix} &= ((1)(3) + (3)(5),\ (1)(-2) + (3)(6), (1)(0) + (3)(0)) \\ &= (18,\ 16,\ 0) \end{aligned}$$

and then compute the second row of *AB*,

$$\begin{aligned} (-2 \quad 4)\begin{pmatrix} 3 & -2 & 0 \\ 5 & 6 & 0 \end{pmatrix} &= ((-2)(3) + (4)(5),\ (-2)(-2) + (4)(6),\ (-2)(0) + (4)(0)) \\ &= (14,\ 28,\ 0). \end{aligned}$$

Solution

$$\begin{aligned}
AB &= \begin{pmatrix} 1 & 3 \\ -2 & 4 \end{pmatrix}\begin{pmatrix} 3 & -2 & 0 \\ 5 & 6 & 0 \end{pmatrix} \\
&= \begin{pmatrix} (1)(3)+(3)(5) & (1)(-2)+(3)(6) & (1)(0)+(3)(0) \\ (-2)(3)+(4)(5) & (-2)(-2)+(4)(6) & (-2)(0)+(4)(0) \end{pmatrix} \\
&= \begin{pmatrix} 18 & 16 & 0 \\ 14 & 28 & 0 \end{pmatrix}.
\end{aligned}$$

The sizes of *A*, *B*, and *AB* are $2 \times 2,\ 2 \times 3,\ 2 \times 3,$ respectively.

Example 2.2.7.

Let

$$A = \begin{pmatrix} 1 & 2 & 4 \\ 2 & 6 & 0 \end{pmatrix} \text{ and } B = \begin{pmatrix} 4 & 1 & 4 \\ 0 & -1 & 3 \\ 2 & 7 & 5 \end{pmatrix}.$$

Compute *AB* and find the sizes of *A*, *B*, and *AB*.

The order of computing *AB* is to compute the first row of *AB*,

$$\begin{aligned}(1\ \ 2\ \ 4)\begin{pmatrix}4 & 1 & 4\\ 0 & -1 & 3\\ 2 & 7 & 5\end{pmatrix} &= ((1)(4)+(2)(0)+(4)(2),\ (1)(1)+(2)(-1)+(4)(7),\ (1)(4)+(2)(3)+(4)(5)\\ &= (12,\ 27,\ 30)\end{aligned}$$

and then compute the second row of *AB*,

$$\begin{aligned}(2\quad 6\quad 0)\begin{pmatrix}4 & 1 & 4\\ 0 & -1 & 3\\ 2 & 7 & 5\end{pmatrix} &= ((2)(4)+(6)(0)+(0)(2),\ (2)(1)+(6)(-1)+(0)(7),\ (2)(4)+(6)(4)+(6)(3)+(0)(5)\\ &= (8,\ -4,\ 26).\end{aligned}$$

Solution

By computation, we have

$$\begin{aligned}AB &= \begin{pmatrix}1 & 2 & 4\\ 2 & 6 & 0\end{pmatrix}\begin{pmatrix}4 & 1 & 4\\ 0 & -1 & 3\\ 2 & 7 & 5\end{pmatrix}\\ &= \begin{pmatrix}(1)(4)+(2)(0)+(4)(2) & (1)(1)+(2)(-1)+(4)(7) & (1)(4)+(2)(3)+(4)(5)\\ (2)(4)+(6)(0)+(0)(2) & (2)(1)+(6)(-1)+(0)(7) & (2)(4)+(6)(3)+(0)(5)\end{pmatrix}\\ &= \begin{pmatrix}12 & 27 & 30\\ 8 & -4 & 26\end{pmatrix}.\end{aligned}$$

The sizes of , , and are $2\times 3,\ 3\times 3,$ and $2\times 3,$ respectively.

Example 2.2.8.

Let

$$A=\begin{pmatrix}2&0&-3\\4&1&5\end{pmatrix} \text{ and } B=\begin{pmatrix}7&-1&4&7\\2&5&0&-4\\-3&1&2&3\end{pmatrix}.$$

Compute *AB* and find the sizes of *A*, *B*, and *AB*.

We always follow the order like the above two examples to compute *AB*, but we do not need to mention the order every time when we compute *AB*.

Solution

$$AB=\begin{pmatrix}2&0&-3\\4&1&5\end{pmatrix}\begin{pmatrix}7&-1&4&7\\2&5&0&-4\\-3&1&2&3\end{pmatrix}=\begin{pmatrix}23&-5&2&5\\15&6&26&39\end{pmatrix}.$$

The sizes of *A*, *B*, and *AB* are $2\times 3,\ 3\times 4,$, and $2\times 4,$ respectively.

In **(2.2.6)** , for each $i=1,\ 2,\ \cdots,\ r,$ the product $A\vec{b}_i$ requires that the number of columns of the matrix *A* be equal to the number of components of the vector $\vec{b}_i$. Hence, the product of two matrices *A* and *B* requires that the number of columns of *A* be equal to the number of rows of *B*. Symbolically,

(2.2.7)

$$(m\ \times\ n)(n\ \times\ r)=m\ \times\ r.$$

The inner numbers must be the same and the outer numbers give the size of the product *AB*. Hence, if the inner numbers are not the same, the product *AB* is not defined.

Example 2.2.9.

1. Let *A*, *B*, and *C* be matrices such that $AB = C.$ Assume that the sizes of *A* and *C* are 5×3 and $5 \times 4.$ respectively. Find the size of *B*.
2. Let
$$A = \begin{pmatrix} 1 & 0 \\ 2 & 1 \\ 3 & 0 \end{pmatrix} \text{ and } B = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}.$$
Are *AB* and *BA* defined? If so, compute them. If not, explain why.
3. Let $A = \begin{pmatrix} 1 & 3 \\ -2 & 4 \end{pmatrix}$ and $B = \begin{pmatrix} 3 & -2 \\ 5 & 6 \end{pmatrix}.$ Are *AB* and *BA* defined? If so, compute them. If not, explain why. Is *AB* equal to *BA*?

Solution

1. Because $AB = C$ and the sizes of *A* and *C* are 5×3 and $5 \times 4,$ respectively, by **(2.2.7)** , the size of *B* is $3 \times 4,$
2. The size of *A* is 3×2 and the size of *B* is $2 \times 2.$ Hence, *AB* is defined because the number of columns of *A* and the number of rows of *B* are the same and equal 2.

$$AB = \begin{pmatrix} 1 & 0 \\ 2 & 1 \\ 3 & 0 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 5 & 8 \\ 3 & 6 \end{pmatrix}.$$

BA is not defined because the number of columns of *B* is 2 and the number of rows of *A* is 3 and they are not the same.

3. *AB* and *BA* are defined because their sizes are 2×2.

$$AB = \begin{pmatrix} 1 & 3 \\ -2 & 4 \end{pmatrix}\begin{pmatrix} 3 & -2 \\ 5 & 6 \end{pmatrix} = \begin{pmatrix} 18 & 16 \\ 14 & 28 \end{pmatrix}$$

and

$$BA = \begin{pmatrix} 3 & -2 \\ 5 & 6 \end{pmatrix}\begin{pmatrix} 1 & 3 \\ -2 & 4 \end{pmatrix} = \begin{pmatrix} 7 & 1 \\ -7 & 39 \end{pmatrix}.$$

Hence, $AB \neq BA$.

Following **(2.2.6)** , the following properties can be easily proved. The proofs are left to the reader.

Theorem 2.2.3.

The following assertions hold:

1. *(Associative law for matrix multiplication)*
   Let $A: = A_{m \times n},\ B: = B_{n \times p}$ *and* $C: = C_{p \times q}$*. Then*

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

2. *(Distributive laws for matrix multiplication)*
   i. *Let* $A: = A_{m \times n},\ B: = B_{n \times p}$ *and* $C: = C_{n \times q}$*. Then*

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

ii. *Let* $A := A_{m\times n}$, $B := B_{m\times n}$ *and* $C := C_{n\times q}$. *Then*

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

3. $(AB)^T = B^T A^T$.

Example 2.2.10.

Let

$$A = \begin{pmatrix} 1 & -3 \\ 0 & 2 \end{pmatrix},\ B = \begin{pmatrix} 2 & -1 & 4 \\ 3 & 1 & 5 \end{pmatrix},\quad \text{and } C = \begin{pmatrix} 0 & -2 & 1 \\ 4 & 3 & 2 \\ -5 & 0 & 6 \end{pmatrix}.$$

Show that $A(BC) = (AB)C$ and $(AB)^T = B^T A^T$.

Solution

$$BC = \begin{pmatrix} 2 & -1 & 4 \\ 3 & 1 & 5 \end{pmatrix}\begin{pmatrix} 0 & -2 & 1 \\ 4 & 3 & 2 \\ -5 & 0 & 6 \end{pmatrix} = \begin{pmatrix} -24 & -7 & 24 \\ -21 & -3 & 35 \end{pmatrix}$$

$$A(BC) = \begin{pmatrix} 1 & -3 \\ 0 & 2 \end{pmatrix}\begin{pmatrix} -24 & -7 & 24 \\ -21 & -3 & 35 \end{pmatrix} = \begin{pmatrix} 39 & 2 & -81 \\ -42 & -6 & 70 \end{pmatrix}$$

$$AB = \begin{pmatrix} 1 & -3 \\ 0 & 2 \end{pmatrix}\begin{pmatrix} 2 & -1 & 4 \\ 3 & 1 & 5 \end{pmatrix} = \begin{pmatrix} -7 & -4 & -11 \\ 6 & 2 & 10 \end{pmatrix}$$

$$(AB)C = \begin{pmatrix} -7 & -4 & -11 \\ 6 & 2 & 10 \end{pmatrix}\begin{pmatrix} 0 & -2 & 1 \\ 4 & 3 & 2 \\ -5 & 0 & 6 \end{pmatrix} = \begin{pmatrix} 39 & 2 & -81 \\ -42 & -6 & 70 \end{pmatrix}$$

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

$$B^TA^T = \begin{pmatrix} 2 & 3 \\ -1 & 1 \\ 4 & 5 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ -3 & 2 \end{pmatrix} = \begin{pmatrix} 2-9 & 0+6 \\ -1-3 & 0+2 \\ 4-15 & 0+10 \end{pmatrix} = \begin{pmatrix} -7 & 6 \\ -4 & 2 \\ -11 & 10 \end{pmatrix}.$$

$$(AB)^T = \begin{pmatrix} -7 & -4 & -11 \\ 6 & 2 & 10 \end{pmatrix}^T = \begin{pmatrix} -7 & 6 \\ -4 & 2 \\ -11 & 10 \end{pmatrix}.$$

Hence, $(AB)^T = B^TA^T$.

Exercises

1. Let $A = \begin{pmatrix} 2 & 1 & 1 \\ 1 & -1 & 2 \end{pmatrix}$, $\vec{X} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$, $\overrightarrow{X_1} = \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix}$, $\overrightarrow{X_2} = \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix}$. Compute $A\vec{X}$, $A\overrightarrow{X_1}$, and $A\overrightarrow{X_2}$.

2. 2. $A = \begin{pmatrix} -2 & -1 \\ 3 & 1 \\ 0 & 1 \end{pmatrix}$, $\vec{X} = \begin{pmatrix} -a \\ 2a \end{pmatrix}$, $\overrightarrow{X_1} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$. Compute $A\vec{X}$ and $A\overrightarrow{X_1}$.

3. Let $A = \begin{pmatrix} 0 & 1 \\ 1 & -1 \\ 2 & 1 \end{pmatrix}$, $\overrightarrow{X_1} = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}$, $\overrightarrow{X_2} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, $\overrightarrow{X_3} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 2 \end{pmatrix}$.

   Determine whether $A\overrightarrow{X_i}$. is defined for each $i = 1, 2, 3$.

4. Let $A = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & -1 \\ 1 & 1 & 2 \end{pmatrix}$, $\vec{X} = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}$, $\vec{Y} = \begin{pmatrix} 3 \\ 0 \\ -1 \end{pmatrix}$. Compute $A(\vec{X} - 3\vec{Y})$.

5. Let $A = \begin{pmatrix} 1 & 1 & -1 \\ 1 & 0 & 2 \\ 2 & 1 & -1 \end{pmatrix}$ and $\vec{X} = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}$. Write $A\vec{X}$ as a linear combination of the column vectors of *A*.

6. We are interested in predicting the population changes among three cities in a country. It is known that in the current year, 20% and 30% of the populations of City 1 will move to Cities 2 and 3, respectively, 10% and 20% of the populations of City 2 will move to Cities 1 and 3, respectively, and 25% and 10%

of the populations of City 3 will move to Cities 1 and 2, respectively. Assume that 200,000, 600,000, and 500,000 people live in Cities, 1, 2, and 3, respectively, in the currently year. Find the populations in the three cities in the following year.

7. Let $A = \begin{pmatrix} 2 & 1 \\ -2 & 3 \end{pmatrix}$ and $B = \begin{pmatrix} 1 & -2 & 1 \\ 3 & 4 & 1 \end{pmatrix}$. Find *AB* and the sizes of *A*, *B*, and *AB*.

8. Let $A = \begin{pmatrix} 0 & 1 & -1 \\ 2 & 3 & 1 \end{pmatrix}$ and $B = \begin{pmatrix} 2 & 1 & -1 \\ 1 & 0 & 1 \\ -1 & 2 & 4 \end{pmatrix}$. Compute *AB* and find the sizes of *A*, *B*, and *AB*.

9. Let $A = \begin{pmatrix} 1 & 0 & -2 \\ 3 & 2 & -1 \end{pmatrix}$ and $B = \begin{pmatrix} 0 & -1 & 2 & 1 \\ -1 & 2 & 1 & -2 \\ 2 & 0 & 0 & 1 \end{pmatrix}$. Compute *AB* and find the sizes of *A*, *B*, and *AB*.

10. 10. Let $A = \begin{pmatrix} 1 & 1 \\ 1 & 3 \\ 3 & 1 \end{pmatrix}$ and $B = \begin{pmatrix} 2 & 1 \\ -2 & 3 \end{pmatrix}$. Are *AB* and *BA* defined?

    If so, compute them. If not, explain why.

11. Let $A = \begin{pmatrix} 2 & -1 \\ -2 & 3 \end{pmatrix}$ and $B = \begin{pmatrix} 0 & -2 \\ 4 & 2 \end{pmatrix}$. Are *AB* and *BA* defined?

    If so, compute them. If not, explain why. Is *AB* equal to *BA*?

12. Let

$$A = \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix}, \quad B = \begin{pmatrix} 0 & -1 & 2 \\ 1 & 1 & 3 \end{pmatrix}, \quad \text{and } C = \begin{pmatrix} 1 & -1 & 1 \\ 0 & 3 & 2 \\ -2 & 0 & 3 \end{pmatrix}.$$

Show that $A(BC) = (AB)C$ and $(AB)^T = B^T A^T$.

13. Two students buy three items in a bookstore. Students 1 and 2 buy 3,2,4 and 6,2,3, respectively. The unit prices and unit taxes are 5, 4, 6 (dollars) and 0.08, 0.07, 0.05, respectively. Use a matrix to show the total prices and taxes paid by each student.
14. Assume that three individuals have contracted a contagious disease and have had direct contact with four people in a second group. The direct contacts can be expressed by a matrix

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

Now, assume that the second group has had a variety of direct contacts with five people in a third group. The direct contacts between groups 2 and 3 are expressed by a matrix

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

 i. Find the total number of indirect contacts between groups 1 and 3.
 ii. How many indirect contacts are between the second individual in group 1 and the fourth individual in group 3?

15. 15. There are three product items, *P*1, *P*2, and *P*3, for sale from a large company. In the first day, 5, 10, and 100 are sought out. The corresponding unit profits are 500, 400, and 20 (in hundreds of dollars) and the corresponding unit taxes are 3, 2, and 1. Find the total profits and taxes in the first day.

2.3 Square matrices

An $n \times n$ matrix is called a square matrix of order n, that is,

(2.3.1)

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}.$$

The entries $a_{11},\ a_{22},\ \cdots,\ a_{nn}$ are said to be on the main diagonal of A. The sum of these entries, denoted by tr(A), is called the trace of A, that is,

(2.3.2)

$$\mathrm{tr}(A) = \sum_{l=1}^{n} a_{ii} = a_{11} + a_{22} + \cdots + a_{nn}.$$

Example 2.3.1.

Let $A = \begin{pmatrix} 15 & 21 \\ 44 & 25 \end{pmatrix}$. Then A is a square matrix of order 2, the numbers 15, 25 are on the main diagonal of A, and tr(A) = 15 + 25 = 40.

Symmetric matrices

A square matrix A is said to be symmetric if

(2.3.3)

$$A^T = A.$$

By **(2.3.3)** we see that A is symmetric if and only if

$$a_{ij} = a_{ji} \text{ for } i, j \in I_n.$$

We can write a symmetric matrix in an explicit form

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{12} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{nn} \end{pmatrix}.$$

From the above matrix, we see that A is symmetric if the ith row and the ith column are the same for each $i \in I_n$.

Example 2.3.2.

1. The following matrices are symmetric.

$$\begin{pmatrix} 7 & -3 \\ -3 & 0 \end{pmatrix} \begin{pmatrix} 1 & 4 & 5 \\ 4 & -3 & 0 \\ 5 & 0 & 7 \end{pmatrix} \begin{pmatrix} 1 & x^2 & 2 \\ x^2 & 0 & x \\ 2 & x & 3 \end{pmatrix}$$

2. The following matrices are not symmetric.

$$\begin{pmatrix} 7 & -3 \\ 2 & 0 \end{pmatrix} \quad \begin{pmatrix} 1 & 4 & 5 \\ 2 & -3 & 0 \\ 5 & 0 & 7 \end{pmatrix} \quad \begin{pmatrix} 2 & x^2+1 & 2 \\ 0 & 1 & 2 \\ 2 & 2 & 2 \end{pmatrix}$$

Definition 2.3.1.

A square matrix is said to be

i. **lower triangular** if all the entries above the main diagonal are zero;
ii. **upper triangular** if all the entries below the main diagonal are zero;
iii. **triangular** if it is lower or upper triangular;
iv. **diagonal** if it is lower and upper triangular;
v. **identity matrix** if it is a diagonal matrix whose entries on the main diagonal are 1.

We denote by *I* or I_n an $n \times n$ identity matrix. Hence,

(2.3.4)

$$I_n = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}.$$

Example 2.3.3.

1. The following matrices are lower triangular.

$$\begin{pmatrix} 0 & 0 \\ 2 & 0 \end{pmatrix} \quad \begin{pmatrix} 1 & 0 & 0 \\ 2 & 0 & 0 \\ 0 & 1 & 1 \end{pmatrix} \quad \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ x & 0 & 0 \end{pmatrix}$$

2. The following matrices are upper triangular.

$$\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

3. The following matrices are diagonal.

$$\begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

4. The following matrices are not triangular matrices.

$$\begin{pmatrix} 1 & 1 & 0 \\ 0 & 2 & 0 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 1 \\ 2 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

5. The following are identity matrices.

$$I_1 = (1),\ I_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},\ I_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix},\ I_4 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

The following theorem gives some properties of triangular matrices. The proofs are left to the reader.

Theorem 2.3.1.

i. *If A is an upper triangular (lower triangular) matrix, then AT is a lower triangular (upper triangular) matrix.*
ii. *If A and B are lower triangular matrices, then AB is a lower triangular matrix.*
iii. *If A and B are upper triangular matrices, then AB is an upper triangular matrix.*

Example 2.3.4.

Let $A = \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix}$ and $B = \begin{pmatrix} 3 & 0 \\ 4 & 1 \end{pmatrix}$. Compute *AB*.

Solution

$$AB = \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix}\begin{pmatrix} 3 & 0 \\ 4 & 1 \end{pmatrix} = \begin{pmatrix} 3 & 0 \\ 10 & 1 \end{pmatrix}.$$

The above example shows that the product of two lower triangular matrices is a lower triangular matrix.

For a square matrix, we can define its power as follows.

(2.3.5)

$$A^0 = I_n,\ A^2 = A \cdot A,\ A^3 = A^2 \cdot\ A,\ \cdots, A^n = A^{n-1} \cdot\ A.$$

Example 2.3.5.

Let $A = \begin{pmatrix} 1 & 2 \\ 1 & 3 \end{pmatrix}$. Find $A^0,\ A^2,$ and A^3.

Solution

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

$$A^2 = \begin{pmatrix} 1 & 2 \\ 1 & 3 \end{pmatrix}\begin{pmatrix} 1 & 2 \\ 1 & 3 \end{pmatrix} = \begin{pmatrix} 3 & 8 \\ 4 & 11 \end{pmatrix},$$

$$A^3 = A^2A = \begin{pmatrix} 3 & 8 \\ 4 & 11 \end{pmatrix}\begin{pmatrix} 1 & 2 \\ 1 & 3 \end{pmatrix} = \begin{pmatrix} 11 & 30 \\ 15 & 41 \end{pmatrix}.$$

We can write a diagonal matrix as

(2.3.6)

$$\operatorname{diag}(a_1, a_2, \cdots, a_n) = \begin{pmatrix} a_1 & 0 & \cdots & 0 \\ 0 & a_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_n \end{pmatrix}.$$

It is easy to verify that for $k \in \mathbb{N}$,

$$\operatorname{diag}(a_1, a_2, \cdots, a_n)^k = \operatorname{diag}(a_1^k, a_2^k, \cdots, a_n^k),$$

that is,

(2.3.7)

$$\begin{pmatrix} a_1 & 0 & \cdots & 0 \\ 0 & a_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_n \end{pmatrix}^k = \begin{pmatrix} a_1^k & 0 & \cdots & 0 \\ 0 & a_2^k & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_n^k \end{pmatrix}.$$

Example 2.3.6.

Find A^5 if

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

Solution

Because A is a diagonal matrix, it follows from **(2.3.7)** that

$$A^5 = \begin{pmatrix} 1^5 & 0 & 0 \\ 0 & (-2)^5 & 0 \\ 0 & 0 & 2^5 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -32 & 0 \\ 0 & 0 & 32 \end{pmatrix}.$$

It is easy to verify that for $r, s \in \mathbb{N}$,

(2.3.8)

$$A^r \cdot A^s = A^{r+s} \text{ and } (A^r)^s = A^{rs}.$$

Using the power of a square matrix, we introduce the notion of a matrix polynomial, which is a polynomial with square matrices as variables. Let

$$P(x) = a_0 + a_1 x + \cdots + a_m x^m$$

be a polynomial of x. Then the polynomial evaluated at a matrix A is

(2.3.9)

$$p(A) = a_0 I_n + a_1 A + a_2 A^2 + \cdots + a_m A^m,$$

which is called a matrix polynomial of degree *m*.

Example 2.3.7.

Let $P(x) = 4 - 3x + 2x^2$ and $A = \begin{pmatrix} -1 & 2 \\ 0 & 3 \end{pmatrix}$. Compute the matrix polynomial *P*(*A*).

Solution

$$\begin{aligned} P(A) &= 4I_2 - 3A + 2A^2 = 4\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - 3\begin{pmatrix} -1 & 2 \\ 0 & 3 \end{pmatrix} + 2\begin{pmatrix} -1 & 2 \\ 0 & 3 \end{pmatrix}^2 \\ &= \begin{pmatrix} 4 & 0 \\ 0 & 4 \end{pmatrix} - \begin{pmatrix} -3 & 6 \\ 0 & 9 \end{pmatrix} + \begin{pmatrix} 2 & 8 \\ 0 & 18 \end{pmatrix} = \begin{pmatrix} 9 & 2 \\ 0 & 13 \end{pmatrix}. \end{aligned}$$

Exercises

1. For each of the following matrices, find its trace and determine whether it is symmetric.

$$A_1 = \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix}, A_2 = \begin{pmatrix} 1 & 4 & 5 \\ 4 & -3 & -1 \\ 5 & 0 & 7 \end{pmatrix}, A_3 = \begin{pmatrix} y & x^3 & 1 \\ x^3 & y & x \\ 1 & x & z \end{pmatrix}.$$

2. Identify which of the following matrices are upper triangular, diagonal, triangular, or an identity matrix.

$$A_1 = \begin{pmatrix} 0 & 2 \\ 0 & 0 \end{pmatrix} A_2 = \begin{pmatrix} 1 & 0 & 0 \\ 2 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} A_3 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

$$A_4 = \begin{pmatrix} 1 & 0 & x \\ 1 & 0 & 0 \\ x & 0 & 0 \end{pmatrix} A_5 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} A_6 = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$A_7 = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} A_8 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 2 & 0 & 1 \end{pmatrix} A_9 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ -1 & 1 & 1 \end{pmatrix}$$

$$A_{10} = \begin{pmatrix} 1 & 0 & 1 \\ 2 & 0 & 0 \\ 1 & 0 & 1 \end{pmatrix} A_{11} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix} A_{12} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}$$

3. Let $A = \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix}$. Find $A^0, A^2, A^3,$ and A^4.

4. Let $A = \begin{pmatrix} 2 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 3 \end{pmatrix}$. Find A^6.

5. Let $P(x) = 1 - 2x - x^2$ and $A = \begin{pmatrix} -1 & 1 \\ 1 & 2 \end{pmatrix}$. Compute $P(A)$.

6. Let $P(x) = x^2 - x - 4$. Compute $P(A)$ and $P(B)$, where

$$A = \begin{pmatrix} 3 & 1 & 0 \\ 3 & -2 & 3 \\ 4 & 5 & -1 \end{pmatrix} \text{ and } B = \begin{pmatrix} 6 & -2 & -1 \\ 3 & 0 & 2 \\ -1 & 3 & 4 \end{pmatrix}.$$

7. Assume that the population of a species has a life span of less than 3 years. The females are divided into three age groups: group 1 contains those of ages less than 1, group 2 contains those of age between 1 to 2, and group 3 contains those of age 2. Suppose that the females in group 1 do not give birth and the average numbers of newborn females produced by one female in group 2 and 3, respectively, are 3 and 2. Suppose that 50% of the females in group 1 survive to age 1 and 80% of the females in group 2 live to age 2. Assume that the numbers of females in groups 1, 2, and 3 are X_1, X_2, and X_3, respectively.
    i. Find the number of females in the three groups in the following year.
    ii. Find the number of females in the three groups in *n* years.
    iii. If the current population distribution $(x_1, x_2, x_3) = (1000{,}500{,}100)$, then use result 2 to predict the population distribution in 2 years.

2.4 Row echelon and reduced row echelon matrices

Arowinan $m \times n$ matrix is said to be a zero row if all the entries in that row are zero. If one of the entries in a row is nonzero, then it is called a nonzero row.

Example 2.4.1.

In the matrix $\begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$, the first two rows are nonzero rows and the last row is a zero row.

The first nonzero entry (starting from the left) in a nonzero row is called the **leading entry** of the row. If a leading entry is 1, it is called a **leading one**.

Example 2.4.2.

Find the leading entries of *A*, where

$$A = \begin{pmatrix} 1 & 6 & 0 & 4 \\ 0 & 2 & 0 & 5 \\ 3 & 0 & 0 & 10 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Solution

The numbers 1, 2, and 3 are the leading entries of *A*.

Definition 2.4.1.

An $m \times n$ matrix *A* is said to be a row echelon matrix if it satisfies the following conditions (1) and (2); or to be a reduced row echelon matrix if it satisfies the following conditions (1), (2), (3), and (4).

1. Each nonzero row (if any) lies above every zero row, that is, all zero rows appear at the bottom of the matrix.
2. For any two successive nonzero rows, the leading entry in the lower row occurs further to the right than the leading entry in the higher row.
3. Each leading entry is 1.
4. For each column containing a leading 1, all entries above the leading 1 are zero.

Remark 2.4.1.

i. For a row echelon matrix, if a column contains a leading entry, then by condition (2), all entries below the leading entry of that column must be zero.
ii. A reduced row echelon matrix must be a row echelon matrix.
iii. For a reduced row echelon matrix, if a column contains a leading 1, then by conditions (2) and (4), all entries except the leading 1 of that column must be zero.

Example 2.4.3.

1. The following matrices are row echelon matrices.

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 2 & 0 & 3 & 5 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 2 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 & 6 & 0 \\ 0 & 0 & 2 & 3 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

2. The following matrices are not row echelon matrices.

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

3. The following matrices are reduced row echelon matrices.

$$\begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 & 6 & 0 \\ 0 & 0 & 1 & 3 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 & 0 & 0 & -1 \\ 0 & 0 & 0 & 1 & 2 \end{pmatrix}.$$

4. The following matrices are row echelon matrices, but not reduced row echelon matrices:

$$\begin{pmatrix} 2 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 2 & 6 & 0 \\ 0 & 0 & 1 & 3 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 & 2 & 0 \\ 0 & 0 & 2 & 1 & 0 \end{pmatrix}.$$

Solution

We only consider (4). These matrices are row echelon matrices. However, the first matrix contains a leading entry 2 that is not a leading 1; column 3 of the second matrix has the leading 1 of the second row but also contains a nonzero entry 2; and the third matrix has a leading entry that is not 1. Hence, they are not reduced row echelon matrices.

Row operations. In many cases, we need to change a matrix to a row echelon matrix or a reduced row echelon matrix and wish that the new matrix and the original matrix share some common properties. Thus, one can obtain properties of the original matrix by studying the row echelon matrix or the reduced row echelon matrix. The powerful tools used to change a matrix to a row echelon matrix or reduced row echelon matrix are the following (elementary) row operations.

1. **First row operation** $R_i(c)$: Multiply the*i*th row of a matrix by a nonzero number *c*;
2. **Second row operation** $R_i(c) + R_j$: Add the *i*th row multiplied by a nonzero number *c* to the *j*th row;
3. **Third row operation** $R_{i,j}$: Interchange the *i*th row and the *j*th row.

Remark 2.4.2.

The first row operation $R_i(c)$ is used to change only the *i*th row, so other rows remain unchanged. The second row operation $R_i(c) + R_j$ is used to change only the *j*th row, so other rows including the *i*th row remain unchanged. The third row operation $R_{i,j}$ is used to interchange only the *i*th row and the *j*th row, so other rows remain unchanged.

Now, we give some examples to show how to use the above row operations to change one matrix to another. These row operations in these examples are used to show how to perform the row operations and are chosen for practice. There are no reasons why we choose these row operations here. You can choose other row operations for practice as well. Later, when we want to change a matrix to a row echelon matrix or reduced row echelon matrix, we must choose suitable row operations to obtain its row echelon matrices.

Example 2.4.4.

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

i. We apply $R_2(-2)$ to change row 2 of *A*, so rows 1 and 3 remain unchanged.

$$A = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 2 & -3 \\ 2 & 1 & 1 \end{pmatrix} R_2(-2)\to \begin{pmatrix} 0 & 1 & 1 \\ -2 & -4 & 6 \\ 2 & 1 & 1 \end{pmatrix}.$$

ii. We apply $R_2(-2) + R_3$ to change row 3 of *A*, so rows 1 and 2 remain unchanged.

$$A = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 2 & -3 \\ 2 & 1 & 1 \end{pmatrix} R_2(-2) + R_3\to \begin{pmatrix} 0 & 1 & 1 \\ 1 & 2 & -3 \\ 0 & -3 & 7 \end{pmatrix}.$$

iii. We apply $R_{1,2}$ to change rows 1 and 2 of *A*, so row 3 remains unchanged.

$$A = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 2 & -3 \\ 2 & 1 & 1 \end{pmatrix} R_{1,2}\to \begin{pmatrix} 1 & 2 & -3 \\ 0 & 1 & 1 \\ 2 & 1 & 1 \end{pmatrix}.$$

Example 2.4.5.

Let

$$A = \begin{pmatrix} 1 & 1 & 1 \\ 2 & -1 & -1 \\ -3 & 4 & 1 \end{pmatrix}.$$

Use the second row operation to change the numbers 2 and -3 in *A* to 0.

Solution

To change the number 2 in the first column and second row of *A*, we use the row operation $R_1(-2)+R_2$ to change the entire second row.

$$A=\begin{pmatrix}1 & 1 & 1\\ 2 & -1 & -1\\ -3 & 4 & 1\end{pmatrix} \xrightarrow{R_1(-2)+R_2} \begin{pmatrix}1 & 1 & 1\\ 0 & -3 & -3\\ -3 & 4 & 1\end{pmatrix}.$$

Similarly, we use $R_1(3)+R_3$ to change the entire third row in order to change the number -3 to 0.

$$\begin{pmatrix}1 & 1 & 1\\ 0 & -3 & -3\\ -3 & 4 & 1\end{pmatrix} \xrightarrow{R_1(3)+R_3} \begin{pmatrix}1 & 1 & 1\\ 0 & -3 & -3\\ 0 & 7 & 4\end{pmatrix}.$$

From the above, we see that row 1 is always kept unchanged when we changed rows 2 and 3. Hence, we can combine the process as follows:

A=(1112−1−1−341)→R1(3)+R3R1(−2)+R2(1110−3−3074).

Example 2.4.6.

Let B=(1110−3−3074). Use the second row operation to change the number 7 in *B* to 0.

Solution

B=(1110−3−3074)R2(73)+R3→(1110−3−300−3).

Combining **Examples 2.4.5** and **2.4.6**, we get

A=(1112−1−1−341)→R1(3)+R3R1(−2)+R2(1110−3−3074)R2(73)+R3→(1110−3−300−3):=C.

Note that the last matrix *C* is a row echelon matrix of *A* and is obtained by performing second row operations three times.

Methods of finding row echelon matrices

After we intuitively get some ideas on how to change a matrix to its echelon matrix from **Examples 2.4.5** and **2.4.6**, we learn how to apply row operations to find a row echelon matrix for a given matrix. To find a row echelon matrix, we need to do several rounds of the same steps, depending on the number of leading entries.

The first round is to find the first leading entry. Here are some key steps to find the first leading entry.

Step 1. Move all zero rows, if any, to the bottom of the matrix. Look at each row to see whether there is a common factor in the row. If there is a common factor in a row, then we would use the first row operation to eliminate the common factor. After eliminating all the common factors in the rows by using the first row operation, the resulting matrix has entries that are smaller than or equal to the corresponding entries of the given matrix. Smaller entries make later computations easier.

Step 2. After eliminating all the common factors, we would start to seek the first leading entry in the first nonzero column of the resulting matrix.

Case I. The first nonzero column of the resulting matrix contains number 1or −1.

i. If the number 1 or −1 is on the first row, then the number 1 or −1 can be used as the first leading entry.
ii. If the number 1 or −1 is not on the first row, then we use the third row operation to interchange one of the rows that contains the number 1 or −1 with row one. So the number 1 or −1 is moved to the first row and can be used as the first leading entry.

Case II. The first nonzero column of the resulting matrix does not contain number 1 or −1.

a. We choose the smallest number of the first nonzero column of the resulting matrix and use the methods in (*i*)and (*ii*) to move the row containing the smallest number to the first row. The smallest number in the first row, in general, can be used as the first leading entry.
b. In some cases, we can continue to change the smallest number to 1 or −1 by using the second row operation. Some examples will be provided later to exhibit the smart method, for example see **Example 2.4.9** .

Step 3. Note that the row one contains the first leading entry 1 or −1 or the smallest number, and there is a column that contains the first leading entry as well. If all the entries below the first leading entry in the column are zero, then the first round is completed. If there are nonzero entries below the first leading entry in the column, then we would use the second row operation and the row one to change each of these nonzero entries to 0. Note that all other entries in the row that contain such a nonzero number must be changed following the second row operation. After all these nonzero entries are changed to 0, the first round is completed.

The second round is to find the second leading entry from the last matrix obtained above.

Step 4. Repeat the above **Steps 1-3** to find the second leading entry below row one of the last matrix obtained in the above Step 3, and then continue the process to find the third, fourth, etc., leading entry, if any, until you get a row echelon matrix.

Example 2.4.7.

Find a row echelon matrix of the matrix

$$A=\begin{pmatrix}0 & 6 & -3\\ 2 & -2 & 6\\ 2 & 0 & 1\end{pmatrix}.$$

Solution

(1) **Step 1.** There are no zero rows in *A*. Look at each row to see whether there is a common factor in the row. Row 1 contains a common factor 3 and row 2 contains a common factor 2. So, we first use the first row operation to eliminate the two common factor.

$$A=\begin{pmatrix}0 & 6 & -3\\ 2 & -2 & 6\\ 2 & 0 & 1\end{pmatrix}\xrightarrow{R_2(\frac{1}{2})\,R_1(\frac{1}{3})}\begin{pmatrix}0 & 2 & -1\\ 1 & -1 & 3\\ 2 & 0 & 1\end{pmatrix}:=B.$$

Step 2. Check the first nonzero column of *B* to see if there is number 1 or −1. The column one of *B* is the first nonzero column, which contains the

number 1 as well. The row that contains the number 1 is not row one, so we need to move it to row one by using the third row operation. Therefore, we interchange rows 1 and 2 of *B* and the number 1 is the first leading entry. We use the symbol a to denote a leading entry *a*.

$$B=\begin{pmatrix}0 & 2 & -1\\ 1 & -1 & 3\\ 2 & 0 & 1\end{pmatrix}\xrightarrow{R_{1,2}}\begin{pmatrix}1 & -1 & 3\\ 0 & 2 & -1\\ 2 & 0 & 1\end{pmatrix}:=C.$$

Step 3. If there are nonzero entries below the first leading entry in the column of *C*, then use the second row operation and row one to change each of these nonzero entries to 0.

We use the symbol ↓ to show the direction of changing all nonzero entries below a leading entry to zero.

C=(1−1302−1201)↓R1(−2)+R3→(1−1302−102−5):=D.

Step 4. Repeat the above steps 1-3 to find the second leading entry below row one of the last matrix *D*. In order to find the second leading entry of *D*, we consider the following matrix below row one of *D*:

D1:=(02−102−5)

We repeat **Steps 1-3**.

Step 1. Look at each row of the matrix D1 to see whether there is a common factor in the row. The answer is that there are no common factors in each row of D1

Step 2. Check the first nonzero column of D1 to see if there are nonzero numbers. The first nonzero column is the second column of D1 and the first and second rows contains a nonzero number 2. So, the number 2 in the first row of D1 is the second leading entry for *D*.

D=(1−1302−102−5).

Step 3. If there are nonzero entries below the second leading entry 2 in the column, then use the second row operation and the row 2 of *D* to make the nonzero entries in the column zero. In *D*, there is a nonzero number 2

in the third row below the second leading entry 2. We change the nonzero number 2 to 0 by using row 2 that contains the second leading entry 2.

D=(1−1302−102−5)↓R2(−1)+R3→(1−1302−100−4).

The last matrix is a row echelon matrix and the number −4 is the third leading entry.

We rewrite the process as follows:

A=(06−32−26201)→R2(12)R1(13)(02−11−13201)R1,2→(1−1302−1201)↓R1(−2)+R3→(1−1302−102−5)↓R2(−1)+R3→(1−1302−100−4).

In the following, we always follow **Steps 1-3** in each round without further statement.

Example 2.4.8.

Find a row echelon matrix of the matrix

A=(1302011100241−1000031).

Solution

A=

(1302011100241−1000031)↓→R1(−2)+R3R1(−1)+R2(130200−21−200−21−5000031)↓R2(−1)+R3→(130200−21−20000−3000031)R3(13)→(130200−21−20000−1000031)↓R3(3)+R4

The last matrix is a row echelon matrix of *A*.

Remark 2.4.3.

Note that the row echelon matrix of a matrix is not unique.

For example, in **Example 2.4.8** , we can continue to use row operations to change the above row echelon matrix to another row echelon matrix such as

(130200−21−20000−1000001)R3(−1)→(130200−21−200001000001).

The last matrix is another row echelon matrix of *A*.

The following example provides a **smart method**, which is mentioned in **Step 2** of **the first round**.

Example 2.4.9.

Find a row echelon matrix of the matrix

A=(2−102030−1102−41−1051030)

Analysis: The first number 2 in the first row of *A* would be the first leading entry. But if we use it as the first leading entry to change all the nonzero numbers 3*,* 2*,* 5 in column one to 0 by using the second row operations and row one, then the problem we encounter is that there are computations of fractions in rows 2 and 4. To avoid the computation of fractions, we can try to use suitable second row operations to get 1 or −1 in the first nonzero column. For the above matrix *A*, we have several choices to get 1 or −1. For example, we can use one of the following:

i. R2(−1)+R1
ii. R1(−1)+R2 and then R1, 2.
iii. R1(−2)+R4 and then R1, 4.

In the following, we use (*i*)toget −1 and take the number −1 as the first leading entry. We also use the smart method to choose the second and third leading entries.

Solution

A=(2−102030−111201−10510−3−1)R2(−1)+R1→(−1−111−130−111201−10510−3−1)↓

→R1(5)+R4R1(3)+R2R1(2)+R3(−1−111−10−324−20−231−20−452−6)R3(−2)+R2→

(−1−111−101−4220−231−20−452−6)↓→R2(2)+R3R2(4)+R4(−1−111−101−42200−55200−11102)R2(−2)+R4→(−1−111−101−42200−55200−10−2)R3, 4→(−1−111−101−42200−10−2

R3(−5)+R4→(−1−111−101−42200−10−2000512).

The last matrix is a row echelon matrix of *A*.

Methods of finding reduced row echelon matrices

The basic idea is to use row operations to change a given matrix *A* to a row echelon matrix *B*, where we change *A* from its top to bottom, and then continue using row operations to change all the nonzero numbers above the leading entries in the columns that contain the leading entries to 0, where we change *B* from its bottom to top. We denote by

l1, l2, ⋯, lr,

the leading entries of the row echelon matrix *B* in order. To find the reduced row echelon matrix of *A*, we first use the second row operation and the row that contains the last leading entry l_r to change all the nonzero numbers above the leading entry l_r from the lower one to the top one to 0. Then we do the same thing to $l_{r-1}, \cdots, l_2$ in order.

We have studied how to change a matrix to a row echelon matrix, so we first give examples demonstrating how to change a row echelon matrix to a reduced row echelon matrix and then provide examples to show how to change a matrix to a reduced row echelon matrix.

Example 2.4.10.

Find the reduced row echelon matrix of the following row echelon matrix

B=(1−1302−100−4).

Solution

(1) Note that the leading entries in order are $l_1=1$, $l_2=2$, and $l_3=-4$. To find the reduced row echelon matrix for the row echelon matrix, the first step is to change all the nonzero numbers above $l_3=-4$ to be zero. The nonzero numbers above $l_3=-4$ are -1 and 3, which are in the second row and the first row of *B*, respectively. We use the second row operations and the row 3 that contains the leading entry $l_3=-4$ to change -1 and 3 to 0 in order. We use the symbol ↑ to show the direction of changing all nonzero entries above a leading entry to zero.

B=(1−1302−100−4)R3(−14)→(1−1302−1001)↑→R3(−3)+R1R3(1)+R2(1−10020001).

(2) The second step is to change the nonzero number above $l_2=2$ to 0 by using row 2 that contains $l_2=2$.

(1−10020001)R2(12)→(1−10010001)↑R2(1)+R1→(100010001).

The last matrix is the reduced row echelon matrix of *B*.

Example 2.4.11.

Find the reduced row echelon matrix of

A=(1302011100241−1000031).

Solution

The row echelon matrix *B* of *A* is given in **Example 2.4.8**. We change the echelon matrix *B* to the reduced echelon matrix.

B=
(130200−21−20000−1000001)R3(−1)→(130200−21−200001000001)↑→R3(−2)+R1R3(2)+R2(130000−21000001000001)R2(−12)→(1300001−12000001000001)↑R2(−3)+R1→(1032

The last matrix is the reduced row echelon matrix of *A*.

Example 2.4.12.

Find the reduced row echelon matrix of

A=(210130314213).

Solution

A=

(210130314213)R2(−1)+R1→(−11−3030314213)↓→R1(4)+R3R1(3)+R2(−11−3003−6106−113)↓R2(−2)+R3→(−11−3003−610011)↑→R3(3)+R1R3(6)+R2(−110303070011)R2(13)→(

The last matrix is the reduced row echelon matrix of *A*.

In **Example 2.4.12**, we employ the smart method used in **Example 2.4.9**. We use the row operation R2(−1)+R1 to change row one of *A* in order to change the leading entry in row one to −1.

By **Remark 2.4.3**, we know that the row echelon matrices of a matrix are not unique. But its row echelon matrix is unique. We give this result below without proof.

Theorem 2.4.1.

Any m×n *matrix can be changed using row operations to a unique reduced row echelon matrix.*

Exercises

1. Which of the following matrices are row echelon matrices?
   A=(1000)B=(9435900074)C=(250090008)D=(30160008860000100000)E=(100001010)F=(141000066)G=(081729800000)H=(000040580961)

2. Which of the following matrices are reduced row echelon matrices?
   A=(1001010100100000)B=(15060001300000100000)C=(1111100011)D=(201010001)E=(10040011300000100000)F=(1002000210)

3. Let A=(10121−1−32−2). Use the second row operation to change the numbers 2 and −3 in the first column of *A* to zero.
4. Let A=(10101−3031). Use the second row operation to change the number 3 in *A* to 0.
5. For each of the following matrices, find its row echelon matrix. Clearly mark every leading entry *a* by using the symbol a and use the symbol ↓ to show the direction of eliminating nonzero entries below leading entries.
   A=(10121−1−32−2)B=(01−21−11−210)C=(210−130325−2−13)

6. For each of the following matrices, find its reduced row echelon matrix. Clearly mark every leading entry *a* by using the symbol a and use the symbols ↓ and ↑ to show the directions of eliminating nonzero entries below or above leading entries.
   A=(2−4201−300−6)B=(11010036−3−6000−2200002)C=(06−32−26201)D=(210130325−2−13)E=(100−11−2112−212)F=(1121111−22211)G=(10111−312−2−5−2−148031−3−70)

2.5 Ranks, nullities, and pivot columns

Let *A* be an $m \times n$ matrix. From **Remark 2.4.3**, we see that the row echelon matrices of *A* are not unique, but the total number of leading entries for each row echelon matrix of *A* is the same. Hence, no matter which row operations are used to change *A* to a row echelon matrix, the total number of the leading entries of each row echelon matrix of *A* is unchanged. This leads us to define the total number of the leading entries as the rank of *A*.

Definition 2.5.1.

Let *A* be an $m \times n$ matrix and let *B* be one of its row echelon matrices.

1. The total number of leading entries of the row echelon matrix *B* is called the rank of *A*, denoted by *r*(*A*).
2. The number $n - r(A)$ is called the nullity of *A*, denoted by null(*A*), that is,
   (2.5.1)
   $$\text{null}(A) = n - r(A).$$
3. The columns of *A* corresponding to the columns of *B* containing the leading entries are called the pivot columns (or pivot column vectors) of *A*.
   According to **Definition 2.5.1**, the rank of *A* is equal to the rank of each of its row echelon matrices. The pivot columns of *A* will be used in **Section 7.3** to find the basis of column space of *A* (see **Theorem 7.3.1**).
   By **Definitions 2.4.1** and **2.5.1**, we obtain the following result.

Theorem 2.5.1.

Let A be an $m \times n$ *matrix and let B be one of its row echelon matrices. Then the following numbers are equal.*

1. *The rank of A.*
2. *The number of leading entries of B.*
3. *The number of rows of B containing the leading entries.*
4. *The number of columns of B containing the leading entries.*

Remark 2.5.1.

By **Theorem 2.5.1** (3), we see that the number of zero rows of *B* equals $m - r(A)$, and if *A* contains *q* zero rows, then $r(A) \leq m - q$. By **Theorem 2.5.1** (4), if *A* contains *k* zero columns, then $r(A) \leq n - k$. .

The following result gives the relation between the rank of a matrix and its size.

Theorem 2.5.2.

Let A be an $m \times n$ *matrix. Then* $r(A) \leq \min\{m, n\}$, *where min{m, n} represents the minimum of m and n.*

Proof

By **Theorem 2.5.1** (3), the number of rows of *B* containing the leading entries is smaller than or equal to *m*. By **Theorem 2.5.1** (1)-(3), $r(A) \leq m$. Similarly, by **Theorem 2.5.1** (4), the number of columns of *B* containing the leading entries is smaller than or equal to *n*, and thus $r(A) \leq n$. This implies that *r*(*A*) is smaller than or equal to the minimum of *m* and *n*.

Example 2.5.1.

Let

$$A = \begin{pmatrix} 1 & -1 & 3 & 2 \\ 0 & 1 & 2 & 1 \\ 2 & -2 & 6 & 4 \end{pmatrix} B = \begin{pmatrix} 1 & 2 & -1 \\ -2 & -4 & 2 \end{pmatrix} C = \begin{pmatrix} 0 & 4 & 2 \\ 0 & 2 & 1 \\ 1 & 1 & 1 \\ 2 & 0 & 1 \end{pmatrix}$$

Show $r(A) \leq 3,\ r(B) \leq 2,$ and $r(C) \leq 3.$

Solution

Because *A* is a 3×4 matrix, by **Theorem 2.5.2** , we have

$r(A) \leq \min\{3, 4\} = 3.$

Similarly, $r(B) \leq \min\{2, 3\} = 2$ and $r(C) \leq \min\{4, 3\} = 3.$

Theorem 2.5.3.

Let A be an $m \times n$ matrix. Assume that the last k rows of A are zero rows. Then $r(A) \leq m - k.$

Proof

Let *C* be the $(m - k) \times n$ matrix consisting of rows 1 to $(m - k)$ of *A*. We use row operations to change *C* to a reduced row echelon matrix denoted by *F*. Let *B* be the $m \times n$ matrix whose first $m - k$ rows are the same as *F* and other rows are zero rows. Then *B* is the row echelon matrix of *A* and by **Theorem 2.5.2** , $r(B) \leq m - k.$. It follows that $r(A) \leq m - k.$

Now, we give the definition of row equivalent matrices.

Definition 2.5.2.

Two $m \times n$ matrices *A* and *B* are said to be row equivalent if *B* can be obtained from *A* by a finite sequence of row operations.

The following result shows that the ranks of two row equivalent matrices are the same.

Theorem 2.5.4.

Let A and B be row equivalent $m \times n$ matrices. Then $r(A) = r(B)$.

The method of finding the rank of a matrix is to follow **Steps 1-3** given in **Section 2.4** for each round until a row echelon matrix is obtained, and then to count the leading entries of the row echelon matrix.

Example 2.5.2.

For each of the following matrices, find its rank, nullity, and pivot columns.

$$A = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} B = \begin{pmatrix} 1 & 2 & -1 \\ 2 & -1 & 3 \end{pmatrix} C = \begin{pmatrix} 0 & 4 & 2 & 1 \\ 0 & 2 & 1 & 1 \\ 1 & 1 & 1 & 4 \\ 2 & 0 & 1 & 3 \end{pmatrix}$$

Solution

Because *A* itself is a reduced row echelon matrix, we do not need to use any row operations. Note that *A* has 2 leading entries, $r(A) = 2$. Because *A* is a 3×4 matrix, $\text{null}(B) = 3 - 2 = 1$. Because the columns 1 and 2 of *A* contain the leading entries, the column vectors $\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$ are the pivot columns of *A*.

Because *B* is not a row echelon matrix, we need to use row operations to find its row echelon matrix.

$$B = \begin{pmatrix} 1 & 2 & -1 \\ 2 & -1 & 3 \end{pmatrix} \Big\downarrow R_1(-2) + R_2 \rightarrow \begin{pmatrix} 1 & 2 & -1 \\ 0 & -5 & 5 \end{pmatrix} : = B^*.$$

The last matrix *B** is a row echelon matrix of *B* and there are two leading entries. Hence, $r(B) = 2$. Because *B* is a 2×3 matrix, $\text{null}(B) = 3 - 2 = 1.$

Because the columns 1 and 2 of *B** contain the leading entries, the columns 1 and 2 of B: $\begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 2 \\ -1 \end{pmatrix}$ are the pivot columns of *B*.

We need to use row operations to change *C* to its row echelon matrix.

$$C = \begin{pmatrix} 0 & 4 & 2 & 1 \\ 0 & 2 & 1 & 1 \\ 1 & 1 & 1 & 4 \\ 2 & 0 & 1 & 3 \end{pmatrix} R_{1,\ 3} \to \begin{pmatrix} 1 & 1 & 1 & 4 \\ 0 & 2 & 1 & 1 \\ 0 & 4 & 2 & 1 \\ 2 & 0 & 1 & 3 \end{pmatrix}\Bigg\downarrow R_1(-2) + R_4 \to$$

$$\begin{pmatrix} 1 & 1 & 1 & 4 \\ 0 & 2 & 1 & 1 \\ 0 & 4 & 2 & 1 \\ 0 & -2 & -1 & -5 \end{pmatrix}\Bigg\downarrow \underset{R_2(1)+R_4}{\overset{R_2(-2)+R_3}{\longrightarrow}} \begin{pmatrix} 1 & 1 & 1 & 4 \\ 0 & 2 & 1 & 1 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & -4 \end{pmatrix}\Bigg\downarrow$$

$$R_3(-4) + R_4 \to \begin{pmatrix} 1 & 1 & 1 & 4 \\ 0 & 2 & 1 & 1 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 \end{pmatrix} : = C^*.$$

The last matrix is a row echelon matrix of *C* and there are three leading entries. Hence, $r(C) = 3$ and $\mathrm{null}(C) = 4 - 3 = 1$. Because the columns 1, 2, 4 of *C** contain the leading entries, the columns 1, 2, 4 of *C*:

$$\begin{pmatrix} 0 \\ 0 \\ 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 4 \\ 2 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 4 \\ 3 \end{pmatrix}$$

are the pivot columns of *C*.

We state the following result without proof. The result shows that the rank of A^T is equal to the rank of *A*.

Theorem 2.5.5.

Let A be an $m \times n$ *matrix. Then*

1. $r(A^T) = r(A)$;
2. $\text{null}(A^T) = m - r(A)$.

For some matrices, you can find the rank of A^T and then use **Theorem 2.5.5** to get the rank of *A*.

Example 2.5.3.

Find the rank of

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

Solution

$$A^T = \begin{pmatrix} 1 & 1 & 1 & 2 \\ 0 & 2 & 1 & -1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & -1 & 0 \end{pmatrix} \xrightarrow{R_3(1)+R_4} \begin{pmatrix} 1 & 1 & 1 & 2 \\ 0 & 2 & 1 & -1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

The last matrix is a row echelon matrix and contains three leading entries. Hence, $r(A^T) = 3$. By **Theorem 2.5.5**, $r(A^T) = r(A^T) = 3$.

In **Example 2.5.3**, if we use row operations on *A* to find the rank of *A*, we would have needed to use five row operations while we only need to use one row operation on A^T to get the rank of *A*.

Exercises

1. For each of the following matrices, find its rank, nullity, and pivot columns. Clearly mark every leading entry *a* by using the symbol a and use the symbol ↓ to show the direction of eliminating nonzero entries below leading entries.

$$A = \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} B = \begin{pmatrix} 0 & 2 & -1 \\ 2 & -1 & 3 \end{pmatrix} C = \begin{pmatrix} 1 & 4 & 2 & 1 \\ 0 & 2 & 1 & 1 \\ 1 & 0 & 1 & 4 \\ 2 & 0 & 1 & -3 \end{pmatrix}$$

$$D = \begin{pmatrix} -1 & 1 & 0 \\ 1 & -2 & 0 \\ 2 & 1 & 1 \\ 0 & -1 & 0 \end{pmatrix} E = \begin{pmatrix} 1 & -2 & -1 & 0 \\ -2 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$$F = \begin{pmatrix} 1 & -1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & -2 \\ 1 & 0 & 0 & 2 & -2 \\ 0 & 0 & -1 & -1 & 5 \end{pmatrix} G = \begin{pmatrix} 2 & -1 & 0 & -1 \\ 0 & 3 & 2 & 7 \\ 3 & 0 & 1 & 2 \\ 5 & -1 & 1 & 1 \end{pmatrix}$$

2.6 Elementary matrices

An elementary matrix is a special type of square matrix that is obtained by performing a single row operation to the identity matrix. Such matrices can be used to derive the algorithm for finding the inverse of an invertible matrix (see **Section 2.7**).

Definition 2.6.1.

An $m \times m$ matrix is called an elementary matrix if it can be obtained by performing a single row operation to the $m \times m$ identity matrix.

Example 2.6.1.

The following matrices are elementary matrices.

$$E_1 = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} E_2 = \begin{pmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} E_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$

$$E_4 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} E_5 = \begin{pmatrix} 1 & 0 & 4 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} E_6 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$

Solution

$$I_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \xrightarrow{R_1(2)} \begin{pmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = E_1.$$

$$I_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \xrightarrow{R_1(-2)+R_2} \begin{pmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = E_2.$$

$$I_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \xrightarrow{R_{2,3}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} = E_3.$$

The matrices $E_1,\ E_2,$ and E_3 are 3×3 elementary matrices. Similarly, $E_4,\ E_5,$ and E_6 are obtained by using the row operations: $R_3(3),\ R_3(4) + R_1,$ and $R_{2,4}$ to I_4, , respectively, and so they are elementary matrices.

The elementary matrices $E_1 - E_6$ in **Example 2.6.1** can be changed back to the corresponding identity matrices by performing a suitable single row operation. For example,

$$E_1 = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \xrightarrow{R_1(\frac{1}{2})} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = I_3.$$

$$E_2 = \begin{pmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \xrightarrow{R_1(2)+R_2} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = I_3.$$

$$E_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \xrightarrow{R_{2,\,3}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = I_3.$$

From the above, we see that the row operations are invertible. Hence, we introduce the definition of inverse row operation.

Definition 2.6.2.

i. $R_i\left(\frac{1}{c}\right)$ is the inverse row operation of $R_i(c)$.
ii. $R_i(-c) + R_j$ is the inverse row operation of $R_i(c) + R_j$.
iii. $R_{i,j}$ is the inverse row operation of $R_{i,j}$.

We use the symbols $E_i(c)$, $E_{i(c)+j}$, and $E_{i,j}$ to denote the elementary matrix obtained by performing the row operation $R_i(c) + R_i(c) + R_j$, and $R_{i,j}$, respectively, to the $m \times m$ identity matrix *I*. Then we have

i. $I_m \xrightarrow{R_i(c)} E_i(c) \xrightarrow{R_i\left(\frac{1}{c}\right)} I_m.$
ii. $I_m \xrightarrow{R_i(c)+R_j} E_{i(c)+j} \xrightarrow{R_1(-c)+R_j} I_m.$
iii. $I_m \xrightarrow{R_{i,j}} E_{i,j} \xrightarrow{R_{i,j}} I_m.$

Example 2.6.2.

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \xrightarrow{R_2(5)} \begin{pmatrix} 1 & 0 \\ 0 & 5 \end{pmatrix} \xrightarrow{R_2\left(\frac{1}{5}\right)} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \xrightarrow{R_1(2) + R_2} \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix} \xrightarrow{R_1(-2) + R_2} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \xrightarrow{R_{1,2}} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \xrightarrow{R_{1,2}} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

It is easy to show

(2.6.1)

$$E_i(c)E_i\left(\frac{1}{c}\right) = I_m; \quad E_{i(c)+j}E_{i\left(\frac{1}{c}\right)+j} = I_m, \quad E_{i,j}E_{i,j} = I_m.$$

The following result shows that if a single row operation is used on an $m \times n$ matrix *A*, then the resulting matrix is equal to the product of the elementary matrix *E* and *A*, where *E* is the matrix obtained by performing the same row operation on the $n \times n$ identity matrix I_n.

Theorem 2.6.1.

Let A be an m × n matrix. Then the following assertions hold.

i. $A \xrightarrow{R_i(c)} B$ *if and only if* $E_i(c)A = B$.

ii. $A \xrightarrow{R_i(c) + R_j} B$ *if and only if* $E_{i(c)+j}A = B$.

iii. $A \quad R_{i,j}\rightarrow \quad B$ *if and only if* $E_{i,j}A = B$.

Proof

i. $$\begin{pmatrix} 1 & \cdots & 0\cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & c\cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0\cdots & 1 \end{pmatrix}\begin{pmatrix} a_{11} & \cdots & a_{1i}\cdots & a_{1n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{i1} & \cdots & a_{ii}\cdots & a_{in} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & \cdots & a_{mi}\cdots & a_{mn} \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & \vdots & \vdots \\ ca_{i1} & ca_{i2} & \cdots & ca_{in} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}.$$

ii. $$\begin{pmatrix} 1 & \cdots & 0\cdots & 0\cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 1\ldots & 0\cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & c\cdots & 1\cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0\cdots & 0\cdots & 1 \end{pmatrix}\begin{pmatrix} a_{11} & \cdots & a_{i1}\cdots & a_{1j}\cdots & a_{1n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{i1} & \cdots & a_{ii}\cdots & a_{ij}\cdots & a_{in} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{j1} & \cdots & a_{ji}\cdots & a_{jj}\cdots & a_{jn} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{m1} & \cdots & a_{mi}\cdots & a_{mj}\cdots & a_{mn} \end{pmatrix}$$

$$= \begin{pmatrix} a_{11} & \cdots & a_{1i}\cdots & a_{1j}\cdots & a_{1n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{i1} & \cdots & a_{ii}\cdots & a_{ij}\cdots & a_{in} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ ca_{i1} + a_{j1} & \cdots & ca_{ii} + a_{ji}\cdots & ca_{ij} + a_{jj}\cdots & ca_{in} + a_{jn} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{m1} & \cdots & a_{mi}\cdots & a_{mj}\cdots & a_{mn} \end{pmatrix}.$$

iii. $$\begin{pmatrix} 1 & \cdots & 0\cdots & 0\cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0\cdots & 1\cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 1\cdots & 0\cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0\cdots & 0\cdots & 1 \end{pmatrix} \begin{pmatrix} a_{11} & \cdots & a_{1i}\cdots & a_{1j}\cdots & a_{1n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{i1} & \cdots & a_{ii}\cdots & a_{ij}\cdots & a_{in} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{j1} & \cdots & a_{ji}\cdots & a_{jj}\cdots & a_{jn} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{m1} & \cdots & a_{mi}\cdots & a_{mj}\cdots & a_{mn} \end{pmatrix}$$

$$= \begin{pmatrix} a_{11} & \cdots & a_{1i} \cdots & a_{1j} \cdots & a_{1n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{j1} & \cdots & a_{ji} \cdots & a_{jj} \cdots & a_{jn} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{i1} & \cdots & a_{ii} \cdots & a_{ij} \cdots & a_{in} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{m1} & \cdots & a_{mi} \cdots & a_{mj} \cdots & a_{mn} \end{pmatrix}.$$

Example 2.6.3.

Let

$$I = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \text{ and } A = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 2 & -3 \\ 2 & 1 & 1 \end{pmatrix}.$$

i. Apply $R_2(-2)$ to *I* and *A*:

$$I = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} R_2(-2) \rightarrow \begin{pmatrix} 1 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 1 \end{pmatrix} = E_1$$

$$A = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 2 & -3 \\ 2 & 1 & 1 \end{pmatrix} R_2(-2) \rightarrow \begin{pmatrix} 0 & 1 & 1 \\ -2 & -4 & 6 \\ 2 & 1 & 1 \end{pmatrix} = B_1.$$

Show that $E_1 A = B_1$.

ii. Apply $R_2(-2) + R_3$ to *I* and *A*:

$$I = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} R_2(-2) + R_3 \rightarrow \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -2 & 1 \end{pmatrix} = E_2$$

$$A = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 2 & -3 \\ 2 & 1 & 1 \end{pmatrix} R_2(-2) + R_3 \rightarrow \begin{pmatrix} 0 & 1 & 1 \\ 1 & 2 & -3 \\ 0 & -3 & 7 \end{pmatrix} = B_2.$$

Show that $E_2 A = B_2$.

iii. Apply $R_{1,\,2}$ to *I* and *A*:

$$I = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} R_{1,\,2}\!\rightarrow \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} = E_3$$

$$A = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 2 & -3 \\ 2 & 1 & 1 \end{pmatrix} R_{1,\,2}\!\rightarrow \begin{pmatrix} 1 & 2 & -3 \\ 0 & 1 & 1 \\ 2 & 1 & 1 \end{pmatrix} = B_3.$$

Show that $E_3 A = B_3$.

Solution

$$E_1 A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} 0 & 1 & 1 \\ 1 & 2 & -3 \\ 2 & 1 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 1 \\ -2 & -4 & 6 \\ 2 & 1 & 1 \end{pmatrix} = B_1.$$

$$E_2 A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -2 & 1 \end{pmatrix}\begin{pmatrix} 0 & 1 & 1 \\ 1 & 2 & -3 \\ 2 & 1 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 2 & -3 \\ 0 & -3 & 7 \end{pmatrix} = B_2.$$

$$E_3 A = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} 0 & 1 & 1 \\ 1 & 2 & -3 \\ 2 & 1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 2 & -3 \\ 0 & 1 & 1 \\ 2 & 1 & 1 \end{pmatrix} = B_3.$$

By using **Theorem 2.6.1** , we can prove the following result, which gives the relation between a matrix and the resulting matrix obtained by using multiple row operations on the matrix.

Theorem 2.6.2.

If we use k row operations on an $m \times n$ matrix A and denote the resulting matrix by B,then

$$B = E_k E_{k-1} \cdots E_2 E_1 A,$$

where $E_1, E_2, \cdots, E_k$ are $m \times m$ elementary matrices corresponding to the row operations used on A.

Proof

We denote the row operations by $r_1, r_2, \cdots, r_k$ in order. Then

$$I \xrightarrow{r_i} E_i \text{ for each } i \in \{1, 2, \cdots, k\}$$

and by **Theorem 2.6.1** , we have

$$A \xrightarrow{r_1} E_1 A \xrightarrow{r_2} E_2(E_1 A) = E_2 E_1 A \cdots \xrightarrow{r_k} E_k\left(E_{k-1} \cdots E_2 E_1 A\right) = B.$$

To understand **Theorem 2.6.2** , we give the following example. We use two row operations to change *A* to *B* as follows:

$$A = \begin{pmatrix} 0 & 2 & -1 \\ 1 & -1 & 3 \\ 2 & 0 & 1 \end{pmatrix} \xrightarrow{R_{1,2}} \begin{pmatrix} 1 & -1 & 3 \\ 0 & 2 & -1 \\ 2 & 0 & 1 \end{pmatrix} \xrightarrow{R_1(-2) + R_3} \begin{pmatrix} 1 & -1 & 3 \\ 0 & 2 & -1 \\ 0 & 2 & -5 \end{pmatrix} = B.$$

$$I = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} R_{1,\,2}\rightarrow \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} = E_1.$$

$$I = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} R_1(-2) + R_3\rightarrow \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -2 & 0 & 1 \end{pmatrix} = E_2.$$

By **Theorem 2.6.2** , we obtain $B = E_2E_1A$. In the following, we directly verify that the result $B = E_2E_1A$ holds.

$$\begin{aligned} E_1E_2A &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -2 & 0 & 1 \end{pmatrix}\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} 0 & 2 & -1 \\ 1 & -1 & 3 \\ 2 & 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -2 & 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & -1 & 3 \\ 0 & 2 & -1 \\ 2 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & -1 & 3 \\ 0 & 2 & -1 \\ 0 & 2 & -5 \end{pmatrix} = B. \end{aligned}$$

We end this section with the ranks of elementary matrices. By **Theorem 2.5.4** , we have the following result.

Corollary 2.6.1.

Let E be an $m \times m$ elementary matrix. Then $r(E) = m$.

Proof

Because E and I are row equivalent and $r(I) = m,$ it follows from **Theorem 2.5.4** that $r(E) = m.$

By **Theorem 2.6.2**, we obtain the following result, which shows that the rank of the product of finitely many elementary matrices is *n*.

Theorem 2.6.3.

Let $E_1, \cdots, E_k$ be $m \times m$ elementary matrices and $E = E_k E_{k-1} \cdots E_2 E_1.$ Then $r(E) = m.$

Proof

By **Theorem 2.6.2**, E and E_1 are row equivalent. By **Theorem 2.5.4** and **Corollary 2.6.1**, $r(E) = r(E_1) = m.$

Exercises

1. Determine which of the following matrices are elementary matrices.

    1. $\begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}$
    2. $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$
    3. $\begin{pmatrix} 1 & 5 \\ 0 & 1 \end{pmatrix}$
    4. $\begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$

5. $\begin{pmatrix} 0 & 2 \\ 2 & 0 \end{pmatrix}$

6. $\begin{pmatrix} 1 & 0 \\ -2 & 1 \end{pmatrix}$

7. $\begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}$

8. $\begin{pmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$

9. $\begin{pmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}$

10. $\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 5 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$

11. $\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$

12. $\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$

2. Use row operations to change the above matrices (1), (2), (3), (6), (7), (8), (10), and (12) to an identity matrix.

3. Let $I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$. Verify the following assertions.

   1. $E_2(5)E_2\left(\frac{1}{5}\right) = I$;
   2. $E_{1(2)+2}E_{1\left(\frac{1}{2}\right)+2} = I$;
   3. $E_{1,\,2}E_{1,\,2} = I$.

4. Find a matrix *E* such that $B = EA$, where

$$A = \begin{pmatrix} 0 & 2 & -1 \\ 1 & -1 & 3 \\ 2 & 0 & 1 \end{pmatrix} \text{ and } B = \begin{pmatrix} 1 & -1 & 3 \\ 0 & 2 & -1 \\ 0 & 0 & -4 \end{pmatrix}.$$

2.7 The inverse of a square matrix

In this section, we study how to determine a square matrix to be invertible by using its rank, and if it is invertible, how to find its inverse. After we study **Section 3.4**, we shall see that determinants of matrices also can be used to determine whether a square matrix is invertible.

Definition 2.7.1.

Let *A* be an $n \times n$ matrix. If there exists an $n \times n$ matrix *B* such that

$$AB = I_n,$$

then *A* is said to be invertible (or nonsingular) and *B* is called the inverse of *A*. We write $B = A^{-1}$. If *A* is not invertible, then *A* is said to be singular.

We only introduce the notion of inverse matrices for square matrices. Therefore, if a matrix is not a square matrix, then it has no inverses.

By **Definition 2.7.1**, if *A* is invertible, then $AA^{-1} = I_n$. Moreover, *A* is not invertible if and only if for each $n \times n$ matrix *B*, $AB \neq I_n$.

We state the following theorem without proof, which provides some equivalent results for inverse matrices.

Theorem 2.7.1.

Let A and B be n × n square matrices. Then the following are equivalent.

i. $AB = I_n$.
ii. $BA = I_n$.
iii. $AB = BA = I_n$.

By **Theorem 2.7.1**, we can show that if *A* is invertible, then its inverse is unique. Indeed, assume that there are two $n \times n$ matrices B_1 and B_2 satisfying $AB_1 = I_n$ and $AB_2 = I_n$. Then, by **Theorem 2.7.1** (*ii*), $B_1A = I_n$. Hence,

$$B_1 = B_1I_n = B_1(AB_2) = (B_1A)B_2 = I_nB_2 = B_2.$$

This shows that the inverse of *A* is unique.

Example 2.7.1.

(1) Show that *B* is an inverse of *A* if

$$A = \begin{pmatrix} 2 & -5 \\ -1 & 3 \end{pmatrix} \text{ and } B = \begin{pmatrix} 3 & 5 \\ 1 & 2 \end{pmatrix}.$$

(2) Show that and $\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$ $\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$ are not invertible.

Solution

(1) Because $AB = \begin{pmatrix} 2 & -5 \\ -1 & 3 \end{pmatrix}\begin{pmatrix} 3 & 5 \\ 1 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = I_2$, by **Definition 2.7.1** , *B* is the inverse of *A*.

(2) Let $\begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix}$ be a 2×2 matrix. Because

$$\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \neq \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

by **Definition 2.7.1** , $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ is not invertible. Because

$$\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} = \begin{pmatrix} b_{11}+b_{21} & b_{12}+b_{22} \\ b_{11}+b_{21} & b_{12}+b_{22} \end{pmatrix} \neq \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

by **Definition 2.7.1** , $\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$ is not invertible.

Example 2.7.2.

Assume that $a_i \neq 0$ for each $i \in I_n$. Then

$$\begin{pmatrix} a_1 & 0 & \cdots & 0 \\ 0 & a_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_n \end{pmatrix}^{-1} = \begin{pmatrix} \frac{1}{a_1} & 0 & \cdots & 0 \\ 0 & \frac{1}{a_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{a_n} \end{pmatrix}.$$

Solution

Because $a_i \neq 0$ for each $i \in I_n$, $\operatorname{diag}\left(a_1^{-1}, a_2^{-1}, \cdots, a_n^{-1}\right)$ exists. It is easy to verity that

$$\mathrm{diag}\left(a_1,\ a_2,\ \cdots,\ a_n\right)\mathrm{diag}\left(a_1^{-1},\ a_2^{-1},\ \cdots,\ a_n^{-1}\right) = I_n.$$

The result follows from **Definition 2.7.1**.

In **Example 2.7.2**, we use **Definition 2.7.1** to verify whether a 2×2 matrix is invertible. But it is not a good method to use **Definition 2.7.1** to verify whether a square matrix with a large size is invertible. Later, we shall study some other useful methods such as using ranks and determinants to determine whether square matrices are invertible.

By **Definition 2.7.1** and **(2.6.1)**, we obtain the following result, which shows that every elementary matrix is invertible.

Theorem 2.7.2.

Every elementary matrix is invertible and its inverse is an invertible elementary matrix.

Theorem 2.7.3.

Assume that B is a row echelon matrix of an $n \times n$ matrix A. Then there exists an invertible $n \times n$ matrix $E = E_kE_{k-1}\cdots E_2E_1$ such that $B = EA,$ where E_i is an $n \times n$ elementary matrix for $i = 1,\ \cdots,\ k$.

Proof

By **Theorem 2.6.2**, there exists an $n \times n$ matrix $E = E_kE_{k-1}\cdots E_2E_1$ such that $B = EA$. By **Theorem 2.7.2**, E_i is an $n \times n$ invertible matrix and by **Theorem 2.7.4**, *E* is invertible.

By **Definition 2.7.1**, we prove the following result, which will be used to prove **Theorem 2.7.3**.

Theorem 2.7.4.

If A and B are $n \times n$ invertible matrices, then AB is invertible and

$$(AB)^{-1} = B^{-1}A^{-1}.$$

Proof

Because

$$(B^{-1}A^{-1})(AB) = B^{-1}(A^{-1}A)B = B^{-1}IB = B^{-1}B = I,$$

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

Now, we show how to find the inverse matrix of a 2×2 matrix *A* given in **(1.2.4)**.

Theorem 2.7.5.

Let $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ be the same as in **(1.2.4)**. *Then the following assertions hold.*

1. *is invertible if and only if $|A| \neq 0$.*

2. *If A is invertible, then*

(2.7.1)

$$A^{-1} = \frac{1}{|A|}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

Proof

Let

(2.7.2)

$$B = \begin{pmatrix} x_1 & x_2 \\ y_1 & y_2 \end{pmatrix}.$$

We find $x_1,\ x_2,\ y_1,\ y_2$ such that $AB = I.$ Because

$$AB = \begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} x_1 & x_2 \\ y_1 & y_2 \end{pmatrix} = \begin{pmatrix} ax_1 + by_1 & ax_2 + by_2 \\ cx_1 + dy_1 & cx_2 + dy_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

we obtain

(2.7.3)

$$ax_1 + by_1 = 1, \quad cx_1 + dy_1 = 0.$$

(2.7.4)

$$ax_2 + by_2 = 0, \quad cx_2 + dy_2 = 1.$$

By **(2.7.3)** , we obtain

$$adx_1 + bdy_1 = d, \quad bcx_1 + bdy_1 = 0.$$

Subtracting the above two equations implies $(ad - bc)x_1 = d$ and $|A|x_1 = d.$ Also, by **(2.7.3)** , we obtain

$$acx_1 + bcy_1 = c, \quad \text{and } acx_1 + ady_1 = 0.$$

Subtracting the above two equations implies $|A|y_1 = -c.$ Similarly, applying the above method to **(2.7.4)** , we obtain $|A|x_2 = -b$ and $|A|y_2 = a.$ Hence, if $AB = I,$ then

$$\left|A\right|x_1 = d,\quad \left|A\right|y_1 = -c,\quad \left|A\right|x_2 = -b,\ \text{ and }\ \left|A\right|y_2 = a,$$

that is, the condition **(2.7.5)** is a necessary condition for $AB = I$.

(1) Assume that *A* is invertible. By **Example 2.7.1** (2), *A* is not a zero matrix. So one of *a*, *b*, *c*, *d* is not zero. This, together with **(2.7.5)**, implies $|A| \neq 0$. Conversely, assume that $|A| \neq 0$. By **(2.7.5)**,

(2.7.6)

$$x_1 = \frac{d}{|A|},\ y_1 = \frac{-c}{|A|},\ x_2 = \frac{-b}{|A|}\ \text{and}\ y_2 = \frac{a}{|A|},$$

that is,

(2.7.7)

$$B = \begin{pmatrix} x_1 & x_2 \\ y_1 & y_2 \end{pmatrix} = \frac{1}{|A|}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

and $AB = I$. Hence, by **Definition 2.7.1**, *A* is invertible.

(2) By the above proof, we see that if *A* is invertible, then $A^{-1} = B$.

Example 2.7.3.

For each of the following matrices, determine whether it is invertible. If so, calculate its inverse.

$$A = \begin{pmatrix} 2 & -4 \\ 1 & 3 \end{pmatrix}\ B = \begin{pmatrix} 1 & 2 \\ -2 & -4 \end{pmatrix}.$$

Solution

Because $|A| = \begin{vmatrix} 2 & -4 \\ 1 & 3 \end{vmatrix} = (2)(3) - (-4)(1) = 10 \neq 0$, by **Theorem 2.7.5**, *A* is invertible and

$$A^{-1} = \frac{1}{|A|}\begin{pmatrix} 3 & 4 \\ -1 & 2 \end{pmatrix} = \frac{1}{10}\begin{pmatrix} 3 & 4 \\ -1 & 2 \end{pmatrix} = \begin{pmatrix} \frac{3}{10} & \frac{2}{5} \\ -\frac{1}{10} & \frac{1}{5} \end{pmatrix}.$$

Because $|B| = \begin{vmatrix} 1 & 2 \\ -2 & -4 \end{vmatrix} = (1)(-4) - (2)(-2) = 0$, *B* is not invertible.

Example 2.7.4.

For each of the following matrices, find all $x \in \mathbb{R}$ such that it is invertible.

$$A = \begin{pmatrix} 1 & x^3 \\ 1 & 1 \end{pmatrix} \quad B = \begin{pmatrix} x & 4 \\ 1 & x \end{pmatrix}$$

Solution

Because if $|A| = \begin{vmatrix} 1 & x^3 \\ 1 & 1 \end{vmatrix} = 1 - (x^3)(1) = 1 - x^3 = 0,$ then $x = 1$. Hence, when $x \neq 1,$ $|A| \neq 0$ and by **Theorem 2.7.5**, *A* is invertible. Because if $|B| = \begin{vmatrix} x & 4 \\ 1 & x \end{vmatrix} = (x)(x) - (4)(1) = x^2 - 4 = 0,$ then $x = -2$ or $x = 2$. Hence, when $x \neq -2$ and $x \neq 2,$ $|B| \neq 0$ and by **Theorem 2.7.5**, *B* is invertible.

In **Theorem 2.7.5** (1), we use the determinants of 2×2 matrices to determine whether they are invertible. The result remains true for $n \times n$ matrices, see **Theorem 3.4.1**. In **Corollary 8.1.3**, we shall show how to use the eigenvalues of a matrix to determine whether it is invertible. Here we show that we can show how to use the rank of an $n \times n$ matrix to determine whether it is invertible.

Theorem 2.7.6.

Let A be an $n \times n$ matrix. Then A is invertible if and only if $r(A) = n$.

Proof

(1) Assume that *A* is invertible. Then there exists an $n \times n$ matrix *B* such that $AB = I.$ If $r(A) < n,$ then by **Theorem 2.7.3**, there exists an invertible matrix *E* such that $C = EA,$ where *C* is the reduced row echelon matrix of *A*. By **Theorem 2.6.3**, $r(E) = n$ and by **Theorem 2.5.4**, $r(C) = r(A) < n.$ This implies that the rows from $r(A) + 1$ to *n* of the reduced row echelon matrix *C* are zero rows and thus, the rows from $r(A) + 1$ to *n* of *CB* are zero rows. By **Theorem 2.5.3**, $r(CB) \leq r(A) < n.$ On the other hand, because $AB = I,$ we have

$$CB = (EA)B = E(AB) = EI = E \text{ and } n = r(E) = r(CB) < n,$$

a contradiction.

(2) Assume that $r(A) = n$ and *B* is the reduced row echelon matrix of *A*. Then $r(B) = n$ and $B = I_n.$ By **Theorem 2.7.3**, there exists an invertible matrix *E* such that $I_n = EA$ and *A* is invertible.

By **Theorem 2.7.5** (1), we see that when $n = 2,$ the determinant of *A* can be used to justify whether *A* is invertible. The result remains true when $n \geq 3,$ see **Theorem 3.4.1** (*i*).

Example 2.7.5.

For each of the following matrices, determine whether it is invertible.

$$A = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix} \quad B = \begin{pmatrix} 1 & 1 & 1 \\ 0 & -1 & 0 \\ 1 & 0 & 1 \end{pmatrix}$$

Solution

$$A = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix} \Bigg\downarrow R_1(-1)+R_3 \rightarrow \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & -1 & 0 \end{pmatrix} \Bigg\downarrow R_2(1)+R_3 \rightarrow \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}.$$

Then *r*(*A*) = 3. By **Theorem 2.7.6**, *A* is invertible.

$$B = \begin{pmatrix} 1 & 1 & 1 \\ 0 & -1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \Bigg\downarrow R_1(-1)+R_3 \rightarrow \begin{pmatrix} 1 & 1 & 1 \\ 0 & -1 & 1 \\ 0 & -1 & 0 \end{pmatrix} \Bigg\downarrow R_2(-1)+R_3 \rightarrow \begin{pmatrix} 1 & 1 & 1 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Then *r*(*B*) = 2. By **Theorem 2.7.6**, *B* is not invertible.

Corollary 2.7.1.

1. *An $n \times n$ matrix is invertible if and only if its reduced row echelon matrix is the identity matrix In.*
2. *An $n \times n$ matrix is not invertible if and only if its row echelon matrix contains at least one zero row.*

Proof

We only prove (1). Assume that *B* is the reduced row echelon matrix of *A*. By **Theorem 2.5.4**, $r(A) = r(B)$. If *A* is invertible, it follows from **Theorem 2.7.6** that $r(A) = n$. Hence, $r(B) = n$ and $B = I_n$. Conversely, if the reduced row echelon matrix of *A* is the identity matrix I_n, then by **Theorem 2.7.3**, there exists an invertible matrix *E* such that $I_n = EA$ and *A* is invertible.

Corollary 2.7.2.

An $n \times n$ matrix is invertible if and only if there exists a finite sequence of elementary matrices $F_1, F_2, \cdots, F_{k-1}, F_k$ such that

$$A = F_1 F_2 \cdots F_{k-1} F_k.$$

Proof

By **Theorem 2.7.3** and **Corollary 2.7.1**, *A* is invertible if and only if there exists an invertible matrix $E = E_k E_{k-1} \cdots E_2 E_1$ such that $I_n = EA$, where E_i is an elementary matrix for $i = 1, \cdots, k$ if and only if

$$A = E^{-1} = (E_k E_{k-1} \cdots E_2 E_1)^{-1} = E_1^{-1} E_2^{-1} \cdots E_{k-1}^{-1} E_k^{-1}.$$

Let $F_i = E_i^{-1}$ for $i \in \{1, 2, \cdots, k\}$. By **Theorem 2.7.2** , F_i is an invertible elementary matrix for $i \in \{1, 2, \cdots, k\}$.

Corollary 2.7.3.

Let A and B be $m \times n$ matrices. Assume that P is an $m \times m$ invertible matrix such that $B = PA$. Then $r(A) = r(B)$.

Proof

Because *P* is an $m \times m$ invertible matrix, by **Corollary 2.7.2** , there exists a finite sequence of $m \times m$ elementary matrices $F_1, F_2, \cdots, F_{k-1}, F_k$ such that $P = F_1 F_2 \cdots F_{k-1} F_k$ and thus, $B = F_1 F_2 \cdots F_{k-1} F_k A$. This shows that we use *k* row operations to change *A* to *B* and *A* and *B* are row equivalent. By **Theorem 2.5.4** , $r(A) = r(B)$.

The following result shows that the converse of **Theorem 2.7.4** holds.

Theorem 2.7.7.

Assume that A and B are $n \times n$ matrices and AB is invertible. Then both A and B are invertible.

Proof

Let $C = AB$. Because *AB* is invertible, by **Theorem 2.7.6** , $r(C) = n$. Let *D* be the reduced row echelon matrix of *A*. By **Theorem 2.7.3** , there exists an invertible matrix *E* such that $D = EA$. Hence,

(2.7.8)

$$EC = EAB = DB.$$

By **Theorem 2.5.4** , $r(A) = r(D)$ and $r(DB) = r(C) = n$. This implies $r(D) = n$. In fact, if $r(D) < n$, then because *D* is a row echelon matrix, by **Corollary 2.7.1** (2), *D* contains at least one zero row, so the last row of *D* must be a zero row. By **(2.2.5)** , we see that the last row of *DB* is a zero row. By **Theorem 2.5.3** , we have $r(DB) < n$, a contradiction. Hence, $r(A) = r(D) = n$. By **Theorem 2.7.6** , *A* is invertible. Now, because *A* is invertible, by **Corollary 2.7.1** (1), $D = I$ and by **(2.7.8)** , $EC = B$. By **Theorem 2.5.4** , $r(B) = r(C) = n$ and by **Theorem 2.7.6** , *B* is invertible.

Methods of finding the inverses

By **Theorem 2.7.3** and **Corollary 2.7.1** , we see that if an $n \times n$ matrix *A* is invertible, then there exists a matrix $E = E_k E_{k-1} \cdots E_2 E_1$ such that $EA = I_n$. Hence, $E = A^{-1}$. Because $EA = I_n$ and $EI_n = E = A^{-1}$, so

$$E(A \mid I_n) = (EA \mid EI_n) = (I_n \mid A^{-1}).$$

This provides a method to find the inverse of *A*.

Theorem 2.7.8.

i. *If we can use row operations to change (A / I_n) to a matrix (I_n / B), then $A^{-1} = B$.*
ii. *If we can use row operations to change (A|I_n) to a row echelon matrix (C|D), where C contains at least one zero row, then A is not invertible.*

Proof

1. Assume that we use row operations to change (A/I_n) to a matrix (I_n/B). From the left sides of $(A\,|\,I_n)$ and $(I_n\,|\,B)$, we see that A and I_n are row equivalent. By Theorems (2.5.4), $r(A) = r(I_n) = n$. By **Theorem 2.7.6**, A is invertible. By **Theorem 2.7.3**, there exists an invertible matrix $E = E_k E_{k-1} \cdots E_2 E_1$ such that $EA = I_n$. This implies $E = A^{-1}$. From the right sides of $(A\,|\,I_n)$ and $(I_n\,|\,B)$, we see that $EI_n = B$. This implies $B = EI_n = E = A^{-1}$.
2. Assume that we use row operations to change $(A\,|\,I_n)$ to the reduced row echelon matrix $(C\,|\,D)$, where C contains at least one zero row. From the left sides of $(A\,|\,I_n)$ and $(C\,|\,D)$, we see that A and C are row equivalent. By **Theorems (2.5.4)** and **2.5.3**, $r(A) = r(C) < n$. It follows from **Theorem 2.7.6** that A is not invertible.

Remark 2.7.1

By **Theorem 2.7.8**, the method to find the inverse of an $n \times n$ matrix A is given below.

i. Change the matrix $(A\,|\,I_n)$ to a row echelon matrix $(C\,|\,D)$. If C contains at least one zero row, then A is not invertible.
ii. If C does not contain any zero rows, then A is invertible and we continue changing the row echelon matrix $(C\,|\,D)$ to the reduced row echelon matrix, which definitely is of the form $(I_n\,|\,A^{-1})$.

Example 2.7.6.

For each of the following matrices, determine whether it is invertible. If so, find its inverse.

$$A = \begin{pmatrix} 2 & -3 \\ -4 & 5 \end{pmatrix} \quad B = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix} \quad C = \begin{pmatrix} 1 & -3 & 4 \\ 2 & -5 & 7 \\ 0 & -1 & 1 \end{pmatrix}$$

Solution

$$(A\,|\,I_2) = \left(\begin{array}{cc|cc} 2 & -3 & 1 & 0 \\ -4 & 5 & 0 & 1 \end{array}\right) \xrightarrow{R_1(2)+R_2} \left(\begin{array}{cc|cc} 2 & -3 & 1 & 0 \\ 0 & -1 & 2 & 1 \end{array}\right) \xrightarrow{R_2(-1)} \left(\begin{array}{cc|cc} 2 & -3 & 1 & 0 \\ 0 & 1 & -2 & -1 \end{array}\right) \xrightarrow{R_2(3)+R_1} \left(\begin{array}{cc|cc} 2 & 0 & -5 & -3 \\ 0 & 1 & -2 & -1 \end{array}\right) \xrightarrow{R_1\left(\frac{1}{2}\right)} \left(\begin{array}{cc|cc} 1 & 0 & -\frac{5}{2} & -\frac{3}{2} \\ 0 & 1 & -2 & -1 \end{array}\right) = (I_2\,|\,A^{-1}).$$

Hence, A is invertible and $A^{-1} = \begin{pmatrix} -\frac{5}{2} & -\frac{3}{2} \\ -2 & -1 \end{pmatrix}$.

$$(B\,|\,I_3) = \left(\begin{array}{ccc|ccc} 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 \end{array}\right) \xrightarrow{R_1(-1)+R_3} \left(\begin{array}{ccc|ccc} 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & -1 & 0 & -1 & 0 & 1 \end{array}\right) \xrightarrow{R_2(1)+R_3} \left(\begin{array}{ccc|ccc} 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & -1 & 1 & 1 \end{array}\right) \xrightarrow[R_3(-1)+R_2]{R_3(-1)+R_1} \left(\begin{array}{ccc|ccc} 1 & 1 & 0 & 2 & -1 & -1 \\ 0 & 1 & 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 & 1 & 1 \end{array}\right) \xrightarrow{R_2(-1)+R_1} \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & -1 & 0 \\ 0 & 1 & 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 & 1 & 1 \end{array}\right) = (I_3\,|\,B$$

Hence B is invertible and $B^{-1} = \begin{pmatrix} 1 & -1 & 0 \\ 1 & 0 & -1 \\ -1 & 1 & 1 \end{pmatrix}$.

$$(C\,|\,I_3) = \left(\begin{array}{ccc|ccc} 1 & -3 & 4 & 1 & 0 & 0 \\ 2 & -5 & 7 & 0 & 1 & 0 \\ 0 & -1 & 1 & 0 & 0 & 1 \end{array}\right) \xrightarrow{R_1(-2)+R_2} \left(\begin{array}{ccc|ccc} 1 & -3 & 4 & 1 & 0 & 0 \\ 0 & 1 & -1 & -2 & 1 & 0 \\ 0 & -1 & 1 & 0 & 0 & 1 \end{array}\right) \xrightarrow{R_2(1)+R_3} \left(\begin{array}{ccc|ccc} 1 & -3 & 4 & 1 & 0 & 0 \\ 0 & 1 & -1 & -2 & 1 & 0 \\ 0 & 0 & 0 & -2 & 1 & 1 \end{array}\right).$$

Because the left side of the last matrix contains a row of zeros, by **Theorem 2.7.8** , C is not invertible.

By **Definition 2.7.1** , it is easy to prove the following result, which provides some properties of inverse matrices. We omit these proofs.

Theorem 2.7.9.

Let A be an invertible $n \times n$ matrix. Then the following assertions hold.

1. $(A^{-1})^{-1} = A$.
2. *If $k \neq 0$, then kA is invertible and* $(kA)^{-1} = \frac{1}{k}A^{-1}$.
3. A^n *is invertible for* $n \in \mathbb{N}$ *and* $(A^n)^{-1} = (A^{-1})^n$. *In this case, we write* $(A^{-n}) = (A^{-1})^n$.
4. A^T *is invertible and* $(A^T)^{-1} = (A^{-1})^T$.
5. *If A is symmetric, then* A^{-1} *is symmetric.*
6. AA^T *and* A^T *A are invertible.*
7. *If A is triangular, then A is invertible if and only if its diagonal entries are all nonzero.*
8. *If A is lower triangular, then* A^{-1} *is lower triangular. If A is upper triangular, then* A^{-1} *is upper triangular.*

Exercises

1. Show that B is an inverse of A if

$$A = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \text{ and } B = \begin{pmatrix} \frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}.$$

2. Use **Definition 2.7.1** to show that $\begin{pmatrix} 1 & -1 \\ -2 & 2 \end{pmatrix}$ is not invertible.

3. For each of the following matrices, determine whether it is invertible. If so, find its inverse.

$$A = \begin{pmatrix} \frac{1}{3} & 0 & 0 \\ 0 & \frac{1}{4} & 0 \\ 0 & 0 & 4 \end{pmatrix} \quad B = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 4 \end{pmatrix}$$

$$C = \begin{pmatrix} 3 & -4 \\ 2 & 3 \end{pmatrix} \quad D = \begin{pmatrix} 2 & 4 \\ -1 & -2 \end{pmatrix}$$

4. Let $A = \begin{pmatrix} 1 & x^2 \\ 1 & 1 \end{pmatrix}$. Find all $x \in \mathbb{R}$ such that A is invertible.

5. Let $A = \begin{pmatrix} 1 & 2 \\ 1 & 3 \end{pmatrix}$ and $B = \begin{pmatrix} 3 & 2 \\ 2 & 2 \end{pmatrix}$. Find A^{-1}, B^{-1}, and $(AB)^{-1}$ and verify $(AB)^{-1} = B^{-1}A^{-1}$.

6. For each of the following matrices, use its rank to determine whether it is invertible.

$$A = \begin{pmatrix} 0 & 2 & -5 \\ 1 & 5 & -5 \\ 1 & 0 & 8 \end{pmatrix} \quad B = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 2 & 1 & 2 & 0 \\ -2 & -3 & 1 & -2 \\ 1 & -1 & 0 & 0 \end{pmatrix} \quad C = \begin{pmatrix} 2 & 1 & 4 \\ 1 & -2 & 4 \\ -2 & -1 & -4 \end{pmatrix}$$

7. For each of the following matrices, use row operations to determine if it is invertible. If so, find its inverse.

$$A = \begin{pmatrix} -1 & 2 \\ 4 & 7 \end{pmatrix} \quad B = \begin{pmatrix} 1 & -3 \\ -4 & 5 \end{pmatrix} \quad C = \begin{pmatrix} 1 & -2 \\ -2 & 4 \end{pmatrix}$$

$$D = \begin{pmatrix} 1 & -1 & 1 \\ 0 & 1 & -1 \\ 2 & 0 & -1 \end{pmatrix} \quad E = \begin{pmatrix} 1 & 1 & 2 \\ 2 & 0 & -5 \\ 3 & 2 & 1 \end{pmatrix} \quad F = \begin{pmatrix} 0 & -8 & -9 \\ 2 & 4 & -1 \\ 1 & -2 & -5 \end{pmatrix}$$

8. Let $A = \begin{pmatrix} -1 & 2 \\ 1 & 3 \end{pmatrix}$. Find A^3, A^{-1}, $(A^{-1})^3$, and verify $(A^3)^{-1} = (A^{-1})^3$.

9. Let $A = \begin{pmatrix} -4 & 1 \\ 3 & 1 \end{pmatrix}$. Find A^{-1}, $(A^T)^{-1}$ and verify that $(A^T)^{-1} = (A^{-1})^T$.

10. Let $A = \begin{pmatrix} 1 & 3 \\ 3 & 2 \end{pmatrix}$. Find A^{-1}. Is A^{-1} symmetric?

Chapter 3 Determinants

3.1 Determinants of 3×3 matrices

The determinant of a 2×2 matrix is defined in **(1.2.4)**. It is applied to introduce the Gram determinant in **Section 1.2** and compute the areas of parallelograms in **Section 1.3**.

In this section, we introduce the determinants of 3×3 matrices. Let

(3.1.1)

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$$

be a 3×3 matrix. The determinant of *A*, denoted by $|A|$, is defined by the number

(3.1.2)

$$\begin{aligned}|A| &= \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} \\ &= (a_{11}a_{22}a_{33} + a_{21}a_{32}a_{13} + a_{31}a_{12}a_{23}) - (a_{13}a_{22}a_{31} + a_{23}a_{32}a_{11} + a_{33}a_{21}a_{12}).\end{aligned}$$

Note that the three numbers in each of the terms are from different rows and different columns of A.

Example 3.1.1.

Calculate the determinant

$$|A| = \begin{vmatrix} 1 & -2 & -1 \\ 0 & 2 & 3 \\ 4 & -5 & -3 \end{vmatrix}.$$

Solution

By **(3.1.3)**, we obtain

$$\begin{aligned}|A| &= [(1)(2)(-3) + (0)(-5)(-1) + (4)(-2)(3)] \\ &- [(-1)(2)(4) + (3)(-5)(1) + (-3)(0)(-2)] = -7.\end{aligned}$$

By **(3.1.3)**, we rewrite $|A|$ as follows.

(3.1.3)

$$\begin{aligned}|A| &= (a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32}) \\ &- (a_{13}a_{22}a_{31} + a_{11}a_{23}a_{32} + a_{12}a_{21}a_{33}).\end{aligned}$$

This gives us another way to compute $|A|$. Indeed, if we write $|A|$ and adjoin it to its first two columns, we get

(3.1.4)

$$\left|\begin{array}{ccc|cc} a_{11} & a_{12} & a_{13} & a_{11} & a_{12} \\ a_{21} & a_{22} & a_{23} & a_{21} & a_{22} \\ a_{31} & a_{32} & a_{33} & a_{31} & a_{32} \end{array}\right.$$

The first three terms in **(3.1.3)** :

$$a_{11}a_{22}a_{33}, \;\; a_{12}a_{23}a_{31}, \;\; a_{13}a_{21}a_{32},$$

can be obtained from **(3.1.4)** in the following way:

Drawing an arrow from a_{11} to a_{33} through a_{22} (from the top left to the bottom right), we see that the product of the three numbers $a_{11},\ a_{22},\ a_{33}$ on the arrow is the first product in **(3.1.3)** . Drawing an arrow from a_{12} to a_{31} through a_{23} and from a_{13} to a_{32} through $a_{21},$ we get the other two products. Similarly, we can get the three terms

$$a_{13}a_{22}a_{31}, \;\; a_{11}a_{23}a_{32}, \;\; a_{12}a_{21}a_{33}$$

by drawing an arrow from a_{13} to a_{31} through a_{22} (from the top right to the bottom left), from a_{11} to a_{32} through a_{23} and from a_{12} to a_{33} through a_{21}.

Example 3.1.2.

Calculate the determinant

$$|A| = \begin{vmatrix} 2 & 4 & 6 \\ -4 & 5 & 6 \\ 7 & -8 & 9 \end{vmatrix}.$$

Solution

By **(3.1.4)**, we obtain

$$|A| = \left|\begin{matrix} 2 & 4 & 6 \\ -4 & 5 & 6 \\ 7 & -8 & 9 \end{matrix}\right| \begin{matrix} 2 & 4 \\ -4 & 5 \\ 7 & -8 \end{matrix} = [(2)(5)(9) + (4)(6)(7) + (6)(-4)(-8)] - [(6)(5)(7) + (2)(6)(-8) + (4)(-4)(9)] = 480.$$

Before we introduce the determinants for matrices with orders greater than 3, we reorganize the six terms in **(3.1.3)**. The basic idea is to use the determinants of 2×2 matrices to compute 3×3 matrices.

(3.1.5)

$$
\begin{aligned}
|A| &= \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = (a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32}) \\
&\quad -(a_{13}a_{22}a_{31} + a_{11}a_{23}a_{32} + a_{12}a_{21}a_{33}) \\
&= a_{11}(a_{22}a_{33} - a_{23}a_{32}) - a_{12}(a_{21}a_{33} - a_{23}a_{31}) + a_{13}(a_{21}a_{32} - a_{22}a_{31}) \\
&= a_{11}\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12}\begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13}\begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix} \\
&= a_{11}M_{11} - a_{12}M_{12} + a_{13}M_{13},
\end{aligned}
$$

where

(3.1.6)

$$
M_{11} = \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix}, \quad M_{12} = \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix}, \quad M_{13} = \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix}.
$$

Remark 3.1.1.

M_{11} is the determinant of the matrix obtained from *A* by deleting row 1 and column 1 of *A*, M_{12} is the determinant of the matrix obtained from *A* by deleting row 1 and column 2, and M_{13} is the determinant of the matrix obtained from *A* by deleting row 1 and column 3.

The expression on the right side of **(3.1.5)** is called an expansion of minors of the determinant $|A|$ corresponding to the first row of *A*.

Note that the signs in **(3.1.5)** change and follow the following rule:

$$(-1)^{1+1},\ (-1)^{1+2},\ (-1)^{1+3}.$$

These powers are obtained from subscripts of $a_{11},\ a_{12},$ and $a_{13},$ respectively.

We define

(3.1.7)

$$A_{11}=(-1)^{1+1}M_{11},\ A_{12}=(-1)^{1+2}M_{12},\ A_{13}=(-1)^{1+3}M_{13}.$$

By **(3.1.5)** , we obtain

(3.1.8)

$$\begin{aligned}|A| &= a_{11}M_{11}-a_{12}M_{12}+a_{13}M_{13}\\ &= a_{11}[(-1)^{1+1}M_{11}]+a_{12}[(-1)^{1+2}M_{12}]+a_{13}[(-1)^{1+3}M_{13}]\\ &= a_{11}A_{11}+a_{12}A_{12}+a_{13}A_{13}.\end{aligned}$$

Definition 3.1.1.

Let M_{ij} be the determinant of the 3×3 matrix obtained from *A* by deleting the *i*th row and the *j*th column of *A*. M_{ij} is called the minor of a_{ij} or the *ij*th minor of *A*.

(3.1.9)

$$A_{ij}=(-1)^{i+j}M_{ij}$$

is called the cofactor of a_{ij} or the *ij*th factor of *A*.

The expression on the right side of **(3.1.8)** is called an expansion of cofactors of the determinant $|A|$ corresponding to the first row of *A*.

Similarly, by reorganizing the six terms in **(3.1.3)**, we obtain the expansion of minors or cofactors of the determinant $|A|$ corresponding to the second row or the third row or each column of *A*. Each of these expressions equals $|A|$.

Theorem 3.1.1.

(3.1.10)

$$\begin{aligned}|A| &= a_{21}A_{21} + a_{22}A_{22} + a_{23}A_{23} = a_{31}A_{31} + a_{32}A_{32} + a_{33}A_{33}\\ &= a_{11}A_{11} + a_{21}A_{21} + a_{31}A_{31} = a_{12}A_{12} + a_{22}A_{22} + a_{32}A_{32}\\ &= a_{13}A_{13} + a_{23}A_{23} + a_{33}A_{33},\end{aligned}$$

Example 3.1.3.

Let

$$A = \begin{pmatrix} 2 & 4 & 6 \\ -4 & 5 & 6 \\ 7 & -8 & 9 \end{pmatrix}.$$

1. Compute $|A|$ by using the minors corresponding to the first row of *A*.
2. Compute $|A|$ by using the cofactors corresponding to the first row of *A*.

Solution

1. By **(3.1.6)**, we have

$$M_{11} = \begin{vmatrix} 5 & 6 \\ -8 & 9 \end{vmatrix} = 93,\ M_{12} = \begin{vmatrix} -4 & 6 \\ 7 & 9 \end{vmatrix} = -78,\ M = \begin{vmatrix} -4 & 5 \\ 7 & -8 \end{vmatrix} = -3.$$

   By **(3.1.5)**, we get

$$|A| = 2M_{11} - 4M_{12} + 6M_{13} = 2(93) - 4(-78) + 6(-3) = 480.$$

2. By **(3.1.7)**, we obtain

$$\begin{aligned} A_{11} &= (-1)^{1+1}M_{11} = M_{11} = 93. \\ A_{12} &= (-1)^{1+2}M_{12} = -M_{12} = -(-78) = 78. \\ A_{13} &= (-1)^{1+3}M_{11} = M_{13} = -3. \end{aligned}$$

   By **(3.1.8)**, $|A| = 2A_{11} + 4A_{12} + 6A_{13} = 2(93) + 4(78) + 6(-3) = 480.$

Example 3.1.4.

Let

$$A = \begin{pmatrix} 2 & 4 & 6 \\ -4 & 5 & 6 \\ 7 & -8 & 9 \end{pmatrix}.$$

1. Find $M_{21}, M_{22}, M_{23}, A_{21}, A_{22}, A_{23}$ and compute $|A|$ using the cofactors.
2. Find $M_{31}, M_{32}, M_{33}, A_{31}, A_{32}, A_{33}$ and compute $|A|$ using the cofactors.

Solution

1.
$$M_{21} = \begin{vmatrix} 4 & 6 \\ -8 & 9 \end{vmatrix} = 36 + 48 = 84,\ A_{21} = (-1)^{2+1} M_{21} = -M_{21} = -84.$$

$$M_{22} = \begin{vmatrix} 2 & 6 \\ 7 & 9 \end{vmatrix} = 18 - 42 = -24,\ A_{22} = (-1)^{2+2} M_{22} = M_{22} = -24.$$

$$M_{23} = \begin{vmatrix} 2 & 4 \\ 7 & 8 \end{vmatrix} = -16 - 28 = -44,\ A_{23} = (-1)^{2+3} M_{23} = -M_{23} = 44.$$

By **Theorem 3.1.1**, we have
$|A| = a_{21}A_{21} + a_{22}A_{22} + a_{23}A_{23} = (-4)(-84) + (5)(-24) + (6)(44) = 480.$

2.

$$M_{31} = \begin{vmatrix} 4 & 6 \\ 5 & 6 \end{vmatrix} = 24 - 30 = -6,\ A_{31} = (-1)^{3+1} M_{31} = M_{31} = -6.$$

$$M_{32} = \begin{vmatrix} 2 & 6 \\ -4 & 6 \end{vmatrix} = 12 + 24 = 36,\ A_{32} = (-1)^{3+2} M_{32} = -M_{32} = -36.$$

$$M_{33} = \begin{vmatrix} 2 & 4 \\ -4 & 5 \end{vmatrix} = 10 + 16 = 26,\ A_{33} = (-1)^{3+3} M_{33} = M_{33} = 26.$$

By **Theorem 3.1.1**, we have

$$\left| A \right| = a_{31}A_{31} + a_{32}A_{32} + a_{33}A_{33} = (7)(-6) + (-8)(-36) + (9)(26) = 480.$$

It is easy to show that the determinant of a 2×2 matrix is equal to the determinant of its transpose.

(3.1.11)

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = \begin{vmatrix} a & b \\ c & d \end{vmatrix}^T = \begin{vmatrix} a & c \\ b & d \end{vmatrix} = ad - bc.$$

The following result shows that it is true for 3×3 matrices.

Theorem 3.1.2.

Let A be the same in **(3.1.1)** *. Show* $|A| = \left|A^T\right|$.

Proof

Using the method stated in **(3.1.3)** , we compute $\left|A^T\right|$ as follows.

(3.1.12)

$$|A^T| = \begin{vmatrix} a_{11} & a_{21} & a_{31} \\ a_{12} & a_{22} & a_{32} \\ a_{13} & a_{23} & a_{33} \end{vmatrix}\begin{matrix} a_{11} & a_{21} \\ a_{12} & a_{22} \\ a_{13} & a_{23} \end{matrix} = (a_{11}a_{22}a_{33} + a_{21}a_{32}a_{13} + a_{31}a_{12}a_{23}) - (a_{31}a_{22}a_{13} + a_{11}a_{32}a_{23} + a_{21}a_{12}a_{33}).$$

Comparing the right sides of **(3.1.3)** and **(3.1.12)** shows $\left|A\right| = \left|A^T\right|$.

It is easy to see that the expansion of cofactors of the determinant $|A|$ corresponding to the *i*th row of *A* is the same as the expansion of minors or cofactors of the determinant $\left|A^T\right|$ corresponding to the *i*th column of A^T. For example, the expansion of cofactors of the determinant $|A|$ corresponding to the first row of *A* is :

$$|A| = a_{11}\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12}\begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13}\begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix}.$$

and the expansion of cofactors of the determinant $\left|A^T\right|$ corresponding to the first row of A^T is

$$|A^T| = a_{11}\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12}\begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13}\begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix}$$

By **(3.1.11)**, we see that the the corresponding minors equal.

The following result shows that the determinants of upper or lower triangular matrices equal the products of entries on the main diagonals.

Theorem 3.1.3.

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ 0 & a_{22} & a_{23} \\ 0 & 0 & a_{33} \end{vmatrix} = \begin{vmatrix} a_{11} & 0 & 0 \\ a_{21} & a_{22} & 0 \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11}a_{22}a_{33} = a_{11}a_{22}a_{33}.$$

Proof

We only prove the second equality because the first equality follows from **Theorem 3.1.2**. By the expansion of cofactors, we have

$$\begin{vmatrix} a_{11} & 0 & 0 \\ a_{21} & a_{22} & 0 \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11}A_{11} + 0A_{12} + 0A_{13} = a_{11}\begin{vmatrix} a_{22} & 0 \\ a_{23} & a_{33} \end{vmatrix} = a_{11}a_{22}a_{33}.$$

Example 3.1.5.

Compute each of the following determinants.

$$|A| = \begin{vmatrix} 3 & -8 & 9 \\ 0 & -2 & 4 \\ 0 & 0 & 4 \end{vmatrix}, \quad |B| = \begin{vmatrix} 6 & 0 & 0 \\ 3 & -1 & 0 \\ 1 & 0 & 5 \end{vmatrix}$$

Solution

By **Theorem 3.1.3** , $|A| = 3(-2)(4) = -24$; $|B| = 6(-1)(5) = -30$.

Exercises

1. Let $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, where $a, b, c, d \in \mathbb{R}$. Show that

$$\left|\lambda I - A\right| = \lambda^2 - \operatorname{tr}(A)\lambda + \left|A\right|$$

for each $\lambda \in \mathbb{R}$.

2. Let $A_1 = \begin{pmatrix} 2 & 4 \\ 3 & -2 \end{pmatrix}$, $A_2 = \begin{pmatrix} 1 & 2 \\ -1 & 4 \end{pmatrix}$, $A_3 = \begin{pmatrix} -1 & 1 \\ -1 & -3 \end{pmatrix}$ For each $i = 1, 2, 3,$ find all $\lambda \in \mathbb{R}$ such that $\left|\lambda I - A_i\right| = 0.$

3. Use **(3.1.3)** and **(3.1.4)** to calculate each of the following determinants.

$$|A| = \begin{vmatrix} 1 & 2 & 3 \\ -2 & 1 & 2 \\ 3 & -1 & 4 \end{vmatrix} \quad |B| = \begin{vmatrix} 2 & -2 & 4 \\ 4 & 3 & 1 \\ 0 & 1 & 2 \end{vmatrix} \quad |C| = \begin{vmatrix} 0 & -1 & 3 \\ 4 & 1 & 2 \\ 0 & 0 & 1 \end{vmatrix}$$

4. Compute the determinant $|A|$ by using its expansion of cofactors corresponding to each of its rows, where

$$|A| = \begin{vmatrix} 1 & 2 & 3 \\ -2 & 1 & 2 \\ 3 & -1 & 4 \end{vmatrix}.$$

5. Compute each of the following determinants.

$$|A| = \begin{vmatrix} 2 & 3 & -4 \\ 0 & 6 & 0 \\ 0 & 0 & 5 \end{vmatrix} \quad |B| = \begin{vmatrix} -2 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 3 & 8 \end{vmatrix}$$

3.2 Determinants of $n \times n$ matrices

Let *A* be an $n \times n$ square matrix given in **(2.3.1)** , that is,

(3.2.1)

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}.$$

Following **Definition 3.1.1** , we introduce the following definition.

Definition 3.2.1.

Let M_{ij} be the determinant of the $(n-1) \times (n-1)$ matrix obtained from *A* by deleting the *i*th row and the *j*th column of *A*. M_{ij} is called the minor of a_{ij} or the *ij*th minor of *A*.

(3.2.2)

$$A_{ij} = (-1)^{i+j} M_{ij}$$

is called the cofactor of a_{ij} or the *ij*th factor of *A*.

Example 3.2.1.

Let

$$A = \begin{pmatrix} 1 & -3 & 5 & 6 \\ 2 & 4 & 0 & 3 \\ 1 & 5 & 9 & -2 \\ 4 & 0 & 2 & 7 \end{pmatrix}.$$

Find M_{32}, M_{24}, A_{32}, and A_{24}.

Solution

$$M_{32} = \begin{vmatrix} 1 & 5 & 6 \\ 2 & 0 & 3 \\ 4 & 2 & 7 \end{vmatrix} = 8, \quad A_{32} = (-1)^{3+2} M_{32} = -M_{32} = -8,$$

$$M_{24} = \begin{vmatrix} 1 & -3 & 5 \\ 1 & 5 & 9 \\ 4 & 0 & 2 \end{vmatrix} = -192, \quad A_{24} = (-1)^{2+4} M_{24} = M_{24} = -192.$$

Definition 3.2.2.

The determinant of *A* is defined by

(3.2.3)

$$|A| = \sum_{k=1}^{n} a_{1k} A_{1k} = a_{11} A_{11} + a_{12} A_{12} + \cdots + a_{1n} A_{1n}.$$

The expression on the right side of **(3.2.3)** is called an expansion of cofactors of the determinant $|A|$ via the first row of *A*.

In **Definition 3.2.2** , we use the cofactors of the first row of *A* to define the determinant $|A|$. Like **Theorems 3.1.1** and 3.1.12, we can use the expansion of cofactors of each row or each column of *A* to compute the determinant $|A|$, but the proof of the result is difficult for $n \geq 4.$ Hence, we state the result without proof.

Theorem 3.2.1.

For each $i,\ j \in I_n,$

$$|A| = \sum_{k=1}^{n} a_{ik} A_{ik} = \sum_{k=1}^{n} a_{kj} A_{kj} = |A^T|.$$

By **Theorem 3.2.1** , we can choose any row of *A* together with its cofactors to compute the determinant of *A*. Hence, the smart choice would be to choose the row or column that contains the most zero entries.

Theorem 3.2.2.

If A has a row of zeros or a column of zeros, then $|A| = 0.$

Proof

Assume that all entries of the *i*th row are zero, that is,

$$a_{i1} = a_{i2} = \cdots = a_{in} = 0.$$

Using the expansion of cofactors of the *i*th row and **Theorem 3.2.1** , we have

$$|A| = a_{i1}A_{i1} + a_{i2}A_{i2} + \cdots + a_{in}A_{in} = 0A_{i1} + 0A_{i2} + \cdots + 0A_{in} = 0.$$

If *A* has a column of zeros, then by **Theorem 3.2.1** , $|A| = 0.$

Example 3.2.2.

Evaluate each of the following determinants.

$$|A| = \begin{vmatrix} 2 & 3 & 5 \\ 0 & 0 & 0 \\ 1 & -2 & 4 \end{vmatrix} \quad |B| = \begin{vmatrix} 0 & 2 & 3 & 4 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 3 & 1 & 2 & 0 \end{vmatrix} \quad |C| = \begin{vmatrix} 1 & 3 & 5 & 2 \\ 0 & -1 & 3 & 4 \\ 2 & 1 & 9 & 6 \\ 3 & 2 & 4 & 8 \end{vmatrix}$$

Solution

Because the matrix *A* contains a zero row, $|A| = 0$. We use the expansion of cofactors of row 3 of *B* to compute $|B|$.

$$|B| = 0A_{31} + 0A_{32} + A_{33} + 0A_{34} = A_{33} = \begin{vmatrix} 0 & 2 & 4 \\ 0 & 1 & 1 \\ 3 & 1 & 0 \end{vmatrix} = -6.$$

We use the expansion of cofactors of row 2 of *C* to compute $|C|$.

$$\begin{aligned} |C| &= 0 + (-1)A_{22} + 3A_{23} + 4A_{24} = M_{22} - 3M_{23} + 4M_{24} \\ &= \begin{vmatrix} 1 & 5 & 2 \\ 2 & 9 & 6 \\ 3 & 4 & 8 \end{vmatrix} - 3 \begin{vmatrix} 1 & 3 & 2 \\ 2 & 1 & 6 \\ 3 & 2 & 8 \end{vmatrix} + 4 \begin{vmatrix} 1 & 3 & 5 \\ 2 & 1 & 9 \\ 3 & 2 & 4 \end{vmatrix} = 160. \end{aligned}$$

Similar to **Theorem 3.1.3** , by **Theorems 3.2.3** and **3.2.1** , we obtain the following result on the determinant of a lower or upper triangular matrix.

Theorem 3.2.3.

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ 0 & a_{22} & a_{23} & \cdots & 0 \\ 0 & 0 & a_{33} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & a_{nn} \end{vmatrix} = \begin{vmatrix} a_{11} & 0 & 0 & \cdots & 0 \\ a_{21} & a_{22} & 0 & \cdots & 0 \\ a_{31} & a_{32} & a_{33} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \cdots & a_{nn} \end{vmatrix} = a_{11}a_{22}\cdots a_{nn}.$$

Proof

Let A be the lower triangular matrix. Using **Theorem 3.2.1** repeatedly, we get

$$|A| = a_{11}\begin{vmatrix} a_{22} & 0 & \cdots & 0 \\ a_{32} & a_{33} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ a_{n2} & a_{n3} & \cdots & a_{nn} \end{vmatrix} = a_{11}a_{22}\begin{vmatrix} a_{33} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ a_{n3} & \cdots & a_{nn} \end{vmatrix} = a_{11}a_{22}\cdots a_{nn}.$$

The second result follows from **Theorem 3.2.1** and the above result.

By **Theorem 3.2.3**, the determinant of the identity matrix *In* is 1.

Example 3.2.3.

Evaluate

$$|A| = \begin{vmatrix} 5 & 0 & 0 & 0 \\ 2 & 3 & 0 & 0 \\ 1 & 6 & 7 & 0 \\ 0 & 0 & 2 & -1 \end{vmatrix} \quad |B| = \begin{vmatrix} 2 & 1 & 7 \\ 0 & 2 & -5 \\ 0 & 0 & 1 \end{vmatrix} \quad |C| = \begin{vmatrix} -2 & 3 & 0 & 1 \\ 0 & 0 & 2 & 4 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 2 \end{vmatrix}$$

Solution

By **Theorem 3.2.3** , we have

$$|A| = \begin{vmatrix} 5 & 0 & 0 & 0 \\ 2 & 3 & 0 & 0 \\ 1 & 6 & 7 & 0 \\ 0 & 0 & 2 & 1 \end{vmatrix} = 5 \times 3 \times 7 \times (-1)1 = 105.$$

$$|B| = \begin{vmatrix} 2 & 1 & 7 \\ 0 & 2 & -5 \\ 0 & 0 & 1 \end{vmatrix} = 2 \times 2 \times 1 = 4.$$

$$|C| = \begin{vmatrix} -2 & 3 & 0 & 1 \\ 0 & 0 & 2 & 4 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 2 \end{vmatrix} = (-2) \times 0 \times 1 \times 2 = 0.$$

Exercises

1. For each of the following matrices, find $M_{32},\ M_{24},\ A_{32},\ A_{24},\ M_{41},\ M_{43},\ A_{41},\ A_{43}$

$$A = \begin{pmatrix} 1 & -1 & 2 & 3 \\ 2 & 2 & 0 & 3 \\ 1 & 2 & 3 & -1 \\ -3 & 0 & 1 & 6 \end{pmatrix} \quad B = \begin{pmatrix} 1 & 0 & 2 & -1 \\ 2 & 0 & 0 & 1 \\ 1 & 0 & 1 & -1 \\ -1 & 0 & 1 & 2 \end{pmatrix}.$$

2. Compute each of the following determinants by using its expansion of cofactors of each row.

$$|A| = \begin{vmatrix} 1 & 0 & 2 & -1 \\ 2 & 0 & 0 & 1 \\ 1 & 0 & 1 & -1 \\ -1 & 0 & 1 & 3 \end{vmatrix} \qquad |B| = \begin{vmatrix} 1 & -1 & 2 & 3 \\ 2 & 2 & 0 & 3 \\ 1 & 2 & 3 & -1 \\ -3 & 0 & 1 & 6 \end{vmatrix}.$$

3. Evaluate $|A|$ by using the expansion of cofactors of a suitable row, where

$$|A| = \begin{vmatrix} 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 1 & 1 & -1 & 0 \end{vmatrix}.$$

4. Evaluate each of the following determinants by inspection.

$$|A| = \begin{vmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 6 \end{vmatrix} \quad |B| = \begin{vmatrix} 2 & 0 & 0 \\ 2 & -2 & 0 \\ 0 & 1 & 2 \end{vmatrix} \quad |C| = \begin{vmatrix} 1 & 2 & 3 \\ 0 & 1 & 2 \\ 0 & 0 & -1 \end{vmatrix}$$

$$|D| = \begin{vmatrix} 1 & -1 & 3 & 5 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 2 & 7 \\ 5 & 6 & 3 & -4 & 6 \\ 2 & -2 & 6 & 10 & 0 \end{vmatrix} \quad |E| = \begin{vmatrix} 1 & -1 & 3 & 5 & 0 \\ 0 & 3 & 0 & 0 & 0 \\ 0 & 0 & -1 & 2 & 7 \\ 0 & 0 & 0 & -4 & 6 \\ 0 & 0 & 0 & 0 & 2 \end{vmatrix}$$

$$|F| = \begin{vmatrix} 1 & 0 & 0 & 0 \\ -2 & 2 & 0 & 0 \\ -1 & 6 & -6 & 0 \\ 4 & 0 & 2 & -1 \end{vmatrix} \quad |G| = \begin{vmatrix} -2 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 2 \end{vmatrix}$$

3.3 Evaluating determinants by row operations

By **Theorem 3.2.3** , we see that the determinant of an upper triangular matrix can be easily calculated. Hence, we can use row operations to change an $n \times n$ matrix *A* to an upper triangular matrix *B* and then evaluate $|A|$ by calculating $|B|$. One can show that there exists a constant $k \in \mathbb{R}$ R such that

(3.3.1)

$$|A| = k|B|.$$

How can we find the number *k*? To do that, we need to know the relations between $|A|$ and $|B|$ if *B* is obtained from *A* by a single row operation.

Theorem 3.3.1.

Let A be an $n \times n$ matrix and c $\in \mathbb{R}$ with c $\neq 0$. Then the following assertions hold.

1. *If* $A \xrightarrow{R_i\left(\frac{1}{c}\right)} B$, *then* $|A| = c|B|$.
2. *If* $A \xrightarrow{R_{i,j}} B$, *then* $|A| = -|B|$.
3. *If* $A \xrightarrow{R_i(c)+R_j} B$, *them* $|A| = |B|$.

Proof

1. Without loss of generalization, we assume that *A* and *B* are of the following forms:

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & \vdots & \vdots \\ ca_{i1} & ca_{i2} & \cdots & ca_{in} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix} \text{ and } B = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{i1} & a_{i2} & \cdots & a_{in} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \vdots & a_{nn} \end{pmatrix}.$$

Then $A \xrightarrow{R_i\left(\frac{1}{c}\right)} B$ and $A_{ik} = B_{ik}$ for each $k \in I_n$ because the *j*th rows of *A* and *B* are the same for $j \neq i$. By **Theorem 3.2.1** , we have

(3.3.2)

$$|A| = \sum_{k=1}^{n} (ca_{ik}) A_{ik} = c \sum_{k=1}^{n} (a_{ik}) A_{ik} = c \sum_{k=1}^{n} (a_{ik}) B_{ik} = c|B|.$$

2. We first prove that the result holds when we interchange two successive rows. Let

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{i1} & a_{12} & \cdots & a_{in} \\ a(i+1)1 & a(i+1)2 & \cdots & a(i+1)n \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}, \ B = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & \vdots & \vdots \\ a(i+1)1 & a(i+1)2 & \cdots & a(i+1)n \\ a_{i1} & a_{i2} & \cdots & a_{in} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}.$$

Then $A \xrightarrow{R_{i,\,(i+1)}} B$, that is, we get *B* by interchanging the *i*th row and the $(i+1)\mathrm{th}$ row of *A*. By **Theorem 3.2.1** , we expand by cofactors via the *i*th row of *A* and the $(i+1)\mathrm{th}$ row of *B* and get

(3.3.3)

$$|A| = \sum_{k=1}^{n} a_{ik} A_{ik} \ \text{ and } |B| = \sum\nolimits_{k=1}^{n} a_{ik} B_{(i+1)k},$$

where A_{ik} and $B_{(i+1)k}$ are the cofactors of *A* and *B*, respectively. We note that the $(i+1)\text{th}$ row of *B* is $(a_{i1} a_{i2} \ \cdots \ a_{in})$. We denote by M_{ik} the minors of *A* and by $U_{(i+1)k}$ the minors of *B*. Then it is easy to see that $M_{ik} = U_{(i+1)k}$ for each $k \in I_n$. Then by **(3.2.2)**, we have

$$B_{(i+1)k} = (-1)^{i+1+k} U_{(i+1)k} = -(-1)^{i+k} M_{ik} = -A_{ik}.$$

This, together with **(3.3.3)**, implies

$$|A| = \sum_{k=1}^{n} a_{ik} A_{ik} = -\sum_{k=1}^{n} a_{ik}(-A_{ik}) = -\sum_{k=1}^{n} a_{ik} B_{(i+1)k} = -|B|.$$

Now we assume

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{i1} & a_{i2} & \cdots & a_{in} \\ \vdots & \vdots & \vdots & \vdots \\ a_{j1} & a_{j2} & \cdots & a_{jn} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix} \text{ and } B = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & \vdots & \vdots \\ a(i-1)1 & a(i-1)2 & \cdots & a(i-1)n \\ a_{j1} & a_{j2} & \cdots & a_{jn} \\ a(i+1)1 & a(i+1)2 & \ldots & a(i+1)n \\ \vdots & \vdots & \vdots & \vdots \\ a(j-1)1 & a(j-1)2 & \cdots & a(j-1)n \\ a_{i1} & a_{12} & \cdots & a_{in} \\ a(j+1)1 & a(j+1)2 & \cdots & a(j+1)n \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}.$$

$A \xrightarrow{R_{i,j}} B$. We can get *B* from *A* by interchanging two successive rows several times by the following ways:

i. Move the *i*th row vector: $(a_{i1}a_{i2}\cdots a_{in})$ to the $(i+1)\mathrm{th}$ row by interchanging the *i*th row and the $(i+1)\mathrm{th}$ row of *A*, then move the vector $(a_{i1}a_{i2}\cdots a_{in})$ from the $i+1\mathrm{th}$ row to the $(i+2)\mathrm{th}$ row by interchanging $(a_{i1}a_{i2}\cdots a_{in})$ and $(a_{(i+2)1}a_{(i+2)2}\cdots a_{(i+2)n})$. By $j-i$ times, we move $(a_{i1}a_{12}\cdots a_{in})$ to the *j*th row. The resulting matrix is

$$C := \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{(i+1)1} & a_{(i-1)2} & \cdots & a_{(i-1)n} \\ a_{(i+1)1} & a_{(i+1)2} & \cdots & a_{(i+1)n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{(j-i)1} & a_{(j-i)2} & \cdots & a_{(j-i)n} \\ a_{j1} & a_{j2} & \cdots & a_{jn} \\ a_{i1} & a_{i2} & \cdots & a_{in} \\ a_{(j+1)1} & a_{(j+1)2} & \cdots & a_{(j+i)n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}$$

and

(3.3.4)

$$\left|A\right| = (-1)^{j-i}\left|C\right|.$$

Note that the vector $(a_{j1}a_{j2}\cdots a_{jn})$ is in the $(j-1)$ row of *C*.

ii. We move $(a_{j1}a_{j2}\cdots a_{jn})$ back to the $(i-2)\mathrm{th}$ row by interchanging the $(j-1)\mathrm{th}$ and $(i-2)\mathrm{th}$ rows of *C*, that is, interchanging the two row vectors $(a_{j1}a_{j2}\cdots a_{jn})$ and $(a_{(j-1)1}a_{(j-1)2}\cdots a_{(j-1)n})$ of *C* Continuing the process, we can move $(a_{j1}a_{j2}\cdots a_{jn})$ back to the *i*th row by $[(j-i)-i]$ times. The resulting matrix exactly is *B*. Moreover, we have

$$\left|C\right| = (-1)^{[(j-1)-i]}\left|B\right| = (-1)^{(j-j)-1}\left|B\right|.$$

This, together with **(3.3.4)**, implies
$|A| = (-1)^{j-i}|C| = |A| = (-1)^{j-i}[(-1)^{(j-j)-1}|B|] = (-1)^{2(j-i)-1}|B| = -|B|$
because $2(j-i)-1$ is an odd number and $(-1)^{2(j-i)-1} = -1.$

3. We consider the following three cases.
   i. We first prove if an $n \times n$ matrix *D* has two equal rows, then $|D| = 0.$ We assume that (3.3.5)

$$D = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & \vdots & \vdots \\ a(i-1)1 & a(i-1)2 & \cdots & a(i-1)n \\ a_{i1} & a_{i2} & \cdots & a_{in} \\ a_{(i+1)1} & a_{(i+1)2} & \cdots & a_{(i+1)n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{(j-1)1} & a_{(j-1)2} & \cdots & a_{(j-1)n} \\ a_{i1} & a_{i2} & \cdots & a_{in} \\ a_{(j+1)1} & a_{(j+1)2} & \cdots & a_{(j+1)n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}.$$

   Note that after interchanging the *i*th and *j*th rows of *D*, the resulting matrix still is *D*. Hence, $|D| = -|D|$. This implies $|D| = 0.$
   ii. Expanding by cofactors of *D* via the *j*th row of *D* given in **(3.3.5)** and using the result (*i*), we obtain

(3.3.6)

$$|D| = \sum_{k=1}^{n} a_{ik} D_{jk} = 0,$$

where D_{jk} is the cofactor of *D* for $k \in I_n$.

iii. Now, we assume that

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{i1} & a_{i2} & \cdots & a_{in} \\ \vdots & \vdots & \vdots & \vdots \\ a_{(j-1)1} & a_{(j-1)2} & \cdots & a_{(j-1)n} \\ a_{j1} & a_{j2} & \cdots & a_{jn} \\ a_{(j+1)1} & a_{(j+1)2} & \cdots & a_{(j+1)n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}$$

and

$$B = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{i1} & a_{i2} & \cdots & a_{in} \\ \vdots & \vdots & \vdots & \vdots \\ a_{(j-1)1} & a_{(j-1)2} & \cdots & a_{(j-1)n} \\ ca_{i1} + a_{j1} & ca_{i2} + a_{i2} & \cdots & ca_{in} + a_{jn} \\ a_{(j+1)1} & a_{(j+1)2} & \cdots & a_{(j+1)n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}.$$

Then $A \xrightarrow{R_{i(c)}+R_j} B$. Let A_{ik} and B_{ik} be the *j*th row cofactors of *A* and *B*, respectively. Then by **(3.3.5)** , (3.3.7), and (3.3.8), we see that

(3.3.9)

$$A_{jk} = B_{jk} = D_{jk}$$

for $k \in I_n$.

Expanding by cofactors of *B* via the *j*th row of *B*, we have by **(3.3.9)** and **(3.3.6)** and **Theorem 3.2.1** ,

$$\begin{aligned} |B| &= \sum_{k=1}^{n}(ca_{ik} + a_{jk})B_{jk} = \sum_{k=1}^{n}(ca_{ik} + a_{jk})A_{jk} = c\sum_{k=1}^{n} a_{ik}D_{jk} + \sum_{k=1}^{n} a_{jk}A_{jk} \\ &= 0 + \sum_{k=1}^{n} a_{jk}A_{jk} = |A|. \end{aligned}$$

By **Theorem 3.3.1**, we see that there exists a number *k* such that $|A| = k|B|$, where *B* is obtained from *A* by several row operations, and *k* can be obtained from the row operations used in **Theorem 3.3.1** (*i*) and (*ii*).

Before we use **Theorem 3.3.1** to compute determinants, we try to understand and digest each result given in **Theorem 3.3.1** and use it to derive some results.

By **Theorem 3.3.1** (1), or **(3.3.2)**, we see that

$$\begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & \vdots & \vdots \\ ca_{i1} & ca_{i2} & \cdots & ca_{in} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix} = c \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{i1} & a_{i2} & \cdots & a_{in} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix}.$$

Hence, we can take the common factor of one row out of the determinant. Similarly, because $|A^T| = |A|$, we can take a common factor of one column out of the determinant.

Example 3.3.1.

Compute each of the following determinants.

$$|A| = \begin{vmatrix} 3 & -3 & 6 \\ 12 & 4 & 16 \\ 0 & -2 & 5 \end{vmatrix} \quad |B| = \begin{vmatrix} -1 & -1 & 12 \\ 2 & 4 & 6 \\ 0 & 1 & 6 \end{vmatrix}.$$

Solution

$$|A| = \begin{vmatrix} 3(1) & 3(-1) & 3(2) \\ (4)(3) & (4)(1) & (4)(4) \\ 0 & -2 & 5 \end{vmatrix} = 3\begin{vmatrix} 1 & -1 & 2 \\ (4)(3) & (4)(1) & (4)(4) \\ 0 & -2 & 5 \end{vmatrix}$$

$$= 3(4)\begin{vmatrix} 1 & -1 & 2 \\ 3 & 1 & 4 \\ 0 & -2 & 5 \end{vmatrix} = (12)(16) = 192.$$

$$|B| = \begin{vmatrix} -1 & -1 & 12 \\ (2)(1) & (2)(2) & (2)(3) \\ 0 & 1 & 6 \end{vmatrix} = 2\begin{vmatrix} -1 & -1 & 12 \\ 1 & 2 & 3 \\ 0 & 1 & 6 \end{vmatrix}$$

$$= \begin{vmatrix} -1 & -1 & (3)(4) \\ 1 & 2 & (3)(1) \\ 0 & 1 & (3)(2) \end{vmatrix} = (2)(3)\begin{vmatrix} -1 & -1 & 4 \\ 1 & 2 & 1 \\ 0 & 1 & 2 \end{vmatrix} = (6)(3) = 18.$$

The following result gives the relation between $|cA|$ and $|A|$.

Theorem 3.3.2.

Let A be an $n \times n$ matrix and c $\in \mathbb{R}$. Then

$$|cA| = c^n\,|A|\,.$$

Proof

The result holds if $c = 0$. We assume that $c \neq 0$. Repeating **(3.3.2)** implies

$$|cA| = \begin{vmatrix} ca_{11} & ca_{12} & \cdots & ca_{1n} \\ ca_{21} & ca_{22} & \cdots & ca_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ ca_{n1} & ca_{n2} & \cdots & ca_{nn} \end{vmatrix} = c \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ ca_{21} & ca_{22} & \cdots & ca_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ ca_{n1} & ca_{n2} & \cdots & ca_{nn} \end{vmatrix} = \cdots$$

$$= c^n \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix} = c^n |A|.$$

Example 3.3.2.

Let *A* be a 3×3 matrix. Assume that $|A| = 6.$ Compute each of the following determinants.

$$|2A| \text{ and } \left|(3A)^T\right|.$$

Solution

By **Theorem 3.3.2** ,

$$|2A| = 2^3 |A| = 8 \times 6 = 48.$$

By **Theorems 3.2.1** and **3.3.2** ,

$$\left|(3A)^T\right| = |3A| = 3^3 |A| = 27(6) = 162.$$

Theorem 3.3.1 (2) means that when interchanging any two different rows of a determinant, the determinant changes its sign.

Example 3.3.3.

Evaluate

$$|A| = \begin{vmatrix} 0 & 0 & 1 \\ 0 & 2 & 3 \\ 1 & -2 & 5 \end{vmatrix}.$$

Solution

$$|A| = \begin{vmatrix} 0 & 0 & 1 \\ 0 & 2 & 3 \\ 1 & -2 & 5 \end{vmatrix} \overset{R_{1,3}}{=\!=} - \begin{vmatrix} 1 & -2 & 5 \\ 0 & 2 & 3 \\ 0 & 0 & 1 \end{vmatrix} = -2.$$

Corollary 3.3.1.

i. *If A has two equal rows, then* $|A| = 0.$
ii. *If A has two equal columns, then* $|A| = 0.$

Proof

The result (*i*) follows from the result (*i*) in **Theorem 3.3.1** (3) and the result (*ii*) follows from **Theorem 3.2.1** and the result (*i*).

Example 3.3.4.

Evaluate

$$A = \begin{vmatrix} 1 & 3 & -2 & 4 \\ 2 & 7 & -4 & 8 \\ 3 & 9 & 1 & 5 \\ 1 & 3 & -2 & 4 \end{vmatrix} \quad B = \begin{vmatrix} -1 & 3 & 2 & 0 & 3 \\ 3 & 1 & 0 & 2 & 1 \\ 3 & 9 & 1 & 5 & 9 \\ 2 & -3 & 1 & 2 & -3 \\ 1 & 3 & 0 & 4 & 3 \end{vmatrix}.$$

Solution

Because A has two equal rows, rows 1 and 4, $|A| = 0.$ Similarly, because B has two equal columns, columns 2 and 5, $|B| = 0.$

Corollary 3.3.1 can be generalized to two proportional rows or columns.

Theorem 3.3.3.

1. *If A has two proportional rows, then* $|A| = 0.$
2. *If A has two proportional columns, then* $|A| = 0.$

Proof

(1) Assume that A is an $n \times n$ matrix and the ith and jth rows are proportional. Without a loss of generalization, we may assume that $1 \leq i < j \leq n$ and $R_i(c) = R_j$ for some $c \neq 0.$ By **Theorem 3.3.1** (1) and **Corollary 3.3.1**, we have

$$|A| = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{i1} & a_{i2} & \cdots & a_{in} \\ \vdots & \vdots & \vdots & \vdots \\ ca_{i1} & ca_{i2} & \cdots & ca_{in} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix} = c \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{i1} & a_{i2} & \cdots & a_{in} \\ \vdots & \vdots & \vdots & \vdots \\ a_{i1} & a_{i2} & \cdots & a_{in} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix} = 0.$$

because the ith and jth rows of the last determinant are the same.

(2) The result follows from **Theorem 3.2.1** and the result (1).

Example 3.3.5.

Evaluate

$$|A| = \begin{vmatrix} 3 & 9 & 1 \\ 1 & 3 & -2 \\ 2 & 6 & -4 \end{vmatrix} \quad |B| = \begin{vmatrix} 1 & 3 & -2 & -2 \\ 3 & 9 & 1 & -6 \\ -2 & 6 & -1 & 4 \\ 1 & 1 & 4 & -2 \end{vmatrix}.$$

Solution

Because rows 2 and 3 of *A* are proportional, by **Theorem 3.3.3** (1), $|A| = 0.$ Because columns 1 and 4 of *B* are proportional, by **Theorem 3.3.3** (2), $|B| = 0.$

Theorem 3.3.1 (3) means that when adding the *i*th row multiplied by a nonzero constant *c* to the *j*th row, the determinant remains unchanged. Using **Theorem 3.3.1** (3) and **Theorem 3.2.2** , one can provide another proof of **Theorem 3.3.3** (1). Indeed, we use the row operation $R_i(-c) + R_j$ on *A*. Then the *j*th row of the resulting matrix denoted by *B* is a zero row. by **Theorem 3.2.2** , $|B| = 0$ and by **Theorem 3.3.1** (3), $|A| = |B| = 0.$

Example 3.3.6.

Compute

$$|A| = \begin{vmatrix} 2 & -3 & 5 \\ 1 & 7 & 2 \\ -4 & 6 & -10 \end{vmatrix}.$$

Solution

$$|A| = \begin{vmatrix} 2 & -3 & 5 \\ 1 & 7 & 2 \\ -4 & 6 & -10 \end{vmatrix} \underline{\underline{R_1(2)+R_3}} \begin{vmatrix} 2 & -3 & 5 \\ 1 & 7 & 2 \\ 0 & 0 & 0 \end{vmatrix} = 0.$$

Now, we use examples to show how to use row operations to calculate determinants. Before calculating the determinant of a matrix, it is best to check the following:

1. Whether the matrix contains a zero row or a zero column.
2. Whether the matrix contains two equal rows or two equal columns.
3. Whether the matrix contains two proportional rows or two proportional columns.
   If one of the above occurs, then the determinant is zero.
4. Whether we should use the transpose to compute the determinant. After checking the above four things, we would follow **Steps 1-3** given in **Section 2.4** for each round and apply **Theorem 3.3.1** .

Example 3.3.7.

Evaluate each of the following determinants.

$$|A| = \begin{vmatrix} 0 & 1 & 5 \\ 3 & -6 & 9 \\ 2 & 6 & 1 \end{vmatrix} |B| = \begin{vmatrix} 2 & 6 & 10 & 4 \\ 0 & -1 & 3 & 4 \\ 2 & 1 & 9 & 6 \\ 3 & 2 & 4 & 8 \end{vmatrix} |C| = \begin{vmatrix} 1 & 0 & 0 & 1 \\ -2 & 1 & 0 & -2 \\ 0 & 2 & 1 & 0 \\ 4 & 0 & 1 & -1 \end{vmatrix}.$$

Solution

$$|A| = \begin{vmatrix} 0 & 1 & 5 \\ 3 & -6 & 9 \\ 2 & 6 & 1 \end{vmatrix} \overset{R_{1,2}}{=\!=} - \begin{vmatrix} 3 & -6 & 9 \\ 0 & 1 & 5 \\ 2 & 6 & 1 \end{vmatrix} = -3 \begin{vmatrix} 1 & -2 & 3 \\ 0 & 1 & 5 \\ 2 & 6 & 1 \end{vmatrix} \overset{R_1(-2)+R_3}{=\!=\!=}$$

$$-3 \begin{vmatrix} 1 & -2 & 3 \\ 0 & 1 & 5 \\ 0 & 10 & -5 \end{vmatrix} \overset{R_2(-10)+R_3}{=\!=\!=} -3 \begin{vmatrix} 1 & -2 & 3 \\ 0 & 1 & 5 \\ 0 & 0 & -55 \end{vmatrix}$$

$$= (-3)(1)(1)(-55) = 165.$$

$$|B| = \begin{vmatrix} 2 & 6 & 10 & 4 \\ 0 & -1 & 3 & 4 \\ 2 & 1 & 9 & 6 \\ 3 & 2 & 4 & 8 \end{vmatrix} = 2 \begin{vmatrix} 1 & 3 & 5 & 2 \\ 0 & -1 & 3 & 4 \\ 2 & 1 & 9 & 6 \\ 3 & 2 & 4 & 8 \end{vmatrix} \xrightarrow[R_1(-3)+R_4]{R_1(-2)+R_3}$$

$$2 \begin{vmatrix} 1 & 3 & 5 & 2 \\ 0 & -1 & 3 & 4 \\ 0 & -5 & -1 & 2 \\ 0 & -7 & -11 & 2 \end{vmatrix} \xrightarrow[R_2(-7)+R_4]{R_2(-5)+R_3} \begin{vmatrix} 1 & 3 & 5 & 2 \\ 0 & -1 & 3 & 4 \\ 0 & 0 & -16 & -18 \\ 0 & 0 & -32 & -26 \end{vmatrix}$$

$$\overset{R_3(-2)+R_4}{=\!=\!=} 2 \begin{vmatrix} 1 & 3 & 5 & 2 \\ 0 & -1 & 3 & 4 \\ 0 & 0 & -16 & -18 \\ 0 & 0 & 0 & 10 \end{vmatrix} = (2)(1)(-1)(-16)(10) = 320.$$

$$|C| = \left|C^T\right| = \begin{vmatrix} 1 & -2 & 0 & 4 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & -2 & 0 & -1 \end{vmatrix} \overset{R_1(-1)+R_4}{=\!=\!=} \begin{vmatrix} 1 & -2 & 0 & 4 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & -5 \end{vmatrix} = -5.$$

By **Theorems 2.6.1** and **3.3.1**, we obtain

Corollary 3.3.2.

Let A be an $n \times n$ matrix and $c \in \mathbb{R}$ with $c \neq 0$. Then the following assertions hold.

i. $|E_i(c)A| = |E_i(c)|\,|A|.$
ii. $\left|E_{i(c)+j}A\right| = \left|E_{i(c)+j}\right|\,|A|.$
iii. $|E_{i,j}A| = |E_{i,j}|\,|A|.$

Proof

i. Let $B = E_i(c)A$. By **Theorem 2.6.1** (*i*), $A \xrightarrow{R_i(c)} B$ and $I_n \xrightarrow{R_i(c)} E_i(c)$. By **Theorem 3.3.1**,

$$|A| = \frac{1}{c}|B| \text{ and } |E_i(c)| = \frac{1}{c}|I_n| = \frac{1}{c}.$$

This implies $c = |E_i(c)|$ and

$$|E_i(c)A| = |B| = c\,|A| = |E_i(c)|\,|A|.$$

ii. Let $B = E_{i(c)+j}A$. By **Theorem 2.6.1** (*iii*), $A \xrightarrow{R_i(c)+R_j} B$ and $I_n \xrightarrow{R_i(c)+R_j} E_{i(c)+j}$.

By **Theorem 3.3.1**,

$$|A| = |B| \text{ and } \left|E_{i(c)+j}\right| = |I_n| = 1.$$

This implies

$$\left|E_{i(c)+j}A\right| = |B| = |A| = \left|E_{i(c)+j}\right| |A| .$$

iii. Let $B = E_{i,j}A$. By **Theorem 2.6.1** (*ii*), $A \xrightarrow{R_{i,j}} B$ and $I_n \xrightarrow{R_{i,j}} E_{i,j}$. By **Theorem 3.3.1** ,

$$|A| = -|B| \text{ and } |E_{i,j}| = -|I_n| = -1.$$

This implies

$$|E_{i,j}A| = -|B| = -|A| = |E_{i,j}| |A| .$$

By using **Corollary 3.3.2** repeatedly, it is easy to prove the following result, which gives the relation among the determinants of an $n \times n$ matrix, its row echelon matrix, and elementary matrices.

Theorem 3.3.4.

Let A and B be $n \times n$ matrices. Assume that there exist elementary matrices $E_1, E_2, \cdots, E_k$ such that $B = E_k E_{k-1} \cdots E_2 E_1 A$. Then

$$|B| = |E_k| |E_{k-1}| \cdots |E_2| |E_1| |A| .$$

By **Theorem 3.3.1** and $|I| = 1$, it is easy to see that the following result on the determinants of elementary matrices holds.

Theorem 3.3.5.

The determinant of every elementary matrix is not zero.

At the end of the section, we generalize the formula **(2.7.1)** of an inverse matrix from a 2×2 matrix to an $n \times n$ matrix for $n \geq 3$.

We introduce the concept of the adjoint matrix of an $n \times n$ matrix.

Definition 3.3.1.

Let *A* be an $n \times n$ matrix. The matrix

(3.3.10)

$$\begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{22} & \cdots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{n1} & A_{n2} & \cdots & A_{nn} \end{pmatrix}$$

is called the matrix of cofactors of *A*. The transpose of the matrix of cofactors of *A* is called the adjoint of *A*, denoted by adj(*A*)

By **Definition 3.3.1** , the adjoint of *A* is the following matrix:

(3.3.11)

$$\operatorname{adj}(A) = \begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{22} & \cdots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{n1} & A_{n2} & \cdots & A_{nn} \end{pmatrix}^T = \begin{pmatrix} A_{11} & A_{21} & \cdots & A_{n1} \\ A_{12} & A_{22} & \cdots & A_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ A_{1n} & A_{2n} & \cdots & A_{nn} \end{pmatrix}.$$

Example 3.3.8.

Find the matrix of cofactors of *A* and adj(*A*), where

$$A=\begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

Solution

Because $A_{11}=M_{11}=d,\ A_{12}=-M_{12}=-c,\ A_{21}=-M_{21}=-b,$ and $A_{22}=M_{22}=a,$ the matrix of cofactors of *A* is

$$\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}=\begin{pmatrix} d & -c \\ -b & a \end{pmatrix}$$

and

$$\mathrm{adj}\,(A)=\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}^T=\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

Note that **(2.7.1)** involves the adjoint matrix of *A* given in **Example 3.3.8**. By **Theorem 2.7.5** and **Example 3.3.8**, we see that

(3.3.12)

$$A\ \mathrm{adj(A)}=\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}=|A|\,I.$$

By **Theorem 3.2.1** and **Corollary 3.3.1**, we generalize **(3.3.12)** to $n\times n$ matrices.

Theorem 3.3.6.

Let A and adj(A) be the same as in ***(3.2.1)*** *and* ***(3.3.11)****, respectively. Then*

(3.3.13)

$$A\ \mathrm{adj}(A) = |A|\, I.$$

Proof

Let $i,\ j \in I_n$ and

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{i1} & a_{i2} & \cdots & a_{in} \\ \vdots & \vdots & \vdots & \vdots \\ a_{j1} & a_{j2} & \cdots & a_{jn} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}.$$

By **Theorem 3.2.1** , we expand $|A|$ via the *j*th row and obtain

(3.3.14)

$$|A| = \sum_{k=1}^{n} a_{jk} A_{jk}.$$

Replacing the *j*th row by the *i*th row of *A*, by **Corollary 3.3.1** (1) we obtain

(3.3.15)

$$\sum_{k=1}^{n} a_{ik} A_{jk} = 0 \qquad \text{for } i < j \text{ or } i > j.$$

Let

$$C=(c_{ij})=A\,\mathrm{adj}(A)$$
$$=\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & \cdots & \vdots \\ a_{i1} & a_{i2} & \cdots & a_{in} \\ \vdots & \vdots & \cdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}\begin{pmatrix} A_{11} & \cdots & A_{j1}\cdots & A_{n1} \\ A_{12} & \cdots & A_{j2}\cdots & A_{n2} \\ \vdots & \vdots & \vdots\vdots & \\ A_{1n} & \cdots & A_{jn}\cdots & A_{nn} \end{pmatrix}.$$

Then

$$c_{ij}=\sum_{k=1}^{n} a_{ik}A_{jk} \quad \text{for } i,\ j\in I_n.$$

By **(3.3.14)** and **(3.3.15)**, $c_{jj}=|A|$ if $i=j$ and $c_{ij}=0$ if $i\neq j$. Hence, $C=\mathrm{diag}(|A|),\ \cdots,|A|)=|A|I.$

By **Theorem 3.3.6**, we generalize **Theorem 2.7.5** to an $n\times n$ matrix.

Theorem 3.3.7.

Let A and adj(*A*) *be the same as in* **(3.2.1)** *and* **(3.3.11)**, *respectively. If* $|A|\neq 0,$ *then*

$$A^{-1}=\frac{1}{|A|}\mathrm{adj}(A).$$

Proof

Because $|A|\neq 0,$ by Theorem 3.3.13, we have

$$A\left(\frac{1}{|A|}\operatorname{adj}(A)\right)=\frac{1}{|A|}[A\operatorname{adj}(A)]=\frac{1}{|A|}(|A|I)=I.$$

The result follows from **Definition 2.7.1** .

Example 3.3.9.

Use **Theorem 3.3.7** to find the inverse of the following matrix

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

Solution

By **(3.1.9)** , we have

$$\begin{aligned}A_{11}&=\begin{vmatrix}1&1\\0&1\end{vmatrix}=1,\ A_{12}=-\begin{vmatrix}0&1\\1&1\end{vmatrix}=1,\ A_{13}=\begin{vmatrix}0&1\\1&0\end{vmatrix}=-1,\\A_{21}&=-\begin{vmatrix}1&1\\0&1\end{vmatrix}=-1,\ A_{22}=\begin{vmatrix}1&1\\1&1\end{vmatrix}=0,\ A_{23}=-\begin{vmatrix}1&1\\1&0\end{vmatrix}=1,\\A_{31}&=\begin{vmatrix}1&1\\1&1\end{vmatrix}=0,\ A_{32}=-\begin{vmatrix}1&1\\0&1\end{vmatrix}=-1,\ A_{33}=\begin{vmatrix}1&1\\0&1\end{vmatrix}=1.\end{aligned}$$

This, together with **(3.3.11)** , implies

$$\operatorname{adj}(A)=\begin{pmatrix}A_{11}&A_{21}&A_{31}\\A_{12}&A_{22}&A_{32}\\A_{13}&A_{23}&A_{33}\end{pmatrix}=\begin{pmatrix}1&-1&0\\1&0&-1\\-1&1&1\end{pmatrix}.$$

By computation, $|A| = 1.$ It follows from **Theorem 3.3.7** that

$$A^{-1} = \frac{1}{|A|}\operatorname{adj}(A) = \operatorname{adj}(A) = \begin{pmatrix} 1 & -1 & 0 \\ 1 & 0 & -1 \\ -1 & 1 & 1 \end{pmatrix}.$$

The inverse of *A* is the same as that of *B* given in **Example 2.7.6**, where a different method is used.

Exercises

1. Evaluate each of the following determinants by inspection.

$$|A| = \begin{vmatrix} 1 & 2 & -3 \\ 0 & 1 & 2 \\ -2 & -4 & 6 \end{vmatrix} \quad |B| = \begin{vmatrix} 2 & -2 & 4 \\ 2 & -2 & 4 \\ 0 & 1 & 2 \end{vmatrix} \quad |C| = \begin{vmatrix} 1 & 2 & 3 \\ -2 & 1 & 2 \\ 1 & -\frac{1}{2} & -1 \end{vmatrix}$$

2. Use row operations to evaluate each of the following determinants.

$$|A| = \begin{vmatrix} 2 & 1 & 4 \\ 1 & -2 & 4 \\ 0 & 2 & 4 \end{vmatrix} \; |B| = \begin{vmatrix} 0 & 2 & 3 \\ -2 & 1 & 2 \\ 3 & -1 & -1 \end{vmatrix} \; |C| = \begin{vmatrix} 1 & 2 & -3 & -2 \\ 2 & 1 & 2 & 0 \\ -2 & -3 & 1 & -2 \\ 1 & -1 & 2 & 0 \end{vmatrix}$$

$$|D| = \begin{vmatrix} 0 & 0 & 0 & 0 \\ -2 & 2 & 0 & 0 \\ 0 & 6 & -6 & 0 \\ 4 & 0 & 2 & -1 \end{vmatrix} \; |E| = \begin{vmatrix} -2 & 0 & 0 & 0 \\ 3 & 3 & 0 & 1 \\ 0 & 1 & -1 & 0 \\ -1 & 0 & 0 & 2 \end{vmatrix}$$

$$|F| = \begin{vmatrix} 0 & -1 & 3 & 5 & 0 \\ 1 & -1 & 2 & 4 & -1 \\ 0 & 0 & 1 & 2 & 7 \\ 1 & 1 & -2 & 3 & 1 \\ 2 & -2 & 6 & 10 & 0 \end{vmatrix} \; |G| = \begin{vmatrix} 1 & -1 & 3 & 5 & 0 \\ 1 & 2 & 0 & 0 & 0 \\ -2 & 0 & -1 & 2 & 1 \\ 3 & 0 & 0 & -6 & 6 \\ 2 & 0 & 0 & 0 & 1 \end{vmatrix}$$

3. Evaluate each of the following determinants by using its transpose.

$$|A| = \begin{vmatrix} 1 & 0 & 0 & 1 \\ 2 & 7 & 0 & 2 \\ 0 & 6 & 3 & 0 \\ 5 & 3 & 1 & 2 \end{vmatrix} \; |B| = \begin{vmatrix} 1 & 0 & 0 & 0 & 0 \\ 2 & 7 & 0 & 2 & 1 \\ 0 & 6 & 3 & 0 & 0 \\ 1 & 3 & 1 & 2 & -1 \\ 0 & 6 & 1 & 2 & 1 \end{vmatrix}.$$

4. Use **Theorem 3.3.7** to find the inverse of the following matrix

$$A = \begin{pmatrix} 0 & 2 & -5 \\ 1 & 5 & -5 \\ 1 & 0 & 8 \end{pmatrix}$$

3.4 Properties of determinants

A criterion that uses ranks of matrices to verify whether $n \times n$ matrices are invertible is **Theorem 2.7.6**, which states that an $n \times n$ matrix *A* is invertible if and only if $r(A) = n$. By **Theorem 2.7.5**, a 2×2 matrix *A* is invertible if and only if its determinant is not zero. The following result shows that this is true for any $n \times n$ matrix. This provides another criterion that uses determinants to verify whether $n \times n$ matrices are invertible.

Theorem 3.4.1.

Let A be an $n \times n$ matrix. Then $|A| \neq 0$ if and only if A is invertible.

Proof

Let *C* be the reduced row echelon matrix of *A*. By **Theorem 2.7.3**, there exist elementary matrices $E_1, E_2, \cdots, E_{k-1}, E_k$ such that

(3.4.1)

$$C = E_k E_{k-1} \cdots E_2 E_1 A.$$

By **Theorem 3.3.4**,

(3.4.2)

$$\left|C\right| = \left|E_k\right|\left|E_{k-1}\right| \cdots \left|E_2\right|\left|E_1\right|\left|A\right|.$$

Assume that $|A| \neq 0$. By **Theorem 3.3.5** and **(3.4.2)**, $|C| \neq 0$. Because C is the reduced row echelon matrix of A, C is an upper triangular matrix and all the entries on the main diagonal that are leading entries are not zero. This implies $r(C) = n$. By **(3.4.1)** and **Theorem 2.5.4**, $r(A) = r(C) = n$. By **Theorem 2.7.6**, A is invertible. Conversely, assume that A is invertible. By **Corollary 2.7.1** (i), $C = I_n$ and $|C| = 1$. It follows from **(3.4.2)** and **Theorem 3.3.5** that $|A| \neq 0$.

In **Example 2.7.5**, we use ranks of matrices, that is, **Theorem 2.7.6**, to verify whether square matrices are invertible. Now we revisit the two matrices and use determinants, that is, **Theorem 3.4.1**, to determine whether they are invertible.

Example 3.4.1.

For each of the following matrices, use its determinant to determine whether it is invertible.

$$A = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix} \qquad B = \begin{pmatrix} 1 & 1 & 1 \\ 0 & -1 & 0 \\ 1 & 0 & 1 \end{pmatrix}$$

Solution

$$\begin{aligned} |A| &= \begin{vmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{vmatrix} \Big\downarrow R_1(-1)+R_3 = \begin{vmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & -1 & 0 \end{vmatrix} \Big\downarrow R_2(1)+R_3 = \begin{vmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{vmatrix} \\ &= 1 \neq 0. \end{aligned}$$

By **Theorem 3.4.1**, A is invertible.

$$|B| = \begin{vmatrix} 1 & 1 & 1 \\ 0 & -1 & 0 \\ 1 & 0 & 1 \end{vmatrix} \Bigg\downarrow R_1(-1)+R_3 \begin{vmatrix} 1 & 1 & 1 \\ 0 & -1 & 0 \\ 0 & -1 & 0 \end{vmatrix} \Bigg\downarrow R_2(-1)+R_3$$

$$\begin{vmatrix} 1 & 1 & 1 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{vmatrix} = 0.$$

By **Theorem 3.4.1** , *A* is not invertible.

Theorem 3.4.2.

Let A and B be $n \times n$ matrices. Then

$$|AB| = |A||B|.$$

Proof

Assume that *C* is the reduced row echelon matrix of *A*. By **Theorem 2.7.3** , there exist elementary matrices $E_1,\ E_2,\ \cdots,\ E_{k-1},\ E_k$ such that

$$C = E_k E_{k-1} \cdots E_2 E_1 A.$$

It follows that

(3.4.3)

$$A = F_1 F_2 \cdots F_{k-1} F_k C,$$

where $F_i = E_i^{-1}$ is an elementary matrix for $i = 1, 2, \cdots, k$. By **Theorem 3.3.4** ,

(3.4.4)

$$\left|A\right| = \left|F_1\right|\left|F_2\right| \cdots \left|F_{k-1}\right|\left|F_k\right|\left|C\right|.$$

If *A* is invertible, then by **Corollary 2.7.1** (*i*), $C = I_n$ By **(3.4.4)**

$$\left|A\right| = \left|F_1\right|\left|F_2\right| \cdots \left|F_{k-1}\right|\left|F_k\right|.$$

By **(3.4.3)** , we obtain

$$AB = F_1 F_2 \cdots F_{k-1} F_k B$$

and by **Theorem 3.3.4** ,

$$\left|AB\right| = \left|F_1 F_2 \cdots F_{k-1} F_k B\right| = \left|F_1\right|\left|F_2\right| \cdots \left|F_{k-1}\right|\left|F_k\right|\left|B\right| = \left|A\right|\left|B\right|.$$

If *A* is not invertible, by **Corollary 2.7.1** (*ii*), the last row of *C* must be a zero row. By **(3.4.4)** , $|A| = 0$. By **(2.2.5)** , we see that the last row of *CB* is a zero row and $|CB| = 0$. It follows that

$$\left|AB\right| = \left|F_1 F_2 \cdots F_{k-1} F_k CB\right| = \left|F_1\right|\left|F_2\right| \cdots \left|F_{k-1}\right|\left|F_k\right|\left|CB\right| = 0.$$

Hence, $|AB| = 0 = |A||B|$.

Example 3.4.2.

ath: 100%

erify that $|AB| = |A||B|$ if

$$A = \begin{pmatrix} 3 & 1 \\ 2 & 1 \end{pmatrix} \quad \text{and } B = \begin{pmatrix} -1 & 3 \\ 5 & 8 \end{pmatrix}.$$

Solution

Because $|A| = 3 - 2 = 1$ and $|B| = -8 - 15 = -23,$ we have

$$|A||B| = (1)(-23) = -23.$$

On the other hand, because

$$AB = \begin{pmatrix} 3 & 1 \\ 2 & 1 \end{pmatrix}\begin{pmatrix} -1 & 3 \\ 5 & 8 \end{pmatrix} = \begin{pmatrix} 2 & 17 \\ 3 & 14 \end{pmatrix},$$

$\left|AB\right| = \begin{vmatrix} 2 & 17 \\ 3 & 14 \end{vmatrix} = -23.$ Hence, $|AB| = |A||B|.$

Now, we generalize **Theorem 2.7.5** to an $n \times n$ matrix.

Theorem 3.4.3.

If A is invertible, then $|A| \neq 0$ *and* $\left|A^{-1}\right| = \frac{1}{|A|}.$

Proof

Because $AA^{-1} = I$ and $|I| = 1,$ it follows from **Theorem 3.4.2** that

$$\left|AA^{-1}\right| = \left|A\right|\left|A^{-1}\right| = \left|I\right| = 1.$$

This implies that $|A| \neq 0$ and $\left|A^{-1}\right| = \frac{1}{|A|}$.

Example 3.4.3.

Let *A* be a 3×3 matrix. Assume that $|A| = 6$. Compute

$$\left|(2A^{-1})^{T}\right|.$$

Solution

By **Theorems 3.2.1**, **3.3.2**, and **3.4.3**, we obtain

$$\left|(2A^{-1})^{T}\right| = \left|2A^{-1}\right| = 2^{3}\left|A^{-1}\right| = 8\frac{1}{|A|} = \frac{8}{6} = \frac{4}{3}.$$

Theorem 3.4.2 shows that the determinant of the product of two $n \times n$ matrices equals the product of the determinants of the two matrices. But in general, it is not true for the addition of two $n \times n$ matrices, that is, in general,

$$|A + B| \neq |A| + |B|.$$

Example 3.4.4.

Let $A = \begin{pmatrix} 1 & 2 \\ 2 & 5 \end{pmatrix}$ and $B = \begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix}$. Show

$$|A + B| \neq |A| + |B|.$$

Solution

Because $A+B=\begin{pmatrix}4 & 3\\ 3 & 8\end{pmatrix}$, $|A+B|=23$. On the other hand, $|A|=1,\ |B|=8$ and $|A|+|B|=9$. Hence, $|A+B|\neq |A|+|B|$.

The following result provides a method to compute $|A|+|B|$, where *A* and *B* are two $n\times n$ matrices and their $(n-1)$ corresponding rows must be the same.

Theorem 3.4.4.

Let

$$A=\begin{pmatrix}a_{11} & a_{12} & \cdots & a_{1n}\\ a_{21} & a_{22} & \cdots & a_{2n}\\ \vdots & \vdots & \ddots & \vdots\\ a_{i1} & a_{i2} & \cdots & a_{in}\\ \vdots & \vdots & \ddots & \vdots\\ a_{n1} & a_{n2} & \cdots & a_{nn}\end{pmatrix} \text{ and } B=\begin{pmatrix}a_{11} & a_{12} & \cdots & a_{1n}\\ a_{21} & a_{22} & \cdots & a_{2n}\\ \vdots & \vdots & \ddots & \vdots\\ b_{i1} & b_{i2} & \cdots & b_{in}\\ \vdots & \vdots & \ddots & \vdots\\ a_{n1} & a_{n2} & \cdots & a_{nn}\end{pmatrix}.$$

Then

(3.4.5)

$$|A| + |B| = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{i1}+b_{i1} & a_{i2}+b_{i2} & \cdots & a_{in}+b_{in} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix}.$$

Proof

We denote by $|C|$ the right-side determinant of **(3.4.5)** . By **Theorem 3.2.1** ,

$\left|A\right| = \sum_{k=1}^{n} a_{ik}A_{ik}$, $\left|B\right| = \sum_{k=1}^{n} b_{ik}A_{ik}$ and

$$|C| = \sum_{k=1}^{n}(a_{ik}+b_{ik})A_{ik} = \sum_{k=1}^{n} a_{ik}A_{ik} + \sum_{k=1}^{n} b_{ik}A_{ik} = |A| + |B|.$$

Example 3.4.5.

Compute $|A| + |B|$ if

$$A = \begin{pmatrix} 1 & 7 & 5 \\ 2 & 0 & 3 \\ 1 & 4 & 7 \end{pmatrix} \text{ and } B = \begin{pmatrix} 1 & 7 & 5 \\ 2 & 0 & 3 \\ 0 & 1 & -1 \end{pmatrix}.$$

Solution

Because the first two rows of A and B are the same, by **Theorem 3.4.4**, we have

$$|A| + |B| = \begin{vmatrix} 1 & 7 & 5 \\ 2 & 0 & 3 \\ 1 & 4 & 7 \end{vmatrix} + \begin{vmatrix} 1 & 7 & 5 \\ 2 & 0 & 3 \\ 0 & 1 & -1 \end{vmatrix} = \begin{vmatrix} 1 & 7 & 5 \\ 2 & 0 & 3 \\ 1+0 & 4+1 & 7-1 \end{vmatrix}$$

$$= \begin{vmatrix} 1 & 7 & 5 \\ 2 & 0 & 3 \\ 1 & 5 & 6 \end{vmatrix} = -28.$$

Exercises

1. For each of the following matrices, use its determinant to determine whether it is invertible.

$$A = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 2 & 1 & 2 & 0 \\ -2 & -3 & 1 & -2 \\ 1 & -1 & 0 & 0 \end{pmatrix} \quad B = \begin{pmatrix} 0 & -1 & 3 & 5 & 2 \\ 1 & -1 & 2 & 4 & -1 \\ 0 & 0 & 1 & 2 & 7 \\ 1 & 1 & -2 & 3 & 1 \\ 0 & -2 & 6 & 10 & 4 \end{pmatrix}$$

$$C = \begin{pmatrix} 2 & 1 & 4 \\ 1 & -2 & 4 \\ -2 & -1 & -4 \end{pmatrix} \quad D = \begin{pmatrix} 0 & 1 & 3 \\ -2 & 1 & 2 \\ 1 & -1 & -1 \end{pmatrix}$$

$$E = \begin{pmatrix} 1 & -1 & 3 & 5 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ -2 & 0 & -1 & 2 & 1 \\ 3 & 0 & 0 & -6 & 6 \\ 2 & 0 & 0 & 0 & 1 \end{pmatrix} \quad F = \begin{pmatrix} 1 & 0 & 0 & 0 \\ -2 & 2 & 0 & 0 \\ 0 & 6 & -6 & 0 \\ 4 & 0 & 2 & -1 \end{pmatrix}$$

2. For each of the following matrices, find all real numbers *x* such that it is invertible.

$$A = \begin{pmatrix} x & 0 & 0 & 0 \\ 2 & x & 2 & 0 \\ -2 & -3 & 1 & -2 \\ 1 & 0 & 0 & 1 \end{pmatrix} \quad B = \begin{pmatrix} 2 & x & 4 \\ 1 & -2 & 4 \\ -2 & -1 & x \end{pmatrix}$$

3. Let A be a 4×4 matrix. Assume that $|A| = 2$. Compute each of the following determinants.

1. $|-3A|$
2. $\left|A^{-1}\right|$
3. $\left|A^{T}\right|$
4. $\left|A^{3}\right|$
5. $\left|\left(2A^{-1}\right)^{T}\right|$
6. $\left|\left(2(-A)^{T}\right)^{-1}\right|$

4. Compute $|A| + |B|$ if $A = \begin{pmatrix} 2 & 1 & 3 \\ 1 & 0 & 3 \\ 1 & 4 & 7 \end{pmatrix}$ and $B = \begin{pmatrix} 2 & 1 & 3 \\ 1 & 0 & 3 \\ 0 & 1 & -1 \end{pmatrix}$.

5. Compute $|A| + |B|$ if

$$A = \begin{pmatrix} 1 & 0 & 0 & 2 \\ -1 & 0 & 1 & 2 \\ 0 & 1 & 4 & 3 \\ -2 & 2 & 0 & 1 \end{pmatrix} \text{ and } B = \begin{pmatrix} 1 & 0 & 0 & 2 \\ 3 & 0 & 0 & 1 \\ 0 & 1 & 4 & 3 \\ -2 & 2 & 0 & 1 \end{pmatrix}.$$

Chapter 4 Systems of linear equations

4.1 Systems of linear equations in *n* variables

A linear equation with *n* variables is of the form

(4.1.1)

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = b,$$

where $a_1, a_2, \cdots, a_n, b$ are constants and $x_1, \cdots, x_n$ are variables.

In particular, a linear equation with two variables can be written as the following form

(4.1.2)

$$ax + by = c,$$

where *a, b, c* are constants and *x,y* are variables. If either *a* or *b* is not zero, then **(4.1.2)** is an equation of a straight line in the *xy*-plane.

A linear equation with three variables can be written in the following form

(4.1.3)

$$ax + by = cz = d,$$

where *a, b, c, d* are constants and *x, y, z* are variables. If one of *a, b, c* is not zero, then **(4.1.3)** is an equation of a plane in $\mathbb{R}^3$. We shall study more about planes in $\mathbb{R}^3$ in **Section 6.3**.

Note that in a linear equation, the power of each variable must be 1 and there are no terms containing the product of two or more variables. Hence, in an equation, if there is a variable whose power is not 1 or if there is a term that contains a product of two or more variables, then the equation is not a linear equation.

Example 4.1.1.

Determine which of the following equations are linear.

1. $0x + 0y = 1;$
2. $x^2 + y^2 = 1;$
3. $XY = 2;$
4. $\sqrt{2}x_1 + x_2 - x_3 = 1.$

Solution

(1) and (4) are linear while (2) and (3) are not linear.

Let $m,\ n \in \mathbb{N}.$ A set of *m* linear equations in the same *n* variables is called a system of linear equations or a linear system, which is of the form:

(4.1.4)

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

The matrix

(4.1.5)

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

is called the **coefficient matrix** of **(4.1.4)** and the matrix

(4.1.6)

$$(A \mid \vec{b}) \left(\begin{array}{cccc|c} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{array}\right)$$

is called the **augmented matrix** of **(4.1.4)** , where $\vec{b} = (b_1,\ b_2,\ \cdots,\ b_m)^T$.

Example 4.1.2.

For each of the following systems of linear equations, find its coefficient and augmented matrices.

1. $\begin{cases} x_1 - x_2 & = & 7 \\ x_1 + x_2 & = & 5 \end{cases}$

2. (2) $\begin{cases} x_1 + x_2 + 3x_3 = 0 \\ 2x_1 + 4x_2 - 3x_3 = 0 \\ 3x_1 + 6x_2 - 5x_3 = 0 \end{cases}$

3. $\begin{cases} 2x_1 + 3x_3 = 2 \\ 4x_2 - 3x_3 = 1 \\ x_1 + 2x_2 - x_3 = 0 \end{cases}$

Solution

1. $A = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$ and $(A\,|\,\vec{b}) = \left(\begin{array}{cc|c} 1 & -1 & 7 \\ 1 & 1 & 5 \end{array}\right)$.

2. $A = \begin{pmatrix} 1 & 1 & 3 \\ 2 & 4 & -3 \\ 3 & 6 & -5 \end{pmatrix}$ and $(A\,|\,\vec{b}) = \left(\begin{array}{ccc|c} 1 & 1 & 3 & 0 \\ 2 & 4 & -3 & 0 \\ 3 & 6 & -5 & 0 \end{array}\right)$.

3. $A=\begin{pmatrix}2&0&3\\0&4&-3\\1&2&-1\end{pmatrix}$ and $(A\,|\,\vec{b})=\left(\begin{array}{ccc|c}2&0&3&2\\0&4&-3&1\\1&2&-1&0\end{array}\right)$.

It is useful to write **(4.1.4)** into other expressions. Let

$$\vec{X}=(x_1,\ x_2,\ \cdots,\ x_n)^T\in\mathbb{R}^n \text{ and } \vec{b}=(b_1,\ b_2,\ \cdots,\ b_m)^T.$$

By **(2.2.2)** , the system **(4.1.4)** can be written into the matrix form

(4.1.7)

$$A\vec{X}=\vec{b}.$$

The system **(4.1.7)** with $\vec{b}=\vec{0}$, that is,

(4.1.8)

$$A\vec{X}=\vec{0},$$

is called a homogeneous system. If $\vec{b}\neq\vec{0}$, it is called a nonhomogeneous system.

Let $\overrightarrow{a_1},\ \overrightarrow{a_2},\ \cdots,\ \overrightarrow{a_n}$ be the column vectors of *A* given in **(4.1.5)** . By **(2.2.3)** , **(4.1.4)** can be written into the linear combination form:

(4.1.9)

$$x_1 \overrightarrow{a_1} + x_2 \overrightarrow{a_2} + \cdots + x_n \overrightarrow{a_n} = \vec{b}.$$

Example 4.1.3.

Consider the system of linear equations

(4.1.10)

$$\begin{cases} x_1 + x_2 + 3x_3 = 2 \\ 2x_1 + 4x_2 - 3x_3 = 1 \\ 3x_1 + 6x_2 + 5x_3 = 0 \end{cases}$$

Write the above system into the forms **(4.1.7)** and **(4.1.9)** .

Solution

Let

$$A = \begin{pmatrix} 1 & 1 & 3 \\ 2 & 4 & -3 \\ 3 & 6 & 5 \end{pmatrix}, \vec{X} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}, \text{ and } \vec{b} = \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix}.$$

Then the system **(4.1.10)** can be rewritten as $A\vec{X} = \vec{b}$. Let

$$\overrightarrow{a_1} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \overrightarrow{a_2} = \begin{pmatrix} 1 \\ 4 \\ 6 \end{pmatrix}, \overrightarrow{a_3} = \begin{pmatrix} 3 \\ -3 \\ 5 \end{pmatrix}, \text{ and } \vec{b} = \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix}.$$

Then the system **(4.1.10)** can be rewritten as $x_1\overrightarrow{a_1} + x_2\overrightarrow{a_2} + x_3\overrightarrow{a_3} = \vec{b}$.

Now, we introduce the definition of a solution for the system **(4.1.4)**.

Definition 4.1.1.

If $(s_1, s_2, \cdots, s_n)$ satisfies each of the linear equations in **(4.1.4)**, that is,

(4.1.11)

$$\begin{cases} a_{11}s_1 + a_{12}s_2 + \cdots + a_{1n}s_n = b_1 \\ a_{21}s_1 + a_{22}s_2 + \cdots + a_{2n}s_n = b_2 \\ \vdots \\ a_{m1}s_1 + a_{m2}s_2 + \cdots + a_{mn}s_n = b_m, \end{cases}$$

then $(s_1, s_2, \cdots, s_n)$ is called a solution of **(4.1.4)**. Solving a system of linear equations means finding all solutions of the system.

Note that any solution $(s_1, s_2, \cdots, s_n)$ of **(4.1.4)** must satisfy each equation of **(4.1.4)**. If $(s_1, s_2, \cdots, s_n)$ does not satisfy one of the equations of **(4.1.4)**, it is not a solution of **(4.1.4)**.

Also, note that if we consider **(4.1.4)**, $(s_1, s_2, \cdots, s_n)$ can be treatedasa point in $\mathbb{R}^n$ or as a vector in $\mathbb{R}^n$.

Example 4.1.4.

For each of the following points:

$$P_1(6,\ -1),\ P_2(8,\ 1),\ P_3(2,\ 3),\ P_4(0,\ 0),$$

verify whether it is a solution of the system of linear equations

$$\begin{cases} x-y &= 7, \\ x+y &= 5. \end{cases}$$

Solution

$(6,\ -1)$ satisfies both equations, so $P_1(6,\ -1)$ is a solution. (8, 1) satisfies the first equation but does not satisfy the second one, so $P_2(8,\ 1)$ is not a solution. (2,3) satisfies the second equation but does not satisfy the the first one, so $P_3(2,\ 3)$ is not a solution. (0,0) satisfies neither of the two equations, so $P_4(0,\ 0)$ is not a solution.

Example 4.1.5.

Consider the following system of linear equations.

(4.1.12)

$$\begin{cases} x_1+x_2+3x_3=2 \\ x_1-x_2+x_3=-3 \\ 2x_1+4x_3=-1 \end{cases}$$

Verify that $(-\frac{5}{2}, \frac{3}{2}, 1)$ is a solution of **(4.1.12)** while $(1, 3, -1)$ is not a solution.

Solution

Let $x_1 = \frac{3}{2}$, $x_2 = \frac{3}{2}$, and $x_3 = 1$. Then

$$\begin{cases} x_1 + x_2 + 3x_3 = -\frac{5}{2} + \frac{3}{2} + 3(1) = 2, \\ x_1 - x_2 + x_3 = -\frac{5}{2} - \frac{3}{2} + 1 = -3, \\ 2x_1 + 4x_3 = 2(-\frac{5}{2}) + 4(1) = -1. \end{cases}$$

Hence, $\left(-\frac{5}{2}, \frac{3}{2}, 1\right)$ satisfies each of the three equations in **(4.1.12)** and is a solution of **(4.1.12)** .

Let $x_1 = 1$, $x_2 = 3$, and $x_3 = -1$. Then

$$x_1 + x_2 + 3x_3 = 1 + 3 + 3(-1) = 1 \neq 2$$

Hence,$(1, 3, -1)$ does not satisfy the first equation of **(4.1.12)** , and thus it is not a solution of **(4.1.12)** .

Note that $(x_1, x_2, x_3) = (1, 3, -1)$ satisfies the second equation but not the third one of **(4.1.12)** .

By **Definition 4.1.1** , we see that the zero vector $\vec{0}$ in $\mathbb{R}^n$ is always a solution of the homogeneous system **(4.1.8)** . Sometimes, the zero solution of **(4.1.8)** is called a trivial solution. A nonzero solution of **(4.1.8)** is called a nontrivial solution.

We denote by N_A the set of all solutions of **(4.1.8)**, that is,

(4.1.13)

$$N_A = \left\{\vec{X} \in \mathbb{R}^n : A\vec{X} = \vec{0}\right\}.$$

Definition 4.1.2.

The set N_A is called the solution space of the homogeneous system **(4.1.8)** or the nullspace of the matrix *A*. The nullity of *A* is called the dimension of the solution space N_A, denoted by dim (N_A), that is,

(4.1.14)

$$\dim(N_A) = \text{null}(A).$$

A solution space that only contains the zero vector is said to be a zero space.

Definition 4.1.3.

The linear system **(4.1.4)** is said to be consistent if it has at least one solution, and to be inconsistent if it has no solutions.

By **Definition 4.1.3**, a homogeneous system is consistent because it has at least a trivial solution.

By **(4.1.9)** and **Theorem 1.4.3**, we obtain the following relation among the consistency of a linear system, the linear combination, and the spanning space.

Theorem 4.1.1.

Let $S = \left\{\overrightarrow{a}_1, \cdots, \overrightarrow{a}_n\right\}$ *be the same as in* **(1.4.2)** , $A = \left\{\overrightarrow{a}_1, \cdots, \overrightarrow{a}_n\right\}$ *be the same as in* **(4.1.5)** *, and* $\vec{b} = (b_1, b_2, \cdots, b_m)^T$. *Then the following assertions are equivalent.*

1. $A\vec{X} = \vec{b}$ *is consistent.*
2. $\vec{b}$ *is a linear combination of S.*
3. $\vec{b} \in$ span *S*.

Proof

If (1) holds, then by **Definition 4.1.3** , there exists $\vec{X} = (x_1, x_2, \cdots, x_n)$ such that $A\vec{X} = \vec{b}$. By **(2.2.3)** , **(4.1.9)** holds and thus, the result (2) holds. Conversely, if (2) holds, then there exist $\vec{X} = (x_1, x_2, \cdots, x_n)$ such that **(4.1.9)** holds. By **(2.2.3)** , $A\vec{X} = \vec{b}$ and the result (1) holds. This shows that (1) and (2) are equivalent. By **Theorem 1.4.3** , the results (2) and (3) are equivalent.

By **Theorem 4.1.1** , we see that the system **(4.1.4)** is inconsistent if and only if $\vec{b}$ is not a linear combination of *S* and if and only if $\vec{b} \notin$ span *S*.

For a system of two linear equations, it is not hard to solve it. Here we give examples to show how to solve such systems and provide geometric meanings that exhibit the solution structures.

Example 4.1.6.

Solve the following system.

(4.1.15)

$$\begin{cases} x_1 - x_2 = 7, \\ x_1 + x_2 = 5. \end{cases}$$

Solution

Adding the two equations implies $2x_1 = 12$ and $x_1 = 6$. Substituting $x_1 = 6$ into the first equation implies $x_2 = x_1 - 7 = 6 - 7 = -1,$ so $(x_1, x_2) = (6, -1)$ is a solution of the system.

In geometric terms, each of the equations in **(4.1.15)** represents a line and the solution $(6, -1)$ is the intersection point of the two lines. In fact, the solution $(6, -1)$ is the unique solution of the system.

Example 4.1.7.

Solve the following system

$$\begin{cases} x_1 + x_2 = 1, \\ 2x_1 + 2x_2 = 2 \end{cases}$$

and express the solution with a linear combination.

Solution

The two equations are the same and the solutions of the first equation are the solutions of the system. Let $x_2 = t \in \mathbb{R}.$ Then $x_1 = 1 - t$ and

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} + t\begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

is a solution of the system for each $t \in \mathbb{R}$.

In geometry, the two equations represent the same line and every point on the line is a solution of the system. Hence, the system has infinitely many solutions.

Example 4.1.8.

Solve the following system

$$\begin{cases} x_1 + x_2 & = & 1, \\ x_1 + x_2 & = & 2. \end{cases}$$

Solution

It is obvious that any $(x_1,\ x_2)$ does not satisfy both equations. Hence, the above system has no solutions.

In geometric terms, the two equations represent two parallel lines. Hence, the above system has no solutions.

The above three examples show that a system of two linear equations with two variables has only one solution, or infinitely many solutions, or no solutions. The same fact holds for a system of linear equations of two or more variables. We shall study the general case in **Section 4.5** .

The methods used to solve a system of two linear equations are not suitable for a general system **(4.1.4)** . In Sections 4.2–4.4, we shall study approaches to solve **(4.1.4)** or **(4.1.7)** , where we do not often mention the linear combinations and spanning spaces. In Sections 4.5 and 4.6, following **Theorem 4.1.1** , we shall study how to determine whether $A\vec{X} = \vec{b}$ is consistent, $\vec{b}$ is a linear combination of *S*, and $\vec{b} \in$ span *S*.

Exercises

1. Determine which of the following equations are linear.
   a. $x + 2y + z = 6$
   b. $2x - 6y + z = 1$
   c. $-\sqrt{2}x + 6^{-\frac{2}{3}}y = 4 - 3z$
   d. $3x_1 + 2x_2 + 4x_3 + 5x_4 = 1$
   e. $2xy + 3yz + 5z = 8$

2. For each of the following systems of linear equations, find its coefficient and augmented matrices.

   1. $\begin{cases} 2x_1 - x_2 & = & 6 \\ 4x_1 + x_2 & = & 3 \end{cases}$

   2. $\begin{cases} -x_1 + 2x_2 + 3x_3 = 4 \\ 3x_1 + 2x_2 - 3x_3 = 5 \\ 2x_1 + 3x_2 - x_3 = 1 \end{cases}$

   3. $\begin{cases} x_1 - 3x_3 - 4 = 0 \\ 2x_2 - 5x_3 - 8 = 0 \\ 3x_1 + 2x_2 - x_3 = 4 \end{cases}$

3. For each of the following augmented matrices, find the corresponding linear system.

$$\left(\begin{array}{ccc|c} 1 & 2 & 3 & 1 \\ -1 & 1 & 0 & -2 \\ 0 & -1 & 1 & 0 \end{array}\right) \quad \left(\begin{array}{cc|c} 2 & 1 & 1 \\ 0 & -1 & -2 \\ 0 & 3 & -1 \end{array}\right)$$

4. For each of the following linear systems, write it into the form $A\vec{X} = \vec{b}$ and express $\vec{b}$ as a linear combination of the column vectors of *A*.

a. $\begin{cases} -x_1 + x_2 + 2x_3 = 3 \\ 2x_1 + 6x_2 - 5x_3 = 2 \\ -3x_1 + 7x_2 - 5x_3 = -1 \end{cases}$

b. $\begin{cases} x_1 - x_2 + 4x_3 = 0 \\ -2x_1 + 4x_2 - 3x_3 = 0 \\ 3x_1 + 6x_2 - 8x_3 = 0, \end{cases}$

c. $\begin{cases} x_1 - 2x_2 + x_3 = 0 \\ -x_1 + 3x_2 - 4x_3 = 0 \end{cases}$

5. For each of the following points: $P_1\left(6,\ -\frac{1}{2}\right)$, $P_2(8,\ 1)$, $P_3(2, 3)$, and $P_4(0,\ 0)$, verify whether it is a solution of the system.

$$\begin{cases} x - 2y = 7, \\ x + 2y = 5. \end{cases}$$

6. Consider the system of linear equations

$$\begin{cases} -x_1 + x_2 + 2x_3 = 3 \\ 2x_1 + 6x_2 - 5x_3 = 2 \\ -3x_1 + 7x_2 - 5x_3 = -1 \end{cases}$$

1. Verify that $(1,\ 3,\ -1)$ is a solution of the system.
2. Verify that $\left(\frac{5}{6},\ \frac{7}{6},\ \frac{4}{3}\right)$ is a solution of the system.
3. Determine if the vector $\vec{b} = (3,\ 2,\ -1)^T$ is a linear combination of the column vectors of the coefficient matrix of the system.
4. Determine if the vector $\vec{b} = (3,\ 2,\ -1)^T$ belongs to the spanning space of the column vectors of the coefficient matrix of the system.

7. Solve each of the following systems.

   1. $\begin{cases} x_1 - x_2 = 6 \\ x_1 + x_2 = 5 \end{cases}$

   2. $\begin{cases} x_1 + 2x_2 = 1 \\ 2x_1 + x_2 = 2 \end{cases}$

   3. $\begin{cases} x_1 + 2x_2 = 1 \\ x_1 + 2x_2 = 2 \end{cases}$

4.2 Gaussian Elimination

Before we study Gaussian elimination, we study how to solve **(4.1.7)** :

(4.2.1)

$$A\vec{X} = \vec{b},$$

where the coefficient matrix *A* is assumed to be a row echelon matrix.

Recall that for each $i \in I_n,$ the *i*th column of the coefficients matrix *A* is from the coefficients of the variable x_i in the system. Because *A* is a row echelon matrix, some columns of *A* would contain leading entries.

A variable x_i is said to be a basic variable if the *i*th column of *A* contains a leading entry, and to be a free variable if the *i*th column of *A* does not contain a leading entry.

The method to solve such a system **(4.2.1)** is to treat the free variables, if any, as parameters, and then solve each linear equation for the basic variable starting from the last equation to the first one, where the parameters and the obtained values of the basic variables are substituted. The method used to solve such a system is called **back-substitution.**

Example 4.2.1.

Consider the system of the linear equations

(1)

$$x_1 + 2x_2 - 3x_3 = 4$$

(2)

$$2x_2 - 6x_3 = 5$$

(3)

$$-x_3 = 2$$

a. Find the coefficient matrix and the augmented matrix of the system.
b. Find the basic variables and free variables of the system.
c. Solve the system using back-substitution.
d. Determine if the vector $\vec{b} = (4,\ 5,\ 2)^T$ is a linear combination of the column vectors of the coefficient matrix of the system.

Solution

a. The coefficient and the augmented matrices are

$$A = \begin{pmatrix} 1 & 2 & -3 \\ 0 & 2 & -6 \\ 0 & 0 & -1 \end{pmatrix},\ (A\,|\,\vec{b}) = \left(\begin{array}{ccc|c} 1 & 2 & -3 & 4 \\ 0 & 2 & -6 & 5 \\ 0 & 0 & -1 & 2 \end{array}\right).$$

b. *A* is a 3×3 row echelon matrix. The columns 1,2 3 of *A* contain leading entries and thus, $x_1,\ x_2,$ and x_3 are basic variables. There are no free variables.
c. To solve the system, we first consider equation (3). Solving equation (3) for the basic variable $x_3,$ we get $x_3 = -2$. Substituting $x_3 = -2$ into equation (2) and solving equation (2) for the basic variable $x_2,$ we obtain

$$x_2 = \frac{1}{2}(5 + 6x_3) = \frac{1}{2}[5 + 6(-2)] = -\frac{7}{2}.$$

Substituting $x_2 = -\frac{7}{2}$ and $x_3 = -2$ into equation (1) and solving equation (1) for the basic variable x_1, we obtain

$$x_1 = 4 - 2x_2 + 3x_3 = 4 - 2(-\frac{7}{2}) + 3(-2) = 5.$$

Hence, $(x_1, x_2, x_3) = \left(5, -\frac{7}{2}, -2\right)$ is a solution of the system.

d. Because the system has a solution, the vector $\vec{b} = (4, 5, 2)^T$ is a linear combination of the column vectors of *A*, namely,

$$\vec{b} = 5\vec{a_1} - \frac{7}{2}\vec{a_2} - 2\vec{a_3},$$

where $\vec{a_1} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$, $\vec{a_2} = \begin{pmatrix} 2 \\ 2 \\ 0 \end{pmatrix}$, and $\vec{a_3} = \begin{pmatrix} -3 \\ -6 \\ -1 \end{pmatrix}$.

Example 4.2.1 shows that if $r(A) = r(A\,|\,\vec{b}) = n$, where $n = 3$ is the number of variables, then the system has a unique solution.

Example 4.2.2.

Consider the linear system

(1)

$$x_1 - x_2 - 3x_3 + x_4 = 1$$

(2)

$$-x_2 - x_3 + x_4 = 2$$

(3)

$$-x_4 = 3$$

a. Find the coefficient matrix and the augmented matrix of the system.
b. Find the basic variables and free variables of the system.
c. Solve the system using back-substitution and express its solution as a linear combination.
d. Determine if the vector $\vec{b} = (1,\ 2,\ 3)^T$ is a linear combination of the column vectors of the coefficient matrix of the system.

Solution

a. The coefficient and the augmented matrices are

$$A = \begin{pmatrix} 1 & -1 & -3 & 1 \\ 0 & -1 & -1 & 1 \\ 0 & 0 & 0 & -1 \end{pmatrix}, \ (A \,|\, \vec{b}) = \left(\begin{array}{cccc|c} 1 & -1 & -3 & 1 & 1 \\ 0 & -1 & -1 & 1 & 2 \\ 0 & 0 & 0 & -1 & 3 \end{array}\right).$$

b. *A* is a row echelon matrix. The columns 1,2,4 of *A* contain leading entries and thus, $x_1,\ x_2,$ and x_4 are basic variables. The third column of A does not contain any leading entries, so x_3 is a free variable.

c. Because x_3 is a free variable, let $x_3 = t$, where $t \in \mathbb{R}$ is an arbitrary real number. Solving equation (3) for the basic variable x_4 implies $x_4 = -3$. Substituting $x_4 = -3$ into equation (2) and solving equation (2) for the basic variable x_2, we obtain

$$x_2 = -2 - x_3 + x_4 = -2 - t + (-3) = -5 - t.$$

Substituting $x_2 = -5 - t,\ x_3 = t$ and $x_4 = -3$ into equation (1) and solving equation (1) for the basic variable x_1, we obtain

$$x_1 = 1 + x_2 + 3x_3 - x_4 = 1 + (-5 - t) + 3t - (-3) = -1 + 2t.$$

Hence, $(x_1,\ x_2,\ x_3,\ x_4) = (-1 + 2t,\ \ -5\ - t,\ t,\ -3)$ is a solution of the system for each $t \in \mathbb{R}$. We can write the solution as a linear combination:

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} -1+2t \\ -5-t \\ 0+t \\ -3+0t \end{pmatrix} = \begin{pmatrix} -1 \\ -5 \\ 0 \\ -3 \end{pmatrix} + \begin{pmatrix} 2t \\ -t \\ t \\ 0t \end{pmatrix} = \begin{pmatrix} -1 \\ -5 \\ 0 \\ -3 \end{pmatrix} + t\begin{pmatrix} 2 \\ -1 \\ 1 \\ 0 \end{pmatrix}.$$

d. Because the system has infinitely many solutions, $\vec{b} = (1,\ 2,\ 3)^T$ is a linear combination of the column vectors of *A*.

The above example shows that if $r(A) = r(A\,|\,\vec{b}) = 3 < n$, where $n = 4$ is the number of variables, then the system has infinitely many solutions.

Example 4.2.3.

Consider the following system

$x_1 - x_2 - 3x_3 + x_4 - x_5 = 2$ (1)

$-x_3 + x_4 = 3$ (2)

$-x_4 + x_5 = 1$ (3)

a. Find the coefficient matrix and the augmented matrix of the system.
b. Find the basic variables and free variables of the system.
c. Solve the system using back-substitution and express its solutions as a linear combination.
d. Determine if the vector $\vec{b} = (2,\ 3,\ 1)^T$ is a linear combination of the column vectors of the coefficient matrix of the system.

Solution

a. The coefficient and the augmented matrices are

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

$$(A\,|\,\vec{b}) = \left(\begin{array}{ccccc|c} 1 & -1 & -3 & 1 & -1 & 2 \\ 0 & 0 & -1 & 1 & 0 & 3 \\ 0 & 0 & 0 & -1 & 1 & 1 \end{array}\right).$$

b. *A* is a 3×5 row echelon matrix. The columns 1,3,4 of *A* contain leading entries and thus, $x_1,\ x_3$, and x_4 are basic variables. The columns 2,5 of *A* do not contain any leading entries, so x_2 and x_5 are free variables.

c. Because x_2 and x_5 are free variables, let $x_2 = s$ and $x_5 = t,$ where $s,\ t \in \mathbb{R}.$ Substituting $x_5 = t$ into equation (3) and solving equation (3) for the basic variable $x_4,$ we obtain $x_4 = -1 + x_5 = -1 + t.$ Substituting $x_4 = -1 + t$ into equation (2) and solving equation (2) for the basic variable $x_3,$ we obtain

$$x_3 = -3 + x_4 = -3 + (-1 + t) = -4 + t.$$

Substituting $x_2 = s,\ x_3 = -4 + t,\ x_4 = -1 + t$ and $x_5 = t$ into equation (1) and solving equation (1) for the basic variable $x_1,$ we obtain

$$x_1 = 2 + x_2 + 3x_3 - x_4 + x_5 = 2 + s + 3(-4 + t) - (-1 + t) + t = -9 + s + 3t.$$

Hence,

$$(x_1,\ x_2,\ x_3,\ x_4,\ x_5) = (-9 + s + 3t,\ s,\ \ -4 + t,\ \ -1 + t,\ t)$$

is a solution of the system for each $s,\ t \in \mathbb{R}.$
We write the solution as a linear combination:

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} = \begin{pmatrix} -9+s+3t \\ s \\ -4+t \\ -1+t \\ t \end{pmatrix} = \begin{pmatrix} -9+s+3t \\ 0+s+0t \\ -4+0s+t \\ -1+0s+t \\ 0+0s+t \end{pmatrix}$$

$$= \begin{pmatrix} -9 \\ 0 \\ -4 \\ -1 \end{pmatrix} + \begin{pmatrix} s \\ s \\ 0 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 3t \\ 0t \\ t \\ t \\ t \end{pmatrix}$$

$$= \begin{pmatrix} -9 \\ 0 \\ -4 \\ -1 \\ 0 \end{pmatrix} + s\begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} + t\begin{pmatrix} 3 \\ 0 \\ 1 \\ 1 \\ 1 \end{pmatrix}.$$

d. Because the system has infinitely many solutions, $\vec{b} = (2,\ 3,\ 1)^T$ is a linear combination of the column vectors of *A*.

Example 4.2.3 also shows that if $r(A) = r(A\,|\,\vec{b}) = 3 < n,$ where $n = 5$ is the number of variables, then the system has infinitely many solutions.

Example 4.2.4.

Consider each of the following linear systems

i. $$\begin{cases} x_1 - x_2 - 3x_3 = 1 \\ -x_2 + x_3 = 3 \\ 0x_3 = 1 \end{cases}$$

ii. $$\begin{cases} x_1 - x_2 - 3x_3 = 1 \\ -x_2 + x_3 = 3 \\ 2x_3 = -1 \\ 0x_3 = 5 \end{cases}$$

a. Find the coefficient matrix and the augmented matrix of the system.
b. Find the basic variables and free variables of the system.
c. Solve the system.

Solution

i.

a. The coefficient and the augmented matrices are

$$A = \begin{pmatrix} 1 & -1 & -3 \\ 0 & -1 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \ (A \mid \vec{b}) = \left(\begin{array}{ccc|c} 1 & -1 & -3 & 1 \\ 0 & -1 & 1 & 3 \\ 0 & 0 & 0 & 1 \end{array}\right).$$

b. *A* is a 3×3 row echelon matrix. The columns 1,2 of *A* contain leading entries and thus, x_1 and x_2 are basic variables. The column 3 of *A* does not contain any leading entries, so x_3 is a free variable.

c. Because the last equation implies $0 = 1,$ which shows that any (x_1, x_2, x_3) doesn't satisfy the equation. Hence, the system has no solutions.

ii.

a. The coefficient and the augmented matrices are

$$A = \begin{pmatrix} 1 & -1 & -3 \\ 0 & -1 & 1 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{pmatrix}, \ (A \mid \vec{b}) = \left(\begin{array}{ccc|c} 1 & -1 & -3 & 1 \\ 0 & -1 & 1 & 3 \\ 0 & 0 & 2 & -1 \\ 0 & 0 & 0 & 5 \end{array}\right).$$

b. *A* is a 4×3 row echelon matrix. The columns 1,2,3 of *A* contain leading entries and thus, $x_1, x_2,$ and x_3 are basic variables. There are no free variables.

c. Because the last equation implies $0 = 5,$ which shows that any (x_1, x_2, x_3) doesn't satisfy the equation. Hence, the system has no solutions.

Note that in **Example 4.2.4** (*i*), $r(A) = 2 < 3 = n,$ where *n* is the number of variables, and $r(A \mid \vec{b}) = 3.$ **Example 4.2.4** (*i*) shows that if $r(A) < r(A \mid \vec{b}),$ then the system has no solutions, and even when a system contains free variables, the system may have no solutions. In **Example 4.2.4** (*ii*), $r(A) = 3 = n,$ where *n* is the number of variables and $r(A \mid \vec{b}) = 4.$ **Example 4.2.4** (*ii*) also shows that if $r(A) < r(A \mid \vec{b}),$ then the system has no solutions, and even when $r(A) = n,$ the system may have no solutions.

Now, we study how to solve the general system $A\vec{X} = \vec{b}$ given in **(4.1.7)** , where *A* is not a row echelon matrix.

Let *B* be a row echelon matrix of *A*. By **Theorems 2.6.2** and **2.7.3**, there exists an invertible matrix *E* such that B $B = EA$. By **(4.1.7)**, we obtain

(4.2.2)

$$B\vec{X} = (EA)\vec{X} = E\vec{b} := \vec{c}.$$

Theorem 4.2.1.

(4.1.7) and (4.2.2) are equivalent.

Proof

We have shown above that if $\vec{X}$ is a solution of **(4.1.7)**, then it is a solution of **(4.2.2)**. Conversely, noting that *E* is invertible, we obtain that if $\vec{X}$ is a solution of **(4.2.2)**, then it is a solution of **(4.1.7)**. Hence, **(4.1.7)** and **(4.2.2)** are equivalent.

Let $(A\,|\,\vec{b})$ be the augmented matrix of **(4.1.4)** given in **(4.1.6)**. Note that **(4.2.2)** can be expressed as

$$E(A\,|\,\vec{b}) = (EA\,|\,E\vec{b}) = (B\,|\,\vec{c}).$$

Hence, to solve **(4.1.7)**, we first use row operations to change $(A\,|\,\vec{b})$ to $(B\,|\,\vec{c})$ and then to solve $B\vec{X} = \vec{c}$, where *B* is a row echelon matrix, by using back-substitution. By **Theorem 4.2.1**, the solutions of **(4.2.2)** are the solutions of **(4.1.7)**. This method is called **Gaussian elimination**.

Example 4.2.5.

Solve each of the following systems by using Gaussian elimination.

a. $\begin{cases} 2x_1 + 4x_2 + 6x_3 = 18 \\ 4x_1 + 5x_2 + 6x_3 = 24 \\ 3x_1 + x_2 - 2x_3 = 4 \end{cases}$

b. $\begin{cases} 2x_1 + 2x_2 - 2x_3 = 4 \\ 3x_1 + 5x_2 + x_3 = -8 \\ -4x_1 - 7x_2 - 2x_3 = 13 \end{cases}$

c. $\begin{cases} x_1 + 2x_2 - x_3 + 3x_4 = 5 \\ -2x_1 + 3x_2 - 5x_3 + x_4 = 4 \end{cases}$

d. $\begin{cases} x_1 + 2x_2 + 3x_3 = 4 \\ 4x_1 + 7x_2 + 6x_3 = 17 \\ 2x_1 + 5x_2 + 12x_3 = 10 \end{cases}$

Solution

a.

$$(A\,|\,\vec{b}) = \left(\begin{array}{ccc|c} 2 & 4 & 6 & 18 \\ 4 & 5 & 6 & 24 \\ 3 & 1 & -2 & 4 \end{array}\right) \xrightarrow{R_1(\frac{1}{2})} \left(\begin{array}{ccc|c} 1 & 2 & 3 & 9 \\ 4 & 5 & 6 & 24 \\ 3 & 1 & -2 & 4 \end{array}\right)$$

$$\underset{R_1(-3)+R_3}{\overset{R_1(-4)+R_2}{\longrightarrow}} \left(\begin{array}{ccc|c} 1 & 2 & 3 & 9 \\ 0 & -3 & -6 & -12 \\ 0 & -5 & -11 & -23 \end{array}\right)$$

$$\xrightarrow{R_2(-\frac{1}{3})} \left(\begin{array}{ccc|c} 1 & 2 & 3 & 9 \\ 0 & 1 & 2 & 4 \\ 0 & -5 & -11 & -23 \end{array}\right)$$

$$\xrightarrow{R_2(5)+R_3} \left(\begin{array}{ccc|c} 1 & 2 & 3 & 9 \\ 0 & 1 & 2 & 4 \\ 0 & 0 & -1 & -3 \end{array}\right) = (B\,|\,\vec{c}).$$

The system of linear equations corresponding to $(B\,|\,\vec{c})$ is

(1)

$$x_1 + 2x_2 + 3x_3 = 9$$

(2)

$$x_2 + 2x_3 = 4$$

(3)

$$-x_3 = -3$$

Now, we use back-substitution to solve the above system. The variables $x_1,\ x_2,$ and x_2 are basic variables. Solving equation (3) for basic variable x_3 implies $x_3 = 3$. Substituting $x_3 = 3$ into equation (2) and solving equation (2) for the basic variable $x_2,$ we obtain $x_2 = 4 - 2x_3 = 4 - 2(3) = -2$. Substituting $x_2 = -2$ and $x_3 = 3$ into equation (1) and solving equation (1) for the basic variable $x_1,$ we obtain

$$x_1 = 9 - 2x_2 - 3x_3 = 9 - 2(-2) - 3(3) = 4.$$

Hence $(x_1,\ x_2,\ x_3) = (4,\ -2,\ 3)$ is a solution of the system.

$$(A\,|\,\vec{b}) = \left(\begin{array}{ccc|c} 2 & 2 & -2 & 4 \\ 3 & 5 & 1 & -8 \\ -4 & -7 & -2 & 13 \end{array}\right) R_1(\tfrac{1}{2})\to \left(\begin{array}{ccc|c} 1 & 1 & -1 & 2 \\ 3 & 5 & 1 & -8 \\ -4 & -7 & -2 & 13 \end{array}\right)\Bigg\downarrow$$

b. $$\underset{R_1(4)+R_3}{\overset{R_1(-3)+R_2}{\longrightarrow}} \left(\begin{array}{ccc|c} 1 & 1 & -1 & 2 \\ 0 & 2 & 4 & -14 \\ 0 & -3 & -6 & 21 \end{array}\right) R_2(\tfrac{1}{2})\to \left(\begin{array}{ccc|c} 1 & 1 & -1 & 2 \\ 0 & 1 & 2 & -7 \\ 0 & -3 & -6 & 21 \end{array}\right)\Bigg\downarrow$$

$$R_2(3) + R_3\to \left(\begin{array}{ccc|c} 1 & 1 & -1 & 2 \\ 0 & 1 & 2 & -7 \\ 0 & 0 & 0 & 0 \end{array}\right) = (B\,|\,\vec{c}).$$

The system of linear equations corresponding to $(B\,|\,\vec{c})$ is
(1)

$$x_1+x_2-x_3=2$$

(2)

$$x_2+2x_3=-7.$$

The variables x_1 and x_2 are the basic variables and x_3 is a free variable. Let $x_3=t,$ where $t\in\mathbb{R}.$ Substituting $x_3=t$ into equation (2) and solving equation (2) for basic variable $x_2,$ we have $x_2=-7-2x_3=-7-2t.$ Substituting $x_2=-7-2t$ and $x_3=t$ into equation (1) and solving equation (1) for the basic variable x_1, we obtain

$$x_1=2-x_2+x_3=2-(-7-2t)+t=9+3t.$$

We write the solution as a linear combination:

$$\begin{pmatrix}x_1\\x_2\\x_3\end{pmatrix}=\begin{pmatrix}9+3t\\-7-2t\\t\end{pmatrix}=\begin{pmatrix}9\\-7\\0\end{pmatrix}+t\begin{pmatrix}3\\-2\\1\end{pmatrix}.$$

c. $(A|\vec{b})=\left(\begin{array}{cccc|c}1&2&-1&3&5\\-2&-3&-5&1&4\end{array}\right)\xrightarrow{R_1(2)+R_2}\left(\begin{array}{cccc|c}1&2&-1&3&5\\0&7&-7&7&14\end{array}\right)\xrightarrow{R_2(\frac{1}{7})}\left(\begin{array}{cccc|c}1&2&-1&3&5\\0&1&-1&1&2\end{array}\right)=(B|\vec{c}).$
The system of linear equations corresponding to $(B|\vec{c})$ is
(1)

$$x_1+2x_2-x_3+3x_4=5$$

(2)

$$x_2-x_3+x_4=2.$$

The variables x_1 and x_2 are the basic variables and x_3 and x_4 are free variables. Let $x_3=s$ and $x_4=t$, where $s,\ t\in\mathbb{R}$. Substituting $x_3=s$ and $x_4=t$ into equation (2) and solving equation (2) for the basic variable x_2, we obtain $x_2=2+x_3-x_4=2+s-t$. Substituting $x_2=2+s-t$, $x_3=s$ and $x_4=t$ into equation (1) and solving equation (1) for the basic variable x_1, we obtain

$$x_1=5-2x_2+x_3-3x_4=5-2(2+s-t)+s-3t=1-s-t.$$

We write the above solution as a linear combination:

$$(x_1x_2x_3x_4)=(1-s-t2+s-tst)=(1200)+s(-1110)+(-1-101).$$

d. $(A|\vec{b})=$
$(1234762512|41710)\downarrow\xrightarrow{R_1(-2)+R_3 \; R_1(-4)+R_2}(1230-1-6016|412)\downarrow\xrightarrow{R_2(1)+R_3}(1230-1-6000|413)=$
$(B|\vec{c})$.

The system of linear equations corresponding to $(B|\vec{c})$ is

(1)

$$x_1+2x_2+3x_3=4$$

(2)

$$-x_2-6x_3=1$$

(3)

$$0x_1+0x_2+0x_3=3.$$

The last equation (3) reads $0=3$; this is impossible. Thus the original system has no solutions.

Exercises

1. Consider each of the following systems.
   i. $\{ x_1+x_2-3x_3=2-x_2-6x_3=4-x_3=1$
   ii. $\{ x_1-2x_2+3x_3+x_4=-1x_2-x_3+x_4=4-x_4=2$
   iii. $\{ -x_1-2x_2-2x_3+x_4-3x_5=4-2x_3+3x_4=1-4x_4+2x_5=5$
   iv. $\{ -3x_1-2x_2-2x_3=2x_2-x_3=20x_3=4$

   a. Find the coefficient and augmented matrices of the system.
   b. Find the basic variables and free variables of the system.

c. Solve the system by using back-substitution if it is consistent, and express its solution as a linear combination.

d. Determine if the vector $\vec{b}$ in the right side of the system is a linear combination of the column vectors of the coefficient matrix of the system.

2. Solve each of the following systems by using Gaussian elimination and express its solutions as linear combinations if they have infinitely many solutions.

   a. $\{ x_1-2x_2+3x_3=62x_1+x_2+4x_3=5-3x_1+x_2-2x_3=3$
   b. $\{ x_1+x_2-2x_3=32x_1+3x_2-x_3=-4-2x_1-3x_2-x_3=4$
   c. $\{ x_1+2x_2+3x_3=44x_1+7x_2+6x_3=172x_1+5x_2+12x_3=7$
   d. $\{ x_1+2x_2-x_3+3x_4=53x_1-x_2+4x_3+2x_4=1-x_1+5x_2-6x_3+4x_4=9$
   e. $\{ x_1-2x_2=32x_1-x_2=0-x_1+4x_2=-1$
   f. $\{ -x_1+2x_2-3x_3=-42x_1-x_2+2x_3=2x_1+x_2-x_3=2$

4.3 Gauss-Jordan Elimination

When we use Gaussian elimination to solve the system $A\vec{X} = \vec{b}$, the augmented matrix $(A\,|\,\vec{b})$ is changed only to a row echelon matrix. In fact, we can continue changing the row echelon matrix to a reduced row echelon matrix $(C\,|\,\vec{d})$ and then solve the system $C\vec{X} = \vec{d}$, where *C* is a reduced row echelon matrix. This method is called **Gauss-Jordan elimination**. The advantage of using Gauss-Jordan elimination to solve $A\vec{X} = \vec{b}$ is that when we solve the system $C\vec{X} = \vec{d}$, each equation in the system only contains one basic variable and can be easily solved. By **Theorem 4.2.1**, we see that $A\vec{X} = \vec{b}$ and $C\vec{X} = \vec{d}$ are equivalent, that is, the solutions of $C\vec{X} = \vec{d}$ are the solutions of $A\vec{X} = \vec{b}$.

Example 4.3.1.

Solve each of the following systems by using Gauss-Jordan elimination.

a. $\begin{cases} 2x_1 + 4x_2 + 6x_3 = 18 \\ 4x_1 + 5x_2 + 6x_3 = 24 \\ 3x_1 + x_2 - 2x_3 = 4 \end{cases}$

b. $\begin{cases} x_1 + x_2 - x_3 = 1 \\ -3x_1 + 5x_2 - x_3 = -1 \\ 3x_1 + 7x_2 - 5x_3 = 4 \end{cases}$

c. $\begin{cases} 2x_2+3x_3+x_4=4 \\ x_1+x_3-2x_4=2 \\ -3x_1+5x_2-5x_4=-5 \end{cases}$

d. $\begin{cases} x_1+x_2-2x_3=1 \\ -2x_1-x_2+x_3=-2 \end{cases}$

Solution

a.

$$(A\,|\,\vec{b}) = \left(\begin{array}{ccc|c} 2 & 4 & 6 & 18 \\ 4 & 5 & 6 & 24 \\ 3 & 1 & -2 & 4 \end{array}\right) R_1(\tfrac{1}{2})\rightarrow \left(\begin{array}{ccc|c} 1 & 2 & 3 & 9 \\ 4 & 5 & 6 & 24 \\ 3 & 1 & -2 & 4 \end{array}\right) \downarrow \begin{matrix} R_1(-4)+R_2 \\ \rightarrow \\ R_1(-3)+R_3 \end{matrix}$$

$$\left(\begin{array}{ccc|c} 1 & 2 & 3 & 9 \\ 0 & -3 & -6 & -12 \\ 0 & -5 & -11 & -23 \end{array}\right) R_2\left(-\tfrac{1}{3}\right)\rightarrow \left(\begin{array}{ccc|c} 1 & 2 & 3 & 9 \\ 0 & 1 & 2 & 4 \\ 0 & -5 & -11 & -23 \end{array}\right) \downarrow R_2(5)+R_3\rightarrow$$

$$\left(\begin{array}{ccc|c} 1 & 2 & 3 & 9 \\ 0 & 1 & 2 & 4 \\ 0 & 0 & -1 & -3 \end{array}\right) R_3(-1)\rightarrow \left(\begin{array}{ccc|c} 1 & 2 & 3 & 9 \\ 0 & 1 & 2 & 4 \\ 0 & 0 & 1 & 3 \end{array}\right) \uparrow \begin{matrix} R_3(-2)+R_2 \\ \rightarrow \\ R_3(-3)+R_1 \end{matrix}$$

$$\left(\begin{array}{ccc|c} 1 & 2 & 0 & 0 \\ 0 & 1 & 0 & -2 \\ 0 & 0 & 1 & 3 \end{array}\right) \uparrow R_2(-2)+R_1\rightarrow \left(\begin{array}{ccc|c} 1 & 0 & 0 & 4 \\ 0 & 1 & 0 & -2 \\ 0 & 0 & 1 & 3 \end{array}\right).$$

From the last augmented matrix, we obtain $x_1 = 4,\ x_2 = -2,$ and $x_3 = 3$. Hence, $(x_1,\ x_2,\ x_3) = (4,\ -2,\ 3)$ is a solution of the system.

b.

$$(A\,|\,\vec{b}) = \left(\begin{array}{ccc|c} 1 & 1 & -1 & 1 \\ -3 & 5 & -1 & -1 \\ 3 & 7 & -5 & 4 \end{array}\right) \xrightarrow[R_1(-3)+R_3]{R_1(3)+R_2} \left(\begin{array}{ccc|c} 1 & 1 & -1 & 1 \\ 0 & 8 & -4 & 2 \\ 0 & 4 & -2 & 1 \end{array}\right)$$

$$R_2(\tfrac{1}{2})\to \left(\begin{array}{ccc|c} 1 & 1 & -1 & 1 \\ 0 & 4 & -2 & 1 \\ 0 & 4 & -2 & 1 \end{array}\right) R_2(-1)+R_3\to \left(\begin{array}{ccc|c} 1 & 1 & -1 & 1 \\ 0 & 4 & -2 & 1 \\ 0 & 0 & 0 & 0 \end{array}\right)$$

$$R_2(\tfrac{1}{4})\to \left(\begin{array}{ccc|c} 1 & 1 & -1 & 1 \\ 0 & 1 & -\frac{1}{2} & \frac{1}{4} \\ 0 & 0 & 0 & 0 \end{array}\right) R_2(-1)+R_1\to \left(\begin{array}{ccc|c} 1 & 0 & -\frac{1}{2} & \frac{3}{4} \\ 0 & 1 & -\frac{1}{2} & \frac{1}{4} \\ 0 & 0 & 0 & 0 \end{array}\right).$$

The system of linear equations corresponding to the last augmented matrix is

$$\begin{cases} x_1 - \frac{1}{2}x_3 = \frac{3}{4} \\ x_2 - \frac{1}{2}x_3 = \frac{1}{4}, \end{cases}$$

where $x_1,\ x_2$ are basic variables and x_3 is a free variable. Let $x_3 = t \in \mathbb{R}$.

Then

$$(x_1,\ x_2,\ x_3) = (\frac{3}{4} + \frac{1}{2}t,\ \frac{1}{4} + \frac{1}{2}t,\ t)$$

is a solution of the system.

$$(A\,|\,\vec{b}) = \left(\begin{array}{cccc|c} 0 & 2 & 3 & 1 & 4 \\ 1 & 0 & 1 & -2 & 2 \\ -3 & 5 & 0 & -5 & -5 \end{array}\right) R_{1,\,2}\!\rightarrow \left(\begin{array}{cccc|c} 1 & 0 & 1 & -2 & 2 \\ 0 & 2 & 3 & 1 & 4 \\ -3 & 5 & 0 & -5 & -5 \end{array}\right)\Big\downarrow$$

$$R_3(3) + R_3\!\rightarrow \left(\begin{array}{cccc|c} 1 & 0 & 1 & -2 & 2 \\ 0 & 2 & 3 & 1 & 4 \\ 0 & 5 & 3 & -11 & 1 \end{array}\right) R_2(-2) + R_3\!\rightarrow$$

c. $$\left(\begin{array}{cccc|c} 1 & 0 & 1 & -2 & 2 \\ 0 & 2 & 3 & 1 & 4 \\ 0 & 1 & -3 & -13 & -7 \end{array}\right) R_{2,\,3}\!\rightarrow \left(\begin{array}{cccc|c} 1 & 0 & 1 & -2 & 2 \\ 0 & 1 & -3 & -13 & -7 \\ 0 & 2 & 3 & 1 & 4 \end{array}\right)\Big\downarrow$$

$$R_2(-2) + R_3\!\rightarrow \left(\begin{array}{cccc|c} 1 & 0 & 1 & -2 & 5 \\ 0 & 1 & -3 & 13 & -3 \\ 0 & 0 & 9 & 27 & 18 \end{array}\right) R_3(\tfrac{1}{9})\!\rightarrow$$

$$\left(\begin{array}{cccc|c} 1 & 0 & 1 & -2 & 5 \\ 0 & 1 & -3 & 13 & -3 \\ 0 & 0 & 1 & 3 & 2 \end{array}\right)\Big\uparrow \;\underset{R_3(-1)+R_1}{\overset{R_3(3)+R_2}{\longrightarrow}} \left(\begin{array}{cccc|c} 1 & 0 & 0 & -5 & 3 \\ 0 & 1 & 0 & 22 & 3 \\ 0 & 0 & 1 & 3 & 2 \end{array}\right).$$

The system of linear equations corresponding to the last matrix is

$$\begin{cases} x_1 - 5x_4 = 3 \\ x_2 - 22x_4 = 3 \\ x_3 + 3x_4 = 2, \end{cases}$$

where $x_1,\ x_2,$ and x_3 are basic variables and x_3 is a free variable. Let $x_4 = t \in \mathbb{R}.$ Then

$$(x_1,\ x_2,\ x_3,\ x_4) = (3 + 5t,\ 3 + 22t,\ 2 - 3t,\ t)$$

is a solution of the system.

d. $$(A\,|\,\vec{b}) = \left(\begin{array}{ccc|c} 1 & 1 & -2 & 1 \\ -2 & -1 & 1 & -2 \end{array}\right) \downarrow R_1(2) + R_2 \rightarrow \left(\begin{array}{ccc|c} 1 & 1 & -2 & 1 \\ 0 & 1 & -3 & 0 \end{array}\right) \uparrow$$

$$R_2(-1) + R_1 \rightarrow \left(\begin{array}{ccc|c} 1 & 0 & 1 & 1 \\ 0 & 1 & -3 & 0 \end{array}\right) = (B\,|\,\vec{c}).$$

The system of linear equations corresponding to $(B \mid \vec{c})$ is

$$\begin{cases} x_1 + x_3 = 1 \\ x_2 - 3x_3 = 0 \end{cases}$$

where x_1 and x_2 are basic variables and x_3 is a free variable. Let $x_3 = t.$ Then $x_1 = 1 - x_3 = 1 - t$ and $x_2 = 3x_3 = 3t.$ Hence,

(4.3.1)

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 1-t \\ 3t \\ t \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} -t \\ 3t \\ t \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + t\begin{pmatrix} -1 \\ 3 \\ 1 \end{pmatrix}.$$

Remark 4.3.1.

The system *a*) in **Example 4.3.1** is the same as the system (*a*) in Example **(4.2.5)** and we have used two methods to solve the system. The solution **(4.3.1)** of the system *d*) in **Example 4.3.1** is the parametric equation of the line of intersections of the two planes

$$x_1 + x_2 - 2x_3 = 1 \quad \text{and} \ -2x_1 - x_2 + x_3 = -2.$$

We shall study the lines and planes in more detail in **Sections 6.2** and **6.3** .

For a homogeneous system **(4.1.8)** , we can use Gauss-Jordan elimination to solve it. Assume that we use row operations to change the augmented matrix $(A\,|\,\vec{0})$ to the reduced row echelon matrix denoted by $(B\,|\,\vec{0})$ from which we see that there are null(*A*) free variables.

The following result gives the expression of the solution space of **(4.1.8)** .

Theorem 4.3.1.

Let k = null(*A*). *Let* $x_{n1},\ x_{n2},\ \cdots,\ x_{nk}$ *be free variables of* **(4.1.8)** *,where* $1 \le n_1 < n_2 < \cdots < n_k \le n.$ *Then there exist k vectors*

(4.3.2)

$$\vec{v}_1,\ \vec{v}_2,\ \cdots,\ \vec{v}_k \ \ in\ \mathbb{R}^m$$

that satisfy the following properties.

(P_1) *for each* $i \in \{1, 2, \cdots, k\}$, *the ith component of* $\overrightarrow{v_i}$ *is* 1 *and the component on* n_j *of* $\overrightarrow{v_i}$ *is zero for* $j \neq i$ *and* $j \in \{1, 2, \cdots, k\}$.

$(P_1) N_A = \operatorname{span}\left\{\overrightarrow{v_1}, \overrightarrow{v_2}, \cdots, \overrightarrow{v_k}\right\}.$

Proof

Let $x_{n_i} = t_i$ for $i \in \{1, 2, \cdots, k\}$. We solve the system corresponding to the augmented matrix $(B \mid \vec{0})$ and we can rewrite the solution into the following form

(4.3.3)

$$\vec{X} = t_1 \overrightarrow{v_1} + t_2 \overrightarrow{v_2} + \cdots + t_k \overrightarrow{v_k},$$

where $\left\{\overrightarrow{v_1}, \overrightarrow{v_2}, \cdots, \overrightarrow{v_k}\right\}$ satisfies (P_1). By **(4.3.3)** , we see that (P_2) holds.

Definition 4.3.1.

Let $\alpha_1, \cdots, \alpha_k$ be nonzero numbers. The vectors

$$\alpha_1 \overrightarrow{v_1}, \alpha_2 \overrightarrow{v_2}, \cdots, \alpha_k \overrightarrow{v_k}$$

are called the basis vectors of the solution space N_A, where the vectors $\vec{v_1}, \vec{v_2}, \cdots, \vec{v_k}$ are given in **(4.3.2)** .

Example 4.3.2.

For each of the following homogeneous systems, solve it by using Gauss-Jordan elimination, write its solution as a linear combination, and find its solution space N_A and $\dim(N_A)$.

a. $$\begin{cases} 2x_1 + 4x_2 + 6x_3 = 0 \\ 4x_1 + 5x_2 + 6x_3 = 0 \\ 3x_1 + x_2 - 2x_3 = 0. \end{cases}$$

b. $$\begin{cases} x_1 + 2x_2 - x_3 = 0 \\ 3x_1 - 3x_2 + 2x_3 = 0 \\ -x_1 - 11x_2 + 6x_3 = 0. \end{cases}$$

c. $$\begin{cases} x_1 - 2x_2 - 3x_3 + x_4 = 0 \\ x_2 - 2x_3 + 2x_4 = 0. \end{cases}$$

Solution

a.

$$(A \mid \vec{b}) = \left(\begin{array}{ccc|c} 2 & 4 & 6 & 0 \\ 4 & 5 & 6 & 0 \\ 3 & 1 & -6 & 0 \end{array}\right) \xrightarrow{R_1\left(\frac{1}{2}\right)} \left(\begin{array}{ccc|c} 1 & 2 & 3 & 0 \\ 4 & 5 & 6 & 0 \\ 3 & 1 & -2 & 0 \end{array}\right) \underset{R_1(-3)+R_3}{\overset{R_1(-4)+R_2}{\longrightarrow}}$$

$$\left(\begin{array}{ccc|c} 1 & 2 & 3 & 0 \\ 0 & -3 & -6 & 0 \\ 0 & -5 & -11 & 0 \end{array}\right) \xrightarrow{R_2\left(-\frac{1}{3}\right)} \left(\begin{array}{ccc|c} 1 & 2 & 3 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & -5 & -11 & 0 \end{array}\right) \xrightarrow{R_2(5)+R_3}$$

$$\left(\begin{array}{ccc|c} 1 & 2 & 3 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & -1 & 0 \end{array}\right) \xrightarrow{R_3(-1)} \left(\begin{array}{ccc|c} 1 & 2 & 3 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 1 & 0 \end{array}\right) \underset{R_3(-3)+R_1}{\overset{R_3(-2)+R}{\longrightarrow}}$$

$$\left(\begin{array}{ccc|c} 1 & 2 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{array}\right) \xrightarrow{R_2(-2)+R_1} \left(\begin{array}{ccc|c} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{array}\right) = (D \mid \vec{0}).$$

The system corresponding to the augmented matrix $(D \mid \vec{0})$ only has the zero solution (0, 0, 0). Hence, the original system has only the zero solution. The solution space is $N_A = \left\{\vec{0}\right\}$. Because $r(A) = 3,\ \text{null}(A) = 3 - 3 = 0$. Hence, $\dim(N_A) = 0$.

$$(A\,|\,\vec{b}) = \left(\begin{array}{ccc|c} 1 & 2 & -1 & 0 \\ 3 & -3 & 2 & 0 \\ -1 & -11 & 6 & 0 \end{array}\right) \xrightarrow[R_1(1)+R_3]{R_1(-3)+R_2} \left(\begin{array}{ccc|c} 1 & 2 & -1 & 0 \\ 0 & -9 & 5 & 0 \\ 0 & -9 & 5 & 0 \end{array}\right)$$

b.

$$\xrightarrow{R_2(-1)+R_3} \left(\begin{array}{ccc|c} 1 & 2 & -1 & 0 \\ 0 & -9 & 5 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right) \xrightarrow{R_2(-\frac{1}{9})}$$

$$\left(\begin{array}{ccc|c} 1 & 2 & -1 & 0 \\ 0 & 1 & -\frac{5}{9} & 0 \\ 0 & 0 & 0 & 0 \end{array}\right) \xrightarrow{R_2(-2)+R_1} \left(\begin{array}{ccc|c} 1 & 0 & \frac{1}{9} & 0 \\ 0 & 1 & -\frac{5}{9} & 0 \\ 0 & 0 & 0 & 0 \end{array}\right) = (D\,|\,\vec{0}).$$

The system corresponding to the augmented matrix $(D\,\big|\,\vec{0})$ is

$$\begin{cases} x_1 + \frac{1}{9}x_3 & = & 0 \\ x_2 - \frac{5}{9}x_3 & = & 0. \end{cases}$$

This implies $x_1 = -\frac{1}{9}x_3$ and $x_2 = \frac{5}{9}x_3$. Let $x_3 = t \in \mathbb{R}$. Then $x_1 = -\frac{1}{9}t$ and $x_2 = \frac{5}{9}t$. We write the solution as a linear combination as follows:

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} -\frac{1}{9}t \\ \frac{5}{9}t \\ t \end{pmatrix} = t\begin{pmatrix} -\frac{1}{9} \\ \frac{5}{9} \\ 1 \end{pmatrix}.$$

Let $\vec{v_1} = (-\frac{1}{9}, \frac{5}{9}, 1)$. Then the solution space is

$$N_A = \left\{t\vec{v_1} : t \in \mathbb{R}\right\} = \operatorname{span}\{\vec{v_1}\}.$$

Because $r(A) = 2$, $\operatorname{null}(A) = 3 - 2 = 1$. Hence, $\dim(N_A) = 1$.

c. $(A\,|\,\vec{b}) = \left(\begin{array}{cccc|c} 1 & -2 & -3 & 1 & 0 \\ 0 & 1 & -2 & 2 & 0 \end{array}\right) \uparrow \xrightarrow{R_2(2)+R_1} \left(\begin{array}{cccc|c} 1 & 0 & -7 & 5 & 0 \\ 0 & 1 & -2 & 2 & 0 \end{array}\right).$

The system corresponding to the last augmented matrix is

$$\begin{cases} x_1 - 7x_3 + 5x_4 & = & 0 \\ x_2 - 2x_3 + 2x_4 & = & 0. \end{cases}$$

This implies $x_1 = 7x_3 - 5x_4$ and $x_2 = 2x_3 - 2x_4$. Let $x_3 = s$ and $x_4 = t$ for $s, t \in \mathbb{R}$. Then $x_1 = 7s - 5t$ and $x_2 = 2s - 2t$. We write the solution as a linear combination of basis vectors as follows:

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 7s - 5t \\ 2s - 2t \\ s \\ t \end{pmatrix} = s\begin{pmatrix} 7 \\ 2 \\ 1 \\ 0 \end{pmatrix} + t\begin{pmatrix} -5 \\ -2 \\ 0 \\ 1 \end{pmatrix}.$$

Let $\overrightarrow{v_1} = (7,\ 2,\ 1,\ 0)$ and $\overrightarrow{v_2} = (-5,\ -2,\ 0,\ 1)$. Then

$$N_A = \left\{s\overrightarrow{v_1} + t\overrightarrow{v_2} : s,\ t \in \mathbb{R}\right\} = \operatorname{span}\left\{\overrightarrow{v_1},\ \overrightarrow{v_2}\right\}.$$

Because $r(A) = 2,\ \operatorname{null}(A) = 4 - 2 = 2$ and $\dim(N_A) = 2.$

Example 4.3.3.

There are three types of food provided to a lake to support three species of fish. Each fish of Species 1 weekly consumes an average of one unit of Food 1, two units of Food 2, and one unit of Food 3. Each fish of Species 2 consumes, each week, an average of two units of Food 1, five units of Food 2, and 3 units of Food 3. The average weekly consumption for a fish of Species 3 is two units of Food 1, two unit of Food 2, and six units of Food 3. Each week, 10000 units of Food 1, 15000 units of Food 2, and 25000 units of Food 3 are supplied to the lake. Assuming that all food is eaten, how many fish of each species can coexist in the lake?

Solution

We denote by $x_1,\ x_2,$ and x_3 the number of fish of the three species in the lake. Using the information in the problem, we see that x_1 fish of Species 1 consume x_1 units of Food 1, x_2 fish of Species 2 consume $2x_2$ units of Food 1, and x_3 fish of Species 3 consume $2x_3$ units of Food 1. Hence, $x_1 + 2x_2 + 2x_3 = 10000 = \text{total}$ weekly supply of Food 1. Similarly, we can establish the equations for the other two foods. This leads to the following system

$$\begin{cases} x_1 + 2x_2 + 2x_3 = 10000 \\ 2x_1 + 5x_2 + 2x_3 = 15000 \\ 2x_1 + 3x_2 + 6x_3 = 25000. \end{cases}$$

We solve the above system by using Gauss-Jordan elimination.

$$(A \mid \vec{b}) = \left(\begin{array}{ccc|c} 1 & 2 & 2 & 10000 \\ 2 & 5 & 2 & 15000 \\ 2 & 3 & 6 & 25000 \end{array}\right) \underset{R_1(-2)+R_3}{\overset{R_1(-2)+R_2}{\longrightarrow}} \left(\begin{array}{ccc|c} 1 & 2 & 2 & 10000 \\ 0 & 1 & -2 & -5000 \\ 0 & -1 & 2 & 5000 \end{array}\right)$$

$$\overset{R_2(1)+R_3}{\longrightarrow} \left(\begin{array}{ccc|c} 1 & 2 & 2 & 10000 \\ 0 & 1 & -2 & -5000 \\ 0 & 0 & 0 & 0 \end{array}\right) \overset{R_2(-21)+R_1}{\longrightarrow}$$

$$\left(\begin{array}{ccc|c} 1 & 0 & 6 & 2000 \\ 0 & 1 & -2 & -5000 \\ 0 & 0 & 0 & 0 \end{array}\right).$$

The system of linear equations corresponding to $(D \mid \vec{d})$ is

$$\begin{cases} x_1 + 6x_3 &= 20000 \\ x_2 - 2x_3 &= -5000, \end{cases}$$

where $x_1,\ x_2$ are basic variables and x_3 is a free variable. Let $x_3 = t \in \mathbb{R}$. Then

$$\begin{cases} x_1 &= 20000 - 6t \\ x_2 &= -5000 + 2t \\ x_3 &= t, \end{cases}$$

which is a solution of the system.

Noting that the number of fish must be nonnegative, we must have $x_1 \geq 0,\ x_2 \geq 0,$ and $x_3 \geq 0$. Hence,

$$\begin{cases} x_1 &= 2000 - 6t \geq 0 \\ x_2 &= -5000 + 2t \geq 0 \\ x_3 &= t \geq 0. \end{cases}$$

Solving the above system of inequalities, we get $2500 \leq t \leq 20000/6 \approx 3333.3$. Because *t* is an integer, so we take $2500 \leq t \leq 3333$. This means that the population that can be supported by the lake with all food consumed is

$$\begin{cases} x_1 = 20000 - 6x_3 \\ x_2 = -5000 + 2x_3 \\ 2500 \leq x_3 \leq 3333. \end{cases}$$

For example, if we take $x_3 = 2500,$ then $x_1 = 20000 - 6(2500) = 5000$ and $x_2 = -5000 + 2(2500) = 0$. This means that if there are 2500 fish of Species 3, then there are no fish of Species 2. Otherwise, the lake cannot support all three species with the current supply of food. If we take $x_3 = 3000,$ then $x_1 = 20000 - 6(3000) = 2000$ and $x_2 = -5000 + 2(3000) = 1000.$ This provides the suggestion that with the current supply of food, if 2000 fish of Species 1, 1000 fish of Species 2, and 3000 fishes of Species 3 can coexist in the lake.

Example 4.3.4 (Leontief input-output model with two industries)

In an economic system with steel and automobile industries, let x_1 and x_2 denote the outputs of the steel and automobile industries, respectively, and b_1 and b_2 represent the external demand from outside the steel and automobile industries. Let a_{11} represent the internal demand placed on the steel industry by itself, that is, the steel industry needs a_{11} units of the output of the steel industry to produce one unit of its output. Thus, $a_{11}x_1$ is the total amount the steel industry needs from itself. Let a_{12} represent the internal demand placed on the steel industry by the automobile industry, that is, the automobile industry needs a_{12} units of the output of the steel industry to produce one unit of output of the automobile industry. $a_{12}x_2$ is the total amount the automobile industry needs from the steel industry. Thus, the total demand on the steel industry is

(4.3.4)

$$a_{11}x_1 + a_{12}x_2 + b_1.$$

Let a_{21} represent the internal demand placed on the automobile industry by the steel industry, that is, the steel industry needs a_{21} units of the output of the automobile industry to produce one unit of output of the steel industry. Hence, $a_{21}x_1$ is the total amount the steel industry needs from the automobile industry. Let a_{22} represent the internal demand placed on the automobile industry by itself, that is, the automobile industry needs a_{22} units of the output of the automobile industry to produce one unit of its output. Hence, $a_{22}x_2$ is the total amount the automobile industry needs from itself. Thus, the total demand on the automobile industry is

(4.3.5)

$$a_{21}x_1 + a_{22}x_2 + b_2.$$

Now we assume that the output of each industry equals its demand, that is, there is no overproduction. Then, by hypotheses, **(4.3.4)** and **(4.3.5)** , we obtain the following system

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + b_1 = x_1 \\ a_{21}x_1 + a_{22}x_2 + b_2 = x_2, \end{cases}$$

or equivalently,

(4.3.6)

$$\begin{cases} (1 - a_{11})x_1 - a_{12}x_2 = b_1, \\ -a_{21}x_1 + (1 - a_{22})x_2 = b_2. \end{cases}$$

Example 4.3.5.

In the economic system given in **Example 4.3.4** , suppose that the external demands are 100 and 200, respectively. Suppose that $a_{11} = \frac{9}{10}$, $a_{12} = \frac{1}{2}$, $a_{21} = 110$, and $a_{22} = 15$. Find the output in each industry such that supply exactly equals demand.

Solution

By **(4.3.6)** , we obtain

(4.3.7)

$$\begin{cases} \left(1 - \frac{9}{10}\right)x_1 - \frac{1}{2}x_2 = 10, \\ -\frac{1}{10}x_1 + \left(1 - \frac{1}{5}\right)x_2 = 20. \end{cases}$$

$$(A \mid \vec{b}) = \left(\begin{array}{cc|c} \frac{1}{10} & -\frac{1}{2} & 100 \\ -\frac{1}{10} & \frac{4}{5} & 200 \end{array}\right) \underset{R_2(10)}{\overset{R_1(10)}{\longrightarrow}} \left(\begin{array}{cc|c} 1 & -5 & 100 \\ -1 & 8 & 200 \end{array}\right) \xrightarrow{R_1(1)+R_2}$$

$$\left(\begin{array}{cc|c} 1 & -5 & 100 \\ 0 & 3 & 300 \end{array}\right) \xrightarrow{R_2(\frac{1}{3})} \left(\begin{array}{cc|c} 1 & -5 & 100 \\ 0 & 1 & 100 \end{array}\right) \xrightarrow{R_2(5)+R_1}$$

$$\left(\begin{array}{cc|c} 1 & 0 & 600 \\ 0 & 1 & 100 \end{array}\right).$$

Hence, $x_1 = 600$ and $x_2 = 100$. The outputs needed for supply to equal demand are 600 and 100 for the steel and automobile industries, respectively.

Exercises

1. Solve each of the following systems using Gauss-Jordan elimination.

a. $\begin{cases} x_1 - x_2 + x_3 = 2 \\ 2x_1 - 3x_2 + x_3 = -1 \end{cases}$

b. $$\begin{cases} 3x_1 - x_2 + 3x_3 = -1 \\ -4x_1 + 2x_2 - 4x_3 = 2 \end{cases}$$

c. $$\begin{cases} x_1 + 3x_2 - x_3 = 2 \\ 4x_1 - 6x_2 + 6x_3 = 14 \\ -3x_1 - x_2 - 2x_3 = 3 \end{cases}$$

d. $$\begin{cases} x_1 - 2x_2 + 2x_3 = 2 \\ -2x_1 + 3x_2 - 4x_3 = -2 \\ -3x_1 + 4x_2 + 6x_3 = 0 \end{cases}$$

e. $$\begin{cases} x_2 - x_3 = 1 \\ -x_1 + 3x_2 - x_3 = -2 \\ 3x_1 + 3x_2 - 2x_3 = 4 \end{cases}$$

f. $$\begin{cases} x_1 + x_2 + 2x_3 = 1 \\ -2x_1 - 4x_2 - 5x_3 = -1 \\ 3x_1 + 6x_2 + 5x_3 = 0 \end{cases}$$

g. $$\begin{cases} 2x_2 + 3x_3 - 2x_4 = 3 \\ 2x_1 + 3x_3 + 2x_4 = 1 \\ -3x_1 + x_2 - 2x_3 = 3 \end{cases}$$

h. $\begin{cases} x_1 - x_2 + 2x_3 = -1 \\ -x_1 + 3x_2 - 4x_3 = -3 \\ x_2 - x_3 = -2 \\ x_1 - 2x_2 + 3x_3 = 1 \end{cases}$

2. For each of the following homogeneous systems, solve it by using Gauss-Jordan elimination, write its solution as a linear combination, and find its solution space N_A and $\dim(N_A)$.

   a. $\begin{cases} x + 2y - z = 0 \\ 2x - y + 3y = 0 \end{cases}$

   b. $\begin{cases} 2x - y + 3z = 0 \\ 4x - 2y + 6z = 0 \\ -6x + 3y - 9z = 0 \end{cases}$

   c. $\begin{cases} x_1 + 4x_2 + 6x_3 = 0 \\ 2x_1 + 6x_2 + 3x_3 = 0 \\ -3x_1 + x_2 + 8x_3 = 0. \end{cases}$

   d. $\begin{cases} -x_1 + 2x_2 + x_3 = 0 \\ 3x_1 - 3x_2 - 9x_3 = 0 \\ -2x_1 - x_2 + 12x_3 = 0. \end{cases}$

e. $$\begin{cases} 2x_1 + 2x_2 - x_3 + x_5 = 0 \\ -x_1 - x_2 + 2x_3 - 3x_4 + x_5 = 0 \\ x_1 + x_2 - 2x_3 - x_5 = 0 \\ x_3 + x_4 + x_5 = 0 \end{cases}$$

f. $$\begin{cases} x_1 + x_2 - 2x_3 = 0 \\ -3x_1 + 2x_2 - x_3 = 0 \\ -2x_1 + 3x_2 - 3x_3 = 0 \\ 4x_1 - x_2 - x_3 = 0. \end{cases}$$

4.4 Inverse matrix method and Cramer's rule

Both Gaussian elimination and Gauss-Jordan elimination can be used to solve a system of linear equations with an $m \times n$ coefficient matrix, where *m* and *n* do not necessarily equal. Of course, they can be used to solve the following system

(4.4.1)

$$A\vec{X} = \vec{b},$$

where *A* is an $n \times n$ matrix. However, if *A* is an invertible matrix, then there are two other methods that can be used to solve such systems. One method is to use the inverse matrix of *A* to solve the system, called the inverse matrix method, and another is to use the determinant of *A* to solve the system, called Cramer's rule.

We start with the inverse matrix method.

Theorem 4.4.1.

If A is an invertible $n \times n$ matrix, then for each $\vec{b} \in \mathbb{R}^n$, the system **(4.4.1)** *has the unique solution*

(4.4.2)

$$\vec{X} = A^{-1}\vec{b}.$$

Proof

Because *A* is invertible, A^{-1} exists and $\vec{X}$ given in **(4.4.2)** is defined. With $\vec{X}$ given in **(4.4.2)**, because

$$A\vec{X} = A(A^{-1}\vec{b}) = (AA^{-1})\vec{b} = I_n\vec{b} = \vec{b},$$

where I_n is the identity matrix, and $\vec{X}$ given in **(4.4.2)** is a solution of **(4.4.1)**. Now we assume that **(4.4.1)** has a solution $\vec{Y}$, that is, $A\vec{Y} = \vec{b}$. Multiplying both sides of $A\vec{Y} = \vec{b}$ by A^{-1}, we obtain

$$A^{-1}(A\vec{Y}) = A^{-1}\vec{b}.$$

This implies $\vec{Y} = A^{-1}\vec{b}$ and $\vec{Y} = \vec{X}$. Hence, **(4.4.1)** has only one solution $\vec{X}$ given in **(4.4.2)**.

Example 4.4.1.

Solve the following system using the inverse matrix method

$$\begin{cases} x_1 + 2x_2 + x_3 = 2 \\ 2x_1 + 5x_2 + 2x_3 = 6 \\ x_1 - x_3 = 2 \end{cases}$$

Solution

From the system, we obtain the coefficient matrix

$$A = \begin{pmatrix} 1 & 2 & 1 \\ 2 & 5 & 2 \\ 1 & 0 & -1 \end{pmatrix} \text{ and } \vec{b} = \begin{pmatrix} 2 \\ 6 \\ 2 \end{pmatrix}.$$

We find A^{-1} by using the method given in **Section 2.7**.

$$(A\,|\,I_3) = \left(\begin{array}{ccc|ccc} 1 & 2 & 1 & 1 & 0 & 0 \\ 2 & 5 & 2 & 0 & 1 & 0 \\ 1 & 0 & -1 & 0 & 0 & 1 \end{array}\right) \underset{R_1(-1)+R_3}{\overset{R_i(-2)+R_2}{\longrightarrow}} \left(\begin{array}{ccc|ccc} 1 & 2 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & -2 & 1 & 0 \\ 0 & -2 & -2 & -1 & 0 & 1 \end{array}\right)$$

$$\overset{R_2(2)+R_3}{\longrightarrow} \left(\begin{array}{ccc|ccc} 1 & 2 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & -2 & 1 & 0 \\ 0 & 0 & -2 & -5 & 2 & 1 \end{array}\right) \overset{R_3(-\frac{1}{2})}{\longrightarrow} \left(\begin{array}{ccc|ccc} 1 & 2 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & -2 & 1 & 0 \\ 0 & 0 & 1 & \frac{5}{2} & -1 & -\frac{1}{2} \end{array}\right)$$

$$\overset{R_3(-1)+R_1}{\longrightarrow} \left(\begin{array}{ccc|ccc} 1 & 2 & 0 & -\frac{3}{2} & 1 & \frac{1}{2} \\ 0 & 1 & 0 & -2 & 1 & 0 \\ 0 & 0 & 1 & \frac{5}{2} & -1 & -\frac{1}{2} \end{array}\right) \overset{R_2(-2)+R_1}{\longrightarrow}$$

$$\left(\begin{array}{ccc|ccc} 1 & 0 & 0 & \frac{5}{2} & -1 & \frac{1}{2} \\ 0 & 1 & 0 & -2 & 1 & 0 \\ 0 & 0 & 1 & \frac{5}{2} & -1 & -\frac{1}{2} \end{array}\right) = (I_3\,|\,A^{-1}).$$

By **(4.4.2)** ,

$$\vec{X} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = A^{-1}\vec{b} = \begin{pmatrix} \frac{5}{2} & -1 & \frac{1}{2} \\ -2 & 1 & 0 \\ \frac{5}{2} & -1 & -\frac{1}{2} \end{pmatrix}\begin{pmatrix} 2 \\ 6 \\ 2 \end{pmatrix} = \begin{pmatrix} 0 \\ 2 \\ -2 \end{pmatrix}$$

is the solution of the system.

By **Theorems 4.4.1** and **3.4.1**, we have

Corollary 4.4.1.

If either $|A| \neq 0$ *or* $r(A) = n,$ *then* **(4.4.1)** *has a unique solution for each* $\vec{b} \in \mathbb{R}^n$.

Example 4.4.2.

Find all the numbers $a \in \mathbb{R}$ such that the following system has a unique solution for each $(b_1, b_2) \in \mathbb{R}^2$.

$$\begin{cases} x + y = b_1 \\ 2x + ay = b_2 \end{cases}$$

Solution

Because $|A| = \begin{vmatrix} 1 & 1 \\ 2 & a \end{vmatrix} = a - 2,$ we have if $a \neq 2,$ $|A| \neq 0$. By **Corollary 4.4.1**, if $a \neq 2,$ then the system has a unique solution for each $(b_1, b_2) \in \mathbb{R}^2$.

Example 4.4.3.

Show that the following system has a unique solution for each $(b_1,\ b_2,\ b_3) \in \mathbb{R}^3$.

$$\begin{cases} x_1 + 2x_2 + x_3 = b_1 \\ 2x_1 + 5x_2 + 2x_3 = b_2 \\ x_1 - x_3 = b_3 \end{cases}$$

Solution

We find the rank of *A*.

$$A = \begin{pmatrix} 1 & 2 & 1 \\ 2 & 5 & 2 \\ 1 & 0 & -1 \end{pmatrix} \xrightarrow[R_1(-1)+R_3]{R_{(-2)}+R_2} \begin{pmatrix} 1 & 2 & 1 \\ 0 & 1 & 0 \\ 0 & -2 & -2 \end{pmatrix} \xrightarrow{R_2(2)+R_3}$$

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

Hence, $r(A) = 3$. By **Corollary 4.4.1**, the system has a unique solution for each $(b_1,\ b_2,\ b_3) \in \mathbb{R}^3$.

By **Corollary 4.4.1**, we see that if $|A| \neq 0,$ then the system **(4.4.1)** has a unique solution. The Cramer's rule we study below gives the expression of the unique solution by using determinants.

Theorem 4.4.2. (Cramer's rule)

Let A be an $n \times n$ matrix given in **(3.2.1)**. *If $|A| \neq 0,$ then the unique solution of* **(4.4.1)** *is*

(4.4.3)

$$x_1 = \frac{\left|A_1\right|}{|A|},\ \ x_2 = \frac{\left|A_2\right|}{|A|},\ \cdots,\ \ x_n = \frac{\left|A_n\right|}{|A|},$$

where A_i is obtained by replacing the ith column of A by $\vec{b}$, namely,

(4.4.4)

$$A_i = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1(i-1)} & b_1 & a_{1(i+1)} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2(i-1)} & b_2 & a_{2(i+2)} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{n(i-1)} & b_n & a_{n(i+1)} & \cdots & a_{nn} \end{pmatrix}.$$

Proof

By **Theorem 3.2.1** and **(4.4.4)**, we have

(4.4.5)

$$|A_i| = \sum_{k=1}^{n} b_k A_{kj}.$$

By **(3.3.11)**,

(4.4.6)

$$\operatorname{adj}(A)\vec{b} = \begin{pmatrix} A_{11} & A_{21}\cdots & A_{n1} \\ \vdots & \vdots & \vdots \\ A_{1i} & A_{2i}\cdots & A_{ni} \\ \vdots & \vdots & \vdots \\ A_{1n} & A_{2n}\cdots & A_{nn} \end{pmatrix} \begin{pmatrix} b_1 \\ \vdots \\ b_i \\ \vdots \\ b_n \end{pmatrix} = \begin{pmatrix} \sum_{k=1}^{n} b_k A_{k1} \\ \vdots \\ \sum_{k=1}^{n} b_k A_{ki} \\ \vdots \\ \sum_{k=1}^{n} b_k A_{km} \end{pmatrix}.$$

By **Theorems 3.3.7** and **4.4.1**, the solution of system **(4.4.1)** equals

(4.4.7)

$$\vec{X} = A^{-1}\vec{b} = \frac{1}{|A|}\operatorname{adj}(A)\vec{b} \quad \text{for each } \vec{b} \in \mathbb{R}^n.$$

By **(4.4.5)**, **(4.4.6)**, and **(4.4.7)**, we see that **(4.4.3)** holds.

Example 4.4.4.

Solve each of the following systems by using Cramer's rule.

a. $\begin{cases} 2x - 3y = 7 \\ x + 5y = 1 \end{cases}$

b. $\begin{cases} x_1 + 2x_3 = 6 \\ -3x_1 + 4x_2 + 6x_3 = 30 \\ -x_1 - 2x_2 + 3x_3 = 8 \end{cases}$

Solution

a. *a*)Let

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = \begin{pmatrix} 2 & -3 \\ 1 & 5 \end{pmatrix} \quad \text{and } \vec{b} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} 7 \\ 1 \end{pmatrix}.$$

By **(4.4.4)** , we have

$$A_1 = \begin{pmatrix} b_1 & a_{12} \\ b_2 & a_{22} \end{pmatrix} = \begin{pmatrix} 7 & -3 \\ 1 & 5 \end{pmatrix} \text{ and } A_2 = \begin{pmatrix} a_{11} & b_1 \\ a_{21} & b_2 \end{pmatrix} = \begin{pmatrix} 2 & 7 \\ 1 & 1 \end{pmatrix}.$$

By computation, $\left|A\right| = 13$, $\left|A_1\right| = 38$, and $\left|A_2\right| = -5$. By **Theorem 4.4.2** ,

$$(x, y) = \left(\frac{38}{13}, -\frac{5}{13}\right)$$

is the solution of the system.

b. Let

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 2 \\ -3 & 4 & 6 \\ -1 & -2 & 3 \end{pmatrix} \text{ and } \vec{b} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} = \begin{pmatrix} 6 \\ 30 \\ 8 \end{pmatrix}.$$

By **(4.4.4)**, we obtain

$$A_1 = \begin{pmatrix} b_1 & a_{12} & a_{13} \\ b_2 & a_{22} & a_{23} \\ b_3 & a_{32} & a_{33} \end{pmatrix} = \begin{pmatrix} 6 & 0 & 2 \\ 30 & 4 & 6 \\ 8 & -2 & 3 \end{pmatrix}$$

$$A_2 = \begin{pmatrix} a_{11} & b_1 & a_{13} \\ a_{21} & b_2 & a_{23} \\ a_{31} & b_3 & a_{33} \end{pmatrix} = \begin{pmatrix} 1 & 6 & 2 \\ -3 & 30 & 6 \\ -1 & 8 & 3 \end{pmatrix}$$

$$A_3 = \begin{pmatrix} a_{11} & a_{12} & b_1 \\ a_{21} & a_{22} & b_2 \\ a_{31} & a_{32} & b_3 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 6 \\ -3 & 4 & 30 \\ -1 & -2 & 8 \end{pmatrix}.$$

By computation,

$$|A| = 44, \ \left|A_1\right| = -40, \ \left|A_2\right| = 72, \text{ and } \left|A_3\right| = 152,$$

and

$$x_1 = \frac{|A_1|}{|A|} = \frac{-40}{44} = -\frac{10}{11},\ x_2 = \frac{|A_2|}{|A|} = \frac{72}{44} = \frac{18}{11},\ x_3 = \frac{|A_3|}{|A|} = \frac{152}{44} = \frac{38}{11}.$$

Hence, by **(4.4.3)** , $\left(-\frac{10}{11}, \frac{18}{11}, \frac{38}{11}\right)$ is the solution of the system.

Exercises

1. Solve each of the following systems using the inverse matrix method.

a. $\begin{cases} 3x_1 + 4x_2 = 7 \\ x_1 - x_2 = 14 \end{cases}$

b. $\begin{cases} x_1 + x_2 + x_3 = 2 \\ 2x_1 + 3x_2 + x_3 = 2 \\ x_1 - x_3 = 1 \end{cases}$

c. $\begin{cases} x_1 - 2x_2 + 2x_3 = 1 \\ 2x_1 + 3x_2 + 3x_3 = -1 \\ x_1 + 6x_3 = -4 \end{cases}$

d. $\begin{cases} x_1 + 2x_2 + 3x_3 = 0 \\ 2x_1 + 5x_2 + 3x_3 = 1 \\ x_1 + 8x_3 = -1 \end{cases}$

e. $\begin{cases} x_1 - 2x_3 + x_4 = 0 \\ x_2 + x_3 + 2x_4 = 1 \\ -x_1 - x_2 + x_3 = -1 \\ 2x_1 + 9x_4 = 1 \end{cases}$

2. Show that if $a \neq 3,$ then the following system has a unique solution for $(b_1,\ b_2) \in \mathbb{R}^2$.

$$\begin{cases} x - y = b_1 \\ ax - 3y = b_2 \end{cases}$$

3. For each $(b_1,\ b_2,\ b_3) \in \mathbb{R}^3,$ show that the following system has a unique solution.

$$\begin{cases} x_1 + 2x_3 = b_1 \\ -3x_1 + 4x_2 + 6x_3 = b_2 \\ -x_1 - 2x_2 + 3x_3 = b_3 \end{cases}$$

4. Solve each of the following systems by using Cramer's rule.

a. $\begin{cases} x - 3y = 6 \\ 2x + 5y = -1 \end{cases}$

b. $\begin{cases} x_1 + 2x_3 = 2 \\ -2x_1 + 3x_2 + x_3 = 16 \\ -x_1 - 2x_2 + 4x_3 = 5 \end{cases}$

c. $\begin{cases} x_1 + x_2 - x_3 = 1 \\ 2x_1 + x_2 - x_3 = 0 \\ x_2 - 2x_3 = -2 \end{cases}$

d. $\begin{cases} x_1 + x_2 + x_3 = 0 \\ 2x_1 - x_2 - x_3 = 1 \\ x_2 + 2x_3 = 2 \end{cases}$

e. $\begin{cases} x_1 - x_4 = 0 \\ x_2 + x_3 = 0 \\ x_1 - x_2 = 0 \\ x_3 - x_4 = 1 \end{cases}$

4.5 Consistency of systems of linear equations

In Sections 4.2–4.4, we study the methods to find all the solutions of systems of linear equations. However, in some cases, for example, see Section 4.6, we do not want the exact solutions of the systems but we just want to know whether the systems have solutions, that is, whether the systems are consistent or inconsistent.

In this section, we use ranks of *r*(*A*) and $(A\,|\,\vec{b})$ to determine whether a linear system **(4.1.7)** , that is,

(4.5.1)

$$A\vec{X} = \vec{b},$$

where the coefficient matrix *A* is an $m \times n$ matrix, has a unique solution, infinitely many solutions,, or no solutions without solving the system.

Theorem 4.5.1.

1. $r(A) = r(A\,|\,\vec{b}) = n$ *if and only if the system* **(4.5.1)** *has a unique solution.*
2. $r(A) = r(A\,|\,\vec{b}) < n$ *if and only if the system* **(4.5.1)** *has infinitely many solutions.*
3. $r(A) < r(A\,|\,\vec{b})$ *if and only if the system* **(4.5.1)** *has no solutions.*

Proof

Note that *A* is an $m \times n$ matrix and $(A\,|\,\vec{b})$ is an $m \times (n+1)$ matrix. Use row operations to change $(A\,|\,\vec{b})$ to a row echelon matrix, denoted by $(B\,|\,\vec{c})$. Then *B* is a row echelon matrix of *A* and

$$r(A) = r(B) \quad \text{and } r(A\,|\,\vec{b}) = r(B\,|\,\vec{c}).$$

By **Theorem 4.2.1**, the linear system corresponding to $(B \mid \vec{c})$ is equivalent to **(4.5.1)**. We first prove that the following assertions hold:

i. If $r(A) = r(A \mid \vec{b}) = n,$ then the system **(4.5.1)** has a unique solution. Indeed, if $r(A) = r(A \mid \vec{b}) = n,$ then by **Theorem 2.5.2**, we have $n \le m$ and $r(B) = r(B \mid \vec{c}) = n.$ We write $B = (b_{ij})_{m \times n}.$ This implies that $b_{ii} \neq 0$ for $i = 1,\ 2,\ \cdots,\ n.$ If $m > n,$ then all rows of the matrix $(B \mid \vec{c})$ from $n + 1$ to m are zero rows. Using back-substitution to solve the system $B\vec{X} = \vec{c},$ we obtain only one solution.

ii. If $r(A) = r(A \mid \vec{b}) < n,$ then the system **(4.5.1)** has infinitely many solutions.

Indeed, because $r(A) = r(A \mid \vec{b}) < n, r(B) = r(B \mid \vec{c}) < n.$ This implies that if $r(B) < m,$ then all rows of the matrix $(B \mid \vec{c})$ from $r(B) + 1$ to *m* are zero rows and the system $B\vec{X} = \vec{c}$ has $n - r(B)$ free variables. Hence, $B\vec{X} = \vec{c}$ has infinitely many solutions.

iii. If $r(A) < r(A \mid \vec{b}),$ then the system **(4.5.1)** has no solutions. In fact, if $r(A) < r(A \mid \vec{b}),$ then $r(B) < r(B \mid \vec{c})$ and $c_j \neq 0,$ where $j = r(B) + 1.$ This implies that the *j*th equation of the system $B\vec{X} = \vec{c}$ is

$$0x_1 + \cdots + 0x_n = c_j,$$

from which we see that the system $B\vec{X} = \vec{c}$ has no solutions.

Note that $r(A) \le r(A \mid \vec{b})$ and $r(A) \le \min\{n,\ m\} \le n.$ Hence, there are only three cases: $r(A) < r(A \mid \vec{b}),$ $r(A) < r(A \mid \vec{b}) = n,$ and $r(A) < r(A \mid \vec{b}) < n,$ from which we see that the converses of *(i)-(iii)* hold.

By **Theorem 4.5.1**, we obtain the following result, which gives the structure of the solutions of **(4.1.4)**.

Corollary 4.5.1.

One of the following assertions must occur.

1. **(4.5.1)** *has a unique solution.*
2. **(4.5.1)** *has infinitely many solutions.*
3. **(4.5.1)** *has no solutions.*

Proof

Because $r(A) \le r(A \mid \vec{b})$ and $r(A) \le n,$ there are only three cases: $r(A) = r(A \mid \vec{b}) = n,$ $r(A) = r(A \mid \vec{b}) < n,$ and $r(A) = r(A \mid \vec{b}).$ The result follows from **Theorem 4.5.1** .

Example 4.5.1.

Use **Theorem 4.5.1** to determine whether the following systems have a unique solution, infinitely many solutions, or no solutions.

a. $\begin{cases} 2x_1 + 4x_2 + 6x_3 = 18 \\ 4x_1 + 5x_2 + 6x_3 = 24 \\ 3x_1 + x_2 - 2x_3 = 4 \end{cases}$

b. $\begin{cases} 2x_1 + 2x_2 - 2x_3 = 4 \\ 3x_1 + 5x_2 + x_3 = -8 \\ -4x_1 - 7x_2 - 2x_3 = 13 \end{cases}$

c. $\begin{cases} x_1 + 2x_2 + 3x_3 = 4 \\ 4x_1 + 7x_2 + 6x_3 = 17 \\ 2x_1 + 5x_2 + 12x_3 = 10 \end{cases}$

d. $\begin{cases} x_1 - x_2 = 1 \\ -2x_1 - 5x_2 = 5 \\ -3x_1 + 10x_2 = -10 \end{cases}$

Solution

a. By the solution of **Example 4.2.5** *a),* we see that the augmented matrix $(A \mid \vec{b})$ is changed to the row echelon matrix

$$(B \mid \vec{c}) = \begin{pmatrix} 1 & 2 & 3 & 9 \\ 0 & 1 & 2 & 4 \\ 0 & 0 & -1 & -3 \end{pmatrix}$$

from which we see that $r(A) = r(B \mid \vec{c}) = 3 = n$. By **Theorem 4.5.1** (1), the system has a unique solution.

b. By the solution of **Example 4.2.5** b), we see that the augmented matrix $(A \mid \vec{b})$ is changed to the row echelon matrix

$$(B \mid \vec{c}) = \begin{pmatrix} 1 & 1 & -1 & 2 \\ 0 & 1 & 2 & -7 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

from which we see that $r(A) = r(B \mid \vec{c}) = 2 < n$. By **Theorem 4.5.1** (2), the system has infinitely many solutions.

c. By the solution of **Example 4.2.5** d), $(A \mid \vec{b})$ is changed to the row echelon matrix

$$(B \mid \vec{c}) = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 0 & -1 & -6 & 1 \\ 0 & 0 & 0 & 3 \end{pmatrix}$$

from which we see that $r(A) = 2 < 3 = r(B \mid \vec{c})$. By **Theorem 4.5.1** (3), the system has no solutions.

d.

$$(A \mid \vec{b}) = \begin{pmatrix} 1 & -1 & 1 \\ -2 & -5 & 5 \\ -3 & 10 & -10 \end{pmatrix} \underset{R_1(-3)+R_3}{\xrightarrow{R_1(2)+R_2}} \begin{pmatrix} 1 & -1 & 1 \\ 0 & -7 & 7 \\ 0 & 7 & -7 \end{pmatrix}$$

$$\xrightarrow{R_2(1)+R_3} \begin{pmatrix} 1 & -1 & 1 \\ 0 & -7 & 7 \\ 0 & 0 & 0 \end{pmatrix}.$$

Hence, we have $r(A) = r(B \mid \vec{c}) = 2 = n \leq 3 = m.$ By **Theorem 4.5.1** (1), the system has a unique solution.

The following result shows that if $n > m,$ then it is impossible for **(4.5.1)** to have a unique solution. Hence, if **(4.5.1)** has a unique solution, then $n \leq m,$ see **Example 4.5.1** (a) and (d).

Corollary 4.5.2.

If $n > m,$ then **(4.5.1)** *either has infinitely many solutions or no solutions.*

Proof

Because $n > m,$ by **Theorem 4.5.1** (1), **(4.5.1)** doesn't have a unique solution. Hence, by **Corollary 4.5.1** , we see that **(4.5.1)** either has infinitely many solutions or no solutions.

The ranks of *A* and $(A \mid \vec{b})$ are not directly involved in the conditions of **Corollary 4.5.2** . Hence, one can easily determine that **(4.5.1)** doesn't have a unique solution by comparing *m* and *n*.

Example 4.5.2.

For each of the following systems, determine whether it has a unique solution.

a. $\begin{cases} x_1 - x_2 - 3x_3 - 4x_4 = 1 \\ -x_1 + x_2 - x_3 = -1 \end{cases}$

b. $\begin{cases} x_1 + x_2 + x_3 = 1 \\ x_1 + x_2 + x_3 = 2 \end{cases}$

Solution

a. Because $m = 2 < 4 = n,$ by **Corollary 4.5.2** , the system doesn't have a unique solution.
b. Because $m = 2 < 3 = n,$ by **Corollary 4.5.2** , the system doesn't have a unique solution.

By **Theorem 4.5.1** , we obtain the following result, which can be used to determine whether a linear system is consistent.

Corollary 4.5.3.

1. **(4.5.1)** *is consistent if and only if* $r(A) = r(A \mid \vec{b})$.
2. **(4.5.1)** *is inconsistent if and only if* $r(A) < r(A \mid \vec{b})$.

Applications of **Corollary 4.5.3** will be given in **Section 4.6**.

Note that a homogeneous system **(4.1.8)** is always consistent because it has a zero solution. Hence, for a homogeneous system, the third case of **Theorem 4.5.1** never occurs. By **Theorem 4.5.1**, we obtain the following result on homogeneous systems.

Theorem 4.5.2.

1. $A\vec{X} = \vec{0}$ has a unique solution if and only if $r(A) = n$.
2. (2)$A\vec{X} = \vec{0}$ has infinitely many solutions if and only if $r(A) = n$.

Remark 4.5.1.

By **Theorem 4.5.2** (2), if $n > m,$ then the homogeneous system $A\vec{X} = \vec{0}$ has infinitely many solutions because by **Theorem 2.5.2**, $r(A) \leq \min\{m,\ n\} = m < n$.

By **Theorems 2.7.6** and **3.4.1**, we obtain the following result for a homogeneous system $A\vec{X}=\vec{0}$, where *A* is an $n\times n$ matrix.

Theorem 4.5.3.

Let A be an $n\times n$ *matrix. Then the following are equivalent.*

1. *The system* $A\vec{X}=\vec{0}$ *only has the zero solution.*
2. $r(A)=n$.
3. $|A|\neq 0$.
4. *A is invertible.*

Example 4.5.3.

Determine whether each of the following systems has a unique solution or infinitely many solutions.

a. $\begin{cases} x_1-2x_2=0 \\ -2x_1+3x_2=0 \\ -3x_1+4x_2=0. \end{cases}$

b. $\begin{cases} x_1+3x_2-x_3=0\\ -x_1+x_2+6x_3=0\\ -3x_1-x_2+2x_3=0. \end{cases}$

c. $\begin{cases} x_1-2x_2-x_3+x_4=0\\ x_2-x_3+2x_4=0. \end{cases}$

d. $\begin{cases} x_1-3x_2+x_3=0\\ -x_1+2x_2-2x_3=0\\ 2x_1-5x_2+3x_3=0\\ 3x_1-7x_2+5x_3=0. \end{cases}$

Solution

a. Because

$$A=\begin{pmatrix}1&-2\\-2&3\\-3&4\end{pmatrix}\xrightarrow[R_1(2)+R_2]{R_1(3)+R_3}\begin{pmatrix}1&-2\\0&-1\\0&-2\end{pmatrix}\xrightarrow{R_2(-2)+R_3}\begin{pmatrix}1&-2\\0&-1\\0&0\end{pmatrix},$$

$r(A)=2$. Note that $m=3$ and $n=2$. Hence, $r(A)=2=n$ and by **Theorem 4.5.2** (1), the system has a unique solution.

b. Because

$$|A|=\begin{vmatrix}1&3&-1\\-1&1&6\\-3&-1&2\end{vmatrix}\begin{matrix}1&3\\-1&1\\-3&-1\end{matrix}=(2-54-1)-(3-6-6)=-44\neq 0,$$

it follows from **Theorem 4.5.3** that the system has a unique solution.

c. Because $m=2<4=n$, it follows from **Theorem 4.5.2** (2) that the system has infinitely many solutions.

d. Because

$$A=\begin{pmatrix}1&-3&1\\-1&2&-2\\2&-5&3\\3&-7&5\end{pmatrix}\xrightarrow[\substack{R_1(1)R_2\\R_1(-2)+R_3}]{R_1(-3)+R_4}\begin{pmatrix}1&-3&1\\0&-1&-1\\0&1&1\\0&2&2\end{pmatrix}\xrightarrow[R_2(1)+R_3]{R_2(2)+R_4}\begin{pmatrix}1&-3&1\\0&-1&-1\\0&0&0\\0&0&0\end{pmatrix},$$

$r(A)=2$. Hence, $n=3<4=m$ and $r(A)=2<3=n$. It follows from **Theorem 4.5.2** (2) that the system has infinitely many solutions.

In Section 4.4, we show that if A is an $n\times n$ matrix and $r(A)=n$, then the system $A\vec{X}=\vec{b}$ has a unique solution for every $\vec{b}\in\mathbb{R}^n$, see **Corollary 4.4.1**. Now we generalize the result to the general system **(4.5.1)**.

Let A be an $m\times n$ matrix. We study the following two problems.

i. How to determine $A(\mathbb{R}^n)=\mathbb{R}^m$, where $A(\mathbb{R}^n)$ is the same as in **(2.2.1)**, that is, how to determine whether $A\vec{X}=\vec{b}$ is consistent for every $\vec{b}\in\mathbb{R}^m$?

ii. If $A(\mathbb{R}^n)\neq\mathbb{R}^m$, that is, $A\vec{X}=\vec{b}$ is not consistent for every $\vec{b}\in\mathbb{R}^m$? then how to find all the vector $\vec{b}$ in $\mathbb{R}^m$ such that $A\vec{X}=\vec{b}$ is consistent? or how to find all the vector $\vec{b}$ in $\mathbb{R}^m$ such that $A\vec{X}=\vec{b}$ is inconsistent?

By **Corollary 4.5.3**, we see that for a given vector $\vec{b}\in\mathbb{R}^m$, $A\vec{X}=\vec{b}$ is consistent if $r(A)=r(A|\vec{b})$. Here we want $A\vec{X}=\vec{b}$ to be consistent for every $\vec{b}\in\mathbb{R}^m$, that is, for every $\vec{b}\in\mathbb{R}^m$, $r(A)=r(A|\vec{b})$.

The following result shows that $r(A)=r(A|\vec{b})$ for every $\vec{b}\in\mathbb{R}^m$ is equivalent to $r(A)=m$.

Theorem 4.5.4.

1. *The following assertions are equivalent.*
   i. *The system* **(4.5.1)** *is consistent for each* $\vec{b}\in\mathbb{R}^m$.
   ii. $A(\mathbb{R}^n)=\mathbb{R}^m$.
   iii. $r(A)=m$.
2. *There exists a vector* $\vec{b}\in\mathbb{R}^m$ *such that the system* **(4.5.1)** *has no solutions in Rn if and only if* $r(A)<m$.

Proof

It is obvious that the results (i) and (ii) are equivalent, and the results (i) and (2) are equivalent. So we only prove that results (i) and (iii) are equivalent. Assume that (i) holds. We change $(A|\vec{b})$ to $(B|\vec{c})$ by using row operations, where *B* is a row echelon matrix of *A*. By **Theorem 4.2.1**, for each $\vec{c}\in\mathbb{R}^m$, the linear system

(4.5.2)

$$B\vec{X}=\vec{c}$$

has at least one solution in $\mathbb{R}^n$. If we assume $r(A)<m$, then $r(B)=r(A)<m$, the rows from $r(A)+1$ to *m* of *B* are zero rows. Let $\vec{c_0}\in\mathbb{R}^m$ be a vector whose $(r(A)+1)$th component is 1 and other components are zero. Then $r(B|\vec{c_0})=r(B)+1$ and $r(B)<r(B|\vec{c_0})$. By **Theorem 4.5.1** (3), we obtain that the following system

$$B\vec{X}=\vec{c_0}$$

has no solutions in $\mathbb{R}^n$. Because the system **(4.5.1)** is equivalent to the system **(4.5.2)**, it follows that the system **(4.5.1)** has no solutions in $\mathbb{R}^n$, which contradicts the fact that system **(4.5.1)** is consistent for each $\vec{b}\in\mathbb{R}^m$. Hence, $r(A)=m$ and (iii) holds.

Conversely, we assume $r(A)=m$. We use row operations to change A to a row echelon matrix B. Then $r(B)=r(A)=m$. For each $\vec{c}\in\mathbb{R}^m$,

$$m=r(B)\le r(B|\vec{c})\le \min\{m, n+1\}=m.$$

Hence, $r(B)=r(B|\vec{c})$. By **Corollary 4.5.3**, **(4.5.2)** is consistent. By **Theorem 4.2.1**, the system **(4.5.1)** is consistent for each $\vec{b}\in\mathbb{R}^m$ and (i) holds.

Theorem 4.5.4 provides a criterion to justify whether a linear system is consistent, that is, whether a linear system has at least one solution.

The following result gives criteria to determine whether a linear system has infinitely many solutions or a unique solution.

Theorem 4.5.5.

1. *The system* **(4.5.1)** *has infinitely many solutions for each* $\vec{b}\in\mathbb{R}^m$ *if and only if* $r(A)=m<n$.
2. *The system* **(4.5.1)** *has a unique solution in* $\mathbb{R}^n$ *for each* $\vec{b}\in\mathbb{R}^m$ *if and only if* $m=n$, $r(A)=n$ *if and only if* $m=n$ *and* $|A|\neq 0$.

Proof

1. Assume that the system **(4.5.1)** has infinitely many solutions for each $\vec{b}\in\mathbb{R}^m$. By **Theorem 4.5.4** (1), $r(A)=m$. If $m=n$, then *A* is an $n\times n$ matrix and is invertible. By **Corollary 4.4.1**, the system **(4.5.1)** with $m=n$ has a unique solution for each $\vec{b}\in\mathbb{R}^m$, which contradicts the fact that the system **(4.5.1)** has infinitely many solutions for each $\vec{b}\in\mathbb{R}^m$. Hence, $m<n$.
   Conversely, assume that $r(A)=m$ and $m<n$. Let $\vec{b}\in\mathbb{R}^m$. We change $(A|\vec{b})$ to $(B|\vec{c})$ by using row operations, where *B* is a row echelon matrix of *A*. Because $r(A)=m$, by **Theorem 4.2.1**,
   (4.5.3)
   $$B\vec{X}=\vec{c}$$
   has at least one solution for every $\vec{c}\in\mathbb{R}^m$. Because $r(A)=m$, we have $r(B)=m$. Because $m<n$, $v(B)=n-m\geq 1$, so the system **(4.5.3)** has at least one free variable. Hence, the system **(4.5.3)** has infinitely many solutions. It follows from **Theorem 4.2.1** that the system **(4.5.1)** has infinitely many solutions.
2. If $m=n$, $r(A)=n$, then by **Theorems 4.5.4**, the system **(4.5.1)** has a unique solution in $\mathbb{R}^n$ for each $\vec{b}\in\mathbb{R}^m=\mathbb{R}^n$.
   Conversely, assume that the system **(4.5.1)** has a unique solution in $\mathbb{R}^n$ for each $\vec{b}\in\mathbb{R}^m$. By **Theorem 4.5.4** (1), $r(A)=m$ and $m\leq n$. Because the system **(4.5.1)** has a unique solution in $\mathbb{R}^n$ for each $\vec{b}\in\mathbb{R}^m$, by the above result (1), $m=n$; otherwise, if $m<n$, then the system **(4.5.1)** would have infinitely many solutions in $\mathbb{R}^n$ for each $\vec{b}\in\mathbb{R}^m$, a contradiction.

Corollary 4.5.2 shows that if **(4.5.1)** has a unique solution, then $n \le m$. But **Theorem 4.5.5** (2) shows that if the system **(4.5.1)** has a unique solution in $\mathbb{R}^n$ for each $\vec{b} \in \mathbb{R}^m$, then A must be a square matrix. In other words, if A is an $m \times n$ matrix with $m \ne n$, then there exists a vector $\vec{b} \in \mathbb{R}^m$ such as $A\vec{X} = \vec{b}$ has infinitely many solutions or has no solutions.

Example 4.5.4.

Consider the following system

$$\begin{cases} x_1 - x_2 + x_3 = b_1 \\ 3x_1 - 2x_2 - x_3 = b_2 \end{cases}$$

1. Determine if the system is consistent for all $b_1, b_2 \in \mathbb{R}$.
2. Determine if the system has infinitely many solutions for all $b_1, b_2 \in \mathbb{R}$.

Solution

Because

$$A = \begin{pmatrix} 1 & -1 & 1 \\ 3 & -2 & -1 \end{pmatrix} \xrightarrow{R_1(-3)+R_2} \begin{pmatrix} 1 & -1 & 1 \\ 0 & 1 & -4 \end{pmatrix},$$

$r(A) = 2$. Because A is a 2×3 matrix, $m = 2$ and $n = 3$.

1. Because $r(A) = 2 = m$, by **Theorem 4.5.4** (1), the system is consistent for all $b_1, b_2 \in \mathbb{R}$.
2. Because $r(A) = 2 = m$ and $m = 2 < 3 = n$, by **Theorem 4.5.5** (1), the system has infinitely many solutions for all $b_1, b_2 \in \mathbb{R}$.

Example 4.5.5.

For each of the following systems, determine whether the system is consistent for all $b_1, b_2, b_3 \in \mathbb{R}$. If not, find conditions on b_1, b_2, b_3 such that the system is consistent.

a. $\begin{cases} x_1 - x_2 + 2x_3 + x_4 = b_1 \\ 2x_1 + 4x_2 - 4x_3 - x_4 = b_2 \\ 3x_1 + 3x_2 - 2x_3 = b_3 \end{cases}$

b. $\begin{cases} x_1 - x_2 + 2x_3 = b_1 \\ -3x_1 + 2x_2 + x_3 = b_2 \\ 4x_1 - 3x_2 + x_3 = b_3. \end{cases}$

c. $\begin{cases} x_1 + x_2 + 2x_3 = b_1 \\ x_1 + x_3 = b_2 \\ 3x_1 + 4x_2 + 7x_3 = b_3 \end{cases}$

Solution

a. (4.5.4)

0%

(A|b→)=
(1−121b124−4−1b233−20b3)→R1(−3)+R3R1(−2)+R2(1−121b106−8−3b2−2b106−8−3b3−3b1)→R2(−1)+R3(1−121b106−8−3b2−2

From the left side of the last matrix in **(4.5.4)**, we have r(A)=2<3=m. By **Theorem 4.5.4** (2), the system is not consistent for all b1, b2, b3∈ℝ.
To choose b1, b2, b3∈ℝ such that the system with these b1, b2, b3 are consistent, by **Corollary 4.5.3**, we need r(A)=(A|b→). By **(4.5.4)**, we see that if b3−b2−b1=0, then r(A)=(A|b→). Hence, the condition on b1, b2, b3 is b3−b2−b1=0, that is, when b3=b1+b2, the system is consistent.

b. (4.5.5)

(A|b→)=
(1−12b1−321b24−31b3)→R1(−4)+R3R1(3)+R2(1−12b10−17b2+3b101−7b3−4b1)→R2(1)+R3(1−12b10−17b2+3b1000b3+b2−b1).

From the left side of the last matrix in **(4.5.5)**, we have r(A)=2<3=m. It follows from **Theorem 4.5.4** (2) that the system is not consistent for all b1, b2, b3∈ℝ. By **(4.5.5)**, we see that if b3+b2−b1=0, then r(A)=2=(A|b→). Hence, when b3+b2−b1≠0, the system is consistent.

c. (A|b→)=
(112b1101b2347b3)→R1(−3)+R3R1(−1)+R2(112b10−1−1b2−b1011b3−3b1)→R2(1)+R3(112b1011b1−b2000−4b1+b2+b3)=
(B|c→).

By **Theorem 4.5.1**, if −4b1+b2+b3=0, the system is consistent.

Exercises

1. For each of the following systems, determine whether it has a unique solution, infinitely many solutions, or no solutions.
   a. { x1−2x2+3x3=62x1+x2+4x3=5−3x1+x2−2x3=3
   b. { x1+x2−2x3=32x1+3x2−x3=−4−2x1−3x2x3=4
   c. { x1+2x2+3x3=44x1+7x2+6x3=172x1+5x2+12x3=7
   d. { x1−2x2−x3+2x4=0−x2−2x3+x4=0
   e. { x1−x2+3x3=2−2x1+2x2−6x3=52x1−2x2+6x3=4
   f. { x1−x2−3x3−4x4=1−x1+x2−x3=−1

2. For each of the following systems, determine if it has a unique solution.

a. { 2x1−3x2−x3−5x4=−1−3x1+6x2−2x3+x4=3x1+x4=−2

b. $\begin{cases} x_1-3x_2+2x_3=-1 \\ x_1+4x_2-x_3=2 \end{cases}$

3. For each of the following homogeneous systems, determine whether it has a unique solution or infinitely many solutions.
   a. $\begin{cases} x_1-2x_2-x_3=0 \\ -x_1+3x_2+2x_3=0 \\ -x_1+x_2=0 \\ x_2+x_3=0. \end{cases}$
   b. $\begin{cases} x_1+x_2-2x_3=0 \\ 2x_1-x_2+4x_3=0 \\ -x_1-2x_2-3x_3=0. \end{cases}$
   c. $\begin{cases} x_1-x_2+x_3=0 \\ 2x_2-4x_3+x_3=0. \end{cases}$
   d. $\begin{cases} x_1-x_2+2x_3=0 \\ -2x_1-2x_2+x_3=0 \\ -x_1-3x_2+3x_3=0 \\ 3x_1+x_2+x_3=0. \end{cases}$

4. Consider the following system

$$\begin{cases} 6x_1-4x_2=b_1 \\ 3x_1-2x_2=b_2 \end{cases}$$

   1. Find b_1 and b_2 such that the system is consistent.
   2. Find b_1 and b_2 such that the system is inconsistent.
   3. Determine whether the system with $\vec{b}=(4, 1)^T$ is inconsistent.

5. Determine if the following system is consistent for all $b_1, b_2, b_3 \in \mathbb{R}$.

$$\begin{cases} x_1-x_2+2x_3+x_4=b_1 \\ 2x_1+4x_2-4x_3=b_2 \\ -3x_1+4x_3-x_4=b_3 \end{cases}$$

6. Find conditions on b_1, b_2, b_3 such that the following system is consistent.

$$\begin{cases} x_1-x_2+2x_3=b_1 \\ 2x_1+4x_2-4x_3=b_2 \\ -3x_1-2x_3=b_3 \end{cases}$$

7. Determine whether the following system has infinitely many solutions for all $b_1, b_2 \in \mathbb{R}$.

$$\begin{cases} x_1-2x_2+2x_3=b_1 \\ -3x_1+3x_2+x_3=b_2 \end{cases}$$

4.6 Linear combination, spanning space and consistency

From **Theorem 4.1.1**, we see three equivalent assertions about linear systems, linear combinations and spanning spaces. By **Examples 1.4.2**, **1.4.4**, **4.2.1**, **4.2.2**, and **4.2.3**, we see that we need to solve the system $A\vec{X} = \vec{b}$ to determine whether one of the three assertions holds. However, **Corollary 4.5.3** provides another method to justify whether these assertions hold by comparing the two ranks, *r(A)* and $r(A \mid \vec{b})$, without solving the system $A\vec{X} = \vec{b}$

Combining **Theorem 4.1.1** and **Corollary 4.5.3**, we obtain the following criteria.

Theorem 4.6.1.

Let $S := \left\{\overrightarrow{a_1}, \cdots, \overrightarrow{a_n}\right\}$ *be the same as in* $A = (\overrightarrow{a_1}, \cdots, \overrightarrow{a_n})$ *be the same as in* **(4.1.5)**, *and* $\vec{b} = (b_1, b_2, \cdots, b_m)^T$. *Then the following assertions are equivalent.*

1. $A\vec{X} = \vec{b}$ *is consistent.*
2. $\vec{b}$ *is a linear combination of S.*
3. $\vec{b} \in \operatorname{span} S$.
4. $r(A) = r(A \mid \vec{b})$.

By **Theorem 4.6.1** (4), we see that if $\vec{b} = \vec{0}$, then $r(A) = r(A \mid \vec{0}) = r(A \mid \vec{b})$. It follows that $\vec{0}$ is a linear combination of S. This provides another proof of **Theorem 1.4.1**.

Example 4.6.1.

Let $S = \left\{ \overrightarrow{a_1}, \overrightarrow{a_2}, \overrightarrow{a_3} \right\}$ and $A = (\overrightarrow{a_1}, \overrightarrow{a_2}, \overrightarrow{a_3})$, where

$$\overrightarrow{a_1} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \quad \overrightarrow{a_2} = \begin{pmatrix} -2 \\ 1 \\ -1 \end{pmatrix}, \quad \overrightarrow{a_3} = \begin{pmatrix} 3 \\ 4 \\ 7 \end{pmatrix}, \quad \vec{b} = \begin{pmatrix} 6 \\ 5 \\ 11 \end{pmatrix}.$$

1. Determine whether $A\vec{X} = \vec{b}$ is consistent, where $\vec{X} = (x_1, x_2, x_3)^T$.
2. Determine whether the vector $\vec{b}$ is a linear combination of *S*.
3. Determine whether the vector $\vec{b} \in \operatorname{span} S$.

Solution

We find *r(A)* and $r(A \mid \vec{b})$.

$$(A \mid \vec{b}) = \begin{pmatrix} 1 & -2 & 3 & 6 \\ 2 & 1 & 4 & 5 \\ 3 & -1 & 7 & 11 \end{pmatrix} \xrightarrow[R_1(-3)+R_3]{R_1(-2)+R_2} \begin{pmatrix} 1 & -2 & 3 & 6 \\ 0 & 5 & -2 & -7 \\ 0 & 5 & -2 & -7 \end{pmatrix}$$

$$\xrightarrow{R_2(-1)+R_3} \begin{pmatrix} 1 & -2 & 3 & 6 \\ 0 & 5 & -2 & -7 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Hence, $r(A) = r(A \mid \vec{b}) = 2$. By **Theorem 4.6.1** , we have the following conclusions:

1. The system $A\vec{X} = \vec{b}$ is consistent.
2. The vector $\vec{b}$ is a linear combination of *S*.
3. The vector $\vec{b} \in \operatorname{span} S$.

Example 4.6.2.

Let *S* and *A* be the same as in **Example 4.6.1** and let $\overrightarrow{b_1} = (6,\ 5,\ 10)^T$.

1. Determine whether $A\vec{X} = \vec{b}$ is consistent, where $\vec{X} = (x_1,\ x_2,\ x_3)^T$.
2. Determine whether the vector $\vec{b}$ is a linear combination of *S*.
3. Determine whether the vector $\overrightarrow{b_1} \in \operatorname{span} S$.

Solution

We find *r(A)* and $r(A\,|\,\vec{b})$.

$$(A\,|\,\vec{b}) = \begin{pmatrix} 1 & -2 & 3 & 6 \\ 2 & 1 & 4 & 5 \\ 3 & -1 & 7 & 10 \end{pmatrix} \xrightarrow[R_1(-3)+R_3]{R_1(-2)+R_2} \begin{pmatrix} 1 & -2 & 3 & 6 \\ 0 & 5 & -2 & -7 \\ 0 & 5 & -2 & -8 \end{pmatrix}$$

$$\xrightarrow{R_2(-1)+R_3} \begin{pmatrix} 1 & -2 & 3 & 6 \\ 0 & 5 & -2 & -7 \\ 0 & 0 & 0 & -1 \end{pmatrix}.$$

Hence, $r(A) = 2$ and $r(A \mid \overrightarrow{b_1}) = 3$. Because $r(A) < r(A \mid \overrightarrow{b_1})$, by **Theorem 4.6.1** , we have the conclusions.

1. The system $A\vec{X} = \overrightarrow{b_1}$ is inconsistent.
2. The vector $\vec{b}$ is not a linear combination of *S*.
3. The vector $\vec{b} \notin \text{span } S$.

Example 4.6.3.

Let $S = \left\{ \begin{pmatrix} 6 \\ 3 \end{pmatrix}, \begin{pmatrix} -4 \\ -2 \end{pmatrix} \right\}$.

1. Determine whether $\vec{b} = (2,\ 1)$ is in span *S*.
2. Determine whether $\vec{b} = (1,\ 2)$ is in span *S*.

Solution

Let $\vec{b} = (b_1,\ b_2)^T$.

$$(A \,|\, \vec{b}) = \begin{pmatrix} 6 & -4 & b_1 \\ 3 & -2 & b_2 \end{pmatrix} \overset{R_1\left(\frac{1}{6}\right)}{\longrightarrow} \begin{pmatrix} 1 & -\frac{2}{3} & \frac{b_1}{6} \\ 3 & -2 & b_2 \end{pmatrix} \overset{R_1(-3)+R_2}{\longrightarrow}$$

$$\begin{pmatrix} 1 & -\frac{2}{3} & \frac{b_1}{6} \\ 0 & 0 & b_2 - \frac{b_1}{2} \end{pmatrix}.$$

1. If $b_2 - \frac{b_1}{2} = 0$, then $r(A) = (A \,|\, \vec{b})$ and by **Theorem 4.6.1** , $\vec{b} \in \text{span } S$. Hence, if $\vec{b} = (b_1, b_2) = (2, 1)$, then $b_2 - \frac{b_1}{2} = 1 - \frac{2}{2} = 0$ and $\vec{b} = (2, 1) \in \text{span } S$.
2. If $b_2 - \frac{b_1}{2} \neq 0$, then $r(A) < (A \,|\, \vec{b})$ and by **Theorem 4.6.1** , $\vec{b} \notin \text{span } S$. Hence, if $\vec{b} = (b_1, b_2) = (1, 2)$, then $b_2 - \frac{b_1}{2} = 2 - \frac{1}{2} = \frac{3}{2} \neq 0$ and $\vec{b} = (1, 2) \notin \text{span } S$.

Example 4.6.4.

Let $\overrightarrow{a_1} = (1, 1, 3)^T$, $\overrightarrow{a_2} = (1, 0, 4)^T$, $\overrightarrow{a_3} = (2, 1, 7)^T$, and $\vec{b} = (b_1, b_2, b_3)^T$.

1. Find conditions on b_1, b_2, b_3 such that $\vec{b} \in \text{span}\left\{\overrightarrow{a_1}, \overrightarrow{a_2}, \overrightarrow{a_3}\right\}$.
2. Find conditions on b_1, b_2, b_3 such that $\vec{b} \notin \text{span}\left\{\overrightarrow{a_1}, \overrightarrow{a_2}, \overrightarrow{a_3}\right\}$.

3. Find a vector $\vec{b} = (b_1,\ b_2,\ b_3)^T$ such that $\vec{b} \notin \text{span}\left\{\vec{a_1},\ \vec{a_2},\ \vec{a_3}\right\}$.

Solution

Let $\vec{b} = (b_1,\ b_2,\ b_3)$ and $A = (\vec{a_1}\vec{a_2}\vec{a_3})$.

$$(A \mid \vec{b}) = \begin{pmatrix} 1 & 1 & 2 & b_1 \\ 1 & 0 & 1 & b_2 \\ 3 & 4 & 7 & b_3 \end{pmatrix} \xrightarrow[R_1(-3)+R_3]{R_1(-1)+R_2} \begin{pmatrix} 1 & 1 & 2 & b_1 \\ 0 & -1 & -1 & b_2 - b_1 \\ 0 & 1 & 1 & b_3 - 3b_1 \end{pmatrix}$$

$$\xrightarrow{R_2(1)+R_3} \begin{pmatrix} 1 & 1 & 2 & b_1 \\ 0 & 1 & 1 & b_1 - b_2 \\ 0 & 0 & 0 & -4b_1 + b_2 + b_3 \end{pmatrix} = (B \mid \vec{c}).$$

1. If $-4b_1 + b_2 + b_3 = 0,$ that is, $b_3 = 4b_1 - b_2,$ then $r(A) = (A \mid \vec{b}).$ By **Theorem 4.6.1** , $\vec{b} \in \text{span}\left\{\vec{a_1},\ \vec{a_2},\ \vec{a_3}\right\}$.

2. If $b_3 \neq 4b_1 - b_2 \neq 0,$ then $r(A) < (A \mid \vec{b}).$ By **Theorem 4.6.1** , $\vec{b} \notin \text{span}\left\{\vec{a_1},\ \vec{a_2},\ \vec{a_3}\right\}$.

3. Let $b_1 = b_2 = 0$ and $b_3 = 1$. Then $b_3 \neq 4b_1 - b_2.$ Hence $\vec{b} = (0,\ 0,\ 1) \notin \text{span}\left\{\vec{a_1},\ \vec{a_2},\ \vec{a_3}\right\}$.

Theorem 4.6.1 provides a criterion to determine whether a given vector $\vec{b}$ belongs to span *S*. In the following, we study how to determine whether every vector $\vec{b}$ in $\mathbb{R}^m$ belongs to span *S*, that is, to determine whether $\text{span } S = \mathbb{R}^m$. By **(2.2.1)**, we have $\text{span } S = A(\mathbb{R}^n)$. This, together with

Theorem 4.5.4 (1), implies that $\text{span } S = A(\mathbb{R}^n) = \mathbb{R}^m$ if and only if the system $A\vec{X} = \vec{b}$ is consistent for every $\vec{b} \in \mathbb{R}^m$. This leads us to obtain the following criterion to determine whether $\text{span } S = \mathbb{R}^m$ by using ranks.

Theorem 4.6.2.

The following assertions are equivalent.

1. $\text{span } S = \mathbb{R}^m$
2. *The system* $A\vec{X} = \vec{b}$ *is consistent for every vector* $\vec{b}$ *in* $\mathbb{R}^m$.
3. $r(A) = m$.

Proof

By **Theorems 4.5.4** and **4.6.1**, $\text{span } S = \mathbb{R}^m$ if and only if every vector $\vec{b}$ in $\mathbb{R}^n$ belongs to span S if and only if the system $A\vec{X} = \vec{b}$ is consistent for every vector $\vec{b}$ in $\mathbb{R}^n$ if and only if $r(A) = m$.

Example 4.6.5.

Determine whether $\text{span } S = \mathbb{R}^3$, where

$$S = \left\{ \begin{pmatrix} 1 \\ -4 \\ 4 \end{pmatrix}, \begin{pmatrix} -2 \\ 5 \\ -5 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \right\}.$$

Solution

$$A = \begin{pmatrix} 1 & -2 & 1 \\ -4 & 5 & 2 \\ 4 & -5 & 3 \end{pmatrix} \underset{R_1(-4)+R_3}{\overset{R_1(4)+R_2}{\longrightarrow}} \begin{pmatrix} 1 & -2 & -1 \\ 0 & -3 & 6 \\ 0 & 3 & -1 \end{pmatrix} \overset{R_2(1)+R_3}{\longrightarrow}$$

$$\begin{pmatrix} 1 & -2 & -1 \\ 0 & -3 & 6 \\ 0 & 0 & 5 \end{pmatrix}.$$

Hence, $r(A) = 3$. By **Theorem 4.6.2** , $\text{span } S = \mathbb{R}^3$.

Example 4.6.6.

Determine whether span $\text{span } S = \mathbb{R}^3$, where

$$S = \left\{ \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}, \begin{pmatrix} -2 \\ 0 \\ -2 \end{pmatrix}, \begin{pmatrix} 0 \\ 3 \\ -3 \end{pmatrix} \right\}.$$

Solution

$$A = \begin{pmatrix} 1 & 0 & -2 & 0 \\ 0 & 1 & 0 & 3 \\ 1 & -1 & -2 & -3 \end{pmatrix} \xrightarrow{R_1(-1)+R_3} \begin{pmatrix} 1 & 0 & -2 & 0 \\ 0 & 1 & 0 & 3 \\ 0 & -1 & 0 & -3 \end{pmatrix} \xrightarrow{R_2(1)+R_3}$$

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

Hence, $r(A) = 2$. Because $r(A) = 2 < 3$, by **Theorem 4.6.2** , $\operatorname{span} S \neq \mathbb{R}^3$.

Exercises

1. Let $S = \left\{\vec{a}_1, \vec{a}_2, \vec{a}_3, \vec{a}_4\right\}$ and $A = (\vec{a}_1, \vec{a}_2, \vec{a}_3, \vec{a}_4)$, where

$$\vec{a}_1 = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \quad \vec{a}_2 = \begin{pmatrix} -2 \\ 1 \\ -1 \end{pmatrix}, \quad \vec{a}_3 = \begin{pmatrix} 3 \\ 4 \\ 7 \end{pmatrix}, \quad \vec{a}_1 = \begin{pmatrix} -1 \\ -4 \\ -5 \end{pmatrix},$$

$$\vec{b} = \begin{pmatrix} 2 \\ 2 \\ 4 \end{pmatrix}.$$

1. Determine whether $A\vec{X} = \vec{b}$ is consistent, where $\vec{X} = (x_1, x_2, x_3, x_4)^T$.
2. Determine whether the vector $\vec{b}$ is a linear combination of S.

3. Determine whether the vector $\vec{b} \in \operatorname{span} S$.

2. Let $\overrightarrow{a_1} = (1, \ -2, \ 2)^T$, $\overrightarrow{a_2} = (-1, \ 0, \ -10)^T$, $\overrightarrow{a_3} = (1, \ -1, \ 6)^T$, and $\vec{b} = (b_1, \ b_2, \ b_3)^T$.

1. Find conditions on $b_1, \ b_2, \ b_3$ such that $\vec{b} \in \operatorname{span}\left\{\overrightarrow{a_1}, \ \overrightarrow{a_2}, \ \overrightarrow{a_3}\right\}$.

2. Find conditions on $b_1, \ b_2, \ b_3$ such that $\vec{b} \notin \operatorname{span}\left\{\overrightarrow{a_1}, \ \overrightarrow{a_2}, \ \overrightarrow{a_3}\right\}$.

3. Find a vector $\vec{b} = (b_1, \ b_2, \ b_3)^T$ such that $\vec{b} \notin \operatorname{span}\left\{\overrightarrow{a_1}, \ \overrightarrow{a_2}, \ \overrightarrow{a_3}\right\}$.

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

1. Determine whether $\operatorname{span} S = \mathbb{R}^2$.
2. Determine whether $\vec{b} = (4, \ 2)^T$ is in span *S*.
3. Determine whether $\vec{b} = (6, \ 1)^T$ is in span *S*.

4. Determine whether $\operatorname{span} S = \mathbb{R}^2$, where

i. $S = \left\{\begin{pmatrix} -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ -2 \end{pmatrix}, \begin{pmatrix} 3 \\ 5 \end{pmatrix}\right\}$.

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

iii. $S = \left\{\begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 3 \\ -2 \end{pmatrix}, \begin{pmatrix} 4 \\ 1 \end{pmatrix}\right\}$.

5. Determine whether $\operatorname{span} S = \mathbb{R}^3$, where

$$S = \left\{ \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ 5 \\ 0 \end{pmatrix}, \begin{pmatrix} 3 \\ 4 \\ 5 \end{pmatrix} \right\}.$$

6. Determine whether $\operatorname{span} S = \mathbb{R}^4$, where

$$S = \left\{ \begin{pmatrix} 1 \\ 0 \\ 4 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 0 \\ -1 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix} \right\}.$$

Chapter 5 Linear transformations

5.1 Linear transformations

In this section, we use the product of a matrix and a vector studied in Section 2.4 to define linear transformations.

Definition 5.1.1.

Let *A* be an $m \times n$ matrix given in **(2.1.1)** . A linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$, denoted by *T*, is defined by

(5.1.1)

$$T(\overrightarrow{X}) = A\overrightarrow{X} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{n2} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix},$$

where $\overrightarrow{X} = (x_1, x_2, \cdots, x_n)^T) \in \mathbb{R}^n$. The vector $T(\overrightarrow{X})$ is said to be the image of $\overrightarrow{X}$ under *T* and $\overrightarrow{X}$ is called the inverse image of $T(\overrightarrow{X})$. The space $\mathbb{R}^n$ is the domain of *T* and $\mathbb{R}^m$ is the codomain of *T*. The matrix *A* is said to be the standard matrix for *T* and *T* is called the multiplication by *A*. A linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^n$ is said to be a linear operator.

If T is a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$, then we say that T maps $\mathbb{R}^n$ into $\mathbb{R}^m$, which is denoted by the symbol $T : \mathbb{R}^n \to \mathbb{R}^m$.

When $m = n = 1,$ the linear transformation is a function $T : \mathbb{R} \to \mathbb{R}$ defined by $T(x) = a_{11}x.$ Hence, linear transformations are generalizations of such functions to higher dimensional spaces.

By **Definition 5.1.1** , we see that every matrix determines a linear transformation. Hence, for every linear transformation, there is a unique matrix corresponding to it. Sometimes, we write T as T_A in order to emphasize that the linear transformation is determined by the standard matrix A.

Example 5.1.1.

For each of the following matrices

$$A_1 = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 0 & 1 \end{pmatrix}, A_2 = \begin{pmatrix} 1 & 0 & 1 \\ -1 & 1 & 2 \\ 3 & -2 & 0 \end{pmatrix},$$

determine the domain and codomain of $T_{A_i}, i = 1, 2.$ Moreover, compute $T_{A_1}(1, 1, 1)$ and $T_{A_2}(-1, 0, 1)$.

Solution

Because the size of A_1 is $2 \times 3,$ the domain and codomain of T_{A_1} are $\mathbb{R}^3$ and $\mathbb{R}^2$, respectively.

$$T_{A_1} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 6 \\ 2 \end{pmatrix}.$$

Because the size of A_2 is $3 \times 3,$ both the domain and codomain of T_{A_2} are $\mathbb{R}^3$.

$$T_{A_2}\begin{pmatrix}-1\\0\\1\end{pmatrix}=\begin{pmatrix}1&0&1\\-1&1&2\\3&-2&0\end{pmatrix}\begin{pmatrix}-1\\0\\1\end{pmatrix}=\begin{pmatrix}0\\3\\-3\end{pmatrix}.$$

Example 5.1.2.

Let $A=\begin{pmatrix}1&4&7\\2&5&8\\3&6&9\end{pmatrix}$. Find $T(\overrightarrow{e_1}),T(\overrightarrow{e_2}),T(\overrightarrow{e_3})$, where $\overrightarrow{e_1},\overrightarrow{e_2},\overrightarrow{e_3}$ are the standard vectors in $\mathbb{R}^3$.

Solution

$$\begin{aligned}T(\overrightarrow{e_1}) &= T\begin{pmatrix}1\\0\\0\end{pmatrix}=\begin{pmatrix}1&4&7\\2&5&8\\3&6&9\end{pmatrix}\begin{pmatrix}1\\0\\0\end{pmatrix}=\begin{pmatrix}1\\2\\3\end{pmatrix}.\\ T(\overrightarrow{e_1}) &= T\begin{pmatrix}0\\1\\0\end{pmatrix}=\begin{pmatrix}1&4&7\\2&5&8\\3&6&9\end{pmatrix}\begin{pmatrix}0\\1\\0\end{pmatrix}=\begin{pmatrix}4\\5\\6\end{pmatrix}.\\ T(\overrightarrow{e_1}) &= T\begin{pmatrix}0\\0\\1\end{pmatrix}=\begin{pmatrix}1&4&7\\2&5&8\\3&6&9\end{pmatrix}\begin{pmatrix}0\\0\\1\end{pmatrix}=\begin{pmatrix}7\\8\\9\end{pmatrix}.\end{aligned}$$

By **Example 5.1.2** , we see that the column vectors of the standard matrix *A* are the images of the standard vectors in $\mathbb{R}^3$ under *T*. In fact, the result holds for any linear transformations.

Theorem 5.1.1.

Let $T:\mathbb{R}^n\to\mathbb{R}^m$ be a linear transformation and $\overrightarrow{e_1},\cdots,\overrightarrow{e_n}$ be the standard vectors in $\mathbb{R}^n$ given in **(1.1.1)** *. Then the standard matrix A of T is given by*

$$A=\left(T(\overrightarrow{e_1}),T(\overrightarrow{e_2}),\cdots,T(\overrightarrow{e_n})\right).$$

By **(5.1.1)** , we obtain the following expressions for linear transformations.

(5.1.2)

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

or

(5.1.3)

$$T(x_1,\cdots,x_n)=(a_{11}x_1+\cdots+a_{1n}x_n,\cdots,a_{m1}x_1+\cdots+a_{mn}x_n).$$

Example 5.1.3.

For each of the following linear transformations, express it in form **(5.1.1)** .

a. $T(x,y,z)=(\frac{1}{2}x,y,4z)$;
b. $T(x,y,z)=(\frac{1}{2}x+\frac{1}{2}z,\frac{1}{3}y+\frac{1}{3}z)$.

Solution

a. $T\begin{pmatrix}x\\y\\z\end{pmatrix}=\begin{pmatrix}\frac{1}{2}&0&0\\0&1&0\\0&0&4\end{pmatrix}\begin{pmatrix}x\\y\\z\end{pmatrix}$;

b. $T\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & \frac{1}{3} & \frac{1}{3} \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix}.$

Definition 5.1.2.

A linear operator $T : \mathbb{R}^n \to \mathbb{R}^n$ satisfying

(5.1.4)

$$T(\overrightarrow{X}) = k\overrightarrow{X} \qquad \text{for each } \overrightarrow{X} \in \mathbb{R}^n$$

is called a contraction with factor *k* if $0 \leq k < 1,$ and a dilation with factor *k* if $k \geq 1.$

By **Definition 5.1.2** , we see that the standard matrix for a contraction or dilation with factor *k* is a diagonal matrix $kI_n.$

By **Theorem 2.2.1** , we obtain the following properties.

Proposition 5.1.1.

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ *be a linear transformation and let* $\overrightarrow{X}, \overrightarrow{Y} \in \mathbb{R}^n$ *and* $\alpha, \beta \in \mathbb{R}.$ *Then*

(5.1.5)

$$T(\alpha\overrightarrow{X} + \beta\overrightarrow{Y}) = \alpha T(\overrightarrow{X}) + \beta T(\overrightarrow{Y}).$$

Proof

Let *A* be the $m \times n$ standard matrix of *T*. Then by **Definition 5.1.1** , we have $T(\overrightarrow{X}) = A\overrightarrow{X}$ and $T(\overrightarrow{Y}) = A\overrightarrow{Y}$. By **Theorem 2.2.1** , we have

$$T(\alpha \overrightarrow{X} + \beta \overrightarrow{Y}) = A(\alpha \overrightarrow{X} + \beta \overrightarrow{Y}) = \alpha(A\overrightarrow{X}) + \beta(A\overrightarrow{Y}) = \alpha T(\overrightarrow{X}) + \beta T(\overrightarrow{Y})$$

and **(5.1.5)** holds.

Example 5.1.4.

Let $\overrightarrow{X} = (1,1,-1)$, $\overrightarrow{Y} = (-1,2,1)$, and

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

Compute $T_A(3\overrightarrow{X} - 2\overrightarrow{Y})$.

Solution

By **Proposition 5.1.1**, we have

$$\begin{aligned} T_A(3\overrightarrow{X} - 2\overrightarrow{Y}) &= 3T_A(\overrightarrow{X}) - 2T_A(\overrightarrow{Y}) \\ &= 3\begin{pmatrix} 0 & 1 & 1 \\ 1 & 2 & -1 \\ 1 & -1 & 3 \end{pmatrix}\begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix} - 2\begin{pmatrix} 0 & 1 & 1 \\ 1 & 2 & -1 \\ 1 & -1 & 3 \end{pmatrix}\begin{pmatrix} -1 \\ 2 \\ 1 \end{pmatrix} \\ &= 3\begin{pmatrix} 0 \\ 4 \\ -3 \end{pmatrix} - 2\begin{pmatrix} 3 \\ 2 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 12 \\ -9 \end{pmatrix} - \begin{pmatrix} 6 \\ 4 \\ 0 \end{pmatrix} = \begin{pmatrix} -6 \\ 8 \\ 9 \end{pmatrix}. \end{aligned}$$

Now, we introduce operations on linear transformations.

Definition 5.1.3.

1. If $S, T : \mathbb{R}^n \to \mathbb{R}^m$ are linear transformations, then we define the following operations:
   i. (Addition) $(T+S)(\overrightarrow{X}) = T(\overrightarrow{X}) + S(\overrightarrow{X})$.
   ii. (Subtraction) $(T-S)(\overrightarrow{X}) = T(\overrightarrow{X}) - S(\overrightarrow{X})$.
   iii. (Scalar multiplication) $(\alpha T)(\overrightarrow{X}) = \alpha(T(\overrightarrow{X}))$ for $\alpha \in \mathbb{R}$.

2. If $T : \mathbb{R}^n \to \mathbb{R}^m$ and $S : \mathbb{R}^m \to \mathbb{R}^l$ are linear transformations, then we define the following composition of *T* and *S*:
   (5.1.6)

$$(ST)(\overrightarrow{X}) = S(T(\overrightarrow{X})).$$

By **Definition 5.1.1** and the above operations, we see that if *A* and *B* are the standard matrices for $T_A, T_B : \mathbb{R}^n \to \mathbb{R}^m$, respectively, then $A+B, A-B, \alpha A$ are the standard matrices for $T_A + T_B, T_A - T_B, \alpha T_A$, respectively, that is,

(i)′ $(T_A + T_B)(\overrightarrow{X}) = (A+B)(\overrightarrow{X})$.

(ii)′ $(T_A - T_B)(\overrightarrow{X}) = (A-B)(\overrightarrow{X})$.

(iii)′ $(\alpha T_A)(\overrightarrow{X}) = (\alpha A)(\overrightarrow{X})$ for $\alpha \in \mathbb{R}$.

If *A* and *B* are the standard matrices for $T_A : \mathbb{R}^n \to \mathbb{R}^m$ and $T_B : \mathbb{R}^m \to \mathbb{R}^l$, respectively, then *BA* is the standard matrix for $T_B T_A$, that is,

$$(T_B T_A)(\overrightarrow{X}) = (BA)(\overrightarrow{X}).$$

Note that the standard matrix for $T_B T_A$ is *BA* instead of *AB* and the composition linear transformation $T_B T_A$ is defined under the condition that the domain of T_B must be the same as the codomain of T_A.

Example 5.1.5.

Compute $(2T-3S)(x_1,x_2)$, where

$$T(x_1,x_2)=(3x_1-2x_2,-x_1-x_2,2x_1+3x_2),$$
$$S(x_1,x_2)=(x_1+x_2,x_1-x_2,-2x_1+4x_2).$$

Solution

(Method 1). We write T and S as follows.

$$T\begin{pmatrix}x_1\\x_2\end{pmatrix}=\begin{pmatrix}3&-2\\-1&-1\\2&3\end{pmatrix}\begin{pmatrix}x_1\\x_2\end{pmatrix}\text{ and }S\begin{pmatrix}x_1\\x_2\end{pmatrix}=\begin{pmatrix}1&1\\1&-1\\-2&4\end{pmatrix}\begin{pmatrix}x_1\\x_2\end{pmatrix}.$$

Then

$$\begin{aligned}(2T-3S)\begin{pmatrix}x_1\\x_2\end{pmatrix} &= 2T\begin{pmatrix}x_1\\x_2\end{pmatrix}-3S\begin{pmatrix}x_1\\x_2\end{pmatrix}\\ &= 2\begin{pmatrix}3&-2\\-1&-1\\2&3\end{pmatrix}\begin{pmatrix}x_1\\x_2\end{pmatrix}-3\begin{pmatrix}1&1\\1&-1\\-2&4\end{pmatrix}\begin{pmatrix}x_1\\x_2\end{pmatrix}\\ &= [2\begin{pmatrix}3&-2\\-1&-1\\2&3\end{pmatrix}-3\begin{pmatrix}1&1\\1&-1\\-2&4\end{pmatrix}]\begin{pmatrix}x_1\\x_2\end{pmatrix}\\ &= \begin{pmatrix}3&-7\\-5&1\\10&-6\end{pmatrix}\begin{pmatrix}x_1\\x_2\end{pmatrix}=\begin{pmatrix}3x_1-7x_2\\-5x_1+x_2\\10x_1-6x_2\end{pmatrix}.\end{aligned}$$

(Method 2) By **Definition 5.1.3** , we have

$$\begin{aligned}(2T-3S)(x_1,x_2) &= 2T(x_1,x_2)-3S(x_1,x_2)\\ &= 2(3x_1-2x_2,-x_1-x_2,2x_1+3x_2)-3(x_1+x_2,x_1-x_2,-2x_1+4x_2)\\ &= (6x_1-4x_2,-2x_1-2x_2,4x_1+6x_2)-(3x_1+3x_2,3x_1-3x_2,-6x_1+12x_2)\\ &= (3x_1-7x_2,-5x_1+x_2,10x_1+6x_2).\end{aligned}$$

Example 5.1.6.

Let T_A, T_B be linear transformations given by

$$T_A(x_1,x_2)=(x_1+x_2,2x_1-3x_2,3x_1-2x_2)$$

and

$$T_B(x_1,x_2,x_3)=(x_1-x_2+x_3,x_1-2x_3).$$

Find $(T_BT_A)(x_1,x_2)$.

Solution

Because

$$T_A\begin{pmatrix}x_1\\x_2\end{pmatrix}=\begin{pmatrix}1&1\\2&-3\\3&-2\end{pmatrix}\begin{pmatrix}x_1\\x_2\end{pmatrix}\qquad \text{and}$$
$$T_B\begin{pmatrix}x_1\\x_2\\x_3\end{pmatrix}=\begin{pmatrix}1&-1&1\\1&0&-2\end{pmatrix}\begin{pmatrix}x_1\\x_2\\x_3\end{pmatrix},$$

we have

$$\begin{aligned}(T_B T_A)\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} &= \begin{pmatrix} 1 & -1 & 1 \\ 1 & 0 & -2 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 2 & -3 \\ 3 & -2 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \\ &= \begin{pmatrix} 2 & 2 \\ -5 & 5 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}.\end{aligned}$$

Exercises

1. For each of the following matrices,

$$A_1 = \begin{pmatrix} 2 & -2 & 4 & -1 \\ -1 & 2 & 3 & 0 \\ 0 & 1 & 2 & -1 \end{pmatrix} \qquad A_2 = \begin{pmatrix} -2 & 2 & 1 \\ 1 & 1 & -3 \\ 1 & 0 & -2 \end{pmatrix}$$

$$A_3 = \begin{pmatrix} 1 & -2 & 0 \\ -2 & 3 & 1 \\ 3 & -2 & 3 \\ 1 & 2 & -1 \end{pmatrix} \qquad A_4 = \begin{pmatrix} -1 & 1 & 1 \\ 2 & 0 & -2 \\ -1 & -1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$$

   determine the domain and codomain of $T_{A_i}, i = 1, 2, 3, 4.$ Moreover, compute $T_{A_3}(0, 1, 1, 2), T_{A_2}(1, -2, 1), T_{A_3}(-1, 0, 1), T_{A_4}(-1, 1, 1).$
2. Let $A = \begin{pmatrix} 1 & 2 \\ -1 & 3 \end{pmatrix}$. Find $T_A\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $T_A\begin{pmatrix} 0 \\ 1 \end{pmatrix}$.
3. Let $A = \begin{pmatrix} -1 & 2 & 1 \\ 3 & 0 & 6 \end{pmatrix}$. Find $T_A(\overrightarrow{e_1}), T_A(\overrightarrow{e_2}), T_A(\overrightarrow{e_3})$, where $\overrightarrow{e_1}, \overrightarrow{e_2}, \overrightarrow{e_3}$ are the standard vectors in $\mathbb{R}^3$.
4. For each of the following linear transformations, express it in **(5.1.1)** .
   a. $T(x_1, x_2) = (x_1 - x_2, 2x_1 + x_2, 5x_1 - 3x_2).$
   b. $T(x_1, x_2, x_3, x_4) = (6x_1 - x_2 - x_3 + x_4, x_2 + x_3 - 2x_4, 3x_3 + x_4, 5x_4).$
   c. $T(x_1, x_2, x_3) = (4x_1 - x_2 + 2x_3, 3x_1 - x_3, 2x_1 + x_2 + 5x_3).$
   d. $T(x_1, x_2, x_3) = (8x_1 - x_1 + x_3, x_2 - x_3, 2x_1 + x_2, x_2 + 3x_3).$

5. Let $\overrightarrow{X} = (2, -2, -2, 1)$, $\overrightarrow{Y} = (1, -2, 2, 3)$ and

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

Compute $T_A(\overrightarrow{X} - 3\overrightarrow{Y})$.

6. Let $S, T : \mathbb{R}^3 \to \mathbb{R}^3$ be linear operators such that $T(1,2,3) = (1,-4)$ and $S(1,2,3) = (4,9)$. Compute $(5T + 2S)(1,2,3)$.

7. Let $S, T : \mathbb{R}^3 \to \mathbb{R}^4$ be linear transformations such that $T(1,2,3) = (1,1,0,-1)$ and $S(1,2,3) = (2,-1,2,2)$. Compute $(3T - 2S)(1,2,3)$.

8. Compute $(T + 2S)(x_1, x_2, x_3, x_4)$, where

$$T(x_1, x_2, x_3, x_4) = (x_1 - x_2 + x_3, x_1 - x_2 + 2x_3 - x_4),$$
$$S(x_1, x_2, x_3, x_4) = (x_2 - x_3 + 2x_4, x_1 + x_3 - x_4).$$

9. For each pair of linear transformations T_A and T_B, find $T_B T_A$.
   a. $T_A(x_1, x_2) = (x_1 + x_2, x_1 - x_2), T_B(x_1, x_2) = (x_1 - x_2, 2x_1 - 3x_2, 3x_1 + x_2)$.
   b. $T_A(x_1, x_2, x_3) = (-x_1 - x_2 + x_3, x_1 + x_2), T_B(x_1, x_2) = (x_1 - x_2, 2x_1 + 3x_2)$.
   c. $T_A(x_1, x_2, x_3) = (-x_1 + x_2 - x_3, x_1 - 2x_2, x_3), T_B(x_1, x_2, x_3) = (x_1 - x_2 + x_3, 2x_1 + 3x_2 + x_3)$.

5.2 Onto, one to one, and inverse transformations

Let $T: \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation with standard matrix *A* given in **(5.1.1)** . We collect all images together as a set denoted by $T(\mathbb{R}^n)$, that is, all the images $T(\vec{X})$ as $\vec{X}$ takes all the vectors in $\mathbb{R}^n$.

Definition 5.2.1.

Let $T: \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. The set

(5.2.1)

$$T(\mathbb{R}^n) = \left\{T(\vec{X}) \in \mathbb{R}^m : \vec{X} \in \mathbb{R}^n\right\}$$

is called the range of *T*.

By **Definitions 5.1.1** and **5.2.1** and **Theorem 2.2.2** , we obtain the following result, which establishes the relations among the range of T, $A(\mathbb{R}^n)$ given in **(2.2.1)** and the spanning space given in **(1.4.8)** .

Theorem 5.2.1.

Let $S = \left\{\vec{a}_1, \vec{a}_2, \cdots, \vec{a}_n\right\}$ *be the set of the column vectors of A given in* **(2.1.2)** *Then*

(5.2.2)

$$T_A(\mathbb{R}^n) = A(\mathbb{R}^n) = \text{span } S.$$

By **Theorems 4.6.1** and **5.2.1**, we obtain the following result, which uses the ranks of $r(A)$ and $r(A\,|\,\vec{b})$ to determine whether $\vec{b} \in T(\mathbb{R}^n)$.

Theorem 5.2.2.

Let S be the same as in ***Theorem 5.2.1****. Then the following assertions are equivalent.*

i. $\vec{b} \in T(\mathbb{R}^n)$;
ii. $A\vec{X} = \vec{b}$ *is consistent;*
iii. $\vec{b}$ *is a linear combination of S;*
iv. $\vec{b} \in$ span *S;*
v. $r(A) = r(A\,|\,\vec{b})$,

Example 5.2.1.

Let

$T(x_1, x_2, x_3) = (x_1 - x_2,\ -2x_1 + 2x_2 + x_3,\ -x_1 + x_2 + x_3)$ and $\vec{b} = (, , 0,\ -1)$. Determine whether $\vec{b} \in T(\mathbb{R}^3)$.

Solution

$$A=\begin{pmatrix}1 & -1 & 0 & 1\\ -2 & 2 & 1 & 0\\ -1 & 1 & 1 & -1\end{pmatrix}\underset{R_1(1)+R_3}{\overset{R_1(2)+R_2}{\longrightarrow}}\begin{pmatrix}1 & -1 & 0 & 1\\ 0 & 0 & 1 & 2\\ 0 & 0 & 1 & 0\end{pmatrix}R_2(-1)+R_3\rightarrow$$

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

This implies $r(A)=2<3=r(A\,|\,\vec{b})$. By **Theorem 5.2.2** , $\vec{b}\notin T(\mathbb{R}^3)$.

Definition 5.2.2.

A linear transformation $T:\mathbb{R}^n\to\mathbb{R}^m$ is said to be onto if $T(\mathbb{R}^n)=\mathbb{R}^m$.

By **Theorem 5.2.1** , *T* is onto if and only if span $S=\mathbb{R}^m$. By **Theorem 4.6.2** , we obtain the following result, which uses the rank of *A* to determine whether *T* is onto.

Theorem 5.2.3.

Let $T:\mathbb{R}^n\to\mathbb{R}^m$ be a linear transformation with standard matrix A. Then the following assertions are equivalent.

i. *T is onto.*
ii. span $S=\mathbb{R}^m$.
iii. *The system $T(\vec{X})=\vec{b}$ is consistent for each $\vec{b}\in\mathbb{R}^m$.*
iv. $r(A)=m$.

By **Theorem 2.5.2** , $r(A)\le\min\{m,n\}$. Hence, if $T:\mathbb{R}^n\to\mathbb{R}^m$ is onto, then by **Theorem 5.2.3** ,

$$m=r(A)\le\min\{m,n\}\le n.$$

Hence, the necessary condition for T to be onto is $m \leq n$ and if $m > n$, then T is not onto.

Example 5.2.2.

Determine which of the following linear transformations are onto.

a. $T(x_1, x_2, x_3, x_4) = (x_1 + 2x_2 - 2x_3 + x_4, 3x_2 + 4x_3, x_1 + 2x_3 + x_4)$.
b. $T(x_1, x_2, x_3,) = (x_1 + x_2, \ -x_1 + 2x_2 + 4x_3, 2x_1 - x_2 + 3x_3, 5x_1 + 3x_2 + 6x_3)$.

Solution

a.
$$A = \begin{pmatrix} 1 & 2 & -2 & 1 \\ 0 & 3 & 4 & 0 \\ 1 & 0 & 2 & 1 \end{pmatrix} \xrightarrow{R_1(-1)+R_3} \begin{pmatrix} 1 & 2 & -2 & 1 \\ 0 & 3 & 4 & 0 \\ 0 & -2 & 4 & 0 \end{pmatrix} \xrightarrow{R_3(1)+R_2}$$
$$\begin{pmatrix} 1 & 2 & -2 & 1 \\ 0 & 1 & 8 & 0 \\ 0 & -2 & 4 & 0 \end{pmatrix} \xrightarrow{R(2)+R_3} \begin{pmatrix} 1 & 2 & -2 & 1 \\ 0 & 1 & 8 & 0 \\ 0 & 0 & 20 & 0 \end{pmatrix}.$$

Hence, $r(A) = 3 = m$. By **Theorem 5.2.3** , T is onto.

b. Because the standard matrix A of T is a 4×3 matrix, by **Theorem 5.2.3** , T is not onto because $r(A) \leq 3 = \min\{4, 3\} < 4 = m$.

In **(4.1.13)** , we introduce the solution space N_A of a homogeneous system $A\vec{X} = \vec{0}$. For a linear transformation T, the solution space of $T\vec{X} = \vec{0}$ is called a kernel for T and is denoted by ker(T), that is,

(5.2.3)

$$\ker(T) = \left\{ \vec{X} \in \mathbb{R}^n : T(\vec{X}) = \vec{0} \right\}.$$

The kernel ker(*T*) is the set of all inverse images of zero vector $\vec{0}$ in $\mathbb{R}^m$ under *T*. If *A* is the standard matrix of *T*, then by **(4.1.13)** and **Definition 5.1.1**, we have

(5.2.4)

$$\ker(T_A) = N_A.$$

Definition 5.2.3.

A linear transformation $T: \mathbb{R}^n \to \mathbb{R}^m$ is said to be one to one if $T(\vec{X}) = \vec{0}$ implies $\vec{X} = \vec{0}$.

Because $T(\vec{X}) = A\vec{X}$, if *T* is one to one, then $A\vec{X} = \vec{0}$ implies $\vec{X} = \vec{0}$. That is, *T* one to one if and only if the system $A\vec{X} = \vec{0}$ only has the zero solution. The latter can be justified by **Theorem 4.5.2** (1). Hence, we obtain the following criterion, which uses the rank of *A* to determine whether *T* is one to one.

Theorem 5.2.4.

Let T : $\mathbb{R}^n \to \mathbb{R}^m$ *be a linear transformation with standard matrix A. Then the following assertions hold.*

i. *T is one to one.*
ii. $A\vec{X} = \vec{0}$ *only has the zero solution.*
iii. $r(A) = n$.

If *T* : $R^n \to \mathbb{R}^m$ is one to one, then by **Theorems 2.5.2** and **5.2.4**,

$$n = r(A) \leq \min\{m, n\} \leq m.$$

Hence, what's necessary for *T* to be one to one is $n \leq m$ and if $n > m$, then *T* is not one to one.

Example 5.2.3.

Determine which of the following linear transformations are one to one.

a. $T(x, y) = (x - y, 2x + y, 4x - y)$.

b. $T(x, y) = (x - 2y, 2x - 4y, 4x - 8y)$.

c. $T(x, y, z) = (x - y - z, 2x + y + 3y)$.

Solution

a. Because

$$A_{3\times 2} = \begin{pmatrix} 1 & -1 \\ 2 & 1 \\ 4 & -1 \end{pmatrix} \underset{R_1(-4)+R_3}{\overset{R_1(-2)+R_2}{\longrightarrow}} \begin{pmatrix} 1 & -1 \\ 0 & 3 \\ 0 & 3 \end{pmatrix} \overset{R_2(-1)+R_3}{\longrightarrow} \begin{pmatrix} 1 & -1 \\ 0 & 3 \\ 0 & 0 \end{pmatrix},$$

$r(A_{3\times 2}) = 2$ and by **Theorem 5.2.4**, *T* is one to one.

b. Because

$$A_{3\times 2} = \begin{pmatrix} 1 & -2 \\ 2 & -4 \\ 4 & -8 \end{pmatrix} \underset{R_1(-4)+R_3}{\overset{R_1(-2)+R_2}{\longrightarrow}} \begin{pmatrix} 1 & -2 \\ 0 & 0 \\ 0 & 0 \end{pmatrix},$$

$r(A_{3\times 2}) = 1 < 2$ and by **Theorem 5.2.4**, *T* is not one to one.

c. Because $r(A_{2\times 3}) \leq \min\left\{2, 3\right\} = 2 < n = 3,$ by **Theorem 5.2.4**, *T* is not one to one.

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. By **Theorem 5.2.3**, *T* is onto if and only if $r(A) = m$ and by **Theorem 5.2.4**, *T* is one to one if and only if $r(A) = n.$ Hence, when $m = n,$ we see that onto and one to one linear transformations are the same. This, together with **Theorems 2.7.6** and **2.7.5**, implies the following result that uses the rank or determinant of *A* to determine whether a linear operator $T : \mathbb{R}^n \to \mathbb{R}^n$ is onto or one to one.

Theorem 5.2.5.

Let T be a linear operator from $\mathbb{R}^n$ to $\mathbb{R}^n$ with standard matrix A. Then the following assertions are equivalent.

i. *T is one to one.*
ii. *T is onto.*
iii. $|A| \neq 0.$
iv. $r(A) = n.$
v. *A is invertible.*

Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be a linear operator with standard matrix *A*. If *A* is invertible, then A^{-1} exists. The linear operator with standard matrix A^{-1} is called the inverse of *T*, denoted by T^{-1}, that is,

(5.2.5)

$$T^{-1}(\vec{X}) = A^{-1}\vec{X}.$$

Example 5.2.4.

Determine whether the following linear operators are invertible. If so, find their inverses.

a. $T(x, y) = (x\cos\theta - y\sin\theta, x\sin\theta + y\cos\theta),$ where $\theta \in \mathbb{R}$ is an angle in radians.
b. $T(x_1, x_2) = (2x_1 + x_2, 3x_1 + 4x_2).$
c. $T(x_1, x_2, x_3) = (x_1 + x_2 - x_3, -x_1 + 2x_2, x_2 + x_3).$

Solution

a. Because

$$|A| = \begin{vmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{vmatrix} = \cos^2\theta + \sin^2\theta = 1 \neq 0,$$

by **Theorem 5.2.5** , T is invertible. By **Theorem 2.7.5** (2),

$$A^{-1} = \frac{1}{|A|} = \begin{vmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{vmatrix} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}.$$

By **(5.2.5)** ,

$$T^{-1}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}.$$

b. Because $|A| = \begin{vmatrix} 2 & 1 \\ 3 & 4 \end{vmatrix} = 5 \neq 0,$ by **Theorem 5.2.5** , T is invertible. By **Theorem 2.7.5** (2),

$$A^{-1} = \frac{1}{|A|}\begin{pmatrix} 4 & -1 \\ -3 & 2 \end{pmatrix} = \begin{pmatrix} \frac{4}{5} & -\frac{1}{5} \\ -\frac{3}{5} & \frac{2}{5} \end{pmatrix}.$$

By **(5.2.5)** , we have

$$T^{-1}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} \frac{4}{5} & -\frac{1}{5} \\ -\frac{3}{5} & \frac{2}{5} \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}.$$

c. By computation,

$$|A| = \begin{vmatrix} 1 & 1 & -1 \\ -1 & 2 & 0 \\ 0 & 1 & 1 \end{vmatrix} = 4 \neq 0.$$

Hence, T is invertible.

$$(A\,|\,I_3) = \begin{pmatrix} 1 & 1 & -1 & 1 & 0 & 0 \\ -1 & 2 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \end{pmatrix} \xrightarrow{R_1(1)+R_2} \begin{pmatrix} 1 & 1 & -1 & 1 & 0 & 0 \\ 0 & 3 & -1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \end{pmatrix}$$

$$\xrightarrow{R_{2,3}} \begin{pmatrix} 1 & 1 & -1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 3 & -1 & 1 & 1 & 0 \end{pmatrix} \xrightarrow{R_2(-3)+R_3}$$

$$\begin{pmatrix} 1 & 1 & -1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & -4 & 1 & 1 & -3 \end{pmatrix} \xrightarrow{R(-\frac{1}{4})}$$

$$\begin{pmatrix} 1 & 1 & -1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & -\frac{1}{4} & -\frac{1}{4} & \frac{3}{4} \end{pmatrix} \xrightarrow[R_3(1)+R_1]{R_3(-1)+R_2} \begin{pmatrix} 1 & 1 & 0 & \frac{3}{4} & -\frac{1}{4} & \frac{3}{4} \\ 0 & 1 & 0 & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ 0 & 0 & 1 & -\frac{1}{4} & -\frac{1}{4} & \frac{3}{4} \end{pmatrix}$$

$$\xrightarrow{R(-1)+R_1} \begin{pmatrix} 1 & 0 & 0 & \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \\ 0 & 1 & 0 & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ 0 & 0 & 1 & -\frac{1}{4} & -\frac{1}{4} & \frac{3}{4} \end{pmatrix} = (I_3\,|\,A^{-1}).$$

By (5.2.5) , we have

$$T^{-1}\begin{pmatrix}x_1\\x_2\\x_3\end{pmatrix}=\begin{pmatrix}\frac{1}{2} & -\frac{1}{2} & \frac{1}{2}\\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4}\\ -\frac{1}{4} & -\frac{1}{4} & \frac{3}{4}\end{pmatrix}\begin{pmatrix}x_1\\x_2\\x_3\end{pmatrix}.$$

Exercises

1. For each of the following linear transformations, determine whether it is onto.
   a. $T(x_1, x_2, x_3) = (2x_1 - 2x_2 + x_3, 3x_1 - 2x_2 + x_3, x_1 + 2x_2 - 4x_3, 2x_1 - 3x_2 + 2x_3)$.
   b. $T(x_1, x_2, x_3) = (x_1 + 2x_2 - x_3,\ -x_1 - 2x_2,\ -x_1 + x_2 + 2x_3)$.
   c. $T(x_1, x_2, x_3, x_4) = (x_1 - 2x_2 + x_4, 3x_1 + 2x_4, x_1 - 2x_3 + 3x_4)$.
   d. $T(x_1, x_2, x_3, x_4) = (x_1 - x_2 + x_3, x_1 - x_4, 2x_1 - x_2 + x_3 - x_4)$.

2. Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be defined by

$$T(x, y) = (x - y, 2x - 2y).$$

3. Show that T is not onto and find a vector $\vec{b} \in \mathbb{R}^2$ such that $\vec{b} \notin T(\mathbb{R}^2)$.
4. For each of the following linear transformations, determine whether it is one to one.
   a. $T(x_1, x_2, x_3) = (x_1 - x_2 + 2x_3,\ -x_1 - x_2, x_1 + x_2 - x_3,\ -2x_1 - x_2 + x_3)$.
   b. $T(x_1, x_2, x_3) = (-x_1 + x_2 + x_3, x_1 - x_2 - 2x_3,\ +x_1 + x_2)$.
   c. $T(x_1, x_2, x_3, x_4) = (x_1 - x_2 + x_3, 2x_1 + x_2 + x_3, x_1 + x_2 + 2x_4)$.
   d. $T(x_1, x_2, x_3) = (x_1 - x_2 + 3x_3, x_1 - 2x_3, 2x_1 - x_2 + x_3)$.

5. For each of the following linear operators, determine whether it is invertible. If so, find its inverse.
   a. $T(x_1, x_2) = (x_1 + x_2, 6x_1 + 7x_2)$.
   b. $T(x_1, x_2) = (3x_1 - 9x_2, 4x_1 - 12x_2)$.
   c. $T(x_1, x_2, x_3) = (x_1 + 3x_2 + 2x_3, x_2 - x_3, x_3)$.
   d. $T(x_1, x_2, x_3) = (x_1 + 3x_2, 3x_1 + 7x_2 - 4x_3, x_1 + 5x_2 + 5x_3)$.

5.3 Linear operators in $\mathbb{R}^2$ and $\mathbb{R}^3$

In this section, we seek some special linear operators in $\mathbb{R}^2$ and in $\mathbb{R}^3$, called projection, reflection, and rotation operators.

Projection operators in $\mathbb{R}^2$

Theorem 5.3.1.

Let $T: \mathbb{R}^2 \to \mathbb{R}^2$ be the orthogonal projection on a line l passing through the origin (0, 0) *and let the angle between l and the x-axis be θ (see* **Figure 5.3** *). Then*

(5.3.1)

$$T\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \cos^2\theta & \sin\theta\cos\theta \\ \cos\theta\sin\theta & \sin^2\theta \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}.$$

Proof

From **Figure 5.3** , we see that

(5.3.2)

$$(w_1, w_2) = (r\cos\theta, r\sin\theta).$$

We prove the following assertion.

(5.3.3)

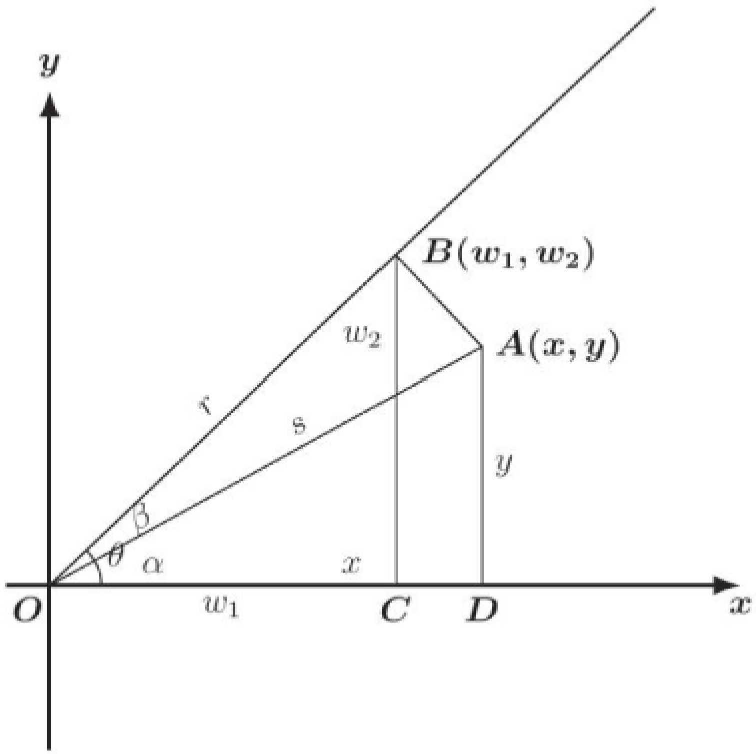
$$r = x\cos\theta + y\sin\theta.$$

Indeed, by **Figure 5.3** , $r = s\cos\beta = s\cos(\theta - \alpha)$ and

(5.3.4)

$$(x, y) = (s\cos\alpha, s\sin\alpha).$$

Figure 5.1: Projection operator



Because $\cos(\theta - \alpha) = \cos\alpha\cos\theta + \sin\alpha\sin\theta$, by **(5.3.4)** we have $r = s\cos(\theta - \alpha) = (s\cos\alpha)\cos\theta + (s\sin\alpha)\sin\theta = x\cos\theta + y\sin\theta$ and **(5.3.3)** holds. By **(5.3.2)** and **(5.3.3)** , we obtain

$$\begin{aligned}(w_1, w_2) &= (r\cos\theta, r\sin\theta) = ((x\cos\theta + y\sin\theta)\cos\theta, (x\cos\theta + y\sin\theta)\sin\theta) \\ &= (x\cos^2\theta + y\sin\theta\cos\theta, x\sin\theta\cos\theta + y\sin^2\theta)\end{aligned}$$

and **(5.3.1)** holds.

The result **(5.3.3)** can be proved by using the dot product. Indeed, by **Figure 5.3** and **(5.3.2)**, we see that

$$\overrightarrow{BA} = (x - r\cos\theta, y - r\sin\theta).$$

Because $\overrightarrow{BA} \perp \overrightarrow{OB}$, the dot product $\overrightarrow{BA} \cdot \overrightarrow{OB} = 0$, that is,

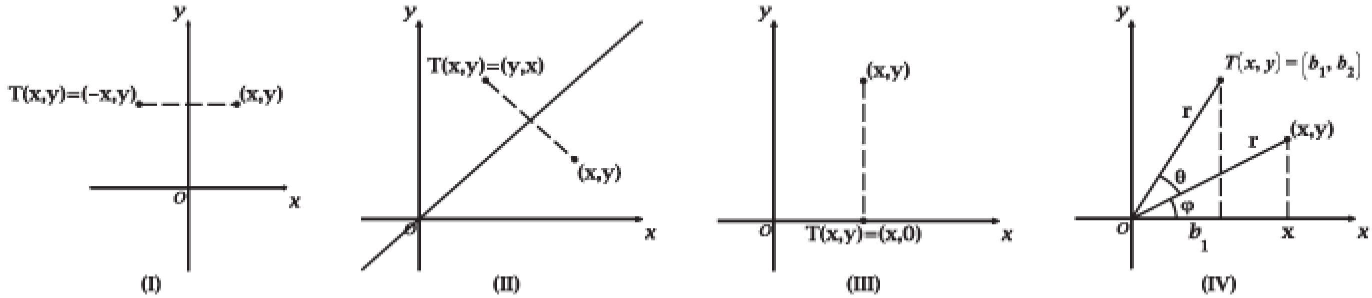
$$(x - r\cos\theta, y - r\sin\theta)\begin{pmatrix} r\cos\theta \\ r\sin\theta \end{pmatrix} = 0.$$

This, together with $\sin^2\theta + \cos^2\theta = 1$, implies

$$\begin{aligned}0 &= (x - r\cos\theta)(r\cos\theta) + (y - r\sin\theta)(r\sin\theta) \\ &= r[(x - r\cos\theta)\cos\theta + (y - r\sin\theta)\sin\theta] \\ &= r\left[x\cos\theta + y\sin\theta - r(\cos^2\theta + \sin^2\theta)\right] = r[x\cos\theta + y\sin\theta - r].\end{aligned}$$

This implies **(5.3.3)**.

Figure 5.2: (I), (II),(III),(IV)



By **Theorem 5.3.1** with $\theta = 0, \frac{\pi}{2}, \frac{\pi}{4},$ we obtain

Example 5.3.1.

i. The projection operator on the *x*-axis is

$$T\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x \\ 0 \end{pmatrix}.$$

ii. The projection operator on the *y*-axis is

$$T\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ y \end{pmatrix}.$$

iii. The projection operator on the line $y = x$ is

$$T\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \frac{x+y}{2} \\ \frac{x+y}{2} \end{pmatrix}.$$

Solution

i. The projection operator on the *x*-axis is

$$T\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \cos^2 0 & \sin 0 \cos 0 \\ \cos 0 \sin 0 & \sin^2 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x \\ 0 \end{pmatrix}.$$

ii. The projection operator on the *y*-axis is

$$T\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \cos^2 \frac{\pi}{2} & \sin\frac{\pi}{2}\cos\frac{\pi}{2} \\ \cos\frac{\pi}{2}\sin\frac{\pi}{2} & \sin^2\frac{\pi}{2} \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ y \end{pmatrix}.$$

iii. The projection operator on the line is

$$T\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \cos^2 \frac{\pi}{4} & \sin\frac{\pi}{4}\cos\frac{\pi}{4} \\ \cos\frac{\pi}{4}\sin\frac{\pi}{4} & \sin^2\frac{\pi}{4} \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \frac{x+y}{2} \\ \frac{x+y}{2} \end{pmatrix}.$$

Reflection operators in $\mathbb{R}^2$

Theorem 5.3.2.

Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ *be the reflection operator about a line l passing through the origin* (0, 0) *and let the angle between l and the x-axis be* θ *(see* ***Figure 5.3*** *). Then*

(5.3.5)

$$T\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \cos 2\theta & \sin 2\theta \\ \sin 2\theta & -\cos 2\theta \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}.$$

Proof

By the midpoint formula **(1.2.14)** and **Theorem 5.3.1**, from **Figure 5.3**, we have

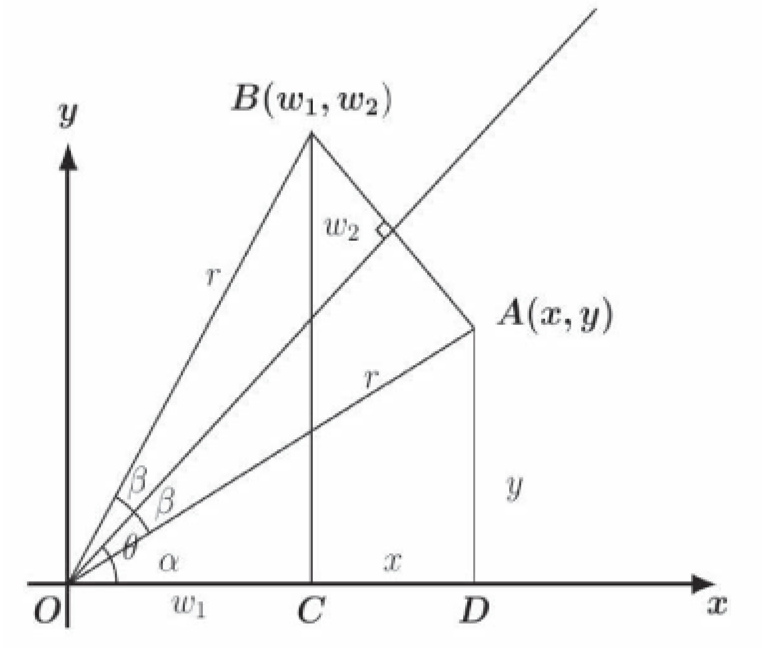
$$\begin{pmatrix} \frac{w_1 + x}{2} \\ \frac{w_2 + y}{2} \end{pmatrix} = \begin{pmatrix} \cos^2\theta & \sin\theta\cos\theta \\ \cos\theta\sin\theta & \sin^2\theta \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}.$$

This implies

$$\begin{aligned} T\begin{pmatrix} x \\ y \end{pmatrix} &= \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} = -\begin{pmatrix} x \\ y \end{pmatrix} + 2\begin{pmatrix} \cos^2\theta & \sin\theta\cos\theta \\ \cos\theta\sin\theta & \sin^2\theta \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} \\ &= \begin{pmatrix} 2\cos^2\theta - 1 & 2\sin\theta\cos\theta \\ 2\cos\theta\sin\theta & 2\sin^2\theta - 1 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}. \end{aligned}$$

This, together with $\cos 2\theta = 2\cos^2\theta - 1 = 1 - 2\sin^2\theta$ and $\sin 2\theta = 2\sin\theta\cos\theta,$ implies **(5.3.5)**.

Figure 5.3: Reflection operator



By **Theorem 5.3.2** with $\theta = 0, \frac{\pi}{2}, \frac{\pi}{4}$, we obtain

Example 5.3.2.

i. The reflection about the *x*-axis is

$$T\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x \\ -y \end{pmatrix}.$$

ii. The reflection about the *y*-axis (see **Figure 6.2** (II)) is

$$T\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} -x \\ y \end{pmatrix}.$$

iii. The reflection about the line $y = x$ (see **Figure 6.2** (II)) is

$$T\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} y \\ x \end{pmatrix}.$$

Solution

i. The reflection about the *x*-axis is

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

ii. The reflection about the *y*-axis is

$$T\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \cos 2\left(\frac{\pi}{2}\right) & \sin 2\left(\frac{\pi}{2}\right) \\ \sin 2\left(\frac{\pi}{2}\right) & -\cos 2\left(\frac{\pi}{2}\right) \end{pmatrix} = \begin{pmatrix} \cos\pi & \sin\pi \\ \sin\pi & -\cos\pi \end{pmatrix}$$

$$= \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} -x \\ y \end{pmatrix}.$$

iii. The reflection about the line $y = x$ is

$$T\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \cos 2\left(\frac{\pi}{4}\right) & \sin 2\left(\frac{\pi}{4}\right) \\ \sin 2\left(\frac{\pi}{4}\right) & -\cos 2\left(\frac{\pi}{4}\right) \end{pmatrix} = \begin{pmatrix} \cos\frac{\pi}{2} & \sin\frac{\pi}{2} \\ \sin\frac{\pi}{2} & -\cos\frac{\pi}{2} \end{pmatrix}$$

$$= \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} y \\ x \end{pmatrix}.$$

Example 5.3.3.

1. Find the linear operator on $\mathbb{R}^2$ that reflects about the line $y = x$ and then projects on the *y*-axis.
2. Find the linear operator on $\mathbb{R}^2$ that projects on the *y*-axis and then reflects about the line $y = x$.

Solution

The reflection about the line $y = x$ is $T_1(x, y) = (y, x)$ and the projection on the *y*-axis is $T_2(x, y) = (0, y)$.

1. The operator on $\mathbb{R}^2$ is $T_2T_1,$ that is,

$$T_2T_1(x, y) = T_2(T_1(x, y)) = T_2(y, x) = (0, x).$$

2. The operator on $\mathbb{R}^2$ is T_1T_2 and $T_1T_2(x, y) = T_1(0, y) = (y, 0)$.

The rotation operators in $\mathbb{R}^2$ through angle θ

A rotation operator on $\mathbb{R}^2$ is an operator that rotates each vector in $\mathbb{R}^2$ through a fixed angle θ (see **Figure 6.2** (IV)).

Theorem 5.3.3.

The rotation operator in $\mathbb{R}^2$ through an angle θ is

(5.3.6)

$$T\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}.$$

Proof

We rotate (*x*, *y*) to (b_1, b_2) through a counterclockwise angle θ Then

(5.3.7)

$$T(x, y) = (b_1, b_2).$$

Let *r* be the length of the segment from the origin to the point (*x*, *y*) and let ϕ denote the angle between the half-line beginning at the origin (0, 0) and passing through (*x*, *y*) and the *x*-axis. Then $(x, y) = (r\cos\phi, r\sin\phi)$ and

$$\begin{aligned} \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} &= \begin{pmatrix} r\cos(\theta+\phi) \\ r\sin(\theta+\phi) \end{pmatrix} = \begin{pmatrix} r\cos\theta\cos\phi - r\sin\theta\sin\phi \\ r\sin\theta\cos\phi + r\sin\phi\cos\theta \end{pmatrix} \\ &= \begin{pmatrix} x\cos\theta - y\sin\theta \\ x\sin\theta + y\cos\theta \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} \end{aligned}$$

and **(5.3.6)** holds.

Example 5.3.4.

Find the rotation operator with the rotation angle $\frac{\pi}{6}$ and the image of $x = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ under the operator.

Solution

By **(5.3.6)** with $\theta = \frac{\pi}{6}$,

$$T\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \cos\frac{\pi}{6} & -\sin\frac{\pi}{6} \\ \sin\frac{\pi}{6} & \cos\frac{\pi}{6} \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \frac{\sqrt{3}}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{\sqrt{3}}{2} \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}.$$

and $T\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} \frac{\sqrt{3}}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{\sqrt{3}}{2} \end{pmatrix}\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} \frac{\sqrt{3}-1}{2} \\ \frac{1+\sqrt{3}}{2} \end{pmatrix}.$

Example 5.3.5.

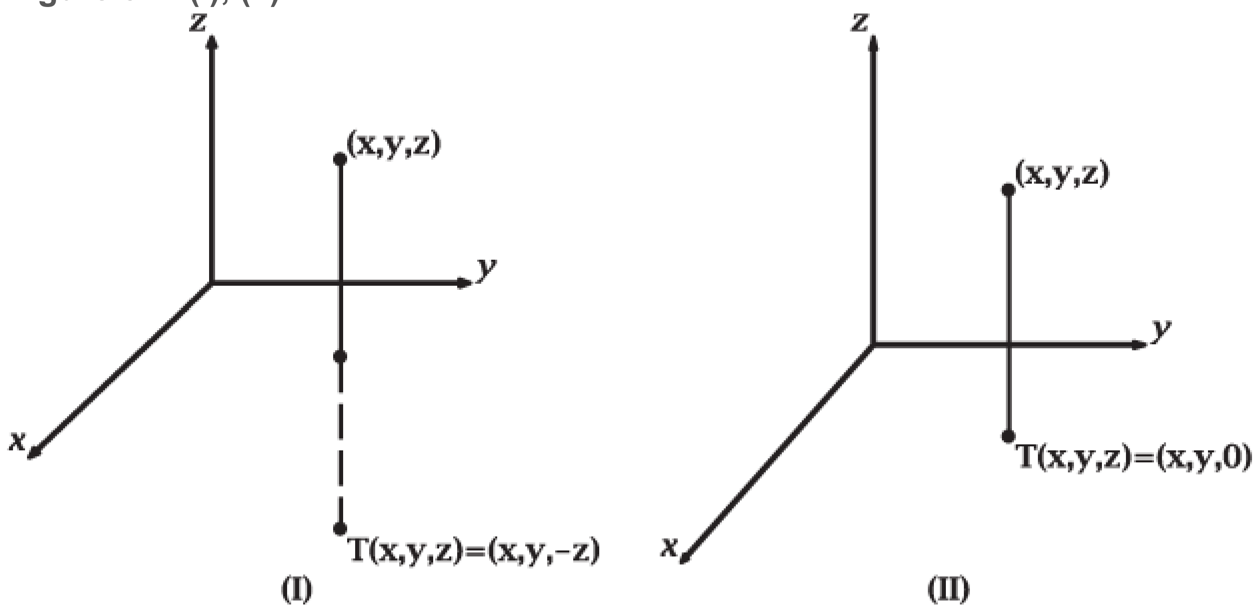
Find the standard matrix for the linear operator on $\mathbb{R}^2$ that rotates by θ_1 and then rotates by θ_2.

Solution

By **(5.3.6)** with $\theta = \theta_1$ and $\theta = \theta_2$, the standard matrix, denoted by *A*, for the linear operator is

$$\begin{aligned} A &= \begin{pmatrix} \cos\theta_2 & -\sin\theta_2 \\ \sin\theta_2 & \cos\theta_2 \end{pmatrix} \begin{pmatrix} \cos\theta_1 & -\sin\theta_1 \\ \sin\theta_1 & \cos\theta_1 \end{pmatrix} \\ &= \begin{pmatrix} \cos\theta_2\cos\theta_1 - \sin\theta_2\sin\theta_1 & -\cos\theta_2\sin\theta_1 - \sin\theta_2\cos\theta_1 \\ \sin\theta_2\cos\theta_1 + \cos\theta_2\sin\theta_1 & -\sin\theta_2\sin\theta_1 + \cos\theta_2\cos\theta_1 \end{pmatrix} \\ &= \begin{pmatrix} \cos(\theta_1+\theta_2) & -\sin(\theta_2+\theta_1) \\ \sin(\theta_2+\theta_1) & \cos(\theta_1+\theta_2) \end{pmatrix}. \end{aligned}$$

Figure 5.4: (I), (II)



Projection operators in $\mathbb{R}^3$

Theorem 5.3.4.

i. *The projection on the xy-plane (see* ***Figure 6.3*** *(II)) is*

$$T\begin{pmatrix}x\\y\\z\end{pmatrix}=\begin{pmatrix}1&0&0\\0&1&0\\0&0&0\end{pmatrix}\begin{pmatrix}x\\y\\z\end{pmatrix},$$

ii. *The projection on the xz-plane is*

$$T\begin{pmatrix}x\\y\\z\end{pmatrix}=\begin{pmatrix}1&0&0\\0&0&0\\0&0&1\end{pmatrix}\begin{pmatrix}x\\y\\z\end{pmatrix}.$$

iii. *The projection on the yz-plane is*

$$T\begin{pmatrix}x\\y\\z\end{pmatrix}=\begin{pmatrix}0&0&0\\0&1&0\\0&0&1\end{pmatrix}\begin{pmatrix}x\\y\\z\end{pmatrix}.$$

Reflection operators in $\mathbb{R}^3$

Theorem 5.3.5.

1. *The reflection about the xy-plane (see* ***Figure 6.3*** *(I)) is*

$$T\begin{pmatrix}x\\y\\z\end{pmatrix}=\begin{pmatrix}1&0&0\\0&1&0\\0&0&-1\end{pmatrix}\begin{pmatrix}x\\y\\z\end{pmatrix}.$$

2. *The reflection about the xz-plane is*

$$T\begin{pmatrix}x\\y\\z\end{pmatrix}=\begin{pmatrix}1&0&0\\0&-1&0\\0&0&1\end{pmatrix}\begin{pmatrix}x\\y\\z\end{pmatrix}.$$

3. *The reflection about the yz-plane is*

$$T\begin{pmatrix}x\\y\\z\end{pmatrix}=\begin{pmatrix}-1&0&0\\0&1&0\\0&0&1\end{pmatrix}\begin{pmatrix}x\\y\\z\end{pmatrix}.$$

The rotation operators in $\mathbb{R}^3$ through angle θ of rotation

The axis of rotation is a ray emanating from the origin. The rotation is counterclockwise looking toward the origin along the axis of rotation.

An angle of rotation is said to be positive if the rotation is counterclockwise looking toward the origin along the axis of rotation, and said to be negative if the rotation is clockwise.

Theorem 5.3.6.

1. *The rotation operator in $\mathbb{R}^3$ about the positive z-axis with a positive angle θ of rotation is* (5.3.8)

$$T\begin{pmatrix}x\\y\\z\end{pmatrix}=\begin{pmatrix}\cos\theta&-\sin\theta&0\\\sin\theta&\cos\theta&0\\0&0&1\end{pmatrix}\begin{pmatrix}x\\y\\z\end{pmatrix}.$$

2. *The rotation operator in* $\mathbb{R}^3$ *about the positive x-axis with a positive angle* θ *of rotation is*
   (5.3.9)

$$T\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix}.$$

3. *The rotation operator in* $\mathbb{R}^3$ *about the positive y-axis with a positive angle* θ *of rotation is*
   (5.3.10)

$$T\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix}.$$

4. *If the axis of rotation is parallel to the vector* $\vec{u} = (a, b, c)$, *then the rotation operator with a positive angle* θ *is* $T(x, y, x) = (b_1, b_2, b_3)$, *where*

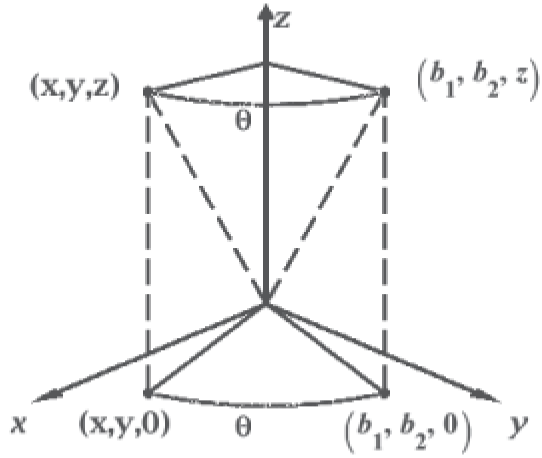
$$\begin{aligned} b_1 &= \left[a^2(1-\cos\theta)+\cos\theta\right]x + [ab(1-\cos\theta)-c\sin\theta]y + [ac(1-\cos\theta)+b\sin\theta]z, \\ b_2 &= [ab(1-\cos\theta)+c\sin\theta]x + \left[b^2(1-\cos\theta)+\cos\theta\right]y + [bc(1-\cos\theta)-a\sin\theta]z, \\ b_3 &= [ac(1-\cos\theta)-b\sin\theta]x + [bc(1-\cos\theta)+a\sin\theta]y + \left[c^2(1-\cos\theta)+\cos\theta\right]z. \end{aligned}$$

Proof

1. The rotation operator in the *xy*-plane in $\mathbb{R}^2$ is

$$T(x, y) = (x\cos\theta - y\sin\theta, x\sin\theta + y\cos\theta).$$

Figure 5.5: (I)



2. So, the rotation operator in the *xy*-plane in $\mathbb{R}^3$ (see **Figure 6.4** (I)) is

$$T(x, y, 0) = (x\cos\theta - y\sin\theta, x\sin\theta + y\cos\theta, 0).$$

   Hence, the rotation operator in $\mathbb{R}^3$ about the positive *z*-axis with θ is

$$T(x, y, z) = (x\cos\theta - y\sin\theta, x\sin\theta + y\cos\theta, z)$$

   and **(5.3.8)** holds.
3. The rotation operator in the *yz*-plane in $\mathbb{R}^2$ is obtained by replacing *x* and *y* in (5.3) by *y* and *z*, respectively, that is,

$$T(y, z) = (y\cos\theta - z\sin\theta, y\sin\theta + z\cos\theta).$$

   So, the rotation operator in the *yz*-plane in $\mathbb{R}^3$ is

$$T(0, y, z) = (0, y\cos\theta - z\sin\theta, y\sin\theta + z\cos\theta).$$

Hence, the rotation operator in $\mathbb{R}^3$ about the positive *x*-axis with θ is

$$T(x, y, z) = (x, y\cos\theta - z\sin\theta, y\sin\theta + z\cos\theta).$$

and **(5.3.9)** holds.

4. The rotation operator in the *zx*-plane in $\mathbb{R}^2$ is obtained by replacing *x* and *y* in (5.3) by *z* and *x*, respectively, that is,

$$T(z, x) = (z\cos\theta - x\sin\theta, z\sin\theta + x\cos\theta)$$

or equivalently

$$T(x, z) = (z\sin\theta + x\cos\theta, z\cos\theta - x\sin\theta).$$

The rotation operator in the *zx*-plane in $\mathbb{R}^3$ is

$$T(x, 0, z) = (z\sin\theta + x\cos\theta, 0, z\cos\theta - x\sin\theta).$$

The rotation operator in $\mathbb{R}^3$ about the positive *y*-axis with θ is

$$T(x, y, z) = (z\sin\theta + x\cos\theta, y, z\cos\theta - x\sin\theta)$$

and **(5.3.10)** holds.

5. We omit the derivation of the linear operator.
It is easy to see that **Theorem 5.3.6** (1), (2), and (3) can be obtained by **Theorem 5.3.6** (4) with $(a, b, c) = (1, 0, 0), (0, 1, 0), (0, 0, 1),$ respectively.

Example 5.3.6.

Find the standard matrix for the linear operator on $\mathbb{R}^3$ that rotates by θ about the positive *z*-axis and then reflects about the *yz*-plane and then projects on the *xy*-plane.

Solution

Let *A*, *B*, and *C* be the standard matrices, respectively, for the linear operator that rotates by θ about the positive *z*-axis, reflects about the *yz*-plane, and projects on the *xy*-plane. Then

$$A = \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}, B = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, C = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Hence, the standard matrix for the required operator is

$$\begin{aligned} CBA &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} -\cos\theta & \sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 0 \end{pmatrix}. \end{aligned}$$

Exercises

1. In **Theorem 5.3.1** , if $\theta = \frac{\pi}{6}, \frac{\pi}{3}$, find $T\begin{pmatrix} 4 \\ -4 \end{pmatrix}$.
2. Find the linear operator $T = T_2 T_1$ on $\mathbb{R}^2$, where T_1 is a dilation with factor $k = 2$ on $\mathbb{R}^2$ and T_2 is a projection on the line $y = x$.
3. 1. Find the linear operator on $\mathbb{R}^2$ that contracts with factor $\frac{1}{2}$ and then reflects about the line $y = x$.

2. Find the linear operator on $\mathbb{R}^2$ that reflects about the *y*-axis and then reflects about the *x*-axis, and the linear operator on $\mathbb{R}^2$ that reflects about the *x*-axis and then reflects about the *y*-axis.

4. Find the rotation operator with the rotation angle of $\frac{\pi}{4}$ and the image of $x = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$ under the operator.
5. Find the linear operator on $\mathbb{R}^2$ that rotates by $\frac{\pi}{3}$ radians about the origin.
6. Find the linear operator on $\mathbb{R}^3$ that rotates by $\frac{\pi}{4}$ radians about the *z*-axis and then reflects about the *yz*-plane.
7. Prove **Theorem 5.3.1** by using **Theorem 5.1.1** .
8. Prove **Theorem 5.3.2** by using a method similar to the proof of **Theorem 5.3.1** .
9. Prove **Theorem 5.3.2** by using **Theorem 5.1.1** .

Chapter 6 Lines and planes in $\mathbb{R}^3$

6.1 Cross product and volume of a parallelepiped

In Section 1.6, we see that the dot product of two vectors in $\mathbb{R}^n$ is a real number but not a vector. Now, we define a product of two vectors in $\mathbb{R}^3$ called the cross product, which is a vector in $\mathbb{R}^3$.

Let $\vec{a} = (x_1,\ x_2,\ x_3)$ and $\vec{b} = (y_1,\ y_2,\ y_3)$ be two nonzero vectors in $\mathbb{R}^3$. We want to find a vector $\vec{n} = (a,\ b,\ c)$ in $\mathbb{R}^3$ such that $\vec{n} \perp \vec{a}$ and $\vec{n} \perp \vec{b}$. By **Theorem 1.2.6** , $\vec{n} \cdot \vec{a} = 0$ and $\vec{n} \cdot \vec{b} = 0$. This implies that $\vec{n}$ satisfies the following system of linear equations

(6.1.1)

$$\begin{cases} x_1 a + x_2 b + x_3 c = 0 \\ y_1 a + y_2 b + y_3 c = 0. \end{cases}$$

Lemma 6.1.1.

Let $\vec{a} = (x_1,\ x_2,\ x_3)$ *and* $\vec{b} = (y_1,\ y_2,\ y_3)$ *be two nonzero vectors in* $\mathbb{R}^3$*. Then*

(6.1.2)

$$(a,\ b,\ c) = \left(\begin{vmatrix} x_2 & x_3 \\ y_2 & y_3 \end{vmatrix},\ -\begin{vmatrix} x_1 & x_3 \\ y_1 & y_3 \end{vmatrix},\ \begin{vmatrix} x_1 & x_2 \\ y_1 & y_2 \end{vmatrix} \right)$$

is a solution of **(6.1.1)** .

Substituting *a, b, c* given in **(6.1.2)** to **(6.1.1)** shows that (*a, b, c*) defined in **(6.1.2)** is a solution of **(6.1.1)** . Here, we provide a proof of **Lemma 6.1.1** that shows where the solution (*a, b, c*) given in **(6.1.2)** comes from.

Proof

If $\vec{a}$ is parallel to $\vec{b}$, then the three determinants in **(6.1.2)** are zero, and $(a,\ b,\ c) = (0,\ 0,\ 0)$ is a solution of **(6.1.1)** . If $\vec{a}$ is not parallel to $\vec{b}$, then one of the three determinants in **(6.1.2)** is not zero.

Without loss of generalization, we assume that $c = \begin{vmatrix} x_1 & x_2 \\ y_1 & y_2 \end{vmatrix} \neq 0$. Solving the following system by using Cramer's rule

$$\begin{cases} x_1 a + x_2 b = -x_3 c \\ y_1 a + y_2 b = -y_3 c, \end{cases}$$

we obtain

$$a = \frac{1}{c}\begin{vmatrix} -x_3 c & x_2 \\ -y_3 c & y_2 \end{vmatrix} = \begin{vmatrix} x_3 & y_3 \\ x_2 & y_2 \end{vmatrix} \text{ and } b = \frac{1}{c}\begin{vmatrix} x_1 & -x_3 c \\ y_1 & -y_3 c \end{vmatrix} = \begin{vmatrix} x_1 & y_1 \\ x_3 & y_3 \end{vmatrix}.$$

Hence, **(6.1.2)** is a solution of **(6.1.1)** .

Definition 6.1.1.

Let $\vec{a} = (x_1, x_2, x_3)$ and $\vec{b} = (y_1, y_2, y_3)$ be two vectors. The cross product $\vec{a} \times \vec{b}$ of $\vec{a}$ and $\vec{b}$ is defined by

(6.1.3)

$$\vec{a} \times \vec{b} = \left(\begin{vmatrix} x_2 & x_3 \\ y_2 & y_3 \end{vmatrix}, \ - \begin{vmatrix} x_1 & x_3 \\ y_1 & y_3 \end{vmatrix}, \ \begin{vmatrix} x_1 & x_2 \\ y_1 & y_2 \end{vmatrix} \right).$$

Let $\vec{i} = (1, 0, 0)$, $\vec{j} = (0, 1, 0)$, , and $\vec{k} = (0, 0, 1)$ be the standard vectors in $\mathbb{R}^3$. Following the minor expansion, we rewrite $\vec{a} \times \vec{b}$ into the following two forms:

(6.1.4)

$$\begin{aligned} \vec{a} \times \vec{b} &= \begin{vmatrix} x_2 & x_3 \\ y_2 & y_3 \end{vmatrix} \vec{i} - \begin{vmatrix} x_1 & x_3 \\ y_1 & y_3 \end{vmatrix} \vec{j} + \begin{vmatrix} x_1 & x_2 \\ y_1 & y_2 \end{vmatrix} \vec{k} \\ &= \begin{vmatrix} \vec{i} & \vec{j} & \vec{k} \\ x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \end{vmatrix}. \end{aligned}$$

Let $\vec{c} = (z_1, z_2, z_3)$. By **Theorem 3.1.1** and **(6.1.3)** , we have

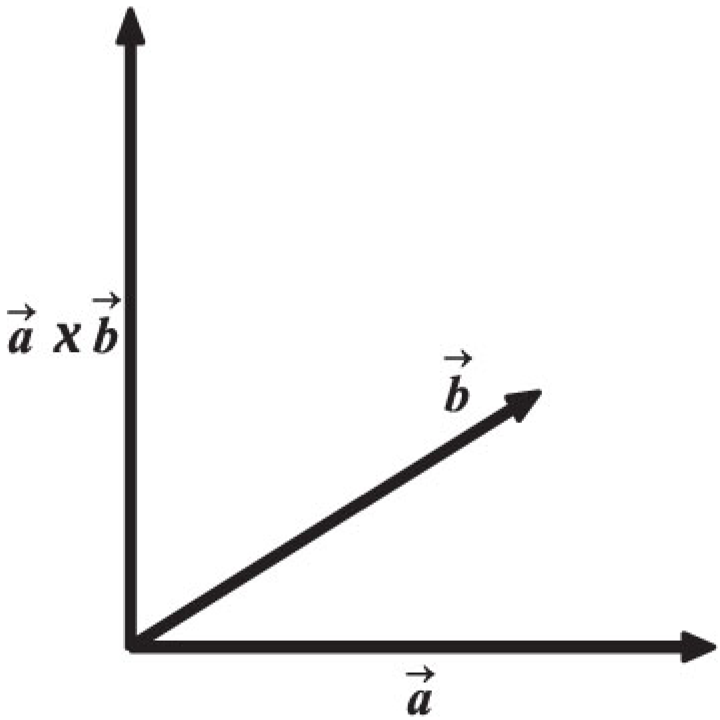
(6.1.5)

$$(\vec{a} \times \vec{b}) \cdot \vec{c} = \begin{vmatrix} x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \\ z_1 & z_2 & z_3 \end{vmatrix}.$$

The value $(\vec{a} \times \vec{b}) \cdot \vec{c}$ is called the scalar triple product of $\vec{a}$, $\vec{b}$, $\vec{c}$.

By **(6.1.5)** and **Corollary 3.3.1**, we see that $(\vec{a} \times \vec{b}) \cdot \vec{a} = 0$ and $(\vec{a} \times \vec{b}) \cdot \vec{b} = 0$. By **Theorem 1.2.6**, the cross product $\vec{a} \times \vec{b}$ is perpendicular to both vectors $\vec{a}$ and $\vec{b}$ (see **Figure 6.1** (I)), that is,

Figure 6.1: (I)



(6.1.6)

$$(\vec{a} \times \vec{b}) \perp \vec{a} \quad \text{and } (\vec{a} \times \vec{b}) \perp \vec{b}.$$

Example 6.1.1.

Let $\vec{a} = (1,\ 2,\ -2)$ and $\vec{b} = (3,\ 0,\ 1)$. Find $\vec{a} \times \vec{b}$ and verify that $(\vec{a} \times \vec{b}) \perp \vec{a}$ and $(\vec{a} \times \vec{b}) \perp \vec{b}$.

Solution

$$\vec{a} \times \vec{b} = \begin{vmatrix} \vec{i} & \vec{j} & \vec{k} \\ 1 & 2 & -2 \\ 3 & 0 & 1 \end{vmatrix} = \begin{vmatrix} 2 & -2 \\ 0 & 1 \end{vmatrix}\vec{i} - \begin{vmatrix} 1 & -2 \\ 3 & 1 \end{vmatrix}\vec{j} + \begin{vmatrix} 1 & 2 \\ 3 & 0 \end{vmatrix}\vec{k}$$

$$= 2\vec{i} - 7\vec{j} - 6\vec{k} = (2,\ -7,\ 6).$$

Because

$$(\vec{a} \times \vec{b}) \cdot \vec{a} = (2)(1) + (-7)(2) + (-6)(-2) = 0$$

and

$$(\vec{a} \times \vec{b}) \cdot \vec{b} = (2)(3) + (-7)(0) + (-6)(-1) = 0$$

Hence, $(\vec{a} \times \vec{b}) \perp \vec{a}$ and $(\vec{a} \times \vec{b}) \perp \vec{b}$.

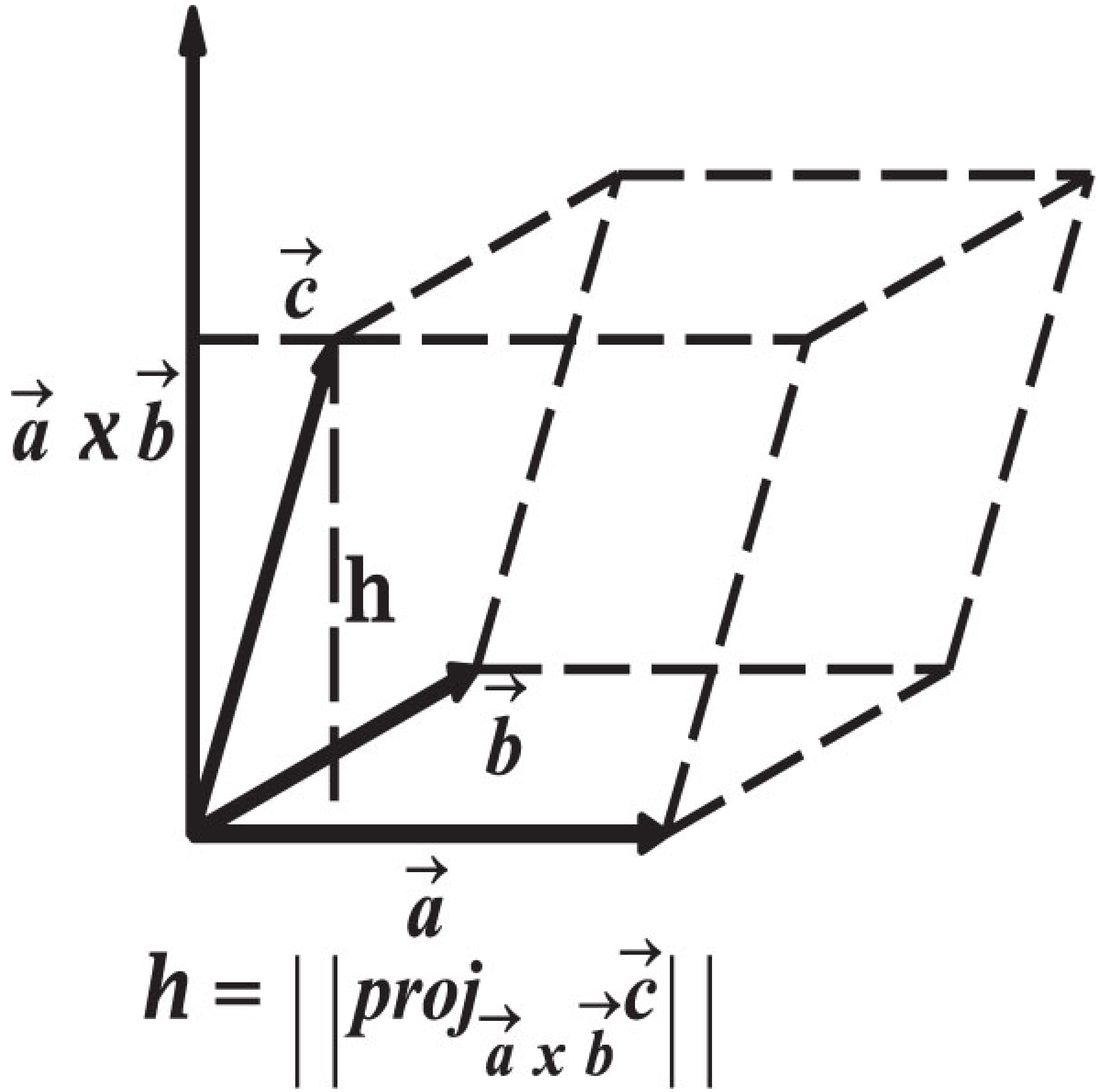
By Corollary 1.3.1 (2) and Definition 6.1.3, we obtain the following result, which shows that the norm of $\vec{a} \times \vec{b}$ is equal to the area of a parallelogram in $\mathbb{R}^3$ determined by $\vec{a}$ and $\vec{b}$.

Theorem 6.1.1.

Let $\vec{a} = (x_1, x_2, x_3)$ and $\vec{b} = (y_1, y_2, y_3)$ be two vectors. Then the area of the parallelogram determined by $\vec{a}$ and $\vec{b}$ is $\left|\left|\vec{a} \times \vec{b}\right|\right|$.

Let $\vec{a}, \vec{b},$ and $\vec{c}$ be three vectors in $\mathbb{R}^3$ that are not in the same plane but their initial points are the same. Then these vectors determine a parallelepiped (see **Figure 6.2** (I)). The purpose of this section is to derive formulas for the volume of such a parallelepiped. It is known that the volume of a parallelepiped is equal to the product of the area of its base and its height. The base of the parallelepiped is determined by any two of the vectors $\vec{a}, \vec{b},$ and $\vec{c}$ and is a parallelogram in $\mathbb{R}^3$. Here we use the base determined by $\vec{a}$ and $\vec{b}$. By **Theorem 6.1.1** , the area of the parallelogram determined by $\vec{a}$ and $\vec{b}$ is $\left|\left|\vec{a} \times \vec{b}\right|\right|$. To find the height of the parallelepiped, we use the projection of the vector $\vec{c}$ on the vector $\vec{a} \times \vec{b}$ because the norm of the projection is equal to the height of the parallelepiped (see **Figure 6.2** (I)). Based on the above ideas, we show the following two formulas for the volume of the parallelepiped determined by $\vec{a}, \vec{b},$ and $\vec{c}$.

Figure 6.2: (I)



Theorem 6.1.2.

Let $\vec{a} = (x_1, x_2, x_3)$ *and* $\vec{b} = (y_1, y_2, y_3)$, $\vec{c} = (z_1, z_2, z_3)$ *be three vectors. Then the volume of the parallelepiped determined by the three vectors is given by*

(6.1.7)

$$V = |\vec{c} \cdot (\vec{a} \times \vec{b})| = \left\| \begin{matrix} x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \\ z_1 & z_2 & z_3 \end{matrix} \right\|.$$

Proof

By **(6.1.5)** , we only prove $V = \left| \vec{c} \cdot (\vec{a} \times \vec{b}) \right|$. Let $\vec{n} = \vec{a} \times \vec{b}$. By **Theorem 6.1.1** , the area of the parallelogram determined by $\vec{a}$ and $\vec{b}$ is $\left| \left| \vec{n} \right| \right|$. The height of the parallelepiped is equal to the norm of the projection of $\vec{c}$ on $\vec{a} \times \vec{b}$ (see **Figure 6.2** (I)). By **(1.3.1)** ,

$$h = \|\mathrm{proj}_{\vec{n}} \vec{c}\| = \left\| \frac{\vec{c} \cdot \vec{n}}{\|\vec{n}\|^2} \vec{n} \right\| = \frac{\left| \vec{c} \cdot \vec{n} \right|}{\|\vec{n}\|}.$$

Hence,

$$V = \|\vec{n}\| h = \|\vec{n}\| \frac{|\vec{c} \cdot \vec{n}|}{\|\vec{n}\|} = |\vec{c} \cdot \vec{n}| = |\vec{c} \cdot (\vec{a} \times \vec{b})|.$$

Example 6.1.2.

Let $\vec{a} = (3, \ -2, \ -5)$, $\vec{b} = (1, \ 4, \ -4)$, and $\vec{c} = (0, \ -3, \ -2)$. Use the two formulas in **Theorem 6.1.2** to find the volume of the parallelepiped determined by $\vec{a}$, $\vec{b}$, and $\vec{c}$.

Solution

By **(6.1.3)** , we have

$$\vec{b}\times\vec{c} = \begin{vmatrix} \vec{i} & \vec{j} & \vec{k} \\ 1 & 4 & -4 \\ 0 & -3 & -2 \end{vmatrix} = \begin{vmatrix} 4 & -4 \\ -3 & -2 \end{vmatrix}\vec{i} - \begin{vmatrix} 1 & -4 \\ 0 & -2 \end{vmatrix}\vec{j} + \begin{vmatrix} 1 & 4 \\ 0 & -3 \end{vmatrix}\vec{k}$$

$$= (-8-12)\vec{i} - (-2-0)\vec{j} + (-3-0)\vec{k} = (-20,\ 2,\ -3).$$

Hence, we obtain

$$V = |\vec{a}\cdot(\vec{b}\times\vec{c})| = |(3)(-20)+(-2)(2)+(-5)(-3)| = |-49|.$$

By **(3.1.4)** , we have

$$V = \left| \begin{vmatrix} 3 & -2 & -5 \\ 1 & 4 & -4 \\ 0 & -3 & -2 \end{vmatrix} \begin{matrix} 3 & -2 \\ 1 & 4 \\ 0 & -3 \end{matrix} \right| = |-49| = 49.$$

Exercises

1. Find $\vec{a}\times\vec{b}$ and verify $(\vec{a}\times\vec{b})\perp\vec{a}$ and $(\vec{a}\times\vec{b})\perp\vec{b}$.
    1. $\vec{a} = (1,\ 2,\ 3);\ \vec{b} = (1,\ 0,\ 1),$
    2. $\vec{a} = (1,\ 0,\ -3),\ \vec{b} = (-1,\ 0,\ -2),$
    3. $\vec{a} = (0,\ 2,\ 1);\ \vec{b} = (-5,\ 0,\ 1),$
    4. $\vec{a} = (1,\ 1,\ -1);\ \vec{b} = (-1,\ 1,\ 0),$

2. Let $\vec{a} = (1,\ -2,\ 3),\ \vec{b} = (-1,\ -4,\ 0),$ and $\vec{c} = (2,\ 3,\ 1).$ Find the scalar triple product of $\vec{a},\ \vec{b},\ \vec{c}$.
3. Find the area of the parallelogram determined by $\vec{a}$ and $\vec{b}$.
    1. $\vec{a} = (1,\ 0,\ 1),\ \vec{b} = (-1,\ 0,\ 1);$

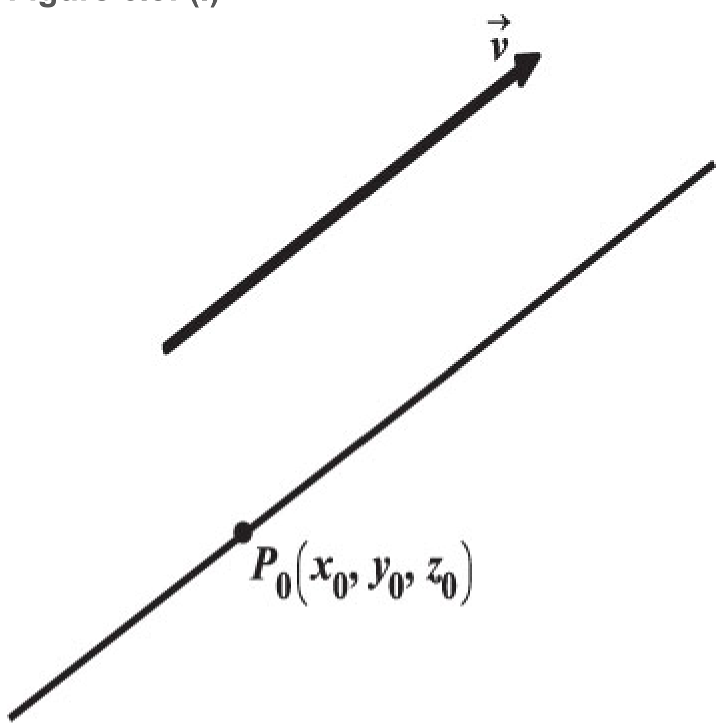
2. $\vec{a} = (1,\ 0,\ 0)$, $\vec{b} = (-1,\ 1,\ -2)$.

4. Find the volume of the parallelpiped determined by $\vec{a}$, $\overrightarrow{b}$, and $\vec{c}$.
   1. $\vec{a} = (1,\ -1,\ 0)$, $\vec{b} = (-1,\ 0,\ 2)$, $\vec{c} = (0,\ -1,\ -1)$.
   2. $\vec{a} = (-1,\ -1,\ 0)$, $\vec{b} = (-1,\ 1,\ -2)$, $\vec{c} = (-1,\ 1,\ 1)$.

6.2 Equations of lines in $\mathbb{R}^3$

In the *xy*-plane, we can obtain the equation of a (straight) line if we know a point on the line and the slope of the line. The slope of the line provides the direction of the line. In $\mathbb{R}^3$, the basic idea is the same. To establish an equation of a line in $\mathbb{R}^3$, we need a point on the line and a vector parallel to the line (see **Figure 6.3** (I)).

Figure 6.3: (I)



Definition 6.2.1.

A vector $\vec{v}$ in $\mathbb{R}^3$ is said to be parallel to a line in $\mathbb{R}^3$ if it is parallel to $\overrightarrow{PQ}$ for any two different points *P* and *Q* on the line.

Note that if P_1 and Q_1 are two other different points on the line, then $\overrightarrow{PQ}$ is parallel to $\overrightarrow{P_1Q_1}$ because $\overrightarrow{PQ}$ and $\overrightarrow{P_1Q_1}$ have the same or opposite direction. By **Definition 1.1.6** , if a vector is parallel to $\overrightarrow{PQ}$, it is parallel to $\overrightarrow{P_1Q_1}$. Hence, in **Definition 6.2.1** , $\vec{v}$ is parallel to $\overrightarrow{PQ}$ for any two different points *P* and *Q* on the line if and only if there exist two different points *P* and *Q* on the line such that $\vec{v}$ is parallel to $\overrightarrow{PQ}$.

Definition 6.2.2.

Let $P_0(x_0, y_0, z_0)$be a point in $\mathbb{R}^3$ and let $\vec{v} = (a,\ b,\ c)$ be a nonzero vector in $\mathbb{R}^3$. Then the set of all points *P*(*x*, *y*, *z*) in $\mathbb{R}^3$ satisfying $\overrightarrow{P_0P} \parallel \vec{v}$ constitutes a line in $\mathbb{R}^3$.

By **Definitions 6.2.1** and **6.2.2** , we see that a line passing through the point P_0 that is parallel to $\vec{v}$ is the same line in **Definition 6.2.2** .

The following result gives a parametric equation of a line passing through a point P_0 that is parallel to a given vector $\vec{v}$.

Theorem 6.2.1.

Let $P_0(x_0,\ y_0,\ z_0)$ be a point in $\mathbb{R}^3$ and let $\vec{v} = (a,\ b,\ c)$ be a nonzero vector in $\mathbb{R}^3$. Then the parametric equation of the line passing through P_0 that is parallel to $\vec{v}$ is

(6.2.1)

$$\begin{cases} x = x_0 + ta \\ y = y_0 + tb & -\infty < t < \infty \\ z = z_0 + tc \end{cases}$$

Proof

Let *P*(*x*, *y*, *z*) be any point on the line. Then $\overrightarrow{P_0P} = (x - x_0,\ y - y_0,\ z - z_0)$ and $\overrightarrow{P_0P}$ is parallel to $\vec{v}$ $\overrightarrow{P_0P} = (x - x_0,\ y - y_0,\ z - z_0)$. By **Definition 1.1.6**, there exists a constant $t \in \mathbb{R}$ that depends on *P* such that $\overrightarrow{P_0P} = t\vec{v}$. This, together with **Definition 1.1.4**, implies **(6.2.1)** holds.

By **(6.2.1)**, we see that the solution **(4.3.1)** of the system *d*) in **Example 4.3.1** is a parametric equation of a line.

By the proof of **Theorem 6.2.1**, we see that for each point (*x, y, z*) on the line, there exists a unique number $t \in \mathbb{R}$ such that **(6.2.1)** holds. Conversely, by **(6.2.1)**, we see that for each $t \in \mathbb{R}$, there exists a unique point (*x, y, z*)on the line that satisfies **(6.2.1)**. Hence, there is a one to one relation between the points on the line and the numbers $t \in \mathbb{R}$.

The steps to find a parametric equation of a line are:

1. Step 1. Find a point $P_0(x_0,\ y_0,\ z_0)$ on the line.
2. Step 2. Find a vector $\vec{v}$ that is parallel to the line.
3. Step 3. Write the equation by using **(6.2.1)**.

Example 6.2.1.

a. Find the parametric equation of the line passing through the point $P_0(1,\ 2,\ 3)$ that is parallel to the vector $\vec{v} = (2,\ 3,\ -1)$.
b. Where does the line intersect the *yz*-plane?

Solution

a. By **(6.2.1)**, the parametric equation of the line is
(6.2.2)

$$x = 1 + 2t, \quad y = 2 + 3t, \quad z = 3 - t.$$

b. Because the line intersects the *yz*-plane, $x = 0$. By the first equation of **(6.2.2)** with $x = 0$, we obtain $t = -\frac{1}{2}$. Substituting $t = -\frac{1}{2}$ into the second and third equations of **(6.2.2)** implies that $y = \frac{1}{2}$ and $z = \frac{7}{2}$. Hence, the intersection point is $\left(0,\ \frac{1}{2},\ \frac{7}{2}\right)$.

Example 6.2.2.

Find an equation of the line that passes through $P_1(1,\ -1,\ 1)$ and $P_2(2,\ 1,\ 0)$.

Solution

Let $P_0(x_0,\ y_0,\ z_0) = P_1(1,\ -1,\ 1)$ and $\vec{v} = (a,\ b,\ c) = \overrightarrow{P_1P_2} = (1,\ 2,\ -1)$. By **(6.2.1)**, we have $x = 1 + t,\ y = -1 + 2t,\ z = 1 - t$.

Now, we discuss the distance from a point to a given line in $\mathbb{R}^3$, that is, the distance between the point and the intersection point of the given line and the line passing through the point that is orthogonal to the given line.

Theorem 6.2.2.

The distance D from $Q(x_1, y_1, z_1)$ to the line **(6.2.1)** *is*

$$D = \frac{\|\vec{v} \times \overrightarrow{PQ}\|}{\|\vec{v}\|} \text{ for any point } P \text{ on the line.}$$

Proof

Let *P* be a point on the line. Then $D = \|\overrightarrow{PQ} - \text{proj}_{\vec{v}}\overrightarrow{PQ}\|$. It follows from **(1.3.4)** and **Theorem 6.1.1** that

$$D = \|\overrightarrow{PQ} - \text{proj}_{\vec{v}}\overrightarrow{PQ}\| = \frac{\|\vec{v} \times \overrightarrow{PQ}\|}{\|\vec{v}\|}.$$

Example 6.2.3.

Find the distance from the point $Q(1, 0, -1)$ to the line

(6.2.3)

$$x = t, \quad y = -1 + t, \quad z = 1 - t.$$

Solution

Let $t = 0$. Then $P(0, -1, 1)$ is on the line and $\overrightarrow{PQ} = (1, 1, -2)$. By **(6.2.3)** , $\vec{v} = (1, 1, -1)$, which is parallel to the line. By computation, $\|\vec{v}\| = \sqrt{3}$,

$$\vec{v} \times \overrightarrow{PQ} = \left(\begin{vmatrix} 1 & -1 \\ 1 & -2 \end{vmatrix}, -\begin{vmatrix} 1 & -1 \\ 1 & -2 \end{vmatrix}, \begin{vmatrix} 1 & 1 \\ 1 & 1 \end{vmatrix} \right) = (-1, 1, 0)$$

and $\|\vec{v} \times \overrightarrow{PQ}\| = \sqrt{2}$. By **Theorem 6.2.2** ,

$$D = \frac{\|\vec{v} \times \overrightarrow{PQ}\|}{\|\vec{v}\|} = \frac{\sqrt{6}}{3}.$$

By **Theorem 6.2.2** , we derive the distance from a point in $\mathbb{R}^2$ to a line

(6.2.4)

$$Ax + By = C,$$

where $A \neq 0$ or $B \neq 0$ in the *xy*-plane.

Theorem 6.2.3.

The distance D from $Q(x_1, y_1)$ *to the line* **(6.2.4)** *is*

(6.2.5)

$$D = \frac{\left|Ax_1 + By_1 - C\right|}{\sqrt{A^2 + B^2}}.$$

Proof

Without loss of generalization, we assume that $A \neq 0$. Let $y = t$. Solving **(6.2.4)** for *x*, we obtain

$$x = \frac{C}{A} - \frac{B}{A}t.$$

Hence, the parametric equation of the line **(6.2.4)** in $\mathbb{R}^3$ is

(6.2.6)

$$x = \frac{C}{A} - \frac{B}{A}t, \quad y = t, \quad z = 0 + 0t.$$

Note that *D* is equal to the distance from $Q^*(x_1, y_1, 0)$ to the line **(6.2.6)**. From **(6.2.6)**, $\vec{v} = \left(-\frac{B}{A}, 1, 0\right)$ and the point $P^*\left(\frac{C}{A}, 0, 0\right)$ is on the line **(6.2.6)**. By computation,

$$\|\vec{v}\| = \frac{\sqrt{A^2+B^2}}{|A|} \quad \text{and } \|\vec{v} \times \overrightarrow{P^*Q^*}\| = \frac{\left|Ax_1+By_1-C\right|}{|A|}.$$

By **Theorem 6.2.2**,

$$D = \frac{\|\vec{v} \times \overrightarrow{P^*Q^*}\|}{\|\vec{v}\|} = \frac{\left|Ax_1 + By_1 - C\right|}{\sqrt{A^2 + B^2}}.$$

Example 6.2.4.

Find the distance from Q(1, 1) to the line $y = x + 1$.

Solution

Rewrite $y = x + 1$ as $-x + y = 1$. By **Theorem 6.2.3**, we have

$$D = \frac{\left|Ax_1 + By_1 - C\right|}{\sqrt{A^2 + B^2}} = \frac{|(-1)(1) + 1 - 1|}{\sqrt{(-1)^2 + 1^2}} = \frac{1}{\sqrt{2}}.$$

Exercises

1. a. Find the parametric equation of the line passing through the point $P_0(-1,\ 2,\ 0)$ that is parallel to the vector $\vec{v} = (1,\ -1,\ 1)$.
   b. Where does the line intersect the *xy*-plane?

2. Find the parametric equation of the line that passes through $P_1(2,\ 1,\ -1)$ and $P_2(-2,\ 3,\ -5)$.
3. Find the distance from the point $Q(-1,\ 2,\ 1)$ to the line

$$x = 1 + t, \quad y = 2 - 4t, \quad z = -3 + 2t.$$

4. Find the distance from the point $Q(-1,\ 2)$ to the line $y = -2x + 4$.

6.3 Equations of planes in $\mathbb{R}^3$

In **Section 6.2**, we see that the parametric equation of a line in $\mathbb{R}^3$ is established according to a given point on the line and a given vector that is parallel to the line. In this section, we establish an equation of a plane in $\mathbb{R}^3$.

Definition 6.3.1.

Let $P_0(x_0, y_0, z_0)$ be a point in $\mathbb{R}^3$ and $\vec{n} = (a, b, c)$ a vector in $\mathbb{R}^3$. Then the set of all points *P*(*x, y, z*)in $\mathbb{R}^3$ satisfying

$$\vec{n} \cdot \overrightarrow{P_0P} = 0$$

constitutes a plane in $\mathbb{R}^3$. This vector $\vec{n}$ is called a normal vector of the plane.

Let $P_0(x_0, y_0, z_0)$ be a point in $\mathbb{R}^3$ and let $\vec{n} = (a, b, c)$ be a normal vector of a plane in $\mathbb{R}^3$. We derive the equation of the plane passing through $P_0(x_0, y_0, z_0)$ and having the normal vector $\vec{n} = (a, b, c)$, which is called the point-normal form equation of the plane.

Theorem 6.3.1.

The equation of the plane passing through $P_0(x_0, y_0, z_0)$ *with the normal vector* $\vec{n} = (a, b, c)$ *is*

(6.3.1)

$$ax + by + cz = d,$$

where

(6.3.2)

$$d = (a,\ b,\ c) \cdot \begin{pmatrix} x_0 \\ y_0 \\ z_0 \end{pmatrix} = ax_0 + by_0 + cz_0.$$

Proof

Let *P*(*x, y, z*) be a point in the plane. Then

$$\overrightarrow{P_0P} = (x - x_0,\ y - y_0,\ z - z_0).$$

By **Definition 6.3.1** , $\vec{n} \cdot \overrightarrow{P_0P} = 0,$ which implies

$$a(x - x_0) + b(y - y_0) + c(z - z_0) = 0.$$

This implies $ax + by + cz = ax_0 + by_0 + cz_0$

Note that equation **(6.3.1)** is the same as **(6.3.1)** , so it is a linear equation with three variables *x, y, z*. The coefficients *a, b, c* of *x, y, z* of **(6.3.1)** are the components of a normal vector of the plane.

Example 6.3.1.

Find a normal vector of the plane $2x + y - z = 3.$.

Solution

By **(6.3.1)**, $\vec{n} = (2,\ 1,\ -1)$.

By **Definition 6.3.1**, we see that the normal vector $\vec{n}$ is orthogonal to the vector $\overrightarrow{P_0P}$ for every point *P* in the plane. Moreover, if $\vec{v}$ is a normal vector of the plane, then $k\vec{n}$ also is a normal vector of the plane for every $k \in \mathbb{R}$ with $k \neq 0$.

The following result shows that the normal vector $\vec{n}$ is orthogonal to the vector $\overrightarrow{PQ}$ for any two different points *P* and *Q* in the plane.

Theorem 6.3.2.

Let $\vec{n}$ be the normal vector of a plane that passes through a point $P_0(x_0,\ y_0,\ z_0)$. Then $\vec{v}$ is orthogonal to any vector $\overrightarrow{PQ}$, where P and Q are any two different points in the plane.

Proof

Let *P* and Q be any two different points in the plane. By **Definition 6.3.1**, $\vec{n} \cdot \overrightarrow{P_0P} = 0$ and $\vec{n} \cdot \overrightarrow{P_0Q} = 0$. Because $\overrightarrow{PQ} = \overrightarrow{P_0Q} - \overrightarrow{P_0P}$, we have

$$\vec{n} \cdot \overrightarrow{PQ} = (\vec{n} \cdot \overrightarrow{P_0Q}) - (\vec{n} \cdot \overrightarrow{P_0P}) = 0.$$

.

Hence $\vec{n}$ is orthogonal to $\overrightarrow{PQ}$.

Definition 6.3.2.

A vector $\vec{v}$ in $\mathbb{R}^3$ is said to be orthogonal to a plane if it is orthogonal to $\overrightarrow{PQ}$ for any two different points *P* and *Q* in the plane. A vector $\vec{v}$ in $\mathbb{R}^3$ is said to be orthogonal to a line in $\mathbb{R}^3$ if it is orthogonal to $\overrightarrow{PQ}$ for any two different points *P* and *Q* on the line.

By **Definition 6.3.2**, we see that a vector in $\mathbb{R}^3$ is orthogonal to a plane if and only if it is orthogonal to any line in the plane. Also, by **Definitions 6.3.1** and **6.3.2** and **Theorem 6.3.2**, we see that a normal vector of a plane is orthogonal to the plane as well as any line in the plane.

Definition 6.3.3.

A vector in $\mathbb{R}^3$ is said to be parallel to a plane in $\mathbb{R}^3$ if there exists a line in the plane that is parallel to the vector.

By **Definitions 6.2.1** and **6.3.3**, we obtain the following results. Its proof is straightforward and is left to the reader.

Proposition 6.3.1.

1. *If* $P_1(x_1, y_1, z_1)$ *and* $P_2(x_2, y_2, z_2)$ *are two different points in a plane, then the vector* $\overrightarrow{P_1P_2}$ *is parallel to the plane.*
2. *A vector* $\vec{n}$ *is orthogonal to a plane if and only if it is orthogonal to any vector that is parallel to the plane.*
3. *If both* $\vec{a}$ and $\vec{b}$ *are parallel to a plane and* $\vec{a}$ *is not parallel to* $\vec{b},$ *then the cross product* $\vec{a} \times \vec{b}$ *is orthogonal to the plane.*

Now, we consider the following two planes linear system

(6.3.3)

$$\begin{cases} a_1x + b_1y + c_1z = d_1 \\ a_2x + b_2y + c_2z = d_2, \end{cases}$$

which is a system of two linear equations.

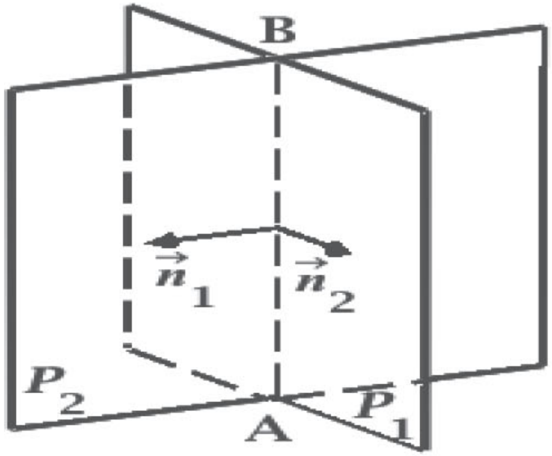
By **Theorem 4.5.1** , **(6.3.3)** has infinitely many solutions or no solutions because $r(A \mid \vec{b}) \leq 2 < 3 = n$, where A and $(A \mid \vec{b})$ are the coefficient matrix and the augmented matrix of **(6.3.3)** , respectively. Geometrically, the two planes are the same or intersect in a line or are parallel. Hence, if they are not parallel, then they intersect in a straight line.

Definition 6.3.4.

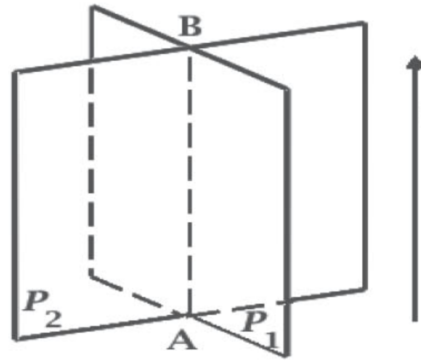
Two planes are said to be parallel if their normal vectors are parallel, and to be orthogonal if their normal vectors are orthogonal.

By **Definitions 1.1.6** and **1.2.5** and **Theorem 1.2.6** , we obtain the following results, (see **Figure 6.4** (III)).

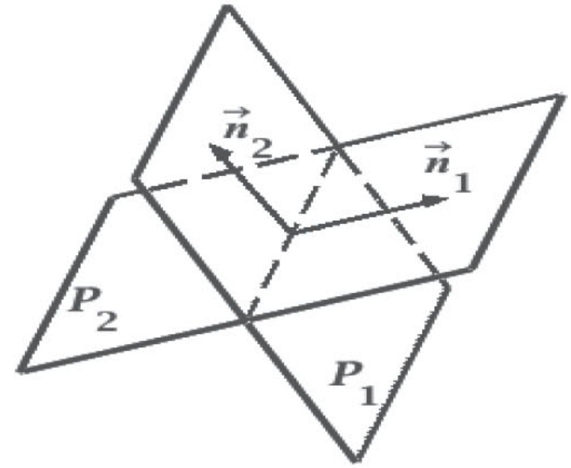
Figure 6.4: (I), (II), (III)



(I)



(II)



(III)

Theorem 6.3.3.

Let $\vec{n}_1$ *and* $\vec{n}_2$ *be the normal vectors of two planes* p_1 *and* p_2, *respectively. Then the following assertions hold.*

i. $p_1 \parallel p_2$ *if and only if there exists* $k \in \mathbb{R}$ *such that* $\vec{n}_2 = k\vec{n}_1$.

ii. $p_1 \perp p_2$ *if and only if* $\vec{n}_1 \cdot \vec{n}_2 = 0$.

Example 6.3.2.

For each pair of the following planes, determine whether it is parallel or orthogonal.

a. $-x - y + z = 2, \quad 3x + 3y - 3z = 5;$
b. $2x - y + z = 1, \quad 4x - 2y + 2z = 2;$
c. $2x - 2y + 3z = 1, \quad x + y + z = 2;$
d. $2x - y + z = 2, \quad x + 3y + z = 2.$

Solution

a. Let $\overrightarrow{n_1} = (-1, \; -1, \; 1)$ and $\overrightarrow{n_2} = (4, \; -2, \; 2)$. Then $\overrightarrow{n_2} = -3\overrightarrow{n_1}$ and $\overrightarrow{n_1} \| \overrightarrow{n_2}$. The two planes are parallel.

b. Let $\overrightarrow{n_1} = (2, \; -1, \; 1)$ and $\overrightarrow{n_2} = (4, \; -2, \; 2)$. Then $\overrightarrow{n_2} = 2\overrightarrow{n_1}$ and $\overrightarrow{n_1} \| \overrightarrow{n_2}$. By **Theorem 6.3.3** , the two planes are parallel.

c. Let $\overrightarrow{n_1} = (2, \; -2, \; 3)$ and $\overrightarrow{n_2} = (1, \; 1, \; 1)$. Then $\overrightarrow{n_2} \neq k\overrightarrow{n_1}$ for any $k \in \mathbb{R}$ and by **Theorem 6.3.3** , the two planes are not parallel. Because $\overrightarrow{n_1} \times \overrightarrow{n_2} = 3 \neq 0,$ the two planes are not orthogonal.

d. Let $\overrightarrow{n_1} = (2, \; -1, \; 1)$ and $\overrightarrow{n_2} = (1, \; 3, \; 1)$. Because $\overrightarrow{n_1} \times \overrightarrow{n_2} = 0,$ the two planes are orthogonal.

As mentioned in **Remark 4.3.1** , the two nonparallel planes $x_1 + x_2 - 2x_3 = 1$ and $-2x_1 - x_2 + x_3 = -2$ in the system *d*) in **Example 4.3.1** have a line of intersection.

Example 6.3.3.

Find the parametric equation of the line of intersection of the planes $x - y - z = 1$ and $x + 2y - 4z = 2.$

Analysis

We need to find a point on the line and a vector parallel to the line as we mentioned above. The cross product of the normal vectors of the two planes is parallel to the line of intersection of the two planes, so we would choose it. But how to find a point on the line of intersection of the two planes? To find a point on the line of intersection, we must solve the system of the two equations. When we solve the system, we shall see that we automatically obtain the equation of the line of intersection. Hence, the methodology to find the equation of the line of intersection of the two planes is simply to solve the system of the two equations.

Solution

Consider the following system

(6.3.4)

$$\begin{cases} x-y-z=1 \\ x+2y-4z=4. \end{cases}$$

$$(A\,|\,\vec{b}) = \begin{pmatrix} 1 & -1 & -1 & 1 \\ 1 & 2 & -4 & 4 \end{pmatrix} R_1(-1)+R_2 \rightarrow \begin{pmatrix} 1 & -1 & -1 & 1 \\ 0 & 3 & -3 & 3 \end{pmatrix}$$

$$R_2\left(\frac{1}{3}\right) \rightarrow \begin{pmatrix} 1 & -1 & -1 & 1 \\ 0 & 1 & -1 & 1 \end{pmatrix} R_2(1)+R_1 \rightarrow \begin{pmatrix} 1 & 0 & -2 & 2 \\ 0 & 1 & -1 & 1 \end{pmatrix}.$$

The system corresponding to the last augmented matrix is

$$\begin{cases} x-2z=2 \\ y-z=1, \end{cases}$$

where *x, y* are basic variables and *z* is a free variable. Let $z=t$. Then $y=1+t$ and $x=2+2t$. The parametric equation of the line of intersection of the planes is $x=2+2t,\ y=1+t,\ z=t$.

In **Example 6.3.3**, we solve the system **(6.3.4)** to obtain the parametric equation of the line of intersection of the two planes. From this line, we can obtain a point, say, $P_0(2,\ 1,\ 0)$ on the line, and a vector, say, $\vec{v}=(2,\ 1,\ 1)$ that is parallel to the line of intersection. In some cases, we do not want any points on the the line of intersection, but we only want a vector that is parallel to the line of intersection.

The following result provides such a vector that is parallel to the line of intersection of two planes (see **Figure 6.4** (I)), where we do not need to solve the system of the two linear equations.

Theorem 6.3.4.

Assume that the two planes given in **(6.3.3)** *are not parallel. Then* $\vec{n}_1 \times \vec{n}_2$ *is parallel to the intersection line of the two planes.*

Proof

We denote by p_1 and p_2 the two planes and by *l* the line of intersection of the two planes. Because $\vec{n}_1$ is perpendicular to the plane p_1, it is perpendicular to the line *l*. Similarly, $\vec{n}_2$ is perpendicular to the line *l*. Hence, *l* is perpendicular to $\vec{n}_1$ and $\vec{n}_2$. This implies that *l* is parallel to $\vec{n}_1 \times \vec{n}_2$.

By **Theorem 6.3.4**, we obtain the following result (see **Figure 6.4** (II)).

Theorem 6.3.5.

A vector is parallel to two nonparallel planes if and only if it is parallel to the intersection line of the two nonparallel planes.

Proof

If a vector is parallel to two nonparallel planes, then it is orthogonal to the normal vectors of the two planes and thus, it is parallel to $\vec{n}_1 \times \vec{n}_2$. It follows from **Theorem 6.3.4** that the vector is parallel to the intersection line of the two planes. Conversely, if a vector is parallel to the intersection line of two planes, then by **Definition 6.3.3**, the vector is parallel to each of the two planes.

Example 6.3.4.

Find an equation of the line passing through *P*(1, 1, 1) that is parallel to the planes $2x+y-z=1$ and $x-y+z=-2$.

Analysis

Before we give the solution, we analyze the problem of finding the equation of the line. Recall the steps for finding an equation of a line given in **Section 6.2** . To establish an equation of a line, we need to find a point on the line and a vector parallel to the line. Because the line passes through *P*(1, 1, 1), the point *P*(1, 1, 1) is on the line that we need. Next, we need to find a vector that is parallel to the line. Because the line we seek is assumed to be parallel to the two planes $2x+y-z=1$ and $x-y+z=-2$, by **Theorem 6.3.5** , the line is parallel to the intersection line of the two planes. By **Theorem 6.3.4** , the cross product of the two normal vectors of the two planes is parallel to the intersection line of the two planes and thus, it is parallel to the line we seek. Hence, we can choose a vector that is parallel to the cross product of the two normal vectors.

Solution

Let $\overrightarrow{n_1}=(2,\ 1,\ -1)$ and $\overrightarrow{n_2}=(1,\ 1,\ -1)$ be the normal vectors of the two planes. Because $\overrightarrow{n_1}$ is not parallel to $\overrightarrow{n_2}$, the two planes intersect in a line. By **Theorem 6.3.4** , $\overrightarrow{n_1}\times\overrightarrow{n_2}$ is parallel to the line of intersection. By computation, we have $\overrightarrow{n_1}\times\overrightarrow{n_2}=(0,\ -3,\ -3)=-3(0,\ 1,\ 1)$. Let $\vec{v}=(0,\ 1,\ 1)$. Then $\vec{v}\|\overrightarrow{n_1}\times\overrightarrow{n_2}$. By **(6.2.1)** ,

$$x=1,\quad y=1+t,\quad z=1+t$$

is an equation of the line.

Now, we study how to find an equation of a plane.

The steps to find an equation of a plane are

1. Step 1. Find a point $P_0(x_0,\ y_0,\ z_0)$ in the plane.
   The point $P_0(x_0,\ y_0,\ z_0)$ is often given directly or it is from a line in the plane. If a line is parallel to the plane, then we would not take any points from this line as the point $P_0(x_0,\ y_0,\ z_0)$.
2. Step 2. Find a vector $\vec{n}$ that is perpendicular to the plane, that is, find a normal vector of the plane.
   By **Proposition 6.3.1** (3), to find a normal vector $\vec{n}$ of the plane, we need to find two nonparallel vectors $\vec{a}$ and $\vec{b},$ which are parallel to the plane, and choose $\vec{n} = k\vec{a} \times \vec{b}$ for some $k \neq 0$. In particular, choose $\vec{n} = \vec{a} \times \vec{b}$.
3. Step 3. Write the equation by using **(6.3.1)** .

Example 6.3.5.

Find the equation of the plane passing through $(1,\ -1,\ -2)$ and perpendicular to $\vec{n} = (2,\ 1,\ -1)$.

Solution

$2x + y - z = (2,\ 1,\ -1) \cdot (1,\ -1,\ -2) = 3.$

Example 6.3.6.

Find the equation of the plane passing through

$$P_1(-1,\ 1,\ 1),\ \ P_2(-2,\ 0,\ 1),\ \ P_3(0,\ 1,\ -2).$$

Solution

1. Step 1. We can choose one of the three points as a point in the plane, say, $P_0(x_0,\ y_0,\ z_0) = P(-1,\ 1,\ 1)$.
2. Step 2. To find the normal vector $\vec{n},$ we need to find two nonparallel vectors $\vec{a}$ and $\vec{b},$ both of which are parallel to the plane. By **Proposition 6.3.1** (1), $\overrightarrow{P_1P_2} = (-1,\ -1,\ 0)$ and

$\overrightarrow{P_1P_3} = (1,\ 0,\ -3)$ are parallel to the plane. It is obvious that $\overrightarrow{P_1P_2}$ is not parallel to $\overrightarrow{P_1P_3}$ because $(-1,\ -1,\ 0) \neq k(1,\ 0,\ -3)$ for any $k \in \mathbb{R}$. By computation, we have

$$\begin{aligned} \overrightarrow{P_1P_2} \times \overrightarrow{P_1P_3} &= \begin{vmatrix} \vec{i} & \vec{j} & \vec{k} \\ -1 & -1 & 0 \\ 1 & 0 & -3 \end{vmatrix} \\ &= \left(\begin{vmatrix} -1 & 0 \\ 0 & -3 \end{vmatrix},\ -\begin{vmatrix} -1 & 0 \\ 1 & -3 \end{vmatrix},\ \begin{vmatrix} -1 & -1 \\ 1 & 0 \end{vmatrix} \right) \\ &= (3,\ -3,\ 1). \end{aligned}$$

We choose $\vec{n} = \overrightarrow{P_1P_2} \times \overrightarrow{P_1P_3} = (3,\ -3,\ 1)$ as the normal vector of the plane.

3. Step 3. By **(6.3.1)** and **(6.3.2)** ,

$$3x - 3y + z = (3,\ -3,\ 1) \cdot (-1,\ 1,\ 1) = -5.$$

Example 6.3.7.

Find an equation for the plane containing the line

(6.3.5)

$$x = 1 + t,\quad y = -1 - t,\quad z = -3 + 2t$$

and that is parallel to the line of intersection of the planes $x - y - z = 1$ and $x - 2y + z = 3$.

Analysis

To establish an equation of a plane, we need to find a point in the plane and a normal vector of the plane. From the given line **(6.3.5)** , we can obtain two things: one is the point $(1,\ -1,\ -3)$ when $t \neq 0$ and another is the vector $\vec{v} = (1,\ -1,\ 2)$, which is parallel to the line, so $\vec{v} = (1,\ -1,\ 2)$, is parallel to the plane we seek. By **Theorem 6.3.5** , the cross product $\overrightarrow{n_1} \times \overrightarrow{n_2}$ of the normal vectors of the two planes is parallel to the intersection line of the two planes. By the assumption, the plane we seek is parallel to the line of intersection of the planes. Hence, $\overrightarrow{n_1} \times \overrightarrow{n_2}$ is parallel to the plane. The cross product $\vec{v} \times (\overrightarrow{n_1} \times \overrightarrow{n_2})$ is the normal vector for the plane.

Solution

Let $\overrightarrow{n_1} = (1,\ -1,\ -1)$ and $\overrightarrow{n_2} = (1,\ -2,\ 1)$. By computation we have

$$\overrightarrow{n_1} \times \overrightarrow{n_2} = \left(\begin{vmatrix} -1 & -1 \\ -2 & 1 \end{vmatrix},\ -\begin{vmatrix} 1 & -1 \\ 1 & 1 \end{vmatrix},\ \begin{vmatrix} 1 & -1 \\ 1 & -2 \end{vmatrix} \right) = (-3,\ -2,\ -1).$$

Let $\vec{v} = (1,\ -1,\ 2)$. Then

$$\begin{aligned} \vec{v} \times (\overrightarrow{n_1} \times \overrightarrow{n_2}) &= \left(\begin{vmatrix} -1 & 2 \\ -2 & -1 \end{vmatrix},\ -\begin{vmatrix} 1 & 2 \\ -3 & -1 \end{vmatrix},\ \begin{vmatrix} 1 & -1 \\ -3 & -2 \end{vmatrix} \right) \\ &= (5,\ -5,\ -5) = 5(1,\ -1,\ -1). \end{aligned}$$

Let $P_0(x_0,\ y_0,\ z_0) = (1,\ -1,\ -3)$ and $\vec{n} = (1,\ -1,\ -1)$. By **(6.3.1)** and **(6.3.2)** , the equation of the plane is

$$x - y - z = (1,\ -1,\ -1) \cdot (1,\ -1,\ -3) = 5.$$

Example 6.3.8.

Find an equation for the plane through $P(-1, -1, -1)$ that contains the line of intersection of the planes $x-y-z=1$ and $x-2y+z=3$.

Analysis

From the line of intersection of the planes, we can obtain a point $P_0(x_0, y_0, z_0)$ and a vector $\vec{v}$. Because the line of intersection of the planes is contained in the plane we seek, so $P_0(x_0, y_0, z_0)$ is in the plane and $\vec{v}$ is parallel to the plane. To find the normal vector, we need another vector that is is parallel to the plane. Because $P_0(x_0, y_0, z_0)$ and $P(-1, -1, -1)$ are in the plane, so $\overrightarrow{P_0P}$ is parallel to the plane. The normal vector $\vec{n} = \vec{v} \times \overrightarrow{P_0P}$. Because we need a point from the line of intersection of the planes, we must solve the system of the two planes to get it.

Solution

We solve the following system

$$\begin{cases} x-y-z=1 \\ x-2y+z=3. \end{cases}$$

$$(A\,|\,\vec{b}) = \begin{pmatrix} 1 & -1 & -1 & 1 \\ 1 & -2 & 1 & 3 \end{pmatrix} R_1(-1)+R_2 \rightarrow \begin{pmatrix} 1 & -1 & -1 & 1 \\ 0 & -1 & 2 & 2 \end{pmatrix} R_2(-1) \rightarrow$$

$$\begin{pmatrix} 1 & -1 & -1 & 1 \\ 0 & 1 & -2 & -2 \end{pmatrix} R_2(1)+R_1 \rightarrow \begin{pmatrix} 1 & 0 & -3 & -1 \\ 0 & 1 & -2 & -2 \end{pmatrix}.$$

The system corresponding to the last augmented matrix is

$$\begin{cases} x - 3z & = -1 \\ y - 2z & = -2, \end{cases}$$

where *x* and *y* are basic variables and *z* is a free variable. Let $z = t$. Then

$$x = -1 + 3z = -1 + 3t \quad \text{and } y = -2 + 2z = -2 + 2t.$$

From the above equation of the line, we obtain $P_0(x_0,\ y_0,\ z_0) = (-1,\ -2,\ 0),\ \vec{v} = (3,\ 2,\ 1)$.

$$\overrightarrow{P_0P} = (-1,\ -1,\ -1) - (-1,\ -2,\ 0) = (0,\ 1,\ -1),$$

and

$$\begin{aligned} \vec{v} \times \overrightarrow{P_0P} &= \left(\begin{vmatrix} 2 & 1 \\ 1 & -1 \end{vmatrix},\ -\begin{vmatrix} 3 & 1 \\ 0 & -1 \end{vmatrix},\ \begin{vmatrix} 3 & 2 \\ 0 & 4 \end{vmatrix} \right) \\ &= (-3,\ 3,\ 3) = -3(1,\ -1,\ -1). \end{aligned}$$

Let $\vec{n} = (1,\ -1,\ -1)$. Then by **(6.3.1)** and **(6.3.2)**, the equation of the plane is

$$x - y - z = (1,\ -1,\ -1) \cdot (-1,\ -2,\ 0) = 1.$$

Example 6.3.9.

Find the equation of the plane passing through the point $P(-1,\ 1,\ 1)$ and perpendicular to the planes $2x - y + z = 1$ and $x + y - z = 2$.

Solution

Let $\overrightarrow{n_1} = (2,\ -1,\ 1)$ and $\overrightarrow{n_2} = (1,\ 1,\ -1)$ be the normal vectors of the two given planes and $\vec{n}$ be the normal vector of the plane we must find. By **Definition 6.3.4**, $\vec{n}$ is perpendicular to both $\overrightarrow{n_1}$ and $\overrightarrow{n_2}$ and thus, it is parallel to $\overrightarrow{n_1} \times \overrightarrow{n_2}$. By computation, we have

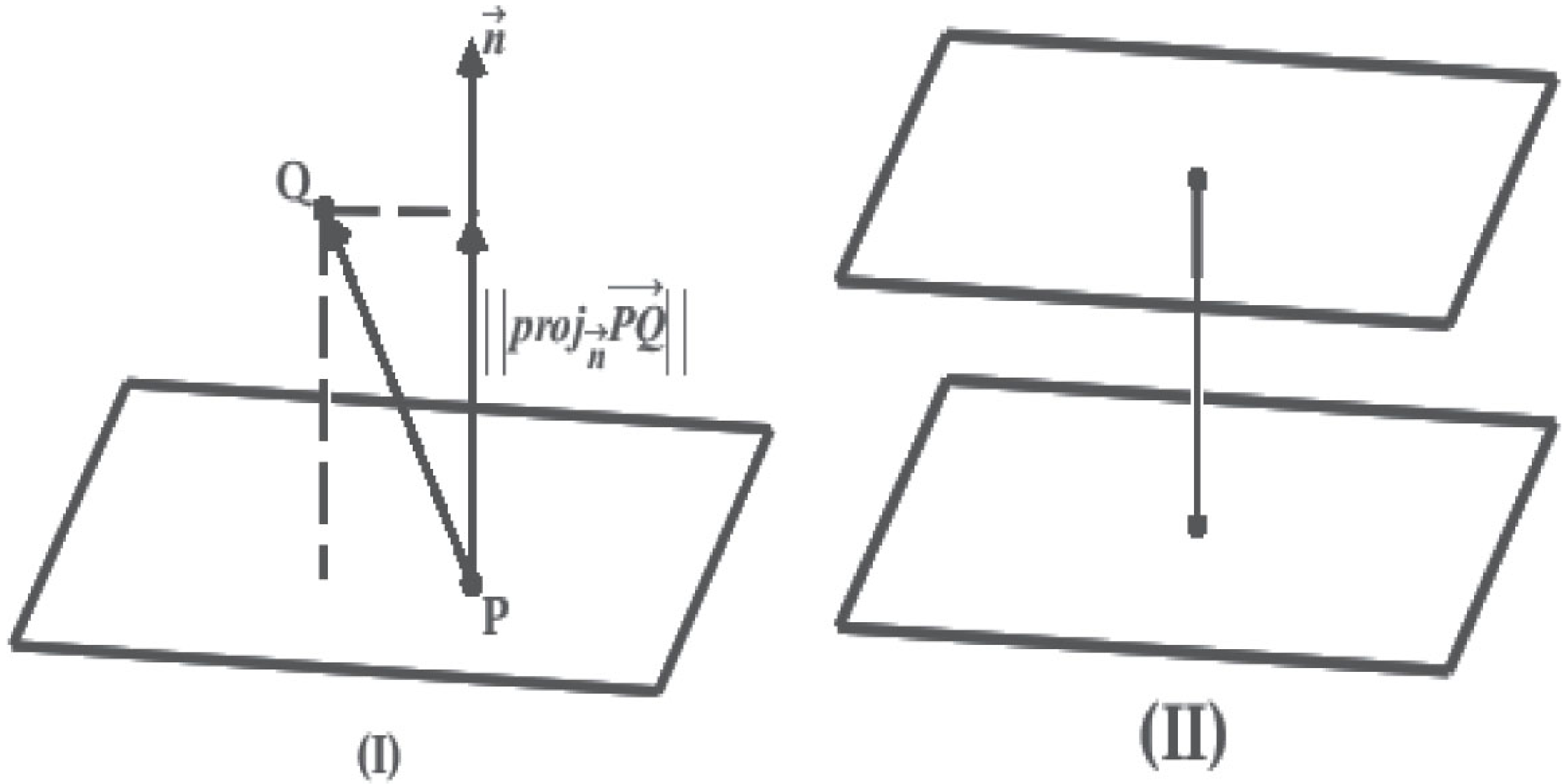
$$\begin{aligned} \overrightarrow{n_1} \times \overrightarrow{n_2} &= \left(\begin{vmatrix} -1 & 1 \\ 1 & -1 \end{vmatrix}, \ - \begin{vmatrix} 2 & 1 \\ 1 & -1 \end{vmatrix}, \ \begin{vmatrix} 2 & -1 \\ 1 & 1 \end{vmatrix} \right) \\ &= (0,\ 3,\ 3) = 3(0,\ 1,\ 1). \end{aligned}$$

Let $\vec{n} = (0,\ 1,\ 1)$. Then by **(6.3.1)** and **(6.3.2)**, the equation of the plane is

$$y + z = (0,\ 1,\ 1) \cdot (-1,\ 1,\ 1) = (0)(-1) + (1)(1) + (1)(1) = 2.$$

We end this section by deriving the formula for the (perpendicular) distance from a point to a plane (see **Figure 6.5** (I)).

Figure 6.5: (I), (II)



Definition 6.3.5.

The distance from a point $Q(x_1, y_1, z_1)$ to a plane

$$ax + by + cz = d$$

is the distance between the point Q and the intersection point of the straight line passing through Q and perpendicular to the plane.

Theorem 6.3.6.

The distance D from a point $Q(x_1, y_1, z_1)$ *to the plane* $ax + by + cz = d$ *is*

(6.3.6)

$$D = \frac{\left|ax_1 + by_1 + cz_1 - d\right|}{\sqrt{a^2 + b^2 + c^2}}.$$

Proof

Let $\vec{n} = (a,\ b,\ c)$ be the normal vector of the plane. Let $P(x_0,\ y_0,\ z_0)$ be a point in the plane. Because $\vec{n}$ is parallel to the line passing through Q and perpendicular to the plane, the projection of $\overrightarrow{PQ}$ to $\vec{n}$ equals the projection of $\overrightarrow{PQ}$ to the line. The norms of the two projections are the same. Hence, we have

(6.3.7)

$$D = \|\mathrm{Proj}_{\vec{n}}\overrightarrow{PQ}\| = \frac{\left|\vec{n} \cdot \overrightarrow{PQ}\right|}{\|\vec{n}\|}$$

Noting that $ax_0 + by_0 + cz_0 = d,$ by **(6.3.7)** , we obtain

$$\begin{aligned} D &= \frac{\left|\vec{n} \cdot \overrightarrow{PQ}\right|}{\|\vec{n}\|} = \frac{\left|a(x_1 - x_0) + b(y_1 - y_0) + c(z_1 - z_0)\right|}{\sqrt{a^2 + b^2 + c^2}} \\ &= \frac{\left|ax_1 + by_1 + cz_1 - (ax_0 + by_0 + cz_0)\right|}{\sqrt{a^2 + b^2 + c^2}} = \frac{\left|ax_1 + by_1 + cz_1 - d\right|}{\sqrt{a^2 + b^2 + c^2}} \end{aligned}$$

and **(6.3.6)** holds.

Example 6.3.10.

Find the distance from the point $Q(0, 0, 0)$ to the plane $x + y + z = 1$.

Solution

By **(6.3.6)**, we have

$$D = \frac{\left|ax_0 + by_0 + cz_0 - d\right|}{\sqrt{a^2 + b^2 + c^2}} = \frac{|(1)(0) + (1)(0) + (1)(0) - 1|}{\sqrt{1^2 + 1^2 + 1^2}} = \frac{|-1|}{\sqrt{3}} = \frac{\sqrt{3}}{3}.$$

The distance between two parallel planes is the distance from any point Q in one plane to the other (see **Figure 6.5** (II)). Hence, to find the distance between two parallel planes, you need to find a point in one plane. Usually, one takes $y = z = 0$ and solves the equation of the plane for x. If the solution is denoted by x_0, then the point $Q(x_0, 0, 0)$ is in the plane.

Example 6.3.11.

Find the distance between the following parallel planes

$$\begin{aligned} x + y - z &= 1 \ (1) \\ 2x + 2y - 2z &= 3. \ (2) \end{aligned}$$

Solution

In equation (1), we let $y = z = 0$. Then $x = 1$ and $Q(1, 0, 0)$ is a point in plane (1). The distance from $Q(1, 0, 0)$ to plane (2) equals the distance between the two planes. By **(6.3.6)**, we have

$$D = \frac{\left|ax_0 + by_0 + cz_0 - d\right|}{\sqrt{a^2 + b^2 + c^2}} = \frac{|(2)(1) + (2(0) + (-2)(0) - 3|}{\sqrt{2^2 + 2^2 + (-2)^2}} = \frac{\sqrt{3}}{6}.$$

Exercises

1. Find a normal vector for each of the following planes.
   a. $x + 2y - z = 1,$
   b. $4x - 2y - 3z = 5,$
   c. $-x - 6y + z - 1 = 0,$
   d. $2x + 3z - 4 = 0.$

2. Determine whether the following planes are parallel or orthogonal.
   a. $x - y + z = 2, \quad 2x - 2y + 2z = 5,$
   b. $2x - y + z = 1, \quad x + y - z = 6,$
   c. $2x - y + 3z = 2, \quad x + 2y + z = 4,$
   d. $4x - 2y + 6z = 3, \quad -2x + y - 3z = 5$

3. For each pair of the following planes, find a parametric equation of the line of intersection of the two planes.
   a. $x - y + z = 2, \quad 2x - 3y + z = -1;$
   b. $3x - y + 3z = -1, \quad -4x + 2y - 4z = 2,$
   c. $x - 3y - 2z = 4, \quad 3x + 2y - 4z = 6.$

4. Find an equation for the line through $P(2, \ -3, \ 0)$ that is parallel to the planes $2x + 2y + z = 2$ and $x - 3y = 5.$
5. Find an equation for the line through $P(2, \ -3, \ 0)$ that is perpendicular to the two lines
   $-1 + y, \ y = 2 - 2t, \ z = 1 - t$ and $x = 1 - t, \ y = 1 - t, \ z = 2 + t.$

6. Find an equation of the plane passing through the point P and perpendicular to the vector $\vec{n}$.
   a. $P(1,\ -1,\ 0),\ \vec{n} = (1,\ 3,\ 5)$
   b. $P(1,\ -1,\ 3),\ \vec{n} = (-1,\ 0,\ -1)$
   c. $P(0,\ -1,\ 2),\ \vec{n} = (-1,\ 1,\ -1)$
   d. $P(1,\ 0,\ 0),\ \vec{n} = (1,\ 1,\ -1)$

7. Find an equation for the plane through $P(2,\ 7,\ -1)$ that is parallel to the plane $4x - y + 3z = 3$.
8. Find an equation of the plane passing through the following given three points:
   a. $P_1(1,\ 2,\ -1),\quad P_2(2,\ 3,\ 1),\quad P_3(3,\ -1,\ 2)$
   b. $P_1(1,\ 0,\ 1)\quad P_2(0,\ 1,\ -1)\quad P_3(1,\ 1,\ -2)$

9. Find an equation for the plane containing the line $x = 3 + 6t,\ y = 4,\ z = t$ and that is parallel to the line of intersection of the planes $2x + y + z = 1$ and $x - 2y + 3z = 2$.
10. Find an equation for the plane through $P(1, 4, 4)$ that contains the line of intersection of the planes $x - y + 3z = 5$ and $2x + 2y + 7z = 0$.
11. Find an equation of the plane passing through the point $Q(1,\ 2,\ -1)$ and perpendicular to the planes $x + y + z = 2$ and $-x + 2y + 3z = 5$.
12. Find an equation for the plane through $P(1, 1, 3)$ that is perpendicular to the line

$$x = 2 - 3t, \quad y = 1 + t, \quad z = 2t.$$

13. Find an equation for the plane that contains the line $x = 3 + t,\ y = 5,\ z = 5 + 2t,$ and is perpendicular to the plane $x + y + z = 4$.
14. Find the distance from the point to the plane.
   a. $P(1,\ -2,\ -1),\ x - y + 2z = -1;$
   b. $P(0,\ 1,\ 2),\ 2x + y + 3z = 4;$
   c. $P(1,\ 0,\ 1),\ 2x - 5y + z = 3;$
   d. $P(-1,\ -1,\ 1),\ -x + 2y + 3z = 1.$

15. Find the distance between the given parallel planes.

a. $x - y + 2z = -3, \quad 3x - 3y + 6z = 1$

b. $x + y - z = 1, \quad 2x + 2y - 2z = -5$

Chapter 7 Bases and dimensions

7.1 Linear independence

Let

$$S = \left\{\overrightarrow{a_1},\ \cdots,\ \overrightarrow{a_n}\right\} \text{ and } A = (\overrightarrow{a_1},\ \cdots,\ \overrightarrow{a_n})$$

be the same as in **(1.4.2)** and **(4.1.5)** , respectively. Let $\mathbb{R}^n \to \mathbb{R}^m$ be the linear transformation with standard matrix *A* given by

(7.1.1)

$$T(\vec{X}) = A\vec{X},$$

where $\vec{X} = (x_1,\ x_2,\ \cdots,\ x_n)^T$. By **(2.2.3)** , we have

(7.1.2)

$$\vec{A}X = x_1\overrightarrow{a_1} + x_2\overrightarrow{a_2} + \cdots + x_n\overrightarrow{a_n}.$$

By **Theorem 4.5.2**, we see that a homogeneous system $A\vec{X} = \vec{0}$ has either only zero solution or infinitely many solutions. This, together with **(7.1.2)**, implies that the homogeneous system

(7.1.3)

$$x_1\overrightarrow{a_1} + x_2\overrightarrow{a_2} + \cdots + x_n\overrightarrow{a_n} = \vec{0}$$

has either only zero solution or infinitely many solutions. Hence, either of the following assertions holds.

1. If **(7.1.3)** holds, then $x_1 = x_2 = \cdots = x_n = 0$.
2. There exists $x_1, x_2, \cdots, x_n \in \mathbb{R}$ not all zero such that **(7.1.3)** holds.

Hence, we introduce the following definition on linear dependence and independence.

Definition 7.1.1.

If there exist $x_1, x_2, \cdots, x_n \in \mathbb{R}$ not all zero such that

(7.1.4)

$$x_1\overrightarrow{a_1} + x_2\overrightarrow{a_2} + \cdots + x_n\overrightarrow{a_n} = \vec{0},$$

then *S* is said to be a linearly dependent set or the vectors $\overrightarrow{a_1}, \overrightarrow{a_2}, \cdots, \overrightarrow{a_n}$ are said to be linearly dependent. If *S* is not linearly dependent, then *S* is said to be linearly independent.

By **Theorems 4.5.2** and **5.2.4**, we obtain the following criteria on linear independence.

Theorem 7.1.1.

The following assertions are equivalent.

1. *S is linearly independent.*
2. *The homogeneous system $A\vec{X} = \vec{0}$ has only zero solution.*
3. *The linear transformation T given in* **(7.1.1)** *is one to one.*
4. $r(A) = n.$

By **Theorem 7.1.1**, we see that *S* is linearly dependent if and only if the homogeneous system $A\vec{X} = \vec{0}$ has infinitely many solutions if and only if $r(A) < n.$ Moreover, if $n > m,$ then *S* is linearly dependent because $r(A) \leq \min\{m,\ n\} = m < n.$ If *S* contains a zero vector, then *S* is linearly dependent because *A* contains a zero column and $r(A) < n.$ In particular, $\{\vec{a}\}$ is linearly dependent if $\vec{a}$ is a zero vector in $\mathbb{R}^m,$ and linearly independent if $\vec{a}$ is a nonzero vector in $\mathbb{R}^m.$

Example 7.1.1.

1. Determine whether $S = \left\{\overrightarrow{a_1},\ \overrightarrow{a_2},\ \overrightarrow{a_3}\right\}$ is linearly independent, where

$$\overrightarrow{a_1} = (1,\ 1,\ -1),\quad \overrightarrow{a_2} = (2,\ -1,\ -11),\quad \overrightarrow{a_3} = (-1,\ 2,\ 10).$$

2. Determine whether $S = \left\{\overrightarrow{a_1},\ \overrightarrow{a_2},\ \overrightarrow{a_3},\ \overrightarrow{a_4}\right\}$ is linearly independent, where

$$\overrightarrow{a_1} = (1,\ 0,\ -1),\quad \overrightarrow{a_2} = (2,\ -1,\ -10),\quad \overrightarrow{a_3} = (-1,\ 2,\ 8),\quad \overrightarrow{a_4} = (0,\ 2,\ -5).$$

Solution

1. Because

$$A=\begin{pmatrix}1 & 2 & -1\\ 1 & -1 & 2\\ -1 & -11 & 10\end{pmatrix}\overset{R_1(-1)+R_2}{\underset{R_1(1)+R_3}{\longrightarrow}}\begin{pmatrix}1 & 2 & -1\\ 0 & -3 & 3\\ 0 & -9 & 9\end{pmatrix}$$

$$\overset{R_2(-3)+R_3}{\longrightarrow}\begin{pmatrix}1 & 2 & -1\\ 0 & -3 & 3\\ 0 & 0 & 0\end{pmatrix},$$

$r(A)=2<3=n.$ By **Theorem 7.1.1**, S is linearly dependent.

2. Let $A=(\vec{a}_1\vec{a}_2\vec{a}_3\vec{a}_4).$ By **Theorem 2.5.2**, we have

$$r(A)\le\min\{m,\ n\}=m=3<4=n.$$

By **Theorem 7.1.1**, S is linearly dependent.

The following result gives the relation between linear dependence and linear combination.

Theorem 7.1.2.

S is linearly dependent if and only if there exists a vector $\vec{a}_i$ in S such that $\vec{a}_i$ is a linear combination of the rest of vectors in S.

Proof

By **Definition 7.1.1**, we see that if S is linearly dependent, then there exist $x_1,\ \cdots,\ x_n,$ not all zero, such that **(7.1.4)** holds. If we assume that *xi* is the nonzero number, then **(7.1.4)** implies that

$$x_i\vec{a}_i=-x_1\vec{a}_1-\cdots-x_{i-1}\vec{a}_{i-1}-x_{i+1}\vec{a}_{i+1}-\cdots-x_n\vec{a}_n$$

and

$$\vec{a}_i = -\frac{x_1}{x_i}\vec{a}_1 - \cdots - \frac{x_{i-1}}{x_i}\vec{a}_{i-1} - \frac{x_{i+1}}{x_i}\vec{a}_{i+1} - \cdots - \frac{x_n}{x_i}\vec{a}_n.$$

Hence, $\vec{a}_i$ is a linear combination of $\vec{a}_1, \cdots, \vec{a}_{i-1}, \vec{a}_{i+1}, \cdots, \vec{a}_n.$.

Conversely, if $\vec{a}_i$ is a linear combination of $\vec{a}_1, \cdots, \vec{a}_{i-1}, \vec{a}_{i+1}, \cdots, \vec{a}_n,$ then there exist $y_1, \cdots, y_{i-1}, y_{i+1}, \cdots, y_n$ such that

$$\vec{a}_i = y_1\vec{a}_1 + \cdots + y_{i-1}\vec{a}_{i-1} + y_i + 1\vec{a}_{i+1} + \cdots + y_n\vec{a}_n.$$

If $n > m,$ then by **Theorem 7.1.1** , *S* is linearly dependent. If $n \leq m,$ then for each $j \neq i,$ we can use row operations $R_j(-y_j) + R_i$ on the transpose A^T of *A*. Hence, the *i*th row of the resulting matrix denoted by *B* is a zero row. Moreover, A^T and *B* are row equivalent. By **Theorem 2.5.5** (*i*) and **Theorem 2.5.4** , $r(A) = r(A^T) = r(B).$ Because *B* contains a zero row, $r(B) < n.$ It follows that $r(A) < n.$ By **Theorem 7.1.1** , *S* is linearly dependent.

By **Theorem 7.1.2** , we obtain the following result, which shows that two nonzero vectors are linearly dependent if and only if they are parallel.

Corollary 7.1.1.

Let $S = \left\{\vec{a}_1, \vec{a}_2\right\}$ *be a set of two nonzero vectors in* $\mathbb{R}^m$. *Then S is linearly dependent if and only if* $\vec{a}_1$ *and* $\vec{a}_2$ *are parallel.*

Proof

By **Theorem 7.1.2** , S is linearly dependent if and only if there exists $k \in \mathbb{R}$ such that $\vec{a}_1 = k\vec{a}_2$ or $\vec{a}_2 = k\vec{a}_1$. By **Definition 1.1.6** , the latter holds if and only if $\vec{a}_1$ and $\vec{a}_2$ are parallel.

The following result shows that if S is linearly dependent, then the set obtained by adding any finitely many vectors into S is linearly dependent.

Corollary 7.1.2.

Let $T = \left\{\vec{b}_1, \cdots, \vec{b}_k\right\} \subset \mathbb{R}^m$. *Then if S is linearly dependent, then*

$$S \cup T := \left\{\vec{a}_1, \cdots, \vec{a}_n, \vec{b}_1, \cdots, \vec{b}_k\right\}$$

is linearly dependent.

Proof

Because S is linearly dependent, one of vectors in S, say $\vec{a}_i$, is a linear combination of the rest of the vectors in S. By **Theorem 1.4.2** , $\vec{a}_i$ is a linear combination of the rest of the vectors in S and all vectors in T. It follows that one of the vectors in $S \cup T$ is a linear combination of the rest of the vectors in $S \cup T$. The result follows from **Theorem 7.1.2** .

Example 7.1.2.

Determine whether $\left\{\vec{a_1},\ \vec{a_2},\ \vec{a_3}\right\}$ is linearly dependent, where

$$\vec{a_1} = (2,\ 4,\ 6),\quad \vec{a_2} = (1,\ 0,\ 1),\quad \vec{a_3} = (1,\ 2,\ 3).$$

Solution

Because $\vec{a_1} = 2\vec{a_3}$, $\left\{\vec{a_1},\ \vec{a_3}\right\}$ is linearly dependent. By **Corollary 7.1.2**, $\left\{\vec{a_1},\ \vec{a_2},\ \vec{a_3}\right\}$ is linearly dependent.

By **Theorems 7.1.2** and **5.2.2**, we obtain

Theorem 7.1.3.

Let $A = (\vec{a_1}, \cdots, \vec{a_n})$ be the same as in **(4.1.5)**. *Let $\vec{b} \in \mathbb{R}^m$. If $\vec{b}$ is a linear combination of $\vec{a_1},\ \vec{a_2},\ \cdots,\ \vec{a_n}$, then*

$$\left\{\vec{a_1},\ \vec{a_2},\ \cdots,\ \vec{a_n},\ \vec{b}\right\}$$

is linear dependent and $r(A) = r(A \mid \vec{b})$.

The following result shows that any nonempty subset of linearly independent vectors is linearly independent.

Theorem 7.1.4.

Let S_1 be a nonempty subset of S. If S is linearly independent, so is S_1.

Proof

The proof is by contradiction. Assume that S_1 is not linearly independent, that is, S_1 is linearly dependent. By **Theorem 7.1.2**, one of the vectors in S_1 is a linear combination of the rest of the vectors in S_1. By **Corollary 7.1.2**, *S* is linearly dependent, which contradicts the assumption that *S* is linearly independent. The contradiction shows that the assumption that S_1 is linearly dependent is false. Hence, S_1 is linearly independent.

Example 7.1.3.

Let $\overrightarrow{e_1} = (1, 0, 0)$ and $\overrightarrow{e_3} = (0, 0, 1)$. Then $\left\{\overrightarrow{e_1}, \overrightarrow{e_3}\right\}$ is linearly independent.

Solution

Let $I_3 = (\overrightarrow{e_1}\overrightarrow{e_2}\overrightarrow{e_3})$, where $\overrightarrow{e_2} = (0, 1, 0)$. Then $r(I_3) = 3$. It follows from **Theorem 7.1.1** that $S = \left\{\overrightarrow{e_1}, \overrightarrow{e_2}, \overrightarrow{e_3}\right\}$ is linearly independent. By **Theorem 7.1.4**, $S_1 = \left\{\overrightarrow{e_1}, \overrightarrow{e_3}\right\}$ is linearly independent.

If $n = m$, then by **Theorems 4.5.3** and **7.1.1**, we obtain the following result, which uses determinants to determine linear independence.

Theorem 7.1.5.

Let A be an $n \times n$ matrix and let S be the set of column vectors of A. Then the following assertions hold.

1. *S is linearly dependent if and only if $|A| = 0$ if and only if $r(A) < n$.*
2. *S is linearly independent if and only if $|A| \neq 0$ if and only if $r(A) = n$.*

Example 7.1.4.

Let $\vec{e}_1, \vec{e}_2, \vec{e}_2, \cdots, \vec{e}_n$ be the standard vectors in $\mathbb{R}^n$ given in **(1.1.1)** . Show that $S = \left\{\vec{e}_1, \vec{e}_2, \cdots, \vec{e}_n\right\}$ is linearly independent.

Solution

Because $\left|A\right| = \left|I_n\right| = 2 \neq 0,$ it follows from **Theorem 7.1.5** (2) that *S* is linearly independent.

Example 7.1.5.

Determine whether *S* given in **Example 7.1.1** is linearly independent by using **Theorem 7.1.5** .

Solution

$$|A| = \begin{vmatrix} 1 & 2 & -1 \\ 1 & -1 & 2 \\ -1 & -11 & 10 \end{vmatrix} \overset{R_1(1)+R_2}{\underset{R_1(-1)+R_2}{=}} \begin{vmatrix} 1 & 2 & -1 \\ 0 & -3 & 3 \\ 0 & -9 & 9 \end{vmatrix} = 0,$$

it follows from **Theorem 7.1.5** (1) that *S* is linearly dependent.

Exercises

1. Let $\vec{a}_1 = (2, 4, 6)^T, \vec{a}_2 = (0, 0, 0)^T,$ and $\vec{a}_3 = (1, 2, 3)^T.$ Determine whether $\left\{\vec{a}_1, \vec{a}_2, \vec{a}_3\right\}$ is linearly dependent.

2. Determine whether $\left\{ \overrightarrow{a_1},\ \overrightarrow{a_2},\ \overrightarrow{a_3},\ \overrightarrow{a_4} \right\}$ is linearly dependent, where

$$\overrightarrow{a_1} = \begin{pmatrix} 1 \\ 2 \\ 1 \\ 0 \end{pmatrix},\quad \overrightarrow{a_2} = \begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \end{pmatrix},\quad \overrightarrow{a_3} = \begin{pmatrix} 0 \\ 2 \\ 0 \\ 0 \end{pmatrix},\quad \overrightarrow{a_4} = \begin{pmatrix} 0 \\ 2 \\ 0 \\ 1 \end{pmatrix}.$$

3. Determine whether $S = \left\{ \overrightarrow{a_1},\ \overrightarrow{a_2},\ \overrightarrow{a_3},\ \overrightarrow{a_4} \right\}$ is linearly independent, where

$$\overrightarrow{a_1} = \begin{pmatrix} 1 \\ 3 \\ -4 \end{pmatrix},\quad \overrightarrow{a_2} = \begin{pmatrix} 7 \\ 9 \\ 8 \end{pmatrix},\quad \overrightarrow{a_3} = \begin{pmatrix} -1 \\ -2 \\ 1 \end{pmatrix},\quad \overrightarrow{a_4} = \begin{pmatrix} 6 \\ -2 \\ 5 \end{pmatrix}.$$

4. Determine that $S = \left\{ \overrightarrow{a_1},\ \overrightarrow{a_2},\ \overrightarrow{a_3} \right\}$ is linearly independent, where

$$\overrightarrow{a_1} = \begin{pmatrix} -1 \\ 0 \\ -1 \end{pmatrix},\quad \overrightarrow{a_2} = \begin{pmatrix} 3 \\ -1 \\ -6 \end{pmatrix},\quad \overrightarrow{a_3} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}.$$

7.2 Bases of spanning and solution spaces

Let

(7.2.1)

$$S = \left\{\overrightarrow{a}_1, \overrightarrow{a}_2, \cdots, \overrightarrow{a}_n\right\}$$

be a set of vectors in $\mathbb{R}^m$. In **Section 1.4** , we introduced the following spanning space (see **Definition 1.4.2**)

(7.2.2)

$$V: = \text{span } S = \left\{x_1\overrightarrow{a}_1 + x_2\overrightarrow{a}_2 + \cdots + x_n\overrightarrow{a}_n : x_i \in \mathbb{R},\ i \in I_n\right\}.$$

In this spanning space, the vectors $\overrightarrow{a}_1, \overrightarrow{a}_2, \cdots, \overrightarrow{a}_n$ are not required to be linearly independent. In some cases, we are interested in the case when these vectors *are* linearly independent. Hence, we introduce the notion of a basis.

Definition 7.2.1.

If *S* is linearly independent, then *S* is called a basis of *V*; *n* is called the dimension of *V*, denoted by $\dim V = n$; and *V* is said to be an *n* dimensional subspace of $\mathbb{R}^m$.

Example 7.2.1.

Let

(7.2.3)

$$E: = \{\vec{e}_1, \vec{e}_2, \cdots, \vec{e}_n\}$$

be the set of standard vectors in $\mathbb{R}^n$ given in **(1.1.1)** or **Example 7.1.4**. Then the following assertions hold.

1. $\mathbb{R}^n = \text{span } S.$
2. *S* is linearly independent.
3. $\dim(\mathbb{R}^n) = n.$

Solution

(1) follows from **(1.4.9)** and (2) follows from **Example 7.1.4**. By **Definition 7.2.1**, the set *E* of the standard vectors constitutes a basis of $\mathbb{R}^n$, $\dim(\mathbb{R}^n) = n$, and (3) holds.

The basis *E* in **(7.2.3)** is called the standard basis of $\mathbb{R}^n$. Because the dimension of $\mathbb{R}^n$ is *n*, we call $\mathbb{R}^n$ an *n*-dimensional (Euclidean) space.

Using vectors in *S*, we define a matrix as follows.

(7.2.4)

$$A = (\vec{a}_1 \vec{a}_2 \cdots \vec{a}_n).$$

By **Definition 7.2.1** and **Theorems 7.1.1** and **7.1.5**, we obtain the following result.

Theorem 7.2.1.

Let S and A be the same as in **(7.2.1)** *and* **(7.2.4)**. *Then the following assertions hold.*

i. *S is a basis of V if and only if* $r(A) = n$.
ii. *If* $n = m$, *then S is a basis of V if and only if* $|A| \neq 0$ *if and only if* $r(A) = n$ *n if and only if A is invertible.*

Example 7.2.2.

1. Let $\vec{a}_1 = (1, 0, 1)^T$ and $\vec{a}_2 = (1, 1, 1)^T$. Show that $S = \left\{\vec{a}_1, \vec{a}_2\right\}$ is a basis of span *S* and $\dim(\text{span } S) = 2$.
2. Let $\vec{a}_1 = (1, 0)^T$, $\vec{a}_2 = (1, 1)^T$, and $\vec{a}_3 = (2, -3)^T$. Determine whether $S = \left\{\vec{a}_1, \vec{a}_2, \vec{a}_3\right\}$ is a basis of span *S*.

Solution

1. Let $A = (\vec{a}_1 \vec{a}_2)$. Then

$$A^T = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 1 \end{pmatrix} \xrightarrow{R_1(-1)+R_2} \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$

and $r(A^T) = 2$. By **Theorem 2.5.5**, $r(A) = r(A^T) = 2$ and by **Theorem 7.2.1**, *S* is a basis of span *S*. By **Definition 7.2.1**, $\dim(\text{span } S) = 2$.

2. Note that $n = 3$ and $m = 2$. Because $n > m$, it follows from **Theorem 7.2.1** that *S* is not a basis of span *S*.

Theorem 7.2.2.

Assume that S is a basis of span *S. Then for each* $\vec{b} \in \operatorname{span} S$, *there exists a unique vector* $(x_1, x_2, \cdots, x_n)$ *such that*

(7.2.5)

$$\vec{b} = x_1\vec{a_1} + x_2\vec{a_2} + \cdots x_n\vec{a_n}.$$

Proof

Let $A = (\vec{a_1}\vec{a_2}\cdots\vec{a_n})$. By **Theorem 7.2.1**, $r(A) = n \le m$. Because $\vec{b} \in \operatorname{span} S$, by **(7.2.2)**, the system **(7.2.5)** has a solution. By **Theorem 4.5.1** (3), $r(A) = r(A \mid \vec{b}) = n$. By **Theorem 4.5.1** (1), **(7.2.5)** has a unique solution.

Note that the bases of span *S* are not unique. For example, if $S = \left\{\vec{a_1}, \vec{a_2}, \cdots, \vec{a_n}\right\}$ is a basis of span *S* and $\beta_i \ne 0$ for $i = 1, 2, \cdots, n$, then

$$T: = \left\{\beta_1\vec{a_1}, \beta_2\vec{a_2}, \cdots, \beta_n\vec{a_n}\right\}$$

is a basis of span *S* because *T* is linearly independent and span $S = \operatorname{span} T$.

Now, we assume that *S* is a basis and $\left\{\overrightarrow{b_1}, \overrightarrow{b_2}, \cdots, \overrightarrow{b_k}\right\}$ in span *S* is another basis such that

(7.2.6)

$$\operatorname{span}\left\{\overrightarrow{a_1}, \overrightarrow{a_2}, \cdots, \overrightarrow{a_n}\right\} = \operatorname{span}\left\{\overrightarrow{b_1}, \overrightarrow{b_2}, \cdots, \overrightarrow{b_k}\right\}.$$

The question is that if **(7.2.6)** holds, is *k* equal to *n*? Equivalently, is the dimension of a spanning space unique?

In the following, we shall prove $k = n,$ so we give a positive answer to the above question.

Lemma 7.2.1.

Let S be the same as in **(7.2.1)** *and let* $T = \left\{\overrightarrow{b_1}, \overrightarrow{b_2}, \cdots, \overrightarrow{b_k}\right\}$ *be a set of vectors in* span *S. Let* $A_{m\times n} = (\overrightarrow{a_1}\overrightarrow{a_2}\cdots\overrightarrow{a_n})$ *and* $B_{m\times n} = (\overrightarrow{b_1}\overrightarrow{b_2}\cdots\overrightarrow{b_n}).$ *Then the following assertions hold.*

i. *There exists an* $n \times k$ *matrix* $C_{n\times k}$ *such that* $B_{m\times k} = A_{m\times n}C_{n\times k}.$
ii. *If* $k = n$ *n and S and T are linearly independent, then there exists a unique invertible* $n \times n$ *matrix* $C_{n\times n}$ *such that*
   (7.2.7)

$$B_{m\times n} = A_{m\times n}C_{n\times n}.$$

Proof

i. Because $\vec{b}_i \in \operatorname{span} S$ for $i \in I_k$, by **(2.2.3)**, there exists a vector $\vec{C}_i := (c_{1i}, c_{2i}, \cdots, c_{ni})^T$ such that
(7.2.8)

$$\vec{b}_i = c_{1i}\vec{a}_1 + c_{2i}\vec{a}_2 + \cdots + c_{ni}\vec{a}_n = A\vec{C}_i.$$

Let $C_{n\times k} = (\vec{C}_1\vec{C}_2\cdots\vec{C}_k)$. By **(2.2.6)** and **(7.2.8)**, we have

$$B_{m\times k} = (\vec{b}_1\vec{b}_2\cdots\vec{b}_k) = (A_{m\times k}\vec{C}_1 A_{m\times k}\vec{C}_2\cdots A_{m\times k}\vec{C}_k) = A_{m\times n}C_{n\times k}.$$

ii. Because *S* is linearly independent, by **Theorem 7.2.1**, there exists a unique vector $\vec{C}_i := (c_{1i}, c_{2i}, \cdots, c_{ni})^T$ such that **(7.2.8)** holds for $i \in I_n$ and **(7.2.7)** with $k = n$ holds. We will prove that $C_{n\times n}$ is invertible by contradiction. If not, then there exists a nonzero vector $\vec{X}_0 \in \mathbb{R}^n$ such that $C_{n\times n}\vec{X}_0 = \vec{0}$. Because $B_{m\times n} = A_{m\times n}C_{n\times n}$, it follows that

$$B_{m\times n}\vec{X}_0 = A_{m\times n}C_{n\times n}\vec{X}_0 = A\vec{0} = \vec{0}.$$

By **Theorem 7.1.1**, *T* is linearly dependent, which contradicts the hypothesis that *T* is linearly independent. Hence, $C_{n\times n}$ is invertible.

Lemma 7.2.2.

Let S and T be the same as in ***Lemma 7.2.1****. Assume that S is linearly independent. Then the following assertions hold.*

i. *If $k > n$, then T is linearly dependent.*
ii. *If T is linearly independent, then $k \leq n$.*

Proof

Note that (*i*) and (*ii*) are equivalent, so we only prove (*i*). Let

$$A_{m \times n} = (\vec{a}_1 \vec{a}_2 \cdots \vec{a}_n) \text{ and } B_{m \times k} = (\vec{b}_1 \vec{b}_2 \cdots \vec{b}_k).$$

By **Lemma 7.2.1** (*i*), there exists an $n \times k$ matrix $C_{n \times k}$ such that $B_{m \times k} = A_{m \times n} C_{n \times k}$. Because $k > n$, it follows from **Theorem 4.5.2** (2) that the homogeneous system $C_{n \times k} \vec{X} = \vec{0}$ has infinitely many solutions. This implies that the homogeneous system

$$B_{m \times k} \vec{X} = A_{m \times n} C_{n \times k} \vec{X} = A_{m \times n} \vec{0} = \vec{0}$$

has infinitely many solutions. By **Theorem 7.1.1**, *T* is linearly dependent.

By **Lemma 7.2.2** (*ii*), we see that the number of linearly independent vectors in span *S* is less than or equal to the number of vectors in *S*.

The following result shows that if *T* is linearly independent and $\text{span } T = \text{span } S$, then $k = n$.

Theorem 7.2.3.

Let S and T be the same as in ***Lemma 7.2.1****. Assume that S and T are linearly independent. Then $k = n$ if and only if*

(7.2.9)

$$\text{span}\left\{\vec{a}_1,\ \vec{a}_2,\ \cdots,\ \vec{a}_n\right\} = \text{span}\left\{\vec{b}_1,\ \vec{b}_2,\ \cdots,\ \vec{b}_k\right\}.$$

Proof

Assume that **(7.2.9)** holds. We will prove $k = n$. By **(7.2.9)** , *T* is the set of vectors in span *S*. By **Lemma 7.2.2** (*ii*), $k \le n$. Similarly, by **(7.2.9)** , *S* is the setofvectorsinspan *T* and by **Lemma 7.2.2** (*ii*), $n \le k$. Hence, $k = n$. Now, we assume that $k = n$. In this case, $T: = \left\{\vec{b}_1,\ \vec{b}_2,\ \cdots,\ \vec{b}_n\right\}$ is a set of vectors in the spanning space span *S*. We prove that **(7.2.9)** holds. Because $\vec{b}_i \in \text{span } S$ for each $i \in \{1,\ 2,\ \cdots,\ n\}$, every vector $\vec{b} \in \text{span } T$ implies that $\vec{b} \in \text{span } S$. We will prove that for $\vec{b} \in \text{span } S$, $\vec{b} \in \text{span } T$. Let

$$A_{m \times n} = (\vec{a}_1\vec{a}_2\cdots\vec{a}_n) \text{ and } B_{m \times n} = (\vec{b}_1\vec{b}_2\cdots\vec{b}_n).$$

For each $\vec{b} \in \text{span } S$, by **(7.2.2)** , there exists a vector $\vec{Y}: = (y_1,\ y_2,\ \cdots, y_n) \in \mathbb{R}^n$ such that $A_{m \times n}\vec{Y} = \vec{b}$. By **Lemma 7.2.1** (*ii*), there exists a unique invertible $n \times n$ matrix $C_{n \times n}$ such that $A_{m \times n} = B_{m \times n}C_{n \times n}^{-1}$. Let $\vec{X} = C_{n \times n}^{-1}\vec{Y}$. Then

$$B\vec{X} = B_{m \times n}C_{n \times n}^{-1}\vec{Y} = A_{m \times n}\vec{Y} = \vec{b}.$$

This implies $\vec{b} \in \text{span } T$. Hence, $\text{span } T = \text{span } S$.

By **Theorem 7.2.3** , we see that the dimension of span *S* is independent of the choices of linearly independent vectors whose spanning space is span *S* and thus, the dimension of a spanning space is unique. Moreover, the spanning space of any *n* linearly independent vectors in span *S* is equal to span *S*.

Theorem 7.2.3 requires all vectors of *T* to be in span *S*. If this condition is not satisfied, then span *T* might not be equal to span *S*.

Example 7.2.3.

Let $S=\left\{\begin{pmatrix}1\\0\end{pmatrix}\right\}$ and $T=\left\{\begin{pmatrix}0\\1\end{pmatrix}\right\}$. Show that $\text{span } T\neq \text{span } S$.

Solution

Note that

$$\text{span } S=\left\{\begin{pmatrix}t\\0\end{pmatrix}: t\in\mathbb{R}\right\} \text{ and span } T=\left\{\begin{pmatrix}0\\s\end{pmatrix}: s\in\mathbb{R}\right\}.$$

It is obvious that $\begin{pmatrix}0\\1\end{pmatrix}\notin \text{span } S$ and thus, $\text{span } T\neq \text{span } S$.

However, in **Theorem 7.2.3** , if $m=n,$ then we do not need to verify the condition that all vectors of *T* are in span *S* because this condition is satisfied automatically.

Corollary 7.2.1.

Let $S=\left\{\vec{a}_1,\ \vec{a}_2,\ \cdots,\ \vec{a}_n\right\}$ *and* $T=\left\{\vec{b}_1,\ \vec{b}_2,\ \cdots,\ \vec{b}_n\right\}$ *be sets of linearly independent vectors in* $\mathbb{R}^n$. *Then* $\text{span } T=\text{span } S=\mathbb{R}^n$.

Proof

By **Theorem 7.2.3** (1), $\mathbb{R}^n = \operatorname{span}\left\{\vec{e_1}, \vec{e_2}, \cdots, \vec{e_n}\right\}$. Because all vectors of *S* are in $\mathbb{R}^n$, they are in span *E*. By **Theorem 7.2.3**, $\operatorname{span} S = \operatorname{span} E = \mathbb{R}^n$. Similarly, we can prove that span $T = \mathbb{R}^n$. Hence, span $T = \operatorname{span} S$.

By **Corollary 7.2.1**, we see that any set of *n* linearly independent vectors in $\mathbb{R}^n$ constitutes a basis of $\mathbb{R}^n$.

Theorem 7.2.4.

Let $S = \left\{\vec{a_1}, \vec{a_2}, \cdots, \vec{a_n}\right\}$ *be a basis of* span *S and assume that* $S \neq \mathbb{R}^m$. *Then* $S_1 = \left\{\vec{a_1}, \vec{a_2}, \cdots, \vec{a_n}, \vec{b}\right\}$ *is linearly independent for each* $\vec{b} \notin \operatorname{span} S$.

Proof

Let *A* be the same as in **(7.2.4)**. Then by **Theorems 7.2.1** and **2.5.2**,

(7.2.10)

$$n = r(A) \leq r(A \,|\, \vec{b}) \leq n + 1.$$

By **(1.4.9)**, $\mathbb{R}^m = \operatorname{span}\left\{\vec{e_1}, \vec{e_2}, \cdots, \vec{e_m}\right\}$. Because $\operatorname{span} S \neq \mathbb{R}^m$, by **Theorem 7.2.3**, $n < m$ and thus $n + 1 \leq m$. Let $\vec{b} \notin \operatorname{span} S$. Then the system

(7.2.11)

$$\vec{b} = x_1\vec{a_1} + x_2\vec{a_2} + \cdots x_n\vec{a_n}$$

has no solutions. By **Theorem 4.5.1** (3), $r(A) < r(A\,|\,\vec{b})$. By **(7.2.10)**, $r(A\,|\,\vec{b}) = n + 1$. By **Theorem 7.1.1** (*i*), S_1 is linearly independent.

By **Theorem 7.2.4**, if $n < m$, then we can extend a set of linearly independent vectors $S = \left\{\vec{a}_1,\ \vec{a}_2,\ \cdots,\ \vec{a}_n\right\}$ toabasisof $\mathbb{R}^m$. Let $A = (\vec{a}_1\vec{a}_2\cdots\vec{a}_n)$ be the same as in **(7.2.4)**. We use methods given in Sections 4.7 or 4.8 to find a vector $\vec{b}_1 \in \mathbb{R}^m$ such that the system $A\vec{X} = \vec{b}_1$ has no solutions (Section 4.7) or such that $\vec{b}_1 \in \operatorname{span} S$ (Section 4.8). By **Theorem 7.2.4**, $S_1 := \left\{S,\ \vec{b}_1\right\} = \left\{\vec{a}_1,\ \vec{a}_2,\ \cdots,\ \vec{a}_n,\ \vec{b}_1\right\}$ is linearly independent. If $n + 1 = m$, then S_1 isabasisof $\mathbb{R}^m$. If $n + 1 < m$, then repeating the process and using S_1, we obtain a vector $\vec{b}_2$ such that $\vec{b}_2 \notin \operatorname{span} S_1$ and $S_2 = \left\{S_1,\ \vec{b}_2\right\}$ is linearly independent. We repeat the process until we find S_{m-n}.

Example 7.2.4.

Let $\vec{a}_1 = (1,\ 0,\ 0,\ 1)^T$ and $\vec{a}_2 = (1,\ -1,\ 1,\ -1)^T$. Find a vector $\vec{b} \in \mathbb{R}^4$ such that $S = \left\{\vec{a}_1,\ \vec{a}_2,\ \vec{b}\right\}$ is a basis of span *S*.

Solution

Let $A = (\vec{a}_1\vec{a}_2)$ and $\vec{b} = (b_1,\ b_2,\ b_3,\ b_4)^T$. Then

$$(A\,|\,\vec{b}) \;=\; \begin{pmatrix} 1 & 1 & b_1 \\ 0 & -1 & b_2 \\ 0 & 1 & b_3 \\ 1 & -1 & b_4 \end{pmatrix} \xrightarrow{R_1(-1)+R_4} \begin{pmatrix} 1 & 1 & b_1 \\ 0 & -1 & b_2 \\ 0 & 1 & b_3 \\ 0 & -2 & b_4 - b_1 \end{pmatrix}$$

$$\xrightarrow[R_2(1)+R_3]{R_2(-2)+R_4} \begin{pmatrix} 1 & 1 & b_1 \\ 0 & -1 & b_2 \\ 0 & 0 & b_3 + b_2 \\ 0 & 0 & b_4 - b_1 - 2b_2 \end{pmatrix} = (B\,|\,\vec{c}).$$

If either $b_3 + b_2 \neq 0,$ or $b_4 - b_1 - 2b_2 \neq 0,$ then $r(A) = 2$ and $r(A\,|\,\vec{b}) \geq 3$. Hence, $r(A) < r(A\,|\,\vec{b}).$ It follows from **Theorem 4.5.1** (3) that if either $b_3 + b_2 \neq 0,$ or $b_4 - b_1 - 2b_2 \neq 0,$ then the system has no solutions. Hence we can choose $b_1 = b_4 = 0,\ b_2 = b_3 = 1$. Then $b_3 + b_2 \neq 0$. Let $\vec{b} = (0,\ 1,\ 1,\ 0)^T$. By **Theorem 7.2.4** , $S = \left\{ \vec{a}_1,\ \vec{a}_2,\ \vec{b} \right\}$ is a basis of span *S*.

By **Theorem 7.2.4** , we prove the following result.

Theorem 7.2.5.

Let S and A be the same as in **(7.2.1)** *and* **(7.2.4)** *. Let* $T = \left\{ \vec{b}_{n_1},\ \vec{b}_{n_2},\ \cdots,\ \vec{b}_{n_r} \right\}$ *be a set of r vectors in* span *S. Assume that T is linearly independent and* $r(A) = r.$ *Then the following assertions hold.*

i. $\text{span } T = \text{span } S$.

ii. *Each of the vectors in S but not in T is a linear combination of vectors in T.*

Proof

i. It is obvious that every vector in span *T* is in span *S*. We will prove that every vector in span *S* must belong to span *T*. In fact, if not, then there exists a vector $\vec{b} \in \text{span } S$ and $\vec{b} \in \text{span } T$. Because *T* is linearly independent, by **Theorem 7.2.4** , the following set

$$\left\{ \overrightarrow{b_{n_1}}, \overrightarrow{b_{n_2}}, \cdots, \overrightarrow{b_{n_r}}, \vec{b} \right\}$$

is linearly independent. Hence, we have $r(A) \geq r + 1$, which contradicts the condition $r(A) = r$.

ii. For every vector $\vec{b}$ in *S*, because $\text{span } T = \text{span } S$, $\vec{b} \in \text{span } T$. Hence, $\vec{b}$ is a linear combination of vectors in *T*. In particular, result (*ii*) holds.

The following result provides a method to choose basis vectors from $\text{span}\left\{ \overrightarrow{a_1}, \overrightarrow{a_2}, \cdots, \overrightarrow{a_n} \right\}$.

Theorem 7.2.6.

Let $\left\{ \overrightarrow{a_1}, \overrightarrow{a_2}, \cdots, \overrightarrow{a_n} \right\}$ *be a set of vectors in* $\mathbb{R}^m$ *and let A be the same as in* **(7.2.4)** *. Assume that*

$r(A) = r \geq 1$. *Then there exists a subset* $\left\{ \overrightarrow{a_{n_1}}, \overrightarrow{a_{n_2}}, \cdots, \overrightarrow{a_{n_r}} \right\}$ *of* $\left\{ \overrightarrow{a_1}, \overrightarrow{a_2}, \cdots, \overrightarrow{a_n} \right\}$.

$$\text{span}\left\{\vec{a}_1,\ \vec{a}_2,\ \cdots,\ \vec{a}_n\right\} = \text{span}\left\{\vec{a}_{n_1},\ \vec{a}_{n_2},\ \cdots,\ \vec{a}_{n_r}\right\}.$$

Proof

Let *S* be the same as in **(7.2.1)** . Because $r(A) = r \geq 1,$ there is at least one nonzero vector in *S*. We choose one nonzero vector in *S*, denoted by $\vec{a}_{n_1}$. Let $I_n: = \left\{1,\ 2,\ \cdots,\ n\right\}$ and we use the notation:

$$i \in I_n \backslash \left\{n_1\right\} \quad \text{means that } i \in I_n \text{ with } i \neq n_i.$$

1. If $\vec{a}_i \in \text{span}\left\{\vec{a}_{n_1}\right\}$ for $i \in I_n \backslash \left\{n_1\right\},$ then there exists $y_i \in \mathbb{R}$ such that

   (7.2.12)

   $$y_i\vec{a}_{n_1} = \vec{a}_i.$$

   Let $T = \left\{\vec{a}_{n_1}\right\}.$ Then $\text{span } S = \text{span } T.$ Indeed, it is obvious that span *T* is a subset of span *S*. We prove that span *S* is a subset of span *T*. Let $\vec{b} \in \text{span } S.$ Then there exist $x_1,\ x_2,\ \cdots, x_n \in \mathbb{R}$ such that

   $$\begin{aligned} \vec{b} &= x_1a_1 + x_2a_2 + \cdots + x_{n_1}a_{n_1} + \cdots + x_na_n \\ &= x_1y_1a_{n_1} + x_2y_2a_{n_2} + \cdots + x_{n_1}a_{n_1} + \cdots + x_ny_na_{n_1} \\ &= (x_1y_1 + x_2y_2 + \cdots + x_{n_1} + \cdots + x_ny_n)a_{n_1}. \end{aligned}$$

This shows $\vec{b} \in \operatorname{span} T$ and $\operatorname{span} S = \operatorname{span} T$. Now, we prove that $r(A) = 1$. By **(7.2.12)**, $\overrightarrow{a_1}$ is a linear combination of $\overrightarrow{a_{n_1}}$. By **Theorem 7.1.3**, $\left\{\overrightarrow{a_{n_1}}, \overrightarrow{a_1}\right\}$ is linearly dependent and

$$r((\overrightarrow{a_{n_1}}\overrightarrow{a_1})) = r((\overrightarrow{a_{n_1}}) = 1.$$

Because $y_2\overrightarrow{a_{n_1}} = \overrightarrow{a_2}$, by **Theorem 1.4.2**, $\overrightarrow{a_2}$ is a linear combination of $\left\{\overrightarrow{a_{n_1}}, \overrightarrow{a_1}\right\}$. By **Theorem 7.1.3**, $\left\{\overrightarrow{a_{n_1}}, \overrightarrow{a_1}, \overrightarrow{a_2}\right\}$ is linearly dependent and

$$r((\overrightarrow{a_{n_1}}, \overrightarrow{a_1}, \overrightarrow{a_2})) = r((\overrightarrow{a_{n_1}}\overrightarrow{a_1})) = r((\overrightarrow{a_{n_1}}) = 1.$$

Repeating the process implies

$$r((\overrightarrow{a_{n_1}}\overrightarrow{a_1}\overrightarrow{a_2}\cdots\overrightarrow{a_{j-1}}\overrightarrow{a_{j+1}}\cdots\overrightarrow{a_n})) = 1.$$

Let $B = (\overrightarrow{a_{n_1}}, \overrightarrow{a_1}, \overrightarrow{a_2}, \cdots\overrightarrow{a_{j-1}}, \overrightarrow{a_{j+1}}, \cdots, \overrightarrow{a_l})$. By the third row operations, it is easy to prove $r(B^T) = r(A^T)$. By **Theorem 2.5.5**, $r(B) = r(A) = 1$.

2. In the above, we have proved that if $\overrightarrow{a_i} \in \text{span}\left\{\overrightarrow{a_{n_1}}\right\}$ for each $i \in I_n := \left\{1, 2, \cdots, n\right\}$ with $i \neq n_1$, then $r(A) = 1$. Hence, if $r(A) > 1$, there exists $a_{n_2} \in S \setminus \left\{\overrightarrow{a_{n_1}}\right\}$ such that $a_{n_2} \notin \text{span}\left\{\overrightarrow{a_{n_1}}\right\}$. By **Theorem 7.2.4**, $\left\{\overrightarrow{a_{n_1}}, \overrightarrow{a_{n_2}}\right\}$ is linearly independent and by **Theorem 7.1.1**, $r((\overrightarrow{a_{n_1}}, \overrightarrow{a_{n_2}})) = 2$. Without loss of generalization, we assume that $n_1 < n_2$. If $\overrightarrow{a_i} \in \text{span}\left\{\overrightarrow{a_{n_1}}, \overrightarrow{a_{n_2}}\right\}$ for each $i \in I_n \setminus \left\{n_1, n_2\right\}$, then for each $i \in I_n \setminus \left\{n_1, n_2\right\}$, there exists $y_1, y_{i2} \in \mathbb{R}$ such that
(7.2.13)

$$y_{i1}\overrightarrow{a_{n_1}} + y_{i2}\overrightarrow{a_{n_2}} = \overrightarrow{a_i}.$$

Let $T = \left\{\overrightarrow{a_{n_1}}, \overrightarrow{a_{n_2}}\right\}$. Then $\text{span } S = \text{span } T$. We prove that span S is a subset of span T. Let $\vec{b} \in \text{span } S$. Then there exist $x_1, x_2, \cdots, x_n \in \mathbb{R}$ such that

$$\begin{aligned}\vec{b} &= x_1 a_1 + x_2 a_2 + \cdots + x_{n_1} a_{n_1} + \cdots + x_{n_2} a_{n_2} + \cdots + x_n a_n \\ &= \sum_{i=1,\ i \neq n_1,\ n_2}^{n} x_i a_i + x_{n_1} a_{n_1} + x_{n_2} a_{n_2} \\ &= \sum_{i=1,\ i \neq n1,\ n_2}^{n} x_i (y_{i1} \vec{a_{n_1}} + y_{i2} \vec{a_{n_2}}) + x_{n_1} a_{n_1} + x_{n_2} a_{n_2} \\ &= \left[(\sum_{i=1,\ i \neq n1,\ n_2}^{n} x_i y_{i1}) + x_{n_1} \right] \vec{a_{n_1}} + \left[(\sum_{i=1,\ i \neq n1,\ n_2}^{n} x_i y_{i2}) + x_{n_2} \right] a_{n_2}.\end{aligned}$$

This shows $\vec{b} \in \operatorname{span} T$ and $\operatorname{span} S = \operatorname{span} T$. Now, we prove that $r(A) = 2$. By **(7.2.13)** , $\vec{a_1}$ is a linear combination of $\left\{ \vec{a_{n_1}},\ \vec{a_{n_1}} \right\}$. By **Theorem 7.1.3** , $\left\{ \vec{a_{n_1}},\ \vec{a_{n_2}},\ \vec{a_1} \right\}$ is linearly dependent and

$$r((\vec{a_{n_1}} \vec{a_{n_2}} \vec{a_1})) = r((\vec{a_{n_1}} \vec{a_{n_2}}) = 2.$$

Because $y_{21} \vec{a_{n_1}} + y_{22} \vec{a_{n_2}} = \vec{a_2}$, by **Theorem 1.4.2** , $\vec{a_2}$ is a linear combination of $\vec{a_{n_1}},\ \vec{a_{n_2}},\ \vec{a_1}$. By **Theorem 7.1.3** , $\left\{ \vec{a_{n_1}},\ \vec{a_{n_2}},\ \vec{a_1},\ \vec{a_2} \right\}$ is linearly dependent and

$$r((\vec{a_{n_1}} \vec{a_{n_2}} \vec{a_1} \vec{a_2})) = r((\vec{a_{n_1}} \vec{a_{n_2}} \vec{a_1})) = r((\vec{a_{n_1}} \vec{a_{n_2}}) = 2.$$

Repeating the process implies

$$r((\vec{a}_{n_1}\vec{a}_{n_2}\vec{a}_1,\ \vec{a}_2,\ \cdots\vec{a}_{n_1-1},\ \vec{a}_{n_1+1}\cdots\vec{a}_{n_2-1},\ \vec{a}_{n_2+1}\cdots\vec{a}_n)) = 2.$$

Let $B = (\vec{a}_{n_1}\vec{a}_{n_2}\vec{a}_1,\ \vec{a}_2,\ \cdots\vec{a}_{n_1-1},\ \vec{a}_{n_1+1}\cdots\vec{a}_{n_2-1},\ \vec{a}_{n_2+1}\cdots\vec{a}_n)$. Then, by the third row operations, it is easy to prove $r(B^T) = r(A^T)$. By **Theorem 2.5.5** , $r(B) = r(A) = 2$.

Finally, we apply linear independence to study bases of the solution spaces of the homogenous system $A\vec{X} = \vec{0}$.

Theorem 7.2.7.

Let A be the same as in **(7.2.4)** *. Assume that* $\mathrm{null}(A) = k$. *Then the basis vectors* $\vec{v}_1,\ \vec{v}_2,\ \cdots,\ \vec{v}_k$ *given in* ***Definition 4.3.1*** *are linearly independent.*

Proof

Assume that there exist $x_1,\ \cdots,\ x_k \in \mathbb{R}$ such that

(7.2.14)

$$x_1\vec{v}_1 + x_2\vec{v}_2 + \cdots + x_k\vec{v}_k = \vec{0}.$$

Let $\vec{v} = x_1\vec{v}_1 + x_2\vec{v}_2 + \cdots + x_k\vec{v}_k$. By **Theorem 4.3.1** (P_1), the components on $n_1,\ n_2\cdots,\ n_k(1 \le n_1 < n_2 < \cdots < n_k \le n)$ are $x_1,\ x_2,\ \cdots,\ x_k$, respectively. By **(7.2.14)** , $\vec{v} = \vec{0}$ implies that $x_1 = x_2 = \cdots = x_k = 0$ and thus, $\vec{v}_1,\ \vec{v}_2,\ \cdots,\ \vec{v}_k$ are linearly independent.

By **Theorem 7.2.7** and **Definition 7.2.1**, we see that the set of vectors $\vec{v}_1, \vec{v}_2, \cdots, \vec{v}_k$ given in **Definition 4.3.1** is a basis of the solution space N_A. As mentioned above, the bases of a spanning space are not unique; hence, the solution space N_A has many other bases.

By **Theorems 7.2.3** and **7.2.7**, we will prove the following result, which shows that the number of vectors in all bases of N_A is equal to the nullity null(*A*) of the matrix *A*.

Theorem 7.2.8.

*Let $\vec{v}_1, \vec{v}_2, \cdots, \vec{v}_k$ be the basis vectors of N_A given in **Definition 4.3.1** and let $\vec{u}_1, \vec{u}_2, \cdots, \vec{u}_l$ be vectors in N_A. Assume that the following conditions hold.*

1. $\vec{u}_1, \vec{u}_2, \cdots, \vec{u}_l$ *are linearly independent.*
2. $N_A = \operatorname{span}\left\{\vec{u}_1, \vec{u}_2, \cdots, \vec{u}_l\right\}$.

Then $l = k$.

Proof

By **Theorem 7.2.7**, the basis vectors $\vec{v}_1, \vec{v}_2, \cdots, \vec{v}_k$ are linearly independent. By conditions (1) and (2) and **Theorem 7.2.3**, $l = k$.

Theorem 7.2.8 shows that any set of *k*-linearly independent vectors in $N_A = \operatorname{span}\left\{\vec{v}_1, \vec{v}_2, \cdots, \vec{v}_k\right\}$ constitutes a basis of N_A. We shall use **Theorem 7.2.8** in the proof of **Theorem 8.2.4**.

Exercises

1. Let $\overrightarrow{a_1} = (1, \; -1, \; 1, \; 0)$, and $\overrightarrow{a_2} = (1, \; 0, \; 1, \; 1)$. Determine whether $S = \left\{ \overrightarrow{a_1}, \; \overrightarrow{a_2} \right\}$ is a basis of span S. If so, find dim(span S).
2. Determine whether $S = \left\{ \overrightarrow{a_1}, \; \overrightarrow{a_2}, \; \overrightarrow{a_3} \right\}$ is a basis of span S, where

   i. $\overrightarrow{a_1} = (1, \; 0, \; -1, \; 1)$, $\overrightarrow{a_2} = (0, \; 0, \; 1, \; 1)$ and $\overrightarrow{a_3} = (-2, \; 1, \; 0, \; 1)$.

   ii. $\overrightarrow{a_1} = (1, \; 0, \; 1, \; 1)$, $\overrightarrow{a_2} = (0, \; 0, \; 1, \; -1)$ and $\overrightarrow{a_3} = (1, \; 0, \; 2, \; 0)$.

3. Let $\overrightarrow{a_1} = (1, \; -1)$, $\overrightarrow{a_2} = (0, \; 1)$, $\overrightarrow{a_3} = (-3, \; 2)$ and $\overrightarrow{a_4} = (1, \; 0)$. Determine whether $S = \left\{ \overrightarrow{a_1}, \; \overrightarrow{a_2}, \; \overrightarrow{a_3}, \; \overrightarrow{a_4} \right\}$ is a basis of span S.
4. Let $\overrightarrow{a_1} = (1, \; 0, \; 1)$ and $\overrightarrow{a_2} = (1, \; 1, \; 1)$. Find a vector $S = \left\{ \overrightarrow{a_1}, \; \overrightarrow{a_2}, \; \vec{b} \right\}$ is a basis of span S.
5. Use **Theorem 7.1.1** to prove **Theorem 7.2.2**.
6. Use **Theorem 7.1.1** to prove **Theorem 7.2.4**.

7.3 Methods of finding bases of spanning spaces

Let *S* and *A* be the same as in **(7.2.1)** and **(7.2.4)** and $B = A^T$. Let *T* be the linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$ with standard matrix *A*.

Definition 7.3.1.

The spanning space span *S* is said to be the column space of *A*, and to be the row space of *B*.

We denote by C_A and R_B the column space of *A* and the row space of *B*, respectively. Then

(7.3.1)

$$C_A = R_B = \text{span } S = A(\mathbb{R}^n) = T(\mathbb{R}^n).$$

Note that in general, $C_A \neq R_A$.

If the set *S* given in **(7.2.1)** is not linearly independent, then it is not a basis of span *S*. Our purpose is to find linearly independent vectors in $\mathbb{R}^m$, say $T = \left\{ \vec{b}_1, \ \vec{b}_2, \ \cdots, \ \vec{b}_r \right\}$ in $\mathbb{R}^m$, such that span $T =$ span S, that is,

(7.3.2)

$$\text{span}\left\{ \vec{b}_1, \ \vec{b}_2, \ \cdots, \ \vec{b}_r \right\} = \text{span}\left\{ \vec{a}_1, \ \vec{a}_2, \ \cdots, \ \vec{a}_n \right\}.$$

By **Definition 7.2.1**, *T* is a basis of span *S*.

Usually, we seek bases with all the vectors satisfying one of the following requirements:

1. All the basis vectors in *T* belong to *S*.
2. The basis vectors in *T* may or may not be in *S*.

We first prove two results, which provide the methods to find basis vectors that satisfy the above requirement (1) or (2).

We start with bases that satisfy the above requirement (1). Let

7.3.3

$$C=\begin{pmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ c_{21} & c_{22} & \cdots & c_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ c_{m1} & c_{m2} & \cdots & c_{mm} \end{pmatrix}$$

be a row echelon matrix and $\overrightarrow{C_i}$ the *i*th column vector of *C*.

By **Theorems 7.1.1** and **7.2.5** , we prove the following result, which shows that the column vectors with the leading entries of a row echelon matrix *C* form a basis for the column space of *C*.

Lemma 7.3.1.

Let C be a row echelon matrix and let $\overrightarrow{C_{n_1}}, \overrightarrow{C_{n_2}} \cdots, \overrightarrow{C_{n_r}}$ *be the column vectors of C containing leading entries. Then the following assertions hold.*

i. $\left\{\overrightarrow{C_{n_1}}, \overrightarrow{C_{n_2}} \cdots, \overrightarrow{C_{n_r}}\right\}$ *is linearly independent.*

ii. $\operatorname{span}\left\{\overrightarrow{C_{n1}},\ \overrightarrow{C_{n2}}\cdots,\ \overrightarrow{C_{nr}}\right\}=\operatorname{span}\left\{\overrightarrow{C_1},\ \overrightarrow{C_2}\cdots,\ \overrightarrow{C_n}\right\}.$

Proof

i. Let $M=(\overrightarrow{C_{n_1}}\overrightarrow{C_{n_2}}\cdots\overrightarrow{C_{n_r}}).$ By **Theorem 2.5.1** (4), $r(M)=r.$ This, together with **Theorem 7.1.1** (1), implies that $\left\{\overrightarrow{C_{n_1}},\ \overrightarrow{C_{n_2}}\cdots,\ \overrightarrow{C_{n_r}}\right\}$ is linearly independent.

ii. By the result (*i*) and **Theorem 7.2.5** , the result (*ii*)holds.

By **Lemma 7.3.1** , we prove the following result, which shows that if *C* is a row echelon matrix of *A* and a set of column vectors of *C* formsabasisfor the column space of *C*, then the corresponding column vectors of *A* form a basis for the column space of *A*, which is equal to span *S*.

Theorem 7.3.1.

Let S be the same as in **(7.2.1)** *and* $A=(\overrightarrow{a_1}\overrightarrow{a_2}\cdots\overrightarrow{a_n})$ *be the same as in* **(7.2.4)** *. Let* $C=(\overrightarrow{C_1}\overrightarrow{C_2}\cdots\overrightarrow{C_n})$ *be a row echelon matrix of A. Let* $\overrightarrow{C_{n_1}},\ \overrightarrow{C_{n_2}},\ \cdots,\ \overrightarrow{C_{n_r}}$ *be the column vectors of C containing leading entries. Then the set of the pivot columns of A, that is,* $\left\{\overrightarrow{a_{n_1}},\ \overrightarrow{a_{n_2}},\ \cdots,\ \overrightarrow{a_{n_r}}\right\}$ *is a basis of* span *S.*

Proof

Because *C* is a row echelon matrix of *A* and there are *r* columns vectors of *C* that contain the leading entries of *C*, by **Theorem 2.5.1** , $r(A)=r(C)=r.$ By **Lemma 7.3.1** , we have

(7.3.4)

$$\operatorname{span}\left\{\overrightarrow{C_{n1}},\ \overrightarrow{C_{n2}},\ \cdots,\ \overrightarrow{C_{nr}}\right\}=\operatorname{span}\left\{\overrightarrow{C_1},\ \overrightarrow{C_2},\ \cdots,\ \overrightarrow{C_n}\right\}.$$

It follows from **Theorem 2.7.3** that there exists an invertible matrix *F* such that $A = FC$. Hence, by **(2.2.6)** , we have

$$A = (\overrightarrow{a_1}\overrightarrow{a_2}\cdots\overrightarrow{a_n}) = F(\overrightarrow{C_1}\overrightarrow{C_2}\cdots\overrightarrow{C_n}) = F(\overrightarrow{FC_1}\overrightarrow{FC_2}\cdots\overrightarrow{FC_n})$$

and thus,

(7.3.5)

$$\overrightarrow{a_i} = \overrightarrow{FC_i} \quad \text{for } i \in I_n.$$

Let $A^* = (\overrightarrow{a_{n_1}}\overrightarrow{a_{n_2}}\cdots\overrightarrow{a_{n_r}})$ and $C^* = (\overrightarrow{C_{n_1}}\overrightarrow{C_{n_2}}\cdots\overrightarrow{C_{n_r}})$. By **(7.3.5)** and **(2.2.6)** , $A^* = (\overrightarrow{a_{n_1}}\overrightarrow{a_{n_2}}\cdots\overrightarrow{a_{n_r}}) = (\overrightarrow{FC_{n_1}}\overrightarrow{FC_{n_2}}\cdots\overrightarrow{FC_{n_r}}) = F(\overrightarrow{C_{n_1}}\overrightarrow{C_{n_2}}\cdots\overrightarrow{FC_{n_r}}) = FC^*$. Because *F* is invertible, by **Corollary 2.7.3** ,

$$r(A^*) = r(C^*) = r(C) = r(A) = r.$$

By **Theorem 7.1.1** (1), $S^* = \left\{\overrightarrow{a_{n_1}},\ \overrightarrow{a_{n_2}},\ \cdots,\ \overrightarrow{a_{n_r}}\right\}$ is linearly independent.

By **(7.3.5)** and **(7.3.4)** , we have

$$\begin{aligned} \text{span } S &= \text{ span}\left\{\overrightarrow{a_1},\ \overrightarrow{a_2},\ \cdots,\ \overrightarrow{a_n}\right\} = \text{span}\left\{\overrightarrow{FC_1},\ \overrightarrow{FC_2},\ \cdots,\ \overrightarrow{FC_n}\right\} \\ &= F\,\text{span}\left\{\overrightarrow{C_1},\ \overrightarrow{C_2},\ \cdots,\ \overrightarrow{C_n}\right\} = F\,\text{span}\left\{\overrightarrow{C_{n_1}},\ \overrightarrow{C_{n_2}},\ \cdots,\ \overrightarrow{EC_{n_r}}\right\} \\ &= \text{span}\left\{\overrightarrow{FC_{n_1}},\ \overrightarrow{FC_{n_2}},\ \cdots,\ \overrightarrow{FC_{n_r}}\right\} = \text{span } S^*.\overrightarrow{a_{n_2}},\ \cdots,\ \overrightarrow{a_{n_r}}\right\}. \end{aligned}$$

By **Definition 7.2.1**, S^* is a basis of span S^* and thus a basis of span *S*.

Next, we work with bases that satisfy the above requirement (2). We first give the following result, which shows that if *B* and *D* are row equivalent (see **Definition 2.5.2**), then all the row vectors of *D* are contained in the spanning space of the row vectors of *B*.

Lemma 7.3.2.

Assume that B and D are row equivalent. Then, each row of D is a linear combination of the row vectors of B.

Proof

Because *B* and *D* are row equivalent, by **Theorem 2.7.3**, there exists an invertible matrix *F* such that $D = FB$. By **Theorem 2.2.3**, we have

$$D^T = B^T F^T = AF^T,$$

where *A* is the same as in **(7.2.4)**.

Let $\overrightarrow{d_i}$ be the *i*th row vector of *D* for $i \in I_n$ and $\overrightarrow{X_i} = (x_{1i}, \ \cdots x_{ni})^T$ be the *i*th column vector of F^T for $i \in I_n$. By **(2.2.6)**, we have for $i \in I_n$,

$$\overrightarrow{d_i}^T = A\overrightarrow{X_i}^T = x_{1i}\overrightarrow{a_1} + x_{2i}\overrightarrow{a_2} + \cdots + x_{ni}\overrightarrow{a_n}$$

and the result follows.

Theorem 7.3.2.

Assume that D is a row-echelon matrix of B. Then the row vectors with the leading entries of D form a basis for R_B.

Proof

Let $\overrightarrow{d_1},\ \overrightarrow{d_2},\ \cdots,\ \overrightarrow{d_n}$ be the row vectors of *D*. Because *D* is a row-echelon matrix, we assume that $\overrightarrow{d_1},\ \overrightarrow{d_2},\ \cdots,\ \overrightarrow{d_r}$ are the row vectors of *D* with the leading entries. Then other rows from $\overrightarrow{d_{r+1}}$ to $\overrightarrow{d_n}$ are zero rows if $r < n$. Let $D^* = (\overrightarrow{d_1}\,\overrightarrow{d_2}\cdots\overrightarrow{d_r})$. Because $r(D^*) = r$, by **Theorem 7.1.1** , the set D^* is linearly independent. Then by **Theorem 2.5.4** and **Theorem 2.5.1** (3),

$$r(B) = r(D) = r(D^*) = r.$$

Because *D* is a row-echelon matrix of *B*, it follows from **Lemma 7.3.2** that

$$\overrightarrow{d_i} \in \operatorname{span}\left\{\overrightarrow{a_1},\ \overrightarrow{a_2},\ \cdots,\ \overrightarrow{a_n}\right\} = \operatorname{span} S \quad \text{for } i = 1,\ 2,\ \cdots,\ r,$$

where *S* is the same as in **(7.2.1)** . By **Theorem 7.2.5** ,

$$\operatorname{span}\left\{\overrightarrow{d_1},\ \overrightarrow{d_2},\ \cdots,\ \overrightarrow{d_r}\right\} = \operatorname{span}\left\{\overrightarrow{a_1},\ \overrightarrow{a_2},\ \cdots,\ \overrightarrow{a_n}\right\}$$

and $\left\{\overrightarrow{d_1},\ \overrightarrow{d_2},\ \cdots,\ \overrightarrow{d_r}\right\}$ is a basis of R_B.

By **Theorems 7.3.1** and **7.3.2** , we see that there are two methods to find the bases.

Method 1. Finding bases of span *S* satisfying the above requirement (1) is equivalent to finding the pivot columns of *A* (see Section 2.9). Hence, we use row operations to change *A* to a row echelon matrix

$$C = (\overrightarrow{C_1}\,\overrightarrow{C_2}\cdots\overrightarrow{C_n}).$$

Let $\overrightarrow{C_{n_1}},\ \overrightarrow{C_{n_2}},\ \cdots,\ \overrightarrow{C_{n_r}}$ be the column vectors of *C* containing leading entries. We shall prove that the corresponding column vectors (pivot columns) in *A*

$$\left\{\overrightarrow{a_{n_1}},\ \overrightarrow{a_{n_2}},\ \cdots,\ \overrightarrow{a_{n_r}}\right\}$$

are a basis of span *S*.

Method 2. To find bases of span *S* that satisfy the above requirement (2), we use row operations to change *B* to a row echelon matrix *D*. Let $\overrightarrow{d_{m1}},\ \overrightarrow{d_{m2}},\ \cdots,\ \overrightarrow{d_{mr}}$ be the column vectors of *D* containing leading entries. Then, we shall prove that

$$\left\{\overrightarrow{d_{m1}},\ \overrightarrow{d_{m2}},\ \cdots,\ \overrightarrow{d_{mr}}\right\}$$

is the basis vector of span *S*.

Example 7.3.1.

Find a basis for the column space of *C*, where

$$C = \begin{pmatrix} 1 & 1 & 2 & 1 \\ 0 & 0 & -1 & -3 \\ 0 & 0 & 0 & 8 \end{pmatrix}.$$

Solution

We denote the column vectors of *C* by $\overrightarrow{C_i},\ i = 1,\ 2,\ 3,\ 4,$ that is,

$$C = (\overrightarrow{C_1}\overrightarrow{C_2}\overrightarrow{C_3}\overrightarrow{C_4}).$$

The column vectors of *C* containing the leading entries are $\overrightarrow{C_1},\ \overrightarrow{C_3},\ \overrightarrow{C_4}$, and by **Theorem 7.3.1** ,
$\left\{\overrightarrow{C_1},\ \overrightarrow{C_3},\ \overrightarrow{C_4}\right\}$ is a basis for the column space of *C*.

Example 7.3.2.

Find a set of column vectors of *A* that forms a basis for C_A and $\dim(C_A)$, where

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

Solution

We denote the column vectors of *A* by $\overrightarrow{a_i},\ i = 1,\ 2,\ 3,\ 4.$ Then

$$A = (\overrightarrow{a_1}\overrightarrow{a_2}\overrightarrow{a_3}\overrightarrow{a_4}).$$

$$A = \begin{pmatrix} 1 & 1 & 2 & 1 \\ 1 & 1 & 1 & -2 \\ 2 & 2 & 1 & 1 \end{pmatrix} \xrightarrow[R_1(-2)+R_3]{R_1(-1)+R_2} \begin{pmatrix} 1 & 1 & 2 & 1 \\ 0 & 0 & -1 & -3 \\ 0 & 0 & -3 & -1 \end{pmatrix}$$
$$\xrightarrow{R_2(-3)+R_3} \begin{pmatrix} 1 & 1 & 2 & 1 \\ 0 & 0 & -1 & -3 \\ 0 & 0 & 0 & 8 \end{pmatrix} = C,$$

where *C* is the same as in **Example 7.3.1** . From the matrix *C*, we see that the column vectors $\overrightarrow{C_1},\ \overrightarrow{C_3},\ \overrightarrow{C_4}$ contain leading entries of *C*. The pivot column vectors in *A* are $\overrightarrow{a_1},\ \overrightarrow{a_3},\ \overrightarrow{a_4}$, that is,

$$\overrightarrow{a_1} = \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix}, \overrightarrow{a_3} = \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix}, \overrightarrow{a_4} = \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix}.$$

By **Theorem 7.3.1** , the column vectors $\overrightarrow{a_1}, \overrightarrow{a_3}, \overrightarrow{a_4}$ of *A* formabasisof C_A and

$$\operatorname{span}\left\{\overrightarrow{a_1}, \overrightarrow{a_3}, \overrightarrow{a_4} = \operatorname{span}\left\{\overrightarrow{a_1}, \overrightarrow{a_2}, \overrightarrow{a_3}, \overrightarrow{a_4}\right\}.$$

It is obvious that $\dim(C_A) = 3$.

Example 7.3.3.

Find a basis for the row space of the following matrix

$$D = \begin{pmatrix} 1 & 0 & 0 & 4 \\ 0 & 1 & 0 & 7 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Solution

Note that *D* is a row echelon matrix. By **Theorem 7.3.2** , the row vectors with the leading entries in *D* form a basis for R_D. Let $\overrightarrow{d_1} = (1, 0, 0, 4)^T$, $\overrightarrow{d_2} = (0, 1, 0, 7)^T$, and $\overrightarrow{d_3} = (0, 0, 0, -1)^T$. Then $\overrightarrow{d_1}, \overrightarrow{d_2}, \overrightarrow{d_3}$ form a basis of R_D.

Example 7.3.4.

Find the set of basis vectors for R_B, where

$$B=\begin{pmatrix}1 & -1 & 3\\ 2 & 0 & 4\\ -1 & -3 & 1\end{pmatrix}.$$

Solution

We use row operations to change *B* to a row echelon matrix.

$$B=\begin{pmatrix}1 & -1 & 3\\ 2 & 0 & 4\\ -1 & -3 & 1\end{pmatrix}\xrightarrow[R_1(1)+R_3]{R_1(-2)+R_2}\begin{pmatrix}1 & -1 & 3\\ 0 & 2 & -2\\ 0 & -4 & 4\end{pmatrix}\xrightarrow{R_2(2)+R_3}$$
$$\begin{pmatrix}1 & -1 & 3\\ 0 & 2 & 2\\ 0 & 0 & 0\end{pmatrix}=D.$$

Let $\overrightarrow{d_1}=(1,\ -1,\ 3),\ \overrightarrow{d_2}=(0,\ 2,\ 2).$ Then by **Theorem 7.3.2** , $\overrightarrow{d_1},\ \overrightarrow{d_2}$ are basis vectors for R_B.

Note that $\overrightarrow{d_1}$ is a row vector of *B*, but $\overrightarrow{d_2}$ is not a row vector of *B*, so the basis $\left\{\overrightarrow{d_1},\ \overrightarrow{d_2}\right\}$ does not satisfy the requirement (1).

Example 7.3.5.

Let $S=\left\{\overrightarrow{a_1},\ \overrightarrow{a_2},\ \overrightarrow{a_3},\ \overrightarrow{a_4}\right\}$, where

$$\overrightarrow{a_1}=\begin{pmatrix}1\\ 2\\ -3\end{pmatrix},\ \overrightarrow{a_2}=\begin{pmatrix}-2\\ 0\\ 4\end{pmatrix},\ \overrightarrow{a_3}=\begin{pmatrix}0\\ 4\\ -2\end{pmatrix},\ \overrightarrow{a_4}=\begin{pmatrix}-2\\ -4\\ 6\end{pmatrix}.$$

a. Use **Theorem 7.3.1** to find a subset of S that forms a basis for span S.
b. Use **Theorem 7.3.2** to find another basis for span S.

Solution

a. Because we seek a subset of S that forms a basis for span S, we need to put the vectors in S as the column vectors of a matrix A, that is,

$$A=\begin{pmatrix} 1 & -2 & 0 & -2 \\ 2 & 0 & 4 & -4 \\ -3 & 4 & -2 & 6 \end{pmatrix}.$$

We use row operations to change A into a row echelon matrix.

$$A=\begin{pmatrix} 1 & -2 & 0 & -2 \\ 2 & 0 & 4 & -4 \\ -3 & 4 & -2 & 6 \end{pmatrix} \xrightarrow[R_1(3)+R_3]{R_1(-2)+R_2} \begin{pmatrix} 1 & -2 & 0 & -2 \\ 0 & 4 & 4 & 0 \\ 0 & -2 & -2 & 0 \end{pmatrix}$$
$$\xrightarrow{R_2\left(\frac{1}{2}\right)+R_3} \begin{pmatrix} 1 & -2 & 0 & -2 \\ 0 & 4 & 4 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}=C.$$

The column vectors $\overrightarrow{C_1}$, $\overrightarrow{C_2}$ of C contain the leading entries of C. The corresponding column vectors of A are $\overrightarrow{a_1}$, $\overrightarrow{a_2}$. By **Theorem 7.3.1** , $\left\{\overrightarrow{a_1}, \overrightarrow{a_2}\right\}$ forms a basis of span S.

b. We define a matrix B by putting the vectors in S as its row vectors

$$B=\begin{pmatrix} 1 & 2 & -3 \\ -2 & 0 & 4 \\ 0 & 4 & -2 \\ -2 & -4 & 6 \end{pmatrix}.$$

We use row operations to change *B* to a row echelon matrix *D*.

$$B=\begin{pmatrix}1&2&-3\\-2&0&4\\0&4&-2\\-2&-4&6\end{pmatrix}\xrightarrow[R_1(2)+R_4]{R_1(2)+R_2}\begin{pmatrix}1&2&-3\\0&4&-2\\0&4&-2\\0&0&0\end{pmatrix}\xrightarrow{R_2(-1)+R_3}$$
$$\begin{pmatrix}1&2&-3\\0&4&-2\\0&0&0\\0&0&0\end{pmatrix}=D.$$

Let $\overrightarrow{d_1}=(1,\ 2,\ -3),\ \overrightarrow{d_2}=(0,\ 4,\ -2).$ It follows from **Theorem 7.3.2** that $\left\{\overrightarrow{d_1},\ \overrightarrow{d_2}\right\}$ forms a basis for span *S*.

Exercises

1. Find a basis for the column space of *C*, where

$$C=\begin{pmatrix}1&-3&4&-2&5&4\\0&0&1&3&-2&-6\\0&0&0&0&1&5\\0&0&0&0&0&0\end{pmatrix}.$$

2. For each of the following matrices, find a set of its column vectors that forms a basis for its column space.

$$A_1 = \begin{pmatrix} 0 & 2 & -1 \\ 1 & -1 & 3 \\ 2 & 0 & 1 \end{pmatrix} \qquad A_2 = \begin{pmatrix} 1 & 3 & 0 & 2 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 2 & 4 & 1 & -1 & 0 \\ 0 & 0 & 0 & 3 & 1 \end{pmatrix}$$

$$A_3 = \begin{pmatrix} 1 & -3 & 4 & -2 & 5 & 4 \\ 2 & -6 & 9 & -1 & 8 & 2 \\ 2 & -6 & 9 & -1 & 9 & 7 \\ -1 & 3 & -4 & 2 & -5 & -4 \end{pmatrix}.$$

3. For each of the following row echelon matrices, find the basis and dimension of its row space and use the basis vectors to denote the row space.

$$D_1 = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad D_2 = \begin{pmatrix} 0 & 1 & -2 & 0 & 1 \\ 0 & 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \quad D_3 = \begin{pmatrix} 1 & 2 & 5 & 0 & 3 \\ 0 & 1 & 3 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

4. Find a set of basis vectors for R_B and $\dim R_B$, where

$$B = \begin{pmatrix} 1 & -1 & 3 \\ 2 & 0 & 4 \\ -1 & -3 & 1 \end{pmatrix}.$$

5. Let $S = \left\{\overrightarrow{a_1}, \overrightarrow{a_2}, \overrightarrow{a_3}, \overrightarrow{a_4}, \overrightarrow{a_5}\right\}$, where

$$\overrightarrow{a_1} = (1, -2, 0, 3)^T, \overrightarrow{a_2} = (2, -5, -3, 6)^T, \overrightarrow{a_3} = (0, 1, 3, 0)^T,$$
$$\overrightarrow{a_4} = (2, -1, 4, -7)^T, \overrightarrow{a_5} = (5, -8, 1, 2)^T.$$

a. Find a subset of S that forms a basis for the spanning space span S.
b. Find another basis for span S.

7.4 Coordinates and orthonormal bases

In **Section 7.2**, we see that $\mathbb{R}^n$ has a standard basis *E* given in **(7.2.3)** and **Theorem 7.2.1** (*ii*) provides a necessary and sufficient condition for a set of *n* vectors in $\mathbb{R}^n$ to be a basis of $\mathbb{R}^n$. Moreover, any set of *n* linearly dependent vectors in $\mathbb{R}^n$ constitutes a basis of $\mathbb{R}^n$.

Let $S = \left\{ \overrightarrow{a_1}, \overrightarrow{a_2}, \cdots, \overrightarrow{a_n} \right\}$ be a basis of $\mathbb{R}^n$. Then

$$\mathbb{R}^n = \operatorname{span}\left\{ \overrightarrow{a_1}, \overrightarrow{a_2}, \cdots, \overrightarrow{a_n} \right\}.$$

By **Theorem 7.2.2** with $n = m$, we obtain the following result.

Theorem 7.4.1.

If $S = \left\{ \overrightarrow{a_1}, \overrightarrow{a_2}, \cdots, \overrightarrow{a_n} \right\}$ *is a basis of* $\mathbb{R}^n$, *then for each* $\vec{b} \in \mathbb{R}^n$, *there exists a unique vector* $(x_1, x_2, \cdots, x_n)$ *such that*

(7.4.1)

$$\vec{b} = x_1\overrightarrow{a_1} + x_2\overrightarrow{a_2} + \cdots x_n\overrightarrow{a_n}.$$

Definition 7.4.1.

The numbers $x_1, x_2, \cdots, x_n$ in **(7.4.1)** are said to be the coordinates of the vector $\vec{b}$ relative to the basis *S*. We write

(7.4.2)

$$(\vec{b})_S = (x_1, x_2, \cdots, x_n).$$

If $\vec{b} = (b_1, b_2, \cdots, b_n) \in \mathbb{R}^n$, then the coordinates $b_1, b_2, \cdots, b_n$ of $\vec{b}$ are the coordinates of $\vec{b}$ relative to the standard basis *E*. Hence, according to **(7.4.2)**, we have

(7.4.3)

$$\vec{b} = (\vec{b})_E.$$

Let *A* be the matrix whose column vectors are $\vec{a}_1, \vec{a}_2, \cdots, \vec{a}_n$. Then we can rewrite **(7.4.1)** as

(7.4.4)

$$(\vec{b})_E = A(\vec{b})_S.$$

By **Theorem 7.2.1** (*ii*), *A* is invertible. Hence, we obtain

(7.4.5)

$$(\vec{b})_S = A^{-1}(\vec{b})_E = A^{-1}\vec{b} \quad \text{for } \vec{b} \in \mathbb{R}^n.$$

Example 7.4.1.

Let $S = \left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \end{pmatrix} \right\}$ be a set of vectors in $\mathbb{R}^2$.

i. Show that S is a basis of $\mathbb{R}^2$.
ii. Let $\vec{b} = (2, \ -4)$. Find $(\vec{b})_S$.

Solution

i. Let $A = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$. Then $|A| = 1 + 1 = 2 \neq 0$. By **Theorem 7.2.1** (*ii*), S is a basis of $\mathbb{R}^2$.

ii. By **Theorem 2.7.5**, $A^{-1} = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}$. By **(7.4.5)**,

$$(\vec{b})_S = A^{-1}\vec{b} = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}\begin{pmatrix} 2 \\ -4 \end{pmatrix} = \begin{pmatrix} -1 \\ -3 \end{pmatrix},$$

that is, $(\vec{b})_S = (-1, \ -3)$ are the coordinates of the vector $\vec{b} = (2, \ -4)$ relative to the basis S.

In **Example 7.4.1**, we used the inverse matrix method to find the coordinates $(\vec{b})_S$. In fact, all the other methods given in **Chapter 4**, including Gaussian elimination, Gauss-Jordan elimination, and Cramer's rule, can be employed to solve the system $A\vec{X} = \vec{b}$.

Example 7.4.2.

Let $\overrightarrow{a_1} = (1, \ 0, \ 1)$, $\overrightarrow{a_2} = (0, \ 1, \ 1)$, and $\overrightarrow{a_3} = (1, \ 1, \ 0)$.

i. Show that $S=\left\{\overrightarrow{a_1},\ \overrightarrow{a_2},\ \overrightarrow{a_3}\right\}$ is a basis of $\mathbb{R}^3$.

ii. Let $\vec{b}=(1,\ 1,\ 1)^T$. Find $(\vec{b})_S$.

Solution

i. Because $|A| = \begin{vmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \end{vmatrix} = -2 \neq 0$, by **Theorem 7.2.1** (*ii*), $\left\{\overrightarrow{a_1},\ \overrightarrow{a_2},\ \overrightarrow{a_3}\right\}$ is a basis of $\mathbb{R}^3$.

ii.

$$(A\,|\,\vec{b}) = \begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 \end{pmatrix} \xrightarrow{R_1(-1)+R_3} \begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 1 & -1 & 0 \end{pmatrix} \xrightarrow{R_2(-1)+R_3}$$

$$\begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & -2 & -1 \end{pmatrix} \xrightarrow{R_3\left(-\frac{1}{2}\right)} \begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & \frac{1}{2} \end{pmatrix} \xrightarrow[R_3(-1)+R_1]{R_3(-1)+R_2}$$

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

Hence, $(\vec{b})_S = \left(\frac{1}{2},\ \frac{1}{2},\ \frac{1}{2}\right)$.

In **Example 7.4.2**, the coordinates of $\vec{b} = (1,\ 1,\ 1)$ relative to the basis *S* are $\left(\frac{1}{2},\ \frac{1}{2},\ \frac{1}{2}\right)$ while its coordinates relative to the standard basis *E* are (1, 1, 1).

Equation **(7.4.5)** gives the relation between the coordinates of a vector $\vec{b}$ relative to a basis *S* and the standard basis *E*.

Now, we study the relations between the coordinates of a vector $\vec{b}$ relative to two bases of $\mathbb{R}^n$, that is, we shall generalize the relation given in **(7.4.5)** by replacing the standard basis *E* by any basis of $\mathbb{R}^n$.

By **Theorem 7.2.1** (*ii*) and **Lemma 7.2.1** (*ii*) with $m = n = k$, we obtain the following result, which gives the change of bases.

Theorem 7.4.2.

Let $S = \left\{\overrightarrow{a_1},\ \overrightarrow{a_2},\ \cdots,\ \overrightarrow{a_n}\right\}$ *and* $T = \left\{\overrightarrow{b_1},\ \overrightarrow{b_2},\ \cdots,\ \overrightarrow{b_n}\right\}$ *be two bases of* $\mathbb{R}^n$. *Let* $A = (\overrightarrow{a_1}\overrightarrow{a_2}\cdots\overrightarrow{a_n})$, $B = (\overrightarrow{b_1}\overrightarrow{b_2}\cdots\overrightarrow{b_n})$. *Then A and B are invertible and there exists a unique invertible* $n \times n$ *matrix*

(7.4.6)

$$C=\begin{pmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ c_{21} & c_{22} & \cdots & c_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ c_{n_1} & c_{n_2} & \cdots & c_{nn} \end{pmatrix}$$

such that $C=A^{-1}B$ *and*

(7.4.7)

$$\begin{cases} \vec{b}_1 = c_{11}\vec{a}_1 + c_{21}\vec{a}_2 + \cdots + c_{n_1}\vec{a}_n \\ \vec{b}_2 = c_{12}\vec{a}_1 + c_{22}\vec{a}_2 + \cdots + c_{n_2}\vec{a}_n \\ \vdots \\ \vec{b}_n = c_{1n}\vec{a}_1 + c_{2n}\vec{a}_2 + \cdots + c_{nn}\vec{a}_n. \end{cases}$$

Definition 7.4.2.

Let $S=\left\{\vec{a}_1, \vec{a}_2, \cdots, \vec{a}_n\right\}$ and $T=\left\{\vec{b}_1, \vec{b}_2, \cdots, \vec{b}_n\right\}$ be two bases of $\mathbb{R}^n$. The matrix *C* given in **(7.4.6)** is said to be the transition matrix from the basis *T* to the basis *S*.

The following result gives the change of coordinates under two bases.

Theorem 7.4.3.

Let $S = \left\{ \overrightarrow{a_1}, \overrightarrow{a_2}, \cdots, \overrightarrow{a_n} \right\}$ *and* $T = \left\{ \overrightarrow{b_1}, \overrightarrow{b_2}, \cdots, \overrightarrow{b_n} \right\}$ *be two bases of* $\mathbb{R}^n$. *Then for each vector* $\vec{b} \in \mathbb{R}^n$,

(7.4.8)

$$(\vec{b})_S = C(\vec{b})_T.$$

Proof

Let A and B be the same as in **Theorem 7.4.2** . Because S and T are bases of $\mathbb{R}^n$, by **(7.4.4)** , we have for each vector $\vec{b} \in \mathbb{R}^n$,

(7.4.9)

$$(\vec{b})_E = A(\vec{b})_S \text{ and } (\vec{b})_E = B(\vec{b})_T.$$

By **Theorem 7.4.2** and **(7.4.9)** , we have

$$(\vec{b})_S = A^{-1}(\vec{b})_E = A^{-1}(B(\vec{b})_T) = (A^{-1}B)(\vec{b})_T = C(\vec{b})_T$$

and **(7.4.8)** holds.

Example 7.4.3.

Let $S = \left\{ \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ -2 \end{pmatrix} \right\}$ and $T = \left\{ \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \end{pmatrix} \right\}$. Assume that $(\vec{b})_T = \begin{pmatrix} 5 \\ -10 \end{pmatrix}$. Find $(\vec{b})_S$.

Solution

Let $A = \begin{pmatrix} 2 & 1 \\ 1 & -2 \end{pmatrix}$ and $B = \begin{pmatrix} 1 & -1 \\ 2 & 1 \end{pmatrix}$. By **Theorem 2.7.5** ,

$$A^{-1} = \frac{1}{-5}\begin{pmatrix} -2 & -1 \\ -1 & 2 \end{pmatrix} = \begin{pmatrix} \frac{2}{5} & \frac{1}{5} \\ \frac{1}{5} & -\frac{2}{5} \end{pmatrix}.$$

Hence,

$$C = A^{-1}B = \begin{pmatrix} \frac{2}{5} & \frac{1}{5} \\ \frac{1}{5} & -\frac{2}{5} \end{pmatrix}\begin{pmatrix} 1 & -1 \\ 2 & 1 \end{pmatrix} = \begin{pmatrix} \frac{4}{5} & -\frac{1}{5} \\ -\frac{3}{5} & -\frac{3}{5} \end{pmatrix}.$$

By **Theorem 7.4.3** , we have

$$(\vec{b})_S = C(\vec{b})_T = \begin{pmatrix} \frac{4}{5} & -\frac{1}{5} \\ -\frac{3}{5} & -\frac{3}{5} \end{pmatrix}\begin{pmatrix} 5 \\ -10 \end{pmatrix} = \begin{pmatrix} 6 \\ 3 \end{pmatrix}.$$

Let $S = \left\{ \vec{a}_1,\ \vec{a}_2,\ \cdots,\ \vec{a}_n \right\}$ be a basis of $\mathbb{R}^n$. In general, these vectors may not be orthogonal. In some cases, we are interested in those bases whose basis vectors *are* orthogonal.

Definition 7.4.3.

Let $S = \left\{ \vec{a}_1,\ \vec{a}_2,\ \cdots,\ \vec{a}_n \right\}$ be a basis of $\mathbb{R}^n$.

i. If *S* satisfies the following condition:

$$\vec{a}_i \perp \vec{a}_j \quad \text{for } i \neq j \text{ and } i,\ j \in \{1,\ 2,\ \cdots,\ n\},$$

then *S* is called an orthogonal basis of $\mathbb{R}^n$.

ii. If *S* is an orthogonal basis and satisfies

$$\left\| \vec{a}_i \right\| = 1 \quad \text{for } i = 1,\ 2,\ \cdots,\ n,$$

then *S* is called an orthonormal basis.

Example 7.4.4.

Determine which of the following bases are orthogonal bases and orthonormal bases.

$$S_1 = \left\{ \begin{pmatrix} -1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \end{pmatrix} \right\} \quad S_2 = \left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \end{pmatrix} \right\}$$

$$S_3 = \left\{ \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}, \begin{pmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{pmatrix} \right\} \; S_4 = \left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\}$$

Solution

S_1 is not an orthogonal basis because $(-1,\ 1) \cdot \begin{pmatrix} 1 \\ -1 \end{pmatrix} = -2 \neq 0.$

S_2 is an orthogonal basis because $(1,\ 1) \cdot \begin{pmatrix} 1 \\ -1 \end{pmatrix} = 0,$ but S_2 is not an orthonormal basis because $\left|\left|(1,\ 1)\right|\right| = \sqrt{2} \neq 1.$

S_3 is an orthonormal basis because $\left(\frac{1}{\sqrt{2}},\ \frac{1}{\sqrt{2}}\right) \cdot \begin{pmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{pmatrix} = 0$ and

$$\left\| \left(\frac{1}{\sqrt{2}},\ \frac{1}{\sqrt{2}}\right) \right\| = 1 \quad \text{and} \left\| \left(\frac{1}{\sqrt{2}},\ -\frac{1}{\sqrt{2}}\right) \right\| = 1.$$

S_4 is an orthonormal basis. In fact, because it is easy to verify that $\vec{e_1} \cdot \vec{e_2} = 0,\ \vec{e_3} = 0$ and $\vec{e_2} \cdot \vec{e_3} = 0.$

Hence, *S* is an orthogonal basis of $\mathbb{R}^3$. Because $\left|\left|\vec{e_1}\right|\right| = 1,\ \left|\left|\vec{e_2}\right|\right| = 1$ and $\left|\left|\vec{e_3}\right|\right| = 1,$ *S* is an orthonormal basis of $\mathbb{R}^3$.

In **Example 7.4.4** , the basis S_4 is the standard basis of $\mathbb{R}^3$. In fact, the standard basis *E* given in **(7.2.1)** is an orthonormal basis of $\mathbb{R}^n$.

The coordinates of a vector $\vec{b}$ in $\mathbb{R}^n$ relative to an orthogonal basis or an orthonormal basis can be obtained by using the methods given in Section 8.4. However, because the basis vectors in an orthogonal basis or an orthonormal basis are orthogonal, we can provide formulas for their coordinates.

Theorem 7.4.4.

i. *If* $S = \left\{\vec{a_1}, \vec{a_2}, \cdots, \vec{a_n}\right\}$ *is an orthogonal basis of* $\mathbb{R}^n$, *then*

(7.4.10)

$$(\vec{b})_S = \left(\frac{\vec{b} \cdot \vec{a_1}}{\left|\left|\vec{a_1}\right|\right|^2}, \frac{\vec{b} \cdot \vec{a_2}}{\left|\left|\vec{a_2}\right|\right|^2}, \cdots, \frac{\vec{b} \cdot \vec{a_n}}{\left|\left|\vec{a_n}\right|\right|^2}\right).$$

ii. *If* $S = \left\{\vec{a_1}, \vec{a_2}, \cdots, \vec{a_n}\right\}$ *is an orthonormal basis of* $\mathbb{R}^n$, *then*

0%

(7.4.11)

$$(\vec{b})_S = (\vec{b} \cdot \overrightarrow{a_1},\ \vec{b} \cdot \overrightarrow{a_2},\ \cdots,\ \vec{b} \cdot \overrightarrow{a_n}).$$

Proof

Because $S = \left\{\overrightarrow{a_1},\ \overrightarrow{a_2},\ \cdots,\ \overrightarrow{a_n}\right\}$ is a basis of $\mathbb{R}^n$, by **Theorem 7.4.1** , for each $\vec{b} \in \mathbb{R}^n$, there exists a unique solution $(x_1,\ x_2,\ \cdots,\ x_n)$ such that

$$\vec{b} = x_1\overrightarrow{a_1} + x_2\overrightarrow{a_2} + \cdots + x_i\overrightarrow{a_i} + \cdots + x_n\overrightarrow{a_n}.$$

This implies that for each $i \in I_n$,

(7.4.12)

$$(\vec{b} \cdot \overrightarrow{a_i}) = x_1(\overrightarrow{a_1} \cdot \overrightarrow{a_i}) + x_2(\overrightarrow{a_2} \cdot \overrightarrow{a_i}) + \cdots + x_i(\overrightarrow{a_i} \cdot \overrightarrow{a_i}) + \cdots + x_n(\overrightarrow{a_n} \cdot \overrightarrow{a_i}).$$

i. Because *S* is an orthogonal basis of $\mathbb{R}^n$,

$$\overrightarrow{a_j} \cdot \overrightarrow{a_i} = 0 \quad \text{for } j \neq i \text{ and } i,\ j \in I_n.$$

This, together with **(7.4.12)** , implies that for $i \in I_n$,

$$\vec{b} \cdot \overrightarrow{a_i} = x_i(\overrightarrow{a_i} \cdot \overrightarrow{a_j}) = x_i \left|\left|\overrightarrow{a_i}\right|\right|^2$$

and

$$x_i = \frac{\vec{b} \cdot \vec{a_i}}{\left|\left|\vec{a_i}\right|\right|^2} \quad \text{for } i \in I_n.$$

Hence, **(7.4.10)** holds.

ii. If S is an orthonormal basis, then $\left|\left|\vec{a_i}\right|\right| = 1$ for $i \in I_n$. This, together with **(7.4.10)**, implies that **(7.4.11)** holds.

Example 7.4.5.

Let $\vec{b} = (0, 1)^T$. Find $(\vec{b})_S$ and $(\vec{b})_T$, where

$$S = \left\{\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \end{pmatrix}\right\} \text{ and } T = \left\{\begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}, \begin{pmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{pmatrix}\right\}.$$

Solution

By **Example 7.4.4**, S is an orthogonal basis but not an orthonormal basis. Let $\vec{a_1} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ and $\vec{a_2} = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$. Then

$$\vec{b} \cdot \overrightarrow{a_1} = (0,\ 1)\begin{pmatrix} 1 \\ 1 \end{pmatrix} = 1,\quad \left|\left|\overrightarrow{a_1}\right|\right|^2 = 2.$$

and

$$\vec{b} \cdot \overrightarrow{a_2} = (0,\ 1)\begin{pmatrix} -1 \\ 1 \end{pmatrix} = 1,\quad \left|\left|\overrightarrow{a_2}\right|\right|^2 = 2.$$

By **(7.4.10)**, we obtain $(\vec{b})_S = \left(\frac{1}{2},\ \frac{1}{2}\right)$.

By **Example 7.4.4**, T is an orthonormal basis. Let

$$\overrightarrow{a_1} = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} \text{ and } \overrightarrow{a_2} = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{pmatrix}.$$

Then

$$\vec{b} \cdot \overrightarrow{a_1} = (0,\ 1)\begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} = \frac{1}{\sqrt{2}} \text{ and } \vec{b} \cdot \overrightarrow{a_2} = (0,\ 1)\begin{pmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{pmatrix} = -\frac{1}{\sqrt{2}}.$$

By **(7.4.11)** , $(\vec{b})_S = \left(\frac{1}{2},\ -\frac{1}{2}\right)$.

The following result is the well-known Gram-Schmidt process, which shows that all bases of $\mathbb{R}^n$ can be changed to orthogonal bases and orthonormal bases of $\mathbb{R}^n$. We omit its proof.

Theorem 7.4.5 (Gram-Schmidt process)

Let $S = \left\{\vec{a}_1,\ \vec{a}_2,\ \cdots,\ \vec{a}_n\right\}$ *be a basis of* $\mathbb{R}^n$. *Let*

$$\vec{b}_1 = \vec{a}_1,$$

$$\vec{b}_2 = \vec{a}_2 - \frac{(\vec{a}_2 \cdot \vec{b}_1)}{\left\|\vec{b}_1\right\|^2}\vec{b}_1,$$

$$\vec{b}_3 = \vec{a}_3 - \frac{(\vec{a}_3 \cdot \vec{b}_1)}{\left\|\vec{b}_1\right\|^2}\vec{b}_1 - \frac{(\vec{a}_3 \cdot \vec{b}_1)}{\left\|\vec{b}_2\right\|^2}\vec{b}_2,$$

$$\vdots$$

$$\vec{b}_n = \vec{a}_n - \frac{(\vec{a}_n \cdot \vec{b}_1)}{\left\|\vec{b}_1\right\|^2}\vec{b}_1 - \frac{(\vec{a}_n \cdot \vec{b}_2)}{\left\|\vec{b}_2\right\|^2}\vec{b}_2 - \cdots - \frac{(\vec{b}_n \cdot \vec{b}_{n-1})}{\left\|\vec{b}_{n-1}\right\|^2}\vec{b}_{n-1}.$$

Then

i. $\left\{\vec{b}_1, \vec{b}_2, \cdots, \vec{b}_n\right\}$ *is an orthogonal basis of* $\mathbb{R}^n$.

ii. $\left\{\dfrac{\vec{b}_1}{\|\vec{b}_1\|}, \dfrac{\vec{b}_2}{\|\vec{b}_2\|}, \cdots, \dfrac{\vec{b}_n}{\|\vec{b}_n\|}\right\}$ *is an orthonormal basis of* $\mathbb{R}^n$.

Example 7.4.6.

Let $\vec{a}_1 = (1, 0, 1)$, $\vec{a}_2 = (0, 1, 1)$, and $\vec{a}_3 = (1, 1, 0)$ be the same as in **Example 7.4.2** .

1. Find an orthogonal basis $S_1 = \left\{\vec{b}_1, \vec{b}_2, \vec{b}_3\right\}$ by using the Gram-Schmidt process. Moreover, if $\vec{b} = (1, 1, 1)^T$, find $(\vec{b})_{S_1}$.
2. Find an orthonormal basis $S_2 := \left\{\vec{q}_1, \vec{q}_2, \vec{q}_3\right\}$ by normalizing S_1. Moreover, if $\vec{b} = (1, 1, 1)^T$, find $(\vec{b})_{S_2}$.

Solution

By **Example 7.4.2** (*i*), *S* is a basis of $\mathbb{R}^3$. We use **Theorem 7.4.5** to find S_1 and S_2.

1. Let $\vec{b}_1 = \vec{a}_1 = (1, 0, 1)$ and let

$$\vec{b}_2 = \vec{a}_2 - \frac{(\vec{a}_2 \cdot \vec{b}_1)}{\left\|\vec{b}_1\right\|^2}\vec{b}_1.$$

To find $\vec{b}_2$, we need to compute $\vec{a}_2 \cdot \vec{b}_1$ and $\left\|\vec{b}_1\right\|^2$. Because $\vec{a}_2 \cdot \vec{b}_1 = (0,\ 1,\ 1) \cdot (1,\ 0,\ 1) = 0$ and $\left\|\vec{b}_1\right\|^2 = 2$, we have

$$\vec{b}_2 = (0,\ 1,\ 1) - \frac{1}{2}(1,\ 0,\ 1) = (0,\ 1,\ 1) - \left(\frac{1}{2},\ 0,\ \frac{1}{2}\right) = \left(-\frac{1}{2},\ 1,\ \frac{1}{2}\right).$$

Let

$$\vec{b}_3 = \vec{a}_3 - \frac{(\vec{a}_3 \cdot \vec{b}_1)}{\left\|\vec{b}_1\right\|^2}\vec{b}_1 - \frac{(\vec{a}_3 \cdot \vec{b}_2)}{\left\|\vec{b}_2\right\|^2}\vec{b}_2.$$

To find $\vec{b}_3$, we need to compute $\vec{a}_3 \cdot \vec{b}_1$, $\vec{a}_3 \cdot \vec{b}_2$ and $\left\|\vec{b}_2\right\|^2$. Because $\vec{a}_3 \cdot \vec{b}_1 = 1$, $\vec{a}_3 \cdot \vec{b}_2 = \frac{1}{2}$, and $\left\|\vec{b}_2\right\|^2 = \frac{3}{2}$, we obtain

$$\overrightarrow{b_3} = (1,\ 1,\ 0) - \frac{1}{2}(1,\ 0,\ 1) - \frac{\frac{1}{2}}{\frac{3}{2}}\left(-\frac{1}{2},\ 1,\ \frac{1}{2}\right) = \left(\frac{2}{3},\ \frac{2}{3},\ -\frac{2}{3}\right).$$

Hence, $\overrightarrow{b_1} = (1,\ 0,\ 1)$, $\overrightarrow{b_2} = \left(-\frac{1}{2},\ 1,\ \frac{1}{2}\right)$, $\overrightarrow{b_3} = \left(\frac{2}{3},\ \frac{2}{3},\ -\frac{2}{3}\right)$ form an orthogonal basis of $\mathbb{R}^3$.

By computation, $\left|\left|\overrightarrow{b_1}\right|\right| = \sqrt{2}$, $\left|\left|\overrightarrow{b_2}\right|\right| = \frac{\sqrt{6}}{2}$ and $\left|\left|\overrightarrow{b_3}\right|\right| = \frac{2\sqrt{3}}{3}$, $(\vec{b} \cdot \overrightarrow{b_1}) = 2$, $\vec{b} \cdot \overrightarrow{b_2} = 1$, and $\vec{b} \cdot \overrightarrow{b_3} = \frac{2}{3}$. By **Theorem 7.4.4** ,

$$(\vec{b})_{S_1} = \left(\frac{\vec{b} \cdot \overrightarrow{b_1}}{\left|\left|\overrightarrow{b_1}\right|\right|^2},\ \frac{\vec{b} \cdot \overrightarrow{b_2}}{\left|\left|\overrightarrow{b_2}\right|\right|^2},\ \frac{\vec{b} \cdot \overrightarrow{b_3}}{\left|\left|\overrightarrow{b_3}\right|\right|^2}\right) = \left(\frac{2}{2},\ \frac{\frac{2}{3}}{\frac{4}{3}},\ \frac{\frac{2}{3}}{\frac{4}{3}}\right) = \left(1,\ \frac{2}{3},\ \frac{1}{2}\right).$$

2. By computation, we obtain

$$\vec{q_1} := \frac{\vec{b}_1}{\left\|\vec{b}_1\right\|} = \frac{\sqrt{2}}{2}(1,\ 0,\ 1) = \left(\frac{\sqrt{2}}{2},\ 0,\ \frac{\sqrt{2}}{2}\right),$$

$$\vec{q_2} := \frac{\vec{b}_2}{\left\|\vec{b}_2\right\|} = \frac{\sqrt{6}}{3}\left(-\frac{1}{2},\ 1,\ \frac{1}{2}\right) = \left(-\frac{\sqrt{6}}{6},\ \frac{\sqrt{6}}{3},\ \frac{\sqrt{6}}{6}\right),$$

$$\vec{q_3} := \frac{\vec{b}_3}{\left\|\vec{b}_3\right\|} = \frac{\sqrt{3}}{2}\left(\frac{2}{3},\ \frac{2}{3},\ -\frac{2}{3}\right) = \left(\frac{\sqrt{3}}{3},\ \frac{\sqrt{3}}{3},\ -\frac{\sqrt{3}}{3}\right).$$

Hence, $S_2 = \left\{\vec{q_1},\ \vec{q_2},\ \vec{q_3}\right\}$ forms an orthonormal basis of $\mathbb{R}^3$ and

$$(\vec{b})_{S_2} = (\vec{b}\cdot\vec{q_1},\ \vec{b}\cdot\vec{q_2},\ \vec{b}\cdot\vec{q_3}) = \left(\sqrt{2},\ \frac{\sqrt{6}}{3},\ \frac{\sqrt{3}}{3}\right).$$

Exercises

1. For each of the following pairs of vectors, verify that it is a basis of $\mathbb{R}^2$ and find the coordinates of the vector $\vec{b} = (x,\ y)^T$.

$$S_1 = \left\{ \begin{pmatrix} -1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right\} \quad S_2 = \left\{ \begin{pmatrix} -3 \\ 2 \end{pmatrix}, \begin{pmatrix} 2 \\ 3 \end{pmatrix} \right\}$$

$$S_3 = \left\{ \begin{pmatrix} 5 \\ -6 \end{pmatrix}, \begin{pmatrix} 4 \\ -1 \end{pmatrix} \right\} \quad S_4 = \left\{ \begin{pmatrix} 3 \\ -2 \end{pmatrix}, \begin{pmatrix} -2 \\ 5 \end{pmatrix} \right\}$$

2. Let $\overrightarrow{a_1} = (1,\ 0,\ 1)^T,\ \overrightarrow{a_2} = (0,\ 1,\ 1)^T,$ and $\overrightarrow{a_3} = (1,\ 1,\ 0)^T.$

   i. Show that $S = \left\{ \overrightarrow{a_1},\ \overrightarrow{a_2},\ \overrightarrow{a_3} \right\}$ is a basis of $\mathbb{R}^3$.

   ii. Let $\vec{b} = (1,\ 1,\ 1)^T.$ Find $(\vec{b})_S$.

3. Let $\overrightarrow{a_1} = (1,\ 0,\ 0)^T,\ \overrightarrow{a_2} = (2,\ 4,\ 0)^T,$ and $\overrightarrow{a_3} = (3,\ -1,\ 5)^T.$

   i. Show that $S = \left\{ \overrightarrow{a_1},\ \overrightarrow{a_2},\ \overrightarrow{a_3} \right\}$ is a basis of $\mathbb{R}^3$.

   ii. Let $\vec{b} = (2,\ -1,\ -2).$ Find $(\vec{b})_S$.

4. Let $S_1 = \left\{ \begin{pmatrix} -1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right\}$ and $S_2 = \left\{ \begin{pmatrix} -3 \\ 2 \end{pmatrix}, \begin{pmatrix} 2 \\ 3 \end{pmatrix} \right\}$.

   i. If $(\vec{b})_{S_2} = (2,\ -1),$ find $(\vec{b})_{S_1}$.

   ii. If $(\vec{b})_{S_1} = (13,\ -26),$ find $(\vec{b})_{S_2}$.

5. If $(\vec{b})_T = (1,\ 2,\ 3),$ find $(\vec{b})_S,$ where

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

$$T=\left\{\begin{pmatrix}1\\0\\0\end{pmatrix},\begin{pmatrix}0\\1\\-1\end{pmatrix},\begin{pmatrix}1\\1\\1\end{pmatrix}\right\}.$$

6. Let $\overrightarrow{b_1}=(0,\ 1,\ 0),\ \overrightarrow{b_2}=\left(-\frac{4}{5},\ 0,\ \frac{3}{5}\right)$, and $\overrightarrow{b_3}=\left(\frac{3}{5},\ 0,\ \frac{4}{5}\right)$.

i. Show that $S:=\left\{\overrightarrow{b_1},\ \overrightarrow{b_2},\ \overrightarrow{b_3}\right\}$ is an orthonormal basis of $\mathbb{R}^3$.

ii. Let $\vec{b}=(1,\ -1,\ 1)$. Find $(\vec{b})_S$.

7. Let $\overrightarrow{a_1}=(1,\ 1,\ 1),\ \overrightarrow{a_2}=(0,\ 1,\ 1),\ \overrightarrow{a_3}=(0,\ 0,\ 1)$, and $\vec{b}=(-1,\ 1,\ -1)^T$.

1. Find an orthogonal basis $S_1:=\left\{\overrightarrow{b_1},\ \overrightarrow{b_2},\ \overrightarrow{b_3}\right\}$ of $\mathbb{R}^3$ by using the Gram-Schmidt process and $(\vec{b})_{S_1}$.

2. Find an orthonormal basis $S_2:=\left\{\overrightarrow{q_1},\ \overrightarrow{q_2},\ \overrightarrow{q_3}\right\}$ of $\mathbb{R}^3$ by normalizing S_1 and $(\vec{b})_{S_2}$.

Chapter **8** Eigenvalues and Diagonalizability

8.1 Eigenvalues and eigenvectors

Let *A* be an $n \times n$ matrix given in **(3.2.1)**, that is,

(8.1.1)

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n_1} & a_{n_2} & \cdots & a_{nn} \end{pmatrix}.$$

Definition 8.1.1.

A real number λ is called an eigenvalue of *A* if there exists a nonzero vector $\overrightarrow{X}$ in $\mathbb{R}^n$ such that

(8.1.2)

$$A\overrightarrow{X} = \lambda\overrightarrow{X}.$$

The vector $\overrightarrow{X}$ is said to be an eigenvector of the matrix *A* corresponding to the eigenvalue λ.

Note that we restrict our study on eigenvalues in $\mathbb{R}$, the set of real numbers. In fact, eigenvalues could be complex numbers, which we shall study in **Chapter 10**. Therefore, eigenvalues in the field of complex numbers can be further studied after **Chapter 10**.

An eigenvalue could be positive, zero, or negative. But, according to **Definition 8.1.1**, an eigenvector must be a nonzero vector. In other words, the zero vector is not an eigenvector although it is a solution of **(8.1.2)**.

From **(8.1.2)** and **Theorems 4.5.2** and **4.5.3**, we see that *A* has a zero eigenvalue if and only if the homogeneous system $A\overrightarrow{X} = 0$ has a nonzero solution if and only if $r(A) < n$ if and only if $|A| = 0$ if and only if *A* is not invertible. Hence, *A* is invertible if and only if all the eigenvalues of *A* are nonzero, see **Corollary 8.1.3** for another proof.

By **(8.1.2)**, we see that $\lambda \in \mathbb{R}$ is an eigenvalue of *A* if and only if

(8.1.3)

$$(\lambda I - A)\overrightarrow{X} = \overrightarrow{0}$$

has a nonzero solution in $\mathbb{R}^n$.

By **Theorem 4.5.3**, we obtain the following result, which shows that we can use determinants to find the eigenvalues.

Theorem 8.1.1.

$\lambda \in \mathbb{R}$ *is an eigenvalue of A if and only if*

(8.1.4)

$$p(\lambda) := |\lambda I - A| = 0.$$

Note that the determinant

$$p(\lambda) = \begin{vmatrix} \lambda - a_{11} & -a_{12} & \cdots & -a_{1n} \\ -a_{21} & \lambda - a_{22} & \cdots & -a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ -a_{n_1} & -a_{n_2} & \cdots & \lambda - a_{nn} \end{vmatrix}$$

is a polynomial of degree *n*, which is called the characteristic polynomial of *A*. The equation $p(\lambda) = 0$ is called the *characteristic equation* of *A*.

The following result can be proved by using the fundamental theorem of algebra from complex analysis.

Lemma 8.1.1.

Any polynomial

$$p(z) = a_0 + a_1 z + a_2 z^2 + \cdots + a_n z^n \quad (a_n \neq 0)$$

of degree $(n \geq 1)$ *can be expressed as a product of the form*

$$p(z) = c(z - z_1)(z - z_2) \cdots (z - z_n),$$

where c and $z_j (j \in I_n)$ *are complex numbers (see* ***Chapter 10*** *for the study of complex numbers).*

Note that some of these complex numbers $z_1,\ z_2,\ \cdots,\ z_n$ may be the same. By **Lemma 8.1.1** , $p(\lambda)$ can be rewritten as follows.

(8.1.5)

$$p(\lambda) = (\lambda - \lambda_1)^{r_1} (\lambda - \lambda_2)^{r_2} \cdots (\lambda - \lambda_k)^{r_k},$$

where λ_i is a real or complex number, $\lambda_i \neq \lambda_j$ for $i \neq j$, and r_i is a positive integer for $i,\ j \in \{1,\ \cdots,\ k\}$ satisfying

(8.1.6)

$$r_1 + r_2 + \cdots + r_k = n.$$

Hence, the characteristic equation $p(\lambda) = 0$ has *n* (real or complex) roots including repeated roots, which are those roots λ_i with $r_i > 1$.

Definition 8.1.2.

The number r_i is called the *algebraic multiplicity* of the eigenvalue λ_i for each $i = 1,\ 2,\ \cdots,\ k$.

In **Definition 8.1.1**, if we allow the eigenvalues to be complex numbers, then the eigenvectors would take complex numbers as their components, so complex vector space $\mathbb{C}^n$, complex matrices, complex determinants, etc., are involved. But from now on, we restrict our discussion of eigenvalues and eigenvectors to real numbers. Hence, we shall only consider matrices *A* such that all of $\lambda_1,\ \cdots,\ \lambda_k$ in **(8.1.5)** are real. But we give one question involving complex eigenvalues in question 8 of **Exercise 9.2**.

For each λ_i, let

(8.1.7)

$$E_{\lambda_i} := N(\lambda_i I - A) = \left\{ \overrightarrow{X} \in \mathbb{R}^n : (\lambda_i I - A)\overrightarrow{X} = \overrightarrow{0} \right\}.$$

The set E_{λ_i} is called the eigenspace of *A*. It is obvious that the eigenspace E_{λ_i} of *A* is exactly the solution space of $(\lambda_i I - A)\overrightarrow{X} = \overrightarrow{0}$ or the nullspace of $\lambda_i I - A$, see **(4.1.13)**. Hence, the eigenspace E_{λ_i} contains all eigenvectors corresponding to the eigenvalue λ_i as well as the zero vector in $\mathbb{R}^n$, which is not an eigenvector (as mentioned above).

Definition 8.1.3.

The dimension $\dim(E_{\lambda_i})$ of the eigenspace E_{λ_i} is called the *geometric multiplicity* of the eigenvalue λ_i for each $i = 1,\ 2,\ \cdots,\ k$.

By **Definition 8.1.1**, the eigenspace E_{λ_i} contains a nonzero vector, so the geometric multiplicity is greater than or equal to 1.

We emphasize that there are two numbers, one being the algebraic multiplicity r_i given in **(8.1.5)** and another the geometric multiplicity $\dim(E_{\lambda_i})$, corresponding to each eigenvalue λ_i.

The following result shows that the geometric multiplicity is less than or equal to the algebraic multiplicity. The proof is omitted because it involves additional properties of determinants.

Theorem 8.1.2.

Let λ_i *and* r_i *be the same as in* **(8.1.5)** *for* $i \in \{1,\ 2,\ \cdots,\ k\}$. *Then*

$$1 \leq \dim(E_{\lambda_i}) \leq r_i.$$

As a special case of **Theorem 8.1.2**, we obtain the following result, which shows that if the algebraic multiplicity of an eigenvalue is 1, then its geometric multiplicity is 1.

Corollary 8.1.1.

Let λ_i *and* r_i *bethesameasin* **(8.1.5)** *for* $i \in \{1,\ 2,\ \cdots,\ k\}$, *Then if* $r_j = 1$ *for some* $j \in \{1,\ 2,\ \cdots,\ k\}$, *then*

$$\dim(E_{\lambda_i}) = r_j = 1.$$

By **Theorem 8.1.2** and **(8.1.6)**, we obtain the following result, which shows that the sum of the geometric multiplicities is less than or equal to *n*.

Corollary 8.1.2.

Let λ_i *and* r_i *bethesameasin* **(8.1.5)** *. Then*

$$k \le \sum_{i=1}^{k} \dim(E_{\lambda_i}) \le n.$$

The following result gives the eigenvalues of triangular (upper, lower, or diagonal) matrices.

Theorem 8.1.3.

Let A be an $n \times n$ *triangular (upper, lower, or diagonal) matrix. Then the eigenvalues of A are all the entries on the main diagonal of A.*

Proof

We prove this only when *A* is an upper triangular matrix. Let

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ 0 & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{nn} \end{pmatrix}$$

be an upper triangular matrix. Then

$$\begin{aligned} |\lambda I - A| &= \begin{vmatrix} \lambda - a_{11} & -a_{12} & \cdots & -a_{1n} \\ 0 & \lambda - a_{22} & \cdots & -a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda - a_{nn} \end{vmatrix} \\ &= (\lambda - a_{11})(\lambda - a_{22}) \cdots (\lambda - a_{nn}). \end{aligned}$$

Solving $|\lambda I - A| = 0$ implies $\lambda_1 = a_{11},\ \lambda_2 = a_{22},\ \cdots,\ \lambda_n = a_{nn}$.

Example 8.1.1.

For each of the following matrices, find its eigenvalues and determine its algebraic multiplicity.

$$A = \begin{pmatrix} -1 & -1 & 0 \\ 0 & 2 & 3 \\ 0 & 0 & 3 \end{pmatrix}, \quad B = \begin{pmatrix} 4 & -1 & 0 \\ 0 & 4 & 3 \\ 0 & 0 & -3 \end{pmatrix}.$$

Solution

By **Theorem 8.1.3** , the eigenvalues of *A* are $\lambda_1 = -1,\ \lambda_2 = 2,$ and $\lambda_3 = 3.$ The algebraic multiplicity of each of $\lambda_1,\ \lambda_2,\ \lambda_3$ is 1 because

$$|\lambda I - A| = (\lambda - \lambda_1)(\lambda - \lambda_2)(\lambda - \lambda_3).$$

All the eigenvalues of *B* are $\lambda_1 = \lambda_2 = 4$ and $\lambda_3 = -3.$ The algebraic multiplicity for the eigenvalue 4 is 2 and the algebraic multiplicity for the eigenvalue 3 is 1 because $\left|\lambda I - A\right| = (\lambda - 4)^2(\lambda + 3)$.

The procedure for finding eigenvalues and eigenvectors is given below.

1. Step 1. Find $p(\lambda) = |\lambda I - A|$.
2. Step 2. Find all roots $\lambda_1,\ \lambda_2,\ \cdots,\ \lambda_n$ of $p(\lambda) = 0.$ Some of these eigenvalues may be the same.
3. Step 3. For each eigenvalue λ_i, solve the corresponding homogenous system

$$(\lambda_i I - A)\overrightarrow{X} = \overrightarrow{0}$$

   and find the basis vectors by using the methods in **Section 4.4** .
4. Step 4. Give the eigenspace E_{λ_i} using the basis vectors.

Example 8.1.2.

Let

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

1. Find the eigenvalues and eigenspaces of *A*.
2. Find the algebraic and geometric multiplicities for each eigenvalue.

Solution

1. Because $|\lambda I - A| = \begin{vmatrix} \lambda - 2 & 0 \\ 0 & \lambda - 2 \end{vmatrix} = (\lambda - 2)^2 = 0,$ we obtain a repeated eigenvalue $\lambda_1 = \lambda_2 = 2.$

   For $\lambda_1 = \lambda_2 = 2,\ (\lambda_1 I - A)\overrightarrow{X} = \overrightarrow{0}$ becomes

   $$\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

   The corresponding system is $0x_1 + 0x_2 = 0,$ where x_1 and x_2 are free variables. Let $x_1 = t$ and $x_2 = s$. Then

   $$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} t \\ s \end{pmatrix} = t\begin{pmatrix} 1 \\ 0 \end{pmatrix} + s\begin{pmatrix} 0 \\ 1 \end{pmatrix} = t\overrightarrow{v_1} + s\overrightarrow{v_2},$$

   where $\overrightarrow{v_1} = (1,\ 0)^T$ and $\overrightarrow{v_2} = (0,\ 1)^T$ is a solution of $0x_1 + 0x_2 = 0.$ Moreover,

   $$E_{\lambda_1} = E_{\lambda_2} = \operatorname{span}\left\{\overrightarrow{v_1},\ \overrightarrow{v_2}\right\}.$$

2. The algebraic multiplicity for the eigenvalue $\lambda_1 = \lambda_2 = 2$ is 2 because 2 is the power of the term $\lambda - 2$ in the characteristic polynomial $\left|\lambda I - A\right| = (\lambda - 2)^2$. Because the eigenspace E_{λ_1} has two basis vectors $\overrightarrow{v_1}$ and $\overrightarrow{v_2}$, $\dim(E_{\lambda_1}) = 2$. Hence, the geometric multiplicity of the eigenvalue 2 is 2.

Example 8.1.3.

Let $A = \begin{pmatrix} 0 & 0 & -1 \\ 1 & -2 & 1 \\ 0 & 0 & 3 \end{pmatrix}$. Find the eigenvalues and eigenspaces of *A*.

Solution

$$p(\lambda) = |\lambda I - A| = \begin{vmatrix} \lambda & 0 & 1 \\ -1 & \lambda + 2 & -1 \\ 0 & 0 & \lambda - 3 \end{vmatrix} = \lambda(\lambda + 2)(\lambda - 3).$$

Solving $p(\lambda) = \lambda(\lambda + 2)(\lambda - 3) = 0$ gets $\lambda_1 = 0,\ \lambda_2 = -2,$ and $\lambda_3 = 3.$

For $\lambda_1 = 0,$

$$\lambda_1 I - A = \begin{pmatrix} 0 & 0 & 1 \\ -1 & 2 & -1 \\ 0 & 0 & -3 \end{pmatrix}.$$

We solve the following system.

$$\begin{pmatrix} 0 & 0 & 1 \\ -1 & 2 & -1 \\ 0 & 0 & -3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

$$(A|\overrightarrow{0}) = \left(\begin{array}{ccc|c} 0 & 0 & 1 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & 0 & -3 & 0 \end{array}\right) \xrightarrow{R_1,\, 2} \left(\begin{array}{ccc|c} -1 & 2 & -1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & -3 & 0 \end{array}\right)$$

$$\xrightarrow[R_2(1)+R_1]{R_2(3)+R_3} \left(\begin{array}{ccc|c} -1 & 2 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right).$$

The system corresponding to the last augmented matrix is

$$\begin{cases} -x_1 + 2x_2 = 0 \\ \qquad\qquad x_3 = 0, \end{cases}$$

where x_1 and x_3 are basic variables and x_2 is a free variable. Let $x_2 = t$. Then $x_2 = 2t$. Let $\overrightarrow{v_1} = (2,\ 1,\ 0)^T$. Then $E_{\lambda_1} = \text{span}\left\{\overrightarrow{v_1}\right\}$.

For $\lambda_2 = -2$,

$$\lambda_2 I - A = \begin{pmatrix} -2 & 0 & 1 \\ -1 & 0 & -1 \\ 0 & 0 & -5 \end{pmatrix}.$$

We solve the following system.

$$\begin{pmatrix} -2 & 0 & 1 \\ -1 & 0 & -1 \\ 0 & 0 & -5 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

$$(A|\overrightarrow{0}) = \left(\begin{array}{ccc|c} -2 & 0 & 1 & 0 \\ -1 & 0 & -1 & 0 \\ 0 & 0 & -5 & 0 \end{array}\right) \xrightarrow{R_{1,\,2}} \left(\begin{array}{ccc|c} -1 & 0 & -1 & 0 \\ -2 & 0 & 1 & 0 \\ 0 & 0 & -5 & 0 \end{array}\right)$$

$$\xrightarrow{R_1(-2)+R_2} \left(\begin{array}{ccc|c} -1 & 0 & -1 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & -5 & 0 \end{array}\right) \xrightarrow{R_1\left(\frac{1}{3}\right)} \left(\begin{array}{ccc|c} -1 & 0 & -1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & -5 & 0 \end{array}\right)$$

$$\xrightarrow{R_2(5)+R_3} \left(\begin{array}{ccc|c} -1 & 0 & -1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right) \xrightarrow{R_2(1)+R_1} \left(\begin{array}{ccc|c} -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right).$$

The system corresponding to the last augmented matrix is

$$\begin{cases} x_1 = 0 \\ x_3 = 0, \end{cases}$$

where x_1 and x_3 are basic variables and x_2 is a free variable. Let $x_2 = t$ and $\overrightarrow{v_2} = (0,\ 1,\ 0)^T$. Then $E_{\lambda_2} = \mathrm{span}\left\{\overrightarrow{v_2}\right\}$.

For $\lambda_3 = 3$,

$$\lambda_3 I - A = \begin{pmatrix} 3 & 0 & 1 \\ -1 & 5 & -1 \\ 0 & 0 & 0 \end{pmatrix}.$$

We solve the following system.

$$\begin{pmatrix} 3 & 0 & 1 \\ -1 & 5 & -1 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

$$(A|\overrightarrow{0}) = \left(\begin{array}{ccc|c} 3 & 0 & 1 & 0 \\ -1 & 5 & -1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right) \xrightarrow{R_{1,\,2}} \left(\begin{array}{ccc|c} -1 & 5 & -1 & 0 \\ 3 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right)$$

$$\xrightarrow{R_1(3)+R_2} \left(\begin{array}{ccc|c} -1 & 5 & -1 & 0 \\ 0 & 15 & -2 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right) \xrightarrow{R_2\left(\frac{1}{15}\right)}$$

$$\left(\begin{array}{ccc|c} -1 & 5 & -1 & 0 \\ 0 & 1 & -\frac{2}{15} & 0 \\ 0 & 0 & 0 & 0 \end{array}\right) \xrightarrow{R_2(-5)+R_1} \left(\begin{array}{ccc|c} -1 & 0 & -\frac{1}{3} & 0 \\ 0 & 1 & -\frac{2}{15} & 0 \\ 0 & 0 & 0 & 0 \end{array}\right).$$

The system corresponding to the last augmented matrix is

$$\begin{cases} -x_1 - \frac{1}{3}x_3 = 0 \\ x_2 - \frac{2}{15}x_3 = 0, \end{cases}$$

where x_1 and x_2 are basic variables, and x_3 is a free variable. Let $x_3 = 15t$ and $\overrightarrow{v_3} = (-5,\ 2,\ 15)^T$. Then $E_{\lambda_3} = \text{span}\left\{\overrightarrow{v_3}\right\}$.

We shall give more examples of finding eigenvalues and eigenspaces in the next section.

Now, we turn our attention to the relations between *A* and its eigenvalues. Recall that the trace of *A* is given by

$$\text{tr}(A) = a_{11} + \cdots + a_{nn},$$

see **(2.3.2)** in **Section 2.2** .

The matrix *A* and its eigenvalues have the following relationships, which show that the sum of all the eigenvalues equals the trace of *A* and the product of all the eigenvalues equals the determinant of *A*.

Theorem 8.1.4.

Let $\lambda_1, \cdots, \lambda_n$ *be all the eigenvalues of the matrix A given in* **(8.1.1)** *. Then the following assertions hold.*

1. $\lambda_1 + \cdots + \lambda_n = \operatorname{tr}(A)$.
2. $\lambda_1 \times \cdots \times \lambda_n = \operatorname{tr}(A)$.

Proof

We only prove the case when $n = 2$. Let $A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$. Then

$$\begin{aligned} p(\lambda) &= \begin{vmatrix} \lambda - a_{11} & -a_{12} \\ -a_{21} & \lambda - a_{22} \end{vmatrix} = (\lambda - a_{11})(\lambda - a_{22}) - a_{12}a_{21} \\ &= \lambda^2 - (a_{11} + a_{22})\lambda + (a_{11}a_{22} - a_{12}a_{21}) = \lambda^2 - \operatorname{tr}(A)\lambda + |A|. \end{aligned}$$

It follows that

(8.1.8)

$$p(\lambda) = \lambda^2 - \operatorname{tr}(A)\lambda + |A|.$$

On the other hand, assume that λ_1 and λ_2 are solutions of $p(\lambda) = 0$. Then

(8.1.9)

$$p(\lambda) = (\lambda - \lambda_1)(\lambda - \lambda_2) = \lambda^2 - (\lambda_1 + \lambda_2)\lambda + \lambda_1\lambda_2.$$

By **(8.1.8)** and **(8.1.9)**, we see that $\lambda_1 + \lambda_2 = \text{tr}(A)$ and $\lambda_1\lambda_2 = |A|$.

Example 8.1.4.

Let *A* be the same as in **Example 8.1.3**. Find $\text{tr}(A)$ and $|A|$ using **Theorem 8.1.4**.

Solution

By **Example 8.1.3**, $\lambda_1 = 0, \ \lambda_2 = -2,$ and $\lambda_3 = 3$. By **Theorem 8.1.4**, $\text{tr}(A) = \lambda_1 + \lambda_2 + \lambda_3 = 1$ and $|A| = \lambda_1\lambda_2\lambda_3 = 0$.

By **Theorem 8.1.4** (2), we obtain the following result, which provides a criterion for a matrix to be invertible (see **Corollary 2.7.1** and **Theorem 3.4.1** for other criteria).

Corollary 8.1.3.

A square matrix A is invertible if and only if all its eigenvalues are nonzero.

By **Corollary 8.1.3** or equation **(8.1.4)**, we see that if zero is an eigenvalue of *A*, then *A* is not invertible.

Example 8.1.5.

Assume that *A* is a 2×2 matrix and $2I - A$ and $3I - A$ are not invertible. Find $|A|$ and $\text{tr}(A)$.

Solution

Because $2I - A$ and $3I - A$ are not invertible, $|2I - A| = 0$ and $|3I - A| = 0$. Hence, 2 and 3 are the eigenvalues of *A*. By **Theorem 8.1.4**, $|A| = 2 \times 3 = 6$ and $\text{tr}(A) = 2 + 3 = 5$.

Theorem 8.1.5.

1. *Let λ be an eigenvalue of A and let $\overrightarrow{X}$ be an eigenvector of A. Then for each integer k, λ^k is an eigenvalue of A^k and $\overrightarrow{X}$ is an eigenvalue of A^k.*
2. *Let $\phi(x) = a_0 + a_1 x + \cdots + a_m x^m$ be a polynomial of degree m and λ be an eigenvalue of A. Then $\phi(\lambda)$ is an eigenvalue of $\phi(A)$.*
3. *Assume that A is invertible and λ is an eigenvalue of A. Then $\frac{1}{\lambda}$ is an eigenvalue of A^{-1}.*

Proof

1. Because $A\overrightarrow{X} = \lambda\overrightarrow{X}$,

$$A^2\overrightarrow{X} = A(A\overrightarrow{X}) = A(\lambda\overrightarrow{X}) = \lambda A\overrightarrow{X} = \lambda(\lambda\overrightarrow{X}) = \lambda^2\overrightarrow{X}.$$

   Repeating the process implies that $A^k\overrightarrow{X} = \lambda^k\overrightarrow{X}$. The result follows.
2. Because λ is an eigenvalue of *A*, then for each positive integer *k*,

$$\lambda^k\overrightarrow{X} = A^k\overrightarrow{X} \text{ for some vector } \overrightarrow{X} \neq \overrightarrow{0}.$$

   Hence,

$$\begin{aligned} \phi(\lambda)I - \phi(A) &= (a_0 + a_1\lambda + \cdots + a_m\lambda^m)I - (a_0 I + a_1 A + \cdots + a_m A^m) \\ &= a_1(\lambda I - A) + a_2(\lambda^2 I - A^2) + \cdots + a_m(\lambda^m I - A^m). \end{aligned}$$

   It follows that

$$(\phi(\lambda)I - \phi(A))\overrightarrow{X} = a_1(\lambda I - A)\overrightarrow{X} + a_2(\lambda^2 I - A^2)\overrightarrow{X} + \cdots + a_m(\lambda^m I - A^m)\overrightarrow{X} = \overrightarrow{0}$$

   and $\phi(\lambda)$ is an eigenvalue of $\phi(A)$.

3. If λ is an eigenvalue of *A*, then $A\overrightarrow{X} = \lambda\overrightarrow{X}$ for some vector $\overrightarrow{X} \neq \overrightarrow{0}$. Because *A* is invertible, A^{-1} exists and

$$A^{-1}(A\overrightarrow{X}) = A^{-1}(\lambda\overrightarrow{X}) = \lambda A^{-1}(\overrightarrow{X}).$$

This implies that $\overrightarrow{X} = \lambda A^{-1}(\overrightarrow{X})$ and $\frac{1}{\lambda}\overrightarrow{X} = A^{-1}(\overrightarrow{X})$. Hence, $\frac{1}{\lambda}$ is an eigen-value of A^{-1}.

Example 8.1.6.

Let *A* be a 3×3 matrix. Assume that 1, 2, and 3 are the eigenvalues of *A*.

1. Find all eigenvalues of $2A + 3A^2$.
2. Find all eigenvalues of A^{-1}.

Solution

Let $\phi(x) = 2x + 3x^2$. Then $\phi(A) = 2A + 3A^2$. Because 1, 2, 3 are the eigenvalues of *A*, it follows from **Theorem 8.1.5** that $\phi(1),\ \phi(2)$, and $\phi(3)$ are the eigenvalues of $\phi(A)$. By computation, we have

$$\phi(1) = 2 + 3 = 5,\ \phi(2) = 2(2) + 3(2)^2 = 16 \text{ and } \phi(3) = 2(3) + 3(3)^2 = 33.$$

Hence, all the eigenvalues of $2A + 3A^2$ are 5, 16, and 33.

(2) By **Theorem 8.1.5** (3), all the eigenvalues of A^{-1} are $1,\ \frac{1}{2}$, and $\frac{1}{3}$.

Exercises

1. For each of the following matrices, find its eigenvalues and determine which eigenvalues are repeated eigenvalues.

$$A = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix} \quad B = \begin{pmatrix} 2 & 0 & 0 \\ 1 & 1 & 0 \\ 2 & -3 & -4 \end{pmatrix} \quad C = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

2. For each of the following matrices, find its eigenvalues and eigenspaces.

$$A = \begin{pmatrix} 5 & 7 \\ 3 & 1 \end{pmatrix} \; B = \begin{pmatrix} 3 & -1 \\ 1 & 5 \end{pmatrix} \; C = \begin{pmatrix} 3 & 0 \\ 0 & 3 \end{pmatrix} \; D = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \\ 1 & 1 & 2 \end{pmatrix}$$

3. Let *A* be a 3×3 matrix. Assume that $2I - A,\ 3I - A,$ and $4I - A$ are not invertible.
   1. Find $|A|$ and tr(*A*).
   2. Prove that $I + 5A$ is invertible.

4. Assume that the eigenvalues of a 3×3 matrix *A* are 1, 2, 3. Compute $\left|A^3 - 2A^2 + 3A + I\right|$.

8.2 Diagonalizable matrices

By **(2.3.7)**, we see that it is easy to compute D^k, where *D* is a diagonal matrix. For a nondiagonal matrix *A*, is it possible to find a suitable diagonal matrix *D* such that A^k can be computed by D^k? The answer is yes if *A* is diagonalizable.

Definition 8.2.1.

An $n \times n$ matrix *A* is said to be diagonalizable if there exists an $n \times n$ invertible matrix *P* such that $P^{-1}AP$ is a diagonal matrix. *P* is said to diagonalize *A*.

By **Definition 8.2.1** and **Theorem 3.4.1** (*i*), we obtain the following result.

Theorem 8.2.1.

An $n \times n$ matrix A is diagonalizable if and only if there exist an $n \times n$ matrix P with $|P| \neq 0$ and a diagonal matrix D such that

(8.2.1)

$$AP = PD.$$

The following result shows that if *A* is diagonalizable, then

(8.2.2)

$$D = \operatorname{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n) \quad \text{and } P = (\overrightarrow{p_1} \cdots \overrightarrow{p_n}),$$

where λ_i is the eigenvalue of *A* and the column vector $\overrightarrow{p_i}$ of *P* is the eigen-vector of *A* corresponding to λ_i for $i \in I_n$.

Theorem 8.2.2.

An $n \times n$ matrix A is diagonalizable if and only if it has n linearly independent eigenvectors. In this case, the matrix P in **(8.2.2)** *diagonalizes A and the diagonalizable matrix is the matrix D in* **(8.2.2)** *.*

Proof

Assume that *A* has *n* linearly independent eigenvectors: $\overrightarrow{p_1}, \cdots, \overrightarrow{p_n}$ such that

$$A\overrightarrow{p_i} = \lambda_i \overrightarrow{p_i} \quad \text{for } i = 1, \cdots, n.$$

Let *D* and *P* be the same as in **(8.2.2)** . Then *P* is invertible and

$$\begin{aligned} AP &= A(\overrightarrow{p_1} \cdots \overrightarrow{p_n}) = (A\overrightarrow{p_1} \cdots A\overrightarrow{p_n}) = (\lambda_1 \overrightarrow{p_1} \cdots \lambda_n \overrightarrow{p_n}) \\ &= (\overrightarrow{p_1} \cdots \overrightarrow{p_n}) \operatorname{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n) = PD. \end{aligned}$$

By **Theorem 8.2.1** , *A* is diagonalizable. Conversely, if *A* is diagonalizable, then $AP = PD$, where *P* and *D* are given by **(8.2.2)** and *P* is invertible. Because

$$AP = A(\overrightarrow{p_1} \cdots \overrightarrow{p_n}) = (A\overrightarrow{p_1} \cdots A\overrightarrow{p_n}) \quad \text{and } PD = (\lambda_1 \overrightarrow{p_1} \cdots \lambda_n \overrightarrow{p_n}),$$

it follows that

$$A\vec{p}_i = \lambda_i \vec{p}_i \quad \text{for } i = 1, \cdots, n.$$

Because P is invertible, by **Theorems 7.1.5** and **3.4.1**, $\left\{\vec{p}_1, \cdots, \vec{p}_n\right\}$ is linearly independent.

Remark 8.2.1.

1. By **(8.2.2)** and the proof of **Theorem 8.2.2**, we see that the matrices D and P depend on the order of eigenvalues and both follow the same order of the eigenvalues. Hence, different orders of eigenvalues give different matrices D and P. Hence, the choices of D and P are not unique (see **Remark 8.2.3**).
2. If there is more than one linearly independent eigenvector corresponding to one eigenvalue, then different orders of these eigenvectors give different matrices P, but the diagonal matrix D remains unchanged.

By **Theorem 8.2.2**, we see that, to determine whether A is diagonalizable, we need linearly independent eigenvectors of A and the number of linearly independent eigenvectors.

The following result provides the linear independence of the eigenvectors of a matrix.

Theorem 8.2.3.

Let λ_i and r_i be the same as in **(8.1.5)** *for $i \in \{1, 2, \cdots, k\}$. Assume that $\vec{pi1}, \vec{pi2}, \cdots, \vec{pir_i}$ are linearly independent eigenvectors corresponding to λ_{r_i} for $i \in I_n$. Then the set of all eigenvectors*

$$\overrightarrow{p11},\ \overrightarrow{p12},\ \cdots,\ \overrightarrow{p1r_1};\ \overrightarrow{p21},\ \overrightarrow{p22},\ \cdots,\ \overrightarrow{p2r_2};\ \cdots;\ \overrightarrow{pk1},\ \overrightarrow{pk2},\ \cdots,\ \overrightarrow{pkr_k}$$

is linearly independent.

Proof

Assume that there exist $x_{i1},\ x_{i2},\ \cdots xir_i \in \mathbb{R}$ for $i \in I_k$ such that

(8.2.3)

$$\sum_{i=1}^{k}\left(\sum_{j=1}^{r_i} x_{ij}\overrightarrow{pi_j}\right) = \vec{0}.$$

We prove below that $x_{ij} = 0$ for $j \in \left\{1,\ 2,\ \cdots,\ r_i\right\}$ and $i \in I_k$. Indeed, for $i \in I_k$, let

(8.2.4)

$$\overrightarrow{q_i} = \sum_{j=1}^{r_i} x_{ij}\overrightarrow{p_{ij}}.$$

Then **(8.2.3)** becomes $\sum_{i=1}^{k}\overrightarrow{q_i} = \vec{0}$. Let $i \in I_k$ be fixed. Then

(8.2.5)

$$\overrightarrow{q_1} + \cdots + \overrightarrow{q_i} + \cdots + \overrightarrow{q_k} = \vec{0}.$$

Because $A\overrightarrow{p_{ij}} = \lambda_i \overrightarrow{p_{ij}}$ for $j \in \{1, 2, \cdots, r_i\}$, we have for $i \in I_n$,

(8.2.6)

$$A\overrightarrow{q_i} = A\left[\sum_{j=1}^{r_i} x_{ij}\overrightarrow{p_{ij}}\right] = \sum_{j=1}^{r_i} x_{ij}A\overrightarrow{p_{ij}} = \lambda_i \sum_{j=1}^{r_i} x_{ij}\overrightarrow{p_{ij}} = \lambda_i \overrightarrow{q_i}$$

and

$$A(\overrightarrow{q_1} + \cdots + \overrightarrow{q_i} + \cdots + \overrightarrow{q_k}) = \lambda_1\overrightarrow{q_1} + \cdots + \lambda_i\overrightarrow{q_i} + \cdots + \lambda_k\overrightarrow{q_k}.$$

This, together with **(8.2.5)**, implies

(8.2.7)

$$\lambda_1\overrightarrow{q_1} + \cdots + \lambda_i\overrightarrow{q_i} + \cdots + \lambda_k\overrightarrow{q_k} = \vec{0}.$$

Multiplying **(8.2.5)** by λ_k implies

$$\lambda_k\overrightarrow{q_1} + \cdots + \lambda_k\overrightarrow{q_i} + \cdots + \lambda_k\overrightarrow{q_k} = \vec{0}.$$

This, together with **(8.2.7)** implies that

(8.2.8)

$$(\lambda_k-\lambda_1)\overrightarrow{q_1}+\cdots+(\lambda_k-\lambda_i)\overrightarrow{q_i}+\cdots+(\lambda_k-\lambda_{k-1})\overrightarrow{q_{k-1}}=\vec{0}.$$

By **(8.2.6)** and **(8.2.7)** , we obtain

(8.2.9)

$$(\lambda_k-\lambda_1)\lambda_1\overrightarrow{q_1}+\cdots+(\lambda_k-\lambda_i)\lambda_i\overrightarrow{q_i}+\cdots+(\lambda_k-\lambda_{k-1})\lambda_{k-1}\overrightarrow{q_{k-1}}=\vec{0}.$$

Multiplying **(8.2.8)** by λ_{k-1} implies

$$(\lambda_k-\lambda_1)\lambda_{k-1}\overrightarrow{q_1}+\cdots+(\lambda_k-\lambda_i)\lambda_{k-1}\overrightarrow{q_i}+\cdots+(\lambda_k-\lambda_{k-1})\lambda_{k-1}\overrightarrow{q_{k-1}}=\vec{0}.$$

This, together with **(8.2.9)** implies

$$\begin{aligned}(\lambda_k-\lambda_1)(\lambda_{k-1}-\lambda_1)\overrightarrow{q_1}\ &+\ \cdots+(\lambda_k-\lambda_i)(\lambda_{k-1}-\lambda_i)\overrightarrow{q_i}+\cdots\\ &+\ (\lambda_k-\lambda_{k-1})(\lambda_{k-1}-\lambda_{k-2})\overrightarrow{q_{k-2}}=\vec{0}.\end{aligned}$$

Repeating the above process by eliminating the following terms in order:

$$\overrightarrow{q_{k-2}},\ \overrightarrow{q_{k-3}},\ \cdots,\ \overrightarrow{q_{i+1}},\ \overrightarrow{q_{i-1}},\ \cdots,\overrightarrow{q_2},\ \overrightarrow{q_1},$$

we obtain

$$(\lambda_k-\lambda_i)(\lambda_{k-1}-\lambda_i)\cdots(\lambda_{i+1}-\lambda_i)(\lambda_{i-1}-\lambda_i)\cdots(\lambda_2-\lambda_i)(\lambda_1-\lambda_i)\overrightarrow{q_i}=\vec{0}.$$

Because $\lambda_j \neq \lambda_i$ for $j \neq i$, we have $\overrightarrow{q_i} = \vec{0}$. By **(8.2.4)** , $\sum_{j=1}^{r_i} x_{ij}\overrightarrow{p_{ij}} = \vec{0}$. Because $\overrightarrow{pi1}, \overrightarrow{pi2}, \cdots, \overrightarrow{pir_i}$ are linearly independent,

$$x_{ij} = 0 \quad \text{for } j \in \left\{1,\ 2,\ \cdots,\ r_i\right\} \text{ and } i \in I_k.$$

The result follows.

Theorem 8.2.3 shows that for each eigenvalue λ_i, we can find linearly independent eigenvectors. For example, for each eigenvalue λ_i, we can take all the basis vectors of E_{λ_i} and put them together, then the set of all the vectors is linearly independent.

The following result gives the largest number of linearly independent eigenvectors of *A*.

Theorem 8.2.4.

Let λ_i and r_i be the same as in **(8.1.5)** *for $i \in \{1,\ 2,\ \cdots,\ k\}$. Then the number $\sum_{i=1}^{k} \dim(E_{\lambda_i})$ is the largest number of linearly independent eigenvectors of A.*

Solution

Let $\overrightarrow{p_1}, \overrightarrow{p_2}, \cdots, \overrightarrow{p_r}$ be linearly independent eigenvectors of *A*. We regroup the eigenvectors corresponding to $\lambda_1, \lambda_2, \cdots, \lambda_k$ as

$$\left\{\overrightarrow{p11},\ \overrightarrow{p12},\ \cdots,\ \overrightarrow{p1s_1}\right\};\ \left\{\overrightarrow{p21},\ \overrightarrow{p22},\ \cdots,\ \overrightarrow{p2s_2}\right\};\ \left\{\overrightarrow{pk1},\ \overrightarrow{pk2},\ \cdots,\ \overrightarrow{pks_{r_k}}\right\}.$$

By **Theorem 7.2.8** , we have It follows that .

$$s_i \le \dim E_{\lambda_i} \quad \text{for } i \in \{1,\ 2,\ \cdots,\ k\}.$$

It follows that $\sum_{i=1}^{k} s_i \le \sum_{i=1}^{k} \dim(E_{\lambda_i})$.

Combining **Corollary 8.1.2**, **Theorems 8.2.2**, and **8.2.3**, we obtain the following result, which provides a necessary and sufficient condition for a matrix to be diagonalizable.

Theorem 8.2.5.

Let λ_i and r_i be the same as in **(8.1.5)** $i \in \{1,\ 2,\ \cdots,\ k\}$. *Then the following assertions hold.*

i. *If $\sum_{i=1}^{k} \dim(E_{\lambda_i}) = n$, then A is diagonalizable.*
ii. *If $\sum_{i=1}^{k} \dim(E_{\lambda_i}) < n$, then A is not diagonalizable.*

Proof

By **Corollary 8.1.2**, $k \le \sum_{i=1}^{k} \dim(E_{\lambda_i}) \le n$ and for each eigenvalue λ_i, there are $\dim(E_{\lambda_i})$ linearly independent eigenvectors. By **Theorem 8.2.3**, the set of all these eigenvectors is linearly independent.

i. If $\sum_{i=1}^{k} \dim(E_{\lambda_i}) = n$, then *A* has *n* linearly independent eigenvectors. It follows from **Theorem 8.2.2** that *A* is diagonalizable.
ii. If $\sum_{i=1}^{k} \dim(E_{\lambda_i}) < n$, then *A* does not have *n* linearly independent eigenvectors and by **Theorem 8.2.2**, *A* is not diagonalizable.

The following result provides the methods used to determine whether an $n \times n$ matrix *A* is diagonalizable.

Corollary 8.2.1.

t A b *e an $n \times n$ matrix and let λ_i and r_i be the same as in* **(8.1.5)** *for $i \in \{1,\ 2,\ \cdots,\ k\}$.*

1. *If A has n distinct eigenvalues, then A is diagonalizable.*
2. *If A has repeated eigenvalues:* $\lambda_{n_1},\ \lambda_{n_2},\ \cdots\lambda_{nl}$ *with algebraic multiplicities,* $r_{n_1},\ r_{n_2},\ \cdots r_{nl},$ *respectively, then the following assertions hold.*

   i. *If* $\dim(E_{\lambda_i}) = r_{n_i}$ *for* $i \in \{1,\ 2,\ \cdots,\ l\},$ *then A is diagonalizable.*

   ii. *If there exists* $i_0 \in \left\{1,\ 2,\ \cdots,\ l\right\}$ *such that* $\dim(E_{\lambda_{i_0}}) < r_{n_{i_0}},$ *then A is not diagonalizable.*

Proof

1. Let $\lambda_1,\ \lambda_2,\ \cdots,\lambda_n$ be distinct eigenvalues of *A*. By **Corollary 8.1.1** , $\dim(E_{\lambda_i}) = 1$ 1 for each $i \in I_n$. This implies that $\sum_{i=1}^{k}\dim(E_{\lambda_i}) = n.$ The result follows from **Theorem 8.2.5** (*i*).
2. Note that if $r_j = 1$ for $j \in \{1,\ 2,\ \cdots,\ k\},$ then by **Corollary 8.1.1** , $\dim(E_{\lambda_i}) = 1$. This, together with $\dim(E_{\lambda_i}) = r_{n_i}$ for $i \in \{1,\ 2,\ \cdots,\ l\},$ implies $\dim(E_{\lambda_i}) = r_i$ for all $i \in \{1,\ 2,\ \cdots,\ k\}.$ By **(8.1.6)** ,

$$\sum_{i=1}^{k}\dim(E_{\lambda_i}) = r_1 + r_2 + \cdots + r_k = n.$$

By **Theorem 8.2.5** , *A* is diagonalizable and the result (*i*) holds. If there exists $i_0 \in \left\{1,\ 2,\ \cdots,\ l\right\}$ such that $\dim(E_{\lambda_{n_{i_0}}}) < r_{n_{i_0}},$ then

$$\sum_{i=1}^{k}\dim(E_{\lambda_i}) < r_1 + r_2 + \cdots + r_k = n.$$

By **Theorem 8.2.5** , *A* is not diagonalizable and the result (*ii*)holds.

By **Corollary 8.2.1**, we see that if each of the algebraic multiplicities is equal to the corresponding geometric multiplicity, then *A* is diagonalizable (see (1) and (*i*) in **Corollary 8.2.1**) and if there exists a geometric multiplicity, which is less than the corresponding algebraic multiplicity, then *A* is not diagonalizable (see (*ii*) in **Corollary 8.2.1**).

Example 8.2.1.

For each of the following matrices, determine whether it is diagonalizable.

$$A = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 4 & -17 & 8 \end{pmatrix} B = \begin{pmatrix} 3 & 2 & 4 \\ 2 & 0 & 2 \\ 4 & 2 & 3 \end{pmatrix} C = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 2 & 0 \\ -3 & 5 & 2 \end{pmatrix}$$

Solution

$$\begin{aligned} p(\lambda) &= |\lambda I - A| = \begin{vmatrix} \lambda & -1 & 0 \\ 0 & \lambda & -1 \\ -4 & 17 & \lambda - 8 \end{vmatrix} = \lambda^2(\lambda - 8) - 4 + 17\lambda \\ &= \lambda^2 - 8\lambda^2 + 17\lambda - 4 = (\lambda - 4)(\lambda^2 - 4\lambda + 1) \\ &= (\lambda - 4)[(\lambda - 2)^2 - 3] = (\lambda - 4)[(\lambda - 2)^2 - (\sqrt{3})^2] \\ &= (\lambda - 4)(\lambda - 2 - \sqrt{3})(\lambda - 2 + \sqrt{3}). \end{aligned}$$

Solving $p(\lambda) = 0$ implies that $\lambda_1 = 4,\ \lambda_2 = 2 + \sqrt{3},$ and $\lambda_3 = 2 - \sqrt{3}.$ Hence, *A* has three distinct eigenvalues. By **Corollary 8.2.1** (1), *A* is diagonalizable.

$$\begin{aligned}
p(\lambda) &= |\lambda I - B| = \begin{vmatrix} \lambda-3 & -2 & -4 \\ -2 & \lambda & -2 \\ -4 & -2 & \lambda-3 \end{vmatrix} \\
&= [\lambda(\lambda-3)-16-16]-[16\lambda+4(\lambda-3)+4(\lambda-3)] \\
&= [\lambda^3-6\lambda^2+9\lambda-32]-[24\lambda-24] = \lambda^3-6\lambda^2-15\lambda-8 \\
&= (\lambda-8)(\lambda^2+2\lambda+1) = (\lambda+1)^2(\lambda-8).
\end{aligned}$$

Solving $p(\lambda)=0,$ we have $\lambda_1 = -1$ and $\lambda_2 = 8$. Moreover, $\lambda_1 = -1$ has algebraic multiplicity 2 and $\lambda_2 = 8$ has algebraic multiplicity 1. We find $\dim(E_{\lambda_i})$. For $\lambda_1 = 1,$ the system $\lambda_1 I - B = 0$ becomes

$$\begin{pmatrix} -4 & -2 & -4 \\ -2 & -1 & -2 \\ -4 & -2 & -4 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

We solve the above system.

$$(B\,|\,\vec{0}) = \begin{pmatrix} -4 & -2 & -4 & 0 \\ -2 & -1 & -2 & 0 \\ -4 & -2 & -4 & 0 \end{pmatrix} \xrightarrow{R_1\left(-\frac{1}{2}\right)} \begin{pmatrix} 2 & 1 & 2 & 0 \\ -2 & -1 & -2 & 0 \\ -4 & -2 & -4 & 0 \end{pmatrix}$$

$$\xrightarrow[R_1(4)+R_3]{R_1(2)+R_2} \begin{pmatrix} 2 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

The system corresponding to the last augmented matrix is

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

and has two free variables. Thus, $\dim(E_{\lambda_1}) = 2$ (algebraic multiplicity). By **Corollary 8.2.1** (*i*), *B* is diagonalizable.

Because

$$p(\lambda) = |\lambda I - C| = \begin{vmatrix} \lambda - 1 & 0 & 0 \\ -1 & \lambda - 2 & 0 \\ 3 & -5 & \lambda - 2 \end{vmatrix} = (\lambda - 1)(\lambda - 2)^2 = 0.$$

$\lambda_1 = 1$ has algebraic multiplicity 2 and $\lambda_2 = 2$ has algebraic multiplicity 2. We find $\dim(E_{\lambda_2})$. For $\lambda_2 = 2,$ the system $\lambda_2 I - C = 0$ becomes

$$\begin{pmatrix} 1 & 0 & 0 \\ -1 & 0 & 0 \\ 3 & -5 & 0 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

We solve the above system.

$$(C\,|\,\vec{0}) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 3 & -5 & 0 & 0 \end{pmatrix} \xrightarrow[R_1(-3)+R_3]{R_1(1)+R_2} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & -5 & 0 & 0 \end{pmatrix}$$

$$\xrightarrow{R_{2,3}} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -5 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

The system corresponding to the last augmented matrix has one free variable and thus, $\dim(E_{\lambda_2}) = 1 < 2$ (algebraic multiplicity). By **Corollary 8.2.1** (*ii*), *C* is not diagonalizable.

Example 8.2.2.

Find $x \in \mathbb{R}$ such that *A* is diagonalizable, where

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

Solution

Because

$$\left|\lambda I - A\right| = \begin{vmatrix} \lambda - 2 & 0 & -1 \\ -2 & \lambda - 1 & -x \\ 0 & 0 & \lambda - 1 \end{vmatrix} = (\lambda - 1)^2(\lambda - 2) = 0,$$

$\lambda_1 = 1$ (with algebraic multiplicity 2), $\lambda_2 = 2$.

We find $\dim(E_{\lambda_1})$. For $\lambda_1 = 1,$ the system $\lambda_1 I - A = 0$ becomes

$$\begin{pmatrix} -1 & 0 & -4 \\ -2 & 0 & -x \\ 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

We solve the above system.

$$(A\,|\,\vec{0}) = \begin{pmatrix} -1 & 0 & -1 & 0 \\ -2 & 0 & -x & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \xrightarrow{R_1(-2)+R_2} \begin{pmatrix} -1 & 0 & -1 & 0 \\ 0 & 0 & -x+2 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

If $x = 2,$ then the system corresponding to the last augmented matrix has two free variables and thus, $\dim(E_{\lambda_1}) = 2$ (algebraic multiplicity). By **Corollary 8.2.1** (*i*), *A* is diagonalizable.

In the above examples, we used eigenvalues or eigenvectors of a matrix to determine whether the matrix is diagonalizable. Now, we use **(8.2.2)** and **Theorem 8.2.2** to find the diagonalizable matrix *D* and the invertible matrix *P* that diagonalizes *A*.

Example 8.2.3.

Let $A = \begin{pmatrix} 4 & 2 \\ 3 & 3 \end{pmatrix}.$

1. Find the eigenvalues and eigenspace of *A*.
2. Find an inverse matrix *P* and a diagonal matrix *D* such that

$$A = PDP^{-1}.$$

Solution

1.

$$\begin{aligned} p(\lambda) &= \left|\lambda I - A\right| = \begin{vmatrix} \lambda - 4 & -2 \\ -3 & \lambda - 3 \end{vmatrix} = (\lambda - 4)(\lambda - 3) - 6 \\ &= \lambda^2 - 7\lambda + 12 - 6 = \lambda^2 - 7\lambda + 6 = (\lambda - 1)(\lambda - 6). \end{aligned}$$

Solving $p(\lambda) = (\lambda - 1)(\lambda - 6) = 0$ implies $\lambda_1 = 1$ and $\lambda_2 = 6$.

For $\lambda_1 = 1,\ (\lambda_1 I - A)\vec{X} = \vec{0}$ becomes

$$\begin{pmatrix} 1-4 & -2 \\ -3 & 1-3 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} -3 & -2 \\ -3 & -2 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

The above system is equivalent to the equation

$$3x + 2y = 0,$$

where *x* is a basic variable and *y* is a free variable. Let $y = 3t$. Then $x = -2t$. Hence,

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} -2t \\ 3t \end{pmatrix} = t\begin{pmatrix} -2 \\ 3 \end{pmatrix} = t\overrightarrow{v_1},$$

where $\overrightarrow{v_1} = (-2,\ 1)^T$. Then $E_{\lambda_1} = \left\{ t\overrightarrow{v_1} : t \in \mathbb{R} \right\}$.

For $\lambda_2 = 6,\ (\lambda_2 I - A)\vec{X} = \vec{0}$ becomes

$$\begin{pmatrix} 6-4 & -2 \\ -3 & 6-3 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 2 & -2 \\ -3 & 3 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

The above system is equivalent to

$$x - y = 0,$$

where x is a basic variable and y is a free variable. Let $y = t$. Then $x = t$ and

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} t \\ t \end{pmatrix} = t\begin{pmatrix} 1 \\ 1 \end{pmatrix} = t\overrightarrow{v_2},$$

where $\overrightarrow{v_2} = (1,\ 1)^T$. Then $E_{\lambda_2} = \left\{ t\overrightarrow{v_2} : t \in \mathbb{R} \right\}$.

2. Let $D = \mathrm{diag}(\lambda_1,\ \lambda_1) = \mathrm{diag}(1, 6)$ and $P = (\overrightarrow{v_1}\overrightarrow{v_2}) = \begin{pmatrix} -2 & 1 \\ 3 & 1 \end{pmatrix}$.

   Then by **Theorem 8.2.2** , $A = P^{-1}DP$.

Remark 8.2.2.

In the above solution, we use **Theorem 8.2.2** to obtain $A = P^{-1}DP$. Infact,we cancheck $AP = PD$. Indeed,

$$AP = \begin{pmatrix} 4 & 2 \\ 3 & 3 \end{pmatrix}\begin{pmatrix} -2 & 1 \\ 3 & 1 \end{pmatrix} = \begin{pmatrix} -2 & 6 \\ 3 & 6 \end{pmatrix}$$

and $PD = \begin{pmatrix} -2 & 1 \\ 3 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 6 \end{pmatrix} = \begin{pmatrix} -2 & 6 \\ 3 & 6 \end{pmatrix}$.

Remark 8.2.3.

As we mentioned in **Remark 8.2.1** , the choice of *D* and *P* is not unique. In **Example 8.2.1** , we can make the following choice:

$$D_1 = \text{diag}(\lambda_2, \lambda_1) = \text{diag}(6, 1)=\begin{pmatrix} 6 & 0 \\ 0 & 1 \end{pmatrix}$$

and

$$P_1 = (\overrightarrow{v_2}\overrightarrow{v_1}) = \begin{pmatrix} 1 & -2 \\ 1 & 3 \end{pmatrix}.$$

By direct verification, we have $AP_1 = P_1D_1$ and $\left|P_1\right| \neq 0$.

Note that both *D* and *P* must follow the same order of the choices of eigenvalues. For example, we cannot choose either *D* and P_1 or D_1 and *P* because you can check $AP_1 \neq P_1D$ and $AP \neq PD_1$.

By **Theorem 8.2.2** , we can choose *P* as follows:

$$P = (\alpha\overrightarrow{v_1}\beta\overrightarrow{v_2}),$$

where $\alpha \neq 0$ and $\beta \neq 0$.

In **Example 8.2.3** , the geometric multiplicity of each eigenvalue is 1.

The following example deals with the case when geometric multiplicities of eigenvalues are greater than 1. The matrix below is taken from **Example 8.2.2** with $x = 2$.

Example 8.2.4.

Let

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

1. Find the eigenvalues and eigenspaces of *A*.
2. Find an inverse matrix *P* and a diagonal matrix *D* such that

$$A = PDP^{-1}.$$

Solution

1.

$$p(\lambda) = |\lambda I - A| = \begin{vmatrix} \lambda - 2 & 0 & 1 \\ -2 & \lambda - 1 & -2 \\ 0 & 0 & \lambda - 1 \end{vmatrix} = (\lambda - 1)^2(\lambda - 2).$$

Solving $p(\lambda) = (\lambda - 1)^2(\lambda - 2) = 0$ implies $\lambda_1 = 1$ and $\lambda_2 = 2$. The algebraic multiplicities of λ_1 and λ_2 are 2 and 1, respectively.
For $\lambda_1 = 1,$ the system $\lambda_1 I - A = 0$ becomes

$$\begin{pmatrix} -1 & 0 & -1 \\ -2 & 0 & -2 \\ 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

We solve the above system.

$$(A \mid \vec{0}) = \begin{pmatrix} -1 & 0 & -1 & 0 \\ -2 & 0 & -2 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \overset{R_1(-2)+R_2}{\longrightarrow} \begin{pmatrix} -1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

The system with the last augmented matrix is equivalent to the equation

$$x + 0y + z = 0,$$

where *x* is a basic variable and *y* and *z* are free variables. Let $y = s$ and $z = t.$ Then $x = -t.$ Hence,

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} -t \\ s \\ t \end{pmatrix} = s\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + t\begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} = t\overrightarrow{v_1} + s\overrightarrow{v_2},$$

where $\overrightarrow{v_1} = (0,\ 1,\ 0)^T$ and $\overrightarrow{v_2} = (-1,\ 0,\ 1)^T.$ Then $E_{\lambda_1} = \left\{s\overrightarrow{v_1} + t\overrightarrow{v_2} : s,\ t \in \mathbb{R}\right\}.$

For $\lambda_2 = 2,$ the system $\lambda_2 I - A = 0$ becomes

$$\begin{pmatrix} 0 & 0 & -1 \\ -2 & 1 & -2 \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

We solve the above system.

$$(A \mid \vec{0}) = \begin{pmatrix} 0 & 0 & -1 & 0 \\ -2 & 1 & -2 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \xrightarrow{R_{1,2}} \begin{pmatrix} -2 & 1 & -2 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

$$\xrightarrow[R_2(-2)+R_1]{R_2(1)+R_3} \begin{pmatrix} -2 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

The system with the last augmented matrix is equivalent to the equation

$$\begin{cases} -2x + y = 0 \\ -z = 0, \end{cases}$$

where *x* and *z* are basic variables and *y* is a free variable. Let $y = 2t$. Then $x = t$. Let $\vec{v}_3 = (1, 2, 0)^T$.

Then $E_{\lambda_2} = \left\{ t\vec{v}_3 : t \in \mathbb{R} \right\}$.

2. Let $D = \text{diag}(\lambda_1, \lambda_2, \lambda_3) = \text{diag}(1, 1, 2)$ and

$$P = (\vec{v}_1 \vec{v}_2 \vec{v}_3) = \begin{pmatrix} 0 & -1 & 1 \\ 1 & 0 & 2 \\ 0 & 1 & 0 \end{pmatrix}.$$

Then by **Theorem 8.2.2**, $A = PDP^{-1}$.

Remark 8.2.4.

In **Example 8.2.4**, there are two basis eigenvectors $\vec{v}_1$ and $\vec{v}_2$ corresponding to the eigenvalue $\lambda_1 = 1$. We choose the column vectors of *P* in the order $\vec{v}_1, \vec{v}_2, \vec{v}_3$. By **Theorem 8.2.2**, we can choose the following order $\vec{v}_1, \vec{v}_2, \vec{v}_3$. If we let

$$P_1 = (\vec{v}_2 \vec{v}_1 \vec{v}_3) = \begin{pmatrix} -1 & 0 & 1 \\ 0 & 1 & 2 \\ 1 & 0 & 0 \end{pmatrix},$$

then it is easily verified that $A = P_1 D P_1^{-1}$.

Finally, we give the methods used to compute A^k if *A* is diagonalizable.

Theorem 8.2.6.

If $P^{1}\overline{A}$ $P = D$, *then* $A^k = PD^kP^{-1}$ *for each* $k \in \mathbb{N}$.

Proof

Because $P^{-1}AP = D,\ AP = PD$ and $A = PDP^{-1}$.

$$A^2 = (PDP^{-1})(PDP^{-1}) = PD(P^{-1}P)DP^{-1} = PD^2P^{-1}.$$

Repeating the process implies that the result holds.

Example 8.2.5.

Let

$$A = \begin{pmatrix} 1 & -1 & 4 \\ 3 & 2 & -1 \\ 2 & 1 & -1 \end{pmatrix}.$$

1. Find the eigenvalues and eigenspace of *A*.
2. Find an inverse matrix *P* and a diagonal matrix *D* such that

$$A = PDP^{-1}.$$

3. Compute A^3 by using the result (2).

Solution

1.

$$\begin{aligned} p(\lambda) &= \left| \lambda I - A \right| = \begin{vmatrix} \lambda - 1 & 1 & -4 \\ -3 & \lambda - 2 & 1 \\ -2 & -1 & \lambda + 1 \end{vmatrix} \\ &= (\lambda-1)(\lambda-2)(\lambda+1) - 12 - 2 - 8(\lambda-2) + (\lambda-1) + 3(\lambda+1) \\ &= (\lambda^2-1)(\lambda-2) - 14 - 8\lambda + 16 + \lambda - 1 + 3\lambda + 3 \\ &= \lambda^3 - 2\lambda^2 - 5\lambda + 6 = (\lambda-1)(\lambda^2-\lambda-6) = (\lambda-1)(\lambda-3)(\lambda+2). \end{aligned}$$

Solving $p(\lambda) = 0$, we get three eigenvalues

$$\lambda_1 = -2, \quad \lambda_2 = 1, \text{ and } \lambda_3 = 3.$$

For $\lambda_1 = -2$, $(\lambda_1 I - A)\vec{X} = \vec{0}$ becomes

$$B\vec{X} := \begin{pmatrix} -3 & 1 & -4 \\ -3 & -4 & 1 \\ -2 & -1 & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

Now we use row operations to solve the above system.

$$(B\,|\,\vec{0}) = \begin{pmatrix} -3 & 1 & -4 & 0 \\ -3 & -4 & 1 & 0 \\ -2 & -1 & 1 & 0 \end{pmatrix} \xrightarrow{R_3(-1)+R_1} \begin{pmatrix} -1 & 2 & -3 & 0 \\ -3 & -4 & 1 & 0 \\ -2 & -1 & -1 & 0 \end{pmatrix}$$

$$\xrightarrow[R_1(-3)+R_2]{R_1(-2)+R_3} \begin{pmatrix} -1 & 2 & -3 & 0 \\ 0 & -10 & 10 & 0 \\ 0 & -5 & 5 & 0 \end{pmatrix} \xrightarrow[R_2\left(-\frac{1}{10}\right)]{R_3\left(-\frac{1}{5}\right)} \begin{pmatrix} -1 & 2 & -3 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 1 & -1 & 0 \end{pmatrix}$$

$$\xrightarrow{R_2(-1)+R_3} \begin{pmatrix} -1 & 2 & -3 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \xrightarrow{R_2(-2)+R_1} \begin{pmatrix} -1 & 0 & -1 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

The system corresponding to the last augmented matrix is

$$\begin{cases} x_1 + x_3 &= 0 \\ x_2 - x_3 &= 0, \end{cases}$$

where x_1 and x_2 are basic variables and x_3 is a free variable. Let $x_3 = t$. Then $x_2 = t$ and $x_1 = -t$. Let $\vec{v}_1 = (-1,\ 1,\ 1)^T$. Then

$$E_{\lambda_1} = \left\{ t\vec{v}_1 : t \in \mathbb{R} \right\}.$$

For λ2=1, (λ2I−A)X→=0→ becomes

(01−4−3−11−2−12)(x1x2x3)=(000).

Now we use row operations to solve the above system.

$$(B|\vec{0})=$$
$$\left(\begin{array}{ccc|c}0&1&-4&0\\-3&-1&1&0\\-2&-1&2&0\end{array}\right)\xrightarrow{R_{1,2}}\left(\begin{array}{ccc|c}-3&-1&1&0\\0&1&-4&0\\-2&-1&2&0\end{array}\right)\xrightarrow{R_3(-1)+R_1}\left(\begin{array}{ccc|c}-1&0&-1&0\\0&1&-4&0\\-2&-1&2&0\end{array}\right)\xrightarrow{R_1(-1)}$$
$$\left(\begin{array}{ccc|c}1&0&1&0\\0&1&-4&0\\-2&-1&2&0\end{array}\right)\xrightarrow{R_1(2)+R_3}\left(\begin{array}{ccc|c}1&0&1&0\\0&1&-4&0\\0&-1&4&0\end{array}\right)\xrightarrow{R_2(1)+R_3}\left(\begin{array}{ccc|c}1&0&1&0\\0&1&-4&0\\0&0&0&0\end{array}\right).$$

Hence, we obtain

$$\begin{cases}x_1+x_3=0\\x_2-4x_3=0,\end{cases}$$

where x_1 and x_2 are basic variables and x_3 is a free variable. Let $x_3=t$. Then $x_2=4t$ and $x_1=-t$. Let $\vec{u}_2=(-1, 4, 1)^T$. Then

$$E_{\lambda_2}=\{t\vec{u}_2 : t\in\mathbb{R}\}.$$

For $\lambda_3=3$, $(\lambda_3 I-A)\vec{X}=\vec{0}$ becomes

$$\begin{pmatrix}2&1&-4\\-3&1&1\\-2&-1&4\end{pmatrix}\begin{pmatrix}x_1\\x_2\\x_3\end{pmatrix}=\begin{pmatrix}0\\0\\0\end{pmatrix}.$$

Now we use row operations to solve the above system.

$$(B|\vec{0})\left(\begin{array}{ccc|c}2&1&-4&0\\-3&1&1&0\\-2&-1&4&0\end{array}\right)\xrightarrow{R_2(1)+R_1}\left(\begin{array}{ccc|c}-1&2&-3&0\\-3&1&1&0\\-2&-1&4&0\end{array}\right)\xrightarrow{R_1(-1)}$$
$$\left(\begin{array}{ccc|c}1&-2&3&0\\-3&1&1&0\\-2&-1&4&0\end{array}\right)\xrightarrow[R_1(3)+R_2]{R_1(2)+R_3}\left(\begin{array}{ccc|c}1&-2&3&0\\0&-5&10&0\\0&-5&10&0\end{array}\right)\xrightarrow{R_2(-1)+R_3}\left(\begin{array}{ccc|c}1&-2&3&0\\0&-5&10&0\\0&0&0&0\end{array}\right)\xrightarrow{R_2(-\frac{1}{5})}$$
$$\left(\begin{array}{ccc|c}1&-2&3&0\\0&1&-2&0\\0&0&0&0\end{array}\right)\xrightarrow{R_2(2)+R_1}\left(\begin{array}{ccc|c}1&0&-1&0\\0&1&-2&0\\0&0&0&0\end{array}\right).$$

This implies

$$\begin{cases}x_1-x_3=0\\x_2-2x_3=0,\end{cases}$$

where x_1 and x_2 are basic variables and x_3 is a free variable. Let $x_3=t$. Then $x_2=2t$ and $x_1=t$. Let $\vec{u}_3=(1, 2, 1)^T$. Then

$$E_{\lambda_3}=\{t\vec{u}_3 : t\in\mathbb{R}\}.$$

2. Let $D=\operatorname{diag}(\lambda_1, \lambda_2, \lambda_3)=\operatorname{diag}(-2, 1, 3)$ and

$$P=(\vec{u}_1\ \vec{u}_2\ \vec{u}_3)=(-1111-421-11).$$

Then by **Theorem 8.2.2** , A=PDP−1.

3. By computation, we have

D3=(−200010003)3=(−8000100027).

Now, we find P−1.

(P|I3)=
(−1−11100142010111001)→R1(1)+R2R1(1)+R3(−1−111000331101111001)→R2(13),R3(12)R1(−1)
(11−1−10001113130001120 12)→R3(−1)+R2R3(1)+R1(110−12012010−1613−1200112012)→R3(−1)+R1
(100−13−131010−1613−1200112012)=(I3|P−1).

By **Theorem 8.2.6** , we obtain

A3=PD3P−1=(−1−11142111)(−8000100027)(−13−131−1613−1212012)=(8−127−8454−8127)
(−13−131−1613−1212012)=(11−322294171635).

Exercises

1. For each of the following matrices, determine whether it is diagonalizable.

A=(−124003170058000−2)B=(−1430−110−43).

2. Find x∈ℝ such that *A* is diagonalizable, where

A=(201−41x102).

3. Assume that u→=(1, 1, 0)T is an eigenvector of the matrix

A=(1−113x2y01).

   i. Find x, y∈ℝ R and the eigenvalue corresponding to u→.
   ii. Is *A* diagonalizable?

4. Let

$$A=(1-1-41).$$

1. Find the eigenvalues and eigenspaces of *A*.
2. Find an inverse matrix *P* and a diagonal matrix *D* such that A=PDP−1.

5. Let

$$A=(324202423).$$

1. Find the eigenvalues and eigenspaces of *A*.
2. Find an inverse matrix *P* and a diagonal matrix *D* such that

$$A=PDP-1.$$

3. Compute A2 by using the result (2).

6. Let

$$A=(00-2121103).$$

1. Find the eigenvalues and eigenspaces of *A*.
2. Find an inverse matrix *P* and a diagonal matrix *D* such that A=PDP−1.
3. Compute A2 by using the result (2).

7. Let

$$A=(1-113-32001).$$

1. Find the eigenvalues and eigenspaces of *A*.
2. Find an inverse matrix *P* and a diagonal matrix *D* such that A=PDP−1.
3. Compute A3 by using the result (2).

8. Let

$$A=(01-10).$$

1. Find the eigenvalues and eigenspaces of *A*.
2. Find an inverse matrix *P* and a diagonal matrix *D* such that $A=PDP-1$.
3. Compute $A2$ by using the result (2).

9. Two $n \times n$ matrices *A* and *B* are said to be similar (or *A* is similar to *B*) if there exists an invertible $n \times n$ matrix *P* such that $B=P-1AP$. We write $A\sim B$. Let

   $$A=(1-121),B=(1-211), \text{ and } P=(01-10).$$

   Show that *A* is similar to *B*.
10. Assume that $A\sim B$. Show that the following assertions hold.
    i. $|\lambda I-A|=|\lambda I-B|$.
    ii. *A* and *B* have the same eigenvalues.
    iii. $tr(A)=tr(B)$.
    iv. $|A|=|B|$.
    v. $r(A)=r(B)$.

11. Assume that *A, B, C* are $n \times n$ matrices. Show that the following assertions hold
    i. $A\sim A$.
    ii. If $A\sim B$, then $B\sim A$.
    iii. If $A\sim B$ and $B\sim C$, then $A\sim C$.

12. Assume that *A* and *B* are $n \times n$ matrices and $A\sim B$. Prove that the following assertions hold.
    a. $Ak\sim Bk$ for every positive integer *k*.
    b. Let $\phi(\lambda)=am\lambda m+am-1\lambda m-1+\cdots+a1\lambda+a0$. Then $\phi(A)\sim\phi(B)$.

Chapter 9 Vector spaces

9.1 Subspaces of $\mathbb{R}^n$

Definition 9.1.1.

Let W be a nonempty subset of $\mathbb{R}^n$. W is called a subspace of $\mathbb{R}^n$ if it satisfies the following two conditions:

i. If $\vec{a} = (x_1,\ x_2,\ \cdots, x_n) \in W,$ and $\vec{b} = (y_1,\ y_2,\ \cdots, y_n) \in W,$ then

$$\vec{a} + \vec{b} = (x_1 + y_1,\ x_2 + y_2,\ \cdots, x_n + y_n) \in W.$$

ii. If $\vec{a} = (x_1,\ x_2,\ \cdots, x_n) \in W$ and $k \in \mathbb{R}$ is a real number, then

$$k\vec{a} = (kx_1,\ kx_2,\ \cdots, kx_n) \in W.$$

Note that the above conditions (*i*) and (*ii*) are equivalent to the following condition: for $\vec{a} = (x_1,\ x_2,\ \cdots, x_n) \in W,\ \vec{b} = (y_1,\ y_2,\ \cdots, y_n) \in W,\ \alpha \in \mathbb{R}$ and $\beta \in \mathbb{R}$,

(9.1.1)

$$\alpha\vec{a} + \beta\vec{b} \in W.$$

By **Definition 9.1.1**, we see that $\mathbb{R}^n$ is a subspace of itself, and the set containing only the origin

$$W: \{\underbrace{(0, 0, \cdots, 0)}_{n}\}$$

is a subspace of $\mathbb{R}^n$.

In **Definition 9.1.1** (*ii*), if $k = 0$ and $\vec{a} \in W$, then $k\vec{a} = 0\vec{a} = \vec{0} \in W.$. This shows that every subspace of $\mathbb{R}^n$ contains the zero vector $\vec{0}$, Hence, we have the following remark.

Remark 9.1.1.

Any nonempty subsets of $\mathbb{R}^n$ that do not contain the origin of $\mathbb{R}^n$ are not subspaces of $\mathbb{R}^n$.

The following result shows that the solution space of a homogeneous system (see **(4.1.13)**) is a subspace of $\mathbb{R}^n$.

Theorem 9.1.1.

Let A be an $m \times n$ matrix. Then the solution space

$$N_A = \left\{\vec{X} \in \mathbb{R}^n : A\vec{X} = \vec{0}\right\}$$

is a subspace of $\mathbb{R}^n$.

Proof

Note that $\vec{X} \in N_A$ if and only if $A\vec{X} = \vec{0}$. Let $\vec{a} \in N_A$, $\vec{b} \in N_A$, and $k \in \mathbb{R}$. Then $A\vec{a} = \vec{0}$ and $A\vec{b} = \vec{0}$. Hence, by **Theorem 2.2.1**,

$$A(\vec{a} + \vec{b}) = A\vec{a} + A\vec{b} = \vec{0} + \vec{0} \quad \text{and } A(k\vec{a}) = k\vec{0} = \vec{0}.$$

This implies $\vec{a}+\vec{b} \in N_A$ and $k\vec{a} \in N_A$. By **Definition 9.1.1**, N_A is a subspace of $\mathbb{R}^n$.

The following result shows that the set of solutions of a nonhomogeneous system $A\vec{X}=\vec{b}$ is not a subspace of $\mathbb{R}^n$ because it doesn't contain the zero vector in $\mathbb{R}^n$.

Theorem 9.1.2.

Let A be an $m \times n$ matrix and $\vec{b} \in \mathbb{R}^m$ and let

$$S_A = \left\{\vec{X} \in \mathbb{R}^n : A\vec{X} = \vec{b}\right\}.$$

If $A\vec{X}=\vec{b}$ is consistent and $\vec{b} \neq \vec{0}$, then S_A is not a subspace of $\mathbb{R}^n$.

Proof

Because $A\vec{X}=\vec{b}$ is consistent, $A\vec{X}=\vec{b}$ has at least one solution and the set S_A is a nonempty set. Because $\vec{b} \neq \vec{0}$, $\vec{0}$ is not a solution of $A\vec{X}=\vec{b}$. Hence, $\vec{0} \notin S_A$. By **Theorem 9.1.2**, S_A is not a subspace of $\mathbb{R}^n$.

What types of subsets in $\mathbb{R}^n$ are subspaces of $\mathbb{R}^n$? In the following, we find all the subspaces in $\mathbb{R}^2$ and $\mathbb{R}^3$. The same approach can be used to deal with $\mathbb{R}^n$ with $n \geq 4$. First, we consider subspaces in $\mathbb{R}^2$.

Example 9.1.1.

Show that *W* is a subspace of $\mathbb{R}^2$, where

$$W = \left\{(x,\ y) \in \mathbb{R}^2 : x + 2y = 0\right\}.$$

Solution

Let $A = (1,\ 2)$ and $\vec{X} = (x,\ y)^T$. Then

$$W = \left\{(x,\ y) \in \mathbb{R}^2 : x + 2y = 0\right\} = \left\{(x,\ y) \in \mathbb{R}^2 : A\vec{X} = \vec{0}\right\}.$$

It follows from **Theorem 9.1.1** that *W* is a subspace of $\mathbb{R}^2$.

Note that $x + 2y = 0$ represents the equation of the line passing through the origin (0, 0). By **Theorem 9.1.1**, one can show that any line passing through the origin (0, 0) is a subspace of $\mathbb{R}^2$. The equation of a line passing through the origin (0, 0) is of the form

$$ax + by = 0,$$

where one of *a* and *b* is not zero.

Conclusion: All the subspaces in $\mathbb{R}^2$ are the origin {(0, 0)}, the entire space $\mathbb{R}^2$, and all the lines passing through the origin (0, 0). Any other subsets of $\mathbb{R}^2$ are not subspaces of $\mathbb{R}^2$. We do not prove the result here, but we give two examples that are not subspaces of $\mathbb{R}^2$.

Example 9.1.2.

Show that *W* is not a subspace of $\mathbb{R}^2$, where

$$W = \left\{(x,\ y) \in \mathbb{R}^2 : x + 2y = 1\right\}.$$

Solution

Because $(x,\ y) = (0,\ 0)$ is not a solution of $x + 2y = 1$, $(0,\ 0) \notin W$. By **Remark 9.1.1**, *W* is not a subspace of $\mathbb{R}^2$.

Note that $x + 2y = 1$ represents the equation of the line that does not pass through the origin (0, 0). By **Theorem 9.1.2** , one can see that any line that does not pass through the origin (0, 0) is not a subspace of $\mathbb{R}^2$. The equation of a line that does not pass through the origin (0, 0) is of the form

$$ax + by = c,$$

where one of *a* and *b* is not zero, and $c \neq 0$.

Example 9.1.3.

Show that *W* is not a subspace of $\mathbb{R}^2$, where

$$W = \left\{(x, y) \in \mathbb{R}^2 : y = x^2\right\}.$$

Solution

Let $(x_1, y_1) = (1, 1)$ and $k = 2$. Then $(x_1, y_1) \in W$ and

$$k(x_1, y_1) = 2(1, 1) = (2, 2) \notin W$$

because $2 = y_1 \neq x_1^2 = 2^2 = 4$. Hence, *W* does not satisfy the condition (*ii*) of **Definition 9.1.1** and *W* is not a subspace of $\mathbb{R}^2$.

Example 9.1.4.

Show that *W* is a subspace of $\mathbb{R}^3$, where

$$W = \left\{(x, y, z) \in \mathbb{R}^3 : x + 2y + 3z = 0\right\}.$$

Solution

Let $A = (1,\ 2,\ 3)$ and $\vec{X} = (x,\ y,\ z)^T$. Then

$$W = \left\{(x,\ y,\ z) \in \mathbb{R}^3 : x + 2y + 3z = 0\right\} = \left\{(x,\ y,\ z) \in \mathbb{R}^3 : A\vec{X} = \vec{0}\right\}.$$

It follows from **Theorem 9.1.1** that *W* is a subspace of $\mathbb{R}^2$.

Note that $x + 2y + 3z = 0$ represents the equation of the plane passing through the origin (0, 0, 0). By **Definition 9.1.1** or **Theorem 9.1.1**, one can show that any plane passing through the origin (0, 0, 0) is a subspace of $\mathbb{R}^3$.

The equation of a plane passing through the origin (0, 0, 0) is of the form

$$ax + by + cz = 0,$$

where at least one of *a*, *b*,and *c* is not zero.

Example 9.1.5.

Show that *W* is a subspace of $\mathbb{R}^3$, where

$$W = \left\{(x,\ y,\ z) \in \mathbb{R}^3 : \begin{cases} 3x - 2y + 3z = 0, \\ x - 3y + 5z = 0 \end{cases}\right\}.$$

Solution

Let $A = \begin{pmatrix} 3 & -2 & 3 \\ 1 & -3 & 5 \end{pmatrix}$, $\vec{X} = (x,\ y,\ z)^T$, and $\vec{0} = (0,\ 0,\ 0)^T$. Then

$$W = \left\{(x,\ y,\ z) \in \mathbb{R}^3 : \begin{cases} 3x - 2y + 3z = 0, \\ x - 3y + 5z = 0 \end{cases}\right\} = \left\{(x,\ y,\ z) \in \mathbb{R}^3 : A\vec{X} = \vec{0}\right\}.$$

By **Theorem 9.1.1**, W is a subspace of $\mathbb{R}^3$.

Example 9.1.6.

Show that W is not a subspace of $\mathbb{R}^3$, where

$$W = \left\{(x,\ y,\ z) \in \mathbb{R}^3 : x + 2y + 3z = 1\right\}.$$

Solution

Note that $(x,\ y,\ z) \in W$ if and only if $x + 2y + 3z = 1$. Because $0 + 2(0) + 3(0) \neq 1$, so $(0,\ 0,\ 0) \notin W$ and W is not a subspace of $\mathbb{R}^3$.

Note that $x + 2y + 3z = 1$ represents the equation of the plane that does not pass through the origin (0, 0). In general, one can show that any plane that does not pass through the origin (0, 0, 0) is not a subspace of $\mathbb{R}^3$. The equation of a plane that does not pass through the origin (0, 0) is of the form

$$ax + by + cz = d,$$

where at least one of *a*, *b*, and *c* is not zero, and $d \neq 0$.

Recall that the equation of a line in $\mathbb{R}^3$ is of the form

(9.1.2)

$$\begin{cases} x = x_0 + ta \\ y = y_0 + tb \qquad -\infty < t < \infty \\ z = z_0 + tc, \end{cases}$$

see **Theorem 6.2.1** . Let $\vec{X} = (x, y, z)$, $\overrightarrow{P_0} = (x_0, y_0, z_0)$, and $\vec{v} = (a, b, c)$. Then **(9.1.2)** can be written as

$$\vec{X} = \overrightarrow{P_0} + t\vec{v}, \quad -\infty < t < \infty.$$

If the line does not pass through the origin, then $\overrightarrow{P_0} \neq (0, 0, 0)$. If the line passes through the origin, then $\overrightarrow{P_0} = (0, 0, 0)$ and the equation of the line becomes

$$\vec{X} = t\vec{v}, \quad -\infty < t < \infty.$$

The following example shows that any lines passing through the origin (0, 0, 0) are subspaces of $\mathbb{R}^3$.

Example 9.1.7.

Let $\vec{v} \in \mathbb{R}^3$ be a given nonzero vector. Show that

$$W = \operatorname{span} \vec{v} = \left\{ t\vec{v} : t \in \mathbb{R} \right\}$$

is a subspace of $\mathbb{R}^3$.

Solution

Note that $\vec{a} \in W$ if and only if there exists $t \in \mathbb{R}$ such that $\vec{a} = t\vec{v}$. Let $\vec{a} \in W$, $\vec{b} \in W$, and $k \in \mathbb{R}$. Then there exist t_1, $t_2 \in \mathbb{R}$ such that $\vec{a} = t_1\vec{v}$ and $\vec{b} = t_2\vec{v}$. Hence,

$$\vec{a} + \vec{b} = t_1\vec{v} + t_2\vec{v} = (t_1 + t_2)\vec{v} \in W$$

and

$$k\vec{a} = k(t_1\vec{v}) = (kt_1)\vec{v} \in W.$$

By **Definition 9.1.1**, W is a subspace of $\mathbb{R}^3$.

Similar to $\mathbb{R}^2$, all the subspaces of $\mathbb{R}^3$ contain the origin {(0, 0, 0)}, the entire space $\mathbb{R}^3$, all the planes passing through the origin (0, 0, 0), and all the lines passing through the origin (0, 0, 0). Any other subsets of $\mathbb{R}^3$ are not subspaces of $\mathbb{R}^3$.

In **Example 9.1.7**, we note that $W = \left\{t\vec{v} : t \in \mathbb{R}\right\} = \operatorname{span}\left\{\vec{v}\right\}$. Hence, by **Example 9.1.7** the spanning space of $\vec{v}$ is a subspace of $\mathbb{R}^3$. The following result shows that spanning spaces in $\mathbb{R}^n$ are subspaces of $\mathbb{R}^n$.

Theorem 9.1.3.

Let $\overrightarrow{a_1}, \overrightarrow{a_2}, \cdots, \overrightarrow{a_n} \in \mathbb{R}^m$. *Show that* $\operatorname{span}\left\{\overrightarrow{a_1}, \overrightarrow{a_2}, \cdots, \overrightarrow{a_n}\right\}$ *is a subspace of* $\mathbb{R}^m$.

Proof

Let $W = \operatorname{span}\left\{\overrightarrow{a_1}, \overrightarrow{a_2}, \cdots, \overrightarrow{a_n}\right\}$. Note that $\vec{u} \in W$ if and only if there exist $x_1, x_2, \cdots, x_n \in \mathbb{R}$ such that

$$\vec{u} = x_1\overrightarrow{a_1} + x_2\overrightarrow{a_2} + \cdots + x_n\overrightarrow{a_n}.$$

Let $\vec{u} \in W$, $\vec{v} \in W$, and $k \in \mathbb{R}$. Then there exist $x_1, x_2, \cdots, x_n \in \mathbb{R}$ and $y_1, y_2, \cdots, y_n \in \mathbb{R}$ such that

$$\vec{u} = x_1\overrightarrow{a_1} + x_2\overrightarrow{a_2} + \cdots + x_n\overrightarrow{a_n} \text{ and } \vec{v} = y_1\overrightarrow{a_1} + y_2\overrightarrow{a_2} + \cdots + y_n\overrightarrow{a_n}.$$

Hence,

$$\begin{aligned} \vec{u} + \vec{v} &= x_1\overrightarrow{a_1} + x_2\overrightarrow{a_2} + \cdots + x_n\overrightarrow{a_n} + y_1\overrightarrow{a_1} + y_2\overrightarrow{a_2} + \cdots + y_n\overrightarrow{a_n} \\ &= (x_1 + y_1)\overrightarrow{a_1} + (x_2 + y_2)\overrightarrow{a_2} + \cdots + (x_n + y_n)\overrightarrow{a_n} \in W \end{aligned}$$

and

$$k\vec{u} = (kx_1)\overrightarrow{a_1} + (kx_2)\overrightarrow{a_2} + \cdots + (kx_n)\overrightarrow{a_n} \in W.$$

This shows that *W* is a subspace of $\mathbb{R}^n$.

Exercises

1. Let $W = \left\{(x, y) \in \mathbb{R}^2 : x - y = 0\right\}$. Show that *W* is a subspace of $\mathbb{R}^2$.
2. Let $W = \left\{(x, y) \in \mathbb{R}^2_+ : x + y \leq 1\right\}$. Show that *W* is not a subspace of $\mathbb{R}^2$.
3. Let

$$\mathbb{R}^2_+ = \left\{(x,\ y): x \geq 0 \quad \text{and } y \geq 0\right\}.$$

i. Show that if $\vec{a} \in \mathbb{R}^2_+$ and $\vec{b} \in \mathbb{R}^2_+$, then $\vec{a} + \vec{b} \in \mathbb{R}^2_+$.

ii. Show that $\mathbb{R}^2_+$ is not a subspace of $\mathbb{R}^2$ by using **Definition 9.1.1** .

4. Let

$$W = \mathbb{R}^2_+ \cup (-\mathbb{R}^2_+).$$

i. Show that if $\vec{a} \in W$ and k is a real number, then $k\vec{a} \in W$.

ii. Show that W is not a subspace of $\mathbb{R}^2$ by using **definition 9.1.1** .

5. Let $W = \left\{(x,\ y,\ z) \in \mathbb{R}^3 : x - 2y - z = 0\right\}$. Show that W is a subspace of $\mathbb{R}^3$ by **Definition 9.1.1** and **Theorem 9.1.1** , respectively.

6. Show that W is a subspace of $\mathbb{R}^4$, where

$$W = \left\{(x,\ y,\ z,\ w) \in \mathbb{R}^4 : \begin{cases} x - 2y + 2z - w = 0, \\ -x + 3y + 4z + 2w = 0 \end{cases}\right\}.$$

7. Show that W is a subspace of $\mathbb{R}^3$, where

$$W = \left\{(x,\ y,\ z) \in \mathbb{R}^3 : \begin{cases} x + 2y + 3z = 0, \\ -x + y - 4z = 0, \\ x - 2y + 5z = 0 \end{cases}\right\}.$$

8. Let $W = \left\{(x,\ y,\ z) \in \mathbb{R}^3 : x - 2y - z = 3\right\}$. Show that W is not a subspace of $\mathbb{R}^3$ by **Definition 9.1.1** .
9. Let $W = [0,\ 1]$. Show that W is not a subspace of R.
10. Show that all the subspaces of $\mathbb{R}$ are {0} and $\mathbb{R}$.
11. Let $W = \left\{(0,\ 0,\ \cdots,\ 0)\right\} \in \mathbb{R}^n$. Show that W is a subspace of $\mathbb{R}^n$.
12. Let $\vec{v} \in \mathbb{R}^n$ be a given nonzero vector. Show that

$$W = \left\{t\vec{v} : t \in \mathbb{R}\right\}$$

is a subspace of $\mathbb{R}^n$ by **Definition 9.1.1** .

9.2 Vector Spaces

Before we give the definition of a vector space, we introduce the notion of a set.

Definition 9.2.1.

A set is a collection of objects.

We always use capital letters such as *A, B, C, V, W, X, Y, Z* to denote sets. Each object in a set is called an element. We denote an element by small letters such as *a, b, c, v, w, x, y, z*. We use the symbol $x \in A$ to denote “*x* belongs to *A*” or ”*x* is in *A*” and the symbol $x \notin A$ to denote “*x* does not belong to *A*”or”*x* is not in *A*.” When a set *A* is a collection of objects that satisfies some property *P*, we write

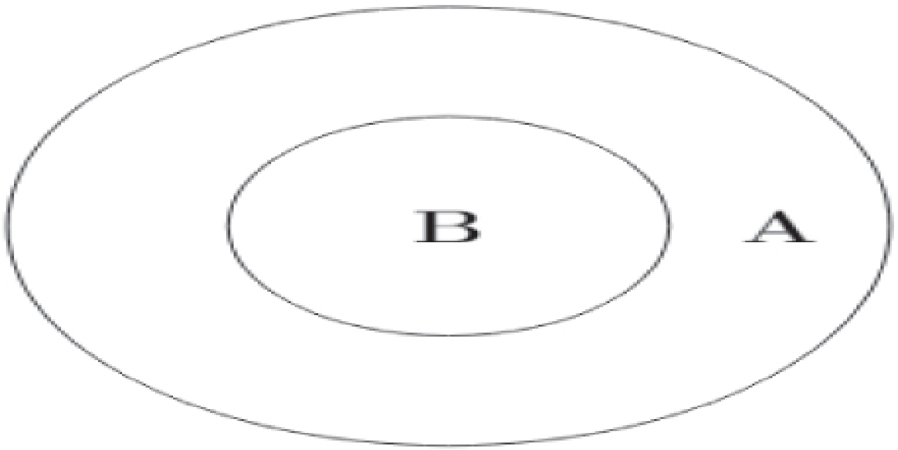
$$A = \{x : x \text{ has the property } P\}.$$

Example 9.2.1.

1. $\mathbb{N} = \{1,\ 2,\ \cdots\},$ the set of natural numbers.
2. $\mathbb{Z} = \{\cdots,\ -2,\ -1,\ 0,\ 1,\ 2,\ \cdots\},$ the set of integers.
3. $\mathbb{R} = (-\infty,\ \infty),$ the set of real numbers.
4. $\mathbb{R}^n = \{(x_1,\ x_2,\ \cdots,\ x_n) : x_i \in \mathbb{R},\quad i = 1,\ 2,\ \cdots,\ n\}.$
5. $\mathbb{Q} = \left\{\frac{p}{q} : p,\ q \in \mathbb{Z},\ q \neq 0\right\},$ the set of rational numbers;
6. $A = \left\{(x,\ y) : x^2 + y^2 = 1\right\}.$
7. $V = \{A_{mn} : A_{mn} \text{ is an } m \times n \text{ matrix}\},$ the set of all $m \times n$ matrices.
8. $C[a,\ b] = \{f : f : [a,\ b] \to \mathbb{R} \quad \text{is a continuous function}\},$ the set of all the continuous functions defined on [*a*, *b*].
9. $V = \{f : f : \mathbb{R} \to \mathbb{R}\text{is a function}\},$ the set of all the functions defined on R.

We use $B \subset A$ to denote that *A* contains *B* or *B* is a subset of *A*, see **Figure 9.1** . It is easy to see that $\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R}$.

Figure 9.1: Subset



A set is a collection of objects in which there may be no operations. If we introduce suitable operations in a set, then such a set is more useful. For example, we introduce addition $+$ and scalar multiplication $\cdot$ together with dot product into $\mathbb{R}^n$. We use the addition $+$ and scalar multiplication $\cdot$ in $\mathbb{R}^n$ to study the linear combination of vectors, spanning space, linear independence, bases, linear transformations, and eigenvalues.

Recall some sets in which we introduce suitable addition and scalar multiplication.

1. In $\mathbb{R}^n$: for $\overrightarrow{a} = (a_1, \ \cdots, a_n)$ and $\overrightarrow{b} = (b_1, \ \cdots, b_n)$, and $k \in \mathbb{R}$, the addition $+$ and scalar multiplication $\cdot$ are defined by
   1. $\overrightarrow{a} + \overrightarrow{b} = (a_1 + b_1, \ \cdots, a_n + b_1)$,
   2. $k \cdot \overrightarrow{a} = (ka_1, \ \cdots, ka_n)$.

2. For $m \times n$ matrices, the addition $+$ and scalar multiplication $\cdot$ are defined as follows. Let $A = (a_{ij}), \ B = (b_{ij})$, and $k \in \mathbb{R}$,
   1. $A + B = (a_{ij} + b_{ij})$,
   2. $k \cdot A = (ka_{ij})$.

3. In the set of all the functions defined on R:

$$V = \{f : f : \mathbb{R} \to \mathbb{R} \quad \text{is a function}\},$$

 the addition and scalar multiplication in *V* are defined as follows:
 1. For $f,\ g \in V,\ (f+g)(x) = f(x) + g(x)$ and $x \in \mathbb{R}$.
 2. For $k \in \mathbb{R}$ and $f \in V,\ (k \cdot f) = kf(x)$ for $x \in \mathbb{R}$.

Let *V* be a nonempty set with addition $+$ and scalar multiplication $\cdot$. Each element in *V* is called a vector. We denote a vector by small letters in bold like $\mathbf{u}$. When $V = \mathbb{R}^n$, then $\mathbf{u} = \overrightarrow{u}$.

Definition 9.2.2.

Let *V* be a nonempty set with addition $+$ and scalar multiplication $\cdot$.. The set *V* is called a vector space if the addition and scalar multiplication satisfy the following ten axioms:

1. If $\mathbf{u},\ \mathbf{v} \in V$, then $\mathbf{u} + \mathbf{v} \in V$	[closure under addition]
2. $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$	[commutative law of vector addition]
3. $\mathbf{u} + (\mathbf{v} + \mathbf{w}) = (\mathbf{u} + \mathbf{v}) + \mathbf{w}$	[associative law of vector addition]
4. There is a vector $\mathbf{0} \in V$ such that $0 + \mathbf{u} = \mathbf{u} + 0 = \mathbf{u}$ for each $\mathbf{u} \in V$.	[**0** is called the additive identity]
5. If $\mathbf{u} \in V$, there is a $-\mathbf{u} \in V$, such that $\mathbf{u} + (-\mathbf{u}) = 0$.	[$-\mathbf{u} \in V$ is called the additive

	inverse of $\mathbf{u}$]
6. If $k \in \mathbb{R}$ and $\mathbf{u} \in V$, then $k\mathbf{u} \in V$.	[closure under scalar multiplication]
7. $k(\mathbf{u} + \mathbf{v}) = k\mathbf{u} + k\mathbf{v}$	[first distributive law]
8. $(k + l)\mathbf{u} = k\mathbf{u} + l\mathbf{u}$	[second distributive law]
9. $k(l\mathbf{u}) = (kl)\mathbf{u}$	[associative law of scalar multiplication]
10. $1\mathbf{u} = \mathbf{u}$	[the scalar 1 is called a multiplicative identity]

Remark 9.2.1.

When we say that *V* is a vector space, we mean:

1. *V* is a set.
2. In V, there are two operations: addition and scalar multiplication.
3. The two operations satisfy the ten axioms.

Remark 9.2.2.

Let *V* be a vector space, $\mathbf{u} \in V$, and $k \in \mathbb{R}$. Then

a. $0\mathbf{u} = \mathbf{0}$;
b. $k\mathbf{0} = \mathbf{0}$;

c. $k\mathbf{u} = \mathbf{0}$, then $k = 0$ or $\mathbf{u} = 0$.

Now, we provide several examples of vector spaces. The first one is the *n*-dimensional Euclidean space.

Example 9.2.2.

The *n*-dimensional Euclidean space $\mathbb{R}^n$ is a vector space with the standard addition $+$ and scalar multiplication $\cdot$ given in **Definition 1.1.5** .

Solution

Let $\overrightarrow{u} = (u_1, u_2, \cdots, u_n) \in \mathbb{R}^n$, $\overrightarrow{v} = (v_1, v_2, \cdots, v_n) \in \mathbb{R}^n$ and $k \in \mathbb{R}$. Then

$$\overrightarrow{u} + \overrightarrow{v} = (u_1 + v_1, u_2 + v_2, \cdots, u_n + v_n) \in \mathbb{R}^n$$

and

$$k\overrightarrow{u} = (ku_1, ku_2, \cdots, ku_n) \in \mathbb{R}^n.$$

Hence, the axioms 1 and 6 hold. It is easy to verify that the other eight axioms given in **Definition 9.2.2** hold. Hence, $\mathbb{R}^n$ is a vector space with the standard operations: $+$ and $\cdot$.

Example 9.2.3.

Let

$$V = \{M_{22} : M_{22} \text{ is a } 2 \times 2 \text{ matrix}\}.$$

Show that *V* is a vector space with the standard matrix addition and matrix scalar multiplication given in **Definition 2.1.4** .

Solution

$$\mathbf{u} = \begin{pmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \end{pmatrix} \in V, \quad \mathbf{v} = \begin{pmatrix} v_{11} & v_{12} \\ v_{21} & v_{22} \end{pmatrix} \in V, \quad \text{and} \quad k,\ l \in \mathbb{R}.$$

We verify that V satisfies the ten axioms given in **Definition 9.2.2** .

1. $\mathbf{u} + \mathbf{v} = \begin{pmatrix} u_{11} + v_{11} & u_{12} + v_{12} \\ u_{21} + v_{21} & u_{22} + v_{22} \end{pmatrix} \in V.$
2. $\mathbf{u} + \mathbf{v} = \begin{pmatrix} u_{11} + v_{11} & u_{12} + v_{12} \\ u_{21} + v_{21} & u_{22} + v_{22} \end{pmatrix} = \begin{pmatrix} v_{11} + v_{11} & v_{12} + v_{12} \\ v_{21} + v_{21} & v_{22} + v_{22} \end{pmatrix} = \mathbf{v} + \mathbf{u}.$
3. If $\mathbf{w} \in V$, then $\mathbf{u} + (\mathbf{v} + \mathbf{w}) = (\mathbf{u} + \mathbf{v}) + \mathbf{w}.$
4. Let $\mathbf{0} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$. Then

$$\begin{aligned} \mathbf{0} + \mathbf{u} &= \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \end{pmatrix} = \begin{pmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \end{pmatrix} = \mathbf{u} \\ \mathbf{u} + \mathbf{0} &= \begin{pmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \end{pmatrix} = \mathbf{u}. \end{aligned}$$

5. $-\mathbf{u} = \begin{pmatrix} -u_{11} & -u_{12} \\ -u_{21} & -u_{22} \end{pmatrix} \in V.$
6. $k\mathbf{u} = \begin{pmatrix} ku_{11} & ku_{12} \\ ku_{21} & ku_{22} \end{pmatrix} \in V.$
7. $$\begin{aligned} k(\mathbf{u} + \mathbf{v}) &= k \begin{pmatrix} u_{11} + v_{11} & u_{12} + v_{12} \\ u_{21} + v_{21} & u_{22} + v_{22} \end{pmatrix} = \begin{pmatrix} ku_{11} + kv_{11} & ku_{12} + kv_{12} \\ ku_{21} + kv_{21} & ku_{22} + kv_{22} \end{pmatrix} \\ &= \begin{pmatrix} ku_{11} & ku_{12} \\ ku_{21} & ku_{22} \end{pmatrix} + \begin{pmatrix} kv_{11} & kv_{12} \\ kv_{21} & kv_{22} \end{pmatrix} = k\mathbf{u} + k\mathbf{v}. \end{aligned}$$
8. $(k + l)\mathbf{u} = k\mathbf{u} + l\mathbf{u}.$

9. $k(l\mathbf{u}) = (kl)\mathbf{u}$.
10. $1\mathbf{u} = \mathbf{u}$.

Hence, *V* is a vector space.

By a method similar to **Example 9.2.3**, one can prove the following result.

Example 9.2.4.

Let *V* be the set of all $m \times n$ matrices, that is,

$$V = \{M_{mn} : M_{mn} \text{ is an } m \times n \text{ matrix}\}$$

with the standard matrix addition and matrix scalar multiplication given in **Definition 2.1.4**. Then *V* is a vector space.

Example 9.2.5.

Let $V = \{f : f : \mathbb{R} \to \mathbb{R} \quad \text{is a function}\}$. The addition and scalar multiplication in *V* are defined as follows:

i. For $f,\ g \in V,\ (f+g)(x) = f(x) + g(x)$ for each $x \in \mathbb{R}$.
ii. For $k \in \mathbb{R}$ and $f \in V,\ (k \cdot f)(x) = kf(x)$ for each $x \in \mathbb{R}$.

Then *V* is a vector space.

Solution

It is obvious that if $f,\ g \in V,$ then $f + g \in V,$ and if $k \in \mathbb{R}$ and $f \in V$, then $kf \in V$. Hence, the axioms (1) and (6) hold. Let $\hat{0}$ denote the zero function defined on $\mathbb{R}$, that is, $\hat{0}(x) = 0$ for each $x \in \mathbb{R}$. Then $\hat{0} + f = f + \hat{0} = f$ for each $f \in V$ and the axiom (4) holds. It is easy to see that other axioms hold.

Example 9.2.6.

In $\mathbb{R}^2$, we introduce the following addition and scalar multiplication:

1. For $\overrightarrow{u} = (u_1,\ u_2),\ \overrightarrow{v} = (v_1,\ v_2),\ \overrightarrow{u} + \overrightarrow{v} = (u_1 + v_1,\ u_2 + v_2)$.
2. For $k \in \mathbb{R}$ and $\overrightarrow{u} = (u_1,\ u_2),\ k \cdot \overrightarrow{u} = (ku_1,\ 0)$.

Then *V* is not a vector space under the above two operations.

Solution

The addition given in (1) is the standard addition, but the scalar multiplication is not the standard scalar multiplication. The axiom (10) does not hold for the scalar multiplication given in (2). For example, let $\overrightarrow{u} = (2,\ 3)$, then by the scalar multiplication given in (2), we have

$$1 \cdot \overrightarrow{u} = (1(2),\ 0) = (2,\ 0) \neq (2,\ 3) = \overrightarrow{u}.$$

Hence, *V* is not a vector space under the two operations (1) and (2).

In a vector space *V*, there are two operations: addition and scalar multiplication, which satisfy 10 axioms. Hence, similar to **Section 1.4**, we can introduce linear combinations, spanning spaces, and similar to Sections 7.1, 7.2 and 7.4, we can introduce linear independence, bases, coordinates. Also, we can introduce linear transformations from one vector space to another, and eigenvalues and eigenvectors of linear transformations. We can generalize the dot product of two vectors and the norm of a vector from $\mathbb{R}^n$ to vector spaces and study some related problems such as those in Sections 1.7, 1.8, and 8.5. We do not discuss these further here in this book. We only generalize the notion of subspaces in $\mathbb{R}^n$ in **Section 10.1** to vector spaces.

Definition 9.2.3.

Let *V* be a vector space and let $W \subset V$ be a nonempty subset of *V*. *W* is called a subspace of *V* if *W* is itself a vector space under the addition and multiplication defined on *V*.

To show that W is a subspace of V, according to **Definition 9.2.3**, we would verify that the addition and multiplication defined on V satisfy the ten axioms given in **Definition 9.2.2**, where V is replaced by W. However, in the following, we show that it suffices to verify that the axioms (1) and (6) hold.

Checking the axioms (1) to (10), we see that each of the six axioms (2), (3), (7)-(10) only depends on the addition and multiplication while the other four axioms (1), (4), (5), and (6) depend on the operations and the set involved. Therefore, if V is a vector space and we consider a nonempty subset W of V under the same addition and multiplication as V, then the six axioms (2), (3), (7)-(10) hold for W because they only depend on addition and multiplication. Therefore, to verify that W is a subspace of V, we only need to verify that the addition and multiplication satisfy the four axioms (1), (4), (5), and (6). The following result shows that if the axioms (1) and (6) hold, then the axioms (4) and (5) hold. Hence, we only need to verify that the axioms (1) and (6) hold to show that W is a subspace of V.

Theorem 9.2.1.

Let V be a vector space and W be a nonempty subset of V. Then W is a subspace of V if and only if the following conditions hold.

a. *if* $\mathbf{u},\ \boldsymbol{v} \in \mathbf{W},$ *then* $\mathbf{u}+\boldsymbol{v} \in \mathbf{W}$.
b. *if* $\mathbf{u} \in \mathbf{W},\ k \in \mathbb{R},$ *then* $k\mathbf{u} \in \mathbf{W}$.

Proof

It is sufficient to show that under (*a*) and (*b*), the axioms (4), (5) to W hold. Indeed, let $k=0$ and $\mathbf{u} \in W.$ Because V is a vector space $k\mathbf{u}=0\mathbf{u}=\mathbf{0}.$ By (*b*), $k\mathbf{u}=\mathbf{0} \in W.$ Therefore, $\mathbf{0} \in W.$ Because V is a vector space, $\mathbf{0}+\mathbf{u}=\mathbf{u}+\mathbf{0}=\mathbf{u}$ for each $\mathbf{u} \in W.$ Hence, the axiom (4) holds. By (*b*), $-\mathbf{u}=(-1)\mathbf{u} \in W$ for each $\mathbf{u} \in W.$ Because V is a vector space, $\mathbf{u}+(-\mathbf{u})=(-\mathbf{u})+\mathbf{u}=\mathbf{0}.$ Hence, the axiom (5) holds.

In **Theorem 9.2.1**, if $V=\mathbb{R}^n,$ then **Theorem 9.2.1** is the same as **Definition 9.1.1**. We refer to **Section 9.1** for the discussion about subspaces in $\mathbb{R}^n.$ In the following, we give examples of subspaces that are not in $\mathbb{R}^n$.

Example 9.2.7.

Let

$$W = \{A : A \text{ is an } n \times n \text{ symmetric matrix}\}.$$

Show that *W* is a vector subspace of $M_{n\times n}$, where $M_{n\times n}$ is the vector space of all $n \times n$ matrices.

Solution

By **(2.3.3)**, *A* is symmetric if and only if $A^T = A$. Let $A,\ B \in W$. Then $A = A^T$ and $B = B^T$. By **Theorem 2.1.1**,

a. $(A+B)^T = A^T + B^T = A + B,$ thus $A + B \in V$;
b. $(kA)^T = kA^T = kA,$ thus $kA \in V$.

It follows that *V* is a subspace of $M_{n\times n}$.

Exercises

1. Let

$$C(\mathbb{R}) = \{f : f : \mathbb{R} \to \mathbb{R} \quad \text{is continuous on } \mathbb{R}\}.$$

Show that $C(\mathbb{R})$ is a vector space under the following operations.

1. $+ : (f+g)(x) = f(x) + g(x)$ for $x \in \mathbb{R}$.
2. $(kf)(x) = k(f(x))$ for $x \in \mathbb{R}$.

2. Let

$$C[a,\ b] = \{f : f : [a,\ b] \to \mathbb{R} \text{ is continuous}\}.$$

Show that *C*[*a, b*] is a vector space under the standard addition and scalar multiplication:

1. For $f,\ g \in C[a,\ b],\ (f+g)(x) = f(x) + g(x)$ for $x \in [a,\ b]$.
2. For $k \in \mathbb{R}$ and $f \in C[a,\ b],\ (k \cdot f)(x) = kf(x)$ for $x \in [a,\ b]$

3. Let *n* be a nonnegative integer and let

$$P_n = \{p : p(x) = a_0 + a_1 x + \cdots + a_n x^n \quad \text{is a polynomial of degree } n\}.$$

Show that P_n is a subspace of $C(\mathbb{R})$.

4. Let

$$V = \{A : A \quad \text{is a } 3 \times 3 \text{ matrix}\}.$$

We define the following addition and scalar multiplication.

1. Addition: For $A,\ B \in V,\ A + B = AB,$ where *AB* is the standard product of *A* and *B* in **Section 2.5** .
2. Scalar multiplication: For $A \in V$ and $k \in \mathbb{R},\ k \cdot A = kA,$ where *kA* is the standard scalar multiplication given in **Definition 2.1.4** .

Then *V* is not a vector space under the above addition and scalar multiplication.

5. Let

$$C^1[a,\ b] = \{f : f' \in C[a,\ b]\},$$

where $f'(x)$ denotes the derivative of *f* at *x*. Show that $C^1[a,\ b]$ is a vector subspace of *C*[*a, b*].

Chapter **10** Complex numbers

10.1 Complex numbers

We want to solve the following quadratic equation

(10.1.1)

$$a\lambda^2 + b\lambda + c = 0,$$

where *a, b, c* are real numbers and $a \neq 0.$ Multiplying **(10.1.1)** by 4*a* implies

$$4a^2\lambda^2 + 4ab\lambda + 4ac = 0.$$

Applying completing the square to the above equation implies

(10.1.2)

$$(2a\lambda + b)^2 = b^2 - 4ac.$$

If $\Delta := b^2 - 4ac \geq 0,$ then

$$2a\lambda + b = \sqrt{b^2 - 4ac} \quad \text{or } 2a\lambda + b = -\sqrt{b^2 - 4ac}.$$

It follows that

(10.1.3)

$$\lambda_1 := \frac{-b+\sqrt{b^2-4ac}}{2a} \quad \text{and } \lambda_2 := \frac{-b-\sqrt{b^2-4ac}}{2a}$$

are all the real roots of the quadratic equation **(10.1.1)** .

If $\Delta < 0$, then **(10.1.1)** has no real roots. Hence, when $\Delta < 0$, we need to extend the set of real numbers R to a larger set denoted by $\mathbb{C}$ such that **(10.1.1)** still has roots in the larger set. What is the set $\mathbb{C}$?

If $\Delta < 0$, then $-(b^2 - 4ac) > 0$ and $\sqrt{-(b^2 - 4ac)}$ is a real number. We introduce a symbol

(10.1.4)

$$i = \sqrt{-1}.$$

Then

$$\sqrt{\Delta} = \sqrt{(-1)[-(b^2 - 4ac)]} = \sqrt{-1}\sqrt{-(b^2 - 4ac)} = i\sqrt{-(b^2 - 4ac)}.$$

By **(10.1.3)** , we obtain two roots of the quadratic equation **(10.1.1)** :

(10.1.5)

$$\lambda_1 = \frac{-b+i\sqrt{-(b^2-4ac)}}{2a} \quad \text{and } \lambda_2 = \frac{-b-i\sqrt{-(b^2-4ac)}}{2a}.$$

Hence, the quadratic equation **(10.1.1)** always has two roots: λ_1 and λ_2 given in **(10.1.5)** .

Example 10.1.1.

Find the roots of

$$\lambda^2 + 4\lambda + 13 = 0.$$

Solution

Because $a = 1,\ b = 4,\ c = 13,$

$$\sqrt{\Delta} = \sqrt{b^2 - 4ac} = \sqrt{4^2 - 4 \times 13} = \sqrt{-36} = \sqrt{-1}\sqrt{36} = 6i.$$

By **(10.1.5)** , $\lambda_1 = \frac{-4+6i}{2} = -2 + 3i$ and $\lambda_2 = \frac{-4-6i}{2} = -2 - 3i$.

By **(10.1.5)** , we see that λ_1 and λ_2 can be written into the following form

$$x + yi,$$

where *x* and *y* are real numbers. If $y \neq 0,$ then $x + yi$ is not a real number. Hence, we introduce

Definition 10.1.1.

The number of the form

$$z = x + yi$$

is called a complex number, where *x* and *y* are real numbers and *i* is given in **(10.1.4)** . *x* is called the real part of *z* and is denoted by $\mathrm{Re}\ z = x$ and *y* is called the imaginary part of *z* and is denoted by $\mathrm{Im}\ z = y$. $i = \sqrt{-1}$ is called the imaginary number and $i^2 = -1.$ We denote by C the set of all complex numbers.

Example 10.1.2.

Find $\mathrm{Re}(-1 + 2i)$ and $\mathrm{Im}(-1 - 2i)$.

Solution

By **Definition 10.1.1** , we have

$$\mathrm{Re}(-1+2i=1) \quad \text{and } \mathrm{Im}(-1-2i)=\mathrm{Im}(-2+i(-2))=-2.$$

The following definition shows that two complex numbers are equal if and only if their corresponding real parts and imaginary parts are the same.

Definition 10.1.2.

Let $z_1 = x_1 + y_1 i$ and $z_2 = x_2 + y_2 i$. Then $z_1 = z_2$ if and only if $x_1 = x_2$ and $y_1 = y_2$.

Example 10.1.3.

Use the following information to find *x* and *y*.

a. $3x + 4i = 6 + 2yi$
b. $(x+y) + (x-y)i = 5 - i$

Solution

a. By **Definition 10.1.2** , $3x = 6$ and $4 = 2y$. This implies that $x = y = 2$.
b. By **Definition 10.1.2** , we have $x + y = 5$ and $x - y = -1$. Solving the system, we obtain $x = 2$ and $y = 3$.

Operations on complex numbers

Definition 10.1.3.

Let $z_1 = x_1 + y_1 i$ and $z_2 = x_2 + y_2 i$.

i. (Addition) $z_1 + z_2 = (x_1 + x_2) + (y_1 + y_2)i$.
ii. (Subtraction) $z_1 - z_2 = (x_1 - x_2) + (y_1 - y_2)i$.
iii. (Multiplication) $z_1 \cdot z_2 = (x_1 x_2 - y_2 y_2) + (x_1 y_2 + y_1 x_2)i$.

iv. (Division) If $z_2 \neq 0$, then

$$\frac{z_1}{z_2} = \left(\frac{x_1x_2 + y_1y_2}{x_2^2 + y_2^2}\right) + \left(\frac{y_2x_2 - x_1y_2}{x_2^2 + y_2^2}\right) i.$$

Remark 10.1.1.

The multiplication (*iii*) can be obtained by the following way.

$$\begin{aligned} z_1z_2 &= (x_1 + y_1i)(x_2 + y_2i) = x_1(x_2 + y_2i) + (y_1i)(x_2 + y_2i) \\ &= x_1x_2 + x_1y_2i + x_2y_1i + y_1y_2i^2 = x_1x_2 + x_1y_2i + x_2y_1i - y_1y_2 \\ &= (x_1x_2 - y_1y_2) + (x_1y_2 + x_2y_1)i. \end{aligned}$$

Hence, we have the following **three formulas**.

1. $(x + yi)(x - yi) = x^2 - (yi)^2 = x^2 + y^2.$
2. $(x + yi)^2 = x^2 + 2x(yi) + (yi)^2 = x^2 + 2xyi - y^2 = (x^2 - y^2) + 2xyi.$
3. $(x - yi)^2 = x^2 - 2x(yi) + (yi)^2 = x^2 - 2xyi - y^2 = (x^2 - y^2) - 2xyi.$

Remark 10.1.2.

By using the above formula (1), the division (*iv*) can be obtained by the following way.

$$\begin{aligned} \frac{z_1}{z_2} &= \frac{x_1+y_1i}{x_2+y_2i} = \frac{(x_1+y_1i)(x_2-y_2i)}{(x_2+y_2i)(x_2-y_2i)} = \frac{x_1x_2-x_1y_2i+x_2y_1i-y_1y_2i^2}{(x_2)^2-(y_2i)^2} \\ &= \frac{(x_1x_2+y_1y_2)+(x_2y_1-x_1y_2)i}{x_2^2+y_2^2} = \frac{x_1x_2+y_1y_2}{x_2^2+y_2^2} + \frac{y_1x_2-x_1y_2}{x_2^2+y_2^2}i. \end{aligned}$$

Example 10.1.4.

Let $z_1 = 4 - 5i$ and $z_2 = -1 + 6i$. Express the following complex numbers in the form $x + yi$.

1. $z_1 + z_2$
2. $z_1 - z_2$
3. $3z_1$
4. $-z_1$
5. $z_1 z_2$
6. $\frac{z_1}{z_2}$

Solution

1. $z_1 + z_2 = (4-5i) + (-1+6i) = (4-1) + (-5+6)i = 3+i.$
2. $z_1 - z_2 = (4-5i) - (-1+6i) = (4-(-1)) + (-5-6)i = 5-11i.$
3. $3z_1 = 3(4-5i) = 3 \times 4 - 3 \times 5i = 12 - 15i.$
4. $-z_1 = -(4-5i) = -4+5i.$
5. $$\begin{aligned} z_1 z_2 &= (4-5i)(-1+6i) = 4(-1+6i) - (5i)(-1+6i) \\ &= -4 + 24i + 5i - 30i^2 = -4 + 29i + 30 = 26 + 29i. \end{aligned}$$
6. $$\begin{aligned} \frac{z_1}{z_2} &= \frac{4-5i}{-1+6i} = \frac{(4-5i)(-1-6i)}{(-1+6i)(-1-6i)} = \frac{-4-24i+5i-30}{(-1)^2-(6i)^2} \\ &= \frac{-34-19i}{1+36} = -\frac{34}{37} - \frac{19}{37}i. \end{aligned}$$

Definition 10.1.4.

Let $z = x + yi$. Then,

$$|z| = \sqrt{x^2 + y^2}$$

is called the modulus of *z* or the absolute value of *z*.

Example 10.1.5.

Let $z = 3 - 4i$. Find $|z|$.

Solution

$$\left|z\right| = \sqrt{3^2 + (-4)^2} = \sqrt{9+16} = \sqrt{25} = 5.$$

The following result gives some properties of the modulus of complex numbers.

Proposition 10.1.1.

Let $z,\ z_1,\ z_2 \in \mathbb{C}$. Then the following assertions hold.

1. $|z_1 z_2| z_1||z_2|$.
2. $|z^n| = |z|^n$ *for* $n \in \mathbb{N}$.
3. *If* $z_2 \neq 0$, *then* $\left|\frac{z_1}{z_2}\right| = \frac{|z_1|}{|z_2|}$.
4. $|\alpha z| = |\alpha||z|$ *for* $\alpha \in \mathbb{R}$.
5. $|z_1 + z_2| \leq |z_1| + |z_2|$.

Example 10.1.6.

Find the modulus of

$$\frac{(1+i)(1-2i)}{i(1-i)(2+3i)}.$$

Solution

$$\left|\frac{(1+i)(1-2i)}{i(1-i)(2+3i)}\right| = \frac{|1+i||1-2i|}{|i||1-i||2+3i|} = \frac{\sqrt{2}\sqrt{5}}{(1)\sqrt{2}\sqrt{13}} = \frac{\sqrt{65}}{13}.$$

Definition 10.1.5.

The complex number $z = x - yi$ is called the conjugate of $z = x + yi$. $\bar{z}$ reads “z bar.”

Example 10.1.7.

For each of the following complex numbers, find its conjugate.

$$z_1 = 3 + 2i; \quad z_2 = -4 - 2i; \quad z_3 = i; \quad z_4 = 4.$$

Solution

$\overline{z_1} = 3 - 2i,\ \overline{z_2} = -4 + 2i,\ \overline{z_3} = -i$ and $\overline{z_4} = 4.$

The conjugate numbers have the following properties.

Proposition 10.1.2.

Let $z,\ z_1,$ and z_2 be complex numbers. Then

i. $\bar{\bar{z}} = z.$
ii. $\bar{z} = z$ *if and only if z is a real number.*
iii. $\bar{z} = -z$ *if and only if z is a pure imaginary.*
iv. $\overline{z_1 + z_2} = \overline{z_1} - \overline{z_2}.$
v. $\overline{z_1 - z_2} = \overline{z_1} - \overline{z_2}.$
vi. $\overline{z_1 \cdot z_2} = \overline{z_1} \cdot \overline{z_2}$ *and* $\overline{\alpha z} = \alpha \bar{z}$ *for* $\alpha \in \mathbb{R}.$
vii. $\overline{\left(\frac{z_1}{z_2}\right)} = \frac{\overline{z_1}}{z_2}.$
viii. $\left|z^2\right| = z\bar{z}.$

Proof

We only prove (*ii*)and (*viii*).

ii. Assume that $\bar{z} = z$. Let $z = x + yi$. Then $\bar{z} = x - yi$ and

$$x + yi = x - yi.$$

This implies that $2yi = 0$ and $y = 0$. Hence, $z = x + 0i = x$ is a real number. Conversely, if *z* is a real number, that is, $z = x + 0i,$ where *x* is a real number, then $\bar{z} = x - 0i = x = z$.

viii. Let $z = x + yi$. Then $\bar{z} = x - yi$ and

$$z\bar{z} = (x + yi)(x - yi) = x^2 - xyi + xyi - y^2 i^2 = x^2 + y^2 = |z^2|.$$

Example 10.1.8.

Verify the following identities.

1. $\overline{\bar{z} + 7i} = z - 7i$
2. $\overline{\left(\frac{i+3\bar{z}}{i-3\bar{z}}\right)} = \frac{i-3z}{i+3z}$

Solution

1. $\overline{\bar{z} + 7i} = \bar{\bar{z}} + \overline{7i} = z - 7i.$
2. $\overline{\left(\frac{i+3\bar{z}}{i-3\bar{z}}\right)} = \frac{\overline{i+3\bar{z}}}{\overline{i-3\bar{z}}} = \frac{\bar{i}+\overline{3\bar{z}}}{\bar{i}-\overline{3\bar{z}}} = \frac{-i+3z}{-i-3z} = \frac{i-3z}{i+3z}.$

Example 10.1.9.

Let $p(\lambda) = \lambda^4 + 2\lambda^3 + \lambda^2 + 5\lambda + 6$. Show that if *z* is a root of $p(\lambda)$, then $\bar{z}$ is a root of $p(\lambda)$,.

Solution

We prove

(10.1.6)

$$\overline{p(z)} = p(\bar{z}) \quad \text{for } z \in \mathbb{C}.$$

Indeed, by **Proposition 10.1.2** , we have

$$\begin{aligned}\overline{p(z)} &= \overline{z^4+2z^3+z^2+5z+6} = \overline{z^4}+\overline{2z^3}+\overline{z^2}+\overline{5z}+\bar{6}\\ &= (\bar{z})^4+2(\bar{z})^3+(\bar{z})^2+5\bar{z}+6 = p(\bar{z}).\end{aligned}$$

If *z* is a root of $p(\lambda)$, then $p(z)=0$. By **(10.1.6)**, we have

$$p(\bar{z}) = \overline{p(z)} = \bar{0} = 0$$

and $\bar{z}$ is a root of $p(\lambda)$.

Similar to **Example 10.1.9**, we can show that if $z \in \mathbb{C}$ is a root of a polynomial of degree *n*, then its conjugate $\bar{z}$ is also a root of this polynomial. This implies that the quadratic equation **(10.1.1)** has either of two real roots, which could possibly be the same or two different complex roots.

Exercises

1. Solve each of the following equations.
   a. $z^2+z+1=0$
   b. $z^2-2z+5=0$
   c. $z^2-8z+25=0$

2. Use the following information to find *x* and *y*.
   a. $2x-4i=4+2yi$
   b. $(x-y)+(x+y)i=2-4i$

3. Let $z_1=2+i$ and $z_2=1-i$. Express each of the following complex numbers in the form $x+yi$, calculate its modulus, and find its conjugate.
   1. z_1+2z_2
   2. $2z_1-3z_2$
   3. z_1z_2

4. $\frac{z_1}{z_2}$

4. Express each of the following complex numbers in the form $x + yi$.
   a. $\overline{\left(\frac{i}{1-i}\right)}$
   b. $\frac{2-i}{(1-i)(2+i)}$
   c. $\frac{1+i}{i(1+i)(2-i)}$
   d. $\frac{1+i}{1-i} - \frac{2-i}{1+i}$

5. Find $i^2,\ i^3,\ i^4,\ i^5,\ i^6$ and compute $\left(-i - i^4 + i^5 - i^6\right)^{1000}$.
6. Let $z = \left(\frac{1-i}{1+i}\right)^8$. Calculate $z^{66} + 2z^{33} - 2$.
7. Find $z \in \mathbb{C}$
   a. $iz = 4 + 3i$
   b. $(1+i)\bar{z} = (1-i)z$

8. Find real numbers *x, y* such that

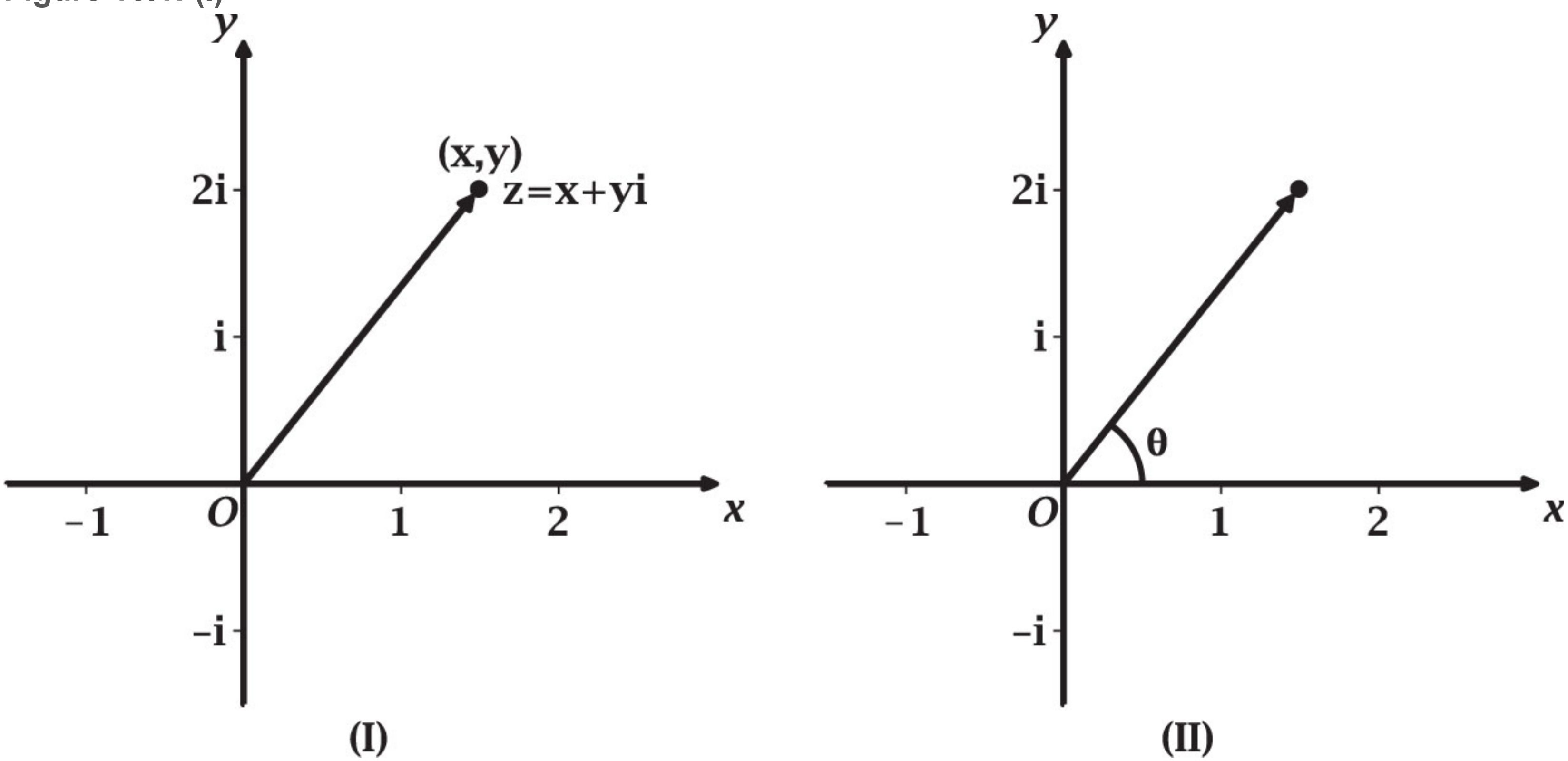
$$\frac{x + 1 + i(y-3)}{5 + 3i} = 1 + i.$$

9. Let $z = \frac{(3+4i)(1+i)^{12}}{i^5(2+4i)^2}$. Find $|z|$.

10.2 Polar and exponential forms

A complex number $z = x + yi$ can be interpreted as a point (*x, y*) or a vector $\begin{pmatrix} x \\ y \end{pmatrix}$ in the *xy*-plane (see **Figure 10.1** (I)).

Figure 10.1: (I)



When complex numbers are viewed in an *xy*-coordinate system, the *x*-axis is called the real axis, the *y*-axis is called the imaginary axis with unit *i*, and the plane is called the complex plane.

Definition 10.2.1.

Let $z = x + yi$ and $z \neq 0$. The angle θ from the positive *x*-axis to the vector $\begin{pmatrix} x \\ y \end{pmatrix}$ is called an argument of *z* and is denoted by

$$\theta = \arg z$$

(See **Figure 10.1** (II)). If θ satisfies

$$-\pi < \theta \leq \pi,$$

then θ is called the principal argument of *z* and is denoted by

$$\theta = \operatorname{Arg} z.$$

The principal argument Arg *z* of *z* is unique and

(10.2.1)

$$\arg z = \operatorname{Arg} z + 2k\pi \quad \text{for } k \in \mathbb{Z},$$

where $\mathbb{Z} = \{0, \ \pm 1, \ \pm 2, \ \cdots\}$ is the set of integers. **Definition 10.2.1** does not give the definitions of an argument and a principal argument for $z = 0$. Hence, if $z = 0$, then *z* does not have an argument nor a principal argument.

Theorem 10.2.1.

Let $x, \ y \in \mathbb{R}$ *be real numbers. Then the following assertions hold.*

1. *If* $z = x$ *and* $x > 0$, *then* $\operatorname{Arg} z = 0$ *and* $\arg z = 2k\pi$;
2. *If* $z = x$ *and* $x < 0$, *then* $\operatorname{Arg} z = \pi$ *and* $\arg z = \pi + 2k\pi$;

3. *If* $z = yi$ *and* $y > 0$, *then* $\mathrm{Arg}z = \frac{\pi}{2}$ *and* $\arg z = \frac{\pi}{2} + 2k\pi$; *and*
4. *If* $z = yi$ *and* $y < 0$, *then* $\mathrm{Arg}z = -\frac{\pi}{2}$ *and* $\arg z = -\frac{\pi}{2} + 2k\pi$, *where* $k \in \mathbb{Z}$.

Example 10.2.1.

For each of the following numbers, find its principal argument and argument.

a. 1
b. 2*i*
c. -1
d. $-2i$

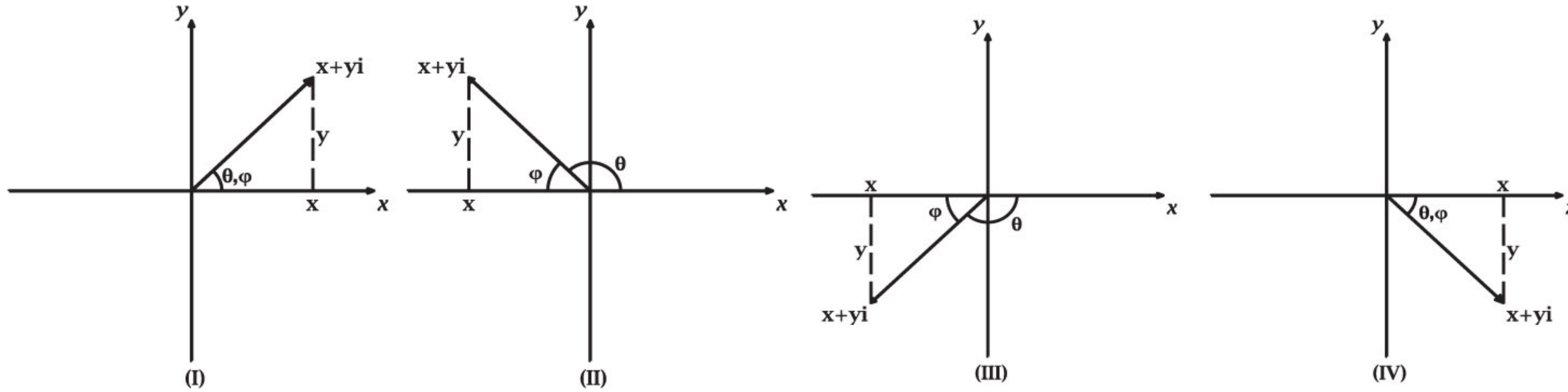
Solution

By **Theorem 10.2.1** , we have

a. $\mathrm{Arg}1 = 0$ and $\arg 1 = 2k\pi$ for $k \in \mathbb{Z}$.
b. $\mathrm{Arg}(2i) = \frac{\pi}{2}$ and $\arg(2i) = \frac{\pi}{2} + 2k\pi$ for $k \in \mathbb{Z}$.
c. $\mathrm{Arg}(-1) = \pi$ and $\arg(-1) = \pi + 2k\pi$ for $k \in \mathbb{Z}$.
d. $\mathrm{Arg}(-2i) = -\frac{\pi}{2}$ and $\arg(-2i) = -\frac{\pi}{2} + 2k\pi$ for $k \in \mathbb{Z}$.

The following result gives the formulas of the principal arguments of complex numbers (see **Figure 10.2** (I), (II), (III), (IV)).

Figure 10.2: (I), (II), (III), (IV)



Theorem 10.2.2.

Assume that $z = x + yi,\ x \neq 0,\ y \neq 0,$ *and* $\psi \in \left(0,\ \frac{\pi}{2}\right)$ *satisfy*

(10.2.2)

$$\tan\psi = \left|\frac{y}{x}\right|.$$

Then the following assertions hold.

1. *If z is in the first quadrant, then* $\mathrm{Arg}z = \psi$ *and* $\arg z = \psi + 2k\pi$.
2. *If z is in the second quadrant, then*

$$\mathrm{Arg}z = \pi - \psi \quad \textit{and}\ \arg z = \pi - \psi + 2k\pi.$$

3. *If z is in the third quadrant, then*

$$\mathrm{Arg}z = \ -\pi - \psi \quad \textit{and}\ \arg z = \ -\pi - \psi + 2k\pi.$$

4. *If z is in the fourth quadrant, then*

$$\text{Arg}z = -\psi \textit{ and } \arg z = -\psi + 2k\pi,$$

where $k \in \mathbb{Z}$.

Note that if $\psi \in \left(0, \frac{\pi}{2}\right)$, then

(10.2.3)

$$\tan\psi = \left|\frac{y}{x}\right| \quad \text{if and only if } \psi = \arctan\left|\frac{y}{x}\right|.$$

Example 10.2.2.

Find Arg *z* and arg *z* for each of the following numbers.

a. $z = 3 + \sqrt{3i}$

b. $z = -2 - 2i.$

Solution

a. Let $z = x + yi = 3 + \sqrt{3i}.$ Then $x = 3$ and $y = \sqrt{3}.$ Let $\psi \in \left(0, \frac{\pi}{2}\right)$

$$\tan\psi = \left|\frac{y}{x}\right| = \frac{\sqrt{3}}{3}.$$

Then $\psi = \frac{\pi}{6}.$ Because $x > 0$ and $y > 0,$ *z* is in the first quadrant and by **Theorem 10.2.2** (1),

$$\operatorname{Arg}z = \psi = \frac{\pi}{6} \quad \text{and } \arg z = \frac{\pi}{6} + 2k\pi \text{ for } k \in \mathbb{Z}.$$

b. Let $z = x + yi = -2 - 2i$. Then $x = -2$ and $y = -2$. Let $\psi \in \left(0, \frac{\pi}{2}\right)$ satisfy

$$\tan\psi = \left|\frac{-2}{-2}\right| = 1.$$

Then $\psi = \frac{\pi}{4}$. Because $x < 0$ and $y < 0$, z is in the third quadrant and by **Theorem 10.2.2** (3),

$$\operatorname{Arg}z = -\pi + \psi = -\pi + \frac{\pi}{4} = -\frac{3}{4}\pi \quad \text{and } \arg z = -\frac{3}{4}\pi + 2k\pi \text{ for } k \in \mathbb{Z}.$$

The symbol $e^{i\theta}$, or $\exp(i\theta)$, is defined by means of Euler's formula as

(10.2.4)

$$\cos\theta + i\sin\theta = e^{i\theta},$$

where θ is measured in radians.

Proposition 10.2.1.

1. $e^{i\theta} = 1$ *if and only if* $\theta = 2k\pi$ for $k \in \mathbb{Z}$.
2. $e^{i\theta} = e^{i\psi}$ if and only if $\theta = \psi + 2k\pi$ for $k \in \mathbb{Z}$.

Proof

1. Assume that $e^{i\theta} = 1$. Then $\cos\theta + i\sin\theta = 1$. This implies that $\cos\theta = 1$ and $\sin\theta = 0$. Hence, $k \in \mathbb{Z}$. Conversely, if $\theta = 2k\pi$ for $k \in \mathbb{Z}$, then

$$e^{i\theta} = e^{i(2k\pi)} = \cos 2k\pi + i\sin 2k\pi = 1.$$

2. It is easy to see that $e^{i\theta} = e^{i\psi}$ if and only if $e^{i(\theta-\psi)} = 1$ if and only if $\theta - \psi = 2k\pi$ for $k \in \mathbb{Z}$.

Theorem 10.2.3.

Let $z = x + yi$ and let θ be an argument of z. Then

$$z = \left|z\right|(\cos\theta + i\sin\theta) = \left|z\right|e^{i\theta}.$$

Proof

Because $x = |x|\cos\theta$ and $y = |z|\sin\theta$,

$$z = x + yi = |z|\cos\theta + (r\sin\theta)i = |z|(\cos\theta + i\sin\theta).$$

The second equality follows from Euler's formula **(10.2.4)** .

For each nonzero complex number $z = x + yi$, there exist a unique number $r = |z| > 0$ and arguments $\theta = \mathrm{Arg}z + 2k\pi$ for $k \in \mathbb{Z}$. We call (r, θ) the polar coordinates of the complex number $z = x + yi$ or the point (*x*, *y*). Conversely, for a polar coordinate (r, θ), we define a complex number *z* by

(10.2.5)

$$z = (r\cos\theta) + (r\sin\theta)i.$$

Then $|z| = r$ and θ is an argument of *z*. Indeed,

$$\left|z\right| = \sqrt{r^2\cos^2\theta + r^2\sin^2\theta} = \sqrt{r^2(\cos^2\theta + \sin^2\theta)} = \sqrt{r^2} = r,$$

where we use the identity: $\cos^2\theta + \sin^2\theta = 1$.

Definition 10.2.2.

Let *z* be a complex number. The following form

(10.2.6)

$$z = r(\cos\theta + i\sin\theta)$$

is called the polar form of z and the following form

(10.2.7)

$$z = re^{i\theta}$$

is called the exponential form of *z*.

It is obvious that if $r = 1$, then

$$z = \cos\theta + i\sin\theta = e^{i\theta}.$$

Such a complex number *z* is called a unit complex number.

By **Theorem 10.2.3** , we see that the polar form and the exponential form of a complex number *z* can be converted into one another easily. Moreover, by **(10.2.5)** , we see that a polar form of a complex number can be easily transferred into the standard form. However, it is not trivial to transfer the standard form of a complex number $z = x + yi$ into a polar form of *z*.

The following are the steps for finding a polar form of a complex number $z = x + yi$.

1. Step 1. Find $|z|$.
2. Step 2. Find Arg *z*.
3. Step 3. Write *z* as $z = r(\cos\theta + i\sin\theta)$, where $\theta = \text{Arg}z$.

In Steps 2 and 3, the principal argument Arg *z* can be replaced by the argument arg *z* of *z*. We always use Arg *z* instead of arg *z* if there are no indications.

Example 10.2.3.

Express $z = 1 - \sqrt{3}i$ in polar and exponential forms.

Solution

1. Step 1. $r = \left| z \right| = \sqrt{1^2 + (-\sqrt{3})^2} = \sqrt{4} = 2.$

2. Step 2. Let $\psi \in \left(0, \frac{\pi}{2}\right)$ such that $\tan\psi = \left|\frac{\sqrt{-3}}{1}\right| = \sqrt{3}.$ Then $\psi = \frac{\pi}{3}.$ Because $x = 1 > 0$ and $y = -\sqrt{3} < 0,$ z is in the second quadrant and by **Theorem 10.2.2** (2),

$$\text{Arg}z = \pi - \psi = \pi - \frac{\pi}{3} = \frac{2\pi}{3}.$$

3. Step 3. $z = 2\left(\cos\left(\frac{2\pi}{3}\right) + i\sin\left(\frac{2\pi}{3}\right)\right).$ Hence, the polar form of $1 - \sqrt{3}i$ is

$$1 - \sqrt{3}i = 2\left(\cos\frac{2\pi}{3} + i\sin\frac{2\pi}{3}\right)$$

and its exponential form is $1-\sqrt{3}i = 2e^{i\frac{2\pi}{3}}$.

Exercises

1. For each of the following numbers, find its principal argument and argument.
   a. -4
   b. $-i$
   c. $\frac{1}{2} - \frac{\sqrt{3}}{2}i$
   d. $1 + \sqrt{3}i$
   e. $-1 + \sqrt{3}i$
   f. $-\sqrt{3} - i$

2. Express each of the following complex numbers in polar and exponent forms.
   a. 1
   b. i
   c. -2
   d. $-3i$
   e. $1 + \sqrt{3}i$
   f. $-1 - i$

10.3 Products in polar and exponential forms

Theorem 10.3.1.

Let $z_1 = r_1(\cos\theta_1 + i\sin\theta_1)$ *and* $z_2 = r_2(\cos\theta_2 + i\sin\theta_2)$. *Then*

$$z_1 z_2 = r_1 r_2[\cos(\theta_1+\theta_2) + i\sin(\theta_1+\theta_2)] = r_1 r_2 e^{i(\theta_1+\theta_2)}$$

and if $z_2 \neq 0$, *then*

$$\frac{z_1}{z_2} = \frac{r_1}{r_2}[\cos(\theta_1-\theta_2) + i\sin(\theta_1-\theta_2)] = \frac{r_1}{r_2}e^{i(\theta_1-\theta_2)}.$$

Proof

$$\begin{aligned} z_1 z_2 &= [r_1(\cos\theta_1 + i\sin\theta_1)][r_2(\cos\theta_2 + i\sin\theta_2)] \\ &= r_1 r_2(\cos\theta_1 + i\sin\theta_1)(\cos\theta_2 + i\sin\theta_2) \\ &= r_1 r_2[(\cos\theta_1\ \cos\theta_2 - \sin\theta_1\ \sin\theta_2) + i(\sin\theta_1\ \cos\theta_2 + \cos\theta_1\ \sin\theta_2)] \\ &= r_1 r_2[\cos(\theta_1+\theta_2) + i\sin(\theta_1+\theta_2)] \\ &= r_1 r_2 e^{i(\theta_1+\theta_2)}. \end{aligned}$$

$$\frac{z_1}{z_2} = \frac{r_1 e^{i\theta_1}}{r_2 e^{i\theta_2}} = \frac{r_1}{r_2}\,\frac{e^{i\theta_1}}{e^{i\theta_2}} = \frac{r_1}{r_2}\,\frac{e^{i\theta_1}\ e^{i\theta 2}}{e^{i\theta_2}\ e^{i\theta_2}} = \frac{r_1}{r_2}\,\frac{e^{i(\theta_1-\theta_2)}}{e^{i0}} = \frac{r_1}{r_2}e^{i(\theta_1-\theta_2)}.$$

Example 10.3.1.

Let $z_1 = 4\left(\cos\frac{\pi}{4} + i\sin\frac{\pi}{4}\right)$ and $z_2 = 2\left(\cos\frac{\pi}{6} + i\sin\frac{\pi}{6}\right)$. Express $z_1 z_2$ and $\frac{z_1}{z_2}$ in polar, exponential, and standard forms.

Solution

$$
\begin{aligned}
z_1 z_2 &= (4)(2)\left[\cos\left(\tfrac{\pi}{4} + \tfrac{\pi}{6}\right) + i\ \sin\left(\tfrac{\pi}{4} + \tfrac{\pi}{6}\right)\right] = 8\left(\cos\tfrac{5\pi}{12} + i\ \sin\tfrac{5\pi}{12}\right).\\
z_1 z_2 &= 8e^{i\frac{5\pi}{12}}.\\
z_1 z_2 &= 8\left[\left(\cos\tfrac{\pi}{4}\cos\tfrac{\pi}{6} - \sin\tfrac{\pi}{4}\sin\tfrac{\pi}{6}\right) + i\left(\sin\tfrac{\pi}{4}\cos\tfrac{\pi}{6} + \cos\tfrac{\pi}{4}\sin\tfrac{\pi}{6}\right)\right]\\
&= 8\left[\left(\frac{\sqrt{2}}{2}\frac{\sqrt{3}}{2} - \frac{\sqrt{2}}{2}\frac{1}{2}\right) + i\left(\frac{\sqrt{2}}{2}\frac{\sqrt{3}}{2} + \frac{\sqrt{2}}{2}\frac{1}{2}\right)\right]\\
&= 2\sqrt{2}[(\sqrt{3}-1) + i(\sqrt{3}+1)].\\
\frac{z_1}{z_2} &= 2\left[\cos\left(\tfrac{\pi}{4} - \tfrac{\pi}{6}\right) + i\ \sin\left(\tfrac{\pi}{4} - \tfrac{\pi}{6}\right)\right] = 2\left[\cos\tfrac{\pi}{12} + i\ \sin\tfrac{\pi}{12}\right].\\
\frac{z_1}{z_2} &= 2e^{i\frac{\pi}{12}}.\\
\frac{z_1}{z_2} &= 2\left[\cos\left(\tfrac{\pi}{4} - \tfrac{\pi}{6}\right) + i\ \sin\left(\tfrac{\pi}{4} - \tfrac{\pi}{6}\right)\right]\\
&= 2\left[\left(\cos\tfrac{\pi}{4}\cos\tfrac{\pi}{6} + \sin\tfrac{\pi}{4}\sin\tfrac{\pi}{6}\right) + i\left(\sin\tfrac{\pi}{4}\cos\tfrac{\pi}{6} - \cos\tfrac{\pi}{4}\sin\tfrac{\pi}{6}\right)\right]\\
&= 2\left[\left(\frac{\sqrt{2}}{2}\frac{\sqrt{3}}{2} + \frac{\sqrt{2}}{2}\frac{1}{2}\right) + i\left(\frac{\sqrt{2}}{2}\frac{\sqrt{3}}{2} - \frac{\sqrt{2}}{2}\frac{1}{2}\right)\right]\\
&= \frac{\sqrt{2}}{2}[(\sqrt{3}+1) + i(\sqrt{3}-1)].
\end{aligned}
$$

By **Theorem 10.3.1** , we obtain the following result on the arguments of products and quotients.

Theorem 10.3.2.

Let $z_1,\ z_2$ be two nonzero complex numbers. Then the following assertions hold.

1. $\arg(z_1 z_2) = \arg\ z_1 + \arg\ z_2 = \operatorname{Arg}\ z_1 + \operatorname{Arg}\ z_2 + 2k\pi$ *for* $\in \mathbb{Z}$.
2. $\arg\left(\frac{z_1}{z_2}\right) = \arg\ z_1 - \arg\ z_2 = \operatorname{Arg}\ z_1 - \operatorname{Arg}\ z_2 + 2k\pi$ for $k \in \mathbb{Z}$.

Proof

We only prove (1). Let $\theta_1 = \arg\ z_1 = \text{Arg}\ z_1 + 2m\pi$ for $m \in \mathbb{Z}$ and $\theta_2 = \arg\ z_2 = \text{Arg}\ z_2 + 2n\pi$ for $n \in \mathbb{Z}$. By **Theorem 10.3.1** ,

$$\begin{aligned} \arg(z_1 z_2) &= \theta_1 + \theta_2 = \arg\ z_1 + \arg\ z_2 = (\text{Arg}\ z_1 + 2m\pi) + (\text{Arg}\ z_2 + 2m\pi) \\ &= (\text{Arg}\ z_1 + \text{Arg}\ z_2) + 2(m+n)\pi = (\text{Arg}\ z_1 + \text{Arg}\ z_2) + 2k\pi, \end{aligned}$$

where $k = m + n$. Because *m, n* can be any values in $\mathbb{Z}$, k can be any integer in $\mathbb{Z}$. Hence,

$$\arg(z_1 z_2) = (\text{Arg}\ z_1 + \text{Arg}\ z_2) + 2k\pi \quad \text{for } k \in \mathbb{Z}.$$

Example 10.3.2.

Let $z_1 = 1 + i$ and $z_2 = -1 - i$. Find $\arg(z_1 z_2)$, $\arg\left(\frac{z_1}{z_2}\right)$, $\arg\left(\frac{z_2}{z_1}\right)$, and $\arg\frac{1}{z_1}$.

Solution

Because

$$\text{Arg}\ z_1 = \text{Arg}(1+i) = \frac{\pi}{4} \text{ and } \text{Arg}\ z_2 = \text{Arg}(-1-i) = -\pi + \frac{\pi}{4} = -\frac{3\pi}{4},$$

it follows from **Theorem 10.3.2** that

$$\begin{aligned} \arg(z_1 z_2) &= \left(\tfrac{\pi}{4} - \tfrac{3\pi}{4}\right) + 2k\pi = -\tfrac{\pi}{2} + 2k\pi \quad \text{for } k \in \mathbb{Z}, \\ \arg\left(\frac{z_1}{z_2}\right) &= \tfrac{\pi}{4} - \left(-\tfrac{3\pi}{4}\right) + 2k\pi = \pi + 2k\pi \quad \text{for } k \in \mathbb{Z}, \\ \arg\left(\frac{z_2}{z_1}\right) &= -\tfrac{3\pi}{4} - \tfrac{\pi}{4} + 2k\pi = -\pi + 2k\pi \quad \text{for } k \in \mathbb{Z}, \end{aligned}$$

and

$$\arg \frac{1}{z_1} = -\frac{\pi}{4} + 2k\pi \quad \text{for } k \in \mathbb{Z}.$$

In **Theorem 10.3.2**, the arguments corresponding to $k = 0$ may not be the principal arguments if they are not in the interval $(-\pi, \pi]$. Therefore, to find $\operatorname{Arg}(z_1 z_2)$ or $\operatorname{Arg}\left(\frac{z_1}{z_2}\right)$, we need to find a suitable integer $k_0 \in \mathbb{Z}$ such that the corresponding argument belongs to $(-\pi, \pi]$, which is the principal argument.

Corollary 10.3.1.

Let z_1, z_2 be two nonzero complex numbers. Then the following assertions hold.

1. $\operatorname{Arg}(z_1 z_2) = \operatorname{Arg}\ z_1 + \operatorname{Arg}\ z_2 + 2k_0\pi,$ where $k_0 \in \mathbb{Z}$ *such that*

$$-\pi < \operatorname{Arg}\ z_1 + \operatorname{Arg}\ z_2 + 2k_0\pi \leq \pi.$$

2. $\operatorname{Arg}\left(\frac{z_1}{z_2}\right) = \operatorname{Arg}\ z_1 - \operatorname{Arg}\ z_2 + 2k_0\pi,$ *where $k_0 \in \mathbb{Z}$ such that*

$$-\pi < \operatorname{Arg}\ z_1 - \operatorname{Arg}\ z_2 + 2k_0\pi \leq \pi.$$

Example 10.3.3.

Let z_1, z_2 be the same as in **Example 10.3.2**. Find $\operatorname{Arg}\left(\frac{z_1}{z_2}\right), \ \operatorname{Arg}\left(\frac{z_2}{z_1}\right),$

Solution

By **Example 10.3.2**, we obtain

$$\arg\left(\frac{z_1}{z_2}\right) = \pi + 2k\pi \ \text{ for } k \in \mathbb{Z}.$$

Only when $k = 0$, the above angle belongs to $(-\pi,\ \pi]$. Hence,

$$\operatorname{Arg}\left(\frac{z_1}{z_2}\right) = \pi.$$

Similarly, we have

$$\arg\left(\frac{z_2}{z_1}\right) = -\pi + 2k\pi \quad \text{for } k \in \mathbb{Z},$$

When $k = 1$, the above angle is in $(-\pi,\ \pi]$. Hence, $\operatorname{Arg}\left(\frac{z_2}{z_1}\right) = \pi$.

Example 10.3.4.

Let $z_1 = -1$ and $z_2 = i$. Use **Corollary 10.3.1** to find $\operatorname{Arg}(z_1 z_2)$

Solution

By **Theorem 10.3.2**, we have

$$\arg(z_1 z_2) = \pi + \tfrac{\pi}{2} + 2k\pi = \tfrac{3\pi}{2} + 2k\pi \quad \text{for } k \in \mathbb{Z}.$$

Because $\frac{3\pi}{2} \notin (-\pi,\ \pi]$, so $\frac{3\pi}{2}$ is not the principal argument of $z_1 z_2$.. When $k = -1$, we have

$$\frac{3\pi}{2} + 2k\pi = \frac{3\pi}{2} - 2\pi = -\frac{\pi}{2} \in (-\pi,\ \pi].$$

Hence, $\operatorname{Arg}(z_1 z_2) = -\frac{\pi}{2}$.

Theorem 10.3.3 (Demoivre's formulas)

1. *Let* $z = r(\cos\theta + i\ \sin\theta)$ *and* $n \in \mathbb{N}$. *Then*

$$z^n = r^n(\cos\ n\theta + i\ \sin\ n\theta).$$

2. *Let* $z = re^{i\theta}$ *and* $n \in \mathbb{N}$. *Then*

$$z^n = r^n e^{i(n\theta)}$$

and if $z \neq 0$, *then*

$$z^{-n} = r^{-n} e^{i(-n\theta)}.$$

Proof

We only prove (*i*) because (*ii*) follows from (*i*). By **Theorem 10.3.1**, we have

$$z^2 = z \cdot z = (r \cdot r)[\cos(2\theta + \theta) + i \sin(2\theta + \theta)] = r^3(\cos(3\theta) + i \sin(3\theta)).$$

and

$$z^3 = z^2 = z \cdot z = (r^2 \cdot r)[\cos(2\theta + \theta) + i \sin(2\theta + \theta)] = r^3(\cos(3\theta) + i \sin(3\theta)).$$

Repeating the process implies that

$$z^n = [r(\cos\theta + i \sin\theta)]^n = r^n(\cos\ n\theta + i \sin\ n\theta).$$

Because $z^{-1} = \frac{1}{z}$,

$$z^{-1} = \frac{1}{z^n} = \frac{1}{r^n e^{i(n\theta)}} = r^{-n} e^{i(-n\theta)}.$$

In **Theorem 10.3.3**, if $r = 1$, then we obtain the following formulas.

Corollary 10.3.2.

1. $(\cos\theta + i\sin\theta)^n = \cos\ n\theta + i\sin\ n\theta$ for $n \in \mathbb{N}$.
2. $\left(e^{i\theta}\right)^n = e^{i(n\theta)}$ *for* $n \in \mathbb{N}$.

Example 10.3.5.

Express $(-2-2i)^4$ in $x+yi$ form.

Solution

1. Step 1. Find the polar form of $z = -2-2i$.

$$r = \left|z\right| = \sqrt{(-2)^2 + (-2)^2} = \sqrt{8} = 2\sqrt{2}.$$

Let $\psi \in \left(0,\ \frac{\pi}{2}\right)$ be such that

$$\tan\psi = \left|\frac{\sqrt{-2}}{-2}\right| = 1.$$

Then $\psi = \frac{\pi}{4}$. Because $x = -2 < 0$ and $y = -2 < 0,\ z$ is in the third quadrant and by **Theorem 10.2.2** (3),

$$\operatorname{Arg} z = -\pi + \psi = -\pi + \frac{\pi}{4} = -\frac{3\pi}{4}.$$

Hence,

$$-2-2i = 2\sqrt{2}\left[\left(\cos\left(-\frac{3\pi}{4}\right) + i\sin\left(-\frac{3\pi}{4}\right)\right]\right..$$

2. Step 2. Use **Theorem 10.3.3** to find the polar form. By **Theorem 10.3.3** , we have

$$\begin{aligned}(-2-2i)^4 &= \left\{2\sqrt{2}\left[\cos\left(-\tfrac{3\pi}{4}\right)+i\sin\left(-\tfrac{3\pi}{4}\right)\right]\right\}^4 \\ &= (2\sqrt{2})^4\left[\cos 4\left(-\tfrac{3\pi}{4}\right)+i\sin 4\left(-\tfrac{3\pi}{4}\right)\right] \\ &= 64[\cos(-3\pi)+i\sin(-3\pi)] = 64[\cos(3\pi)+i0] \\ &= 64\cos\pi = -64.\end{aligned}$$

Exercises

1. Express the following numbers in $x+yi$ form.
    1. $(1+\sqrt{3}i)^3$
    2. $(-1-i)^4$
    3. $(1+i)^{-8}$.
2. Let $z_1 = \frac{1+i}{\sqrt{2}}$ and $z_2 = \sqrt{3}-i$. Find the exponential form of $z_1 z_2$.
3. Let $z = \frac{\sqrt{2}}{2}(1-i)$. Find $z^{100}+z^{50}+1$.
4. Compute the following value

$$z = \left(-\frac{1}{2}+\frac{\sqrt{3}}{2}i\right)^{600} + \left(-\frac{1}{2}-\frac{\sqrt{3}}{2}i\right)^{60}.$$

10.4 Roots of complex numbers

Given a complex number z, and an integer $n \geq 2,$ we want to find a complex number $w \in \mathbb{C}$ such that

(10.4.1)

$$w^n = z.$$

If $w \in \mathbb{C}$ satisfies **(10.4.1)**, we write

(10.4.2)

$$w = z^{\frac{1}{n}} = \sqrt[n]{z}.$$

$z^{\frac{1}{n}}$ is called the n^{th} root of z. It is obvious that the n^{th} root of z is a root of equation **(10.4.1)**.

Theorem 10.4.1.

Let $z = r(\cos\theta + i\sin\theta)$ *Then z has n* n^{th} *roots in exponential form:*

(10.4.3)

$$\sqrt[n]{z} = \sqrt[n]{r}e^{\frac{(\theta+2k)\pi}{n}} \quad for\ k = 0,\ 1,\ 2,\ \cdots,\ n-1,$$

or in polar form

(10.4.4)

$$\sqrt[n]{z} = \sqrt[n]{r}\left(\cos\frac{\theta+2k\pi}{n} + i\sin\frac{\theta+2k\pi}{n}\right) \quad for\ k = 0,\ 1,\ 2,\ \cdots,\ n-1,$$

Proof

Let $w = \rho e^{i\psi}$. Because $z = r(\cos\theta + i\sin\theta) = re^{i\theta}$ and $w^n = z$, it follows from **Theorem 10.3.3** (2) that

$$\rho^n e^{i(n\psi)} = \left(\rho e^{i\psi}\right)^n = re^{i\theta}.$$

By **Proposition 10.2.1** (2), we obtain

$$\rho^n = r \quad \text{and } n\psi = \theta + 2k\pi \text{ for } k \in \mathbb{Z}.$$

This implies that $\rho = \sqrt[n]{r}$ and

$$\psi = \frac{\theta+2k\pi}{n} \quad \text{for } k \in \mathbb{Z}.$$

Hence,

(10.4.5)

$$w = \sqrt[n]{r}e^{\frac{\theta+2k\pi}{n}} \quad \text{for } k \in \mathbb{Z}.$$

Note that for each $k \in \mathbb{Z}$, there exist $j \in \mathbb{Z}$ and $l \in \{0,\ 1,\ 2,\ \cdots,\ n-1\}$ such that

$$k = jn + l.$$

This, together with **(10.4.5)** and **Proposition 10.2.1** (1), implies that

$$w = \sqrt[n]{r}e^{\frac{\theta+2k\pi}{n}} = \sqrt[n]{r}e^{\frac{\theta+2(jn+l)\pi}{n}} = \sqrt[n]{r}e^{\frac{\theta+2l+2jn)\pi}{n}} = \sqrt[n]{r}e^{\frac{(\theta+2l)\pi}{n}}e^{i(2j)}$$
$$= \sqrt[n]{r}e^{\frac{(\theta+2l)\pi}{n}}.$$

Replacing *l* by *k*, we obtain for $k \in \{0,\ 1,\ 2,\ \cdots,\ n-1\}$,

$$w = \sqrt[n]{r}e^{\frac{(\theta+2k)\pi}{n}} = \sqrt[n]{r}\left(\cos\frac{\theta+2k\pi}{n} + i\sin\frac{\theta+2k\pi}{n}\right).$$

Because the right side of **(10.4.4)** depends on *k*, we can write

$$(\sqrt[n]{z})k = \sqrt[n]{r}\left(\cos\frac{\theta+2k\pi}{n} + i\sin\frac{\theta+2k\pi}{n}\right) \quad \text{for } k = 0,\ 1,\ 2,\ \cdots,\ n-1.$$

By **Theorem 10.4.1** , we see that equation **(10.4.1)** has *n* distinct roots given by **(10.4.4)** . We write

(10.4.6)

$$\sqrt[n]{z} = \left\{(\sqrt[n]{z})_0,\ (\sqrt[n]{z})_1,\ \cdots,\ (\sqrt[n]{z})_{n-1}\right\}.$$

Corollary 10.4.1.

Let $z = r(\cos\theta + i\sin\theta)$. *Then*

$$(\sqrt{z})_0 = \sqrt{r}\left(\cos\tfrac{\theta}{2} + i\sin\tfrac{\theta}{2}\right) \quad and\ (\sqrt{z})_1 = -(\sqrt{z})_0.$$

If *z* is a real number, then by **Corollary 10.4.1** ,

$$\sqrt{z} = \sqrt{r}\left(\cos\tfrac{\theta}{2} + i\sin\tfrac{\theta}{2}\right) \quad \text{or } \sqrt{z} = -\sqrt{r}\left(\cos\tfrac{\theta}{2} + i\sin\tfrac{\theta}{2}\right).$$

It means that the square root of a complex number has a positive root $(\sqrt{z})_0$ and a negative root $(\sqrt{z})_1$. Hence, when we consider complex numbers, the square root $\sqrt{z}$ represents two roots, which are

(10.4.7)

$$\sqrt{z} = (\sqrt{z})_0 \quad \text{and } \sqrt{z} = -(\sqrt{z})_0.$$

Example 10.4.1.

Find $\sqrt{-1}$.

Solution

Because $-1 = \cos\pi + i\sin\pi$. It follows from **Corollary 10.4.1** that

$$(\sqrt{-1})_0 = \cos\tfrac{\pi}{2} + i\sin\tfrac{\pi}{2} = i \quad \text{and } (\sqrt{-1})_1 = -i.$$

Therefore, $\sqrt{-1} = \{i, \ -i\}$.

From **Example 10.4.1** , we see that in **(10.1.4)** , $i = \sqrt{-1}$ can be interpreted as the positive root of $\sqrt{-1}$, that is, *i* is the positive root of the following equation

$$w^2 = -1.$$

Example 10.4.2.

Solve $z^2 - 1 - \sqrt{3}i = 0$ for z.

Solution

$z^2 - 1 - \sqrt{3}i = 0$ if and only if $z = \sqrt{1 + \sqrt{3}i}$. Because

$$1+\sqrt{3}i = 2\left(\cos\frac{\pi}{3} + i\sin\frac{\pi}{3}\right),$$

it follows from **Theorem 10.4.1** that

$$z_0 := \left(\sqrt{1+\sqrt{3}i}\right)_0 = \sqrt{2}\left(\cos\frac{\pi}{6} + i\sin\frac{\pi}{6}\right) = \sqrt{2}\left(\frac{\sqrt{3}}{2} + \frac{1}{2}i\right) = \frac{\sqrt{6}}{2} + \frac{\sqrt{2}}{2}i$$

and

$$\begin{aligned} z_1 :&= \left(\sqrt{1+\sqrt{3}i}\right)_1 = \sqrt{2}\left[\cos\left(\tfrac{\pi}{6}+\pi\right) + i\sin\left(\tfrac{\pi}{6}+\pi\right)\right] = \sqrt{2}\left(-\tfrac{\sqrt{3}}{2} - \tfrac{1}{2}i\right) \\ &= -\tfrac{\sqrt{6}}{2} - \tfrac{\sqrt{2}}{2}i. \end{aligned}$$

Hence, z_0 and z_1 are the solutions of $z^2 - 1 - \sqrt{3}i = 0$.

Example 10.4.3.

Find all $\sqrt[3]{-1}$.

Solution

Because $-1 = \cos\pi + i\sin\pi$. It follows from **Theorem 10.4.1** that

$$\begin{aligned} \left(\sqrt[3]{-1}\right)_0 &= \cos\tfrac{\pi}{3} + i\sin\tfrac{\pi}{3} = \tfrac{1}{2} + i\tfrac{\sqrt{3}}{2}. \\ \left(\sqrt[3]{-1}\right)_1 &= \cos\tfrac{\pi+2\pi}{3} + i\sin\tfrac{\pi+2\pi}{3} = \cos\pi + i\sin\pi = -1. \\ \left(\sqrt[3]{-1}\right)_2 &= \cos\tfrac{\pi+4\pi}{3} + i\sin\tfrac{\pi+4\pi}{3} = \cos\left(2\pi - \tfrac{\pi}{3}\right) + i\sin\left(2\pi - \tfrac{\pi}{3}\right) \\ &= \cos\tfrac{\pi}{3} - i\sin\tfrac{\pi}{3} = \tfrac{1}{2} - i\tfrac{\sqrt{3}}{2}. \end{aligned}$$

Therefore, $\sqrt[3]{-1} = \left\{\tfrac{1}{2} + i\tfrac{\sqrt{3}}{2},\ -1,\ \tfrac{1}{2},\ -i\tfrac{\sqrt{3}}{2}\right\}$.

Theorem 10.4.2.

Let $a,\ b,\ c \in \mathbb{C}$ be complex numbers and $a \neq 0.$ Then the solution of the following quadratic equation

(10.4.8)

$$az^2 + bz + c = 0$$

is

(10.4.9)

$$z = \frac{-b + \sqrt{b^2 - 4ac}}{2a}.$$

Proof

The proof is the same as **(10.1.2)** and is omitted.

Note that if $b^2 - 4ac \neq 0,$ then by **(10.4.7)**, we see that $\sqrt{b^2 - 4ac}$ contains two values. Hence, **(10.4.9)** is consistent with **(10.1.3)** or **(10.1.5)**.

Example 10.4.4.

Solve the following equation

$$(1-i)z^2 - 2z + (1+i) = 0.$$

Solution

By **(10.4.9)**, the solution of the equation is

(10.4.10)

$$z=\frac{2+\sqrt{4-4(1-i)(1+i)}}{2(1-i)}=\frac{2+\sqrt{-4}}{2(1-i)}=\frac{1+\sqrt{-1}}{(1-i)}.$$

By **Example 10.4.1** , $\sqrt{-1}=i$ or $\sqrt{-1}=-i$. If $\sqrt{-1}=i$, by **(10.4.10)** , we have

$$z_0=\frac{1+i}{1-i}=\frac{(1+i)^2}{(1-i)(1+i)}=\frac{2i}{2}=i.$$

If $\sqrt{-1}=-i$. by **(10.4.10)** , we have

$$z_1=\frac{1-i}{1-i}=1.$$

Hence, $z_0=i$ and $z_1=1$ are the solutions of $(1-i)z^2-2z+(1+i)=0$.

Exercises

1. Find all roots for each of the following complex numbers.
   a. $\sqrt[4]{1}$
   b. $\sqrt[6]{-1}$
   c. $\sqrt[3]{1-i}$
   d. $\sqrt[3]{-i}$
   e. $(-1+\sqrt{3}i)^{\frac{1}{4}}$

2. Find all solutions of $z^3=-8$.
3. Solve the following equations.
   a. $(1-i)z^2+2z+(1+i)=0$
   b. $z^2+2z+(1-i)=0$

Appendix A Solutions to the Problems

A.1 Euclidean spaces

Section 1.1

1. Write a three-column zero vector, and a three-column nonzero vector, and a five-column zero vector and a five-column nonzero vector.
2. Suppose that the buyer for a manufacturing plant must order different quantities of oil, paper, steel, and plastics. He will order 40 units of oil, 50 units of paper, 80 units of steel, and 20 units of plastics. Write the quantities in a single vector.
3. Suppose that a student's course marks for quiz 1, quiz 2, test 1, test 2, and the final exam are 70, 85, 80, 75, and 90, respectively. Write his marks as a column vector.
4. Let $\vec{a} = \begin{pmatrix} x-2y \\ 2x-y \\ 2z \end{pmatrix}$ and $\vec{b} = \begin{pmatrix} 2 \\ -2 \\ 1 \end{pmatrix}$. Find all $x,\ y,\ z \in \mathbb{R}$ that $\vec{a} = \vec{b}$.
5. $\vec{a} = \begin{pmatrix} |x| \\ y^2 \end{pmatrix}$ and $\vec{b} = \begin{pmatrix} 1 \\ 4 \end{pmatrix}$. Find all $x,\ y \in \mathbb{R}$ such that $\vec{a} = \vec{b}$.
6. $\vec{a} = \begin{pmatrix} x-y \\ 4 \end{pmatrix}$ and $\vec{b} = \begin{pmatrix} 2 \\ x+y \end{pmatrix}$. Find all $x,\ y \in \mathbb{R}$ such that $\vec{a} - \vec{b}$ is a nonzero vector.
7. Let $P(1,\ x^2),\ Q(3,\ 4),\ P_1(4,\ 5),$ and $Q_1(6,\ 1)$. Find all possible values of $x \in \mathbb{R}$ such that $\overrightarrow{PQ} = \overrightarrow{P_1Q_1}$.
8. A company with 553 employees lists each employee's salary as a component of a vector $\vec{a}$ in $\mathbb{R}^{553}$. If a 6% salary increase has been approved, find the vector involving $\vec{a}$ that gives all the new salaries.

9. Let $\vec{a} = \begin{pmatrix} 110 \\ 88 \\ 40 \end{pmatrix}$ denote the current prices of three items at a store. Suppose that the store announces a sale so that the price of each item is reduced by 20%.
   a. Find a three-vector that gives the price changes for the three items.
   b. Find a three-vector that gives the new prices of the three items.

10. Let $\vec{a} = \begin{pmatrix} -1 \\ 2 \\ 3 \end{pmatrix}$, $\vec{b} = \begin{pmatrix} 2 \\ -2 \\ 0 \end{pmatrix}$, $\vec{c} = \begin{pmatrix} 3 \\ 0 \\ -1 \end{pmatrix}$, $\alpha \in \mathbb{R}$. Compute
   i. $2\vec{a} - \vec{b} + 5\vec{c}$;
   ii. $4\vec{a} + \alpha \overrightarrow{b} - 2\vec{c}$

11. Find *x*, *y*, and *z* such that $\begin{pmatrix} 9 \\ 4y \\ 2z \end{pmatrix} + \begin{pmatrix} 3x \\ 8 \\ -6 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$.

12. Let $\vec{u} = (1,\ -1,\ 0,\ 2)$, $\vec{v} = (-2,\ -1,\ 1,\ -4)$, and $\vec{w} = (3,\ 2,\ -1,\ 0)$.
   a. Find a vector $\vec{a} \in \mathbb{R}^4$ such that $2\vec{u} - 3\vec{v} - \vec{a} = \vec{w}$.
   b. Find a vector $\vec{a}$

$$\frac{1}{2}[2\vec{u} - 3\vec{v} + \vec{a}] = 2\vec{w} + \vec{u} - 2\vec{a}.$$

13. Determine whether $\vec{a} \parallel \vec{b}$,
   1. $\vec{a} = (1,\ 2,\ 3)$ and $\vec{b} = (-2,\ -4,\ -6)$.

2. $\vec{a} = (-1,\ 0,\ 1,\ 2)$ and $\vec{b} = (-3,\ 0,\ 3,\ 6)$.
3. $\vec{a} = (1,\ -1,\ 1,\ 2)$ and $\vec{b} = (-2,\ 2,\ 2,\ 3)$.

14. Let $x,\ y,\ a,\ b \in \mathbb{R}$ with $x \neq y$. Let $P(x,\ 2x),\ Q(y,\ 2y),\ P_1(a,\ 2a),$ and $Q_1(b,\ 2b)$ be points in $\mathbb{R}^2$. Show that $\overrightarrow{PQ} \parallel \overrightarrow{P_1Q_1}$.

15. Let $P(x,\ 0),\ Q(3,\ x^2),\ P_1(2x,\ 1),$ and $Q_1(6,\ x^2)$. Find all possible values of $x \in \mathbb{R}$ such that $\overrightarrow{PQ} \parallel \overrightarrow{P_1Q_1}$.

16. Let $\vec{a} = \begin{pmatrix} -1 \\ 2 \\ 3 \end{pmatrix},\ \vec{b} = \begin{pmatrix} 2 \\ -2 \\ 0 \end{pmatrix},\ \vec{c} = \begin{pmatrix} 3 \\ 0 \\ -1 \end{pmatrix},$ and $\alpha \in \mathbb{R}$.

 Compute
 a. $\vec{a} \cdot \vec{b}$;
 b. $\vec{a} \cdot \vec{c}$;
 c. $\vec{b} \cdot \vec{c}$;
 d. $\vec{a} \cdot (\vec{b} + \vec{c})$.

17. Let $\vec{u} = (1,\ -1,\ 0,\ 2),\ \vec{v} = (-2,\ -1,\ 1,\ -4),$ and $\vec{w} = (3,\ 2,\ -1,\ 0)$. Compute $\vec{u} \cdot \vec{v},\ \ \vec{u} \cdot \vec{w},$ and $\vec{w} \cdot \vec{v}$.
18. Assume that the percentages for homework, test 1, test 2, and the final exam for a course are 10%, 25%, 25%, and 40%, respectively. The total marks for homework, test 1, test 2, and the final exam are 10, 50, 50, and 90, respectively. A student's corresponding marks are 8, 46, 48, and 81, respectively. What is the student's final mark out of 100?
19. Assume that the percentages for homework, test 1, test 2, and the final exam for a course are 14%, 20%, 20%, and 46%, respectively. The total marks for homework, test 1, test 2, and the final exam are 14, 60, 60, and 80, respectively. A student's corresponding marks are 12, 45, 51, and 60, respectively. What is the student's final mark out of 100?

20. A manufacturer produces three different types of products. The demand for the products is denoted by the vector $\vec{a} = (10,\ 20,\ 30)$. The price per unit for the products is given by the vector $\vec{b} = (\$200,\ \$150,\ \$100)$. If the demand is met, how much money will the manufacturer receive?
21. A company pays four groups of employees a salary. The numbers of the employees for the four groups are expressed by a vector $\vec{a} = (5,\ 20,\ 40)$. The payments for the groups are expressed by a vector $\vec{b} = (\$100,\ 000,\ \$80,\ 000,\ \$60,\ 000)$. Use the dot product to calculate the total amount of money the company paid its employees.
22. There are three students who may buy Calculus or algebra books. Use the dot product to find the total number of students who buy both calculus and algebra books.
23. There are *n* students who may buy a calculus or algebra books $(n \geq 2)$. Use the dot product to find the total number of students who buy both calculus and algebra books.
24. Assume that a person *A* has contracted a contagious disease and has direct contacts with four people: *P*1, *P*2, *P*3, and *P*4. We denote the contacts by a vector $\vec{a} := (a_{11},\ a_{12},\ a_{13},\ a_{14})$, where if the person *A* has made contact with the person *Pj*, then $a_{1j} = 1$, and if the person *A* has made no contact with the person *Pj*, then $a_{1j} = 0$. Now we suppose that the four people then have had a variety of direct contacts with another individual *B*, which we denote by a vector $\vec{b} := (b_{11},\ b_{21},\ b_{31},\ b_{41})$, where if the person *Pj* has made contact with the person *B*, then $b_{j1} = 1$, and if the person *Pj* has made no contact with the person *B*, then $b_{j1} = 0$.
    i. Find the total number of indirect contacts between *A* and *B*.
    ii. If $\vec{a} = (1,\ 0,\ 1,\ 1)$ and $\vec{a} = (1,\ 1,\ 1,\ 1)$, find the total number of indirect contacts between *A* and *B*.

A.2 Matrices

Section 2.1

1. Find the sizes of the following matrices:

$$A = (2)\ B = \begin{pmatrix} 5 & 8 \\ 1 & 4 \end{pmatrix}\quad C = \begin{pmatrix} 10 & 3 & 8 \\ 10 & 3 & 4 \end{pmatrix}\quad D = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 10 & 2 \\ 0 & 0 & 1 \\ 6 & 9 & 10 \end{pmatrix}$$

2. Use column vectors and row vectors to rewrite each of the following matrices.

$$A = \begin{pmatrix} 3 & 3 & 2 \\ 1 & 8 & 1 \\ 5 & 4 & 10 \end{pmatrix}\quad B = \begin{pmatrix} 1 & 9 & 7 & 2 \\ 3 & 10 & 8 & 7 \\ 4 & 3 & 10 & 4 \end{pmatrix}\quad C = \begin{pmatrix} 9 & 8 & 7 \\ 6 & 5 & 4 \\ 3 & 2 & 1 \\ 0 & -1 & -2 \end{pmatrix}$$

3. Find the transposes of the following matrices:

$$C = (4)\ B = \begin{pmatrix} 3 & 11 & 2 \end{pmatrix}\quad C = \begin{pmatrix} 33 \\ 8 \\ 12 \end{pmatrix}\quad D = \begin{pmatrix} 3 & 9 & -19 \\ -2 & 8 & -7 \\ 5 & 3 & -9 \end{pmatrix}$$

4. Let $A = \begin{pmatrix} 9 & 3 \\ 6 & x^2 \end{pmatrix}$ and $B = \begin{pmatrix} 9 & 3 \\ 6 & 4 \end{pmatrix}$. Find all $x \in \mathbb{R}$ such that $A = B$.

5. Let $C = \begin{pmatrix} 12 & 5 & 25 \\ 19 & 4 & 6 \end{pmatrix}$ and $D = \begin{pmatrix} 12 & 5 & x^2 \\ 19 & 4 & 6 \end{pmatrix}$. Find all $x \in \mathbb{R}$ such that $C = D$.

6. Let $E = \begin{pmatrix} 120 & 25 & 122 \\ 123 & 124 & 125 \\ 126 & 127 & 128 \end{pmatrix}$ and $F = \begin{pmatrix} 120 & x^2 & 122 \\ 123 & 124 & x^3 \\ 26x-4 & 127 & 128 \end{pmatrix}$. Find all $x \in \mathbb{R}$ such that $E = F$.

7. Let $A = \begin{pmatrix} a & b \\ 3x+2y & -x+y \end{pmatrix}$ and $B = \begin{pmatrix} 1 & 1 \\ 3 & 6 \end{pmatrix}$. Find all $a,\ b,\ x,\ y \in \mathbb{R}$ such that $A = B$.

8. Let

$$C = \begin{pmatrix} a+b & 2b-a \\ x-2y & 5x+3y \end{pmatrix} \text{ and } D = \begin{pmatrix} 6 & 0 \\ 8 & 14 \end{pmatrix}.$$

Find all $a,\ b,\ x,\ y \in \mathbb{R}$ such that $C = D$.

9. Let $A = \begin{pmatrix} -2 & 3 & 4 \\ 6 & -1 & -8 \end{pmatrix}$ and $B = \begin{pmatrix} 7 & -8 & 9 \\ 0 & -1 & 0 \end{pmatrix}$. Compute

   i. $A + B$;
   ii. $-A$;
   iii. $4A - 2B$;
   iv. $100A + B$.

10. Let $A = \begin{pmatrix} 9 & 5 & 1 \\ 8 & 0 & 0 \\ 0 & 3 & 2 \end{pmatrix}$, $B = \begin{pmatrix} 2 & 2 & 2 \\ 0 & 0 & 0 \\ 4 & 6 & 8 \end{pmatrix}$, and $C = \begin{pmatrix} 4 & 7 & 0 \\ 0 & 3 & 1 \\ 0 & 0 & 2 \end{pmatrix}$.

Compute

   1. $3A - 2B + C$,
   2. $[3(A+B)]^T$,
   3. $\left(4A + \frac{1}{2}B - C\right)^T$.

11. Find the matrix *A* if $\left[\left(3A^T\right) - \begin{pmatrix} -7 & -2 \\ -6 & 9 \end{pmatrix}^T\right]^T = \begin{pmatrix} -5 & -10 \\ 33 & 12 \end{pmatrix}$.

12. Find the matrix *B* if

$$\left[\frac{1}{2}B+\begin{pmatrix}6 & 3\\ 8 & 3\\ 1 & 4\end{pmatrix}\right]^T-3\begin{pmatrix}-5 & 6\\ 8 & -9\\ -4 & 2\end{pmatrix}^T=\begin{pmatrix}23 & -16 & 17\\ -16 & 26.5 & 2\end{pmatrix}.$$

A.3 Determinants

Section 3.1

1. Let $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, where $a,\ b,\ c,\ d \in \mathbb{R}$. Show that

$$|\lambda I - A| = \lambda^2 - \operatorname{tr}(A)\lambda + |A|$$

   for each $\lambda \in \mathbb{R}$.

2. Let $A_1 = \begin{pmatrix} 2 & 4 \\ 3 & -2 \end{pmatrix}$, $A_2 = \begin{pmatrix} 1 & 2 \\ -1 & 4 \end{pmatrix}$, $A_3 = \begin{pmatrix} -1 & 1 \\ -1 & -3 \end{pmatrix}$. For each $i = 1,\ 2,\ 3$, find all $\lambda \in \mathbb{R}$ such that $|\lambda I - A_i| = 0$.
3. Use **(3.1.3)** and **(3.1.4)** to calculate each of the following determinants.

$$|A| = \begin{vmatrix} 1 & 2 & 3 \\ -2 & 1 & 2 \\ 3 & -1 & 4 \end{vmatrix} \quad |B| = \begin{vmatrix} 2 & -2 & 4 \\ 4 & 3 & 1 \\ 0 & 1 & 2 \end{vmatrix} \quad |C| = \begin{vmatrix} 0 & -1 & 3 \\ 4 & 1 & 2 \\ 0 & 0 & 1 \end{vmatrix}$$

4. Compute the determinant $|A|$ by using its expansion of cofactors corresponding to each of its rows, where

$$|A| = \begin{vmatrix} 1 & 2 & 3 \\ -2 & 1 & 2 \\ 3 & -1 & 4 \end{vmatrix}.$$

5. Compute each of the following determinants.

Not for Distribution

$$|A| = \begin{vmatrix} 2 & 3 & -4 \\ 0 & 6 & 0 \\ 0 & 0 & 5 \end{vmatrix} \qquad |B| = \begin{vmatrix} -2 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 3 & 8 \end{vmatrix}$$

A.4 Systems of linear equations

Section 4.1

1. Determine which of the following equations are linear.
   a. $x + 2y + z = 6$
   b. $2x - 6y + z = 1$
   c. $-\sqrt{2}x + 6^{-\frac{2}{3}}y = 4 - 3z$
   d. $3x_1 + 2x_2 + 4x_3 + 5x_4 = 1$
   e. $2xy + 3yz + 5z = 8$

2. For each of the following systems of linear equations, find its coefficient and augmented matrices
   1. $\begin{cases} 2x_1 - x_2 = 6 \\ 4x_1 + x_2 = 3 \end{cases}$
   2. $\begin{cases} -x_1 + 2x_2 + 3x_3 = 4 \\ 3x_1 + 2x_2 - 3x_3 = 5 \\ 2x_1 + 3x_2 - x_3 = 1 \end{cases}$
   3. $\begin{cases} x_1 - 3x_3 - 4 = 0 \\ 2x_2 - 5x_3 - 8 = 0 \\ 3x_1 + 2x_2 - x_3 = 4 \end{cases}$

3. For each of the following augmented matrices, find the corresponding linear system.

$$\left(\begin{array}{ccc|c} 1 & 2 & 3 & 1 \\ -1 & 1 & 0 & -2 \\ 0 & -1 & 1 & 0 \end{array}\right) \qquad \left(\begin{array}{cc|c} 2 & 1 & 1 \\ 0 & -1 & -2 \\ 0 & 3 & -1 \end{array}\right)$$

4. For each of the following linear systems, write it into the form $A\overrightarrow{X} = \overrightarrow{b}$ and express $\overrightarrow{b}$ as a linear combination of the column vectors of *A*.

a. $\begin{cases} -x_1 + x_2 + 2x_3 = 3 \\ 2x_1 + 6x_2 - 5x_3 = 2 \\ -3x_1 + 7x_2 - 5x_3 = -1 \end{cases}$

b. $\begin{cases} x_1 - x_2 + 4x_3 = 0 \\ -2x_1 + 4x_2 - 3x_3 = 0 \\ 3x_1 + 6x_2 - 8x_3 = 0, \end{cases}$

c. $\begin{cases} x_1 - 2x_2 + x_3 = 0 \\ -x_1 + 3x_2 - 4x_3 = 0 \end{cases}$

5. For each of the following points, $P_1\left(6,\ -\frac{1}{2}\right),\ P_2(8,\ 1),\ P_3(2,3),$ and $P_4(0,\ 0)$, verify whether it is a solution of the system

$$\begin{cases} x - 2y = 7, \\ x + 2y = 5. \end{cases}$$

6. Consider the system of linear equations

$$\begin{cases} -x_1 + x_2 + 2x_3 = 3 \\ 2x_1 + 6x_2 - 5x_3 = 2 \\ -3x_1 + 7x_2 - 5x_3 = -1 \end{cases}$$

1. Verify that $(1,\ 3,\ -1)$ is a solution of the system.
2. Verify that $\left(\frac{5}{6},\ \frac{7}{6},\ \frac{4}{3}\right)$ is a solution of the system.
3. Determine if the vector $\overrightarrow{b} = (3,\ 2,\ -1)^T$ is a linear combination of the column vectors of the coefficient matrix of the system.

4. Determine if the vector $\overrightarrow{b} = (3,\ 2,\ -1)^T$ belongs to the spanning space of the column vectors of the coefficient matrix of the system.

7. Solve each of the following systems.
   1. $\begin{cases} x_1 - x_2 = 6 \\ x_1 + x_2 = 5 \end{cases}$
   2. $\begin{cases} x_1 + 2x_2 = 1 \\ 2x_1 + x_2 = 2 \end{cases}$
   3. $\begin{cases} x_1 + 2x_2 = 1 \\ x_1 + 2x_2 = 2 \end{cases}$

A.5 Linear transformations

Section 5.1

1. For each of the following matrices,

$$A_1 = \begin{pmatrix} 2 & -2 & 4 & -1 \\ -1 & 2 & 3 & 0 \\ 0 & 1 & 2 & -1 \end{pmatrix} \qquad A_2 = \begin{pmatrix} -2 & 2 & 1 \\ 1 & 1 & -3 \\ 1 & 0 & -2 \end{pmatrix}$$

$$A_3 = \begin{pmatrix} 1 & -2 & 0 \\ -2 & 3 & 1 \\ 3 & -2 & 3 \\ 1 & 2 & -1 \end{pmatrix} \qquad A_4 = \begin{pmatrix} -1 & 1 & 1 \\ 2 & 0 & -2 \\ -1 & -1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$$

determine the domain and codomain of $T_{A_i}, i = 1, 2, 3, 4.$ Moreover, compute $T_{A_3}(0, 1, 1, 2), T_{A_2}(1, -2, 1), T_{A_3}(-1, 0, 1), T_{A_4}(-1, 1, 1).$

2. Let $A = \begin{pmatrix} 1 & 2 \\ -1 & 3 \end{pmatrix}$. Find $T_A \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $T_A \begin{pmatrix} 0 \\ 1 \end{pmatrix}$.

3. Let $A = \begin{pmatrix} -1 & 2 & 1 \\ 3 & 0 & 6 \end{pmatrix}$. Find $T_A(\overrightarrow{e_1}), T_A(\overrightarrow{e_2}), T_A(\overrightarrow{e_3})$, where $\overrightarrow{e_1}, \overrightarrow{e_2}, \overrightarrow{e_3}$ are the standard vectors in $\mathbb{R}^3$.

4. For each of the following linear transformations, express it in **(5.1.1)** .
   a. $T(x_1, x_2) = (x_1 - x_2, 2x_1 + x_2, 5x_1 - 3x_2).$
   b. $T(x_1, x_2, x_3, x_4) = (6x_1 - x_2 - x_3 + x_4, x_2 + x_3 - 2x_4, 3x_3 + x_4, 5x_4).$
   c. $T(x_1, x_2, x_3) = (4x_1 - x_2 + 2x_3, 3x_1 - x_3, 2x_1 + x_2 + 5x_3).$
   d. $T(x_1, x_2, x_3) = (8x_1 - x_1 + x_3, x_2 - x_3, 2x_1 + x_2, x_2 + 3x_3).$

5. Let $\overrightarrow{X} = (2, -2, -2, 1), \overrightarrow{Y} = (1, -2, 2, 3)$ and

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

Compute $T_A(\overrightarrow{X} - 3\overrightarrow{Y})$.

6. Let $S, T : \mathbb{R}^3 \to \mathbb{R}^3$ be linear operators such that $T(1,2,3) = (1,-4)$ and $S(1,2,3) = (4,9)$. Compute $(5T + 2S)(1,2,3)$.
7. Let $S, T : \mathbb{R}^3 \to \mathbb{R}^4$ be linear transformations such that $T(1,2,3) = (1,1,0,-1)$ and $S(1,2,3) = (2,-1,2,2)$. Compute $(3T - 2S)(1,2,3)$.
8. Compute $(T + 2S)(x_1, x_2, x_3, x_4)$, where

$$T(x_1, x_2, x_3, x_4) = (x_1 - x_2 + x_3, x_1 - x_2 + 2x_3 - x_4),$$
$$S(x_1, x_2, x_3, x_4) = (x_2 - x_3 + 2x_4, x_1 + x_3 - x_4).$$

9. For each pair of linear transformations T_A and T_B, find $T_B T_A$.
   a. $T_A(x_1, x_2) = (x_1 + x_2, x_1 - x_2), T_B(x_1, x_2) = (x_1 - x_2, 2x_1 - 3x_2, 3x_1 + x_2)$.
   b. $T_A(x_1, x_2, x_3) = (-x_1 - x_2 + x_3, x_1 + x_2), T_B(x_1, x_2) = (x_1 - x_2, 2x_1 + 3x_2)$.
   c. $T_A(x_1, x_2, x_3) = (-x_1 + x_2 - x_3, x_1 - 2x_2, x_3), T_B(x_1, x_2, x_3) = (x_1 - x_2 + x_3, 2x_1 + 3x_2 + x_3)$.

A.6 Lines and planes in $\mathbb{R}^3$

Section 5.4

1. Find $\vec{a} \times \vec{b}$ and verify $(\vec{a} \times \vec{b}) \perp \vec{a}$ and $(\vec{a} \times \vec{b}) \perp \vec{b}$.
    1. $\vec{a} = (1,\ 2,\ 3)$; $\vec{b} = (1,\ 0,\ 1)$,
    2. $\vec{a} = (1,\ 0,\ -3)$, $\vec{b} = (-1,\ 0,\ -2)$,
    3. $\vec{a} = (0,\ 2,\ 1)$; $\vec{b} = (-5,\ 0,\ 1)$,
    4. $\vec{a} = (1,\ 1,\ -1)$; $\vec{b} = (-1,\ 1,\ 0)$,

2. Let $\vec{a} = (1,\ -2,\ 3)$, $\vec{b} = (-1,\ -4,\ 0)$, and $\vec{c} = (2,\ 3,\ 1)$. Find the scalar triple product of $\vec{a}$, $\vec{b}$, $\vec{c}$.
3. Find the area of the parallelogram determined by $\vec{a}$ and $\vec{b}$.
    1. $\vec{a} = (1,\ 0,\ 1)$, $\vec{b} = (-1,\ 0,\ 1)$;
    2. $\vec{a} = (1,\ 0,\ 0)$, $\vec{b} = (-1,\ 1,\ -2)$.

4. Find the volume of the parallelpiped determined by $\vec{a}$, $\vec{b}$, and $\vec{c}$.
    1. $\vec{a} = (1,\ -1,\ 0)$, $\vec{b} = (-1,\ 0,\ 2)$, $\vec{c} = (0,\ -1,\ -1)$.
    2. $\vec{a} = (-1,\ -1,\ 0)$, $\vec{b} = (-1,\ 1,\ -2)$, $\vec{c} = (-1,\ 1,\ 1)$.

Not for Distribution

A.7 Bases and dimensions

Section 7.1

1. Let $\overrightarrow{a_1} = (2,\ 4,\ 6)^T$, $\overrightarrow{a_2} = (0,\ 0,\ 0)^T$, and $\overrightarrow{a_3} = (1,\ 2,\ 3)^T$. Determine whether $\left\{\overrightarrow{a_1},\ \overrightarrow{a_2},\ \overrightarrow{a_3}\right\}$ is linearly dependent.
2. Determine whether $\left\{\overrightarrow{a_1},\ \overrightarrow{a_2},\ \overrightarrow{a_3},\ \overrightarrow{a_4}\right\}$ is linearly dependent, where

$$\overrightarrow{a_1} = \begin{pmatrix} 1 \\ 2 \\ 1 \\ 0 \end{pmatrix},\quad \overrightarrow{a_2} = \begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \end{pmatrix},\quad \overrightarrow{a_3} = \begin{pmatrix} 0 \\ 2 \\ 0 \\ 0 \end{pmatrix},\quad \overrightarrow{a_4} = \begin{pmatrix} 0 \\ 2 \\ 0 \\ 1 \end{pmatrix}.$$

3. Determine whether $S = \left\{\overrightarrow{a_1},\ \overrightarrow{a_2},\ \overrightarrow{a_3},\ \overrightarrow{a_4}\right\}$ is linearly independent, where

$$\overrightarrow{a_1} = \begin{pmatrix} 1 \\ 3 \\ -4 \end{pmatrix},\quad \overrightarrow{a_2} = \begin{pmatrix} 7 \\ 9 \\ 8 \end{pmatrix},\quad \overrightarrow{a_3} = \begin{pmatrix} -1 \\ -2 \\ 1 \end{pmatrix},\quad \overrightarrow{a_4} = \begin{pmatrix} 6 \\ -2 \\ 5 \end{pmatrix}.$$

4. Determine that $S = \left\{\overrightarrow{a_1},\ \overrightarrow{a_2},\ \overrightarrow{a_3}\right\}$ is linearly independent, where

$$\overrightarrow{a_1} = \begin{pmatrix} -1 \\ 0 \\ -1 \end{pmatrix},\quad \overrightarrow{a_2} = \begin{pmatrix} 3 \\ -1 \\ -6 \end{pmatrix},\quad \overrightarrow{a_3} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}.$$

A.8 Eigenvalues and Diagonalizability

Section 8.1

1. For each of the following matrices, find its eigenvalues and determine which eigenvalues are repeated eigenvalues.

$$A = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix} \quad B = \begin{pmatrix} 2 & 0 & 0 \\ 1 & 1 & 0 \\ 2 & -3 & -4 \end{pmatrix} \quad C = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

2. For each of the following matrices, find its eigenvalues and eigenspaces.

$$A = \begin{pmatrix} 5 & 7 \\ 3 & 1 \end{pmatrix} \; B = \begin{pmatrix} 3 & -1 \\ 1 & 5 \end{pmatrix} \; C = \begin{pmatrix} 3 & 0 \\ 0 & 3 \end{pmatrix} \; D = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \\ 1 & 1 & 2 \end{pmatrix}$$

3. Let *A* be a 3×3 matrix. Assume that $2I - A,\ 3I - A,$ and $4I - A$ are not invertible.
    1. Find $|A|$ and tr(*A*).
    2. Prove that $I + 5A$ is invertible.

4. Assume that the eigenvalues of a 3×3 matrix *A* are 1, 2, 3. Compute $\left|A^3 - 2A^2 + 3A + I\right|$.

A.9 Vector spaces

1. Let $W = \{(x,\ y) \in \mathbb{R}^2 : x - y = 0\}$. Show that *W* is a subspace of $\mathbb{R}^2$.
2. Let $W = \{(x,\ y) \in \mathbb{R}^2_+ : x + y \leq 1\}$. Show that *W* is not a subspace of $\mathbb{R}^2$.
3. Let

$$\mathbb{R}^2_+ = \{(x,\ y) : x \geq 0 \quad \text{and } y \geq 0\}.$$

   i. Show that if $\overrightarrow{a} \in \mathbb{R}^2_+$ and $\overrightarrow{b} \in \mathbb{R}^2_+$, then $\overrightarrow{a} + \overrightarrow{b} \in \mathbb{R}^2_+$.
   ii. Show that $\mathbb{R}^2_+$ is not a subspace of $\mathbb{R}^2$ by using **Definition 9.1.1** .

4. Let

$$W = \mathbb{R}^2_+ \cup (-\mathbb{R}^2_+).$$

   i. Show that if $\overrightarrow{a} \in W$ and *k* is a real number, then $k\overrightarrow{a} \in W$.
   ii. Show that *W* is not a subspace of $\mathbb{R}^2$ by using **definition 9.1.1** .

5. Let $W = \{(x,\ y,\ z) \in \mathbb{R}^3 : x - 2y - z = 0\}$. Show that *W* is a subspace of $\mathbb{R}^3$ by **Definition 9.1.1** and **Theorem 9.1.1** , respectively.
6. Show that *W* is a subspace of $\mathbb{R}^4$, where

$$W = \left\{(x,\ y,\ z,\ w) \in \mathbb{R}^4 : \begin{cases} x - 2y + 2z - w = 0, \\ -x + 3y + 4z + 2w = 0 \end{cases}\right\}.$$

7. Show that *W* is a subspace of $\mathbb{R}^3$, where

$$W=\left\{(x,\ y,\ z)\in\mathbb{R}^3:\left\{\begin{array}{l}x+2y+3z=0,\\-x+y-4z=0,\\x-2y+5z=0\end{array}\right.\right\}.$$

8. Let $W=\left\{(x,\ y,\ z)\in\mathbb{R}^3:x-2y-z=3\right\}$. Show that *W* is not a subspace of $\mathbb{R}^3$ by **Definition 9.1.1** .
9. Let $W=[0,\ 1]$. Show that *W* is not a subspace of *R*.
10. Show that all the subspaces of $\mathbb{R}$ are {0} and $\mathbb{R}$.
11. Let $W=\{(0,\ 0,\ \cdots,\ 0)\}\in\mathbb{R}^n$. Show that *W* is a subspace of $\mathbb{R}^n$.
12. Let $\overrightarrow{v}\in\mathbb{R}^n$ be a given nonzero vector. Show that

$$W=\left\{t\overrightarrow{v}:t\in\mathbb{R}\right\}$$

is a subspace of $\mathbb{R}^n$ by **Definition 9.1.1** .

A.10 Complex numbers

Section 10.1

1. Solve each of the following equations.
   a. $z^2 + z + 1 = 0$
   b. $z^2 - 2z + 5 = 0$
   c. $z^2 - 8z + 25 = 0$

2. Use the following information to find *x* and *y*.
   a. $2x - 4i = 4 + 2yi$
   b. $(x - y) + (x + y)i = 2 - 4i$

3. Let $z_1 = 2 + i$ and $z_2 = 1 - i$. Express each of the following complex numbers in the form $x + yi$, calculate its modulus, and find its conjugate.
   1. $z_1 + 2z_2$
   2. $2z_1 - 3z_2$
   3. $z_1 z_2$
   4. $\frac{z_1}{z_2}$

4. Express each of the following complex numbers in the form $x + yi$.
   a. $\overline{\left(\frac{i}{1-i}\right)}$
   b. $\frac{2-i}{(1-i)(2+i)}$
   c. $\frac{1+i}{i(1+i)(2-i)}$
   d. $\frac{1+i}{1-i} - \frac{2-i}{1+i}$

5. Find $i^2,\ i^3,\ i^4,\ i^5,\ i^6$ and compute $\left(-i - i^4 + i^5 - i^6\right)^{1000}$.

6. Let $z = \left(\frac{1-i}{1+i}\right)^8$. Calculate $z^{66} + 2z^{33} - 2$.

7. Find $z \in \mathbb{C}$

   a. $iz = 4 + 3i$

   b. $(1+i)\bar{z} = (1-i)z$

8. Find real numbers *x*, *y* such that

$$\frac{x + 1 + i(y-3)}{5+3i} = 1 + i.$$

9. Let $z = \frac{(3+4i)(1+i)^{12}}{i^5(2+4i)^2}$. Find $|z|$.

Index