MATH 136: Linear Algebra 1 for Honours Mathematics

Fall 2023 Edition

1st edition 2020 by Conrad Hewitt and Francine Vinette

We would like to acknowledge contributions from:
Shane Bauman, Judith Koeller, Anton Mosunov, Graeme Turner, and Owen Woody

Contents

Chapter 1

Vectors in $\mathbb{R}^n$

1.1 Introduction

Linear algebra is used widely in the social sciences, business, science, and engineering. Vectors are used in the sciences for displacement, velocity, acceleration, force, and many other important physical concepts. Informally, a *vector* is an object that has both magnitude and direction. For instance, the velocity of a boat on a river can be represented by a vector that encodes both its magnitude (speed) and its direction of travel.

In mathematics, a vector is represented by a column of numbers, as in the following definition.

Definition 1.1.1
$\mathbb{R}^n$, Vector in $\mathbb{R}^n$

The set $\mathbb{R}^n$ is defined as $\left\{ \vec{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} : x_1, \ldots, x_n \in \mathbb{R} \right\}$.

A **vector** is an element $\vec{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$ of $\mathbb{R}^n$.

REMARK

In these notes, a vector is indicated by an accented right arrow: $\vec{v}$.

Other texts use different notation for vectors, such as $\mathbf{v}$ or $\underline{v}$.

Example 1.1.2

The following are vectors: $\vec{w} = \begin{bmatrix} 2 \\ 3 \end{bmatrix} \in \mathbb{R}^2$, $\quad \vec{v} = \begin{bmatrix} -3 \\ 1 \\ -5 \end{bmatrix} \in \mathbb{R}^3$, and $\vec{u} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_{100} \end{bmatrix} \in \mathbb{R}^{100}$.

We will usually write vectors in **column notation**. It can be more convenient and compact to write a vector in **row notation**, which includes the superscript "T" (short for transpose) and clearly separates the components of the vectors with spaces, as follows.

Definition 1.1.3
Row Notation for a Vector

The **row notation** for the vector $\vec{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$ is $\vec{v} = \begin{bmatrix} v_1 & v_2 & \cdots & v_n \end{bmatrix}^T$.

REMARK

Many texts do not distinguish between $\begin{bmatrix} v_1 & v_2 & \cdots & v_n \end{bmatrix}$ and $\begin{bmatrix} v_1 & v_2 & \cdots & v_n \end{bmatrix}^T$.
For our purposes, this distinction is important, i.e. $\begin{bmatrix} v_1 & v_2 & \ldots & v_n \end{bmatrix} \neq \begin{bmatrix} v_1 & v_2 & \ldots & v_n \end{bmatrix}^T$.

1.2 Algebraic and Geometric Representation of Vectors

A vector's **algebraic representation** consists of a column of numbers, e.g. $\vec{v} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$.

A vector's **geometric representation** consists of a directed line segment in $\mathbb{R}^n$.

Vectors are not **localized**; that is, if two vectors have the same magnitude and direction, but different starting points, we consider those vectors to be equal. For convenience, we often choose the starting point of a vector to be the origin, O.

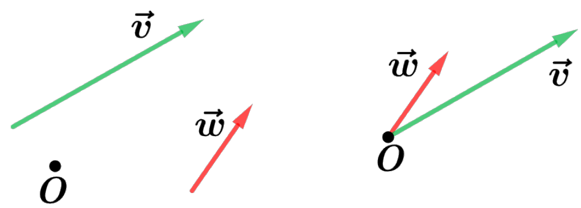


Figure 1.2.1: Vectors and the Origin

Relationship between Geometric and Algebraic Representations

Consider the *geometric representation* for a vector $\vec{v}$ in $\mathbb{R}^n$, which is a directed line segment. We typically consider the initial point of $\vec{v}$ to be the origin O and call its terminal point V (i.e., we use the same letter as the vector). If the point V has coordinates $(v_1, v_2, \ldots, v_n)$, then we write the vector $\vec{v}$ as

$$\vec{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$$

This n-tuple is the *algebraic representation* of $\vec{v}$.

On the other hand, if the algebraic representation of a vector $\vec{v}$ in $\mathbb{R}^n$ is $\begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$, then its geometric interpretation is a directed line segment from the initial point, O, to the terminal point, V, with co-ordinates $(v_1, v_2, \ldots, v_n)$.

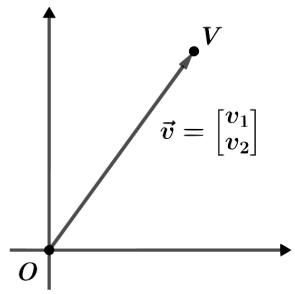


Figure 1.2.2: The geometric representation of $\vec{v} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$ in $\mathbb{R}^2$

We are free to move the vector $\vec{v}$ around, as long as we maintain its magnitude and direction; however, moving it will change both the initial and terminal points. We say that the point V is the terminal point associated with the vector $\vec{v}$ (using the initial point as O), or more simply, we say that V is the point associated with the vector $\vec{v}$.

Example 1.2.1

The point $V = (2, -4, 9)$ is the terminal point associated with the vector $\vec{v} = \begin{bmatrix} 2 \\ -4 \\ 9 \end{bmatrix}$ in $\mathbb{R}^3$.

1.3 Operations on Vectors

Definition 1.3.1
Equality of Vectors in $\mathbb{R}^n$

We say that vectors $\vec{u} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}$ and $\vec{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$ in $\mathbb{R}^n$ are **equal** if $u_i = v_i$ for all $i = 1, \ldots, n$.
We denote this by writing $\vec{u} = \vec{v}$.

The vector equation $\vec{u} = \vec{v}$ is therefore equivalent to a system of n equations, one for each component.

Example 1.3.2

If $\vec{u} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$ and $\vec{v} = \begin{bmatrix} 1 \\ 2 \\ 4 \end{bmatrix}$, then $\vec{u} \neq \vec{v}$ because $u_3 \neq v_3$.

Geometrically, two vectors $\vec{u}$ and $\vec{v}$ are said to be **equal** if they have the same magnitude and the same direction.

Definition 1.3.3
Addition in $\mathbb{R}^n$

Let $\vec{u} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}, \vec{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} \in \mathbb{R}^n$. Then $\vec{u} + \vec{v} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} + \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = \begin{bmatrix} u_1 + v_1 \\ u_2 + v_2 \\ \vdots \\ u_n + v_n \end{bmatrix}$.

Example 1.3.4

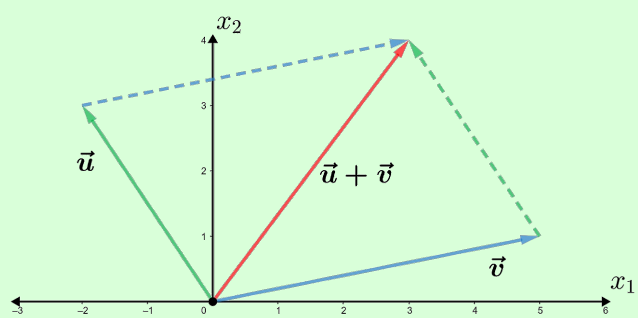
We have $\begin{bmatrix} 2 \\ 4 \\ 6 \end{bmatrix} + \begin{bmatrix} 3 \\ -5 \\ 7 \end{bmatrix} = \begin{bmatrix} 2+3 \\ 4-5 \\ 6+7 \end{bmatrix} = \begin{bmatrix} 5 \\ -1 \\ 13 \end{bmatrix}$.

Geometrically, the vector $\vec{w} = \vec{u} + \vec{v}$ follows the Parallelogram Law, illustrated in the following example with $\vec{v}$ moved so that its initial point is the terminal point of $\vec{u}$.

Example 1.3.5

Add $\vec{u} = \begin{bmatrix} -2 \\ 3 \end{bmatrix}$ to $\vec{v} = \begin{bmatrix} 5 \\ 1 \end{bmatrix}$ in $\mathbb{R}^2$. Show the result geometrically.

Solution: $\vec{u} + \vec{v} = \begin{bmatrix} -2 \\ 3 \end{bmatrix} + \begin{bmatrix} 5 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$.



REMARK

In this course, most of the time we will be labelling axes in $\mathbb{R}^2$ as x_1, x_2 instead of x, y. Similarly, in $\mathbb{R}^3$ we will be labelling axes as x_1, x_2, x_3 instead of x, y, z. We recommend you to follow that practice as well.

Example 1.3.6

Suppose a boat is moving with a velocity of $\vec{u}$ and a person on the boat walks with velocity $\vec{v}$ on (relative to) the boat. An observer on the shore sees the person move with velocity $\vec{w} = \vec{u} + \vec{v}$, and thus sees the combined velocity of the person and the boat.

The following rules of vector addition follow from the properties of addition of real numbers.

Proposition 1.3.7 **(Properties of Vector Addition)**

Let $\vec{u}, \vec{v}, \vec{w} \in \mathbb{R}^n$.

(a) $\vec{u} + \vec{v} = \vec{v} + \vec{u}$ (**symmetry**)

(b) $\vec{u} + \vec{v} + \vec{w} = \vec{u} + (\vec{v} + \vec{w}) = (\vec{u} + \vec{v}) + \vec{w}$ (**associativity**)

(c) There is a **zero vector**, $\vec{0} = \begin{bmatrix} 0 & 0 & \cdots & 0 \end{bmatrix}^T$ in $\mathbb{R}^n$, with the property that $\vec{v} + \vec{0} = \vec{0} + \vec{v} = \vec{v}$.

EXERCISE

Prove Proposition 1.3.7 (Properties of Vector Addition).

For many of the results in this chapter, the proof is left as an exercise for you.

Definition 1.3.8
Additive Inverse

Let $\vec{u} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} \in \mathbb{R}^n$. The **additive inverse** of $\vec{u}$, denoted $-\vec{u}$, is defined as

$$-\vec{u} = \begin{bmatrix} -u_1 \\ -u_2 \\ \vdots \\ -u_n \end{bmatrix}.$$

Example 1.3.9

If $\vec{u} = \begin{bmatrix} 1 \\ -4 \\ 2 \end{bmatrix}$, then $-\vec{u} = \begin{bmatrix} -1 \\ 4 \\ -2 \end{bmatrix}$.

We now notice the key property of the additive inverse.

Proposition 1.3.10 **(Additive Inverse Property)**

If $\vec{u} \in \mathbb{R}^n$, we have

$$\vec{u} - \vec{u} = \vec{u} + (-\vec{u}) = (-\vec{u}) + \vec{u} = \vec{0}.$$

Thus $-\vec{u}$ has the effect of "cancelling" the vector $\vec{u}$ when addition is performed (hence the term *additive* inverse).

REMARK

The vector $-\vec{u}$ has the same magnitude as $\vec{u}$, but the opposite direction.

Definition 1.3.11
Subtraction in $\mathbb{R}^n$

Let $\vec{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$, $\vec{u} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}$ and $-\vec{u} = \begin{bmatrix} -u_1 \\ -u_2 \\ \vdots \\ -u_n \end{bmatrix}$ be vectors in $\mathbb{R}^n$. We define

$$\vec{v} - \vec{u} = \vec{v} + (-\vec{u}) = \begin{bmatrix} v_1 - u_1 \\ v_2 - u_2 \\ \vdots \\ v_n - u_n \end{bmatrix}.$$

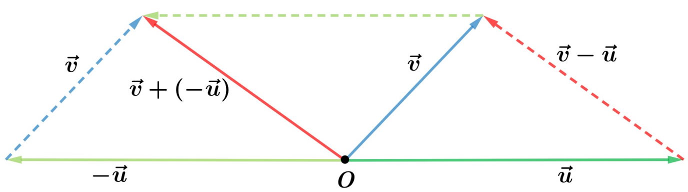


Figure 1.3.3: General geometric interpretation of $\vec{v} - \vec{u}$

Example 1.3.12

We have $\begin{bmatrix} 2 \\ 4 \\ 6 \end{bmatrix} - \begin{bmatrix} 3 \\ -5 \\ 7 \end{bmatrix} = \begin{bmatrix} 2-3 \\ 4+5 \\ 6-7 \end{bmatrix} = \begin{bmatrix} -1 \\ 9 \\ -1 \end{bmatrix}$.

For vectors in $\mathbb{R}^n$ with $n > 1$, we do not define a component-wise "vector multiplication", as we did for addition and subtraction, since such an operation turns out to be of little use in linear algebra. There are different types of "vector products" that we *do* define, however. These will be discussed later in this chapter.

In the meantime, we will define the **scalar multiplication** of a vector by a real number c.

Definition 1.3.13
Scalar Multiplication

Let $c \in \mathbb{R}$ and $\vec{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} \in \mathbb{R}^n$. We define $c\vec{v} = c \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = \begin{bmatrix} cv_1 \\ cv_2 \\ \vdots \\ cv_n \end{bmatrix}$.

We say that the vector $\vec{v}$ is **scaled** by c.

For this reason, real numbers are often referred to as **scalars**.

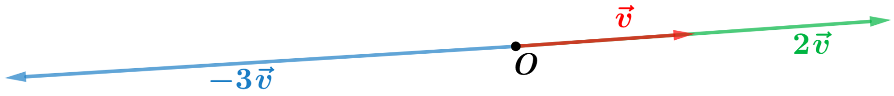


Figure 1.3.4: Scalar multiplication of $\vec{v}$

REMARKS

- When $c > 0$, the vector $c\vec{v}$ points in the direction of $\vec{v}$ and is c times as long.
- When $c < 0$, then $c\vec{v}$ points in the opposite direction to $\vec{v}$ and is $|c|$ times as long.
- The vector $(-1)\vec{v} = -\vec{v}$ has the same length as $\vec{v}$ but points in the opposite direction.

Proposition 1.3.14 **(Properties of Scalar Multiplication)**

Let $c, d \in \mathbb{R}$ and $\vec{u}, \vec{v} \in \mathbb{R}^n$.

(a) $(c+d)\vec{v} = c\vec{v} + d\vec{v}$.

(b) $(cd)\vec{v} = c(d\vec{v})$.

(c) $c(\vec{u} + \vec{v}) = c\vec{u} + c\vec{v}$.

(d) $0\vec{v} = \vec{0}$.

(e) If $c\vec{v} = \vec{0}$, then $c = 0$ or $\vec{v} = \vec{0}$. (**cancellation law**)

Property (e) may look especially familiar, because the following result is often used in mathematics when dealing with real numbers:

For all $a, b \in \mathbb{R}$, if $ab = 0$, then $a = 0$ or $b = 0$.

Definition 1.3.15
Standard Basis for $\mathbb{R}^n$

In $\mathbb{R}^n$, let $\vec{e_i}$ be the vector whose i^{th} component is 1 with all other components 0. The set $\mathcal{E} = \{\vec{e_1}, \vec{e_2}, \ldots, \vec{e_n}\}$ is called the **standard basis for** $\mathbb{R}^n$.

Example 1.3.16

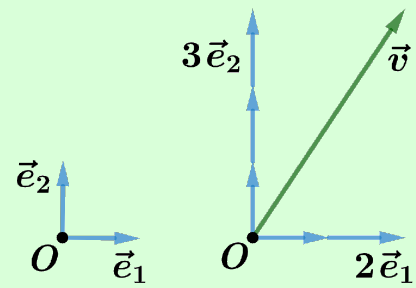
The standard basis for $\mathbb{R}^3$ is $\{\vec{e_1}, \vec{e_2}, \vec{e_3}\} = \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}$.

Definition 1.3.17
Components

If $\vec{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = v_1\vec{e_1} + v_2\vec{e_2} + \ldots + v_n\vec{e_n}$, then we call $v_1, v_2, \ldots, v_n$ the **components of** $\vec{v}$.

Example 1.3.18

In $\mathbb{R}^2$, $\vec{v} = \begin{bmatrix} 2 \\ 3 \end{bmatrix} = 2\vec{e_1} + 3\vec{e_2} = 2\begin{bmatrix} 1 \\ 0 \end{bmatrix} + 3\begin{bmatrix} 0 \\ 1 \end{bmatrix}$.



Example 1.3.19

If $\vec{v} = \begin{bmatrix} 5 \\ 3 \\ -4 \end{bmatrix} \in \mathbb{R}^3$, then $\vec{v} = 5\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + 3\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} - 4\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = 5\vec{e_1} + 3\vec{e_2} - 4\vec{e_3}$.
The components of $\vec{v}$ are 5, 3 and -4.

1.4 Vectors in $\mathbb{C}^n$

Vectors are also defined in $\mathbb{C}^n$, where each component is a complex number. Equality, addition, subtraction, scalar multiplication, and the cancellation law in $\mathbb{C}^n$ are all defined as in the previous propositions, replacing $\mathbb{R}$ with $\mathbb{C}$.

Definition 1.4.1
$\mathbb{C}^n$, Vector in $\mathbb{C}^n$

The set $\mathbb{C}^n$ is defined as $\left\{ \vec{z} = \begin{bmatrix} z_1 \\ \vdots \\ z_n \end{bmatrix} : z_1, \ldots, z_n \in \mathbb{C} \right\}$.

A **vector** is an element $\vec{z} = \begin{bmatrix} z_1 \\ \vdots \\ z_n \end{bmatrix}$ of $\mathbb{C}^n$.

Example 1.4.2

We have $\vec{w} = \begin{bmatrix} 1-2i \\ 7 \end{bmatrix} \in \mathbb{C}^2$ and $\vec{z} = \begin{bmatrix} 1+i \\ 2-3i \\ -4+5i \end{bmatrix} \in \mathbb{C}^3$.

Addition, subtraction, and scalar multiplication in $\mathbb{C}^n$ are defined as in $\mathbb{R}^n$.

Example 1.4.3

The expression $3\begin{bmatrix} 1+i \\ 2+i \end{bmatrix} - 2\begin{bmatrix} 2-i \\ 4+i \end{bmatrix}$ evaluates to

$$3\begin{bmatrix} 1+i \\ 2+i \end{bmatrix} - 2\begin{bmatrix} 2-i \\ 4+i \end{bmatrix} = \begin{bmatrix} 3+3i \\ 6+3i \end{bmatrix} - \begin{bmatrix} 4-2i \\ 8+2i \end{bmatrix} = \begin{bmatrix} -1+5i \\ -2+i \end{bmatrix}.$$

Example 1.4.4 If $c \in \mathbb{C}$ and $\vec{z}, \vec{w} \in \mathbb{C}^n$ and $(c-4-i)(\vec{z}-\vec{w}) = \vec{0}$, then $c = 4+i$ or $\vec{z} = \vec{w}$.

The standard basis $\{\vec{e_1}, \vec{e_2}, ..., \vec{e_n}\}$ for $\mathbb{C}^n$ is the same as the standard basis for $\mathbb{R}^n$.

Definition 1.4.5
Standard Basis for $\mathbb{C}^n$

In $\mathbb{C}^n$, let $\vec{e_i}$ represent the vector whose i^{th} component is 1 with all other components 0. The set $\{\vec{e_1}, \vec{e_2}, ..., \vec{e_n}\}$ is called the **standard basis for** $\mathbb{C}^n$.

Example 1.4.6 The standard basis for $\mathbb{C}^4$ is $\{\vec{e_1}, \vec{e_2}, \vec{e_3}, \vec{e_4}\} = \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \right\}$.

Definition 1.4.7
Components

If $\vec{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = v_1\vec{e_1} + v_2\vec{e_2} + ... + v_n\vec{e_n}$, then we refer to $v_1, v_2, ..., v_n$ as the **components of** $\vec{v}$.

Example 1.4.8 If $\vec{v} = \begin{bmatrix} 5+i \\ 3i \\ -4 \end{bmatrix} \in \mathbb{C}^3$, then $\vec{v} = (5+i)\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + 3i\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} - 4\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = (5+i)\vec{e_1} + 3i\,\vec{e_2} - 4\vec{e_3}$.
The components of $\vec{v}$ are $5+i$, $3i$ and -4.

1.5 Dot Product in $\mathbb{R}^n$

The dot product is a function that takes as input two vectors in $\mathbb{R}^n$ and returns a real number. (Later, we will adapt this idea to vectors in $\mathbb{C}^n$ using a different operation.)

Definition 1.5.1
Dot Product

Let $\vec{u} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}$ and $\vec{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$ be vectors in $\mathbb{R}^n$. We define their **dot product** by

$$\vec{u} \cdot \vec{v} = u_1v_1 + u_2v_2 + \cdots + u_nv_n.$$

Example 1.5.2 We have $\begin{bmatrix} 2 \\ 3 \end{bmatrix} \cdot \begin{bmatrix} 5 \\ 7 \end{bmatrix} = 2(5)+3(7) = 31$ and $\begin{bmatrix} 3 \\ -5 \\ 2 \end{bmatrix} \cdot \begin{bmatrix} -4 \\ 4 \\ -2 \end{bmatrix} = 3(-4)+(-5)(4)+2(-2) = -36$.

Proposition 1.5.3 **(Properties of the Dot Product)**

If $c \in \mathbb{R}$ and $\vec{u}, \vec{v}, \vec{w}$ are vectors in $\mathbb{R}^n$, then:

(a) $\vec{u} \cdot \vec{v} = \vec{v} \cdot \vec{u}$ (**symmetry**)

(b) $(\vec{u} + \vec{v}) \cdot \vec{w} = \vec{u} \cdot \vec{w} + \vec{v} \cdot \vec{w}$

(c) $(c\vec{u}) \cdot \vec{v} = c(\vec{u} \cdot \vec{v})$

(collectively known as **linearity**)

(d) $\vec{v} \cdot \vec{v} \geq 0$, with $\vec{v} \cdot \vec{v} = 0$ if and only if $\vec{v} = \vec{0}$. (**non-negativity**)

An important use of the dot product is determining the length of a vector.

Definition 1.5.4 **Length in $\mathbb{R}^n$, Norm in $\mathbb{R}^n$**

The **length** (or **norm** or **magnitude**) of the vector $\vec{v} \in \mathbb{R}^n$ is $\|\vec{v}\| = \sqrt{\vec{v} \cdot \vec{v}}$.

Example 1.5.5

We have $\left\| \begin{bmatrix} 2 \\ 4 \end{bmatrix} \right\| = \sqrt{\begin{bmatrix} 2 \\ 4 \end{bmatrix} \cdot \begin{bmatrix} 2 \\ 4 \end{bmatrix}} = \sqrt{2(2) + 4(4)} = \sqrt{20} = 2\sqrt{5}$.

Example 1.5.6

We have $\left\| \begin{bmatrix} 2 \\ -3 \\ 4 \end{bmatrix} \right\| = \sqrt{\begin{bmatrix} 2 \\ -3 \\ 4 \end{bmatrix} \cdot \begin{bmatrix} 2 \\ -3 \\ 4 \end{bmatrix}} = \sqrt{2(2) + (-3)(-3) + 4(4)} = \sqrt{29}$.

REMARKS

The notation "$\|$" is used in order to distinguish from the notation for the absolute value of a real number, "$|$".

Recall that when we take the square root, we mean the positive square root, so that $\sqrt{4} = 2$ (not ± 2). Notice that it follows from Property (d) of Proposition 1.5.3 (Properties of the Dot Product) that every non-zero vector has length greater than zero, and the zero vector $\vec{0}$ has a length of zero.

Proposition 1.5.7

If $c \in \mathbb{R}$ and $\vec{v} \in \mathbb{R}^n$, then $\|c\vec{v}\| = |c|\, \|\vec{v}\|$.

Example 1.5.8

We have $\left\| \begin{bmatrix} -3 \\ -6 \end{bmatrix} \right\| = \left\| -3 \begin{bmatrix} 1 \\ 2 \end{bmatrix} \right\| = |-3| \; \left\| \begin{bmatrix} 1 \\ 2 \end{bmatrix} \right\| = 3\sqrt{1(1) + 2(2)} = 3\sqrt{5}$.

Example 1.5.9

We have $\left\| \begin{bmatrix} -2 \\ -6 \\ 4 \end{bmatrix} \right\| = \left\| -2 \begin{bmatrix} 1 \\ 3 \\ -2 \end{bmatrix} \right\| = |-2| \; \left\| \begin{bmatrix} 1 \\ 3 \\ -2 \end{bmatrix} \right\| = 2\sqrt{1^2 + 3^2 + (-2)^2} = 2\sqrt{14}$.

Definition 1.5.10
Unit Vector

We say that $\vec{v} \in \mathbb{R}^n$ is a **unit vector** if $\|\vec{v}\| = 1$.

Example 1.5.11

If $\vec{u} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$, then $\|\vec{u}\| = \sqrt{13}$, so $\vec{u}$ is not a unit vector.

Example 1.5.12

If $\vec{v} = \begin{bmatrix} \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & \frac{2}{\sqrt{6}} \end{bmatrix}^T = \frac{1}{\sqrt{6}} \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}$, then $\|\vec{v}\| = \frac{1}{\sqrt{6}}\sqrt{1^2 + 1^2 + 2^2} = 1$, so $\vec{v}$ is a unit vector.

Definition 1.5.13
Normalization

When $\vec{v} \in \mathbb{R}^n$ is a non-zero vector, we can produce a unit vector

$$\hat{v} = \frac{\vec{v}}{\|\vec{v}\|}$$

in the direction of $\vec{v}$ by scaling $\vec{v}$. This process is called **normalization**.

REMARKS

The vector $\hat{v}$ has the same direction as the vector $\vec{v}$, since $\hat{v} = \frac{\vec{v}}{\|\vec{v}\|} = \frac{1}{\|\vec{v}\|}\vec{v}$.

The vector $\hat{v}$ is a unit vector, since $\|\hat{v}\| = \left\| \frac{1}{\|\vec{v}\|}\vec{v} \right\| = \left| \frac{1}{\|\vec{v}\|} \right| \|\vec{v}\| = 1$.

Example 1.5.14

The unit vector in the direction of $\vec{v} = \begin{bmatrix} 2 \\ -4 \\ -1 \end{bmatrix}$ is the vector $\hat{v} = \frac{\vec{v}}{\|\vec{v}\|} = \frac{1}{\sqrt{21}} \begin{bmatrix} 2 \\ -4 \\ -1 \end{bmatrix}$.

The dot product can also be used to determine the angle between two vectors. In $\mathbb{R}^2$ this can be seen as follows.

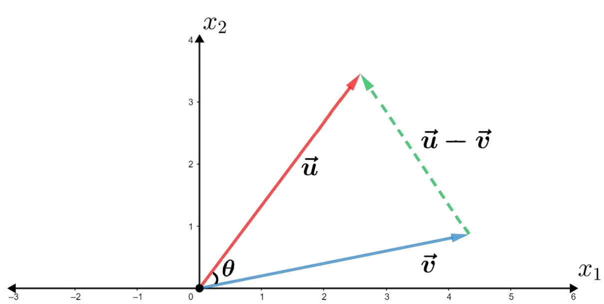


Figure 1.5.5: The angle θ between $\vec{u}$ and $\vec{v}$ in $\mathbb{R}^2$

If θ is the angle between $\vec{u}, \vec{v} \in \mathbb{R}^2$, then the law of cosines says that

$$\|\vec{u} - \vec{v}\|^2 = \|\vec{u}\|^2 + \|\vec{v}\|^2 - 2\|\vec{u}\|\|\vec{v}\| \cos\theta.$$

On the other hand, recalling the definition of the length $\|\cdot\|$ in terms of the dot product, we have

$$\begin{aligned}\|\vec{u} - \vec{v}\|^2 &= (\vec{u} - \vec{v}) \cdot (\vec{u} - \vec{v}) \\ &= \vec{u} \cdot \vec{u} - \vec{u} \cdot \vec{v} - \vec{v} \cdot \vec{u} + \vec{v} \cdot \vec{v} \\ &= \|\vec{u}\|^2 - 2\vec{v} \cdot \vec{u} + \|\vec{v}\|^2.\end{aligned}$$

If we combine the above two equations, we arrive at

$$\vec{u} \cdot \vec{v} = \|\vec{u}\|\|\vec{v}\| \cos\theta.$$

This gives us a relationship between the angle θ between the vectors $\vec{u}, \vec{v} \in \mathbb{R}^2$ and their dot product. This motivates the following definition.

Definition 1.5.15

Angle Between Vectors

Let $\vec{u}$ and $\vec{v}$ be non-zero vectors in $\mathbb{R}^n$. The **angle** θ, in radians $(0 \le \theta \le \pi)$, between $\vec{u}$ and $\vec{v}$ is such that

$$\vec{u} \cdot \vec{v} = \|\vec{u}\|\|\vec{v}\| \cos\theta, \text{ that is, } \theta = \arccos\left(\frac{\vec{u} \cdot \vec{v}}{\|\vec{u}\|\|\vec{v}\|}\right).$$

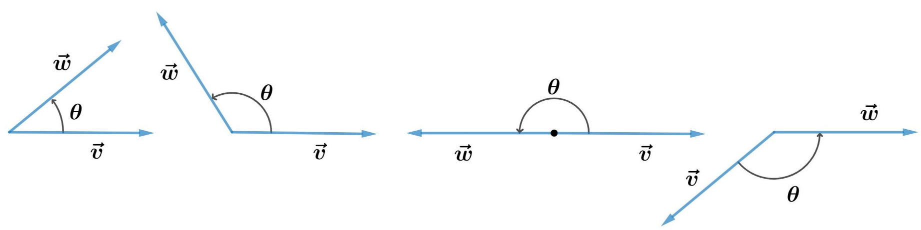


Figure 1.5.6: The angle θ between $\vec{v}$ and $\vec{w}$ is between 0 and π

REMARK

In order for the above definition to be well-defined, we need to know that $\frac{\vec{u} \cdot \vec{v}}{\|\vec{u}\|\|\vec{v}\|}$ lies in in the domain of arccos—otherwise $\arccos\left(\frac{\vec{u} \cdot \vec{v}}{\|\vec{u}\|\|\vec{v}\|}\right)$ is undefined. That is, we need to know that

$$\frac{\vec{u} \cdot \vec{v}}{\|\vec{u}\|\|\vec{v}\|} \in [-1, 1],$$

or, equivalently, that

$$|\vec{u} \cdot \vec{v}| \le \|\vec{u}\|\|\vec{v}\| \text{ for all } \vec{u}, \vec{v} \in \mathbb{R}^n.$$

This inequality is true and is known as the *Cauchy–Schwarz Inequality*. We will take its validity for granted in these notes, and will not provide a proof.

Example 1.5.16

Determine the angle θ between $\begin{bmatrix}1\\4\end{bmatrix}$ and $\begin{bmatrix}-2\\3\end{bmatrix}$ in $\mathbb{R}^2$, rounding your answer to three decimal places throughout.

Solution: $\theta = \arccos\left(\dfrac{\begin{bmatrix}1\\4\end{bmatrix}\cdot\begin{bmatrix}-2\\3\end{bmatrix}}{\left\|\begin{bmatrix}1\\4\end{bmatrix}\right\|\left\|\begin{bmatrix}-2\\3\end{bmatrix}\right\|}\right) = \arccos\left(\dfrac{10}{\sqrt{17}\sqrt{13}}\right) \approx 0.833.$

Example 1.5.17

Determine the angle, θ, between $\begin{bmatrix}1\\3\\5\end{bmatrix}$ and $\begin{bmatrix}2\\-4\\3\end{bmatrix}$ in $\mathbb{R}^3$, rounding your answer to three decimal places throughout.

Solution: $\theta = \arccos\left(\dfrac{\begin{bmatrix}1\\3\\5\end{bmatrix}\cdot\begin{bmatrix}2\\-4\\3\end{bmatrix}}{\left\|\begin{bmatrix}1\\3\\5\end{bmatrix}\right\|\left\|\begin{bmatrix}2\\-4\\3\end{bmatrix}\right\|}\right) = \arccos\left(\dfrac{5}{\sqrt{35}\sqrt{29}}\right) \approx 1.413.$

Definition 1.5.18
Orthogonal in $\mathbb{R}^n$

We say that the two vectors $\vec{u}$ and $\vec{v}$ in $\mathbb{R}^n$ are **orthogonal** (or **perpendicular**) if $\vec{u}\cdot\vec{v} = 0$.

REMARK

Our definition of the angle between two vectors, Definition 1.5.15, is consistent with our definition of orthogonality. Examining our new dot product formula,

$$\vec{v}\cdot\vec{w} = \|\vec{v}\|\|\vec{w}\|\cos\theta,$$

we see that if two vectors are orthogonal, then their dot product is zero, and thus conclude that the angle θ between them is $\frac{\pi}{2}$, since $\cos\frac{\pi}{2} = 0$ and $0 \le \theta \le \pi$.

Example 1.5.19

The vectors $\vec{u} = \begin{bmatrix}1\\5\\-3\end{bmatrix}$ and $\vec{v} = \begin{bmatrix}1\\1\\2\end{bmatrix}$ are orthogonal, since $\vec{u}\cdot\vec{v} = \begin{bmatrix}1\\5\\-3\end{bmatrix}\cdot\begin{bmatrix}1\\1\\2\end{bmatrix} = 0$.

Example 1.5.20

The vectors $\vec{w} = \begin{bmatrix}1\\-3\\5\end{bmatrix}$ and $\vec{x} = \begin{bmatrix}2\\2\\1\end{bmatrix}$ are not orthogonal, since $\vec{w}\cdot\vec{x} = \begin{bmatrix}1\\-3\\5\end{bmatrix}\cdot\begin{bmatrix}2\\2\\1\end{bmatrix} = 1 \neq 0$.

REMARK

Every vector $\vec{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} \in \mathbb{R}^n$ is orthogonal to $\vec{0}$, as $\vec{0} \cdot \vec{v} = 0v_1 + 0v_2 + \cdots + 0v_n = 0$.

Given a vector $\vec{v}$, there are many ways to produce a non-trivial vector that is orthogonal to $\vec{v}$.

Example 1.5.21 A vector orthogonal to $\begin{bmatrix} a \\ b \end{bmatrix}$ in $\mathbb{R}^2$ is $\begin{bmatrix} -b \\ a \end{bmatrix}$, since $\begin{bmatrix} -b \\ a \end{bmatrix} \cdot \begin{bmatrix} a \\ b \end{bmatrix} = 0$.

Example 1.5.22 A vector orthogonal to $\begin{bmatrix} a \\ b \\ c \end{bmatrix}$ in $\mathbb{R}^3$ is $\begin{bmatrix} -b \\ a \\ 0 \end{bmatrix}$, since $\begin{bmatrix} -b \\ a \\ 0 \end{bmatrix} \cdot \begin{bmatrix} a \\ b \\ c \end{bmatrix} = 0$.

EXERCISE

Produce two different vectors that are orthogonal to $\begin{bmatrix} a \\ b \\ c \end{bmatrix}$.

Example 1.5.23 Determine a vector $\vec{u}$ that is orthogonal to both $\vec{v} = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}$ and $\vec{w} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$.

Solution: Let $\vec{u} = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix}$ be such a vector. Then

$$\begin{aligned} &\vec{u} \cdot \vec{v} = 0 \\ &\begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} = 0 \\ &u_1 - u_2 = 0 \\ &u_1 = u_2 \ (*) \end{aligned} \qquad \text{and} \qquad \begin{aligned} &\vec{u} \cdot \vec{w} = 0 \\ &\begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = 0 \\ &u_1 + 2u_2 + 3u_3 = 0 \\ &3u_1 + 3u_3 = 0 \quad \text{substitute from } (*) \\ &u_3 = -u_1 \ (**) \end{aligned}$$

From $(*)$ and $(**)$ we have $\vec{u} = \begin{bmatrix} u_1 \\ u_1 \\ -u_1 \end{bmatrix}$. For any choice of $u_1 \in \mathbb{R}$, the vector $\begin{bmatrix} u_1 \\ u_1 \\ -u_1 \end{bmatrix}$ will be orthogonal to both $\vec{v}$ and $\vec{w}$. This is because by choosing different values of u_1, we change the length of $\vec{u}$, but the line through the origin which contains $\vec{u}$ remains fixed.

For example, with $u_1 = 4$ we get $\vec{u} = \begin{bmatrix} 4 \\ 4 \\ -4 \end{bmatrix}$, in which case $\vec{u} \cdot \vec{v} = \begin{bmatrix} 4 \\ 4 \\ -4 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} = 4 - 4 = 0$ and $\vec{u} \cdot \vec{w} = \begin{bmatrix} 4 \\ 4 \\ -4 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = 4 + 8 - 12 = 0$.

EXERCISE

With $\vec{u} = \begin{bmatrix} u_1 \\ u_1 \\ -u_1 \end{bmatrix}$ from Example 1.5.23, choose different values of $u_1 \in \mathbb{R}$ and convince yourself that the resulting $\vec{u}$ is still orthogonal to both $\vec{v} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$ and $\vec{w} = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}$.

Show that for any choice of $u_1 \in \mathbb{R}$, the resulting $\vec{u}$ is still orthogonal to both $\vec{v}$ and $\vec{w}$.

1.6 Projection, Components and Perpendicular

Definition 1.6.1

Projection of $\vec{v}$ Onto $\vec{w}$

Let $\vec{v}, \vec{w} \in \mathbb{R}^n$ with $\vec{w} \neq \vec{0}$. The **projection of $\vec{v}$ onto $\vec{w}$** is defined by

$$\operatorname{proj}_{\vec{w}}(\vec{v}) = \frac{(\vec{v} \cdot \vec{w})}{\|\vec{w}\|^2}\,\vec{w} = \frac{(\vec{v} \cdot \vec{w})}{\vec{w} \cdot \vec{w}}\,\vec{w}.$$

We also refer to this as the **projection of $\vec{v}$ in the $\vec{w}$ direction**.

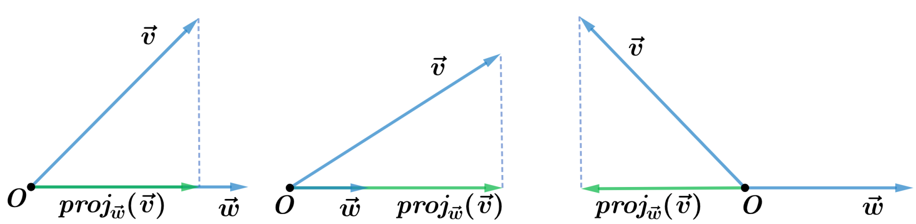


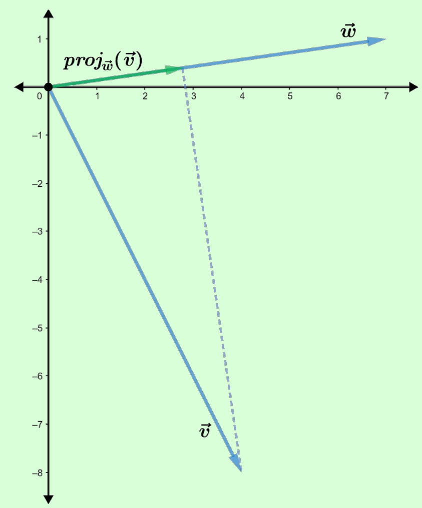
Figure 1.6.7: $\operatorname{proj}_{\vec{w}}(\vec{v})$

Example 1.6.2

Find the projection of $\vec{v} = \begin{bmatrix} 4 \\ -8 \end{bmatrix}$ onto $\vec{w} = \begin{bmatrix} 7 \\ 1 \end{bmatrix}$.

Solution:

$$\begin{aligned}
\text{proj}_{\vec{w}}(\vec{v}) &= \frac{(\vec{v} \cdot \vec{w})}{\|\vec{w}\|^2}\,\vec{w} \\
&= \frac{\begin{bmatrix} 4 \\ -8 \end{bmatrix} \cdot \begin{bmatrix} 7 \\ 1 \end{bmatrix}}{\left\| \begin{bmatrix} 7 \\ 1 \end{bmatrix} \right\|^2} \begin{bmatrix} 7 \\ 1 \end{bmatrix} \\
&= \frac{20}{50} \begin{bmatrix} 7 \\ 1 \end{bmatrix} \\
&= \frac{2}{5}\vec{w}.
\end{aligned}$$



Example 1.6.3

Find the projection of $\vec{v} = \begin{bmatrix} 3 \\ -4 \\ 5 \end{bmatrix}$ onto $\vec{w} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$.

Solution:

$$\text{proj}_{\vec{w}}(\vec{v}) = \frac{(\vec{v} \cdot \vec{w})}{\|\vec{w}\|^2}\,\vec{w} = \frac{\begin{bmatrix} 3 \\ -4 \\ 5 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}}{\left\| \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \right\|^2} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \frac{10}{14} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \frac{5}{7} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}.$$

REMARKS

1. We can write the projection as follows:

$$\text{proj}_{\vec{w}}(\vec{v}) = \frac{(\vec{v} \cdot \vec{w})}{\|\vec{w}\|^2}\,\vec{w} = \left(\vec{v} \cdot \frac{\vec{w}}{\|\vec{w}\|} \right) \frac{\vec{w}}{\|\vec{w}\|} = (\vec{v} \cdot \hat{w})\,\hat{w}.$$

   Thus, *the magnitude of $\vec{w}$ is not relevant to the projection, but the direction of $\vec{w}$ is.*

2. Recall from Definition 1.5.15 that the angle θ between $\vec{v}$ and $\vec{w}$ is given by $\vec{v} \cdot \vec{w} = \|\vec{v}\|\|\vec{w}\| \cos\theta$, giving:

$$\text{proj}_{\vec{w}}(\vec{v}) = \frac{\|\vec{v}\|\|\vec{w}\| \cos\theta}{\|\vec{w}\|^2}\vec{w} = (\|\vec{v}\| \cos\theta)\hat{w}.$$

   You can think of the projection of $\vec{v}$ onto $\vec{w}$ as giving that part of the vector $\vec{v}$ that lies in the $\vec{w}$ direction.

Definition 1.6.4
Component

Let $\vec{v}, \vec{w} \in \mathbb{R}^n$ with $\vec{w} \neq \vec{0}$. We refer to the quantity

$$\|\vec{v}\| \cos\theta = \vec{v} \cdot \hat{w}$$

as the **component** (or **scalar component**) **of** $\vec{v}$ **along** $\vec{w}$.

If the scalar component is negative, then the projection of $\vec{v}$ along $\vec{w}$ will be in the direction of the vector $-\vec{w}$; we can still think of this projection vector as being in the $\vec{w}$ direction.

Example 1.6.5

Determine the component of $\vec{v} = \begin{bmatrix} 3 \\ -4 \\ 5 \end{bmatrix}$ along $\vec{w} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$.

Solution: Note that $\|\vec{w}\| = \sqrt{14}$, so $\hat{w} = \dfrac{1}{\sqrt{14}} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$.

The component of $\vec{v}$ along $\vec{w}$ is $\vec{v} \cdot \hat{w} = \begin{bmatrix} 3 \\ -4 \\ 5 \end{bmatrix} \cdot \dfrac{1}{\sqrt{14}} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \dfrac{10}{\sqrt{14}}$.

Definition 1.6.6
Perpendicular

Let $\vec{v}, \vec{w} \in \mathbb{R}^n$ with $\vec{w} \neq \vec{0}$. The **perpendicular** of $\vec{v}$ onto $\vec{w}$ is defined by

$$\text{perp}_{\vec{w}}(\vec{v}) = \vec{v} - \text{proj}_{\vec{w}}(\vec{v}).$$

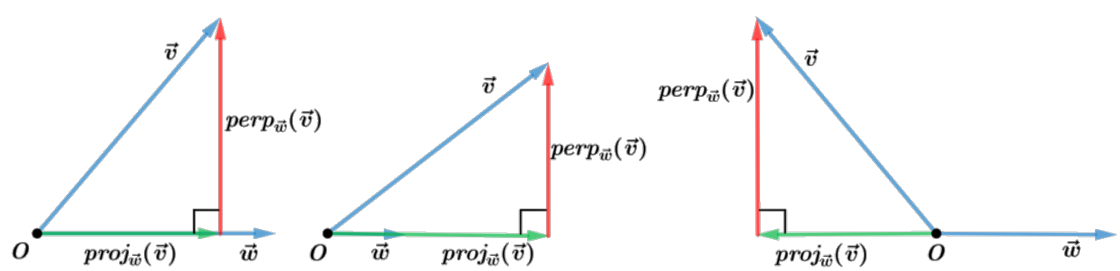


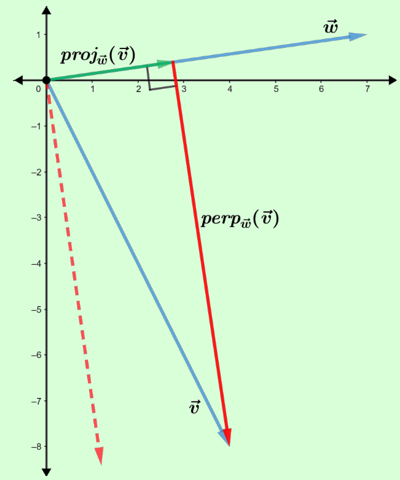
Figure 1.6.8: $\text{perp}_{\vec{w}}(\vec{v})$

Example 1.6.7

Determine the perpendicular of $\vec{v} = \begin{bmatrix} 4 \\ -8 \end{bmatrix}$ onto $\vec{w} = \begin{bmatrix} 7 \\ 1 \end{bmatrix}$.
Then calculate $\text{proj}_{\vec{w}}(\vec{v}) \cdot \text{perp}_{\vec{w}}(\vec{v})$.
Solution:

From [Example 1.6.2], we have $\text{proj}_{\vec{w}}(\vec{v}) = \frac{2}{5}\begin{bmatrix} 7 \\ 1 \end{bmatrix}$, so

$$\begin{aligned} \text{perp}_{\vec{w}}(\vec{v}) &= \vec{v} - \text{proj}_{\vec{w}}(\vec{v}) \\ &= \begin{bmatrix} 4 \\ -8 \end{bmatrix} - \frac{2}{5}\begin{bmatrix} 7 \\ 1 \end{bmatrix} \\ &= \frac{6}{5}\begin{bmatrix} 1 \\ -7 \end{bmatrix} \\ \text{proj}_{\vec{w}}(\vec{v}) \cdot \text{perp}_{\vec{w}}(\vec{v}) &= \left(\frac{2}{5}\begin{bmatrix} 7 \\ 1 \end{bmatrix}\right) \cdot \left(\frac{6}{5}\begin{bmatrix} 1 \\ -7 \end{bmatrix}\right) \\ &= \frac{12}{25}(7(1) + 1(-7)) \\ &= 0. \end{aligned}$$



Example 1.6.8 Determine the perpendicular of $\vec{v} = \begin{bmatrix} 3 \\ -4 \\ 5 \end{bmatrix}$ onto $\vec{w} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$.

Solution: From Example 1.6.3, we have $\text{proj}_{\vec{w}}(\vec{v}) = \frac{5}{7}\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$, so

$$\text{perp}_{\vec{w}}(\vec{v}) = \vec{v} - \text{proj}_{\vec{w}}(\vec{v}) = \begin{bmatrix} 3 \\ -4 \\ 5 \end{bmatrix} - \frac{5}{7}\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \frac{1}{7}\begin{bmatrix} 16 \\ -38 \\ 20 \end{bmatrix}.$$

Proposition 1.6.9 The projection and the perpendicular of a vector $\vec{v}$ onto $\vec{w} \neq \vec{0}$ are orthogonal; that is

$$\text{perp}_{\vec{w}}(\vec{v}) \cdot \text{proj}_{\vec{w}}(\vec{v}) = 0.$$

Proof: First, $\text{perp}_{\vec{w}}(\vec{v})$ is orthogonal to $\vec{w}$, since

$$\begin{aligned} \text{perp}_{\vec{w}}(\vec{v}) \cdot \vec{w} &= (\vec{v} - \text{proj}_{\vec{w}}(\vec{v})) \cdot \vec{w} \\ &= \vec{v} \cdot \vec{w} - (\text{proj}_{\vec{w}}(\vec{v})) \cdot \vec{w} \\ &= \vec{v} \cdot \vec{w} - \left(\frac{(\vec{v} \cdot \vec{w})}{\|\vec{w}\|^2}\vec{w}\right) \cdot \vec{w} \\ &= \vec{v} \cdot \vec{w} - \left(\frac{(\vec{v} \cdot \vec{w})}{\|\vec{w}\|^2}\right)\|\vec{w}\|^2 \\ &= \vec{v} \cdot \vec{w} - \vec{v} \cdot \vec{w} \\ &= 0. \end{aligned}$$

Since $\text{perp}_{\vec{w}}(\vec{v})$ is orthogonal to $\vec{w}$, it is also orthogonal to $\text{proj}_{\vec{w}}(\vec{v})$. $\square$

1.7 Standard Inner Product in $\mathbb{C}^n$

Recall the different ways used to find the magnitude of a real number and the magnitude of a complex number.

$$\text{If } x \in \mathbb{R}, \text{ then the magnitude (absolute value) of } x \text{ is } |x| = \sqrt{x(x)}.$$

$$\text{If } z \in \mathbb{C}, \text{ then the magnitude (modulus) of } z \text{ is } |z| = \sqrt{z(\bar{z})}.$$

Example 1.7.1

Determine the magnitude of $z = 2 - 3i$.

Solution: The magnitude of z is

$$|z| = |2 - 3i| = \sqrt{(2-3i)(\overline{2-3i})} = \sqrt{2^2 + (-3)^2} = \sqrt{13}.$$

One of the most important uses of the dot product on $\mathbb{R}^n$ is in finding the lengths of vectors. When we extend this concept to $\mathbb{C}^n$, we will need to use complex conjugates.

Definition 1.7.2

Standard Inner Product on $\mathbb{C}^n$

The **standard inner product** of $\vec{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}, \vec{w} = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix} \in \mathbb{C}^n$ is

$$\langle \vec{v}, \vec{w} \rangle = v_1\overline{w_1} + v_2\overline{w_2} + \cdots + v_n\overline{w_n}.$$

Notice that this operation is similar to the dot product, in that we multiply components together and add them up. An important difference is the fact that we take the complex conjugate of all of the components that come from the second vector $\vec{w}$.

Example 1.7.3

Let $\vec{v} = \begin{bmatrix} 1+i \\ 1+2i \end{bmatrix}, \vec{w} = \begin{bmatrix} 2+i \\ 1+i \end{bmatrix}$. Evaluate $\langle \vec{v}, \vec{w} \rangle$.

Solution:

$$\begin{aligned} \langle \vec{v}, \vec{w} \rangle &= (1+i)(\overline{2+i}) + (1+2i)(\overline{1+i}) \\ &= (2 - i + 2i + 1) + (1 - i + 2i + 2) \\ &= (3+i) + (3+i) \\ &= 6 + 2i. \end{aligned}$$

Example 1.7.4

Evaluate $\left\langle \begin{bmatrix} 2-3i \\ 1-2i \end{bmatrix}, \begin{bmatrix} -3+4i \\ 3+5i \end{bmatrix} \right\rangle$.

Solution:

$$\left\langle \begin{bmatrix} 2-3i \\ 1-2i \end{bmatrix}, \begin{bmatrix} -3+4i \\ 3+5i \end{bmatrix} \right\rangle = (2-3i)(\overline{-3+4i}) + (1-2i)(\overline{3+5i})$$
$$= (2-3i)(-3-4i) + (1-2i)(3-5i)$$
$$= (-6-8i+9i-12) + (3-5i-6i-10)$$
$$= -25-10i.$$

Proposition 1.7.5 **(Properties of the Standard Inner Product)**

If $c \in \mathbb{C}$ and $\vec{u}$, $\vec{v}$ and $\vec{w}$ are vectors in $\mathbb{C}^n$, then

(a) $\langle \vec{u}, \vec{v} \rangle = \overline{\langle \vec{v}, \vec{u} \rangle}$ **(conjugate symmetry)**

$$\left.\begin{array}{l} \text{(b) } \langle \vec{u}+\vec{v}, \vec{w} \rangle = \langle \vec{u}, \vec{w} \rangle + \langle \vec{v}, \vec{w} \rangle \\ \text{(c) } \langle c\vec{u}, \vec{v} \rangle = c\,\langle \vec{u}, \vec{v} \rangle \end{array}\right\} \quad \left(\begin{array}{l} \text{collectively known as} \\ \textbf{linearity in the first argument}. \end{array} \right)$$

(d) $\langle \vec{v}, \vec{v} \rangle \geq 0$, with $\langle \vec{v}, \vec{v} \rangle = 0$ if and only if $\vec{v} = \vec{0}$. **(non-negativity)**.

As in $\mathbb{R}^n$, we can use Property (d) to calculate lengths of vectors in $\mathbb{C}^n$.

Definition 1.7.6
Length in $\mathbb{C}^n$, Norm in $\mathbb{C}^n$

The **length** (or **norm** or **magnitude**) of the vector $\vec{v} \in \mathbb{C}^n$ is $\|\vec{v}\| = \sqrt{\langle \vec{v}, \vec{v} \rangle}$.

Example 1.7.7

Determine the length of the vector $\vec{v} = \begin{bmatrix} 2-i \\ -3+2i \\ -4-5i \end{bmatrix}$.

Solution:

$$\begin{aligned} \langle \vec{v}, \vec{v} \rangle &= \left\langle \begin{bmatrix} 2-i \\ -3+2i \\ -4-5i \end{bmatrix}, \begin{bmatrix} 2-i \\ -3+2i \\ -4-5i \end{bmatrix} \right\rangle \\ &= (2-i)(\overline{2-i}) + (-3+2i)(\overline{-3+2i}) + (-4-5i)(\overline{-4-5i}) \\ &= 2^2+1^2+3^2+2^2+4^2+5^2 \\ &= 59 \\ \|\vec{v}\| &= \sqrt{\langle \vec{v}, \vec{v} \rangle} = \sqrt{59}. \end{aligned}$$

Note that $\langle \vec{v}, \vec{v} \rangle = \|\vec{v}\|^2 = |2-i|^2 + |-3+2i|^2 + |-4-5i|^2$.

Proposition 1.7.8 **(Properties of the Length)**

Let $c \in \mathbb{C}$ and $\vec{v} \in \mathbb{C}^n$. Then

(a) $\|c\vec{v}\| = |c|\,\|\vec{v}\|$

(b) $\|\vec{v}\| \geq 0$, and $\|\vec{v}\| = 0$ if and only if $\vec{v} = 0$.

Example 1.7.9

Determine $\left\| (-1+2i) \begin{bmatrix} 2-i \\ -3+2i \\ -4-5i \end{bmatrix} \right\|$.

Solution: Using Property (a) in the above Proposition and Example 1.7.7,

$$\left\| (-1+2i) \begin{bmatrix} 2-i \\ -3+2i \\ -4-5i \end{bmatrix} \right\| = |-1+2i| \left\| \begin{bmatrix} 2-i \\ -3+2i \\ -4-5i \end{bmatrix} \right\| = \sqrt{5}\sqrt{59} = \sqrt{295}.$$

Definition 1.7.10
Orthogonal in $\mathbb{C}^n$

We say that the two vectors $\vec{u}$ and $\vec{v}$ in $\mathbb{C}^n$ are **orthogonal** if $\langle \vec{u}, \vec{v} \rangle = 0$.

Example 1.7.11

Is $\vec{v} = \begin{bmatrix} 1+i \\ 1+2i \end{bmatrix}$ orthogonal to $\vec{w} = \begin{bmatrix} 2+i \\ -1-i \end{bmatrix}$?

Solution: Let's calculate the standard inner product of $\vec{v}$ and $\vec{w}$:

$$\begin{aligned} \langle \vec{v}, \vec{w} \rangle &= (1+i)(\overline{2+i}) + (1+2i)(\overline{-1-i}) \\ &= 2 - i + 2i + 1 - 1 + i - 2i - 2 \\ &= 0. \end{aligned}$$

Since $\langle \vec{v}, \vec{w} \rangle = 0$, $\vec{v}$ and $\vec{w}$ are orthogonal.

Example 1.7.12

(a) Find a non-zero vector that is orthogonal to $\vec{w} = \begin{bmatrix} 2+i \\ 1+i \end{bmatrix}$.

(b) Find a unit vector that is orthogonal to $\vec{w} = \begin{bmatrix} 2+i \\ 1+i \end{bmatrix}$.

Solution:
(a) Let $\vec{v} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$ be orthogonal to $\vec{w}$. Then

$$\begin{aligned} \langle \vec{v}, \vec{w} \rangle &= 0 \\ v_1(2-i) + v_2(1-i) &= 0 \\ v_2(1-i) &= v_1(i-2) \\ v_2(1-i)(1+i) &= v_1(i-2)(1+i) \qquad \left(\text{multiply by } \overline{1-i}\right) \\ 2v_2 &= v_1(-3-i) \\ v_2 &= \frac{-3-i}{2} v_1. \quad (*) \end{aligned}$$

By Proposition 1.7.5 (Properties of the Standard Inner Product), if $\langle \vec{v}, \vec{w} \rangle = 0$, then $\langle c\vec{v}, \vec{w} \rangle = c \langle \vec{v}, \vec{w} \rangle = c(0) = 0$; that is, if $\vec{v}$ is orthogonal to $\vec{w}$, then any scalar multiple $c\vec{v}$ is also orthogonal to $\vec{w}$.

Accordingly, by choosing any non-zero value for v_1, we can use $(*)$ to determine a corresponding value for v_2, so that the resulting vector $\begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$ is orthogonal to $\vec{w}$.

For example, choosing $v_1 = 2$ gives $v_2 = \dfrac{-3-i}{2}(2) = -3-i$.

$$\begin{aligned} \text{Since } \langle \vec{v}, \vec{w} \rangle &= \left\langle \begin{bmatrix} 2 \\ -3-i \end{bmatrix}, \begin{bmatrix} 2+i \\ 1+i \end{bmatrix} \right\rangle \\ &= 2(\overline{2+i}) + (-3-i)(\overline{1+i}) \\ &= 2(2-i) + (-3-i)(1-i) \\ &= 4 - 2i - 3 - i + 3i - 1 = 0, \end{aligned}$$

we know that $\vec{v} = \begin{bmatrix} 2 \\ -3-i \end{bmatrix}$ is orthogonal to $\vec{w} = \begin{bmatrix} 2+i \\ 1+i \end{bmatrix}$.

(b) A unit vector in the same direction as $\vec{v} = \begin{bmatrix} 2 \\ -3-i \end{bmatrix}$ is $\hat{v} = \dfrac{1}{\|\vec{v}\|}\,\vec{v}$.

$$\begin{aligned} \|\vec{v}\| = \sqrt{\langle \vec{v}, \vec{v} \rangle} &= \sqrt{2(\overline{2}) + (-3-i)(\overline{-3-i})} \\ &= \sqrt{4+9+1} \\ &= \sqrt{14} \end{aligned}$$

Thus, $\hat{v} = \dfrac{1}{\|\vec{v}\|}\,\vec{v} = \dfrac{1}{\sqrt{14}} \begin{bmatrix} 2 \\ -3-i \end{bmatrix}$ is a unit vector that is orthogonal to $\vec{w}$.

EXERCISE

In (a) of the previous Example, choose a different non-zero value for v_1 and use $(*)$ to determine a value for v_2. Verify that your resulting vector $\begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$ is orthogonal to $\vec{w} = \begin{bmatrix} 2+i \\ 1+i \end{bmatrix}$.

Definition 1.7.13

Projection of $\vec{v}$ Onto $\vec{w}$

Let $\vec{v}, \vec{w} \in \mathbb{C}^n$, with $\vec{w} \neq \vec{0}$. The **projection of $\vec{v}$ onto $\vec{w}$** is defined by

$$\operatorname{proj}_{\vec{w}}(\vec{v}) = \frac{\langle \vec{v}, \vec{w} \rangle}{\|\vec{w}\|^2}\,\vec{w} = \langle \vec{v}, \hat{w} \rangle\, \hat{w}.$$

Example 1.7.14

Find the projection of $\vec{v} = \begin{bmatrix} 1 \\ i \\ 1+i \end{bmatrix}$ onto $\vec{w} = \begin{bmatrix} 1-i \\ 2-i \\ 3+i \end{bmatrix}$.

Solution:

$$\begin{aligned}
\langle \vec{v}, \vec{w} \rangle &= 1(\overline{1-i}) + i(\overline{2-i}) + (1+i)(\overline{3+i}) \\
&= 1(1+i) + i(2+i) + (1+i)(3-i) \\
&= 1 + i + 2i - 1 + 3 - i + 3i + 1 \\
&= 4 + 5i \\
\|\vec{w}\|^2 &= |1-i|^2 + |2-i|^2 + |3+i|^2 \\
&= 1 + 1 + 4 + 1 + 9 + 1 \\
&= 17 \\
\text{Thus, } \operatorname{proj}_{\vec{w}}(\vec{v}) &= \frac{\langle \vec{v}, \vec{w} \rangle}{\|\vec{w}\|^2}\,\vec{w} = \frac{4+5i}{17}\begin{bmatrix} 1-i \\ 2-i \\ 3+i \end{bmatrix}.
\end{aligned}$$

1.8 Fields

So far, we have used the sets $\mathbb{R}$ and $\mathbb{C}$ to choose the components of vectors, but they will play other roles as well. These sets are examples of a **field**.

We will not see any other fields in this course beyond $\mathbb{R}$ and $\mathbb{C}$, but for the interested student, we can give some feel for what a field "is".

Loosely speaking, a field is a set where we can add, subtract, multiply and divide. So, the set of integers, $\mathbb{Z}$, is not a field, since although 2 and 3 are both integers, the quotient $2/3$ is not. The set of positive real numbers $\mathbb{R}^+$ is not a field since although 2 and 3 are positive real numbers, the difference $2-3$ is not. It turns out (and this is not supposed to be obvious) that the ability to add, subtract, multiply and divide are exactly the properties we need of our scalars in order to do the type of algebra that we want.

In much of what we do, there is no need to make an explicit distinction between $\mathbb{R}$ and $\mathbb{C}$, and so we will refer to our universal set as the field $\mathbb{F}$, with the understanding that $\mathbb{F}$ will either be $\mathbb{R}$ or $\mathbb{C}$, so we will assume that $\mathbb{F} \subseteq \mathbb{C}$. Most of the time, the choice of $\mathbb{F}$ will not matter. There are however a few occasions where we need to be explicit about which field we are using.

For example, if we attempt to solve $2x = 5$ over $\mathbb{F}$, the solution is $x = \frac{5}{2}$, whether $\mathbb{F}$ is $\mathbb{R}$ or $\mathbb{C}$. However, if we try to solve $x^2 = -1$ over $\mathbb{F}$, then the field matters. If $\mathbb{F} = \mathbb{R}$, then this equation has no solutions, but if $\mathbb{F} = \mathbb{C}$, then there are two solutions: $x = \pm i$.

We previously saw the definition of the standard inner product in $\mathbb{C}^n$. It will be useful to extend this definition to $\mathbb{F}^n$.

Definition 1.8.1

Standard Inner Product on $\mathbb{F}^n$

The **standard inner product** of $\vec{v}, \vec{w} \in \mathbb{F}^n$ is

$$\langle \vec{v}, \vec{w} \rangle = v_1\overline{w_1} + v_2\overline{w_2} + \cdots + v_n\overline{w_n}.$$

REMARK

- If $\mathbb{F} = \mathbb{C}$, then this is exactly the standard inner product on $\mathbb{C}^n$, from Definition 1.7.2.
- If $\mathbb{F} = \mathbb{R}$, since each $w_i \in \mathbb{R}$, then $\overline{w_i} = w_i$, and this is just the dot product on $\mathbb{R}^n$, as in Definition 1.5.1.

1.9 The Cross Product in $\mathbb{R}^3$

Given two vectors $\vec{u}, \vec{v} \in \mathbb{R}^3$, the cross product returns a vector in $\mathbb{R}^3$ that is orthogonal to both $\vec{u}$ and $\vec{v}$.

Definition 1.9.1
Cross Product

Let $\vec{u} = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix}, \vec{v} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} \in \mathbb{R}^3$. The **cross product** of $\vec{u}$ and $\vec{v}$ is defined to be the vector in $\mathbb{R}^3$ given by

$$\vec{u} \times \vec{v} = \begin{bmatrix} u_2v_3 - u_3v_2 \\ -(u_1v_3 - u_3v_1) \\ u_1v_2 - u_2v_1 \end{bmatrix}.$$

Example 1.9.2

We have $\begin{bmatrix} 2 \\ -3 \\ 5 \end{bmatrix} \times \begin{bmatrix} -2 \\ 1 \\ 4 \end{bmatrix} = \begin{bmatrix} (-3)(4) - (5)(1) \\ -((2)(4) - (5)(-2)) \\ (2)(1) - (-3)(-2) \end{bmatrix} = \begin{bmatrix} -17 \\ -18 \\ -4 \end{bmatrix}$.

REMARK

A useful trick for remembering the definition of the cross product will be given in Section 6.6.

Proposition 1.9.3

(Properties of the Cross Product)

Let $\vec{u}, \vec{v} \in \mathbb{R}^3$ and let $\vec{z} = \vec{u} \times \vec{v}$. Then

(a) $\vec{z} \cdot \vec{u} = 0$ and $\vec{z} \cdot \vec{v} = 0$ ($\vec{z}$ is **orthogonal** to both $\vec{u}$ and $\vec{v}$)

(b) $\vec{v} \times \vec{u} = -\vec{z} = -\vec{u} \times \vec{v}$ **(skew-symmetric)**

(c) If $\vec{u} \neq \vec{0}$ and $\vec{v} \neq \vec{0}$, then $\|\vec{u} \times \vec{v}\| = \|\vec{u}\|\|\vec{v}\| \sin\theta$, where θ is the angle between $\vec{u}$ and $\vec{v}$.

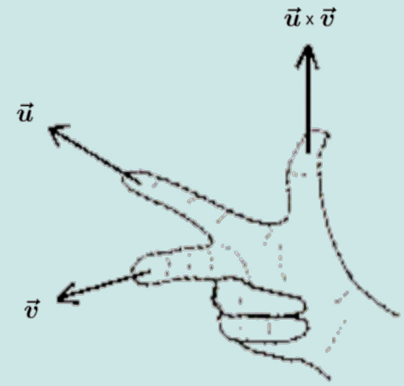
REMARK (Right-Hand Rule)

If $\vec{z}$ is orthogonal to both $\vec{u}$ and $\vec{v}$, then $-\vec{z}$ will also be orthogonal to both $\vec{u}$ and $\vec{v}$, and $-\vec{z}$ has opposite direction to $\vec{z}$.

The **Right-Hand Rule** is a convention used to determine which of these directions $\vec{u} \times \vec{v}$ should have.

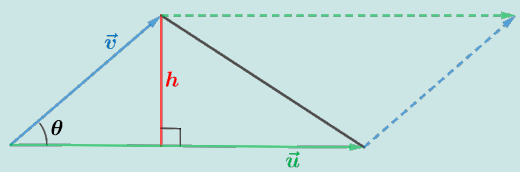
If the pointer finger of your right hand points in the direction of $\vec{u}$, and the middle finger of your right hand points in the direction of $\vec{v}$, then your thumb points in the direction of $\vec{u} \times \vec{v}$.

If you consider $\vec{v} \times \vec{u}$ instead, you would have to turn your hand upside-down, which helps demonstrate the **skew-symmetric** property above!



REMARK (Parallelogram Area via Cross Product)

Consider the parallelogram defined by $\vec{u}$ and $\vec{v}$; its height is $h = \|\vec{v}\| \sin\theta$.



As a result of Property (c) of the above Proposition, its area is

$$\|\vec{u}\|h = \|\vec{u}\|\|\vec{v}\| \sin\theta = \|\vec{u} \times \vec{v}\|.$$

Example 1.9.4 With $\vec{u} = \begin{bmatrix} 2 \\ -3 \\ 5 \end{bmatrix}$ and $\vec{v} = \begin{bmatrix} -2 \\ 1 \\ 4 \end{bmatrix}$, verify Proposition 1.9.3 (Properties of the Cross Product).

Solution: From Example 1.9.2, $\vec{z} = \vec{u} \times \vec{v} = \begin{bmatrix} -17 \\ -18 \\ -4 \end{bmatrix}$.

(a) $\vec{z} \cdot \vec{u} = \begin{bmatrix} -17 \\ -18 \\ -4 \end{bmatrix} \cdot \begin{bmatrix} 2 \\ -3 \\ 5 \end{bmatrix} = -17(2) - 18(-3) - 4(5) = 0.$

$\vec{z} \cdot \vec{v} = \begin{bmatrix} -17 \\ -18 \\ -4 \end{bmatrix} \cdot \begin{bmatrix} -2 \\ 1 \\ 4 \end{bmatrix} = -17(-2) - 18(1) - 4(4) = 0.$

(b) $\vec{v} \times \vec{u} = \begin{bmatrix} -2 \\ 1 \\ 4 \end{bmatrix} \times \begin{bmatrix} 2 \\ -3 \\ 5 \end{bmatrix} = \begin{bmatrix} (1)(5)-(4)(-3) \\ -((-2)(5)-(4)(2)) \\ (-2)(-3)-(1)(2) \end{bmatrix} = \begin{bmatrix} 17 \\ 18 \\ 4 \end{bmatrix} = -\vec{u} \times \vec{v}.$

(c) Let θ be the angle between $\vec{u}$ and $\vec{v}$. Working to three decimal places throughout, $\vec{u} \cdot \vec{v}$ can be used to evaluate θ.

$$\theta = \arccos\left(\frac{\vec{u} \cdot \vec{v}}{\|\vec{u}\|\|\vec{v}\|}\right) = \arccos\left(\frac{\begin{bmatrix} 2 \\ -3 \\ 5 \end{bmatrix} \cdot \begin{bmatrix} -2 \\ 1 \\ 4 \end{bmatrix}}{\left\|\begin{bmatrix} 2 \\ -3 \\ 5 \end{bmatrix}\right\|\left\|\begin{bmatrix} -2 \\ 1 \\ 4 \end{bmatrix}\right\|}\right) = \arccos\left(\frac{13}{\sqrt{798}}\right) \approx 1.093.$$

Thus $\|\vec{u}\|\|\vec{v}\| \sin\theta = \left\|\begin{bmatrix} 2 \\ -3 \\ 5 \end{bmatrix}\right\|\left\|\begin{bmatrix} -2 \\ 1 \\ 4 \end{bmatrix}\right\| \sin\theta \approx \sqrt{798}(0.888) \approx 25.080.$

Since $\vec{u} \times \vec{v} = \begin{bmatrix} -17 \\ -18 \\ -4 \end{bmatrix}$, thus

$\|\vec{u} \times \vec{v}\| = \sqrt{17^2 + 18^2 + 4^2} = \sqrt{629} = 25.080 \approx \|\vec{u}\|\|\vec{v}\| \sin\theta.$

Proposition 1.9.5 **(Linearity of the Cross Product)**

If $c \in \mathbb{R}$ and $\vec{u}$, $\vec{v}, \vec{w} \in \mathbb{R}^3$, then

$$\left.\begin{array}{ll} (a) & (\vec{u}+\vec{v}) \times \vec{w} = (\vec{u} \times \vec{w}) + (\vec{v} \times \vec{w}) \\ (b) & (c\vec{u}) \times \vec{v} = c(\vec{u} \times \vec{v}) \end{array}\right\} \quad \textbf{linearity in the first argument}$$

$$\left.\begin{array}{ll} (c) & \vec{u} \times (\vec{v}+\vec{w}) = (\vec{u} \times \vec{v}) + (\vec{u} \times \vec{w}) \\ (d) & \vec{u} \times (c\vec{v}) = c(\vec{u} \times \vec{v}) \end{array}\right\} \quad \textbf{linearity in the second argument}$$

Proof: We will prove linearity in the second argument; the proof of linearity in the first argument follows by using the skew-symmetry of the cross product.

(c) We have

$$\begin{aligned} \vec{u} \times (\vec{v}+\vec{w}) &= \begin{bmatrix} u_2(v_3+w_3) - u_3(v_2+w_2) \\ -[u_1(v_3+w_3) - u_3(v_1+w_1)] \\ u_1(v_2+w_2) - u_2(v_1+w_1) \end{bmatrix} \\ &= \begin{bmatrix} u_2v_3 - u_3v_2 \\ -[u_1v_3 - u_3v_1] \\ u_1v_2 - u_2v_1 \end{bmatrix} + \begin{bmatrix} u_2w_3 - u_3w_2 \\ -[u_1w_3 - u_3w_1] \\ u_1w_2 - u_2w_1 \end{bmatrix} \\ &= (\vec{u} \times \vec{v}) + (\vec{u} \times \vec{w}). \end{aligned}$$

(d) We have

$$\begin{aligned}\vec{u} \times (c\vec{v}) &= \begin{bmatrix} u_2(cv_3) - u_3(cv_2) \\ -[u_1(cv_3) - u_3(cv_1)] \\ u_1(cv_2) - u_2(cv_1) \end{bmatrix} \\ &= c \begin{bmatrix} u_2v_3 - u_3v_2 \\ -[u_1v_3 - u_3v_1] \\ u_1v_2 - u_2v_1 \end{bmatrix} \\ &= c(\vec{u} \times \vec{v}).\end{aligned}$$

$\square$

REMARK (Why is $\vec{u} \times \vec{v}$ only defined for $\mathbb{R}^3$?)

Given two non-parallel vectors $\vec{u}, \vec{v} \in \mathbb{R}^n$ for $n = 3$, we can notice that there is always exactly one line through the origin that is orthogonal to both $\vec{u}$ and $\vec{v}$. A cross product allows us to compute one non-zero vector on this line. It turns out that any other vector that is orthogonal to both $\vec{u}$ and $\vec{v}$ will always lie on this line, and hence will be a multiple of $\vec{u} \times \vec{v}$. Notice that, when $n \neq 3$, this does not happen.

- In $\mathbb{R}^2$, if $\vec{u} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\vec{v} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$, then the only vector that is orthogonal to both $\vec{u}$ and $\vec{v}$ is $\vec{0}$, which we can consider a trivial choice.
- For vectors $\vec{u}, \vec{v} \in \mathbb{R}^n$ with $n \geq 4$, there are many different lines through the origin that are orthogonal to both $\vec{u}$ and $\vec{v}$, making it difficult to choose one line out of many.

  For example, in $\mathbb{R}^4$, if $\vec{u} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$ and $\vec{v} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}$, then $\vec{y} = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}$ and $\vec{z} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$ are each orthogonal to both $\vec{u}$ and $\vec{v}$, but $\vec{y}$ and $\vec{z}$ lie on different lines through the origin.

Chapter 2

Span, Lines and Planes

2.1 Linear Combinations and Span

Definition 2.1.1 **Linear Combination**

Let $c_1, c_2, \ldots, c_k \in \mathbb{F}$ and let $\vec{v_1}, \vec{v_2}, \ldots \vec{v_k}$ be vectors in $\mathbb{F}^n$. We refer to any vector of the form $c_1\vec{v_1} + c_2\vec{v_2} + \cdots + c_k\vec{v_k}$ as a **linear combination** of $\vec{v_1}, \vec{v_2}, \ldots \vec{v_k}$.

Example 2.1.2

Since $2\begin{bmatrix}2\\0\\5\end{bmatrix} - 5\begin{bmatrix}-1\\1\\3\end{bmatrix} = \begin{bmatrix}9\\-5\\-5\end{bmatrix}$, $\begin{bmatrix}9\\-5\\-5\end{bmatrix}$ is a linear combination of vectors $\begin{bmatrix}2\\0\\5\end{bmatrix}$ and $\begin{bmatrix}-1\\1\\3\end{bmatrix}$ in $\mathbb{R}^3$.

Example 2.1.3

Since $3i\begin{bmatrix}i\\0\end{bmatrix} + 4\begin{bmatrix}1+i\\-4i\end{bmatrix} = \begin{bmatrix}1+4i\\-16i\end{bmatrix}$, $\begin{bmatrix}1+4i\\-16i\end{bmatrix}$ is a linear combination of $\begin{bmatrix}i\\0\end{bmatrix}$ and $\begin{bmatrix}1+i\\-4i\end{bmatrix}$ in $\mathbb{C}^2$.

Example 2.1.4

The vector $\begin{bmatrix}-31\\54\\-10\\23\end{bmatrix} = 2\begin{bmatrix}1\\0\\2\\0\end{bmatrix} - 5\begin{bmatrix}0\\0\\-1\\1\end{bmatrix} + 7\begin{bmatrix}1\\2\\3\\4\end{bmatrix} - 4\begin{bmatrix}10\\-10\\10\\0\end{bmatrix}$ is a linear combination of the four vectors

$$\begin{bmatrix}1 & 0 & 2 & 0\end{bmatrix}^T, \begin{bmatrix}0 & 0 & -1 & 1\end{bmatrix}^T, \begin{bmatrix}1 & 2 & 3 & 4\end{bmatrix}^T, \begin{bmatrix}10 & -10 & 10 & 0\end{bmatrix}^T$$

in $\mathbb{R}^4$.

Example 2.1.5

For any $\vec{v}, \vec{w} \in \mathbb{F}^n$, the zero vector $\vec{0} = 0\vec{v} + 0\vec{w}$ is a linear combination of $\vec{v}$ and $\vec{w}$.

Definition 2.1.6 **Span**

Let $\vec{v_1}, \vec{v_2}, \ldots, \vec{v_k}$ be vectors in $\mathbb{F}^n$. We define the **span** of $\{\vec{v_1}, \vec{v_2}, \ldots, \vec{v_k}\}$ to be the set of all linear combinations of $\vec{v_1}, \vec{v_2}, \ldots, \vec{v_k}$. That is,

$$\mathrm{Span}\{\vec{v_1}, \vec{v_2}, \ldots, \vec{v_k}\} = \{c_1\vec{v_1} + c_2\vec{v_2} + \cdots + c_k\vec{v_k} : c_1, c_2, \ldots, c_k \in \mathbb{F}\}.$$

We refer to $\{\vec{v_1}, \vec{v_2}, \ldots, \vec{v_k}\}$ as a **spanning set** for $\text{Span}\{\vec{v_1}, \vec{v_2}, \ldots, \vec{v_k}\}$. We also say that $\text{Span}\{\vec{v_1}, \vec{v_2}, \ldots, \vec{v_k}\}$ is **spanned by** $\{\vec{v_1}, \vec{v_2}, \ldots, \vec{v_k}\}$.

Example 2.1.7

Consider the vectors $\vec{v_1} = \begin{bmatrix} 2i \\ 3 \\ i \end{bmatrix}, \vec{v_2} = \begin{bmatrix} 2+i \\ -5+2i \\ 6i \end{bmatrix}$ in $\mathbb{C}^3$. Then

$$\text{Span}\left\{ \begin{bmatrix} 2i \\ 3 \\ i \end{bmatrix}, \begin{bmatrix} 2+i \\ -5+2i \\ 6i \end{bmatrix} \right\} = \left\{ c_1 \begin{bmatrix} 2i \\ 3 \\ i \end{bmatrix} + c_2 \begin{bmatrix} 2+i \\ -5+2i \\ 6i \end{bmatrix} : c_1, c_2 \in \mathbb{C} \right\}.$$

Example 2.1.8

Consider the vectors $\vec{v_1} = \begin{bmatrix} 1 \\ 3 \\ 7 \end{bmatrix}, \vec{v_2} = \begin{bmatrix} 2 \\ 5 \\ 7 \end{bmatrix}, \vec{v_3} = \begin{bmatrix} 2 \\ -8 \\ 6 \end{bmatrix}, \vec{v_4} = \begin{bmatrix} 5 \\ 7 \\ -4 \end{bmatrix}$ in $\mathbb{F}^3$. Then

$$\text{Span}\{\vec{v_1}, \vec{v_2}, \vec{v_3}, \vec{v_4}\} = \left\{ c_1 \begin{bmatrix} 1 \\ 3 \\ 7 \end{bmatrix} + c_2 \begin{bmatrix} 2 \\ 5 \\ 7 \end{bmatrix} + c_3 \begin{bmatrix} 2 \\ -8 \\ 6 \end{bmatrix} + c_4 \begin{bmatrix} 5 \\ 7 \\ -4 \end{bmatrix} : c_1, c_2, c_3, c_4 \in \mathbb{F} \right\}.$$

REMARK

In the above example, $\mathbb{F}$ could be either $\mathbb{R}$ or $\mathbb{C}$. If $\mathbb{F} = \mathbb{R}$, then we would choose $c_1, c_2, c_3, c_4 \in \mathbb{R}$. However, if $\mathbb{F} = \mathbb{C}$, then we would choose $c_1, c_2, c_3, c_4 \in \mathbb{C}$, resulting in more vectors being in $\text{Span}\{\vec{v_1}, \vec{v_2}, \vec{v_3}, \vec{v_4}\}$.

Example 2.1.9

Determine whether the vector $\begin{bmatrix} -3 \\ 9 \\ 2 \end{bmatrix}$ is an element of $\text{Span}\left\{ \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} \right\}$.

Solution:

Since $\begin{bmatrix} -3 \\ 9 \\ 2 \end{bmatrix} = 2 \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} - 5 \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}$, we see that $\begin{bmatrix} -3 \\ 9 \\ 2 \end{bmatrix} \in \text{Span}\left\{ \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} \right\}$.

Example 2.1.10

Determine whether $\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$ is in $\text{Span}\left\{ \begin{bmatrix} 2 \\ 0 \\ 5 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 3 \end{bmatrix} \right\}$.

Solution:

Every vector in $\text{Span}\left\{ \begin{bmatrix} 2 \\ 0 \\ 5 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 3 \end{bmatrix} \right\}$ has the form $c_1 \begin{bmatrix} 2 \\ 0 \\ 5 \end{bmatrix} + c_2 \begin{bmatrix} -1 \\ 0 \\ 3 \end{bmatrix} = \begin{bmatrix} 2c_1 - c_2 \\ 0 \\ 5c_1 + 3c_2 \end{bmatrix}$

for some $c_1, c_2 \in \mathbb{F}$. Since the equation $\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 2c_1 - c_2 \\ 0 \\ 5c_1 + 3c_2 \end{bmatrix}$ has no solutions in $c_1, c_2 \in \mathbb{F}$, we conclude that $\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \notin \text{Span}\left\{ \begin{bmatrix} 2 \\ 0 \\ 5 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 3 \end{bmatrix} \right\}$.

2.2 Lines in $\mathbb{R}^2$

There is a nice way to write down the equation of a line in $\mathbb{R}^n$ making use of vectors. We begin with $\mathbb{R}^2$ (otherwise known as the xy-plane) before generalizing to $\mathbb{R}^n$.

We consider the line $\mathcal{L}$ with points (x_1, y_1) and (x_2, y_2), y-intercept at $(0, b)$, with slope $m = \frac{p}{q}$ where $q \neq 0$. We visualize all of this information below.

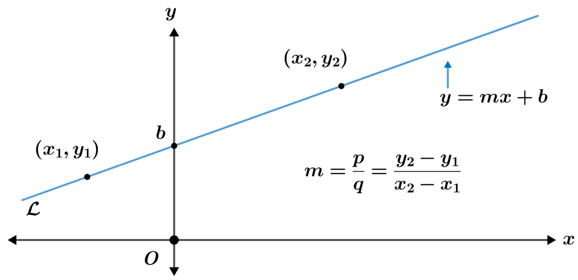


Figure 2.2.1: Characteristics of a line $\mathcal{L}$ in $\mathbb{R}^2$

The table below includes four different forms of equation(s) for $\mathcal{L}$ in the xy-plane, based on given information.

Given Information	**Equation of Line**
Slope m and y-intercept b	$y = mx + b$
Two points (x_1, y_1) and (x_2, y_2)	$\frac{y - y_1}{y_2 - y_1} = \frac{x - x_1}{x_2 - x_1}$
Point (x_1, y_1) and slope m	$y - y_1 = m(x - x_1)$
Point (x_1, y_1) and slope $\frac{p}{q}$ $(q \neq 0)$	$\begin{matrix} x = x_1 + qt \\ y = y_1 + pt \end{matrix}, \quad t \in \mathbb{R}$

In the first 3 forms above, we input a particular value for x to obtain the value of y for the corresponding point on the line.

In the last form above, the variable t is called a **parameter**. We input a particular value for t to obtain the value of the x- and y-coordinates for the corresponding point on the line.

Definition 2.2.1

Parametric Equations of a Line in $\mathbb{R}^2$

Let p, q be fixed real numbers and $q \neq 0$. Then **parametric equations of a line in $\mathbb{R}^2$ through the point (x_1, y_1) with slope $\frac{p}{q}$** are

$$\begin{matrix} x = x_1 + qt \\ y = y_1 + pt \end{matrix}, \quad t \in \mathbb{R}.$$

Each value of t gives a different point on the line. For instance,

- $t = 0$ gives the point $(x, y) = (x_1, y_1)$
- $t = 2$ gives the point $(x, y) = (x_1 + 2q, y_1 + 2p)$
- $t = -5$ gives the point $(x, y) = (x_1 - 5q, y_1 - 5p)$.

As t varies over all real numbers, we generate all the points on the line.

REMARK

Since $q \neq 0$, the expression $\frac{p}{q}$ is always defined. Putting $q = 0$ into the parametric equations of the line gives us $x = x_1$ and $y = y_1 + pt$, a vertical line with undefined slope (we can think of this informally as "infinite slope").

Definition 2.2.2

Vector Equation of a Line in $\mathbb{R}^2$

Let $\begin{bmatrix} q \\ p \end{bmatrix}$ be a non-zero vector in $\mathbb{R}^2$. The expression

$$\vec{\ell} = \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} + t \begin{bmatrix} q \\ p \end{bmatrix}, \ t \in \mathbb{R}$$

is a **vector equation of the line $\mathcal{L}$ in $\mathbb{R}^2$ through $\begin{bmatrix} x_1 \\ y_1 \end{bmatrix}$ with direction $\begin{bmatrix} q \\ p \end{bmatrix}$**.

For any value of $t \in \mathbb{R}$, the expression $\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} + t \begin{bmatrix} q \\ p \end{bmatrix}$ produces a vector in $\mathbb{R}^2$ whose terminal point has coordinates $L = (x_1 + tq, y_1 + tp)$ and is on $\mathcal{L}$.

REMARKS

If we let $\vec{u} = \begin{bmatrix} x_1 \\ y_1 \end{bmatrix}$ and $\vec{v} = \begin{bmatrix} q \\ p \end{bmatrix}$, we can write a vector equation of line $\mathcal{L}$ in $\mathbb{R}^2$ as

$$\vec{\ell} = \vec{u} + t\vec{v}, \ \text{for } t \in \mathbb{R}.$$

- Letting $t = 0$ gives us that the vector $\vec{u}$ is on the line.
- The line passes through the terminal point U associated with $\vec{u}$. The other points on the line move from U in the $\vec{v}$ direction by scalar multiples of $\vec{v}$.
- We say that $\vec{v}$ is **parallel** to the line and that $\vec{v}$ is a **direction vector** to the line.
- The vector $\vec{v}$ is parallel to the line. However, the terminal point V associated with $\vec{v}$ is not usually a point on the line; in fact, V is a point on the line if and only if the vector $\vec{v}$ is a scalar multiple of the vector $\vec{u}$.

Definition 2.2.3
Line in $\mathbb{R}^2$

Let $\vec{u}, \vec{v} \in \mathbb{R}^2$ with $\vec{v} \neq \vec{0}$. We refer to the set of vectors

$$\mathcal{L} = \{\vec{u} + t\vec{v} : t \in \mathbb{R}\}$$

as a **line $\mathcal{L}$ in $\mathbb{R}^2$ through $\vec{u}$ with direction $\vec{v}$.**

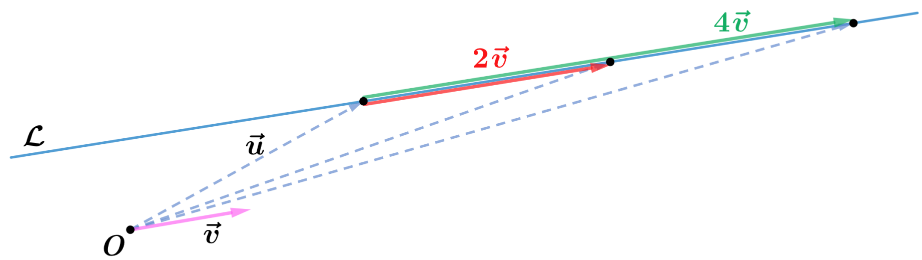


Figure 2.2.2: Line in $\mathbb{R}^2$ through $\vec{u}$ with direction $\vec{v}$

Example 2.2.4

Let $U = (2,3)$ and $V = (1,4)$ be two points on a line $\mathcal{L}$ in $\mathbb{R}^2$. Find a vector equation of $\mathcal{L}$.

Solution:
We can find a vector parallel to the line by taking the difference $\vec{v} - \vec{u}$, where $\vec{u}$ and $\vec{v}$ are the vector representations of points U and V, respectively. This gives $\begin{bmatrix} 1 \\ 4 \end{bmatrix} - \begin{bmatrix} 2 \\ 3 \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$, and a vector equation of the line is therefore $\vec{\ell} = \vec{u} + t(\vec{v} - \vec{u}) = \begin{bmatrix} 2 \\ 3 \end{bmatrix} + t \begin{bmatrix} -1 \\ 1 \end{bmatrix}$ for $t \in \mathbb{R}$.

2.3 Lines in $\mathbb{R}^n$

We now consider lines in $\mathbb{R}^n$. Lines in $\mathbb{R}^n$ can be described using the same vector equations as lines in $\mathbb{R}^2$. The only major differences in describing these lines arise from the implicit number of components in the vectors, which in turn affects the number of parametric equations associated with these lines.

Definition 2.3.1
Vector Equation of a Line in $\mathbb{R}^n$

Let $\vec{u}, \vec{v} \in \mathbb{R}^n$ with $\vec{v} \neq \vec{0}$. The expression

$$\vec{\ell} = \vec{u} + t\vec{v}, \quad t \in \mathbb{R}$$

is a **vector equation of the line $\mathcal{L}$ in $\mathbb{R}^n$ through $\vec{u}$ with direction $\vec{v}$.**

If $\vec{\ell}_1$ and $\vec{\ell}_2$ are lines with direction $\vec{v}_1$ and $\vec{v}_2$, respectively, we say that they have the **same direction** if $c\vec{v}_1 = \vec{v}_2$ for some non-zero $c \in \mathbb{R}$.

For any value of $t \in \mathbb{R}$, the expression $\begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} + t \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$ produces a vector in $\mathbb{R}^n$ whose terminal point has coordinates $L = (u_1 + tv_1, u_2 + tv_2, ..., u_n + tv_n)$ and is on line $\mathcal{L}$.

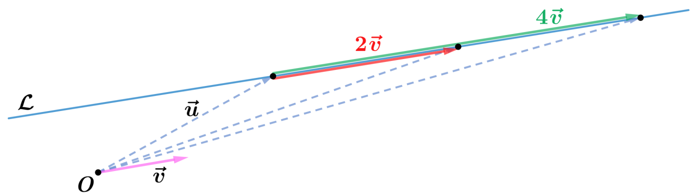


Figure 2.3.3: In $\mathbb{R}^n$, the line $\vec{\ell} = \vec{u} + t\vec{v}, \quad t \in \mathbb{R}$.

REMARKS

- Letting $t = 0$ gives us that the vector $\vec{u}$ is on the line.
- The line passes through the terminal point U associated with $\vec{u}$. The other points on the line move from U in the $\vec{v}$ direction by scalar multiples of $\vec{v}$.
- We say that $\vec{v}$ is **parallel** to the line, and that $\vec{v}$ is a **direction vector** to the line.
- The vector $\vec{v}$ is parallel to the line, however the terminal point V associated with the vector $\vec{v}$, is not usually a point on the line. In fact, V is a point on the line if and only if the vector $\vec{v}$ is a multiple of the vector $\vec{u}$.
- There are many different vector equations that could be used to describe the same line $\mathcal{L}$. We'll see this in Example 2.3.5

Definition 2.3.2

Parametric Equations of a Line in $\mathbb{R}^n$

Let $\vec{u}, \vec{v} \in \mathbb{R}^n$ with $\vec{v} \neq \vec{0}$. Consider the vector equation of the line $\mathcal{L}$ in $\mathbb{R}^n$ given by

$$\vec{\ell} = \vec{u} + t\vec{v}, \quad t \in \mathbb{R}.$$

The **parametric equations of the line $\mathcal{L}$ in $\mathbb{R}^n$ through $\vec{u}$ with direction $\vec{v}$** are

$$\begin{array}{c} \ell_1 = u_1 + tv_1 \\ \ell_2 = u_2 + tv_2 \\ \vdots \\ \ell_n = u_n + tv_n \end{array}, \quad t \in \mathbb{R}.$$

Example 2.3.3

Give a vector equation and a set of parametric equations of the line through $U = (4, -3, 5)$ and in the direction of the vector $\vec{v} = \begin{bmatrix} -2 \\ 4 \\ 1 \end{bmatrix}$.

Solution: A vector equation of the line is

$$\vec{\ell} = \begin{bmatrix} 4 \\ -3 \\ 5 \end{bmatrix} + t \begin{bmatrix} -2 \\ 4 \\ 1 \end{bmatrix}, \; t \in \mathbb{R}.$$

The corresponding parametric equations of the line are:

$$\begin{aligned} \ell_1 &= 4 - 2t \\ \ell_2 &= -3 + 4t \;, \quad t \in \mathbb{R}. \\ \ell_3 &= 5 + t \end{aligned}$$

Definition 2.3.4

Line in $\mathbb{R}^n$

Let $\vec{u}, \vec{v} \in \mathbb{R}^n$ with $\vec{v} \neq \vec{0}$. We refer to the set of vectors

$$\mathcal{L} = \{\vec{u} + t\vec{v} : t \in \mathbb{R}\}$$

as a **line $\mathcal{L}$ in $\mathbb{R}^n$ through $\vec{u}$ with direction $\vec{v}$**.

Example 2.3.5

Let $\mathcal{L} = \left\{ \begin{bmatrix} 4 \\ -3 \\ 5 \end{bmatrix} + t \begin{bmatrix} -2 \\ 4 \\ 1 \end{bmatrix} : t \in \mathbb{R} \right\}$, be the line from Example 2.3.3.

Show that the line $\mathcal{M} = \left\{ \begin{bmatrix} 0 \\ 5 \\ 7 \end{bmatrix} + s \begin{bmatrix} 6 \\ -12 \\ -3 \end{bmatrix} : s \in \mathbb{R} \right\}$ is identical to $\mathcal{L}$.

Solution: First, $-3 \begin{bmatrix} -2 \\ 4 \\ 1 \end{bmatrix} = \begin{bmatrix} 6 \\ -12 \\ -3 \end{bmatrix}$, so $\begin{bmatrix} -2 \\ 4 \\ 1 \end{bmatrix}$ has the same direction as $\begin{bmatrix} 6 \\ -12 \\ -3 \end{bmatrix}$. Thus, $\mathcal{L}$ and $\mathcal{M}$ are parallel. Two parallel lines are the same if and only if they share a common point.

Taking $s = 0$ shows that $\begin{bmatrix} 0 \\ 5 \\ 7 \end{bmatrix}$ is on $\mathcal{M}$. Taking $t = 2$ shows that $\begin{bmatrix} 4 \\ -3 \\ 5 \end{bmatrix} + 2 \begin{bmatrix} -2 \\ 4 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 5 \\ 7 \end{bmatrix}$ is also on $\mathcal{L}$. Thus, lines $\mathcal{L}$ and $\mathcal{M}$ are identical.

EXERCISE

Determine two different vector equations for the line $\vec{\ell} = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix} + t \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, t \in \mathbb{R}$.

Suppose that we are given two distinct points, U and Q, on a line in $\mathbb{R}^n$, with associated vectors, $\vec{u}$ and $\vec{q}$, respectively. We can define $\vec{v} = \vec{q} - \vec{u}$, and this vector gives the direction of the line. Thus, we can still use Definition 2.3.1 for the vector equation of the line:

$$\vec{\ell} = \vec{u} + t(\vec{q} - \vec{u}), \quad t \in \mathbb{R}.$$

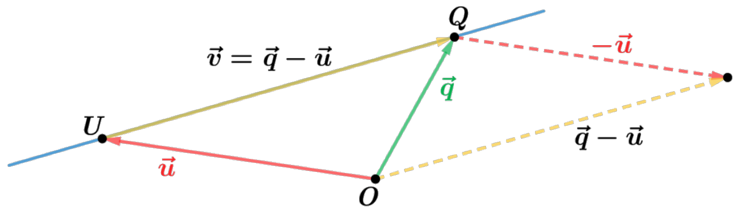


Figure 2.3.4: Line in $\mathbb{R}^n$ with direction $\vec{v} = \vec{q} - \vec{u}$

Example 2.3.6

Determine a vector equation of the line through $U = (2, -3, 5)$ and $Q = (4, -2, 6)$.

Solution: A vector equation of the line through the given points is

$$\vec{\ell} = \begin{bmatrix} 2 \\ -3 \\ 5 \end{bmatrix} + t\left(\begin{bmatrix} 4 \\ -2 \\ 6 \end{bmatrix} - \begin{bmatrix} 2 \\ -3 \\ 5 \end{bmatrix}\right) = \begin{bmatrix} 2 \\ -3 \\ 5 \end{bmatrix} + t\begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix}, t \in \mathbb{R}.$$

We can describe lines through the origin as the span of a single vector. If $\vec{v} \in \mathbb{R}^n$, with $\vec{v} \neq \vec{0}$, then $\text{Span}\{\vec{v}\} = \{\vec{0} + t\vec{v} : t \in \mathbb{R}\}$ is the line through the origin O with $\vec{v}$ as a direction vector.

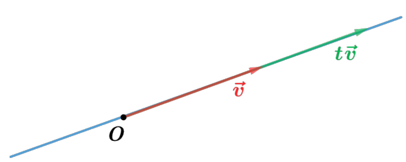


Figure 2.3.5: In $\mathbb{R}^n$, $\text{Span}\{\vec{v}\} = \{\vec{0} + t\vec{v} : t \in \mathbb{R}\}$ is a line through the origin

REMARK

If a line can be expressed as the span of one vector, then the line must pass through the origin.

Example 2.3.7

The set $\text{Span}\left\{\begin{bmatrix} 2 \\ 4 \\ 6 \\ 8 \end{bmatrix}\right\}$ is a line $\mathcal{L}$ in $\mathbb{R}^4$ that goes through the origin and points in the direction of $\begin{bmatrix} 2 \\ 4 \\ 6 \\ 8 \end{bmatrix}$. It is shorthand for $\mathcal{L} = \left\{ t\begin{bmatrix} 2 \\ 4 \\ 6 \\ 8 \end{bmatrix} : t \in \mathbb{R} \right\}$.

2.4 The Vector Equation of a Plane in $\mathbb{R}^n$

In this section we will be working in $\mathbb{R}^n$ where $n \geq 2$.

Definition 2.4.1

Plane in $\mathbb{R}^n$ Through the Origin

Let $\vec{v}, \vec{w}$ be non-zero vectors in $\mathbb{R}^n$ with $\vec{v} \neq c\vec{w}$ for any $c \in \mathbb{R}$. Then

$$\mathcal{P} = \mathrm{Span}\{\vec{v}, \vec{w}\} = \{s\vec{v} + t\vec{w} : s, t \in \mathbb{R}\}$$

is a **plane in $\mathbb{R}^n$ through the origin with direction vectors $\vec{v}$ and $\vec{w}$.**

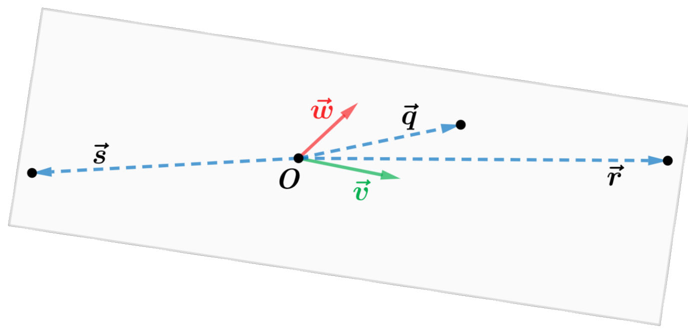


Figure 2.4.6: Plane $\mathcal{P} = \mathrm{Span}\{\vec{v}, \vec{w}\}$ in gray.
Three points on $\mathcal{P}$ are illustrated: $\vec{q} = \vec{v} + \vec{w}$, $\vec{r} = 3\vec{v} + \vec{w}$ and $\vec{s} = -2\vec{v} - \vec{w}$.

REMARKS

- $\mathcal{P}$ contains the points V and W, which are the terminal points associated with the vectors $\vec{v}$ and $\vec{w}$ respectively.
- If a point P with associated vector $\vec{p}$ lies on the plane, then $\vec{p} = s\vec{v} + t\vec{w}$, for some $s, t \in \mathbb{R}$.
- Any plane defined by the span of two vectors must pass through the origin.
- To form a plane, $\vec{v}, \vec{w}$ must be non-zero vectors with $\vec{w} \neq c\vec{v}$ for any $c \in \mathbb{R}$. If exactly one of $\vec{v}$ and $\vec{w}$ is $\vec{0}$, or if $\vec{w} = c\vec{v}$ for some $c \in \mathbb{R}$, then $\mathrm{Span}\{\vec{v}, \vec{w}\}$ is a line, not a plane. If $\vec{v} = \vec{w} = \vec{0}$, then $\mathrm{Span}\{\vec{v}, \vec{w}\} = \vec{0}$; this is not a plane, but rather a single point, the origin.

Definition 2.4.2

Vector Equation of a Plane in $\mathbb{R}^n$ Through the Origin

Let $\vec{v}, \vec{w}$ be non-zero vectors in $\mathbb{R}^n$ with $\vec{v} \neq c\vec{w}$ for any $c \in \mathbb{R}$. The expression

$$\vec{p} = s\vec{v} + t\vec{w}$$

is a **vector equation of the plane in $\mathbb{R}^n$ through the origin with direction vectors $\vec{v}$ and $\vec{w}$.**

Example 2.4.3

Determine a vector equation of the plane in $\mathbb{R}^3$ through the origin with direction vectors $\begin{bmatrix} 2 \\ 4 \\ 6 \end{bmatrix}$ and $\begin{bmatrix} -1 \\ 2 \\ -3 \end{bmatrix}$.

Solution: A vector equation of this plane is

$$\vec{p} = s \begin{bmatrix} 2 \\ 4 \\ 6 \end{bmatrix} + t \begin{bmatrix} -1 \\ 2 \\ -3 \end{bmatrix}, s, t \in \mathbb{R}.$$

Notice that the points $(2, 4, 6)$ and $(-1, 2, -3)$ are on this plane.

Example 2.4.4

Give a vector equation of the plane in $\mathbb{R}^5$ that passes through the origin with direction vectors $\begin{bmatrix} 5 & 4 & 3 & 2 & 1 \end{bmatrix}^T$ and $\begin{bmatrix} -5 & 4 & -3 & 2 & -1 \end{bmatrix}^T$.

Solution: A vector equation of this plane is

$$\vec{p} = s \begin{bmatrix} 5 \\ 4 \\ 3 \\ 2 \\ 1 \end{bmatrix} + t \begin{bmatrix} -5 \\ 4 \\ -3 \\ 2 \\ -1 \end{bmatrix}, s, t \in \mathbb{R}.$$

Notice that the points $(5, 4, 3, 2, 1)$ and $(-5, 4, -3, 2, -1)$ in $\mathbb{R}^5$ are on this plane.

Definition 2.4.5

Plane in $\mathbb{R}^n$

Let $\vec{u} \in \mathbb{R}^n$ and let $\vec{v}$ and $\vec{w}$ be non-zero vectors in $\mathbb{R}^n$ with $\vec{v} \neq c\vec{w}$, for any $c \in \mathbb{R}$. Then

$$\mathcal{P} = \{\vec{u} + s\vec{v} + t\vec{w} : s, t \in \mathbb{R}\}$$

is a **plane in $\mathbb{R}^n$ through $\vec{u}$ with direction vectors $\vec{v}$ and $\vec{w}$**. We say that $\vec{v}$ and $\vec{w}$ are **parallel** to $\mathcal{P}$.

REMARKS

- Letting $s = t = 0$ gives us that the vector $\vec{u}$ is on the plane.
- The plane passes through the terminal point U associated with $\vec{u}$. The other points on the plane move from U in the $\vec{v}$ and the $\vec{w}$ directions by linear combinations of $\vec{v}$ and $\vec{w}$.
- The terminal points, V and W, associated with the vectors, $\vec{v}$ and $\vec{w}$, are not usually on the plane. In fact, V is a point on the plane if and only if $\vec{u} \in \text{Span}\{\vec{v}, \vec{w}\}$; that is, if and only if U lies on the plane through the origin that contains V and W.

Example 2.4.6

Since the plane $\mathcal{M} = \left\{ \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} + s \begin{bmatrix} 2 \\ 4 \\ 6 \end{bmatrix} + t \begin{bmatrix} -1 \\ 2 \\ -3 \end{bmatrix} : s, t \in \mathbb{R} \right\}$ has the same direction vectors as the plane $\mathcal{P} = \left\{ s \begin{bmatrix} 2 \\ 4 \\ 6 \end{bmatrix} + t \begin{bmatrix} -1 \\ 2 \\ -3 \end{bmatrix} : s, t \in \mathbb{R} \right\}$ from Example 2.4.3, then $\mathcal{M}$ is parallel to $\mathcal{P}$. It can be shown that $\mathcal{M}$ does not pass through the origin.

Definition 2.4.7

Vector Equation of a Plane

Let $\vec{u} \in \mathbb{R}^n$ and let $\vec{v}$ and $\vec{w}$ be non-zero vectors in $\mathbb{R}^n$ with $\vec{v} \neq c\vec{w}$, for any $c \in \mathbb{R}$. Then

$$\vec{p} = \vec{u} + s\vec{v} + t\vec{w}, \quad s, t \in \mathbb{R}$$

is a **vector equation of the plane in $\mathbb{R}^n$ through $\vec{u}$ with direction vectors $\vec{v}$ and $\vec{w}$**.

Example 2.4.8

Give a vector equation of a plane in $\mathbb{R}^3$ which passes through the point $(1, -4, 6)$, and has vectors $\begin{bmatrix} 2 \\ 4 \\ 6 \end{bmatrix}$ and $\begin{bmatrix} -1 \\ 2 \\ -3 \end{bmatrix}$ parallel to it.

Solution: A vector equation of this plane is

$$\vec{p} = \begin{bmatrix} 1 \\ -4 \\ 6 \end{bmatrix} + s \begin{bmatrix} 2 \\ 4 \\ 6 \end{bmatrix} + t \begin{bmatrix} -1 \\ 2 \\ -3 \end{bmatrix}, s, t \in \mathbb{R}.$$

Note that the points $(2, 4, 6)$ and $(-1, 2, -3)$ are not on this plane; however, the vectors $\begin{bmatrix} 2 \\ 4 \\ 6 \end{bmatrix}$ and $\begin{bmatrix} -1 \\ 2 \\ -3 \end{bmatrix}$ are parallel to this plane.

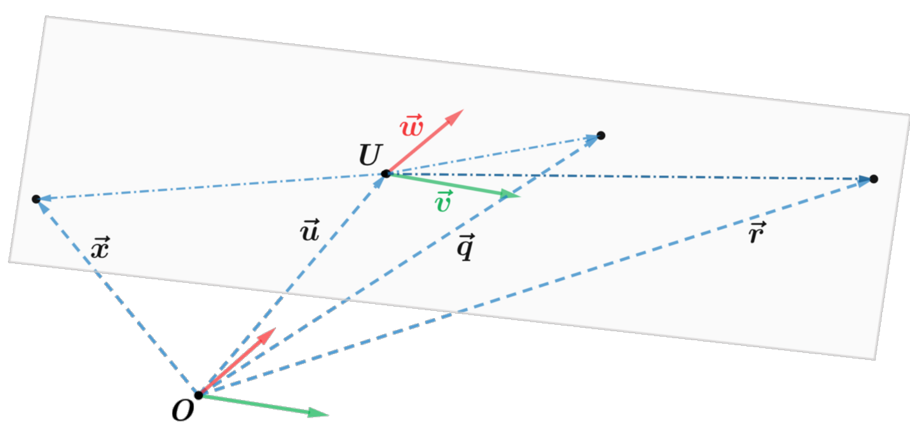


Figure 2.4.7: Plane $\mathcal{P}$ with equation $\vec{p} = \vec{u} + s\vec{v} + t\vec{w}$ in gray.
Plane $\mathcal{P}$ includes $\vec{q} = \vec{u} + \vec{v} + \vec{w}$ and $\vec{r} = \vec{u} + 3\vec{v} + \vec{w}$ and $\vec{x} = \vec{u} - 2\vec{v} - \vec{w}$.

Example 2.4.9

Give a vector equation of a plane $\mathcal{P}$ in $\mathbb{R}^5$ that passes through the point $(1, -4, 5, -2, 7)$ and has direction vectors $\begin{bmatrix} 2 & 4 & 6 & 8 & 10 \end{bmatrix}^T$ and $\begin{bmatrix} -1 & 2 & -3 & 4 & -5 \end{bmatrix}^T$.

Solution: A vector equation of this plane is

$$\vec{p} = \begin{bmatrix} 1 \\ -4 \\ 5 \\ -2 \\ 7 \end{bmatrix} + s \begin{bmatrix} 2 \\ 4 \\ 6 \\ 8 \\ 10 \end{bmatrix} + t \begin{bmatrix} -1 \\ 2 \\ -3 \\ 4 \\ -5 \end{bmatrix}, s, t \in \mathbb{R}.$$

Note that the points $(2, 4, 6, 8, 10)$ and $(-1, 2, -3, 4, -5)$ are not on $\mathcal{P}$. However, the vectors $\begin{bmatrix} 2 & 4 & 6 & 8 & 10 \end{bmatrix}^T$ and $\begin{bmatrix} -1 & 2 & -3 & 4 & -5 \end{bmatrix}^T$ are parallel to $\mathcal{P}$.

Noticing that $\begin{bmatrix} 2 & 4 & 6 & 8 & 10 \end{bmatrix}^T = 2\begin{bmatrix} 1 & 2 & 3 & 4 & 5 \end{bmatrix}^T$, a different vector equation for $\mathcal{P}$ is

$$\vec{p} = \begin{bmatrix} 1 \\ -4 \\ 5 \\ -2 \\ 7 \end{bmatrix} + s \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{bmatrix} + t \begin{bmatrix} -1 \\ 2 \\ -3 \\ 4 \\ -5 \end{bmatrix}, s, t \in \mathbb{R}.$$

REMARK

There are many different vector equations that could be used to describe the same plane $\mathcal{P}$.

EXERCISE

Determine two different vector equations for the plane $\vec{p} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + s \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} + t \begin{bmatrix} 1 \\ -1 \\ -1 \end{bmatrix}, s, t \in \mathbb{R}$.

We can also uniquely define a plane using three points. Let U, Q and R be three non-colinear points in $\mathbb{R}^n$ (that is, three points which do not all lie on the same line), with associated vectors $\vec{u}$, $\vec{q}$, and $\vec{r}$, respectively. Suppose that we want the equation of the unique plane containing these three points. The vectors $\vec{v} = \vec{q} - \vec{u}$ and $\vec{w} = \vec{r} - \vec{u}$ are parallel to the plane, and thus, we can express the equation of the plane as

$$\mathcal{P} = \{\vec{u} + s(\vec{q} - \vec{u}) + t(\vec{r} - \vec{u}) : s, t \in \mathbb{R}\}.$$

Example 2.4.10

Find a vector equation of a plane in $\mathbb{R}^4$ which contains the following points:

$$U = (2, -4, 6, -8),\ Q = (1, 3, -2, -4),\ \text{and } R = (9, 7, 5, 3).$$

Solution: We need to determine two non-parallel direction vectors. We have:

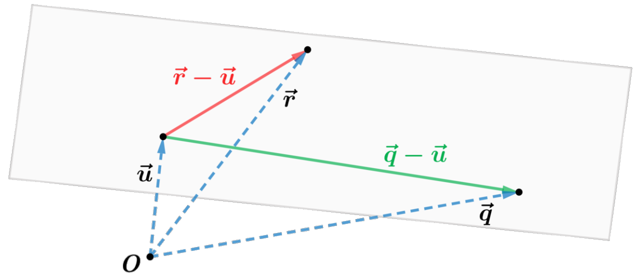


Figure 2.4.8: Plane $\mathcal{P} = \{\vec{u} + s(\vec{q}-\vec{u}) + t(\vec{r}-\vec{u}) : s,t \in \mathbb{R}\}$ in gray.

$$\vec{q}-\vec{u} = \begin{bmatrix}1\\3\\-2\\-4\end{bmatrix} - \begin{bmatrix}2\\-4\\6\\-8\end{bmatrix} = \begin{bmatrix}-1\\7\\-8\\4\end{bmatrix} \text{ and } \vec{r}-\vec{u} = \begin{bmatrix}9\\7\\5\\3\end{bmatrix} - \begin{bmatrix}2\\-4\\6\\-8\end{bmatrix} = \begin{bmatrix}7\\11\\-1\\11\end{bmatrix}.$$

A vector equation of this plane is then

$$\vec{p} = \vec{u} + s(\vec{q}-\vec{u}) + t(\vec{r}-\vec{u}) = \begin{bmatrix}2\\-4\\6\\-8\end{bmatrix} + s\begin{bmatrix}-1\\7\\-8\\4\end{bmatrix} + t\begin{bmatrix}7\\11\\-1\\11\end{bmatrix}, \quad s,t \in \mathbb{R}.$$

2.5 Scalar Equation of a Plane in $\mathbb{R}^3$

In $\mathbb{R}^3$ (and only in $\mathbb{R}^3$!), there is an additional way to generate an equation of a plane, using the cross product.

Let $\vec{v}$ and $\vec{w}$ be non-zero vectors in $\mathbb{R}^3$ with $\vec{v} \neq c\vec{w}$, for any $c \in \mathbb{R}$. We know that the cross product, $\vec{n} = \vec{v} \times \vec{w}$, is orthogonal to any plane to which $\vec{v}$ and $\vec{w}$ are both direction vectors. The vector $\vec{n}$ is referred to as a **normal vector** to such a plane. Suppose that U and P are points on the plane, with associated vectors, $\vec{u}$ and $\vec{p}$, respectively. Then the vector $\vec{p}-\vec{u}$ is parallel to the plane and is therefore orthogonal to $\vec{n}$; that is, $\vec{n} \cdot (\vec{p}-\vec{u}) = (\vec{v} \times \vec{w}) \cdot (\vec{p}-\vec{u}) = 0$.

Definition 2.5.1

Normal Form, Scalar Equation of a Plane in $\mathbb{R}^3$

Let $\mathcal{P}$ be a plane in $\mathbb{R}^3$ with direction vectors $\vec{v}$ and $\vec{w}$ and a normal vector $\vec{n} = \begin{bmatrix}a\\b\\c\end{bmatrix} \neq \vec{0}$.

Let $\vec{u} \in \mathcal{P}$ and $\vec{p} = \begin{bmatrix}x\\y\\z\end{bmatrix} \in \mathcal{P}$ where $\vec{p} \neq \vec{u}$. A **normal form** of $\mathcal{P}$ is given by

$$\vec{n} \cdot (\vec{p}-\vec{u}) = 0.$$

Expanding this, we arrive at a **scalar equation** (or **general form**) of $\mathcal{P}$,

$$ax + by + cz = d,$$

where $d = \vec{n} \cdot \vec{u}$.

REMARK

The plane goes through the origin, O

- if and only if the vector $\vec{0}$ satisfies this equation
- if and only if $(\vec{v} \times \vec{w}) \cdot (\vec{0} - \vec{u}) = 0$
- if and only if $\vec{u} = a\vec{v} + b\vec{w}$, for some $a, b \in \mathbb{R}$.

Geometrically, the plane goes through the origin if and only if both V and W lie on the plane.

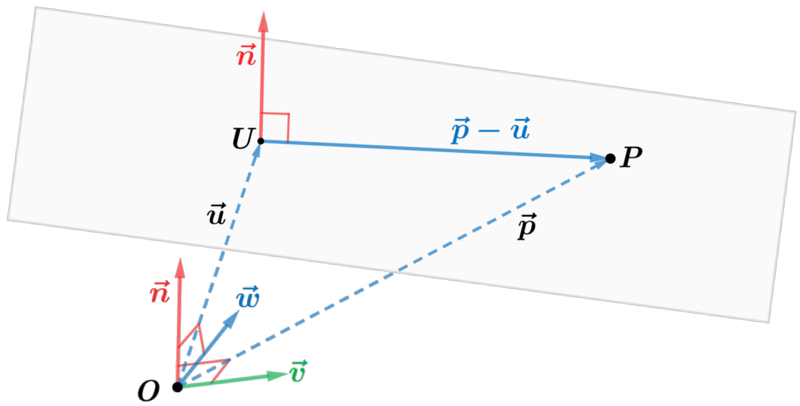


Figure 2.5.9: Let $\vec{n} = \vec{v} \times \vec{w}$. The plane in $\mathbb{R}^3$ with normal form $(\vec{v} \times \vec{w}) \cdot (\vec{p} - \vec{u}) = 0$ is in gray. It passes through $\vec{u}$ and has direction vectors $\vec{v}$ and $\vec{w}$.

Example 2.5.2

Find a scalar equation of the plane in $\mathbb{R}^3$ which passes through the origin and has direction vectors $\begin{bmatrix} 2 \\ 4 \\ 7 \end{bmatrix}$ and $\begin{bmatrix} -1 \\ 2 \\ -3 \end{bmatrix}$.

Solution: Let $\vec{n}$ be a normal to the plane and let $\vec{p} = \begin{bmatrix} x \\ y \\ z \end{bmatrix}$ be any vector on the plane. Then

$$\vec{n} = \begin{bmatrix} 2 \\ 4 \\ 7 \end{bmatrix} \times \begin{bmatrix} -1 \\ 2 \\ -3 \end{bmatrix} = \begin{bmatrix} -26 \\ -1 \\ 8 \end{bmatrix}.$$

A normal form of the plane (going through the origin) is thus given by

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} \cdot \begin{bmatrix} -26 \\ -1 \\ 8 \end{bmatrix} = 0.$$

Expanding this, we obtain a scalar equation of the plane

$$-26x - y + 8z = 0.$$

Note that the two points $(2,4,7)$ and $(-1,2,-3)$ lie on this plane.

Example 2.5.3 Give a scalar equation of the plane in $\mathbb{R}^3$ which passes through the point $(1,-4,3)$ and has the two vectors $\begin{bmatrix}2\\4\\7\end{bmatrix}$ and $\begin{bmatrix}-1\\2\\-3\end{bmatrix}$ parallel to it.

Solution: Let $\vec{n}$ be a normal to the plane and $\vec{p}=\begin{bmatrix}x\\y\\z\end{bmatrix}$ be any vector on the plane.

$$\vec{n}=\begin{bmatrix}2\\4\\7\end{bmatrix}\times\begin{bmatrix}-1\\2\\-3\end{bmatrix}=\begin{bmatrix}-26\\-1\\8\end{bmatrix}.$$

A scalar equation of the plane (not going through the origin) is then given by

$$\begin{aligned}\left(\begin{bmatrix}x\\y\\z\end{bmatrix}-\begin{bmatrix}1\\-4\\3\end{bmatrix}\right)\cdot\begin{bmatrix}-26\\-1\\8\end{bmatrix} &= 0\\ -26(x-1)-(y+4)+8(z-3) &= 0\\ -26x-y+8z &= 2.\end{aligned}$$

Note that the two points $(2,4,7)$ and $(-1,2,-3)$ do not lie on this plane.

REMARKS

1. Examining the definition of a scalar equation of a plane above, we see that the components of a normal vector $\vec{n}=\begin{bmatrix}a\\b\\c\end{bmatrix}$ are recorded as the coefficients of x,y,z in a scalar equation. We can use this information to quickly retrieve a normal form from a scalar equation by identifying a single vector on the plane.

2. We use the language "a scalar equation", "a normal vector", and "a normal form" of a plane $\mathcal{P}$ to emphasize that these are *not unique*.

   - Normal vectors of a plane are simply non-zero vectors that are orthogonal to any vector that is parallel to $\mathcal{P}$. If $\vec{n}$ is a normal vector of $\mathcal{P}$, then so is $c\vec{n}$ for any non-zero scalar $c\in\mathbb{R}$, so $\mathcal{P}$ does not have a unique normal vector (and thus normal form). That said, all normal vectors of $\mathcal{P}$ lie on the same line.
   - A scalar equation of a plane is an equation which all vectors of $\mathcal{P}$ satisfy. We know that multiplying an equation by a non-zero $c\in\mathbb{R}$ does not change the solution set to that equation, so $\mathcal{P}$ does not have a unique scalar equation.

Example 2.5.4

Let $\mathcal{P}$ be a plane in $\mathbb{R}^3$ with scalar equation $2x - y + z = 7$. Give a normal form of $\mathcal{P}$.

Solution: From the scalar equation, we see that a normal vector of $\mathcal{P}$ is $\vec{n} = \begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix}$. To identify a vector on $\mathcal{P}$, we can pick any constants we want for *exactly two* of x, y, z and solve for the third value. Taking $x = y = 0$, we have

$$7 = 2(0) - 0 + z = z,$$

so $\vec{u} = \begin{bmatrix} 0 \\ 0 \\ 7 \end{bmatrix} \in \mathcal{P}$. Therefore, a normal form of $\mathcal{P}$ is

$$\begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix} \cdot \left(\begin{bmatrix} x \\ y \\ z \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \\ 7 \end{bmatrix} \right) = 0.$$

Chapter 3

Systems of Linear Equations

3.1 Introduction

We begin with some introductory examples in which linear systems of equations naturally arise.

Example 3.1.1 At the market, 3 watermelons and 7 kiwis cost \$19 and 2 watermelons and 5 kiwis cost \$13. How much does a watermelon cost and how much does a kiwi cost?

Solution: Let w denote the price of a watermelon and let k denote the price of a kiwi. We convert the given information into a system of two equations.

$$\begin{aligned} 3w + 7k &= 19 \qquad (e_1) \\ 2w + 5k &= 13 \qquad (e_2) \end{aligned}$$

We must find values for w and k that satisfy the equations (e_1) and (e_2) simultaneously.

If we multiply (e_1) by 2 and subtract from that 3 times (e_2), we obtain

$$[2(3) - 3(2)]w + [2(7) - 3(5)]k = 2(19) - 3(13).$$

That is, $-k = -1$, which implies that $k = 1$.

Substituting this value for k into equation (e_1) gives $w = \frac{19-7(1)}{3} = 4$.

Therefore, the price of a watermelon is \$4 and the price of a kiwi is \$1.

Substituting these values into the two equations

$$\begin{aligned} 3(4) + 7(1) &= 19 \\ 2(4) + 5(1) &= 13 \end{aligned}$$

shows that they satisfy both equations.

The crucial step in solving this problem is that of taking the combination $2(e_1) - 3(e_2)$. This step is determined by looking at the coefficients of w in the two equations (e_1) and (e_2), which are 3 and 2, respectively. The combination of $2(e_1) - 3(e_2)$ produces a new equation in which w has been eliminated. We then solve this new equation for the other variable, k. Once we have a value for k, we then substitute it into one of the original two equations and solve for the value of the other variable, w.

Alternatively, we could have considered the combination $5(e_1) - 7(e_2)$, to eliminate the variable k, and then solve for w. We then could substitute its value into either equation, (e_1) or (e_2) in order to find the value of k.

Example 3.1.2

Find the equation of the line that passes through the points $(2, 3)$ and $(-3, 4)$.

Solution: The general equation of a line is $y = mx + b$. We must determine the values of the constants m and b. Since both points lie on the line, they must both satisfy the equation. Therefore,

$$3 = 2m + b \quad \text{and} \quad 4 = -3m + b.$$

Thus, we have a system with two equations and two variables, m and b:

$$\begin{aligned} 2m + b &= 3 \\ -3m + b &= 4 \end{aligned}$$

We will not solve this system now.

Example 3.1.3

Determine whether the vector $\vec{v} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$ lies in Span $\left\{ \begin{bmatrix} -1 \\ 2 \\ -3 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right\}$.

Solution: The vector $\vec{v}$ will lie in the span of the two given vectors if, and only if, there exist scalars $a, b \in \mathbb{R}$, such that

$$\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = a \begin{bmatrix} -1 \\ 2 \\ -3 \end{bmatrix} + b \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix},$$

or equivalently

$$\begin{array}{rcrr} 1 & = & -1a & +b \\ 2 & = & 2a & +b \\ 3 & = & -3a & +b \end{array}.$$

We have a system with three equations and two variables. We will not solve this system now.

3.2 Systems of Linear Equations

Definition 3.2.1

Linear Equation, Coefficient, Constant Term

A **linear equation** in n variables (or unknowns) $x_1, x_2, \ldots, x_n$ is an equation that can be written in the form $a_1x_1 + a_2x_2 + \cdots + a_nx_n = b$, where $a_1, a_2, \ldots, a_n, b \in \mathbb{F}$. The scalars $a_1, a_2, \ldots, a_n$ are the **coefficients** of $x_1, x_2, \ldots, x_n$, respectively, and b is called the **constant term**.

Definition 3.2.2

System of Linear Equations

A **system of linear equations** is a collection of m linear equations in n variables, $x_1, \ldots, x_n$:

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

We will use the convention that a_{ij} is the coefficient of x_j in the i^{th} equation.

Definition 3.2.3

Solve, Solution

We say that the scalars $y_1, y_2, \ldots, y_n$ in $\mathbb{F}$ **solve** a system of linear equations if, when we set $x_1 = y_1, x_2 = y_2, \ldots, x_n = y_n$ in the system, then each of the equations is satisfied:

$$\begin{array}{cccccc} a_{11}y_1 & +a_{12}y_2 & + & \cdots & +a_{1n}y_n & = & b_1 \\ a_{21}y_1 & +a_{22}y_2 & + & \cdots & +a_{2n}y_n & = & b_2 \\ & & & \vdots & & & \\ a_{m1}y_1 & +a_{m2}y_2 & + & \cdots & +a_{mn}y_n & = & b_m \end{array}$$

We also say that the vector $\vec{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$ is a **solution** to the system.

Note that if at least one equation in the system of linear equations is *not* satisfied by the entries in $\vec{y}$, then $\vec{y}$ is not a solution.

Example 3.2.4

Consider the following system of linear equations:

$$\begin{array}{ccccc} x_1 & -2x_2 & +3x_3 & = & 1 \\ 2x_1 & -4x_2 & +6x_3 & = & 2 \\ 3x_1 & -x_2 & +x_3 & = & 3 \end{array}.$$

The vector $\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$ is a solution, as

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

However, $\begin{bmatrix} -2 \\ 0 \\ 1 \end{bmatrix}$ is not a solution, because even though $1(-2) - 2(0) + 3(1) = 1$ and $2(-2) - 4(0) + 6(1) = 2$, we have that $3(-2) - 1(0) + 1(1) = -5 \neq 3$.

Definition 3.2.5
Solution Set

The set of all solutions to a system of linear equations is called the **solution set** to the system.

Example 3.2.6

Solve $2x = 4$. That is, find the solution set to this system of linear equations (a system with one equation).

Solution: The solution to this equation is $x = 2$ and thus, the solution set S is $\{2\}$. We can verify that $2(2) = 4$.

Example 3.2.7

Solve $2x + 4y = 6$.

Solution: We have only one equation for the two unknowns, x and y. There will be an infinite number of solutions to this equation. Indeed, given any real number y, we can solve for x to obtain $x = 3 - 2y$.

In this way we see that x depends on y, but that y can be freely chosen to be any real number. We emphasize this point by writing $y = t$, where the variable $t \in \mathbb{R}$ is called a **parameter**. Then $x = 3 - 2t$, and therefore the solution set, S, can be expressed as

$$S = \left\{ \begin{bmatrix} 3-2t \\ t \end{bmatrix} : t \in \mathbb{R} \right\}.$$

We can verify our solution by setting $y = t$ and $x = 3 - 2t$:

$$2x + 4y = 2(3 - 2t) + 4t = 6 - 4t + 4t = 6,$$

which is the right-hand side of the equation.

Note that declaring $y = t$ as our parameter was an arbitrary choice. We could have just as well let $x = t$ and solved for y in terms of t. The important property to note is that a parameter was needed to describe the solution set S.

Example 3.2.8

Solve the system of linear equations

$$\begin{aligned} 2x + 3y &= 4 \\ 2x + 3y &= 5 \end{aligned}$$

Solution: Since both equations have identical left-hand sides, but different right-hand sides, it will be impossible to find values for x and y that satisfy both of these equations. Therefore, the solutions set is $S = \varnothing$.

Theorem 3.2.9

(The Solution Set to a System of Linear Equations)

The solution set to a system of linear equations is exactly one of the following:

(a) empty (there are no solutions),

(b) contains exactly one element (there is a unique solution), or

(c) contains an infinite number of elements (the solution set has one or more parameters).

We have given examples above to illustrate these three outcomes. However, at this time, we will not give a proof of this result.

Definition 3.2.10 **Inconsistent and Consistent Systems**

If the solution set to a system of linear equations is empty, then we say that the system is **inconsistent**.

If the solution set has a unique solution or infinitely many solutions, then we say that the system is **consistent**.

Example 3.2.11

Solve the system $\begin{matrix} x = 1 \\ x = 2 \end{matrix}$.

Solution: There is no solution and $S = \varnothing$. This system is inconsistent.

Example 3.2.12

Solve the system $\begin{matrix} x + y = 2 \\ x - y = 4 \end{matrix}$.

Solution: There is a unique solution, $x = 3$ and $y = -1$. Therefore, $S = \left\{ \begin{bmatrix} 3 \\ -1 \end{bmatrix} \right\}$. This system is consistent.

Example 3.2.13

Solve the system $x + y = 1$.

Solution: Let $y = t$, where $t \in \mathbb{R}$. Therefore, $x = 1-t$. We obtain infinitely many solutions and $S = \left\{ \begin{bmatrix} 1-t \\ t \end{bmatrix} : t \in \mathbb{R} \right\}$. This system is consistent.

Note that if we had let $x = t$, where $t \in \mathbb{R}$, we would obtain $y = 1 - t$, and $S = \left\{ \begin{bmatrix} t \\ 1-t \end{bmatrix} : t \in \mathbb{R} \right\}$.

These are two different representations of the same set S. Generally there will be many different ways of expressing any given set.

It is usually not obvious whether a system is inconsistent, will have a unique solution, or will have infinitely many solutions. The key idea to solving a system of equations is to manipulate the system into another system of equations which is easier to solve and has the same solution set as the original system.

Definition 3.2.14 **Equivalent Systems**

We say that two linear systems are **equivalent** whenever they have the same solution set.

Given a system of linear equations, we will manipulate it into equivalent systems and we will stop once we have obtained an equivalent system whose solution set is easily obtained.

The question then becomes: what manipulations can we do to a system which will produce an equivalent system?

Example 3.2.15

Consider the system $\begin{array}{rrcl} x & +y & = & 1 \\ 2x & +3y & = & 6 \end{array}$. If we multiply the first equation by zero, we get a new system:

$$0 = 0$$
$$2x + 3y = 6$$

This system is not equivalent to the original system. For example, $x = 3$ and $y = 0$ is a solution to the new system, but it is not a solution to the original system. We have lost information by multiplying the first equation by zero.

Example 3.2.16

Consider the system $\begin{array}{rrrcl} x & +y & +z & = & 3 \\ 2x & +3y & -2z & = & 3 \end{array}$

We could add an additional equation to get a new system.

$$\begin{array}{rrrcl} x & +y & +z & = & 3 \\ 2x & +3y & -2z & = & 3 \\ x & +2y & -3z & = & 5 \end{array}$$

This new system is not equivalent to the old one. For example, $x = 1, y = 1, z = 1$ is a solution to the original system, but it is not a solution to the new system. We have added an extra restriction on the variables by adding in this new equation. When manipulating a system, we must be careful that we neither lose information nor add information.

Definition 3.2.17
Elementary Operations

Consider a system of m linear equations in n variables. The equations are ordered and labelled from e_1 to e_m. The following three operations are known as **elementary operations**.

Equation swap elementary operation: **interchange two equations.**

$$e_i \leftrightarrow e_j \quad \text{Interchange equations } e_i \text{ and } e_j.$$

Equation scale elementary operation: **multiply one equation by a non-zero constant.**

$$e_i \to me_i, m \in \mathbb{F}\backslash\{0\} \quad \text{Replace equation } e_i \text{ by } m \text{ times equation } e_i.$$

Equation addition elementary operation: **add a multiple of one equation to another equation.**

$$\left.\begin{array}{c} e_j \to ce_i + e_j \\ i \neq j, c \in \mathbb{F} \end{array}\right\} \text{Replace equation } e_j \text{ by adding } e_j \text{ and a multiple of equation } e_i.$$

Some sources refer to these operations as **Type I**, **Type II**, and **Type III** operations, respectively.

When performing a series of elementary operations, we will use the convention that e_i and e_j refer to the current system we are working with. Therefore, e_i and e_j will change as the system is manipulated.

Theorem 3.2.18 **(Elementary Operations)**

If a single elementary operation of any type is performed on a system of linear equations, then the system produced will be equivalent to the original system.

In practice, there might be other operations that we want to perform. These operations will not be elementary operations, instead they will be *combinations* of elementary operations. For example,

$$e_j \to ce_i - e_j$$

is not an elementary operation, but rather it is a combination of two elementary operations. Can you determine which ones?

Using these operations will still produce an equivalent system, but they will cause confusion later. *Therefore, when manipulating a system of equations, we will make use of the three elementary operations only.*

Note that if at any point we produce an equation of the form

$$0 = a, \text{ where } a \neq 0,$$

then the system is inconsistent and we stop at once.

Definition 3.2.19 **Trivial Equation**

We refer to the equation $0 = 0$ as the **trivial equation**. Any other equation is known as a **non-trivial** equation.

The trivial equation is always true and it has no information content. If at any point when performing elementary operations we produce the trivial equation, then we move this equation to the end of the system, by performing an equation swap elementary operation. We might be tempted to ignore such equations and not write them down. In practice, we keep any (there may be more than one) equations of this form in our new system, so that we will always have the same number of equations in any of the equivalent systems.

The goal of performing elementary operations is to obtain a system whose solution set is easier to determine. Let us refer to this system as our "final equivalent system." The final equivalent system is not unique and depends on the sequence of elementary operations used.

Example 3.2.20

Solve the following system of three linear equations in three unknowns:

$$\begin{array}{rrrcr} -2x_1 & -4x_2 & -6x_3 & = & 4 \\ 3x_1 & +6x_2 & +10x_3 & = & 6 \\ & x_2 & +2x_3 & = & 5 \end{array}.$$

Solution: Suppose that we perform the elementary operation

$$e_1 \to -\frac{1}{2}e_1 \quad \text{to get the new system} \quad \begin{array}{rrrcr} x_1 & +2x_2 & +3x_3 & = & -2 \\ 3x_1 & +6x_2 & +10x_3 & = & 6 \\ & x_2 & +2x_3 & = & 5 \end{array}.$$

We will now make use of this new equation e_1 to remove the x_1 term from e_2.

$$e_2 \to -3e_1 + e_2 \quad \text{gives} \quad \begin{array}{rrrcr} x_1 & +2x_2 & +3x_3 & = & -2 \\ & & x_3 & = & 12 \\ & x_2 & +2x_3 & = & 5 \end{array}$$

This could be our final equivalent system. From e_2 we obtain that $x_3 = 12$. We could make use of this fact and e_3 to get $x_2 + 2(12) = 5$, so that $x_2 = -19$. We can put the values for x_2 and x_3 into e_1 to get $x_1 + 2(-19) + 3(12) = -2$, so that $x_1 = 0$.

However, instead of doing these substitutions, we may wish to continue applying elementary operations. We perform the operation

$$e_2 \leftrightarrow e_3 \text{ to get} \quad \begin{array}{rrrcr} x_1 & +2x_2 & +3x_3 & = & -2 \\ & x_2 & +2x_3 & = & 5 \\ & & x_3 & = & 12 \end{array}.$$

Then we perform

$$e_2 \to -2e_3 + e_2 \quad \text{to get} \quad \begin{array}{rrrcr} x_1 & +2x_2 & +3x_3 & = & -2 \\ & x_2 & & = & -19 \\ & & x_3 & = & 12 \end{array},$$

$$e_1 \to -3e_3 + e_1 \quad \text{to get} \quad \begin{array}{rrrcr} x_1 & +2x_2 & & = & -38 \\ & x_2 & & = & -19 \\ & & x_3 & = & 12 \end{array},$$

$$e_1 \to -2e_2 + e_1 \quad \text{to finally get} \quad \begin{array}{rrrcr} x_1 & & & = & 0 \\ & x_2 & & = & -19 \\ & & x_3 & = & 12 \end{array}.$$

This is our final equivalent system, and its solution set is

$$S = \left\{ \begin{bmatrix} 0 \\ -19 \\ 12 \end{bmatrix} \right\}.$$

It is clear from this example that there will be plenty of choices on which elementary operations to apply at any stage and also that there will be choices as to when we will stop at our final equivalent system. There is no "best" final equivalent system. However, there are two particular forms that are preferable for the final equivalent system. We will attempt to motivate these two forms and discuss how they can be obtained.

3.3 An Approach to Solving Systems of Linear Equations

Let us consider a system of m linear equations and n unknowns. We normally try to solve for the variables of the system in the alphabetical or numerical order in which they appear. Let us assume that the variables are labelled $x_1, x_2, \ldots, x_n$. We assume that there is at least one equation with the variable x_1 appearing in it with a non-zero coefficient (by relabeling the variables, if necessary).

We would like the first equation to tell us about the first variable, x_1, so that it is of the form

$$a_1x_1 + \cdots = b_1,$$

with $a_1 \neq 0$. We will make use of the first equation to remove the variable x_1 from all the other equations *after* the first equation by performing equation addition elementary operations.

We would like the second equation to tell us about the next variable that can be obtained. Call this x_i. Often this would be the variable x_2, so that $i = 2$, but this is not always the case. The (new) second equation has the form

$$a_ix_i + \cdots = b_2,$$

with $a_i \neq 0$. We will make use of the second equation to remove the variable x_i from all the other equations *after* the second equation by performing equation addition elementary operations.

We repeat this process, moving down through the equations, performing equation swap and equation addition elementary operations as needed. We also move any trivial equations to the end of the system.

We assume that the r^{th} equation is the last non-trivial equation. (It is possible that r is equal to m.) We also assume that this equation tells us about the k^{th} variable, x_k. Note that this equation should not involve the variables which were removed from the preceding equations by the process above. (x_1 from equation 1, x_i from equation 2, etc.) Therefore, we have an equation of the form

$$a_kx_k + \cdots = b_r,$$

with $a_k \neq 0$.

We illustrate the process up to this point with the following example.

Example 3.3.1

Solve the following system of five linear equations in four unknowns:

$$\begin{array}{rrrrcr} x_1 & +2x_2 & & & = & 1 \\ x_1 & +2x_2 & +3x_3 & +x_4 & = & 0 \\ -x_1 & -x_2 & +x_3 & +x_4 & = & -2 \\ & x_2 & +x_3 & +x_4 & = & -1 \\ & -x_2 & +2x_3 & & = & 0 \end{array}.$$

Solution: We will apply elementary operations, usually just one or two at a time. At each stage, we will produce an equivalent system. We will stop the process when we produce an equivalent system whose solution set is relatively easily obtained.

We notice that e_1 has x_1 in it and we will use it to eliminate the variable x_1 from all the equations below it. (If e_1 did not have any x_1 terms, then we would have switched it with one that did.)

We perform the elementary operations

$$e_2 \to -e_1 + e_2$$
$$e_3 \to e_1 + e_3$$

We notice that e_4 and e_5 do not contain x_1 and so they are not modified at this stage. We get

$$\begin{array}{rrrrl} x_1 & +2x_2 & & & = 1 \\ & & 3x_3 & +x_4 & = -1 \\ & x_2 & +x_3 & +x_4 & = -1 \\ & x_2 & +x_3 & +x_4 & = -1 \\ & -x_2 & +2x_3 & & = 0 \end{array} .$$

We notice that e_2 does not have an x_2 variable, but e_3 does. Therefore, we perform the elementary operation

$$e_2 \leftrightarrow e_3$$

to get

$$\begin{array}{rrrrl} x_1 & +2x_2 & & & = 1 \\ & x_2 & +x_3 & +x_4 & = -1 \\ & & 3x_3 & +x_4 & = -1 \\ & x_2 & +x_3 & +x_4 & = -1 \\ & -x_2 & +2x_3 & & = 0 \end{array} .$$

There is an x_2 in e_2 and we can make use of this fact to remove the variable x_2 from the equations below it.

Performing the elementary operations

$$e_4 \to -e_2 + e_4$$
$$e_5 \to e_2 + e_5$$

yields

$$\begin{array}{rrrrl} x_1 & +2x_2 & & & = 1 \\ & x_2 & +x_3 & +x_4 & = -1 \\ & & 3x_3 & +x_4 & = -1 \\ & & & 0 & = 0 \\ & & 3x_3 & +x_4 & = -1 \end{array} .$$

There is an x_3 in e_3 and we make use of this fact to remove x_3 from all the equations below it. In this case, we need only perform

$$e_5 \to -e_3 + e_5$$

to obtain

$$\begin{array}{rrrrl} x_1 & +2x_2 & & & = 1 \\ & x_2 & +x_3 & +x_4 & = -1 \\ & & 3x_3 & +x_4 & = -1 \\ & & & 0 & = 0 \\ & & & 0 & = 0 \end{array} .$$

Notice that the two trivial equations are at the end.

At this point in the process, the form of the equivalent system of equations has a particularly simple appearance. We may choose to stop performing elementary operations at this point, and so this would be our final equivalent system.

Note that a different sequence of elementary operations could produce a different final equivalent system of a similar form. For example, a final equivalent system could involve the equation $6x_3 + 2x_4 = -2$.

Proceeding from the original system to the final equivalent system in this form is known as the **forward elimination phase**.

Once we have a final equivalent system in this form, we can obtain the solution set using a process known as **back substitution**. We use the information given in the r^{th} equation for the variable x_k to then **back substitute** for the variable x_k into all the previous equations in our final equivalent system. Then we repeat this process moving back up the system. Our last step is to use the information given in the second equation for the variable x_i to back substitute for the variable x_i into the first equation (if it appears there).

Let us return to our example.

Example 3.3.1 Our final equivalent system was

$$\begin{array}{rrrrcr} x_1 & +2x_2 & & & = & 1 \\ & x_2 & +x_3 & +x_4 & = & -1 \\ & & 3x_3 & +x_4 & = & -1 \\ & & & 0 & = & 0 \\ & & & 0 & = & 0 \end{array}.$$

We have completed the forward elimination phase and this system is a good choice for a final equivalent system. It will have the same solution set as the original system and we can obtain its solution by back substitution.

We notice that e_4 and e_5 are trivial equations and do not give us any information. We also notice that e_3 gives us a relationship between variables x_3 and x_4. One of these variables can arbitrarily be assigned any real value and the value of the other variable will depend on this arbitrary value.

We will choose to assign x_4 any real value and to emphasize this fact we will let $x_4 = t$, where $t \in \mathbb{R}$.

Therefore, e_3 tells us that

$$x_3 = \frac{1}{3}(-1 - t).$$

Then e_2 tells us that

$$x_2 = -1 - x_3 - x_4 = -1 - \frac{1}{3}(-1 - t) - t = -\frac{2}{3} - \frac{2t}{3}.$$

Then e_1 tells us that

$$x_1 = 1 - 2x_2 = 1 - 2\left(-\frac{2}{3} - \frac{2t}{3}\right) = \frac{7}{3} + \frac{4t}{3}.$$

Therefore, the solution set is

$$S = \left\{ \begin{bmatrix} \frac{7}{3} + \frac{4}{3}t \\ -\frac{2}{3} - \frac{2}{3}t \\ -\frac{1}{3} - \frac{1}{3}t \\ t \end{bmatrix} : t \in \mathbb{R} \right\} = \left\{ \begin{bmatrix} \frac{7}{3} \\ -\frac{2}{3} \\ -\frac{1}{3} \\ 0 \end{bmatrix} + t \begin{bmatrix} \frac{4}{3} \\ -\frac{2}{3} \\ -\frac{1}{3} \\ 1 \end{bmatrix} : t \in \mathbb{R} \right\}.$$

An alternative way to proceed, instead of performing back substitution, is to continue to simplify the system. The following extension yields a unique final equivalent system, which is particularly simple.

First, using equation scale elementary operations, scale each of the non-trivial equations of the current system. Scale the first equation by the factor $\frac{1}{a_1}$, the second equation by the factor $\frac{1}{a_i}$ and so on, scaling the r^{th} equation by the factor $\frac{1}{a_k}$.

Using the r^{th} equation and equation addition elementary operations, we eliminate the variable x_k from all the equations *above* it. We then examine the $(r-1)^{st}$ equation, which will be of the form

$$x_j + \cdots = b_{r-1},$$

for some natural number $j < k$. Using this equation and equation addition elementary operations, we eliminate the variable x_j from all the equations above it.

We repeat this process, moving up through the equations, and performing equation addition elementary operations to eliminate variables. The process ends when we use the second equation to eliminate the variable x_i from the first equation.

These additional steps, performed after the end of the forward elimination phase, are collectively known as the **backward elimination phase**. When we are performing these steps we say that we are doing **backward elimination**.

For a given system, there are many different possibilities for the equivalent system after doing forward elimination, depending on the particular elementary operations used. However, regardless of the elementary operations used, the same final equivalent system will be obtained after performing both forward and backward elimination. We will formally establish this result in Theorem 3.5.8 (Unique RREF).

Once again we return to our example.

Example 3.3.1 Our final equivalent system after the forward elimination phase was

$$\begin{array}{ccccc} x_1 & +2x_2 & & & = & 1 \\ & x_2 & +x_3 & +x_4 & = & -1 \\ & & 3x_3 & +x_4 & = & -1 \\ & & & 0 & = & 0 \\ & & & 0 & = & 0 \end{array}$$

and instead of using back substitution, we continue performing elementary operations.

We observe that the last two equations are trivial equations and e_3 is in terms of both variables x_3 and x_4. We perform the elementary operation

$$e_3 \to \frac{1}{3}e_3$$

to get

$$\begin{array}{cccccc} x_1 & +2x_2 & & & = & 1 \\ & x_2 & +x_3 & +x_4 & = & -1 \\ & & x_3 & +\frac{1}{3}x_4 & = & -\frac{1}{3} \\ & & & 0 & = & 0 \\ & & & 0 & = & 0 \end{array} .$$

We use e_3 to eliminate the variable x_3 from the equations above it, by performing

$$e_2 \to -e_3 + e_2$$

to get

$$\begin{array}{ccccccc} x_1 & +2x_2 & & & = & 1 \\ & x_2 & & +\frac{2}{3}x_4 & = & -\frac{2}{3} \\ & & x_3 & +\frac{1}{3}x_4 & = & -\frac{1}{3} \\ & & & 0 & = & 0 \\ & & & 0 & = & 0 \end{array}.$$

We use e_2 to eliminate variable x_2 from the equation above it by performing

$$e_1 \to -2e_2 + e_1$$

to get

$$\begin{array}{cccccc} x_1 & & & -\frac{4}{3}x_4 & = & \frac{7}{3} \\ & x_2 & & +\frac{2}{3}x_4 & = & -\frac{2}{3} \\ & & x_3 & +\frac{1}{3}x_4 & = & -\frac{1}{3} \\ & & & 0 & = & 0 \\ & & & 0 & = & 0 \end{array}.$$

We have completed the backward elimination phase. We have that $x_4 = t, t \in \mathbb{R}$. From e_1 we obtain that

$$x_1 = \frac{7}{3} + \frac{4}{3}t.$$

From e_2 we obtain that

$$x_2 = -\frac{2}{3} - \frac{2}{3}t.$$

From e_3 we obtain that

$$x_3 = -\frac{1}{3} - \frac{1}{3}t.$$

Therefore, the solution set is

$$S = \left\{ \begin{bmatrix} \frac{7}{3} + \frac{4}{3}t \\ -\frac{2}{3} - \frac{2}{3}t \\ -\frac{1}{3} - \frac{1}{3}t \\ t \end{bmatrix} : t \in \mathbb{R} \right\} = \left\{ \begin{bmatrix} \frac{7}{3} \\ -\frac{2}{3} \\ -\frac{1}{3} \\ 0 \end{bmatrix} + t \begin{bmatrix} \frac{4}{3} \\ -\frac{2}{3} \\ -\frac{1}{3} \\ 1 \end{bmatrix} : t \in \mathbb{R} \right\}.$$

If we use another parameter, $u = \dfrac{t}{3}$, then we remove some of the fractions to get

$$S = \left\{ \begin{bmatrix} \frac{7}{3} + 4u \\ -\frac{2}{3} - 2u \\ -\frac{1}{3} - u \\ 3u \end{bmatrix} : u \in \mathbb{R} \right\} = \left\{ \begin{bmatrix} \frac{7}{3} \\ -\frac{2}{3} \\ -\frac{1}{3} \\ 0 \end{bmatrix} + u \begin{bmatrix} 4 \\ -2 \\ -1 \\ 3 \end{bmatrix} : u \in \mathbb{R} \right\}.$$

Alternatively, we could let $s = \dfrac{t-2}{3}$ and get an even nicer looking solution set

$$S = \left\{ \begin{bmatrix} 5 \\ -2 \\ -1 \\ 2 \end{bmatrix} + s \begin{bmatrix} 4 \\ -2 \\ -1 \\ 3 \end{bmatrix} : s \in \mathbb{R} \right\}.$$

REMARK

Remember that, when a solution set contains at least one parameter, you can obtain particular solutions by setting particular values to parameters. In the case of Example 3.3.1, the solution set of a system of linear equations is given by

$$S = \left\{ \begin{bmatrix} 5 \\ -2 \\ -1 \\ 2 \end{bmatrix} + s \begin{bmatrix} 4 \\ -2 \\ -1 \\ 3 \end{bmatrix} : s \in \mathbb{R} \right\}.$$

By setting $s = -1$, $s = 0$ and $s = 1$, we find that

$$\begin{bmatrix} 1 \\ 0 \\ 0 \\ -1 \end{bmatrix}, \quad \begin{bmatrix} 5 \\ -2 \\ -1 \\ 2 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 9 \\ -4 \\ -2 \\ 5 \end{bmatrix}$$

are particular solutions to the given system of linear equations.

3.4 Solving Systems of Linear Equations Using Matrices

When we are solving a system of equations, we manipulate the system using elementary operations. At each stage, we have an equivalent system of equations. That is, each system has the same solution set. We can stop at any stage with any equivalent system as soon as we find that we can write down the solution set.

In the systems of equations that we have solved, notice that we write down the variables many times in the process, but it turns out they are only really important at the beginning and the end of the process.

Moving forwards, we will only need to write down the coefficients and the constant terms in the intermediate steps. To do this, we will make use of matrices.

Definition 3.4.1
Matrix, Entry

An $m \times n$ **matrix**, A, is a rectangular array of scalars with m rows and n columns. The scalar in the i^{th} row and j^{th} column is the $(i, j)^{th}$ **entry** and is denoted a_{ij} or $(A)_{ij}$. That is,

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

Definition 3.4.2
Coefficient Matrix, Augmented Matrix

For a given system of linear equations,

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

the **coefficient matrix**, A, of the system is the matrix

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

The entry a_{ij} is the coefficient of the variable x_j in the i^{th} equation.

The **augmented matrix**, $[A|\vec{b}]$, of the system is

$$[A|\vec{b}] = \left[\begin{array}{cccc|c} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{array}\right],$$

where $\vec{b}$ is a column whose entries are the constant terms on the right-hand side of the equations.

The augmented matrix is the coefficient matrix, A, with the extra column, $\vec{b}$, added.

In solving a system of equations, we can manipulate the rows of the augmented matrix, which will correspond to manipulating the equations.

Definition 3.4.3
Elementary Row Operation (ERO)

Elementary row operations (EROs) are the operations performed on the coefficient and/or augmented matrix which correspond to the elementary operations performed on the system of equations.

Elementary operation	Equation	Row
Row swap	$e_i \leftrightarrow e_j$	$R_i \leftrightarrow R_j$
Row scale	$e_i \to ce_i, c \neq 0$	$R_i \to cR_i, c \neq 0$
Row addition	$e_i \to ce_j + e_i, i \neq j$	$R_i \to cR_j + R_i, i \neq j$

As with the equation operations, these EROs are referred to in some sources as **Type I**, **Type II**, and **Type III** operations, respectively.

Definition 3.4.4
Zero Row

In a matrix, we refer to a row that has all zero entries as a **zero row**.

Definition 3.4.5
Row Equivalent

If a matrix B is obtained from a matrix A by a finite number of EROs, then we say that B is **row equivalent to** A.

At any point in the process of manipulating the rows of an augmented matrix, we could stop and immediately write down the corresponding system of linear equations. The first column in the matrix corresponds to the variable x_1, the second column in the matrix corresponds to the variable x_2 and so on up to the second-last column in the matrix which corresponds to the variable x_n. The last column in the matrix corresponds to the constants on the right hand side of the system.

The augmented matrix at any stage is row equivalent to the original augmented matrix. Therefore, the system of equations corresponding to this augmented matrix is equivalent to the system of equations corresponding to the original augmented matrix.

Example 3.4.6 Solve the following system (the same system as in Example 3.3.1) using augmented matrices and EROs.

$$\begin{array}{rrrrcr} x_1 & +2x_2 & & & = & 1 \\ x_1 & +2x_2 & +3x_3 & +x_4 & = & 0 \\ -x_1 & -x_2 & +x_3 & +x_4 & = & -2 \\ & x_2 & +x_3 & +x_4 & = & -1 \\ & -x_2 & +2x_3 & & = & 0 \end{array}.$$

Solution: The augmented matrix for this system is

$$\left[\begin{array}{rrrr|r} 1 & 2 & 0 & 0 & 1 \\ 1 & 2 & 3 & 1 & 0 \\ -1 & -1 & 1 & 1 & -2 \\ 0 & 1 & 1 & 1 & -1 \\ 0 & -1 & 2 & 0 & 0 \end{array}\right].$$

We begin by eliminating x_1 from the second and third equations by performing the following EROs

$$R_2 \to -R_1 + R_2$$
$$R_3 \to R_1 + R_3$$

to get

$$\left[\begin{array}{rrrr|r} 1 & 2 & 0 & 0 & 1 \\ 0 & 0 & 3 & 1 & -1 \\ 0 & 1 & 1 & 1 & -1 \\ 0 & 1 & 1 & 1 & -1 \\ 0 & -1 & 2 & 0 & 0 \end{array}\right].$$

The second row does not contain the variable x_2, but the third one does and so we swap these two rows, $R_2 \leftrightarrow R_3$, to get

$$\left[\begin{array}{rrrr|r} 1 & 2 & 0 & 0 & 1 \\ 0 & 1 & 1 & 1 & -1 \\ 0 & 0 & 3 & 1 & -1 \\ 0 & 1 & 1 & 1 & -1 \\ 0 & -1 & 2 & 0 & 0 \end{array}\right].$$

We now make use of the new second row to eliminate x_2 from all rows *after* the second one, by performing the following EROs

$$R_4 \to -R_2 + R_4$$
$$R_5 \to R_2 + R_5$$

to get

$$\left[\begin{array}{rrrr|r} 1 & 2 & 0 & 0 & 1 \\ 0 & 1 & 1 & 1 & -1 \\ 0 & 0 & 3 & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 3 & 1 & -1 \end{array}\right].$$

We use the third row to eliminate x_3 from the rows *after* it, by performing the following ERO

$$R_5 \to -R_3 + R_5$$

to get

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

At this point, we have completed the forward elimination phase using a sequence of EROs equivalent to how we manipulated the equations in Example 3.3.1.

We can write down the system of equations that corresponds to the final augmented matrix that we obtained. This system of equations is equivalent to the original system.

$$\begin{array}{cccccc} x_1 & +2x_2 & & & = & 1 \\ & x_2 & +x_3 & +x_4 & = & -1 \\ & & 3x_3 & +x_4 & = & -1 \\ & & & 0 & = & 0 \\ & & & 0 & = & 0 \end{array}$$

This system is the same system we obtained after the forward elimination phase when we were using elementary operations. As before we can use back substitution to obtain the solution set

$$S = \left\{ \begin{bmatrix} \frac{7}{3} + \frac{4}{3}t \\ -\frac{2}{3} - \frac{2}{3}t \\ -\frac{1}{3} - \frac{1}{3}t \\ t \end{bmatrix} : t \in \mathbb{R} \right\} = \left\{ \begin{bmatrix} \frac{7}{3} \\ -\frac{2}{3} \\ -\frac{1}{3} \\ 0 \end{bmatrix} + t \begin{bmatrix} \frac{4}{3} \\ -\frac{2}{3} \\ -\frac{1}{3} \\ 1 \end{bmatrix} : t \in \mathbb{R} \right\}.$$

Instead of using back substitution we can continue using EROs to obtain an even simpler system.

We scale the third row by performing the following ERO

$$R_3 \to \frac{1}{3} R_3$$

to get

$$\left[\begin{array}{cccc|c} 1 & 2 & 0 & 0 & -1 \\ 0 & 1 & 1 & 1 & -1 \\ 0 & 0 & 1 & \frac{1}{3} & -\frac{1}{3} \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right].$$

Next, we use the third row to eliminate x_3 from the equations above it. We do this by performing the following ERO

$$R_2 \to -R_3 + R_2$$

yielding

$$\left[\begin{array}{cccc|c} 1 & 2 & 0 & 0 & -1 \\ 0 & 1 & 0 & \frac{2}{3} & -\frac{2}{3} \\ 0 & 0 & 1 & \frac{1}{3} & -\frac{1}{3} \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right].$$

The last step of the backward elimination phase is to eliminate x_2 from the first equation, which is achieved by the ERO

$$R_1 \to -2R_2 + R_1$$

yielding

$$\left[\begin{array}{cccc|c} 1 & 0 & 0 & -\frac{4}{3} & \frac{7}{3} \\ 0 & 1 & 0 & \frac{2}{3} & -\frac{2}{3} \\ 0 & 0 & 1 & \frac{1}{3} & -\frac{1}{3} \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right].$$

This augmented matrix has a very simple structure. We could argue that it has the simplest structure of all the augmented matrices we have seen in this example. In the next section we will introduce terminology to formalize this structure.

The system of equations corresponding to the last augmented matrix above is

$$\begin{array}{ccccccc} x_1 & & & -\frac{4}{3}x_4 & = & \frac{7}{3} \\ & x_2 & & +\frac{2}{3}x_4 & = & -\frac{2}{3} \\ & & x_3 & +\frac{1}{3}x_4 & = & -\frac{1}{3} \\ & & & 0 & = & 0 \\ & & & 0 & = & 0 \end{array}$$

This system is the same system we obtained after the forward and backward elimination phases when we were using elementary operations. Therefore, the solution set is

$$S = \left\{ \begin{bmatrix} \frac{7}{3} + \frac{4}{3}t \\ -\frac{2}{3} - \frac{2}{3}t \\ -\frac{1}{3} - \frac{1}{3}t \\ t \end{bmatrix} : t \in \mathbb{R} \right\} = \left\{ \begin{bmatrix} \frac{7}{3} \\ -\frac{2}{3} \\ -\frac{1}{3} \\ 0 \end{bmatrix} + t \begin{bmatrix} \frac{4}{3} \\ -\frac{2}{3} \\ -\frac{1}{3} \\ 1 \end{bmatrix} : t \in \mathbb{R} \right\}.$$

REMARK

In an augmented matrix, a zero row corresponds to the equation $0 = 0$. In the above example, the last two rows of the augmented matrix eventually became zero rows. Although these equations do not give us any information about the solution to the system of equations, we continued to write them down so that we would not forget that our system started with five equations.

If at any point we have a zero row in the coefficient matrix and the last entry in the corresponding row in the augmented matrix is $b \neq 0$, then we can stop and deduce that our system is inconsistent, since one of the equations has become $0 = b$, where $b \neq 0$.

3.5 The Gauss–Jordan Algorithm for Solving Systems of Linear Equations

Consider a system of m linear equations in n variables (unknowns). We assume the variables are labelled $x_1, x_2, \ldots, x_n$. The Gauss–Jordan algorithm is the formalization of the strategy we used in the previous section to solve such a system. In this algorithm, we use EROs to manipulate the augmented matrix into a simpler form from which it is easier to determine the solution of the system. The process of performing EROs on the matrix to bring it into these simpler forms is called **row reduction** or **Gaussian elimination** after Carl Friedrich Gauss (1777-1855) who outlined the process; a similar method for solving systems of linear equations was known to the Chinese around 250 B.C.

In the previous section, we obtained two simpler forms of the augmented matrix. The first was obtained after completing the forward elimination process and is called the **row echelon form**. The second is obtained after completing both the forward and backward elimination processes and is called the **reduced row echelon form**. We now give the formal definitions of these forms in terms of **leading entries** and **pivots**.

Definition 3.5.1
Leading Entry, Leading One

The leftmost non-zero entry in any non-zero row of a matrix is called the **leading entry** of that row. If the leading entry is a 1, then it is called a **leading one**.

Definition 3.5.2
Row Echelon Form

We say that a matrix is in **row echelon form** (REF) whenever both of the following two conditions are satisfied:

1. All zero rows occur as the final rows in the matrix.
2. The leading entry in any non-zero row appears in a column to the right of the columns containing the leading entries of any of the rows above it.

We say that the matrix R is a **row echelon form** of matrix A to mean that R is in row echelon form and that R can be obtained from A by performing a finite number of EROs to A.

Definition 3.5.3
Pivot, Pivot Position, Pivot Column, Pivot Row

If a matrix is in REF, then the leading entries are referred to as **pivots** and their positions in the matrix are called **pivot positions**. Any column that contains a pivot position is called a **pivot column**. Any row that contains a pivot position is called a **pivot row**.

Due to the structure of REF, any pivot column will contain exactly one pivot and any pivot row will contain exactly one pivot.

Example 3.5.4

The matrix
$$\begin{bmatrix} 3 & 2 & -3 & 4 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & -5 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

is in REF, with pivots in the $(1,1)$, $(2,3)$ and $(3,4)$ entries. However, the matrix

$$\begin{bmatrix} -3 & 3 & 5 & 7 \\ 0 & 2 & 6 & 0 \\ 1 & 0 & 1 & -2 \end{bmatrix}$$

is not in REF because the leading entry in the third row (marked in red) is not to the right of the leading entry of one (in fact both of) the rows above it.

Definition 3.5.5

Reduced Row Echelon Form

We say that a matrix is in **reduced row echelon form** (RREF) whenever all of the following three conditions are satisfied:

1. It is in REF.
2. All its pivots are leading ones.
3. The only non-zero entry in a pivot column is the pivot itself.

REMARK

Item "3." from the definition of RREF above implies that all entries above and below a pivot MUST be zero for a matrix in RREF; in other words, any column of a matrix in RREF with two or more non-zero entries is not a pivot column.

Example 3.5.6

The matrix

$$\begin{bmatrix} 3 & 2 & -3 & 4 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & -5 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

from the previous example is in REF, but not in RREF, because its pivots are not leading ones, and moreover, the third and fourth pivot columns contain non-zero entries that are not pivots.

The matrix

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

is in RREF.

Example 3.5.7

List all possible 2×2 matrices that are in RREF.

Solution: Since a 2×2 matrix has two rows and two columns, then the number of pivots is either zero, one or two. If there are zero pivots, then all entries in the matrix must be 0. Therefore, we obtain the matrix

$$\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

If there is one pivot, then this pivot must occur in the first row and the second row must be a zero row. The pivot could occur in the first column or the second column. If the pivot occurs in the first column, then the $(1, 2)$ entry could be anything. Therefore, we obtain an infinite number of matrices of the form

$$\begin{bmatrix} 1 & a \\ 0 & 0 \end{bmatrix}, \ a \in \mathbb{F}.$$

If the pivot occurs in the second column, then the $(1, 1)$ entry must be 0. Therefore, we obtain the matrix

$$\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}.$$

If there are two pivots, then they must occur in the positions $(1, 1)$ and $(2, 2)$ and the other entries must be 0. Therefore, we obtain the matrix

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

If A is not a matrix of all zeros, then there are an infinite number of matrices which are REFs of A. However, the pivot positions of A are unique, as well as is its RREF. At this point, we do not have the tools to prove this result.

Theorem 3.5.8 **(Unique RREF)**

Let A be a matrix with REFs R_1 and R_2. Then R_1 and R_2 will have the same set of pivot positions. Moreover, there is a *unique* matrix R such that R is the RREF of A.

Because the RREF of a given matrix A is unique, we can introduce notation to refer to it.

Definition 3.5.9
RREF(A)

We say that the matrix R is the **reduced row echelon form of matrix** A, and we write $R = \text{RREF}(A)$, if R is in reduced row echelon form and if R can be obtained from A by performing a finite number of EROs to A.

Having laid down the needed definitions, we now outline an algorithm that takes as input a given matrix A and produces a row equivalent matrix R in REF. This algorithm corresponds to the forward elimination process for a consistent system.

ALGORITHM (Obtaining an REF)

1. Consider the leftmost non-zero column of the matrix. Use EROs to obtain a leading entry in the top position of this column. This entry is now a pivot and this row is now a pivot row.

2. Use EROs to change all other entries below the pivot in this pivot column to 0.

3. Consider the submatrix formed by covering up the current pivot row and all previous pivot rows. If there are no more rows or if the only remaining rows are zero rows, we are finished. Otherwise, repeat steps 1 and 2 on the submatrix. Continue in this manner, covering up the current pivot row to obtain a matrix with one less row until no rows remain or we obtain a submatrix with only zero rows.

We can also outline a similar algorithm that takes as input a matrix R in REF and produces the RREF of R. This will correspond to the backward elimination process.

ALGORITHM (Obtaining the RREF from an REF)

Start with a matrix in REF.

1. Select the rightmost pivot column. If the pivot is not already 1, use EROs to change it to 1.

2. Use EROs to change all entries above the pivot in this pivot column to 0.

3. Consider the submatrix formed by covering up the current pivot row and all other rows below it. If there are no more rows, then we are finished. Otherwise, repeat steps 1 and 2 on the submatrix until no rows remain.

The preceding two algorithms can be run in sequence to take a matrix A and produce RREF(A). This process is known as the Gauss–Jordan[1] algorithm.

REMARKS

- If the system is inconsistent, then at some point we will obtain a row of the form $\left[\begin{array}{ccc|c} 0 & \cdots & 0 & b \end{array}\right]$, where $b \neq 0$. This row corresponds to the equation $0 = b$, where $b \neq 0$. As soon as we obtain such a row, we can stop the algorithm and conclude that the system is inconsistent.

- We will not always blindly follow these algorithms when finding an REF or the RREF. There may be other choices along the way that lead to fewer calculations, or we may want to avoid fractions for as long as possible. When determining an REF we often try to obtain leading 1s, even though they are not required, because they are easier to work with.

[1]Named after the aforementioned Gauss and Wilhelm Jordan (1842-1899), a German engineer who popularized this method for solving systems of equations.

We will apply the Gauss–Jordan algorithm to the augmented matrix obtained from a system of linear equations. This will allow us to easily obtain the solution set to the system. The following terminology will be helpful in describing these solution sets.

Definition 3.5.10
Basic Variable, Free Variable

Consider a system of linear equations. Let R be an REF of the coefficient matrix of this system. If the i^{th} column of this matrix contains a pivot, then we call x_i a **basic variable**. Otherwise, we call x_i a **free variable**.

This terminology reflects the fact that we will be able to assign parameters to the free variables, and then we will be able to express the basic variables in terms of these parameters. See the examples below.

Example 3.5.11

Solve the following system.

$$\begin{array}{rrrrcr} -x_1 & -2x_2 & +2x_3 & +4x_4 & = & 3 \\ 2x_1 & +4x_2 & -2x_3 & -2x_4 & = & 4 \\ x_1 & +2x_2 & -x_3 & -2x_4 & = & 0 \\ 2x_1 & +4x_2 & -6x_3 & -8x_4 & = & -4 \end{array}$$

Solution: We begin by giving the augmented matrix for the system.

$$\left[\begin{array}{cccc|c} -1 & -2 & 2 & 4 & 3 \\ 2 & 4 & -2 & -2 & 4 \\ 1 & 2 & -1 & -2 & 0 \\ 2 & 4 & -6 & -8 & -4 \end{array}\right]$$

The leftmost non-zero column is the first column. Therefore, our first goal is to get a leading entry (which will become a pivot) in the top position of this column. It will make our calculations easier if this is a leading one, so we will perform the ERO $R_1 \longleftrightarrow R_3$ to get

$$\left[\begin{array}{cccc|c} 1 & 2 & -1 & -2 & 0 \\ 2 & 4 & -2 & -2 & 4 \\ -1 & -2 & 2 & 4 & 3 \\ 2 & 4 & -6 & -8 & -4 \end{array}\right].$$

Next, we want to change all entries below the pivot in this column to 0. So we perform the following EROs

$$\begin{aligned} &R_2 \to -2R_1 + R_2 \\ &R_3 \to R_1 + R_3 \\ &R_4 \to -2R_1 + R_4 \end{aligned}$$

to get

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

Covering up the first row, the leftmost non-zero column is the third column. We need to get a leading entry (which will become a pivot) in the top position of this column (with the first row covered) and so we need to get a pivot in the (2,3) position. The simplest way to do this is to perform the ERO $R_2 \longleftrightarrow R_3$ to get

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

Next, we want to change all entries below the pivot in this column to 0. So we perform the following ERO

$$R_4 \to 4R_2 + R_4$$

to get

$$\left[\begin{array}{cccc|c} 1 & 2 & -1 & -2 & 0 \\ 0 & 0 & 1 & 2 & 3 \\ 0 & 0 & 0 & 2 & 4 \\ 0 & 0 & 0 & 4 & 8 \end{array}\right].$$

Covering up the first two rows, the leftmost non-zero column is the fourth column. We need to get a leading entry (which will become a pivot) in the top position of this column (with the first two rows covered) and so we need to get a pivot in the (3,4) position. There already is a leading entry there and so no EROs are needed.

Next, we want to change all entries below the pivot in this column to 0. So we perform the following ERO

$$R_4 \to -2R_3 + R_4$$

to get

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

Covering up the first three rows, all remaining rows are zero rows and so we now have an augmented matrix in REF. We write down the system of equations that corresponds to this augmented matrix and solve this system using back substitution.

The corresponding system of equations is

$$\begin{array}{ccccccc} x_1 & +2x_2 & -x_3 & -2x_4 & = & 0 \\ & & x_3 & +2x_4 & = & 3 \\ & & & 2x_4 & = & 4 \\ & & & 0 & = & 0 \end{array}.$$

The third equation tells us that $x_4 = 2$. Substituting this into the second equation, we obtain that $x_3 = -1$. Substituting both of these values into the first equation, we obtain that $x_1 + 2x_2 = 3$. Using our convention, we set x_1 to be a basic variable and x_2 to be a free variable. Therefore, $x_2 = t, t \in \mathbb{R}$ and $x_1 = 3 - 2t$.

Therefore, the solution set is $S = \left\{ \begin{bmatrix} 3-2t \\ t \\ -1 \\ 2 \end{bmatrix} : t \in \mathbb{R} \right\} = \left\{ \begin{bmatrix} 3 \\ 0 \\ -1 \\ 2 \end{bmatrix} + t \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \end{bmatrix} : t \in \mathbb{R} \right\}.$

Instead of stopping at REF and using back substitution, we could continue to do row reduction until we reached an augmented matrix in RREF (the Gauss-Jordan algorithm).

Beginning with

$$\left[\begin{array}{cccc|c} 1 & 2 & -1 & -2 & 0 \\ 0 & 0 & 1 & 2 & 3 \\ 0 & 0 & 0 & 2 & 4 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

we identify the rightmost pivot column, which is the fourth column. We change the pivot to 1 by performing the ERO $R_3 \to \frac{1}{2}R_3$ to get

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

Next, we want to change all entries above this pivot to 0. So we perform the EROs

$$\begin{aligned} R_1 &\to 2R_3 + R_1 \\ R_2 &\to -2R_3 + R_2 \end{aligned}$$

to get

$$\left[\begin{array}{cccc|c} 1 & 2 & -1 & 0 & 4 \\ 0 & 0 & 1 & 0 & -1 \\ 0 & 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right].$$

Covering up the bottom two rows, the rightmost pivot column is the third column. The pivot in this column is already 1 so no ERO is necessary. We want to change all entries above this pivot to 0 and so we perform the ERO

$$R_1 \to R_2 + R_1$$

to get

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

Covering up the bottom three rows, the rightmost pivot column is the first column. The pivot in this column is already 1 so no ERO is necessary.

There are now no more rows to adjust and so our augmented matrix is in RREF.

The corresponding system of equations is

$$\begin{array}{cccccc} x_1 & +2x_2 & & & = & 3 \\ & & x_3 & & = & -1 \\ & & & x_4 & = & 2 \\ & & & 0 & = & 0 \end{array}$$

The second and third equations give us values for x_3 and x_4, respectively. We set x_2 to be a free variable and thus $x_2 = t, t \in \mathbb{R}$, which implies that $x_1 = 3 - 2t$. Therefore, the solution set is $S = \left\{ \begin{bmatrix} 3-2t \\ t \\ -1 \\ 2 \end{bmatrix} : t \in \mathbb{R} \right\} = \left\{ \begin{bmatrix} 3 \\ 0 \\ -1 \\ 2 \end{bmatrix} + t \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \end{bmatrix} : t \in \mathbb{R} \right\}$ which is the same solution set we obtained earlier.

Example 3.5.12 Solve the following system.

$$\begin{array}{rrrcr} 2x_1 & +4x_2 & -2x_3 & = & 2 \\ 3x_1 & +7x_2 & +x_3 & = & 0 \\ 2x_1 & +5x_2 & +2x_3 & = & 1 \end{array}$$

Solution: We begin by giving the augmented matrix for the system.

$$\left[\begin{array}{rrr|r} 2 & 4 & -2 & 2 \\ 3 & 7 & 1 & 0 \\ 2 & 5 & 2 & 1 \end{array}\right]$$

Our first goal is to get a leading entry (which will become a pivot) in the top position of the first column, since is it the leftmost non-zero column. It will actually make our calculations easier if this is a leading one and so we will perform the ERO $R_1 \to \frac{1}{2}R_1$ to obtain

$$\left[\begin{array}{rrr|r} 1 & 2 & -1 & 1 \\ 3 & 7 & 1 & 0 \\ 2 & 5 & 2 & 1 \end{array}\right].$$

Next, we want to change all entries below the pivot in this column to 0. So we perform the following row addition EROs

$$R_2 \to -3R_1 + R_2$$
$$R_3 \to -2R_1 + R_3$$

to obtain

$$\left[\begin{array}{rrr|r} 1 & 2 & -1 & 1 \\ 0 & 1 & 4 & -3 \\ 0 & 1 & 4 & -1 \end{array}\right].$$

Covering up the first row, the leftmost non-zero column is the second column. We need to obtain a leading entry (which will become a pivot) in the top position of this column (with the first row covered) and so we need to get a pivot in the (2,2) position. There already is a pivot there and so no ERO is necessary. Next, we want to change all entries below this pivot to 0 and so we perform the ERO $R_3 \to -R_2 + R_3$ to obtain

$$\left[\begin{array}{rrr|r} 1 & 2 & -1 & 1 \\ 0 & 1 & 4 & -3 \\ 0 & 0 & 0 & 2 \end{array}\right].$$

The third row is of the form $\left[\begin{array}{ccc|c} 0 & 0 & 0 & b \end{array}\right]$, where $b \neq 0$ and so we can stop the algorithm here and declare that this system is inconsistent. (The third row corresponds to the equation $0 = 2$.)

We will include one more example in which we give far less commentary. We will simply outline the EROs that we are using. We will make use of the Gauss–Jordan algorithm (forward and backward elimination).

Example 3.5.13 Solve the following system.

$$\begin{array}{rrrcr} 3x_1 & +5x_2 & +3x_3 & = & 9 \\ -2x_1 & -x_2 & +6x_3 & = & 10 \\ 4x_1 & +10x_2 & -3x_3 & = & -2 \end{array}$$

Solution: The augmented matrix for the system is

$$\left[\begin{array}{ccc|c} 3 & 5 & 3 & 9 \\ -2 & -1 & 6 & 10 \\ 4 & 10 & -3 & -2 \end{array}\right].$$

We determine an REF.

$$R_1 \to R_2 + R_1 \text{ gives } \left[\begin{array}{ccc|c} 1 & 4 & 9 & 19 \\ -2 & -1 & 6 & 10 \\ 4 & 10 & -3 & -2 \end{array}\right].$$

$$\begin{cases} R_2 \to 2R_1 + R_2 \\ R_3 \to -4R_1 + R_3 \end{cases} \text{gives } \left[\begin{array}{ccc|c} 1 & 4 & 9 & 19 \\ 0 & 7 & 24 & 48 \\ 0 & -6 & -39 & -78 \end{array}\right].$$

$$R_2 \to R_3 + R_2 \text{ gives } \left[\begin{array}{ccc|c} 1 & 4 & 9 & 19 \\ 0 & 1 & -15 & -30 \\ 0 & -6 & -39 & -78 \end{array}\right].$$

$$R_3 \to 6R_2 + R_3 \text{ gives } \left[\begin{array}{ccc|c} 1 & 4 & 9 & 19 \\ 0 & 1 & -15 & -30 \\ 0 & 0 & -129 & -258 \end{array}\right].$$

This augmented matrix is in REF. At this point we could use back substitution, but we will continue on to RREF.

$$R_3 \to -\tfrac{1}{129}R_3 \text{ gives } \left[\begin{array}{ccc|c} 1 & 4 & 9 & 19 \\ 0 & 1 & -15 & -30 \\ 0 & 0 & 1 & 2 \end{array}\right].$$

$$\begin{cases} R_1 \to -9R_3 + R_1 \\ R_2 \to 15R_3 + R_2 \end{cases} \text{gives } \left[\begin{array}{ccc|c} 1 & 4 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 2 \end{array}\right].$$

$$R_1 \to -4R_2 + R_1 \text{ gives } \left[\begin{array}{ccc|c} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 2 \end{array}\right].$$

Given this augmented matrix we can see that the solution is $x_1 = 1$, $x_2 = 0$ and $x_3 = 2$. Therefore, the solution set is $S = \left\{ \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} \right\}$.

3.6 Rank and Nullity

Definition 3.6.1

$M_{m\times n}(\mathbb{R})$, $M_{m\times n}(\mathbb{C})$, $M_{m\times n}(\mathbb{F})$

We use $M_{m\times n}(\mathbb{R})$ to denote to the set of all $m \times n$ matrices with real entries.

We use $M_{m\times n}(\mathbb{C})$ to denote to the set of all $m \times n$ matrices with complex entries.

If we do not need to distinguish between real and complex entries we will use $M_{m\times n}(\mathbb{F})$.

REMARK

If $m = n$ it is common to abbreviate $M_{m\times n}(\mathbb{F})$ as $M_n(\mathbb{F})$.

Let us consider a system of m linear equations in n variables. The coefficient matrix of this system, A, belongs to $M_{m\times n}(\mathbb{F})$ and the augmented matrix, $\left[\, A \mid \vec{b} \,\right]$, belongs to $M_{m\times(n+1)}(\mathbb{F})$. In other words, the augmented matrix has exactly one more column than the coefficient matrix.

There are three important questions that we would like to consider about this system.

1. Is the system consistent or inconsistent?
2. If the system is consistent, does it have a unique solution?
3. If the system is consistent and does not have a unique solution, how many parameters are there in the solution set?

We will be able to answer these questions once we have an REF or the RREF of the augmented matrix $\left[\, A \mid \vec{b} \,\right]$ even before we have found the solution set.

Definition 3.6.2

Rank

Let $A \in M_{m\times n}(\mathbb{F})$ such that $\text{RREF}(A)$ has exactly r pivots. Then we say that the rank of A is r, and we write $\text{rank}(A) = r$.

Although there are infinitely many REFs of a matrix A, Theorem 3.5.8 (Unique RREF) tells us that they will all have the same pivot positions, hence the same number of pivots. Since $\text{RREF}(A)$ is an REF of A, all REFs of A will have the same number of pivots as $\text{RREF}(A)$. Therefore, we can use any REF of A to determine its rank.

Proposition 3.6.3

(Rank Bounds)

If $A \in M_{m\times n}(\mathbb{F})$, then $\text{rank}(A) \leq \min\{m, n\}$.

Proof: Since there is at most one pivot in each row, then $\text{rank}(A) \leq m$. Also, since there is at most one pivot in each column, then $\text{rank}(A) \leq n$. $\square$

Comparing the rank of the coefficient matrix of a system of linear equations to the rank of its augmented matrix will tell us whether the system is consistent or inconsistent.

Proposition 3.6.4 **(Consistent System Test)**

Let A be the coefficient matrix of a system of linear equations and let $\left[\,A\,|\,\vec{b}\,\right]$ be the augmented matrix of the system. The system is consistent if and only if $\text{rank}(A) = \text{rank}\left(\left[\,A\,|\,\vec{b}\,\right]\right)$.

Proof: Suppose that we perform a sequence of EROs on $\left[\,A\,|\,\vec{b}\,\right]$ to manipulate it into its RREF, which we will call $\left[\,R\,|\,\vec{c}\,\right]$. If we perform the same sequence of EROs on A, we will obtain the matrix R, which is the RREF of A. The only difference between R and $\left[\,R\,|\,\vec{c}\,\right]$ is that $\left[\,R\,|\,\vec{c}\,\right]$ has an additional column. Therefore, $\text{rank}(A) \le \text{rank}\left(\left[\,A\,|\,\vec{b}\,\right]\right)$.

Assume that the system is inconsistent. Then $\left[\,R\,|\,\vec{c}\,\right]$ contains a row of the form $[0\cdots 0|1]$ and so $\left[\,R\,|\,\vec{c}\,\right]$ has a pivot in its final column. Therefore, $\text{rank}(A) < \text{rank}\left(\left[\,A\,|\,\vec{b}\,\right]\right)$.

Assume that the system is consistent. Then $\left[\,R\,|\,\vec{c}\,\right]$ does not contain a row of the form $[0\cdots 0|1]$ and so R and $[R|\vec{c}]$ have the same pivots. Therefore, $\text{rank}(A) = \text{rank}\left(\left[\,A\,|\,\vec{b}\,\right]\right)$.

In summary, the system is consistent if and only if $\text{rank}(A) = \text{rank}\left(\left[\,A\,|\,\vec{b}\,\right]\right)$. $\square$

Example 3.6.5 Suppose we have a system of linear equations and we perform EROs on the augmented matrix to obtain the following REF.

$$\left[\begin{array}{cccc|c} 2 & -5 & 5 & -7 & 4 \\ 0 & 0 & 4 & -9 & 3 \\ 0 & 0 & 0 & 5 & 2 \\ 0 & 0 & 0 & 0 & 1 \end{array}\right]$$

Determine whether the system is consistent or inconsistent.

Solution:

The rank of the coefficient matrix is 3 and the rank of the augmented matrix is 4. Therefore, the system is inconsistent. The fact that the system is inconsistent is also evident from the fact that the final row of the augmented matrix corresponds to the equation $0 = 1$.

Example 3.6.6 Suppose we have a system of linear equations and we perform EROs on the augmented matrix to obtain the following REF.

$$\left[\begin{array}{cccc|c} 2 & -5 & 5 & -7 & 4 \\ 0 & 0 & 4 & -9 & 3 \\ 0 & 0 & 0 & 5 & 2 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

Determine whether the system is consistent or inconsistent.

Solution: The ranks of the coefficient matrix and the augmented matrix are both 3. Therefore, the system is consistent.

Comparing the ranks of A and $\left[A \mid \vec{b}\right]$ gives us a quick way of determining whether or not a system of linear equations is consistent.

If the system is consistent, then we usually want to know how many parameters, if any, appear in the solution set. We can also answer this question using $\operatorname{rank}(A)$, as our next result shows.

Theorem 3.6.7 **(System Rank Theorem)**

Let $A \in M_{m\times n}(\mathbb{F})$ with $\operatorname{rank}(A) = r$.

(a) Let $\vec{b} \in \mathbb{F}^m$. If the system of linear equations with augmented matrix $[A \mid \vec{b}]$ is consistent, then the solution set to this system will contain $n - r$ parameters.

(b) The system with augmented matrix $\left[A \mid \vec{b}\right]$ is consistent for every $\vec{b} \in \mathbb{F}^m$ if and only if $r = m$.

Proof: (a) Since A has n columns, the system of equations has n variables. Since $\operatorname{rank}(A) = r$, $\operatorname{RREF}(A)$ will have r pivot columns. Therefore, the system will have r basic variables and $n-r$ free variables. In determining the solution set, we assign each free variable a parameter. Therefore, the solution set will contain $n - r$ parameters.

(b) We prove the contrapositive which says that the system with augmented matrix $\left[A \mid \vec{b}\right]$ is inconsistent for some $\vec{b} \in \mathbb{F}^m$ if and only if $r \neq m$.

We begin with the forward direction and assume that the system with augmented matrix $\left[A \mid \vec{b}\right]$ is inconsistent for some $\vec{b} \in \mathbb{F}^m$. Then, as in the proof of Proposition 3.6.4 (Consistent System Test), $\operatorname{RREF}\left(\left[A \mid \vec{b}\right]\right)$ will include a row of the form $[0 \cdots 0|1]$. Therefore, $\operatorname{RREF}(A)$ will include a row of all zeros. Since $\operatorname{RREF}(A)$ has m rows, it follows that $\operatorname{rank}(A) < m$.

For the backward direction we assume that $r \neq m$. By Proposition 3.6.3 (Rank Bounds), we get that $r < m$. If we let $R = \operatorname{RREF}(A)$, then R must include a row of all zeros. Consider the augmented matrix $\left[R \mid \vec{e}_m\right]$, where $\vec{e}_m$ is the vector from the standard basis of $\mathbb{F}^m$ whose m^{th} component is 1 with all other components 0. It corresponds to an inconsistent system. Since all EROs are reversible, we can apply to $\left[R \mid \vec{e}_m\right]$ the reverse of the EROs needed to row reduce A to R. Applying these reverse EROs will result in an augmented matrix of the form $\left[A \mid \vec{b}\right]$, for some $\vec{b} \in \mathbb{F}^m$. The matrix is row equivalent to $\left[R \mid \vec{e}_m\right]$. Therefore, $\left[A \mid \vec{b}\right]$ is the augmented matrix of an inconsistent system.

□

Definition 3.6.8
Nullity

Let $A \in M_{m\times n}(\mathbb{F})$ with $\operatorname{rank}(A) = r$. We define the **nullity** of A, written $\operatorname{nullity}(A)$, to be the integer $n - r$.

Thus Theorem 3.6.7 (System Rank Theorem) shows that, if $\left[\, A \mid \vec{b}\, \right]$ is the augmented matrix of a consistent system, the number of parameters in the solution set of this system is given by nullity(A).

Example 3.6.9 Returning to Example 3.6.6, we have an augmented matrix with the following REF:

$$\left[\begin{array}{cccc|c} 2 & -5 & 5 & -7 & 4 \\ 0 & 0 & 4 & -9 & 3 \\ 0 & 0 & 0 & 5 & 2 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right].$$

How many parameters will there be in the solution set of the corresponding system?

Solution:

We already determined that this system was consistent. The system has four variables and three pivots. Using Theorem 3.6.7 (System Rank Theorem), the number of parameters in the solution set to this system is $4 - 3 = 1$ (i.e. the nullity is 1). The variable x_2 is a free variable because the second column is not a pivot column. All other variables are basic.

3.7 Homogeneous and Non-Homogeneous Systems, Nullspace

We examine a series of examples that will help us state general facts about solving systems of linear equations.

Example 3.7.1 Solve the following system of linear equations:

$$\begin{array}{rrrrcl} x_1 & -2x_2 & -x_3 & +3x_4 & = & 1 \\ 2x_1 & -4x_2 & +x_3 & & = & 5 \\ x_1 & -2x_2 & +2x_3 & -3x_4 & = & 4 \end{array}.$$

Solution:

The augmented matrix $\left[\, A \mid \vec{b}\, \right]$ for the system is

$$\left[\begin{array}{cccc|c} 1 & -2 & -1 & 3 & 1 \\ 2 & -4 & 1 & 0 & 5 \\ 1 & -2 & 2 & -3 & 4 \end{array}\right].$$

We perform the following EROs.

$$\begin{cases} R_2 \to -2R_1 + R_2 \\ R_3 \to -R_1 + R_3 \end{cases} \text{gives} \left[\begin{array}{cccc|c} 1 & -2 & -1 & 3 & 1 \\ 0 & 0 & 3 & -6 & 3 \\ 0 & 0 & 3 & -6 & 3 \end{array}\right].$$

$$R_2 \to \tfrac{1}{3}R_2 \text{ gives } \left[\begin{array}{cccc|c} 1 & -2 & -1 & 3 & 1 \\ 0 & 0 & 1 & -2 & 1 \\ 0 & 0 & 3 & -6 & 3 \end{array}\right].$$

$$R_3 \to -3R_2 + 3R_3 \text{ gives } \left[\begin{array}{cccc|c} 1 & -2 & -1 & 3 & 1 \\ 0 & 0 & 1 & -2 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right].$$

This matrix is in REF and since $\text{rank}(A) = \text{rank}\left(\left[\, A \mid \vec{b} \,\right]\right) = 2$, then we conclude that this system is consistent.

We perform the following ERO.

$$R_1 \to R_1 + R_2 \text{ gives } \left[\begin{array}{cccc|c} 1 & -2 & 0 & 1 & 2 \\ 0 & 0 & 1 & -2 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right].$$

This matrix is in RREF. Since there are four variables and $\text{rank}(A) = 2$, the number of parameters is $4 - 2 = 2$. We let the free variables x_2 and x_4 be s and t respectively, where $s, t \in \mathbb{R}$.

Solving for the basic variables, we get $x_1 = 2+2s-t$ and $x_3 = 1+2t$, for $s, t \in \mathbb{R}$. Therefore, the solution set is

$$S = \left\{ \begin{bmatrix} 2+2s-t \\ s \\ 1+2t \\ t \end{bmatrix} : \ s, t \in \mathbb{R} \right\} = \left\{ \begin{bmatrix} 2 \\ 0 \\ 1 \\ 0 \end{bmatrix} + s \begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \end{bmatrix} + t \begin{bmatrix} -1 \\ 0 \\ 2 \\ 1 \end{bmatrix} : \ s, t \in \mathbb{R} \right\}.$$

Example 3.7.2 Solve the following system of linear equations:

$$\begin{array}{rrrrcl} x_1 & -2x_2 & -x_3 & +3x_4 & = & 1 \\ 2x_1 & -4x_2 & +x_3 & & = & 2 \\ x_1 & -2x_2 & +2x_3 & -3x_4 & = & 3 \end{array}.$$

Solution:

The augmented matrix is:

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

The coefficient matrix for this system is exactly the same as it was for the previous system. So we perform the exact same sequence of EROs to obtain the following augmented matrix. (Only the constant terms will be different.)

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

This matrix is in REF and since $\text{rank}(A) = 2$ is not equal to $\text{rank}\left(\left[\, A \mid \vec{b} \,\right]\right) = 3$, we conclude that this system is inconsistent. (We could also notice that the third row corresponds to the equation $0 = 2$.)

The solution set is $T = \varnothing$.

Example 3.7.3 Solve the following system of linear equations:

$$\begin{array}{rrrrcr} x_1 & -2x_2 & -x_3 & +3x_4 & = & -1 \\ 2x_1 & -4x_2 & +x_3 & & = & 4 \\ x_1 & -2x_2 & +2x_3 & -3x_4 & = & 5 \end{array}.$$

Solution:

Again we observe that we have the same coefficient matrix and thus, we will perform the same sequence of EROs as in the two previous examples to obtain the following augmented matrix.

$$\left[\begin{array}{rrrr|r} 1 & -2 & -1 & 3 & -1 \\ 0 & 0 & 1 & -2 & 2 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right].$$

This matrix is in REF, and since $\operatorname{rank}(A) = \operatorname{rank}\left(\left[\, A \mid \vec{b}\,\right]\right) = 2$, we conclude that this system is consistent. This time the RREF will be

$$\left[\begin{array}{rrrr|r} 1 & -2 & 0 & 1 & 1 \\ 0 & 0 & 1 & -2 & 2 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right].$$

Since there are four variables and $\operatorname{rank}(A) = 2$, then the number of parameters is $4 - 2 = 2$. We let the free variables x_2 and x_4 be s and t respectively, where $s, t \in \mathbb{R}$.

Solving for the basic variables, we get $x_1 = 1+2s-t$ and $x_3 = 2+2t$, for $s, t \in \mathbb{R}$. Therefore, the solution set is

$$U = \left\{ \begin{bmatrix} 1+2s-t \\ s \\ 2+2t \\ t \end{bmatrix} : \ s, t \in \mathbb{R} \right\} = \left\{ \begin{bmatrix} 1 \\ 0 \\ 2 \\ 0 \end{bmatrix} + s \begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \end{bmatrix} + t \begin{bmatrix} -1 \\ 0 \\ 2 \\ 1 \end{bmatrix} : \ s, t \in \mathbb{R} \right\}.$$

Example 3.7.4 Solve the following system of linear equations:

$$\begin{array}{rrrrcr} x_1 & -2x_2 & -x_3 & +3x_4 & = & 0 \\ 2x_1 & -4x_2 & +x_3 & & = & 0 \\ x_1 & -2x_2 & +2x_3 & -3x_4 & = & 0 \end{array}.$$

Solution:

Once again, we observe that we have the same coefficient matrix and thus we will be performing the same sequence of EROs as in the previous three examples. We also note that this system is guaranteed to be consistent since all of the constant terms are originally 0 and no ERO will change this fact. We can always set all of the variables to 0 to obtain a solution. The RREF of this system is

$$\left[\begin{array}{cccc|c} 1 & -2 & 0 & 1 & 0 \\ 0 & 0 & 1 & -2 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right].$$

As before we let the free variables x_2 and x_4 be s and t respectively, where $s, t \in \mathbb{R}$.

Solving for the basic variables, we get $x_1 = 2s - t$ and $x_3 = 2t$, for $s, t \in \mathbb{R}$. Therefore, the solution set is

$$V = \left\{ \begin{bmatrix} 2s-t \\ s \\ 2t \\ t \end{bmatrix} : \ s, t \in \mathbb{R} \right\} = \left\{ s \begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \end{bmatrix} + t \begin{bmatrix} -1 \\ 0 \\ 2 \\ 1 \end{bmatrix} : \ s, t \in \mathbb{R} \right\}.$$

Note that in this final example, the solution set, V, can be written as all linear combinations of the vectors $\begin{bmatrix} 2 & 1 & 0 & 0 \end{bmatrix}^T$ and $\begin{bmatrix} -1 & 0 & 2 & 1 \end{bmatrix}^T$. Therefore, we can write the solution set as the span of these two vectors:

$$V = \left\{ s \begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \end{bmatrix} + t \begin{bmatrix} -1 \\ 0 \\ 2 \\ 1 \end{bmatrix} : s, t \in \mathbb{R} \right\} = \operatorname{Span} \left\{ \begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 2 \\ 1 \end{bmatrix} \right\}.$$

There are several observations we can make by comparing these four systems of linear equations.

The *first* observation is that the last system is the simplest to solve because all the constant terms are zero. We give a name to such a system and its solution set.

Definition 3.7.5

Homogeneous and Non-homogeneous Systems

We say that a system of linear equations is **homogeneous** if all the constant terms on the right-hand side of the equations are zero. Otherwise we say the system is **non-homogeneous**.

Thus, if $\left[A \,|\, \vec{b} \right]$ is the augmented matrix of a system of linear equations, then the system is homogeneous if $\vec{b} = \vec{0}$, and non-homogeneous if $\vec{b} \neq \vec{0}$.

As we noted in Example 3.7.4, a homogeneous system will always be consistent. We can set all of the variables to 0 to obtain a solution. This solution is special in its consideration, and we formally define it below.

Definition 3.7.6

Trivial Solution

For a homogeneous system with variables $x_1, x_2, ..., x_n$, the **trivial solution** is the solution $x_1 = x_2 = \cdots = x_n = 0$.

Moreover, if we collect all the solutions of a homogeneous system into a single set, that set has a special relationship with the matrix A. We define this set formally below.

Definition 3.7.7
Nullspace

The solution set of a *homogeneous* system of linear equations with coefficient matrix A is called the **nullspace** of A and is denoted $\text{Null}(A)$.

Consider the solution sets of the three consistent systems in our previous examples.

$$S = \left\{ \begin{bmatrix} 2 \\ 0 \\ 1 \\ 0 \end{bmatrix} + s \begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \end{bmatrix} + t \begin{bmatrix} -1 \\ 0 \\ 2 \\ 1 \end{bmatrix} \ : \ s, t \in \mathbb{R} \right\}.$$

$$U = \left\{ \begin{bmatrix} 1 \\ 0 \\ 2 \\ 0 \end{bmatrix} + s \begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \end{bmatrix} + t \begin{bmatrix} -1 \\ 0 \\ 2 \\ 1 \end{bmatrix} \ : \ s, t \in \mathbb{R} \right\}.$$

$$V = \left\{ s \begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \end{bmatrix} + t \begin{bmatrix} -1 \\ 0 \\ 2 \\ 1 \end{bmatrix} : s, t \in \mathbb{R} \right\} = \text{Span} \left\{ \begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 2 \\ 1 \end{bmatrix} \right\}.$$

Note that $V = \text{Null}(A)$, where $A = \begin{bmatrix} 1 & -2 & -1 & 3 \\ 2 & -4 & 1 & 0 \\ 1 & -2 & 2 & -3 \end{bmatrix}$ is the common coefficient matrix of all three systems.

The *second* observation is that the solution set for the homogeneous system of equations can be written as the span of a set of vectors. The solution sets for the two non-homogeneous systems of equations cannot be written as the span of a set of vectors because these solution sets do not include $\vec{0}$.

The *third* observation is that the solution sets S, U and V are very similar. To be more precise, the solution sets S and U of the non-homogeneous systems differ by only a constant vector and the part which they have in common is V. We will return to this observation later and formalize it in Theorem 3.11.6 (Solutions to $A\vec{x} = \vec{0}$ and $A\vec{x} = \vec{b}$).

The systems in Examples 3.7.1, 3.7.2, 3.7.3 and 3.7.4 have the same coefficient matrix, A, but different right-hand sides:

$$\begin{bmatrix} 1 \\ 5 \\ 4 \end{bmatrix}, \quad \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \quad \begin{bmatrix} -1 \\ 4 \\ 5 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix},$$

respectively. To solve each of the previous four examples, we performed the same sequence of EROs. We can solve all four systems at once using the following "super-augmented matrix",

$$\left[\begin{array}{cccc|cccc} 1 & -2 & -1 & 3 & 1 & 1 & -1 & 0 \\ 2 & -4 & 1 & 0 & 5 & 2 & 4 & 0 \\ 1 & -2 & 2 & -3 & 4 & 3 & 5 & 0 \end{array}\right],$$

which has RREF
$$\left[\begin{array}{cccc|cccc} 1 & -2 & 0 & 1 & 2 & 1 & 1 & 0 \\ 0 & 0 & 1 & -2 & 1 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 \end{array}\right].$$

We use the RREF of the "super-augmented matrix" to solve the four systems. For the first system, we need the first column after the vertical line.

$$\left[\begin{array}{cccc|c} 1 & -2 & 0 & 1 & 2 \\ 0 & 0 & 1 & -2 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

This corresponds to the system of equations

$$\begin{array}{cccccc} x_1 & -2x_2 & & +x_4 & = & 2 \\ & & x_3 & -2x_4 & = & 1 \\ & & & 0 & = & 0 \end{array}$$

with solution set
$$S = \left\{ \begin{bmatrix} 2 \\ 0 \\ 1 \\ 0 \end{bmatrix} + s \begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \end{bmatrix} + t \begin{bmatrix} -1 \\ 0 \\ 2 \\ 1 \end{bmatrix} \ : \ s, t \in \mathbb{R} \right\}.$$

For the second system, we need the second column after the vertical line.

$$\left[\begin{array}{cccc|c} 1 & -2 & 0 & 1 & 1 \\ 0 & 0 & 1 & -2 & 0 \\ 0 & 0 & 0 & 0 & 2 \end{array}\right]$$

This is an inconsistent system and has solution set $T = \varnothing$.

For the third and fourth systems we need the third and fourth columns after the vertical line, respectively,

$$\left[\begin{array}{cccc|c} 1 & -2 & 0 & 1 & 1 \\ 0 & 0 & 1 & -2 & 2 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right] \text{ and } \left[\begin{array}{cccc|c} 1 & -2 & 0 & 1 & 0 \\ 0 & 0 & 1 & -2 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right].$$

These correspond to solution sets

$$U = \left\{ \begin{bmatrix} 1 \\ 0 \\ 2 \\ 0 \end{bmatrix} + a \begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \end{bmatrix} + b \begin{bmatrix} -1 \\ 0 \\ 2 \\ 1 \end{bmatrix} \ : \ a, b \in \mathbb{R} \right\}$$

and
$$V = \left\{ s \begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \end{bmatrix} + t \begin{bmatrix} -1 \\ 0 \\ 2 \\ 1 \end{bmatrix} \ : \ s, t \in \mathbb{R} \right\} = \text{Span} \left\{ \begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 2 \\ 1 \end{bmatrix} \right\},$$

respectively.

3.8 Solving Systems of Linear Equations over $\mathbb{C}$

The Gauss–Jordan Algorithm works just as well over $\mathbb{C}$ as it does over $\mathbb{R}$. The only new feature is that the arithmetic is more intricate because of the complex numbers. If parameters are introduced in the solution, then these parameters take on all complex values instead of just real values.

REMARK

Recall that for $z = a + bi \in \mathbb{C},\ z \neq 0$, that $\dfrac{1}{z} = \dfrac{\overline{z}}{z\overline{z}} = \dfrac{a - bi}{a^2 + b^2}$.

Example 3.8.1 Solve the following system of linear equations over $\mathbb{C}$.

$$\begin{array}{rlcr} (1+i)\,z_1 & +(-2-3i)z_2 & = & -15i \\ (1+3i)\,z_1 & +(-1-8i)z_2 & = & 15-30i \end{array}$$

Solution:

The augmented matrix is

$$\left[\begin{array}{cc|c} 1+i & -2-3i & -15i \\ 1+3i & -1-8i & 15-30i \end{array}\right].$$

We perform the following EROs.

$$R_1 \to \frac{1}{1+i}R_1 \left(= \left(\frac{1}{2} - \frac{1}{2}i\right) R_1 \right) \text{ gives } \left[\begin{array}{cc|c} 1 & -\frac{5}{2} - \frac{1}{2}i & -\frac{15}{2} - \frac{15}{2}i \\ 1+3i & -1-8i & 15-30i \end{array}\right].$$

$$R_2 \to -(1+3i)R_1 + R_2 \text{ gives } \left[\begin{array}{cc|c} 1 & -\frac{5}{2} - \frac{1}{2}i & -\frac{15}{2} - \frac{15}{2}i \\ 0 & 0 & 0 \end{array}\right].$$

This matrix is in RREF. The rank of the coefficient matrix and the rank of the augmented matrix are both equal to 1 and so the system is consistent. The number of parameters in the solution set is $2 - 1 = 1$ parameter in the solution set. We let the free variable $z_2 = t$ for $t \in \mathbb{C}$. Solving for the basic variable z_1 we get $z_1 = -\frac{15}{2} - \frac{15}{2}i + \left(\frac{5}{2} + \frac{1}{2}i\right)t$. Therefore, the solution set, S, is

$$S = \left\{ \begin{bmatrix} \left(-\frac{15}{2} - \frac{15}{2}i\right) + \left(\frac{5}{2} + \frac{1}{2}i\right)t \\ t \end{bmatrix} : t \in \mathbb{C} \right\} = \left\{ \begin{bmatrix} -\frac{15}{2} - \frac{15}{2}i \\ 0 \end{bmatrix} + \begin{bmatrix} \frac{5}{2} + \frac{1}{2}i \\ 1 \end{bmatrix} t : t \in \mathbb{C} \right\}$$

3.9 Matrix–Vector Multiplication

In this section we will define a notion of matrix–vector multiplication that will be very useful in our study of systems of linear equations. However, before we give the definition, we will consider some different ways that we can partition a matrix.

Definition 3.9.1

Row Vector

A **row vector** is a matrix with exactly one row. For a matrix $A \in M_{m\times n}(\mathbb{F})$, we will denote the i^{th} row of A by $\overrightarrow{\text{row}_i}(A)$. That is,

$$\overrightarrow{\text{row}_i}(A) = \begin{bmatrix} a_{i1} & a_{i2} & \cdots & a_{in} \end{bmatrix}.$$

We can write an $m \times n$ matrix A as

$$A = \begin{bmatrix} \overrightarrow{\text{row}_1}(A) \\ \overrightarrow{\text{row}_2}(A) \\ \vdots \\ \overrightarrow{\text{row}_m}(A) \end{bmatrix}.$$

In other words, we can partition A into m row vectors, each with n components.

Example 3.9.2

Let $A = \begin{bmatrix} 5 & -2 & 0 \\ 2 & 1 & -8 \end{bmatrix}$. Then $\overrightarrow{\text{row}_1}(A) = \begin{bmatrix} 5 & -2 & 0 \end{bmatrix}$ and $\overrightarrow{\text{row}_2}(A) = \begin{bmatrix} 2 & 1 & -8 \end{bmatrix}$.

Similarly, we can think of an $m \times n$ matrix A as $A = \begin{bmatrix} \vec{a_1} & \vec{a_2} & \dots & \vec{a_n} \end{bmatrix}$, where $\vec{a_i} \in \mathbb{F}^m$. In other words, we can partition A into n column vectors, each with m components. We will use the convention, demonstrated here, where an upper case letter refers to the matrix and the corresponding lower case letter with subscripts refers to the columns.

Example 3.9.3

Let $B = \begin{bmatrix} 1 & 2 \\ -3 & -4 \\ 7 & 9 \end{bmatrix}$. Then $B = \begin{bmatrix} \vec{b_1} & \vec{b_2} \end{bmatrix}$, where $\vec{b}_1 = \begin{bmatrix} 1 \\ -3 \\ 7 \end{bmatrix}$ and $\vec{b}_2 = \begin{bmatrix} 2 \\ -4 \\ 9 \end{bmatrix}$.

We now define the product of an $m \times n$ matrix and a (column) vector in $\mathbb{F}^n$.

Definition 3.9.4

Matrix–Vector Multiplication in Terms of the Individual Entries

Let $A \in M_{m\times n}(\mathbb{F})$ and $\vec{x} \in \mathbb{F}^n$. We define the product $A\vec{x}$ as follows:

$$A\vec{x} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \end{bmatrix}.$$

REMARK

Notice that we have defined the product of an $m \times n$ matrix A with a vector $\vec{x}$ only if $\vec{x} \in \mathbb{F}^n$. The expression $A\vec{y}$ is meaningless if $\vec{y} \in \mathbb{F}^k$ with $k \neq n$.

In other words, the number of components of $\vec{x}$ must be equal to the number of columns of A for the product $A\vec{x}$ to be defined.

The resulting product $A\vec{x}$ of $A \in M_{m\times n}(\mathbb{F})$ and $\vec{x} \in \mathbb{F}^n$ is a vector in $\mathbb{F}^m$. The i^{th} component of $A\vec{x}$ is the sum

$$a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n = \sum_{j=1}^{n} a_{ij}x_j = \sum_{j=1}^{n}(\overrightarrow{\text{row}}_i(A))_j x_j.$$

Informally, we can think of the sum above as something analogous to a "dot product" of the i^{th} row of A with the vector $\vec{x}$.

Example 3.9.5

Compute the product $A\vec{x}$ where $A = \begin{bmatrix} 1 & 6 & 1 \\ 3 & 4 & 5 \\ 5 & 2 & -3 \end{bmatrix}$ and $\vec{x} = \begin{bmatrix} 1 \\ -4 \\ 6 \end{bmatrix}$.

Solution: We have coloured the matrix red and the vector blue to emphasize the role their entries play in calculating the product.

First, we calculate the components of the product by calculating "dot products" of the rows of A with $\vec{x}$.

$$\begin{bmatrix} 1 & 6 & 1 \\ 3 & 4 & 5 \\ 5 & 2 & -3 \end{bmatrix} \begin{bmatrix} 1 \\ -4 \\ 6 \end{bmatrix} = \begin{bmatrix} (1)(1) + (6)(-4) + (1)(6) \\ (3)(1) + (4)(-4) + (5)(6) \\ (5)(1) + (2)(-4) + (-3)(6) \end{bmatrix} = \begin{bmatrix} -17 \\ 17 \\ -21 \end{bmatrix}.$$

We also make the observation that

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

which is a linear combination of the columns of A. Therefore, we can re-express matrix–vector multiplication in terms of the columns of A.

Proposition 3.9.6 **(Matrix–Vector Multiplication in Terms of Columns)**

Let $A \in M_{m\times n}(\mathbb{F})$ and $\vec{x} \in \mathbb{F}^n$. Then

$$A\vec{x} = \begin{bmatrix} \vec{a}_1 & \vec{a}_2 & \cdots & \vec{a}_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = x_1\vec{a}_1 + x_2\vec{a}_2 + \cdots + x_n\vec{a}_n.$$

Example 3.9.7

We will compute $\begin{bmatrix} 1 & 6 & 1 \\ 3 & 4 & 5 \\ 5 & 2 & -3 \end{bmatrix} \begin{bmatrix} 1 \\ -4 \\ 6 \end{bmatrix}$ again, this time by thinking of the product as a linear combination of the columns of A:

$$\begin{aligned}\begin{bmatrix}1 & 6 & 1\\ 3 & 4 & 5\\ 5 & 2 & -3\end{bmatrix}\begin{bmatrix}1\\ -4\\ 6\end{bmatrix} &= (1)\begin{bmatrix}1\\ 3\\ 5\end{bmatrix} + (-4)\begin{bmatrix}6\\ 4\\ 2\end{bmatrix} + (6)\begin{bmatrix}1\\ 5\\ -3\end{bmatrix}\\ &= \begin{bmatrix}1\\ 3\\ 5\end{bmatrix} + \begin{bmatrix}-24\\ -16\\ -8\end{bmatrix} + \begin{bmatrix}6\\ 30\\ -18\end{bmatrix}\\ &= \begin{bmatrix}-17\\ 17\\ -21\end{bmatrix}.\end{aligned}$$

Example 3.9.8

Compute the product $A\vec{x}$ where $A = \begin{bmatrix}1+i & 2+2i & 3-i\\ 2+3i & 4+i & 5-2i\end{bmatrix}$ and $\vec{x} = \begin{bmatrix}1\\ 1-i\\ 2-3i\end{bmatrix}$.

Solution: We compute the product in two different ways. We have coloured the matrix red and the vector blue to emphasize the role their entries play in calculating the product.

First, we calculate the components of the product by calculating "dot products" of the rows of A with $\vec{x}$.

$$\begin{aligned}A\vec{x} &= \begin{bmatrix}1+i & 2+2i & 3-i\\ 2+3i & 4+i & 5-2i\end{bmatrix}\begin{bmatrix}1\\ 1-i\\ 2-3i\end{bmatrix}\\ &= \begin{bmatrix}(1+i)(1)+(2+2i)(1-i)+(3-i)(2-3i)\\ (2+3i)(1)+(4+i)(1-i)+(5-2i)(2-3i)\end{bmatrix}\\ &= \begin{bmatrix}8-10i\\ 11-19i\end{bmatrix}.\end{aligned}$$

Second, we calculate the product by thinking of the product as a linear combination of the columns of A.

$$\begin{aligned}A\vec{x} &= \begin{bmatrix}1+i & 2+2i & 3-i\\ 2+3i & 4+i & 5-2i\end{bmatrix}\begin{bmatrix}1\\ 1-i\\ 2-3i\end{bmatrix}\\ &= (1)\begin{bmatrix}1+i\\ 2+3i\end{bmatrix} + (1-i)\begin{bmatrix}2+2i\\ 4+i\end{bmatrix} + (2-3i)\begin{bmatrix}3-i\\ 5-2i\end{bmatrix}\\ &= \begin{bmatrix}1+i\\ 2+3i\end{bmatrix} + \begin{bmatrix}4\\ 5-3i\end{bmatrix} + \begin{bmatrix}3-11i\\ 4-19i\end{bmatrix} = \begin{bmatrix}8-10i\\ 11-19i\end{bmatrix}.\end{aligned}$$

The following useful result can be proved using Proposition 3.9.6 (Matrix–Vector Multiplication in Terms of Columns).

Theorem 3.9.9 **(Linearity of Matrix–Vector Multiplication)**

Let $A \in M_{m\times n}(\mathbb{F})$. Let $\vec{x}, \vec{y} \in \mathbb{F}^n$ and $c \in \mathbb{F}$. Then

(a) $A(\vec{x}+\vec{y}) = A\vec{x} + A\vec{y}$.

(b) $A(c\vec{x}) = cA\vec{x}$.

3.10 Using a Matrix–Vector Product to Express a System of Linear Equations

Consider the system of linear equations

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

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

This system can now be represented as $A\vec{x} = \vec{b}$, where $\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$ and $\vec{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$.

Going forwards, we will also refer to expressions of the form $A\vec{x} = \vec{b}$, with $A \in M_{m\times n}(\mathbb{F})$, $\vec{x} \in \mathbb{F}^m$ and $\vec{b} \in \mathbb{F}^m$, as "a system of linear equations", acknowledging the equivalence of a system of m linear equations (as given above) to the expression $A\vec{x} = \vec{b}$.

Example 3.10.1 The system of equations

$$\begin{array}{rrrrcr} 5x_1 & +4x_2 & -7x_3 & +2x_4 & = & -1 \\ -6x_1 & -2x_2 & +3x_3 & & = & 2 \\ & +3x_2 & +5x_3 & +5x_4 & = & 7 \end{array}$$

can be represented as $A\vec{x} = \vec{b}$, where

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

Using this notation and the linearity of matrix–vector multiplication, we can prove a result that is essentially a corollary of Theorem 3.6.7 (System Rank Theorem).

Proposition 3.10.2 Let $A \in M_{m\times n}(\mathbb{F})$. If for every vector $\vec{e_i}$ in the standard basis of $\mathbb{F}^m$ the system of equations $A\vec{x} = \vec{e_i}$ is consistent, then $\operatorname{rank}(A) = m$.

Proof: We begin by assuming that for every vector $\vec{e_i}$ in the standard basis of $\mathbb{F}^m$, the system of equations $A\vec{x} = \vec{e_i}$ is consistent. Let $\vec{x_i}$ be a solution to $A\vec{x} = \vec{e_i}$.

We know that for any vector $\vec{b} \in \mathbb{F}^m$, we can write $\vec{b}$ as a linear combination of the vectors in the standard basis for $\mathbb{F}^m$. That is, there exists $c_1, \dots, c_m \in \mathbb{F}$ such that

$$\vec{b} = c_1\vec{e_1} + \cdots + c_m\vec{e_m}.$$

Using these scalars, the solutions $\vec{x}_i$ to $A\vec{x} = \vec{e_i}$ and the linearity of matrix-vector multiplication we find that

$$A(c_1\vec{x}_1 + \cdots + c_m\vec{x}_m) = c_1A\vec{x}_1 + \cdots + c_mA\vec{x}_m = c_1\vec{e_1} + \cdots + c_m\vec{e_m} = \vec{b}.$$

Therefore, $c_1\vec{x_1} + \cdots + c_m\vec{x_m}$ is a solution to $A\vec{x} = \vec{b}$. Thus, $A\vec{x} = \vec{b}$ is consistent for every $\vec{b} \in \mathbb{F}^m$. Therefore, by Part (b) of Theorem 3.6.7 (System Rank Theorem), $\operatorname{rank}(A) = m$. □

3.11 Solution Sets to Systems of Linear Equations

Let $A\vec{x} = \vec{b}$ be a system of linear equations. If $\vec{b} = \vec{0}$, then the system is homogeneous. If $\vec{b} \neq \vec{0}$, then the system is non-homogeneous. We will consider the solution sets of homogeneous and non-homogeneous systems with the same coefficient matrix.

As we noted in Section 3.7, a homogeneous system of linear equations is always consistent, since we can always set all of the variables to 0 (the trivial solution). In other words, the solution set to a homogeneous system always contains the zero vector (the trivial solution), and thus, it is never empty. The solution set of a non-homogeneous system does not contain the trivial solution. Non-homogeneous systems may be consistent or inconsistent.

The next result gives another important property of the solution set to a homogeneous system.

Proposition 3.11.1 Let $A\vec{x} = \vec{0}$ be a homogeneous system of linear equations with solution set S. If $\vec{x}, \vec{y} \in S$, and if $c \in \mathbb{F}$, then $\vec{x} + \vec{y} \in S$ and $c\vec{x} \in S$.

Proof: Assume that $\vec{x}, \vec{y} \in S$. Then $\vec{x}$ and $\vec{y}$ are solutions to the homogeneous system and $A\vec{x} = \vec{0}$ and $A\vec{y} = \vec{0}$. Using the linearity of matrix multiplication,

$$A(\vec{x} + \vec{y}) = A\vec{x} + A\vec{y} = \vec{0} + \vec{0} = \vec{0}$$

and

$$A(c\vec{x}) = c(A\vec{x}) = c(\vec{0}) = \vec{0}.$$

Therefore, $\vec{x} + \vec{y} \in S$ and $c\vec{x} \in S$. □

We can combine these two results to state that

$$\text{If } \vec{x}, \vec{y} \in S \text{ and if } c, d \in \mathbb{F}, \text{ then } c\vec{x} + d\vec{y} \in S.$$

We can also extend this result inductively to n vectors in S and n scalars in $\mathbb{F}$. In other words, given n solutions to a homogeneous system of linear equations, then *any linear combination of these solutions* is also a solution to the homogeneous system. We say that S is *closed* under addition and scalar multiplication. This fact, along with the fact that S is non-empty means that S forms a type of object called a *subspace*. We will explore subspaces later in the course.

Example 3.11.2

Consider the homogeneous system given in Example 3.7.4 of Section 3.7. It has coefficient matrix

$$A = \begin{bmatrix} 1 & -2 & -1 & 3 \\ 2 & -4 & 1 & 0 \\ 1 & -2 & 2 & -3 \end{bmatrix}$$

and solution set

$$V = \left\{ s \begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \end{bmatrix} + t \begin{bmatrix} -1 \\ 0 \\ 2 \\ 1 \end{bmatrix} : s, t \in \mathbb{R} \right\}.$$

By choosing $s = 1$, $t = 0$ and then $s = 0$, $t = 1$, we get the solutions $\vec{x} = \begin{bmatrix} 2 & 1 & 0 & 0 \end{bmatrix}^T$ and $\vec{y} = \begin{bmatrix} -1 & 0 & 2 & 1 \end{bmatrix}^T$, respectively. Let $\vec{z}$ be the following linear combination of these two solutions:

$$\vec{z} = 2\vec{x} + 3\vec{y} = \begin{bmatrix} 1 & 2 & 6 & 3 \end{bmatrix}^T.$$

Then

$$A\vec{z} = \begin{bmatrix} 1 & -2 & -1 & 3 \\ 2 & -4 & 1 & 0 \\ 1 & -2 & 2 & -3 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 6 \\ 3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

And so $\vec{z} \in V$ as predicted by Proposition 3.11.1.

Definition 3.11.3

Associated Homogeneous System

Let $A\vec{x} = \vec{b}$, where $\vec{b} \neq \vec{0}$, be a non-homogeneous system of linear equations. The **associated homogeneous system** to this system is the system $A\vec{x} = \vec{0}$.

Example 3.11.4

(a) $\begin{bmatrix} 1 & 2 & 3 \\ -4 & -5 & 7 \\ 2 & -4 & 6 \end{bmatrix} \vec{x} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$ is a homogeneous system of 3 equations in 3 unknowns.

(b) $\begin{bmatrix} 1 & 2 & 3 & 4 \\ -4 & -5 & 7 & -9 \\ 2 & -4 & 6 & -4 \end{bmatrix} \vec{y} = \begin{bmatrix} 2 \\ 4 \\ 6 \end{bmatrix}$ is a non-homogeneous system of 3 equations in 4 unknowns,

with associated homogeneous system of equations $\begin{bmatrix} 1 & 2 & 3 & 4 \\ -4 & -5 & 7 & -9 \\ 2 & -4 & 6 & -4 \end{bmatrix} \vec{y} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$.

Definition 3.11.5
Particular Solution

Let $A\vec{x} = \vec{b}$ be a consistent system of linear equations. We refer to a solution of this system, $\vec{x_p}$, as a **particular solution** to this system.

Since a consistent system $A\vec{x} = \vec{b}$ either has a unique solution or it has an infinite number of solutions, there may be an infinite number of choices for a particular solution.

We can use a particular solution to a non-homogeneous system to find a relationship between the non-homogeneous system and its associated homogeneous system. Specifically, we can express the solution set of a consistent, non-homogeneous system $A\vec{x} = \vec{b}$ in terms of a particular solution to the system and the solution set of its associated homogeneous system $A\vec{x} = \vec{0}$ as follows.

Theorem 3.11.6

(Solutions to $A\vec{x} = \vec{0}$ and $A\vec{x} = \vec{b}$)

Let $A\vec{x} = \vec{b}$, where $\vec{b} \neq \vec{0}$, be a consistent non-homogeneous system of linear equations with solution set $\tilde{S}$. Let $A\vec{x} = \vec{0}$ be the associated homogeneous system with solution set S. If $\vec{x_p} \in \tilde{S}$, then

$$\tilde{S} = \{\vec{x_p} + \vec{x} : \vec{x} \in S\}.$$

Proof: Let $\vec{x_p} \in \tilde{S}$. Therefore, $A\vec{x_p} = \vec{b}$. Let $\vec{x}$ be a solution of its associated homogeneous system. Therefore, $A\vec{x} = \vec{0}$. To show that $\tilde{S} = \{\vec{x_p} + \vec{x} : \vec{x} \in S\}$, we will show that $\tilde{S} \subseteq \{\vec{x_p} + \vec{x} : \vec{x} \in S\}$ and $\{\vec{x_p} + \vec{x} : \vec{x} \in S\} \subseteq \tilde{S}$.

Suppose that $\vec{y} \in \tilde{S}$. Therefore, $A\vec{y} = \vec{b}$. We can write $\vec{y}$ as $\vec{y} = \vec{x_p} + \vec{y} - \vec{x_p}$. Using the linearity of matrix-vector multiplication we have that

$$A(\vec{y} - \vec{x_p}) = A\vec{y} - A\vec{x_p} = \vec{b} - \vec{b} = \vec{0},$$

which implies that $\vec{y} - \vec{x_p} \in S$. Therefore, $\vec{y} \in \{\vec{x_p} + \vec{x} : \vec{x} \in S\}$. Thus, we have shown that $\tilde{S} \subseteq \{\vec{x_p} + \vec{x} : \vec{x} \in S\}$.

Now, suppose that $\vec{z} \in \{\vec{x_p} + \vec{x} : \vec{x} \in S\}$. Then $\vec{z} = \vec{x_p} + \vec{x}_1$ for some $\vec{x}_1 \in S$. Since $\vec{x}_1 \in S$, $A\vec{x}_1 = \vec{0}$. Again using the linearity of matrix-vector multiplication, we have that

$$A\vec{z} = A(\vec{x_p} + \vec{x}_1) = A\vec{x_p} + A\vec{x}_1 = \vec{b} + \vec{0} = \vec{b}.$$

Therefore, $\vec{z} \in \tilde{S}$. Thus, we have shown $\{\vec{x_p} + \vec{x} : \vec{x} \in S\} \subseteq \tilde{S}$. $\square$

The next example illustrates the idea from this theorem.

Example 3.11.7

Let $A = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$. Solve the following systems and represent their solutions sets graphically.

(a) $A\vec{x} = \vec{0}$

(b) $A\vec{x} = \begin{bmatrix} 4 \\ 8 \end{bmatrix}$

(c) $A\vec{x} = \begin{bmatrix} -2 \\ -4 \end{bmatrix}$

Solution: We solve all three systems simultaneously by row-reducing the "super-augmented matrix"

$$\left[\begin{array}{cc|ccc} 1 & 2 & 0 & 4 & -2 \\ 2 & 4 & 0 & 8 & -4 \end{array}\right].$$

We perform the ERO $-2R_1 + R_2$ to obtain the following matrix which is in RREF:

$$\left[\begin{array}{cc|ccc} 1 & 2 & 0 & 4 & -2 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right].$$

Therefore, the solution set to the first system, which is a homogeneous system, is

$$S = \left\{ \begin{bmatrix} -2 \\ 1 \end{bmatrix} t : t \in \mathbb{R} \right\}.$$

The solution to the second system is

$$T = \left\{ \begin{bmatrix} 4 \\ 0 \end{bmatrix} + \begin{bmatrix} -2 \\ 1 \end{bmatrix} t : t \in \mathbb{R} \right\}.$$

The solution to the third system is

$$U = \left\{ \begin{bmatrix} -2 \\ 0 \end{bmatrix} + \begin{bmatrix} -2 \\ 1 \end{bmatrix} t : t \in \mathbb{R} \right\}.$$

We represent these solution sets graphically. The solution sets are labelled in the diagram.

We see that S is a line through the origin, which is to be expected because it is the solution set to a homogeneous system. The solution sets T and U are translations of the set S.

The solution set T is obtained by translating every vector in S to the right by 4 units, which is equivalent to adding $\begin{bmatrix}4\\0\end{bmatrix}$, a particular solution to the system in (b), to every vector in S.

Similarly, the solution set U is obtained by translating every vector in S to the left by 2 units, which is equivalent to adding $\begin{bmatrix}-2\\0\end{bmatrix}$, a particular solution to the system in (c), to every vector in S.

Let's consider another example in $\mathbb{R}^3$.

Example 3.11.8 Let $A = \begin{bmatrix}1&2&3\\2&4&6\\3&6&9\end{bmatrix}$. Solve the following systems and represent their solutions sets graphically.

(a) $A\vec{x} = \vec{0}$

(b) $A\vec{x} = \begin{bmatrix}6\\12\\18\end{bmatrix}$

(c) $A\vec{x} = \begin{bmatrix}-4\\-8\\-12\end{bmatrix}$

Solution: We solve all three systems simultaneously by row-reducing the "super-augmented matrix"

$$\left[\begin{array}{ccc|ccc}1&2&3&0&6&-4\\2&4&6&0&12&-8\\3&6&9&0&18&-12\end{array}\right].$$

We perform the EROs $-2R_1 + R_2$ and $-3R_1 + R_3$ to obtain the following matrix which is in RREF:

$$\left[\begin{array}{ccc|ccc}1&2&3&0&6&-4\\0&0&0&0&0&0\\0&0&0&0&0&0\end{array}\right].$$

Therefore, the solution set to the first system, which is a homogeneous system, is

$$S = \left\{\begin{bmatrix}-2\\1\\0\end{bmatrix}t + \begin{bmatrix}-3\\0\\1\end{bmatrix}s : t,s \in \mathbb{R}\right\} = \text{Span}\left\{\begin{bmatrix}-2\\1\\0\end{bmatrix}, \begin{bmatrix}-3\\0\\1\end{bmatrix}\right\}.$$

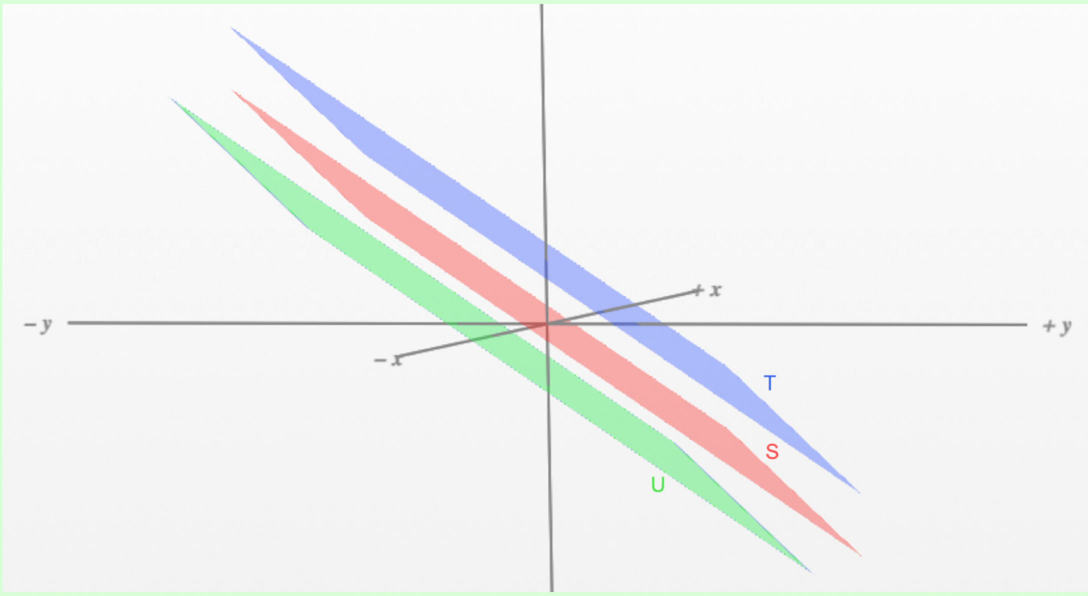
The solution to the second system is

$$T = \left\{\begin{bmatrix}6\\0\\0\end{bmatrix} + \begin{bmatrix}-2\\1\\0\end{bmatrix}t + \begin{bmatrix}-3\\0\\1\end{bmatrix}s : t,s \in \mathbb{R}\right\}.$$

The solution to the third system is

$$U = \left\{ \begin{bmatrix} -4 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix} t + \begin{bmatrix} -3 \\ 0 \\ 1 \end{bmatrix} s : t, s \in \mathbb{R} \right\}.$$

We represent these solution sets graphically. The solution sets are labelled in the diagram



We see that S is a plane through the origin, which is to be expected because it is the solution set to a homogeneous system.

The solution sets T and U are translations of the set S. For T, every vector in S has been translated in the direction of $\begin{bmatrix} 6 \\ 0 \\ 0 \end{bmatrix}$, a particular solution to the system in (b). For U, every vector in S has been translated in the direction of $\begin{bmatrix} -4 \\ 0 \\ 0 \end{bmatrix}$, a particular solution to the system in (c).

In the previous two examples, we see that we can find the solution set U by subtracting from every element in T a particular solution of T and then adding a particular solution of U. Let's focus on the sets in Example 3.11.8.

Let $\vec{z} \in T$, and let $\vec{u}_p = \begin{bmatrix} -4 \\ 0 \\ 0 \end{bmatrix}$ and $\vec{t}_p = \begin{bmatrix} 6 \\ 0 \\ 0 \end{bmatrix}$ be particular solutions of U and T, respectively. From Theorem 3.11.6 (Solutions to $A\vec{x} = \vec{0}$ and $A\vec{x} = \vec{b}$), as $\vec{z} \in T$, we know that

$$\vec{z} = \vec{t}_p + \vec{x}$$

for some $\vec{x} \in S$, the solution set of the associated homogeneous system. From above, we would express

$$\vec{x} = \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix} t + \begin{bmatrix} -3 \\ 0 \\ 1 \end{bmatrix} s \text{ for some } s, t \in \mathbb{R}.$$

As we can certainly write $\vec{u}_p = \vec{u}_p - \vec{t}_p + \vec{t}_p$ and any vector $\vec{u} \in U$ as $\vec{u} = \vec{u}_p + \vec{x}$ for some $\vec{x} \in S$ by Theorem 3.11.6 (Solutions to $A\vec{x} = \vec{0}$ and $A\vec{x} = \vec{b}$), we have

$$\begin{aligned} U &= \{\vec{u}_p + \vec{x} : \vec{x} \in S\} \\ &= \left\{\left(\vec{u}_p - \vec{t}_p + \vec{t}_p\right) + \vec{x} : \vec{x} \in S\right\} \\ &= \{\left(\vec{u}_p - \vec{t}_p\right) + \underbrace{\left(\vec{t}_p + \vec{x}\right)}_{\text{element of } T} : \vec{x} \in S\} \\ &= \left\{\left(\vec{u}_p - \vec{t}_p\right) + \vec{z} : \vec{z} \in T\right\}. \end{aligned}$$

This holds in general and provides a nice way to understand the relationships between parallel lines and between parallel planes. We formalize this discussion below.

Corollary 3.11.9 **(Solutions to $A\vec{x} = \vec{b}$ and $A\vec{x} = \vec{c}$)**

Consider the two consistent, non-homogeneous systems

$$A\vec{x} = \vec{b}, \text{ and } A\vec{x} = \vec{c},$$

where $\vec{b} \neq \vec{c}$. If $\tilde{S}_b$ and $\tilde{S}_c$ are their respective solution sets, with particular solutions $\vec{x_b}$ and $\vec{x_c}$, respectively, then

$$\tilde{S}_c = \{(\vec{x_c} - \vec{x_b}) + \vec{z} : \vec{z} \in \tilde{S}_b\}.$$

We can use this result to help us find the solution set to a consistent, non-homogeneous system, provided we have a particular solution to the system and we have the solution set to another consistent, non-homogeneous system which shares the coefficient matrix. Note that finding a particular solution of a system is not always a simple task.

Example 3.11.10 Using Example 3.11.8, solve the consistent non-homogeneous system $A\vec{x} = \vec{c}$ using the solution to $A\vec{x} = \vec{b}$, where

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 3 & 6 & 9 \end{bmatrix} \text{ and } \vec{b} = \begin{bmatrix} 6 \\ 12 \\ 18 \end{bmatrix}, \ \vec{c} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}.$$

Solution: From Example 3.11.8, we have that the solution set to $A\vec{x} = \begin{bmatrix} 6 & 12 & 18 \end{bmatrix}^T$ is

$$T = \tilde{S}_b = \left\{ \begin{bmatrix} 6 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix} t + \begin{bmatrix} -3 \\ 0 \\ 1 \end{bmatrix} s : t, s \in \mathbb{R} \right\}.$$

By inspection, we can see that $\begin{bmatrix} 1 & 0 & 0 \end{bmatrix}^T$ is a particular solution to $A\vec{x} = \vec{c}$.

Therefore, by Corollary 3.11.9 (Solutions to $A\vec{x} = \vec{b}$ and $A\vec{x} = \vec{c}$), the solution set to $A\vec{x} = \vec{c}$ is

$$\tilde{S}_c = \left\{ \left(\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} - \begin{bmatrix} 6 \\ 0 \\ 0 \end{bmatrix} \right) + \underbrace{\begin{bmatrix} 6 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix} t + \begin{bmatrix} -3 \\ 0 \\ 1 \end{bmatrix} s}_{\vec{z} \in \tilde{S}_b} : t, s \in \mathbb{R} \right\}$$

Note that in the final solution above, simplifying the expression in the set gives us

$$\tilde{S}_c = \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix} t + \begin{bmatrix} -3 \\ 0 \\ 1 \end{bmatrix} s : t, s \in \mathbb{R} \right\} = \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + \vec{x} : \vec{x} \in S \right\},$$

where S is the solution set of the associated homogeneous system, which is in alignment with Theorem 3.11.6 (Solutions to $A\vec{x} = \vec{0}$ and $A\vec{x} = \vec{b}$).

Chapter 4

Matrices

4.1 The Column and Row Spaces of a Matrix

In the previous chapter, we introduced the concept of a matrix, which was built from the coefficients of a system of equations. It turns out that matrices are interesting and powerful mathematical objects all by themselves. In this chapter, we will explore matrices further.

We saw in Proposition 3.9.6 that the matrix–vector product $A\vec{x}$ is a linear combination of the columns of A. We give the set of all linear combinations of the columns of A a name.

Definition 4.1.1
Column Space

Let $A \in M_{m\times n}(\mathbb{F})$. We define the **column space** of A, denoted by $\operatorname{Col}(A)$, to be the span of the columns of A. That is,

$$\operatorname{Col}(A) = \operatorname{Span}\{\vec{a_1}, \vec{a_2}, \ldots, \vec{a_n}\}.$$

Proposition 4.1.2 **(Consistent System and Column Space)**

Let $A \in M_{m\times n}(\mathbb{F})$ and $\vec{b} \in \mathbb{F}^m$. The system of linear equations $A\vec{x} = \vec{b}$ is consistent if and only if $\vec{b} \in \operatorname{Col}(A)$.

Proof: We must prove the statement in both directions. We begin with the forward direction. Assume that there exists a vector $\vec{y} = \begin{bmatrix} y_1 & y_2 & \cdots & y_n \end{bmatrix}^T$ such that $A\vec{y} = \vec{b}$. Thus,

$$\begin{bmatrix} \vec{a_1} & \vec{a_2} & \cdots & \vec{a_n} \end{bmatrix} \begin{bmatrix} y_1 & y_2 & \cdots & y_n \end{bmatrix}^T = \vec{b}.$$

Therefore,

$$y_1\vec{a_1} + y_2\vec{a_2} + \cdots + y_n\vec{a_n} = \vec{b}.$$

Thus, $\vec{b}$ is a linear combination of the columns of the matrix A, and so $\vec{b} \in \operatorname{Col}(A)$.

Now we prove the backward direction. Assume that $\vec{b} \in \operatorname{Col}(A)$. Then

$$\vec{b} = s_1\vec{a_1} + s_2\vec{a_2} + \cdots + s_n\vec{a_n}$$

for some $s_1, s_2, \ldots, s_n \in \mathbb{F}$. Let $\vec{s} = \begin{bmatrix} s_1 & s_2 & \cdots & s_n \end{bmatrix}^T$. Then

$$A\vec{s} = s_1\vec{a_1} + s_2\vec{a_2} + \cdots + s_n\vec{a_n} = \vec{b}.$$

Thus, $A\vec{s} = \vec{b}$, and so $\vec{s}$ is a solution to the system. Therefore, $A\vec{x} = \vec{b}$ is consistent. $\square$

We now begin our examination of a similar space for the rows of the matrix. First, we need the notion of the transpose of a matrix.

Definition 4.1.3 **Transpose**

Let $A \in M_{m\times n}(\mathbb{F})$. We define the **transpose** of A, denoted A^T, by $(A^T)_{ij} = (A)_{ji}$.

If A is an $m \times n$ matrix, then A^T is an $n \times m$ matrix. It is formed by making all the rows of A into the columns of A^T in the order in which they appear.

This definition is the motivation behind the notation that we have used to write a vector from $\mathbb{F}^n$ horizontally: $\vec{v} = \begin{bmatrix} v_1 & \dots & v_n \end{bmatrix}^T$.

Example 4.1.4

If $A = \begin{bmatrix} 1 & 2 & -3 \\ 4 & -5 & 6 \end{bmatrix}$, then $A^T = \begin{bmatrix} 1 & 4 \\ 2 & -5 \\ -3 & 6 \end{bmatrix}$.

In Section 3.9, we introduced the idea of thinking of an $m \times n$ matrix A as

$$A = \begin{bmatrix} \overrightarrow{\text{row}_1}(A) \\ \overrightarrow{\text{row}_2}(A) \\ \vdots \\ \overrightarrow{\text{row}_m}(A) \end{bmatrix}.$$

In other words, we partition A into m row vectors, each with n components. If we transpose these row vectors, we get vectors in $\mathbb{R}^n$. We use these vectors to define the row space of A.

Definition 4.1.5 **Row Space**

Let $A \in M_{m\times n}(\mathbb{F})$. We define the **row space** of A, denoted by $\text{Row}(A)$, to be the span of the transposed rows of A. That is,

$$\text{Row}(A) = \text{Span}\left\{ \left(\overrightarrow{\text{row}_1}(A)\right)^T, \left(\overrightarrow{\text{row}_2}(A)\right)^T, \dots, \left(\overrightarrow{\text{row}_m}(A)\right)^T \right\}.$$

Notice that, if $A \in M_{m\times n}(\mathbb{F})$, $\text{Row}(A)$ is a subset of $\mathbb{F}^n$ while $\text{Col}(A)$ is a subset of $\mathbb{F}^m$.

Example 4.1.6

If $A = \begin{bmatrix} 1 & 4 & 1 \\ 2 & 3 & -1 \\ 3 & 2 & 1 \\ 4 & 1 & -1 \end{bmatrix}$, then $\text{Row}(A) = \text{Span}\left\{ \begin{bmatrix} 1 \\ 4 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 3 \\ -1 \end{bmatrix}, \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 4 \\ 1 \\ -1 \end{bmatrix} \right\}$.

REMARK

Since the transposed rows of A are the columns of A^T, we see that $\text{Row}(A) = \text{Col}(A^T)$.

Applying EROs to a matrix does not affect the row space. We formalize this below.

Proposition 4.1.7

Let $A, B \in M_{m\times n}(\mathbb{F})$. If B is row equivalent to A, then

$$\text{Row}(B) = \text{Row}(A).$$

Proof: Using the remark above, we note that $\text{Row}(A) = \text{Col}(A^T)$ and $\text{Row}(B) = \text{Col}(B^T)$. Therefore, if we can show that $\text{Col}(B^T) = \text{Col}(A^T)$, then the proposition will be proved.

We begin by considering the case where B is obtained from A by applying exactly one ERO.

When we perform a single ERO of any type to A to obtain B, each column of B^T will be a linear combination of the columns of A^T. Therefore, $\text{Col}(B^T) \subseteq \text{Col}(A^T)$.

As we know from Theorem 3.2.18 (Elementary Operations), the operations which reverse EROs are themselves EROs. Therefore, we can obtain A from B by doing a single ERO. Thus, every column of A^T is a linear combination of the columns of B^T. Therefore, $\text{Col}(A^T) \subseteq \text{Col}(B^T)$ and we have that $\text{Col}(B^T) = \text{Col}(A^T)$.

In the case that B is obtained from A by applying multiple EROs, we can apply the above argument multiple times, once for each of the EROs. □

REMARK

One might expect that a similar result would hold for the column space. That is, if $A, B \in M_{m\times n}(\mathbb{F})$, and if B is row equivalent to A, then $\text{Col}(B) = \text{Col}(A)$. However, this result is not true. For example, the matrix $B = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$ is row equivalent to $A = \begin{bmatrix} 0 & 0 \\ 1 & 1 \end{bmatrix}$, but $\text{Col}(B) = \text{Span}\left\{\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right\}$ while $\text{Col}(A) = \text{Span}\left\{\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right\}$.

4.2 Matrix Equality and Multiplication

As a preliminary step toward discussing statements involving algebraic matrix operations, we establish what it means for matrices to be equal.

Definition 4.2.1
Matrix Equality

Let $A \in M_{m\times n}(\mathbb{F})$ and $B \in M_{p\times q}(\mathbb{F})$. We say that A and B are **equal** if

1. A and B have the same size, that is, $m = p$ and $n = q$, and
2. The corresponding entries of A and B are equal, i.e., $a_{ij} = b_{ij}$, for all $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, n$.

We denote this by writing $A = B$.

While this definition seems to suggest that matrix equality should always be determined through exhaustive comparison of matrix entries, we can also uniquely identify matrices based on their action on other objects. Our next theorem gives a very useful criterion for checking that two given matrices of the same size are equal based on their behaviour in matrix-vector products. It requires a preliminary lemma, which is important in its own right.

Lemma 4.2.2 **(Column Extraction)**

Let $A = \begin{bmatrix} \vec{a_1} & \cdots & \vec{a_n} \end{bmatrix} \in M_{m\times n}(\mathbb{F})$. Then $A\vec{e_i} = \vec{a_i}$ for all $i = 1, \ldots, n$.

In other words, the matrix-vector multiplication of A and the i^{th} standard basis vector $\vec{e_i}$ yields the i^{th} column of A.

Proof: According to the definition of matrix-vector multiplication,

$$\begin{aligned} A\vec{e_1} &= & 1\cdot\vec{a_1} + 0\cdot\vec{a_2} + 0\cdot\vec{a_3} + \cdots + 0\cdot\vec{a_n} &&= \vec{a_1}, \\ A\vec{e_2} &= & 0\cdot\vec{a_1} + 1\cdot\vec{a_2} + 0\cdot\vec{a_3} + \cdots + 0\cdot\vec{a_n} &&= \vec{a_2}, \\ &&\vdots&& \\ A\vec{e_n} &= & 0\cdot\vec{a_1} + 0\cdot\vec{a_2} + 0\cdot\vec{a_3} + \cdots + 1\cdot\vec{a_n} &&= \vec{a_n}. \end{aligned}$$

□

Theorem 4.2.3 **(Equality of Matrices)**

Let $A, B \in M_{m\times n}(\mathbb{F})$. Then

$$A = B \text{ if and only if } A\vec{x} = B\vec{x} \text{ for all } \vec{x} \in \mathbb{F}^n.$$

Proof: If $A = B$, then it is clear that $A\vec{x} = B\vec{x}$ for all $\vec{x} \in \mathbb{F}^n$.

Now, let $A = \begin{bmatrix} \vec{a_1} & \cdots & \vec{a_n} \end{bmatrix}$, $B = \begin{bmatrix} \vec{b_1} & \cdots & \vec{b_n} \end{bmatrix}$, and suppose that $A\vec{x} = B\vec{x}$ for all $\vec{x} \in \mathbb{F}^n$. Then we can take $\vec{x}$ equal to the i^{th} standard basis vector $\vec{e_i}$ and conclude that $A\vec{e_i} = B\vec{e_i}$ for all $i = 1, \ldots, n$. But then it follows from Lemma 4.2.2 (Column Extraction) that

$$\vec{a}_i = A\vec{e_i} = B\vec{e_i} = \vec{b}_i.$$

Since the columns of A and B are the same, we conclude that the matrices are equal. □

Recall that we can calculate the matrix–vector product $A\vec{x}$ provided the sizes of A and $\vec{x}$ are compatible: we need $A \in M_{m\times n}(\mathbb{F})$ and $\vec{x} \in \mathbb{F}^n$. We extend this process to multiplying several column vectors, of the correct size, by the same matrix simultaneously.

Definition 4.2.4
Matrix Multiplication

Let $A \in M_{m\times n}(\mathbb{F})$ and $B \in M_{n\times p}(\mathbb{F})$. We define the **matrix product** $AB = C$ to be the matrix $C \in M_{m\times p}(\mathbb{F})$, constructed as follows:

$$C = AB = A\begin{bmatrix} \vec{b_1} & \vec{b_2} & \cdots & \vec{b_p} \end{bmatrix} = \begin{bmatrix} A\vec{b_1} & A\vec{b_2} & \cdots & A\vec{b_p} \end{bmatrix}.$$

That is, the j^{th} column of C, $\vec{c_j}$, is obtained by multiplying the matrix A by the j^{th} column of the matrix B:

$$\vec{c_j} = A\vec{b_j}, \text{ for all } j = 1, \ldots, p.$$

Thus, we can construct the product AB, column by column, by calculating a sequence of matrix–vector products $A\vec{b_j}$.

REMARKS

- In calculating the product AB, the number of columns of the first matrix, A, must be equal to the number of rows of the second matrix, B. If these values do not match, then the product is undefined.
- Matrix multiplication is generally non-commutative, so the order of multiplication matters. In other words, for matrices A and B, we are not guaranteed that $AB = BA$.
- We can immediately connect the definition of matrix multiplication to our understanding of column spaces. As the j^{th} column of $C = AB$ is obtained by multiplying the matrix A by the j^{th} column of the matrix B, the j^{th} column of C must be a linear combination of the columns of A (by Proposition 3.9.6 (Matrix–Vector Multiplication in Terms of Columns)), and therefore is a member of $\mathrm{Col}(A)$.

Example 4.2.5 Given $A = \begin{bmatrix} 1 & 2 \\ 3 & 5 \\ 8 & 7 \end{bmatrix}$ and $B = \begin{bmatrix} \mathbf{-1} & \mathbf{3} \\ \mathbf{2} & \mathbf{-4} \end{bmatrix}$, calculate the products AB and BA if possible.

Solution: First, we note that the number of columns of A is equal to the number of rows of B and so the product is defined. Also, since A has three rows and B has two columns, then the size of the product will be 3×2.

We have coloured the entries of A blue and the entries of B red to emphasize the role their entries play in calculating the product.

$$\begin{aligned}
\begin{bmatrix} 1 & 2 \\ 3 & 5 \\ 8 & 7 \end{bmatrix}\begin{bmatrix} \mathbf{-1} & \mathbf{3} \\ \mathbf{2} & \mathbf{-4} \end{bmatrix} &= \left[\begin{bmatrix} 1 & 2 \\ 3 & 5 \\ 8 & 7 \end{bmatrix}\begin{bmatrix} \mathbf{-1} \\ \mathbf{2} \end{bmatrix} \quad \begin{bmatrix} 1 & 2 \\ 3 & 5 \\ 8 & 7 \end{bmatrix}\begin{bmatrix} \mathbf{3} \\ \mathbf{-4} \end{bmatrix}\right] \\
&= \begin{bmatrix} (1)(\mathbf{-1}) + (2)(\mathbf{2}) & (1)(\mathbf{3}) + (2)(\mathbf{-4}) \\ (3)(\mathbf{-1}) + (5)(\mathbf{2}) & (3)(\mathbf{3}) + (5)(\mathbf{-4}) \\ (8)(\mathbf{-1}) + (7)(\mathbf{2}) & (8)(\mathbf{3}) + (7)(\mathbf{-4}) \end{bmatrix} \\
&= \begin{bmatrix} 3 & -5 \\ 7 & -11 \\ 6 & -4 \end{bmatrix}.
\end{aligned}$$

The product BA is undefined, since B has 2 columns, A has 3 rows and $2 \neq 3$.

Example 4.2.6 Given $A = \begin{bmatrix} 2-i & 1-2i \\ 3-2i & 1-3i \end{bmatrix}$ and $B = \begin{bmatrix} \mathbf{1+i} & \mathbf{-2+i} \\ \mathbf{-1+i} & \mathbf{3+2i} \end{bmatrix}$, calculate the products AB and BA if possible.

Solution: First, we note that the number of columns of A is equal to the number of rows of B and so the product is defined. Also, since A has two rows and B has two columns, then the size of the product will be 2×2.

We have coloured the entries of A blue and the entries of B red to emphasize the role their entries play in calculating the product.

$$
\begin{aligned}
&\begin{bmatrix} 2-i & 1-2i \\ 3-2i & 1-3i \end{bmatrix} \begin{bmatrix} \mathbf{1+i} & \mathbf{-2+i} \\ \mathbf{-1+i} & \mathbf{3+2i} \end{bmatrix} \\
&= \left[\begin{bmatrix} 2-i & 1-2i \\ 3-2i & 1-3i \end{bmatrix} \begin{bmatrix} \mathbf{1+i} \\ \mathbf{-1+i} \end{bmatrix} \quad \begin{bmatrix} 2-i & 1-2i \\ 3-2i & 1-3i \end{bmatrix} \begin{bmatrix} \mathbf{-2+i} \\ \mathbf{3+2i} \end{bmatrix} \right] \\
&= \begin{bmatrix} (2-i)(\mathbf{1+i}) + (1-2i)(\mathbf{-1+i}) & (2-i)(\mathbf{-2+i}) + (1-2i)(\mathbf{3+2i}) \\ (3-2i)(\mathbf{1+i}) + (1-3i)(\mathbf{-1+i}) & (3-2i)(\mathbf{-2+i}) + (1-3i)(\mathbf{3+2i}) \end{bmatrix} \\
&= \begin{bmatrix} 4+4i & 4 \\ 7+5i & 5 \end{bmatrix}.
\end{aligned}
$$

The product BA is also defined since the number of columns of B is equal to the number of rows of A.

$$
\begin{aligned}
&\begin{bmatrix} \mathbf{1+i} & \mathbf{-2+i} \\ \mathbf{-1+i} & \mathbf{3+2i} \end{bmatrix} \begin{bmatrix} 2-i & 1-2i \\ 3-2i & 1-3i \end{bmatrix} \\
&= \left[\begin{bmatrix} \mathbf{1+i} & \mathbf{-2+i} \\ \mathbf{-1+i} & \mathbf{3+2i} \end{bmatrix} \begin{bmatrix} 2-i \\ 3-2i \end{bmatrix} \quad \begin{bmatrix} \mathbf{1+i} & \mathbf{-2+i} \\ \mathbf{-1+i} & \mathbf{3+2i} \end{bmatrix} \begin{bmatrix} 1-2i \\ 1-3i \end{bmatrix} \right] \\
&= \begin{bmatrix} (\mathbf{1+i})(2-i) + (\mathbf{-2+i})(3-2i) & (\mathbf{1+i})(1-2i) + (\mathbf{-2+i})(1-3i) \\ (\mathbf{-1+i})(2-i) + (\mathbf{3+2i})(3-2i) & (\mathbf{-1+i})(1-2i) + (\mathbf{3+2i})(1-3i) \end{bmatrix} \\
&= \begin{bmatrix} -1+8i & 4+6i \\ 12+3i & 10-4i \end{bmatrix}.
\end{aligned}
$$

Notice that even though AB and BA are both defined, we have that $AB \neq BA$.

In practice, we usually construct the product $AB = C$ one entry at a time. The entry $(C)_{ij}$ is in the i^{th} row and the j^{th} column of the product. Therefore, $(C)_{ij} = (\vec{c_j})_i$, where $\vec{c_j} = A\vec{b_j}$. For $A \in M_{m\times n}(\mathbb{F})$, $B \in M_{n\times p}(\mathbb{F})$, and $C \in M_{m\times p}(\mathbb{F})$, the i^{th} entry of $A\vec{b_j}$ is

$$\sum_{k=1}^{n} a_{ik}b_{kj} = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{in}b_{nj}.$$

Thus,

$$(C)_{ij} = (A\vec{b_j})_i = \sum_{k=1}^{n} a_{ik}b_{kj}.$$

Consequently, in order to obtain the $(i,j)^{th}$ entry of AB, we need to multiply the corresponding entries of the i^{th} row of A and the j^{th} column of B and sum them up. We can re-write this last expression in a more suggestive manner:

$$(C)_{ij} = \begin{bmatrix} a_{i1} \\ a_{i2} \\ \vdots \\ a_{in} \end{bmatrix} \cdot \begin{bmatrix} b_{1j} \\ b_{2j} \\ \vdots \\ b_{nj} \end{bmatrix}.$$

REMARK

The equation above shows that the $(i, j)^{th}$ entry of AB may be obtained by taking the dot product of the (transposed) i^{th} row of A with the j^{th} column of B.

Example 4.2.7

Determine the $(2, 2)^{th}$ entry of the product

$$AB = C = \begin{bmatrix} 1 & 2 \\ 3 & 5 \\ 8 & 7 \end{bmatrix} \begin{bmatrix} \mathbf{-1} & \mathbf{3} \\ \mathbf{2} & \mathbf{-4} \end{bmatrix}.$$

Solution: We use the second row of A and the second column of B to calculate

$$c_{22} = \begin{bmatrix} 3 & 5 \end{bmatrix}^T \cdot \begin{bmatrix} \mathbf{3} \\ \mathbf{-4} \end{bmatrix} = \begin{bmatrix} 3 \\ 5 \end{bmatrix} \cdot \begin{bmatrix} \mathbf{3} \\ \mathbf{-4} \end{bmatrix} = 9 - 20 = -11.$$

Example 4.2.8

Determine the $(2, 1)^{th}$ entry of the product

$$CD = E = \begin{bmatrix} 2 - i & 1 - 2i \\ 3 - 2i & 1 - 3i \end{bmatrix} \begin{bmatrix} \mathbf{1 + i} & \mathbf{-2 + i} \\ \mathbf{-1 + i} & \mathbf{3 + 2i} \end{bmatrix}.$$

Solution: We use the second row of C and the first column of D to calculate

$$\begin{aligned} e_{21} &= \begin{bmatrix} 3 - 2i & 1 - 3i \end{bmatrix}^T \cdot \begin{bmatrix} \mathbf{1 + i} \\ \mathbf{-1 + i} \end{bmatrix} \\ &= \begin{bmatrix} 3 - 2i \\ 1 - 3i \end{bmatrix} \cdot \begin{bmatrix} \mathbf{1 + i} \\ \mathbf{-1 + i} \end{bmatrix} \\ &= (3 - 2i)(\mathbf{1 + i}) + (1 - 3i)(\mathbf{-1 + i}) \\ &= 7 + 5i. \end{aligned}$$

4.3 Arithmetic Operations on Matrices

In this section, we will introduce additional operations that can be performed on matrices. We will also see that, while the sets $M_{m \times n}(\mathbb{R})$ and $M_{m \times n}(\mathbb{C})$ are very similar to $\mathbb{R}^n$ and $\mathbb{C}^n$, in certain respects, there also exist quite a few important distinctions between them.

Definition 4.3.1
Sum of Matrices

Let $A, B \in M_{m \times n}(\mathbb{F})$. We define the **matrix sum** $A + B = C$ to be the matrix $C \in M_{m \times n}(\mathbb{F})$ whose $(i, j)^{th}$ entry is $c_{ij} = a_{ij} + b_{ij}$, for all $i = 1, \ldots, m$ and $j = 1, \ldots, n$.

REMARK

Addition is not defined for matrices that do not have the same size.

Example 4.3.2

For $A = \begin{bmatrix} 0 & 2 \\ 3 & 0 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 0 \\ 0 & -4 \end{bmatrix}$, we have the sum $A + B = \begin{bmatrix} 0 & 2 \\ 3 & 0 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & -4 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 3 & -4 \end{bmatrix}$.

Some basic properties of matrix addition and matrix multiplication are given in the next two propositions.

Proposition 4.3.3

(Matrix Addition)

If $A, B, C \in M_{m\times n}(\mathbb{F})$, then

(a) $A + B = B + A$.

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

Proposition 4.3.4

(Matrix Multiplication)

If $A, B \in M_{m\times n}(\mathbb{F})$, $C, D \in M_{n\times p}(\mathbb{F})$ and $E \in M_{p\times q}(\mathbb{F})$, then

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

(b) $A(C + D) = AC + AD$.

(c) $(AC)E = A(CE) = ACE$.

EXERCISE

Prove the previous two Propositions.

Definition 4.3.5
Additive Inverse

Let $A \in M_{m\times n}(\mathbb{F})$. We define the **additive inverse of** A to be the matrix $-A$ whose $(i, j)^{th}$ entry is $-a_{ij}$ for all $i = 1, \ldots, m$ and $j = 1, \ldots, n$.

Example 4.3.6

If $A = \begin{bmatrix} 1 & 2 \\ 3 & -4 \end{bmatrix}$, then the additive inverse of A is $-A = \begin{bmatrix} -1 & -2 \\ -3 & 4 \end{bmatrix}$.

Definition 4.3.7
Zero Matrix

The $m \times n$ **zero matrix** is the matrix $\mathcal{O}_{m\times n} \in M_{m\times n}(\mathbb{F})$ all of whose entries are 0.

REMARK

1. It is important to distinguish the zero matrix $\mathcal{O}_{m\times n}$ from the zero scalar 0 and the zero vector $\vec{0}$.
2. Usually, the context in which the zero matrix is used is sufficient for us to determine its size, in which case we, denote it by $\mathcal{O}$.

Example 4.3.8

We have $\mathcal{O}_{2\times 3} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$ and $\mathcal{O}_{1\times 1} = [0]$. Note that $\mathcal{O}_{1\times 1}$ is technically different from the zero scalar 0.

On the other hand, $\mathcal{O}_{m\times 1} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}$ is the zero vector in $\mathbb{F}^m$.

Proposition 4.3.9

(Properties of the Additive Inverse and the Zero Matrix)

If $A \in M_{m\times n}(\mathbb{F})$, then

(a) $\mathcal{O} + A = A + \mathcal{O} = A$.

(b) $A + (-A) = (-A) + A = \mathcal{O}$.

EXERCISE

Prove the previous Proposition.

Definition 4.3.10

Multiplication of a Matrix by a Scalar

Let $A \in M_{m\times n}(\mathbb{F})$ and $c \in \mathbb{F}$. We define the **product of c and A** to be the matrix $cA \in M_{m\times n}(\mathbb{F})$ whose $(i,j)^{th}$ entry is $(cA)_{ij} = ca_{ij}$ for all $i = 1,\dots,m$ and $j = 1,\dots,n$.

Notice that the product $(-1)A$ (of the scalar -1 and the matrix A) is $-A$, the additive inverse of A. That is,

$$(-1)A = -A.$$

So, our notational choices are compatible. The following proposition gives some additional properties of scalar multiplication.

Proposition 4.3.11

(Properties of Multiplication of a Matrix by a Scalar)

If $A, B \in M_{m\times n}(\mathbb{F})$, $C \in M_{n\times k}(\mathbb{F})$, and $r, s \in \mathbb{F}$, then

(a) $s(A + B) = sA + sB$.

(b) $(r+s)A = rA + sA$.

(c) $r(sA) = (rs)A$.

(d) $s(AC) = (sA)C = A(sC)$. Thus, we may write sAC for any of these three quantities unambiguously.

EXERCISE

Prove the previous Proposition.

Our previous results demonstrate that matrix operations behave like their scalar counterparts. However, there are significant differences. For instance, while it is true that $ab = ba$ for all $a, b \in \mathbb{F}$, we have already seen an example of matrices A and B where $AB \neq BA$, even when both AB and BA are defined. (See Example 4.2.6.)

Two other results that apply to scalars which students are tempted to apply to matrices are the following:

1. If $ab = ac$ and $a \neq 0$, then $b = c$ (cancellation law).
2. $ab = 0$ if and only if $a = 0$ or $b = 0$.

These results do not extend to matrix arithmetic, in general.

Example 4.3.12

1. Let $A = \begin{bmatrix} 0 & 1 \\ 0 & 2 \end{bmatrix}$, $B = \begin{bmatrix} 1 & 1 \\ 3 & 4 \end{bmatrix}$, and $C = \begin{bmatrix} 2 & 5 \\ 3 & 4 \end{bmatrix}$. Then $AB = AC = \begin{bmatrix} 3 & 4 \\ 6 & 8 \end{bmatrix}$ and $A \neq \mathcal{O}$, but $B \neq C$. Thus, the cancellation law does not apply to matrices.

2. Let $A = \begin{bmatrix} 0 & 1 \\ 0 & 2 \end{bmatrix}$ and $B = \begin{bmatrix} 3 & 7 \\ 0 & 0 \end{bmatrix}$. Then $AB = \mathcal{O}$, but $A \neq \mathcal{O}$ and $B \neq \mathcal{O}$.

In Section 4.1 we were introduced to the transpose A^T of a matrix $A \in M_{m\times n}(\mathbb{F})$. Recall that $(A^T)_{ij} = (A)_{ji}$. Our next Proposition describes the interplay between transpose and the other operations on matrices.

Proposition 4.3.13

(Properties of Matrix Transpose)

If $A, B \in M_{m\times n}(\mathbb{F})$, $C \in M_{n\times k}(\mathbb{F})$, and $s \in \mathbb{F}$, then

(a) $(A+B)^T = A^T + B^T$.

(b) $(sA)^T = sA^T$.

(c) $(AC)^T = C^T A^T$.

(d) $(A^T)^T = A$.

REMARK

Pay close attention to Property (c). The transpose of a product is the product of the transposes *in reverse*!

EXERCISE

Prove the previous Proposition.

4.4 Properties of Square Matrices

We turn our attention to matrices that have the same number of rows as columns.

Definition 4.4.1 **Square Matrix**

A matrix $A \in M_{n\times n}(\mathbb{F})$ where the number of rows is equal to the number of columns is called a **square matrix** .

REMARK

For the case where $m = n$, it is common to refer to $M_{n\times n}(\mathbb{F})$ as $M_n(\mathbb{F})$.

Example 4.4.2

The matrix $\begin{bmatrix} 3 & 2 & 1 \\ 1 & 1 & 1 \end{bmatrix}$ is not a square matrix, since it has 2 rows and 3 columns. The matrix $\begin{bmatrix} -2+i & 3-2i \\ 2+2i & 6-7i \end{bmatrix}$ is a square matrix, since it has 2 rows and 2 columns.

Definition 4.4.3 **Upper Triangular**

Let $A \in M_{n\times n}(\mathbb{F})$. We say that A is **upper triangular** if $a_{ij} = 0$ for $i > j$ with $i = 1, \ldots, n$ and $j = 1, \ldots, n$.

Example 4.4.4

The matrices $\begin{bmatrix} 1+i & -7 \\ 0 & 1-i \end{bmatrix}$, $\begin{bmatrix} 3 & -4 & 6 \\ 0 & 5 & 9 \\ 0 & 0 & 7 \end{bmatrix}$, and $\begin{bmatrix} i & 4-i & 6-4i \\ 0 & 1+i & 7i \\ 0 & 0 & 2-4i \end{bmatrix}$ are upper-triangular.

Definition 4.4.5 **Lower Triangular**

Let $B \in M_{n\times n}(\mathbb{F})$. We say that B is **lower triangular** if $b_{ij} = 0$ for $i < j$ with $i = 1, \ldots, n$ and $j = 1, \ldots, n$.

Example 4.4.6

The matrices $\begin{bmatrix} i & 0 \\ 4 & 1-i \end{bmatrix}$, $\begin{bmatrix} 3 & 0 & 0 \\ 3 & 4 & 0 \\ 0 & 7 & 0 \end{bmatrix}$, and $\begin{bmatrix} i & 0 & 0 & 0 \\ 7i & 1+i & 0 & 0 \\ 6 & 2+i & 2-4i & 0 \\ 3-5i & 5 & 0 & 4+3i \end{bmatrix}$ are lower triangular.

REMARKS

- The transpose of an upper (lower) triangular matrix is a lower (upper) triangular matrix, respectively.
- The product of upper (lower) triangular $n \times n$ matrices is upper (lower) triangular, respectively.

Definition 4.4.7 **Diagonal**

We say that an $n \times n$ matrix A is **diagonal** if $a_{ij} = 0$ for $i \neq j$ with $i = 1, \ldots, n$ and $j = 1, \ldots, n$. We refer to the entries a_{ii} as the **diagonal entries** of A, and denote our matrix A by $A = \operatorname{diag}(a_{11}, a_{22}, \ldots, a_{nn})$.

Example 4.4.8

The matrices $\begin{bmatrix} i & 0 \\ 0 & 1-i \end{bmatrix}$, $\begin{bmatrix} 3 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 0 \end{bmatrix}$, and $\begin{bmatrix} i & 0 & 0 & 0 \\ 0 & 1+i & 0 & 0 \\ 0 & 0 & 2-4i & 0 \\ 0 & 0 & 0 & 4+3i \end{bmatrix}$ are diagonal.

Example 4.4.9

If $C = \operatorname{diag}(1, -2, 3)$, then $C = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 3 \end{bmatrix}$.

REMARKS

- Perhaps the most simple example of a diagonal matrix is the $n \times n$ zero matrix $\mathcal{O}_{n \times n}$.
- All diagonal matrices are also upper triangular and lower triangular matrices.

Another simple (and important) example is the identity matrix, which we define as follows.

Definition 4.4.10 **Identity Matrix**

The diagonal matrix $\operatorname{diag}(1, 1, \ldots, 1)$ is called the **identity matrix**, and is denoted by I. If we wish to indicate that the size of the identity matrix is $n \times n$, we add a subscript n and write I_n.

Example 4.4.11

$I_1 = [1]$, $I_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, $I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$, and $I_4 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$.

REMARK

The identity matrix behaves as a **multiplicative identity**: for any $A \in M_{m\times n}(\mathbb{F})$ we have that

$$I_m A = A \quad \text{and} \quad AI_n = A.$$

In particular, this also holds for vectors in $\mathbb{F}^n$, so $I_n \vec{x} = \vec{x}$ for all $\vec{x} \in \mathbb{F}^n$.

When the identity matrix is multiplied by a scalar $c \in \mathbb{F}$, the result is a diagonal matrix whose diagonal entries are all equal to c:

$$cI_n = \mathrm{diag}(c, c, \ldots, c).$$

Example 4.4.12

If $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$, then $AI_2 = I_2 A = A$.

If $B = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 4 & 0 \end{bmatrix}$, then $BI_3 = I_2 B = B$.

Example 4.4.13

If $c \in \mathbb{F}$, then $cI_2 = \begin{bmatrix} c & 0 \\ 0 & c \end{bmatrix}$.

4.5 Elementary Matrices

In this section we will show that performing an elementary row operation (ERO) on a matrix is equivalent to multiplying that matrix on the left by a square matrix of a certain type.

Definition 4.5.1
Elementary Matrices

A matrix that can be obtained by performing a *single* ERO on the identity matrix is called an **elementary matrix**.

If we wish to be more precise, we refer to elementary matrices using the same names as we have for EROs.

Example 4.5.2

The following matrices are examples of elementary matrices:

- $E_1 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$ is a row swap elementary matrix, since performing $R_1 \leftrightarrow R_3$ on I_3 will produce E_1.
- $E_2 = \begin{bmatrix} 5 & 0 \\ 0 & 1 \end{bmatrix}$ is a row scale elementary matrix, since performing $R_1 \to 5R_1$ on I_2 will produce E_2.

- $E_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -2 \\ 0 & 0 & 1 \end{bmatrix}$ is a row addition elementary matrix, since performing $R_2 \to -2R_3 + R_2$ on I_3 will produce E_3.

Elementary matrices are useful because they can be used to carry out EROs. The next two results make this statement precise.

Proposition 4.5.3 Let $A \in M_{m\times n}(\mathbb{F})$ and suppose that a single ERO is performed on it to produce matrix B. Suppose, also, that we perform the same ERO on the matrix I_m to produce the elementary matrix E. Then

$$B = EA.$$

Proof: We will prove the result only for row swap elementary matrices and leave the rest as an exercise. Let A be an $m \times n$ matrix with $m > 1$, and consider its transpose A^T. The columns of A^T are $\overrightarrow{r_1}^T, \ldots, \overrightarrow{r_m}^T$, where $\overrightarrow{r_i} = \overrightarrow{\text{row}_i}(A)$ is the i^{th} row of A. Thus

$$A^T = \begin{bmatrix} \overrightarrow{r_1}^T & \cdots & \overrightarrow{r_i}^T & \cdots & \overrightarrow{r_j}^T & \cdots & \overrightarrow{r_m}^T \end{bmatrix}.$$

Interchanging rows i and j of A yields the matrix B whose transpose is

$$B^T = \begin{bmatrix} \overrightarrow{r_1}^T & \cdots & \overrightarrow{r_j}^T & \cdots & \overrightarrow{r_i}^T & \cdots & \overrightarrow{r_m}^T \end{bmatrix}.$$

On the other hand, let

$$E = \begin{bmatrix} \overrightarrow{e_1} & \cdots & \overrightarrow{e_j} & \cdots & \overrightarrow{e_i} & \cdots & \overrightarrow{e_m} \end{bmatrix}$$

be the elementary matrix that corresponds to the row swap $R_i \leftrightarrow R_j$. One can quickly show that $E = E^T$; it then follows from Lemma 4.2.2 (Column Extraction) that

$$\begin{aligned}
(EA)^T &= A^T E^T \\
&= A^T E \\
&= A^T \begin{bmatrix} \overrightarrow{e_1} & \cdots & \overrightarrow{e_j} & \cdots & \overrightarrow{e_i} & \cdots & \overrightarrow{e_m} \end{bmatrix} \\
&= \begin{bmatrix} A^T\overrightarrow{e_1} & \cdots & A^T\overrightarrow{e_j} & \cdots & A^T\overrightarrow{e_i} & \cdots & A^T\overrightarrow{e_m} \end{bmatrix} \\
&= \begin{bmatrix} \overrightarrow{r_1}^T & \cdots & \overrightarrow{r_j}^T & \cdots & \overrightarrow{r_i}^T & \cdots & \overrightarrow{r_m}^T \end{bmatrix} \\
&= B^T
\end{aligned}$$

Since $B^T = (EA)^T$, we see that $B = EA$ by Property (d) in Proposition 4.3.13 (Properties of Matrix Transpose). □

The previous Proposition tells us that we can perform an ERO on A by multiplying A on the left by the appropriate elementary matrix E. We can also carry out a sequence of EROs by multiplying A on the left by a sequence of appropriate elementary matrices.

Corollary 4.5.4 Let $A \in M_{m\times n}(\mathbb{F})$ and suppose that a finite number of EROs, numbered 1 through k, are performed on A to produce a matrix B. Let E_i denote the elementary matrix corresponding to the i^{th} ERO $(1 \leq i \leq k)$ applied to I_m. Then

$$B = E_k \ldots E_2 E_1 A.$$

Proof: Use the Principle of Mathematical Induction on k and Proposition 4.5.3. □

Pay attention to the order in which the elementary matrices appear in this product.

Example 4.5.5 Let us revisit Example 3.7.1 in Section 3.7. The augmented matrix, B, was given by

$$B = \left[\begin{array}{cccc|c} 1 & -2 & -1 & 3 & 1 \\ 2 & -4 & 1 & 0 & 5 \\ 1 & -2 & 2 & -3 & 4 \end{array}\right].$$

We perform the following EROs:

$$\begin{cases} R_2 \to -2R_1 + R_2 \\ R_3 \to -R_1 + R_3 \end{cases} \text{gives} \left[\begin{array}{cccc|c} 1 & -2 & -1 & 3 & 1 \\ 0 & 0 & 3 & -6 & 3 \\ 0 & 0 & 3 & -6 & 3 \end{array}\right] = C,$$

$$R_2 \to \tfrac{1}{3}R_2 \text{ gives } \left[\begin{array}{cccc|c} 1 & -2 & -1 & 3 & 1 \\ 0 & 0 & 1 & -2 & 1 \\ 0 & 0 & 3 & -6 & 3 \end{array}\right] = D,$$

$$R_3 \to -3R_2 + R_3 \text{ gives } \left[\begin{array}{cccc|c} 1 & -2 & -1 & 3 & 1 \\ 0 & 0 & 1 & -2 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right] = F.$$

For each of the four EROs, we give the corresponding elementary matrix:

$$R_2 \to -2R_1 + R_2 \text{ gives } \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = E_1,$$

$$R_3 \to -R_1 + R_3 \text{ gives } \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix} = E_2.$$

Notice that $C = E_2E_1B$. Further,

$$R_2 \to \tfrac{1}{3}R_2 \text{ gives } \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{3} & 0 \\ 0 & 0 & 1 \end{bmatrix} = E_3,$$

$$R_3 \to -3R_2 + R_3 \text{ gives } \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -3 & 1 \end{bmatrix} = E_4.$$

Notice that $D = E_3C$ and $F = E_4D$.

We can now verify our result by evaluating the product $E_4E_3E_2E_1B = E_4(E_3(E_2(E_1B)))$:

$$\begin{bmatrix}1&0&0\\0&1&0\\0&-3&1\end{bmatrix}\begin{bmatrix}1&0&0\\0&\frac{1}{3}&0\\0&0&1\end{bmatrix}\begin{bmatrix}1&0&0\\0&1&0\\-1&0&1\end{bmatrix}\begin{bmatrix}1&0&0\\-2&1&0\\0&0&1\end{bmatrix}\left[\begin{array}{cccc|c}1&-2&-1&3&1\\2&-4&1&0&5\\1&-2&2&-3&4\end{array}\right]$$
$$=\begin{bmatrix}1&0&0\\0&1&0\\0&-3&1\end{bmatrix}\begin{bmatrix}1&0&0\\0&\frac{1}{3}&0\\0&0&1\end{bmatrix}\begin{bmatrix}1&0&0\\0&1&0\\-1&0&1\end{bmatrix}\left[\begin{array}{cccc|c}1&-2&-1&3&1\\0&0&3&-6&3\\1&-2&2&-3&4\end{array}\right]$$
$$=\begin{bmatrix}1&0&0\\0&1&0\\0&-3&1\end{bmatrix}\begin{bmatrix}1&0&0\\0&\frac{1}{3}&0\\0&0&1\end{bmatrix}\left[\begin{array}{cccc|c}1&-2&-1&3&1\\0&0&3&-6&3\\0&0&3&-6&3\end{array}\right]$$
$$=\begin{bmatrix}1&0&0\\0&1&0\\0&-3&1\end{bmatrix}\left[\begin{array}{cccc|c}1&-2&-1&3&1\\0&0&1&-2&1\\0&0&3&-6&3\end{array}\right]$$
$$=\left[\begin{array}{cccc|c}1&-2&-1&3&1\\0&0&1&-2&1\\0&0&0&0&0\end{array}\right]=F.$$

4.6 Matrix Inverse

Let A be an $m \times n$ matrix. In this section, we aim to explore the following two questions:

- Does there exist an $n \times m$ matrix B such that, for all $\vec{x} \in \mathbb{F}^m$, $AB\vec{x} = \vec{x}$?
- Does there exist an $n \times m$ matrix C such that, for all $\vec{x} \in \mathbb{F}^n$, $CA\vec{x} = \vec{x}$?

In the first case, you can think of the matrix A as the one that "cancels out" the action of B. In the second case, you can think of the matrix C as the one that "cancels out" the action of A.

In view of the Theorem 4.2.3 (Equality of Matrices), let us reformulate our two motivating questions as follows:

- Given an $m \times n$ matrix A, does there exist an $n \times m$ matrix B such that $AB = I_m$?
- Given an $m \times n$ matrix A, does there exist an $n \times m$ matrix C such that $CA = I_n$?

If the matrices B and C exist, we refer to them as a *right inverse* and a *left inverse* of A, respectively. Notice that if B and C exist and all of A, B and C are square, then $AB = CA = I_n$. This case is the most interesting to us, and it motivates the following definition.

Definition 4.6.1
Invertible Matrix

We say that an $n \times n$ matrix A is **invertible** if there exist $n \times n$ matrices B and C such that $AB = CA = I_n$.

The following theorem demonstrates that *if* an $n \times n$ matrix A is invertible, then its left and right inverses are equal.

Proposition 4.6.2 **(Equality of Left and Right Inverses)**

Let $A \in M_{n\times n}(\mathbb{F})$. If there exist matrices B and C in $M_{n\times n}(\mathbb{F})$ such that $AB = CA = I_n$, then $B = C$.

Proof: We have $B = I_nB = (CA)B = C(AB) = CI_n = C$. □

So far, we have learned that if *both* left and right inverses exist, then they must be equal. The next result guarantees that for square matrices the left inverse exists if and only if the right inverse exists.

Theorem 4.6.3 **(Left Invertible Iff Right Invertible)**

For $A \in M_{n\times n}(\mathbb{F})$, there exists an $n \times n$ matrix B such that $AB = I_n$ if and only if there exists an $n \times n$ matrix C such that $CA = I_n$.

Proof: Let A be an $n \times n$ matrix.

To prove the forward direction, suppose that there exists an $n \times n$ matrix B such that $AB = I_n$. Consider the homogeneous system

$$B\vec{x} = \vec{0}.$$

Multiplying both sides of this equation by A on the left, we have that

$$AB\vec{x} = A\vec{0} \implies I_n\vec{x} = \vec{0} \implies \vec{x} = \vec{0}.$$

Thus, the homogeneous system $B\vec{x} = \vec{0}$ has the unique solution $\vec{x} = \vec{0}$. Therefore, its solution set has no parameters, and so it follows from Part (a) of Theorem 3.6.7 (System Rank Theorem) that $\text{rank}(B) = n$. It then follows from Part (b) of Theorem 3.6.7 (System Rank Theorem) that the system $B\vec{x} = \vec{b}$ has a solution for every $\vec{b} \in \mathbb{F}^n$. Thus, for any $\vec{b} \in \mathbb{F}^n$, there exists $\vec{x} \in \mathbb{F}^n$ such that $\vec{b} = B\vec{x}$. Consequently,

$$(BA)\vec{b} = BA(B\vec{x}) = B(AB)\vec{x} = BI_n\vec{x} = B\vec{x} = \vec{b} = I_n\vec{b}.$$

Since the identity $(BA)\vec{b} = I_n\vec{b}$ holds for all $\vec{b} \in \mathbb{F}^n$, it follows from Theorem 4.2.3 (Equality of Matrices) that $BA = I_n$. Taking $C = B$, the result follows.

To prove the backward direction, suppose that there exists an $n \times n$ matrix C such that $CA = I_n$. Then

$$I_n = I_n^T = (CA)^T = A^TC^T.$$

Thus, C^T is the right inverse of A^T, and so it follows from the proof of the forward direction that C^T is also the left inverse of A^T, that is,

$$C^TA^T = I_n.$$

But then

$$I_n = I_n^T = (C^TA^T)^T = (A^T)^T(C^T)^T = AC.$$

Taking $B = C$, the result follows. □

Notice that if a matrix A is invertible, then there exists a *unique* matrix B such that $AB = I_n$. Indeed, suppose that $AB_1 = I_n$ and $AB_2 = I_n$. Since $AB_1 = AB_2$, we see that $B_1AB_1 = B_1AB_2$. By Theorem 4.6.3 (Left Invertible Iff Right Invertible), we know that $B_1A = I_n$, so $B_1 = B_2$. In view of this, consider the following definition.

Definition 4.6.4
Inverse of a Matrix

If an $n \times n$ matrix A is invertible, we refer to the matrix B such that $AB = I_n$ as the **inverse** of A. We denote the inverse of A by A^{-1}. The inverse of A satisfies

$$AA^{-1} = A^{-1}A = I_n.$$

REMARK

The above results tell us that, in order to verify that the matrix B is the inverse of A, it is sufficient to verify that $AB = I_n$. We do not need to also verify that $BA = I_n$.

Example 4.6.5

The matrix $B = \begin{bmatrix} 1+i & i \\ 1 & i \end{bmatrix}$ is the inverse of $A = \begin{bmatrix} -i & i \\ 1 & -1-i \end{bmatrix}$, because

$$AB = \begin{bmatrix} -i & i \\ 1 & -1-i \end{bmatrix} \begin{bmatrix} 1+i & i \\ 1 & i \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Example 4.6.6

The matrix $B = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}$ is **not** the inverse of $A = \begin{bmatrix} 2 & -1 \\ -2 & 2 \end{bmatrix}$, because

$$AB = \begin{bmatrix} 2 & -1 \\ -2 & 2 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} \neq \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

While there exist numerous criteria for the invertibility of a matrix, for now we will outline only three of them.

Theorem 4.6.7

(Invertibility Criteria – First Version)

Let $A \in M_{n\times n}(\mathbb{F})$. The following three conditions are equivalent:

(a) A is invertible.

(b) $\operatorname{rank}(A) = n$.

(c) $\operatorname{RREF}(A) = I_n$.

Proof: $((a) \Rightarrow (b))$:
Suppose that A is invertible. Then there exists a matrix B such that $BA = I_n$. We can show that $\operatorname{rank}(A) = n$ as in the proof of Theorem 4.6.3 (Left Invertible Iff Right Invertible).

$((b) \Rightarrow (c))$:
Suppose that $\operatorname{rank}(A) = n$. Then every row and every column of the $n \times n$ matrix $\operatorname{RREF}(A)$ has a pivot. The only matrix that can satisfy this condition is the $n \times n$ identity matrix I_n.

$((c) \Rightarrow (a))$:
Suppose that $\operatorname{RREF}(A) = I_n$. Then the rank of A is equal to the number of rows of A, which means that the system $A\vec{x} = \vec{b}$ is consistent for every $\vec{b} \in \mathbb{F}^n$. Thus, we let $\vec{b_i}$ be a solution to $A\vec{x} = \vec{e_i}$ for $i = 1, \ldots, n$. If we now define $B = \begin{bmatrix} \vec{b_1} & \cdots & \vec{b_n} \end{bmatrix}$, then by construction

$$AB = A\begin{bmatrix} \vec{b_1} & \cdots & \vec{b_n} \end{bmatrix} = \begin{bmatrix} A\vec{b_1} & \cdots & A\vec{b_n} \end{bmatrix} = \begin{bmatrix} \vec{e_1} & \cdots & \vec{e_n} \end{bmatrix} = I_n$$

From Theorem 4.6.3 (Left Invertible Iff Right Invertible) it follows that there exists a matrix C such that $CA = I_n$. Since $AB = CA = I_n$, we conclude that A is invertible. □

Let us look closely at the last part of the proof, as it describes an algorithm for finding the inverse of an invertible matrix A. Indeed, if we are able to solve each of the equations

$$A\vec{b_1} = \vec{e_1}, \quad A\vec{b_2} = \vec{e_2}, \quad \ldots, \quad A\vec{b_n} = \vec{e_n}$$

for $\vec{b_1}, \ldots, \vec{b_n} \in \mathbb{F}^n$, then $A^{-1} = \begin{bmatrix} \vec{b_1} & \cdots & \vec{b_n} \end{bmatrix}$. Let $R = \operatorname{RREF}(A)$, and suppose that a sequence of EROs $r_1, \ldots, r_k$ is used to row reduce the augmented matrix A to R. Convince yourself that the same sequence of EROs can be used to row reduce *any* system $[A \mid \vec{e_i}]$ to its RREF, $[R \mid \vec{b_i}]$, for some vector $\vec{b_i}$. Consequently, the same sequence of EROs can be used to row reduce the *super-augmented matrix*

$$[A \mid I_n] = \left[\begin{array}{c|ccc} A & \vec{e_1} & \cdots & \vec{e_n} \end{array}\right]$$

into its RREF

$$[R \mid B] = \left[\begin{array}{c|ccc} R & \vec{b_1} & \cdots & \vec{b_n} \end{array}\right]$$

for some $n \times n$ matrix $B = \begin{bmatrix} \vec{b_1} & \cdots & \vec{b_n} \end{bmatrix}$. Notice how by performing row reduction on a super-augmented matrix, we are solving n distinct systems of linear equations *simultaneously*. Of course, this is possible because these systems $A\vec{b_i} = \vec{e_i}$ have equal coefficient matrices. By Theorem 3.6.7 (System Rank Theorem), we have that

- If $\operatorname{rank}(A) \neq n$ (which is equivalent to $R \neq I_n$), there exists an index i such that the system $A\vec{b_i} = \vec{e_i}$ is inconsistent (by Proposition 3.10.2). Consequently, it is impossible to find a matrix B such that $AB = I_n$, which means that A is not invertible.
- If $\operatorname{rank}(A) = n$ (which is equivalent to $R = I_n$), then each system $A\vec{b_i} = \vec{e_i}$ is consistent, and so the matrix B is the inverse A.

We summarize our observations in the following proposition.

Proposition 4.6.8 **(Algorithm for Checking Invertibility and Finding the Inverse)**

The following algorithm allows you to determine whether an $n \times n$ matrix A is invertible, and if it is, the algorithm will provide the inverse of A.

1. Construct a super-augmented matrix $[A \mid I_n]$.
2. Find the RREF, $[R \mid B]$, of $[A \mid I_n]$.
3. If $R \neq I_n$, conclude that A is not invertible. If $R = I_n$, conclude that A is invertible, and that $A^{-1} = B$.

Example 4.6.9 Determine whether the matrix

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$$

is invertible. If it is invertible, find its inverse.

Solution: In order to find the inverse, we need to solve two systems of equations with augmented matrices $[A \mid \vec{e_1}]$ and $[A \mid \vec{e_2}]$. Instead of solving them separately, we will solve them *simultaneously* by forming a super-augmented matrix

$$[A \mid I_2] = \left[\begin{array}{c|cc} A & \vec{e_1} & \vec{e_2} \end{array}\right] = \left[\begin{array}{cc|cc} 1 & 2 & 1 & 0 \\ 3 & 4 & 0 & 1 \end{array}\right]$$

and row reducing it. We have

$$R_2 \to R_2 - 3R_1 \quad \text{gives} \quad \left[\begin{array}{cc|cc} 1 & 2 & 1 & 0 \\ 0 & -2 & -3 & 1 \end{array}\right].$$

Notice that the matrix that we've obtained is in REF. Since it has two pivots, we have that $\operatorname{rank}(A) = 2$, so it must be the case that A is invertible. We continue reducing to RREF.

$$R_2 \to -\tfrac{1}{2}R_2 \quad \text{gives} \quad \left[\begin{array}{cc|cc} 1 & 2 & 1 & 0 \\ 0 & 1 & \frac{3}{2} & -\frac{1}{2} \end{array}\right],$$

$$R_1 \to R_1 - 2R_2 \quad \text{gives} \quad \left[\begin{array}{cc|cc} 1 & 0 & -2 & 1 \\ 0 & 1 & \frac{3}{2} & -\frac{1}{2} \end{array}\right].$$

Thus, we conclude that the inverse of A is

$$A^{-1} = \begin{bmatrix} -2 & 1 \\ \frac{3}{2} & -\frac{1}{2} \end{bmatrix} = -\frac{1}{2}\begin{bmatrix} 4 & -2 \\ -3 & 1 \end{bmatrix}.$$

Let us verify our calculations:

$$\begin{aligned} A^{-1}A &= -\frac{1}{2}\begin{bmatrix} 4 & -2 \\ -3 & 1 \end{bmatrix}\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} = -\frac{1}{2}\begin{bmatrix} 4 \cdot 1 + (-2) \cdot 3 & 4 \cdot 2 + (-2) \cdot 4 \\ (-3) \cdot 1 + 1 \cdot 3 & (-3) \cdot 2 + 1 \cdot 4 \end{bmatrix} \\ &= -\frac{1}{2}\begin{bmatrix} -2 & 0 \\ 0 & -2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}. \end{aligned}$$

Example 4.6.10 Determine whether the matrix

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

is invertible. If it is invertible, find its inverse.

Solution: We consider the super-augmented matrix $[B \mid I_3]$ and reduce it to a matrix in REF.

$$\begin{cases} R_2 \to R_2 - 4R_1 \\ R_3 \to R_3 - 7R_1 \end{cases} \quad \text{gives} \quad \left[\begin{array}{rrr|rrr} 1 & 2 & 3 & 1 & 0 & 0 \\ 0 & -3 & -6 & -4 & 1 & 0 \\ 0 & -6 & -12 & -7 & 0 & 1 \end{array}\right],$$

$$R_2 \to -\frac{1}{3}R_2 \quad \text{gives} \quad \left[\begin{array}{rrr|rrr} 1 & 2 & 3 & 1 & 0 & 0 \\ 0 & 1 & 2 & \frac{4}{3} & -\frac{1}{3} & 0 \\ 0 & -6 & -12 & -7 & 0 & 1 \end{array}\right],$$

$$R_3 \to R_3 + 6R_2 \quad \text{gives} \quad \left[\begin{array}{rrr|lrr} 1 & 2 & 3 & 1 & 0 & 0 \\ 0 & 1 & 2 & \frac{4}{3} & -\frac{1}{3} & 0 \\ 0 & 0 & 0 & 1 & -2 & 1 \end{array}\right].$$

At this point, we have that $\operatorname{rank}(B) = 2$, and so it is not invertible.

Example 4.6.11 Determine whether the matrix

$$C = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 10 \end{bmatrix}$$

is invertible. If it is invertible, find its inverse.

Solution: We consider the super-augmented matrix $[C \mid I_3]$ and row reduce it.

$$\begin{cases} R_2 \to R_2 - 4R_1 \\ R_3 \to R_3 - 7R_1 \end{cases} \quad \text{gives} \quad \left[\begin{array}{rrr|rrr} 1 & 2 & 3 & 1 & 0 & 0 \\ 0 & -3 & -6 & -4 & 1 & 0 \\ 0 & -6 & -11 & -7 & 0 & 1 \end{array}\right],$$

$$R_2 \to -\frac{1}{3}R_2 \quad \text{gives} \quad \left[\begin{array}{rrr|rrr} 1 & 2 & 3 & 1 & 0 & 0 \\ 0 & 1 & 2 & \frac{4}{3} & -\frac{1}{3} & 0 \\ 0 & -6 & -11 & -7 & 0 & 1 \end{array}\right],$$

$$R_3 \to R3 + 6R_2 \quad \text{gives} \quad \left[\begin{array}{rrr|lrr} 1 & 2 & 3 & 1 & 0 & 0 \\ 0 & 1 & 2 & \frac{4}{3} & -\frac{1}{3} & 0 \\ 0 & 0 & 1 & 1 & -2 & 1 \end{array}\right].$$

At this point, we have that $\operatorname{rank}(C) = 3$, and so it is invertible. We thus continue row reducing it to RREF.

$$\begin{cases} R_1 \to R_1 - 3R_3 \\ R_2 \to R_2 - 2R_3 \end{cases} \quad \text{gives} \quad \left[\begin{array}{rrr|rrr} 1 & 2 & 0 & -2 & 6 & -3 \\ 0 & 1 & 0 & -\frac{2}{3} & \frac{11}{3} & -2 \\ 0 & 0 & 1 & 1 & -2 & 1 \end{array}\right],$$

$$R_1 \to R_1 - 2R_2 \quad \text{gives} \quad \left[\begin{array}{rrr|rrr} 1 & 0 & 0 & -\frac{2}{3} & -\frac{4}{3} & 1 \\ 0 & 1 & 0 & -\frac{2}{3} & \frac{11}{3} & -2 \\ 0 & 0 & 1 & 1 & -2 & 1 \end{array}\right].$$

We conclude that the inverse of C is

$$C^{-1} = \begin{bmatrix} -\frac{2}{3} & -\frac{4}{3} & 1 \\ -\frac{2}{3} & \frac{11}{3} & -2 \\ 1 & -2 & 1 \end{bmatrix} = \frac{1}{3}\begin{bmatrix} -2 & -4 & 3 \\ -2 & 11 & -6 \\ 3 & -6 & 3 \end{bmatrix}.$$

Let us verify our calculations.

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 10 \end{bmatrix} \left(\frac{1}{3}\right) \begin{bmatrix} -2 & -4 & 3 \\ -2 & 11 & -6 \\ 3 & -6 & 3 \end{bmatrix} = \frac{1}{3} \begin{bmatrix} 3 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 3 \end{bmatrix} = I_3$$

Example 4.6.12 Determine whether the matrix

$$D = \begin{bmatrix} -1-2\,i & -1-i \\ 1-i & -i \end{bmatrix}$$

is invertible. If it is invertible, find its inverse.

Solution We consider the super-augmented matrix $[D \mid I_2]$ and row reduce it.

$$[D \mid I_2] = \left[\begin{array}{cc|cc} -1-2\,i & -1-i & 1 & 0 \\ 1-i & -i & 0 & 1 \end{array}\right]$$

$$\begin{cases} R_1 \to (-1+2i)R_1 \\ R_2 \to (1+i)R_2 \end{cases} \text{ gives } \left[\begin{array}{cc|cc} 5 & 3-i & -1+2\,i & 0 \\ 2 & 1-i & 0 & 1+i \end{array}\right]$$

$$R_2 \to -2R_1 + 2R_2 \text{ gives } \left[\begin{array}{cc|cc} 5 & 3-i & -1+2\,i & 0 \\ 0 & -1-3\,i & 2-4\,i & 5+5\,i \end{array}\right]$$

At this point, we have that $\operatorname{rank}(D) = 2$ and so D is invertible. We thus continue row reducing to RREF.

$$R_2 \to iR_2 \text{ gives } \left[\begin{array}{cc|cc} 5 & 3-i & -1+2\,i & 0 \\ 0 & 3-i & 4+2\,i & -5+5\,i \end{array}\right]$$

$$R_1 \to R_1 - R_2 \text{ gives } \left[\begin{array}{cc|cc} 5 & 0 & -5 & 5-5\,i \\ 0 & 3-i & 4+2\,i & -5+5\,i \end{array}\right]$$

As $\dfrac{1}{3-i} = \dfrac{3+i}{9+1} = \dfrac{3+i}{10}$, we complete our reduction with

$$\begin{cases} R_1 \to \frac{1}{5}R_1 \\ R_2 \to \frac{1}{10}(3+i)R_2 \end{cases} \text{ gives } \left[\begin{array}{cc|cc} 1 & 0 & -1 & 1-i \\ 0 & 1 & 1+i & -2+i \end{array}\right]$$

Therefore, we conclude that

$$D^{-1} = \begin{bmatrix} -1 & 1-i \\ 1+i & -2+i \end{bmatrix}$$

Let us verify our calculations:

$$\begin{bmatrix} -1 & 1-i \\ 1+i & -2+i \end{bmatrix} \begin{bmatrix} -1-2\,i & -1-i \\ 1-i & -i \end{bmatrix}$$
$$= \begin{bmatrix} 1+2i+(1-i)^2 & 1+i-i(1-i) \\ (1+i)(-1-2i)+(-2+i)(1-i) & -(1+i)^2-i(-2+i) \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

The inverse of a 2×2 matrix may be computed as follows.

Proposition 4.6.13 **(Inverse of a 2×2 Matrix)**

Let $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$. Then A is invertible if and only if $ad - bc \neq 0$. Furthermore, if $ad - bc \neq 0$, then

$$A^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

Proof: If $ad - bc \neq 0$, then we may define a matrix $B \in M_{2\times 2}(\mathbb{F})$ by

$$B = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

A straightforward computation shows that $AB = BA = I_2$. Thus, A is invertible in this case and moreover $B = A^{-1}$, as claimed.

Conversely, assume that $ad - bc = 0$. We will show in this case that A is not invertible by proving that $\operatorname{rank}(A) \neq 2$. We will have to consider two cases: $d = 0$ and $d \neq 0$. If $d = 0$ then

$$A = \begin{bmatrix} a & b \\ c & 0 \end{bmatrix}.$$

and $bc = ad = 0$ (since $ad - bc = 0$ by assumption). So, one of b or c must be 0. If $c = 0$ then A has a row of zeros, so it can have at most one pivot, hence $\operatorname{rank}(A) \neq 2$, and A is not invertible. A similar argument can be applied if $b = 0$: in that case, there can be at most one pivot (in the first column).

Next, if $d \neq 0$, we can perform the following EROs to A:

$$R_1 \to dR_1 \text{ gives } \begin{bmatrix} ad & bd \\ c & d \end{bmatrix}.$$
$$R_1 \to -bR_2 + R_1 \text{ gives } \begin{bmatrix} ad - bc & 0 \\ c & d \end{bmatrix}.$$

Since $ad - bc = 0$, we have thus row reduced A to a matrix with a row of zeros. Hence, as above, A can have at most one pivot, and therefore, it cannot be invertible.

This completes the analysis of all cases. $\square$

REMARK

If $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, then the number $ad - bc$ is called the **determinant** of A. We will discuss it in more detail in Chapter 6.

Example 4.6.14 Find the inverses of the following matrices, if they exist.

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \quad B = \begin{bmatrix} 1 & 2 \\ 3 & 6 \end{bmatrix} \quad C = \begin{bmatrix} -1-2i & -1-i \\ 1-i & -i \end{bmatrix}$$

Solution: The determinant of A is $(1(4) - 2(3)) = -2$, and so A is invertible with

$$A^{-1} = -\frac{1}{2}\begin{bmatrix} 4 & -2 \\ -3 & 1 \end{bmatrix}.$$

The determinant of B is $(1(6) - 2(3)) = 0$, and so B is not invertible.

The determinant of C is

$$(-1-2i)(-i) - (-1-i)(1-i) = -2 + i + 2 = i$$

and so C is invertible. As $\dfrac{1}{i} = -i$, we have

$$C^{-1} = -i\begin{bmatrix} -i & 1+i \\ -1+i & -1-2i \end{bmatrix} = \begin{bmatrix} -1 & 1-i \\ 1+i & -2+i \end{bmatrix}$$

Chapter 5

Linear Transformations

5.1 The Function Determined by a Matrix

Up to this point in the course, we have considered matrices as arrays of numbers which contain important information about systems of linear equations through either the coefficient matrix, the augmented matrix, or both. Moving forward, we will discover that there is much more going on than this. We will make use of multiplication by a matrix to define functions.

Definition 5.1.1 **Function Determined by a Matrix**

Let $A \in M_{m\times n}(\mathbb{F})$. The **function determined by the matrix** A is the function

$$T_A \colon \mathbb{F}^n \to \mathbb{F}^m$$

defined by

$$T_A(\vec{x}) = A\vec{x}.$$

Example 5.1.2

Let $A = \begin{bmatrix} 1 & 4 \\ -2 & -5 \\ 4 & 6 \end{bmatrix}$. If $\vec{x} \in \mathbb{R}^2$, then

$$T_A(\vec{x}) = \begin{bmatrix} 1 & 4 \\ -2 & -5 \\ 4 & 6 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1 + 4x_2 \\ -2x_1 - 5x_2 \\ 4x_1 + 6x_2 \end{bmatrix}.$$

Note that $T_A(\vec{x}) \in \mathbb{R}^3$.

For instance, if $\vec{x} = \begin{bmatrix} -2 \\ 3 \end{bmatrix}$, then $T_A(\vec{x}) = T_A\left(\begin{bmatrix} -2 \\ 3 \end{bmatrix}\right) = \begin{bmatrix} 10 \\ -11 \\ 10 \end{bmatrix}$.

The function T_A takes as input vectors in $\mathbb{R}^2$ and outputs vectors in $\mathbb{R}^3$.

Example 5.1.3

Let $A = \begin{bmatrix} i & 1+2i & 3+2i \\ 2-i & 4 & 2-5i \end{bmatrix}$. If $\vec{z} \in \mathbb{C}^3$, then

$$T_A(\vec{z}) = \begin{bmatrix} i & 1+2i & 3+2i \\ 2-i & 4 & 2-5i \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} = \begin{bmatrix} iz_1 + (1+2i)z_2 + (3+2i)z_3 \\ (2-i)z_1 + 4z_2 + (2-5i)z_3 \end{bmatrix}.$$

Note that $T_A(\vec{z}) \in \mathbb{C}^2$.

For instance, if $\vec{z} = \begin{bmatrix} 3 \\ 2-i \\ 2i \end{bmatrix}$, then $T_A(\vec{z}) = T_A\left(\begin{bmatrix} 3 \\ 2-i \\ 2i \end{bmatrix}\right) = \begin{bmatrix} 12i \\ 24-3i \end{bmatrix}$.

The function T_A takes as input vectors in $\mathbb{C}^3$ and outputs vectors in $\mathbb{C}^2$.

The function T_A determined by a matrix $A \in M_{m\times n}(\mathbb{F})$ has some rather special features. In Theorem 3.9.9 we saw that matrix–vector multiplication is *linear*: that is, for any $\vec{x}, \vec{y} \in \mathbb{F}^n$ and any $c \in \mathbb{F}$, the following two properties hold:

$$A(\vec{x} + \vec{y}) = A\vec{x} + A\vec{y}$$
$$A(c\vec{x}) = cA\vec{x}$$

In view of these properties, the following result holds.

Theorem 5.1.4

(Function Determined by a Matrix is Linear)

Let $A \in M_{m\times n}(\mathbb{F})$ and let T_A be the function determined by the matrix A. Then T_A is linear; that is, for any $\vec{x}, \vec{y} \in \mathbb{F}^n$ and any $c \in \mathbb{F}$, the following two properties hold:

(a) $T_A(\vec{x} + \vec{y}) = T_A(\vec{x}) + T_A(\vec{y})$

(b) $T_A(c\vec{x}) = cT_A(\vec{x})$

Proof:

$$\begin{aligned} T_A(\vec{x} + \vec{y}) &= A(\vec{x} + \vec{y}) \\ &= A(\vec{x}) + A(\vec{y}) \\ &= T_A(\vec{x}) + T_A(\vec{y}) \end{aligned} \qquad \text{and} \qquad \begin{aligned} T_A(c\vec{x}) &= A(c\vec{x}) && \text{(by definition)} \\ &= cA\vec{x} && \text{(by Theorem 3.9.9)} \\ &= cT_A(\vec{x}) && \text{(by definition).} \end{aligned}$$

□

5.2 Linear Transformations

The properties exhibited by T_A in Theorem 5.1.4 (Function Determined by a Matrix is Linear) turn out to be of such importance in linear algebra that we give a name to the class of functions that exhibit these properties.

Definition 5.2.1 **Linear Transformation**

We say that the function $T\colon \mathbb{F}^n \to \mathbb{F}^m$ is a **linear transformation** (or **linear mapping**) if, for any $\vec{x}, \vec{y} \in \mathbb{F}^n$ and any $c \in \mathbb{F}$, the following two properties hold:

1. $T(\vec{x} + \vec{y}) = T(\vec{x}) + T(\vec{y})$ (called **linearity over addition**).
2. $T(c\vec{x}) = cT(\vec{x})$ (called **linearity over scalar multiplication**).

We refer to $\mathbb{F}^n$ here as the **domain** of T and $\mathbb{F}^m$ as the **codomain** of T, as we would for any function.

REMARKS

- In linear algebra, the words **transformation** or **mapping** are often used instead of the word **function**.
- In high school, you studied a variety of non-linear functions, such as $f(x) = x^2 + 1$ or $f(x) = 2^x$. These functions usually had a simple domain and codomain, such as $\mathbb{R}$. In linear algebra, the domains and codomains are more complicated (such as $\mathbb{R}^n$ or $\mathbb{C}^n$), but the functions themselves are rather simple — they are linear. If you go on to take a vector calculus course, you may see more complicated functions and more complicated domains and codomains occurring together.

REMARK

A fundamental example of a linear transformation is the function T_A determined by a matrix A. The linearity of T_A was established in Theorem 5.1.4 (Function Determined by a Matrix is Linear).

From this point forward we shall refer to T_A as the *linear transformation* determined by A, instead of the *function* determined by A.

We now consider some features of linear transformations. We begin with the result below which tells us that, if we want, we can check linearity over addition and linearity over scalar multiplication *simultaneously*, instead of checking them separately. We will give a few examples after the following two results.

Proposition 5.2.2 **(Alternate Characterization of a Linear Transformation)**

Let $T : \mathbb{F}^n \to \mathbb{F}^m$ be a function. Then T is a linear transformation if and only if for any $\vec{x}, \vec{y} \in \mathbb{F}^n$ and any $c \in \mathbb{F}$,

$$T(c\vec{x} + \vec{y}) = cT(\vec{x}) + T(\vec{y}).$$

Proof: We begin with the forward direction. Suppose that T is a linear transformation. Let $\vec{x}, \vec{y} \in \mathbb{F}^n$ and $c \in \mathbb{F}$. Put $\vec{w} = c\vec{x}$. Since T is linear, we have

$$T(\vec{w} + \vec{y}) = T(\vec{w}) + T(\vec{y})$$

and

$$T(\vec{w}) = T(c\vec{x}) = cT(\vec{x}).$$

Thus, we have

$$\begin{aligned} T(c\vec{x} + \vec{y}) &= T(\vec{w} + \vec{y}) \\ &= T(\vec{w}) + T(\vec{y}) \\ &= cT(\vec{x}) + T(\vec{y}). \end{aligned}$$

Now for the backward direction. Suppose that for all $\vec{x}, \vec{y} \in \mathbb{F}^n$ and for all $c \in \mathbb{F}$,

$$T(c\vec{x} + \vec{y}) = cT(\vec{x}) + T(\vec{y}).$$

Taking $c = 1$, we find that for all $\vec{x}, \vec{y} \in \mathbb{F}^n$,

$$T(\vec{x} + \vec{y}) = T(\vec{x}) + T(\vec{y}).$$

Taking $\vec{y} = \vec{0}$, we find that for all $\vec{x} \in \mathbb{F}^n$ and for all $c \in \mathbb{F}$,

$$T(c\vec{x} + \vec{y}) = T(c\vec{x}) = cT(\vec{x}).$$

We conclude that T is a linear transformation. □

Proposition 5.2.3 **(Zero Maps to Zero)**

Let $T : \mathbb{F}^n \to \mathbb{F}^m$ be a linear transformation. Then

$$T(\vec{0}_{\mathbb{F}^n}) = \vec{0}_{\mathbb{F}^m}.$$

Proof: Notice that $\vec{0}_{\mathbb{F}^n} = 0 \cdot \vec{0}_{\mathbb{F}^n}$. Since T is linear,

$$\begin{aligned} T(\vec{0}_{\mathbb{F}^n}) &= T(0 \cdot \vec{0}_{\mathbb{F}^n}) \\ &= 0 \cdot T(\vec{0}_{\mathbb{F}^n}) \\ &= \vec{0}_{\mathbb{F}^m}, \end{aligned}$$

where the last equality follows from the fact that for any $\vec{x} \in \mathbb{F}^m$, $0 \cdot \vec{x} = \vec{0}_{\mathbb{F}^m}$. □

We will now look at some functions $T \colon \mathbb{F}^n \to \mathbb{F}^m$ and determine whether or not they are linear. Towards this end, here are a couple of questions to consider:

- Does the zero vector of the domain get mapped to the zero vector of the codomain? If it does not, then the function is not linear.
- Does the function *look* linear?
  All of its terms must be linear. For example, the terms $2x_1$, $\sqrt{3}x_2$, πx_3, and $-x_4$ are linear, while the terms of the form $-3x_1x_3$, x_1^2 or $x_1x_2x_3$ are *not* linear. If you encounter any such non-linear terms, this indicates that the function is likely not linear. Find a counterexample to definitively prove that the function is not linear.

If the answer to both of these questions is yes, then we use the definition to prove linearity.

Example 5.2.4

Determine whether or not the function $T_1 \colon \mathbb{R}^3 \to \mathbb{R}^2$ defined by

$$T_1\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) = \begin{bmatrix} 2x+3y+4 \\ 6x-7z \end{bmatrix}$$

is a linear transformation.

Solution: This transformation is not linear, because

$$T_1(\begin{bmatrix} 0 & 0 & 0 \end{bmatrix}^T) = \begin{bmatrix} 4 & 0 \end{bmatrix}^T \neq \begin{bmatrix} 0 & 0 \end{bmatrix}^T.$$

Example 5.2.5

Determine whether or not the function $T_2 \colon \mathbb{C}^3 \to \mathbb{C}^2$ defined by

$$T_2\left(\begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix}\right) = \begin{bmatrix} 2z_1+3z_3 \\ z_2 z_3 \end{bmatrix}$$

is a linear transformation.

Solution: The zero vector of $\mathbb{C}^3$ is mapped to the zero vector of $\mathbb{C}^2$ by this function. However, this function does not look linear because of the product $z_2 z_3$. We try to construct a counterexample. The (possibly) troublesome term does not involve z_1, so we will set that to zero. Since the image under the function does involve the other two variables let us see what happens for different values (setting both to 1 or 0 is not usually helpful).

Note that on one hand,

$$T_2\left(\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}\right) + T_2\left(\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 3 \\ 0 \end{bmatrix} = \begin{bmatrix} 3 \\ 0 \end{bmatrix}.$$

On the other hand,

$$T_2\left(\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}\right) = T_2\left(\begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} 3 \\ 1 \end{bmatrix}.$$

Since

$$T_2\left(\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}\right) + T_2\left(\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}\right) \neq T_2\left(\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}\right),$$

the function T_2 is not a linear transformation.

Example 5.2.6

Determine whether or not the function $T_3 \colon \mathbb{R}^2 \to \mathbb{R}^3$ defined by

$$T_3\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} 2x_1+3x_2 \\ 6x_1-5x_2 \\ 2x_1 \end{bmatrix}$$

is a linear transformation.

Solution: As the zero vector of $\mathbb{R}^2$ is mapped to the zero vector of $\mathbb{R}^3$ and we do not detect any non-linear terms, this function could be linear. Let us prove that this is indeed the case.

Let $\vec{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$, $\vec{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \in \mathbb{R}^2$ and $c \in \mathbb{R}$. Then, using the definition of T_3, we have

$$T_3(\vec{x}) = T_3\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} 2x_1 + 3x_2 \\ 6x_1 - 5x_2 \\ 2x_1 \end{bmatrix}$$

and

$$T_3(\vec{y}) = T_3\left(\begin{bmatrix} y_1 \\ y_2 \end{bmatrix}\right) = \begin{bmatrix} 2y_1 + 3y_2 \\ 6y_1 - 5y_2 \\ 2y_1 \end{bmatrix}.$$

We also have

$$c\vec{x} + \vec{y} = \begin{bmatrix} cx_1 \\ cx_2 \end{bmatrix} + \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} cx_1 + y_1 \\ cx_2 + y_2 \end{bmatrix},$$

so that

$$\begin{aligned} T_3\left(c\vec{x} + \vec{y}\right) = T_3\left(\begin{bmatrix} cx_1 + y_1 \\ cx_2 + y_2 \end{bmatrix}\right) &= \begin{bmatrix} 2(cx_1 + y_1) + 3(cx_2 + y_2) \\ 6(cx_1 + y_1) - 5(cx_2 + y_2) \\ 2(cx_1 + y_1) \end{bmatrix} \\ &= c\begin{bmatrix} 2x_1 + 3x_2 \\ 6x_1 - 5x_2 \\ 2x_1 \end{bmatrix} + \begin{bmatrix} 2y_1 + 3y_2 \\ 6y_1 - 5y_2 \\ 2y_1 \end{bmatrix} = cT_3\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) + T_3\left(\begin{bmatrix} y_1 \\ y_2 \end{bmatrix}\right). \end{aligned}$$

That is, $T_3\left(c\vec{x} + \vec{y}\right) = cT_3\left(\vec{x}\right) + T_3\left(\vec{y}\right)$. It follows that T_3 is a linear transformation (by Proposition 5.2.2 (Alternate Characterization of a Linear Transformation)).

Example 5.2.7

Determine whether or not the function $T_4 \colon \mathbb{C}^2 \to \mathbb{C}^2$ defined by

$$T_4\left(\begin{bmatrix} z_1 \\ z_2 \end{bmatrix}\right) = \begin{bmatrix} 2iz_1 + 3z_2 \\ (2-i)z_1 - (1+3i)z_2 \end{bmatrix}.$$

is a linear transformation.

Solution: As the zero vector of $\mathbb{C}^2$ is mapped to the zero vector of $\mathbb{C}^2$ and we do not detect any non-linear terms, this function could be linear. Let us prove that this is indeed the case.

Let $\vec{z} = \begin{bmatrix} z_1 \\ z_2 \end{bmatrix}$, $\vec{w} = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} \in \mathbb{C}^2$, and $c \in \mathbb{C}$. Then, from the definition of T_4, we have

$$T_4(\vec{z}) = T_4\left(\begin{bmatrix} z_1 \\ z_2 \end{bmatrix}\right) = \begin{bmatrix} 2iz_1 + 3z_2 \\ (2-i)z_1 - (1+3i)z_2 \end{bmatrix}$$

and

$$T_4(\vec{w}) = T_4\left(\begin{bmatrix} w_1 \\ w_2 \end{bmatrix}\right) = \begin{bmatrix} 2iw_1 + 3w_2 \\ (2-i)w_1 - (1+3i)w_2 \end{bmatrix}.$$

We also have

$$c\vec{z}+\vec{w}=\begin{bmatrix} cz_1 \\ cz_2 \end{bmatrix}+\begin{bmatrix} w_1 \\ w_2 \end{bmatrix}=\begin{bmatrix} cz_1+w_1 \\ cz_2+w_2 \end{bmatrix},$$

so that

$$\begin{aligned} T_4\left(c\vec{z}+\vec{w}\right) &= T_4\left(\begin{bmatrix} cz_1+w_1 \\ cz_2+w_2 \end{bmatrix}\right)=\begin{bmatrix} 2i(cz_1+w_1)+3(cz_2+w_2) \\ (2-i)(cz_1+w_1)-(1+3i)(cz_2+w_2) \end{bmatrix} \\ &= c\begin{bmatrix} 2iz_1+3z_2 \\ (2-i)z_1-(1+3i)z_2 \end{bmatrix}+\begin{bmatrix} 2iw_1+3w_2 \\ (2-i)w_1-(1+3i)w_2 \end{bmatrix} \\ &= cT_4\left(\begin{bmatrix} z_1 \\ z_2 \end{bmatrix}\right)+T_4\left(\begin{bmatrix} w_1 \\ w_2 \end{bmatrix}\right). \end{aligned}$$

That is, $T_4\left(c\vec{z}+\vec{w}\right)=cT_4\left(\vec{z}\right)+T_4(\vec{w})$. It follows that T_4 is a linear transformation (by Proposition 5.2.2 (Alternate Characterization of a Linear Transformation)).

5.3 The Range of a Linear Transformation and "Onto" Linear Transformations

A linear transformation is a special kind of function, and, as with many other functions, we consider the set of all outputs of a linear transformation.

Definition 5.3.1
Range

Let $T\colon \mathbb{F}^n \to \mathbb{F}^m$ be a linear transformation. We define the **range** of T, denoted $\operatorname{Range}(T)$, to be the set of all outputs of T. That is,

$$\operatorname{Range}(T)=\{T(\vec{x}) : \vec{x}\in\mathbb{F}^n\}.$$

The range of T is a subset of $\mathbb{F}^m$.

Since for any linear transformation $T\colon \mathbb{F}^n \to \mathbb{F}^m$ we have $T(\vec{0}_{\mathbb{F}^n})=\vec{0}_{\mathbb{F}^m}$ (by Proposition 5.2.3 (Zero Maps to Zero)), we have that $\vec{0}_{\mathbb{F}^m}\in\operatorname{Range}(A)$. Therefore, the range of a linear transformation is *never* empty.

Proposition 5.3.2

(Range of a Linear Transformation)

Let $A\in M_{m\times n}(\mathbb{F})$, and let $T_A\colon \mathbb{F}^n \to \mathbb{F}^m$ be the linear transformation determined by A. Then

$$\operatorname{Range}(T_A)=\operatorname{Col}(A).$$

Proof: We will show that these two subsets of $\mathbb{F}^m$ are equal by showing that each set is contained in the other.

First, suppose that $\vec{y}\in\operatorname{Range}(T_A)$. Then there exists $\vec{x}\in\mathbb{F}^n$ such that $T_A(\vec{x})=\vec{y}$. That is, there exists a vector $\vec{x}\in\mathbb{F}^n$ such that $A\vec{x}=\vec{y}$. Since the system $A\vec{x}=\vec{y}$ is consistent if and only if $\vec{y}\in\operatorname{Col}(A)$ (see Proposition 4.1.2 (Consistent System and Column Space)), we conclude that $\vec{y}\in\operatorname{Col}(A)$.

Next, suppose that $\vec{z} \in \text{Col}(A)$. Then there exist scalars $x_1, x_2, \ldots, x_n \in \mathbb{F}$ such that $\vec{z} = x_1\vec{a}_1 + x_2\vec{a}_2 + \cdots + x_n\vec{a}_n$. If we let $\vec{x} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}^T$, then $A\vec{x} = \vec{z}$; that is, $\vec{z} = T_A(\vec{x})$, and so $\vec{z} \in \text{Range}(T_A)$.

Thus, we conclude that $\text{Range}(T_A) = \text{Col}(A)$. □

REMARK (Connection to Systems of Linear Equations)

We have already seen in Proposition 4.1.2 (Consistent System and Column Space) that

$$\text{the system of linear equations } A\vec{x} = \vec{b} \text{ has a solution } \textbf{if and only if } \vec{b} \in \text{Col}(A).$$

We can now write

$$A\vec{x} = \vec{b} \text{ is consistent } \textbf{if and only if } \vec{b} \in \text{Range}(T_A).$$

Example 5.3.3

Determine $\text{Range}(T_A)$, where $A = \begin{bmatrix} 1 & 4 \\ -2 & -5 \\ 4 & 6 \end{bmatrix}$.

Solution: The range of T_A is

$$\text{Range}(T_A) = \text{Col}(A) = \text{Span}\left\{ \begin{bmatrix} 1 \\ -2 \\ 4 \end{bmatrix}, \begin{bmatrix} 4 \\ -5 \\ 6 \end{bmatrix} \right\}.$$

Note that the system of linear equations $A\vec{x} = \vec{b}$ is consistent if and only if $\vec{b} \in \text{Range}(T_A)$.

Example 5.3.4

Determine $\text{Range}(T_A)$, where $A = \begin{bmatrix} i & 1+2i & 3+2i \\ 2-i & 4 & 2-5i \end{bmatrix}$.

Solution: The range of T_A is

$$\text{Range}(T_A) = \text{Span}\left\{ \begin{bmatrix} i \\ 2-i \end{bmatrix}, \begin{bmatrix} 1+2i \\ 4 \end{bmatrix}, \begin{bmatrix} 3+2i \\ 2-5i \end{bmatrix} \right\}.$$

It can be shown that $\text{Range}(T_A) = \mathbb{C}^2$, and thus the system of linear equations $A\vec{x} = \vec{b}$ is consistent for *all* $\vec{b} \in \mathbb{C}^2$.

Linear transformations whose range is equal to the entire codomain deserve a special name.

Definition 5.3.5
Onto

We say that the transformation $T \colon \mathbb{F}^n \to \mathbb{F}^m$ is **onto** (or **surjective**) if $\text{Range}(T) = \mathbb{F}^m$.

Corollary 5.3.6 **(Onto Criteria)**

Let $A \in M_{m \times n}(\mathbb{F})$ and let T_A be the linear transformation determined by the matrix A. The following statements are equivalent:

(a) T_A is onto.

(b) $\mathrm{Col}(A) = \mathbb{F}^m$.

(c) $\mathrm{rank}(A) = m$.

Proof: $((a) \Rightarrow (b))$:
Suppose that T_A is onto. From Proposition 5.3.2 (Range of a Linear Transformation), it follows that

$$\mathrm{Col}(A) = \mathrm{Range}(T_A) = \mathbb{F}^m.$$

$((b) \Rightarrow (c))$:
Suppose that $\mathrm{Col}(A) = \mathbb{F}^m$. The system $A\vec{x} = \vec{b}$ is consistent if and only if $\vec{b} \in \mathrm{Col}(A)$ (see Proposition 4.1.2 (Consistent System and Column Space)). Thus we have that $A\vec{x} = \vec{b}$ is consistent for all $\vec{b} \in \mathbb{F}^m$. From Part (b) of Theorem 3.6.7 (System Rank Theorem), we conclude that $\mathrm{rank}(A) = m$.

$((c) \Rightarrow (a))$:
Suppose that $\mathrm{rank}(A) = m$. From Part (b) of Theorem 3.6.7 (System Rank Theorem), it follows that the system $A\vec{x} = \vec{b}$ is consistent for all $\vec{b} \in \mathbb{F}^m$. By Proposition 4.1.2 (Consistent System and Column Space), $\mathrm{Col}(A) = \mathbb{F}^m$. By Proposition 5.3.2 (Range of a Linear Transformation), $\mathrm{Range}(T_A) = \mathrm{Col}(A) = \mathbb{F}^m$. □

Example 5.3.7 Determine whether or not the linear transformation T_A determined by $A = \begin{bmatrix} 1 & 4 \\ -2 & -5 \\ 4 & 6 \end{bmatrix}$ is onto.

Solution: From Example 5.3.3 we know that

$$\mathrm{Range}(T_A) = \mathrm{Col}(A) = \mathrm{Span}\left\{ \begin{bmatrix} 1 \\ -2 \\ 4 \end{bmatrix}, \begin{bmatrix} 4 \\ -5 \\ 6 \end{bmatrix} \right\}.$$

The range of T_A is only a plane in $\mathbb{R}^3$. Therefore, it cannot be equal to all of $\mathbb{R}^3$, and so $\mathrm{Range}(T_A) \neq \mathbb{R}^3$. There will be some vectors $\vec{b}$ in $\mathbb{R}^3$ for which the equation $T_A(\vec{x}) = \vec{b}$ has no solution. This will happen for any vector which does not lie on the plane determined by $\mathrm{Col}(A)$. One such vector is $\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$. Therefore, the linear transformation T_A is not onto.

5.4 The Kernel of a Linear Transformation and "One-to-One" Linear Transformations

When analyzing functions, we consider the set of inputs whose output is zero.

Definition 5.4.1 **Kernel**

Let $T\colon \mathbb{F}^n \to \mathbb{F}^m$ be a linear transformation. We define the **kernel** of T, denoted $\mathrm{Ker}(T)$, to be the set of inputs of T whose output is zero. That is,

$$\mathrm{Ker}(T) = \left\{ \vec{x} \in \mathbb{F}^n : T(\vec{x}) = \vec{0}_{\mathbb{F}^m} \right\}.$$

The kernel of T is a subset of $\mathbb{F}^n$.

Since for any linear transformation $T\colon \mathbb{F}^n \to \mathbb{F}^m$ we have $T(\vec{0}_{\mathbb{F}^n}) = \vec{0}_{\mathbb{F}^m}$ (by Proposition 5.2.3 (Zero Maps to Zero)), we have $\vec{0}_{\mathbb{F}^n} \in \mathrm{Ker}(T)$. Thus, the kernel of a linear transformation is *never* empty.

Proposition 5.4.2 **(Kernel of a Linear Transformation)**

Let $A \in M_{m\times n}(\mathbb{F})$ and let $T_A\colon \mathbb{F}^n \to \mathbb{F}^m$ be the linear transformation determined by A. Then

$$\mathrm{Ker}(T_A) = \mathrm{Null}(A).$$

Proof: Using the definitions of kernel and nullspace, we find that $\vec{x} \in \mathrm{Ker}(T_A) \iff T_A(\vec{x}) = \vec{0}_{\mathbb{F}^m} \iff A\vec{x} = \vec{0}_{\mathbb{F}^m} \iff \vec{x} \in \mathrm{Null}(A)$. □

REMARK

The kernel of T_A is equal to the solution set of the homogeneous system $A\vec{x} = \vec{0}$.

Definition 5.4.3 **One-to-One**

We say that the transformation $T : \mathbb{F}^n \to \mathbb{F}^m$ is **one-to-one** (or **injective**) if, whenever $T(\vec{x}) = T(\vec{y})$, then $\vec{x} = \vec{y}$.

REMARK

Notice that the statement

$$\text{For all } \vec{x}, \vec{y} \in \mathbb{F}^n, \text{ if } T(\vec{x}) = T(\vec{y}) \text{ then } \vec{x} = \vec{y}$$

is logically equivalent to its contrapositive

$$\text{For all } \vec{x}, \vec{y} \in \mathbb{F}^n, \text{ if} \vec{x} \neq \vec{y} \text{ then } T(\vec{x}) \neq T(\vec{y})$$

Thus, one-to-one linear transformations have the nice property that they map distinct elements of $\mathbb{F}^n$ to distinct elements of $\mathbb{F}^m$.

There is a simple way to check if a linear transformation is one-to-one.

Proposition 5.4.4 **(One-to-One Test)**

Let $T\colon \mathbb{F}^n \to \mathbb{F}^m$ be a linear transformation. Then

$$T \text{ is one-to-one if and only if } \mathrm{Ker}(T) = \{\vec{0}_{\mathbb{F}^n}\}.$$

Proof: Suppose that T is one-to-one. We have noted that $\{\vec{0}_{\mathbb{F}^n}\} \subseteq \mathrm{Ker}(T)$, so it remains to show that $\mathrm{Ker}(T) \subseteq \{\vec{0}_{\mathbb{F}^n}\}$. For this purpose, choose an arbitrary vector $\vec{x}$ in $\mathrm{Ker}(T)$, so that $T(\vec{x}) = \vec{0}_{\mathbb{F}^m}$. Since $T(\vec{0}_{\mathbb{F}^n}) = \vec{0}_{\mathbb{F}^m}$, we have that $T(\vec{x}) = T(\vec{0}_{\mathbb{F}^n})$. Since T is one-to-one, we conclude that $\vec{x} = \vec{0}_{\mathbb{F}^n}$. Thus, $\mathrm{Ker}(T) \subseteq \{\vec{0}_{\mathbb{F}^n}\}$.

Conversely, suppose that $\mathrm{Ker}(T) = \{\vec{0}_{\mathbb{F}^n}\}$. Given any $\vec{x}, \vec{y} \in \mathbb{F}^n$ that satisfy $T(\vec{x}) = T(\vec{y})$, we must show that $\vec{x} = \vec{y}$. To this end, notice that

$$T(\vec{x}) = T(\vec{y}) \implies T(\vec{x}) - T(\vec{y}) = \vec{0}_{\mathbb{F}^m} \implies T(\vec{x} - \vec{y}) = \vec{0}_{\mathbb{F}^m},$$

where the last implication follows from linearity. This shows that $\vec{x} - \vec{y} \in \mathrm{Ker}(T)$. Since by assumption $\mathrm{Ker}(T) = \{\vec{0}_{\mathbb{F}^n}\}$, it follows that $\vec{x} - \vec{y} = \vec{0}_{\mathbb{F}^n}$, and therefore $\vec{x} = \vec{y}$, as required. $\square$

If we know that our linear transformation T is of the form $T = T_A$ for some matrix A, then we can say a bit more.

Corollary 5.4.5 **(One-to-One Criteria)**

Let $A \in M_{m\times n}(\mathbb{F})$ and let T_A be the linear transformation determined by the matrix A. The following statements are equivalent.

(a) T_A is one-to-one.

(b) $\mathrm{Null}(A) = \{\vec{0}_{\mathbb{F}^n}\}$.

(c) $\mathrm{nullity}(A) = 0$.

(d) $\mathrm{rank}(A) = n$.

Proof: $((a) \Rightarrow (b))$:
This follows from Proposition 5.4.2 (Kernel of a Linear Transformation) and Proposition 5.4.4 (One-to-One Test).

$((b) \Rightarrow (c))$:
Suppose that $\mathrm{Null}(A) = \{\vec{0}_{\mathbb{F}^n}\}$. Then the consistent system $A\vec{x} = \vec{0}_{\mathbb{F}^m}$ has a unique solution, namely $\vec{x} = \vec{0}_{\mathbb{F}^n}$. Thus, the solution set of this consistent system has 0 parameters, and so it follows from the definition of nullity that $\mathrm{nullity}(A) = 0$.

$((c) \Rightarrow (d))$:
Suppose that $\mathrm{nullity}(A) = 0$. Since $\mathrm{nullity}(A) = n - \mathrm{rank}(A)$ by definition, we see that $\mathrm{rank}(A) = n$.

$((d) \Rightarrow (a))$:
Suppose that $\operatorname{rank}(A) = n$ and consider the system $A\vec{x} = \vec{0}_{\mathbb{F}^m}$. Note that this system is consistent, because $\vec{x} = \vec{0}_{\mathbb{F}^n}$ is a solution. From Part (a) of Theorem 3.6.7 (System Rank Theorem), it follows that number of parameters in the solution set of this consistent system is $n - \operatorname{rank}(A) = n - n = 0$. We conclude that the solution $\vec{x} = \vec{0}_{\mathbb{F}^n}$ to $A\vec{x} = \vec{0}_{\mathbb{F}^m}$ is unique.

This shows that $\operatorname{Null}(A) = \{\vec{0}_{\mathbb{F}^n}\}$. Thus $\operatorname{Ker}(T_A) = \{\vec{0}_{\mathbb{F}^n}\}$ by Proposition 5.4.2 (Kernel of a Linear Transformation), and therefore T_A is one-to-one by Proposition 5.4.4 (One-to-One Test). □

Example 5.4.6 Determine whether or not the linear transformation $T_A \colon \mathbb{R}^3 \to \mathbb{R}^3$ determined by $A = \begin{bmatrix} 2 & 4 & 10 \\ 4 & -4 & -4 \\ 6 & 8 & 22 \end{bmatrix}$ is one-to-one.

Solution: Let us examine the kernel of T_A, which is the same as looking at the solution set to $A\vec{x} = \vec{0}$. The augmented matrix is

$$\begin{bmatrix} 2 & 4 & 10 & 0 \\ 4 & -4 & -4 & 0 \\ 6 & 8 & 22 & 0 \end{bmatrix},$$

which row reduces to

$$\begin{bmatrix} 1 & 2 & 5 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Thus $\operatorname{rank}(A) = 2 < 3 = n$, and it follows that T_A is not one-to-one by Part (d) of Corollary 5.4.5 (One-to-One Criteria).

By combining Corollary 5.3.6 (Onto Criteria) and Corollary 5.4.5 (One-to-One Criteria) we arrive at the following result concerning square matrices. The proof is left as an exercise. (Compare with Theorem 4.6.7 (Invertibility Criteria – First Version).)

Theorem 5.4.7 **(Invertibility Criteria – Second Version)**

Let $A \in M_{n\times n}(\mathbb{F})$ be a square matrix and let T_A be the linear transformation determined by the matrix A. The following statements are equivalent.

(a) A is invertible.

(b) T_A is one-to-one.

(c) T_A is onto.

(d) $\operatorname{Null}(A) = \{\vec{0}\}$. That is, the only solution to the homogeneous system $A\vec{x} = \vec{0}$ is the trivial solution $\vec{x} = \vec{0}$.

(e) $\operatorname{Col}(A) = \mathbb{F}^n$. That is, for every $\vec{b} \in \mathbb{F}^n$, the system $A\vec{x} = \vec{b}$ is consistent.

(f) $\operatorname{nullity}(A) = 0$.

(g) $\operatorname{rank}(A) = n$.

(h) $\operatorname{RREF}(A) = I_n$.

5.5 Every Linear Transformation is Determined by a Matrix

In Section 5.1 we saw that every matrix $A \in M_{m\times n}(\mathbb{F})$ determines a linear transformation $T_A\colon \mathbb{F}^n \to \mathbb{F}^m$. We will show in this section that, conversely, every linear transformation $T\colon \mathbb{F}^n \to \mathbb{F}^m$ is actually of the form $T = T_A$ for a certain matrix A (that depends on T). In fact, we will show that there is a very natural process that constructs this matrix A from T.

We will make use of the set

$$\mathcal{E} = \left\{ \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \cdots, \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix} \right\} = \{\vec{e_1}, \vec{e_2}, \ldots, \vec{e_n}\}$$

of **standard basis vectors** in $\mathbb{F}^n$ (which we had already met in Definition 1.3.15).

Example 5.5.1 Let us examine the consequences of linearity in the special case when $\mathbb{F}^n = \mathbb{F}^m = \mathbb{F}^2$. Thus suppose that $T\colon \mathbb{F}^2 \to \mathbb{F}^2$ is a linear mapping and let $\vec{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ be a vector in $\mathbb{F}^2$. Then

$$\begin{aligned} T\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) &= T\left(\begin{bmatrix} x_1 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ x_2 \end{bmatrix}\right) \\ &= T\left(x_1 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} 0 \\ 1 \end{bmatrix}\right) \\ &= x_1\, T\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) + x_2\, T\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right) \qquad \text{(by linearity)} \\ &= \begin{bmatrix} T(\vec{e_1}) \; T(\vec{e_2}) \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \\ &= \begin{bmatrix} T(\vec{e_1}) \; T(\vec{e_2}) \end{bmatrix} \vec{x}. \end{aligned}$$

This shows us that the actual effect of the linear transformation can be replicated by the introduction of a matrix $\begin{bmatrix} T(\vec{e_1}) \; T(\vec{e_2}) \end{bmatrix}$.

In addition, this matrix $\begin{bmatrix} T(\vec{e_1}) \; T(\vec{e_2}) \end{bmatrix}$ has columns which are constructed by applying T to the basis vectors $\vec{e_1}$ and $\vec{e_2}$ in $\mathbb{F}^2$. This means that if we know what the linear transformation does to just these two (standard basis) vectors, then we can determine what it does to all vectors in $\mathbb{F}^2$.

Finally, the actual value of $T(\vec{x})$ can be computed by matrix multiplication of this matrix $\begin{bmatrix} T(\vec{e_1}) \; T(\vec{e_2}) \end{bmatrix}$ by the component vector $\vec{x}$. This result extends to higher dimensions.

Definition 5.5.2
Standard Matrix

Let $T: \mathbb{F}^n \to \mathbb{F}^m$ be a linear transformation. We define the **standard matrix** of T, denoted by $[T]_{\mathcal{E}}$, to be $m \times n$ matrix whose columns are the images under T of the vectors in the standard basis of $\mathbb{F}^n$:

$$\begin{aligned}[T]_{\mathcal{E}} &= \begin{bmatrix} T(\vec{e_1}) & T(\vec{e_2}) & \cdots & T(\vec{e_n}) \end{bmatrix} \\ &= \begin{bmatrix} T\left(\begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}\right) & T\left(\begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}\right) & \cdots & T\left(\begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}\right) \end{bmatrix}.\end{aligned}$$

REMARKS

- There is a T to remind us of the linear transformation that we are going to represent.
- The square brackets in $[T]_{\mathcal{E}}$ is our way of indicating that we are converting a linear transformation into a matrix.
- The $\mathcal{E}$ indicates that the standard basis is being used for both the domain and the codomain. Later we will investigate the consequences of using different bases in either or both of these two sets.

Theorem 5.5.3 **(Every Linear Transformation Is Determined by a Matrix)**

Let $T: \mathbb{F}^n \to \mathbb{F}^m$ be a linear transformation and let $[T]_{\mathcal{E}}$ be the standard matrix of T. Then for all $\vec{x} \in \mathbb{F}^n$,

$$T(\vec{x}) = [T]_{\mathcal{E}}\, \vec{x}.$$

That is, $T = T_{[T]_{\mathcal{E}}}$ is the linear transformation determined by the matrix $[T]_{\mathcal{E}}$.

Proof: Let $\vec{x} \in \mathbb{F}^n$ with $\vec{x} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}^T$. Then

$$\begin{aligned} T(\vec{x}) &= T\left(\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}\right) \\ &= T\left(\begin{bmatrix} x_1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ x_2 \\ \vdots \\ 0 \end{bmatrix} + \cdots + \begin{bmatrix} 0 \\ 0 \\ \vdots \\ x_n \end{bmatrix}\right) \\ &= T\left(x_1\, \vec{e_1} + x_2\, \vec{e_2} + \cdots + x_n\, \vec{e_n}\right) \\ &= x_1 T\left(\vec{e_1}\right) + x_2 T\left(\vec{e_2}\right) + \cdots + x_n T\left(\vec{e_n}\right) \qquad \text{(using linearity)} \end{aligned}$$

$$= \begin{bmatrix} T(\vec{e_1}) & T(\vec{e_2}) & \cdots & T(\vec{e_n}) \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

$$= [T]_{\mathcal{E}}\, \vec{x}.$$

$\square$

This result is quite remarkable.

First, the matrix representation provides us with a very elegant and compact way of expressing the linear transformation.

Second, the construction of the standard matrix is very simple. We must only find the images of the n standard basis vectors and build the standard matrix column by column from these images.

Third, once we have this standard matrix, we can find the image of any vector in $\mathbb{F}^n$ under the linear transformation using matrix–vector multiplication.

Thus, if $T\colon \mathbb{F}^n \to \mathbb{F}^m$ is a linear transformation, then once we know $T(\vec{e_1}), \ldots, T(\vec{e_n})$, we can compute $T(\vec{x})$ for *all* $\vec{x} \in \mathbb{F}^n$. This property is incredibly strong and is one of the key features of *linear* transformations.

In particular, linear transformations $T\colon \mathbb{R} \to \mathbb{R}$ are very simple. For example, given $g(x) = mx$ (a linear function through the origin), it is possible to determine the value of m if we are provided with any non-zero point P that falls on the line. It turns out that these functions completely categorize linear transformations $T\colon \mathbb{R} \to \mathbb{R}$, which is our result below.

Proposition 5.5.4 Let $T\colon \mathbb{R} \to \mathbb{R}$ be a linear transformation. Then there is a real number $m \in \mathbb{R}$ such that $T(x) = mx$ for all $x \in \mathbb{R}$.

Proof: We have $T(x) = [T]_{\mathcal{E}}x$, where the standard matrix $[T]_{\mathcal{E}}$ of T is the 1×1 matrix $[T(1)]$. Thus, if we let $m = T(1)$, we find that $T(x) = mx$, as required. $\square$

Note that the above property is not true of non-linear transformations. For example, the (non-linear) function $f\colon \mathbb{R} \to \mathbb{R}$ defined by $f(x) = \sin(x)$ cannot be determined from knowing the value of $\sin(1)$.

To conclude this section, we summarize what we have learned. Given a matrix $A \in M_{m\times n}(\mathbb{F})$, we can define a linear transformation $T_A\colon \mathbb{F}^n \to \mathbb{F}^m$. Conversely, given a linear transformation $T\colon \mathbb{F}^n \to \mathbb{F}^m$, we can construct a matrix $[T]_{\mathcal{E}} \in M_{m\times n}(\mathbb{F})$. What happens if we combine these two processes? That is, what can we say about $T_{[T]_{\mathcal{E}}}$ and $[T_A]_{\mathcal{E}}$?

We answer both these questions and state a couple of useful consequences in the following result.

Proposition 5.5.5 **(Properties of a Standard Matrix)**

Let $A \in M_{m\times n}(\mathbb{F})$, let $T_A \colon \mathbb{F}^n \to \mathbb{F}^m$ be the linear transformation determined by A, and let $T \colon \mathbb{F}^n \to \mathbb{F}^m$ be a linear transformation. Then

(a) $T_{[T]_\mathcal{E}} = T$.

(b) $[T_A]_\mathcal{E} = A$.

(c) T is onto if and only if $\operatorname{rank}([T]_\mathcal{E}) = m$.

(d) T is one-to-one if and only if $\operatorname{rank}([T]_\mathcal{E}) = n$.

Proof: (a) Using the definition of the linear transformation defined by a matrix, we have

$$T_{[T]_\mathcal{E}}\vec{x} = [T]_\mathcal{E}\vec{x} = T\vec{x} \quad \text{for all } \vec{x} \in \mathbb{F}^n,$$

where the last equality follows from Theorem 5.5.3 (Every Linear Transformation Is Determined by a Matrix). Thus, the two functions $T_{[T]_\mathcal{E}}$ and T have the same values at each $\vec{x} \in \mathbb{F}^n$, so they must be equal.

(b) Using the definition of the standard matrix, we have

$$\begin{aligned}
[T_A]_\mathcal{E} &= \begin{bmatrix} T_A(\vec{e_1}) & \cdots & T_A(\vec{e_n}) \end{bmatrix} \\
&= \begin{bmatrix} A\vec{e_1} & \cdots & A\vec{e_n} \end{bmatrix} \\
&= \begin{bmatrix} \vec{a_1} & \cdots & \vec{a_n} \end{bmatrix} \quad \text{(by Lemma 4.2.2 (Column Extraction))} \\
&= A.
\end{aligned}$$

(c) This follows from Corollary 5.3.6 (Onto Criteria).

(d) This follows from Corollary 5.4.5 (One-to-One Criteria).

$\square$

5.6 Special Linear Transformations: Projection, Perpendicular, Rotation and Reflection

In this section, we will revisit the operations of projection and perpendicular discussed in Chapter 1, prove that both of these types of transformations are linear, and determine their standard matrices. We will do the same for the operations of rotation and reflection.

Example 5.6.1 **(Projection onto a line through the origin)**

Let $\vec{w} = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$ be a non-zero vector in $\mathbb{R}^2$. Consider the projection transformation $\operatorname{proj}_{\vec{w}} \colon \mathbb{R}^2 \to \mathbb{R}^2$ defined in Section 1.6.

(a) Prove that $\operatorname{proj}_{\vec{w}}$ is a linear transformation.

(b) Determine the standard matrix of $\operatorname{proj}_{\vec{w}}$.

(c) Determine whether $\operatorname{proj}_{\vec{w}}$ is onto.

(d) Determine whether $\operatorname{proj}_{\vec{w}}$ is one-to-one.

Solution:

(a) Recall that if $\vec{v}$ is a vector in $\mathbb{R}^2$, then the projection of $\vec{v}$ onto $\vec{w}$ is defined by

$$\operatorname{proj}_{\vec{w}}(\vec{v}) = \frac{\vec{v}\cdot\vec{w}}{\|\vec{w}\|^2}\vec{w}.$$

Now let $\vec{u}, \vec{v} \in \mathbb{R}^2$ and let $c_1, c_2 \in \mathbb{R}$. By the linearity properties of the dot product,

$$\begin{aligned}\operatorname{proj}_{\vec{w}}(c_1\vec{u} + c_2\vec{v}) &= c_1\frac{\vec{u}\cdot\vec{w}}{\|\vec{w}\|^2}\vec{w} + c_2\frac{\vec{v}\cdot\vec{w}}{\|\vec{w}\|^2}\vec{w}\\ &= c_1\operatorname{proj}_{\vec{w}}(\vec{u}) + c_2\operatorname{proj}_{\vec{w}}(\vec{v}).\end{aligned}$$

We conclude that $\operatorname{proj}_{\vec{w}}$ is a linear transformation.

(b) To find the standard matrix $[\operatorname{proj}_{\vec{w}}]_{\mathcal{E}}$ of $\operatorname{proj}_{\vec{w}}$, recall that

$$[\operatorname{proj}_{\vec{w}}]_{\mathcal{E}} = \begin{bmatrix}\operatorname{proj}_{\vec{w}}(\vec{e_1}) & \operatorname{proj}_{\vec{w}}(\vec{e_2})\end{bmatrix}.$$

Since

$$\begin{aligned}\operatorname{proj}_{\vec{w}}(\vec{e_1}) &= \frac{\vec{e_1}\cdot\vec{w}}{\|\vec{w}\|^2}\vec{w} = \frac{w_1\cdot 1 + w_2\cdot 0}{w_1^2+w_2^2}\begin{bmatrix}w_1\\w_2\end{bmatrix} = \frac{1}{w_1^2+w_2^2}\begin{bmatrix}w_1^2\\w_1w_2\end{bmatrix}\\ \operatorname{proj}_{\vec{w}}(\vec{e_2}) &= \frac{\vec{e_2}\cdot\vec{w}}{\|\vec{w}\|^2}\vec{w} = \frac{w_1\cdot 0 + w_2\cdot 1}{w_1^2+w_2^2}\begin{bmatrix}w_1\\w_2\end{bmatrix} = \frac{1}{w_1^2+w_2^2}\begin{bmatrix}w_1w_2\\w_2^2\end{bmatrix},\end{aligned}$$

we have

$$[\operatorname{proj}_{\vec{w}}]_{\mathcal{E}} = \frac{1}{w_1^2+w_2^2}\begin{bmatrix}w_1^2 & w_1w_2\\ w_1w_2 & w_2^2\end{bmatrix}.$$

(c) There are two ways in which we can argue why $\operatorname{proj}_{\vec{w}}$ is *not* an onto linear transformation. Thinking geometrically, note that all images of $\operatorname{proj}_{\vec{w}}$ lie on the line $\operatorname{Span}\{\vec{w}\}$ and nowhere else in $\mathbb{R}^2$. Equivalently,

$$\operatorname{Range}(\operatorname{proj}_{\vec{w}}) = \operatorname{Span}\left\{\frac{w_1}{w_1^2+w_2^2}\begin{bmatrix}w_1\\w_2\end{bmatrix},\ \frac{w_2}{w_1^2+w_2^2}\begin{bmatrix}w_1\\w_2\end{bmatrix}\right\} = \operatorname{Span}\left\{\begin{bmatrix}w_1\\w_2\end{bmatrix}\right\} \neq \mathbb{R}^2.$$

(d) Once again, there are two ways in which we can argue why $\operatorname{proj}_{\vec{w}}$ is *not* a one-to-one linear transformation. Thinking geometrically, note that there are many different vectors which have the same image when they are projected onto the line. Equivalently, the kernel is obtained by solving the system

$$\frac{1}{w_1^2+w_2^2}\begin{bmatrix}w_1^2 & w_1w_2\\ w_1w_2 & w_2^2\end{bmatrix}\vec{v} = \vec{0},$$

which row-reduces to

$$\begin{bmatrix}w_1 & w_2\\ 0 & 0\end{bmatrix}\vec{v} = \vec{0}.$$

Since the rank of the coefficient matrix is strictly less than 2, it follows that $\operatorname{proj}_{\vec{w}}$ not one-to-one from Corollary 5.4.5 (One-to-One Criteria).

Note that if you want to project onto a line that does *not* pass through the origin, then you can project on the parallel line through the origin and then translate the solution. Note, however, that this function is not an example of a linear transformation, since $\vec{0}$ would not be mapped to $\vec{0}$.

EXERCISE

Let $\vec{w}$ be a non-zero vector in $\mathbb{R}^n$ and consider the perpendicular transformation $\text{perp}_{\vec{w}}\colon \mathbb{R}^n \to \mathbb{R}^n$ defined in Section 1.6 for $\vec{v} \in \mathbb{R}^n$:

$$\text{perp}_{\vec{w}}(\vec{v}) = \vec{v} - \text{proj}_{\vec{w}}(\vec{v}).$$

(a) Prove that $\text{perp}_{\vec{w}}$ is a linear transformation.

(b) Determine the standard matrix of $\text{perp}_{\vec{w}}$.

(c) Determine whether $\text{perp}_{\vec{w}}$ is onto.

(d) Determine whether $\text{perp}_{\vec{w}}$ is one-to-one.

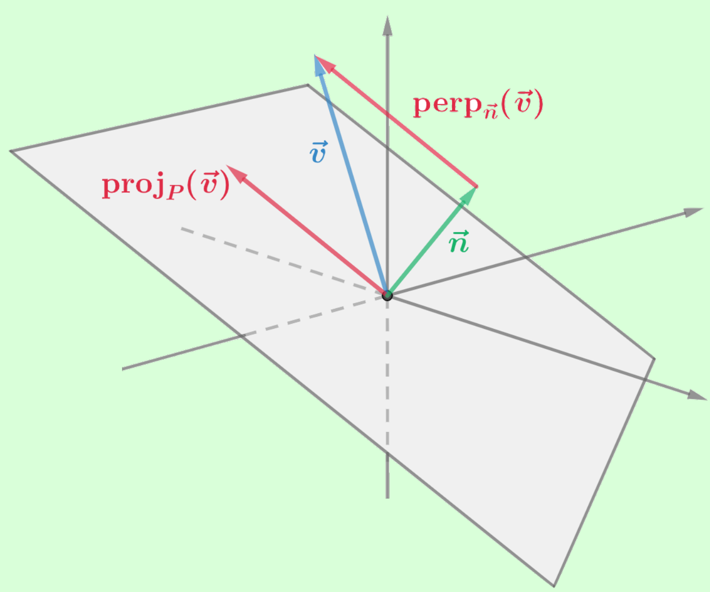
Example 5.6.2 **(Projection onto a plane through the origin)**

We can define the projection map onto a plane through the origin by using a normal vector of the plane. For a plane P with normal vector $\vec{n}$, we know every vector $\vec{w} \in P$ is orthogonal to $\vec{n}$, so we know that $\vec{w} = \text{perp}_{\vec{n}}(\vec{w})$.

Intuitively, for $\vec{v}, \vec{w} \in \mathbb{R}^3$, we think of $\text{proj}_{\vec{w}}(\vec{v})$ as the $\vec{w}$-component of $\vec{v}$; the same analogy can be made with planes. With this intuition in place, we define the projection of $\vec{v}$ onto the plane P as

$$\text{proj}_P(\vec{v}) = \text{perp}_{\vec{n}}(\vec{v}).$$

With this established, consider the projection map $\text{proj}_P\colon \mathbb{R}^3 \to \mathbb{R}^3$ onto a plane P, where P is defined by the scalar equation $3x - 4y + 5z = 0$.



(a) Prove that proj_P is a linear transformation.

(b) Determine the standard matrix of proj_P.

(c) Determine whether proj_P is onto.

(d) Determine whether proj_P is one-to-one.

Solution:

(a) Let $\vec{n} = \begin{bmatrix} 3 & -4 & 5 \end{bmatrix}^T$ be a normal to the plane P. Then $\text{proj}_P(\vec{v}) = \text{perp}_{\vec{n}}(\vec{v})$, and because we know (from the exercise above) that $\text{perp}_{\vec{n}}$ is a linear transformation, we conclude that proj_P is a linear transformation as well.

(b) To find the standard matrix $[\text{proj}_P]_{\mathcal{E}}$ of proj_P, recall that

$$[\text{proj}_P]_{\mathcal{E}} = \begin{bmatrix} \text{proj}_P(\vec{e_1}) & \text{proj}_P(\vec{e_2}) & \text{proj}_P(\vec{e_3}) \end{bmatrix}$$

Since

$$\begin{aligned}
\text{proj}_P(\vec{e_1}) &= \text{perp}_{\vec{n}}(\vec{e_1}) \\
&= \vec{e_1} - \text{proj}_{\vec{n}}(\vec{e_1}) \\
&= \vec{e_1} - \frac{\vec{e_1} \cdot \vec{n}}{\|\vec{n}\|^2}\vec{n} \\
&= \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} - \frac{1 \cdot 3 + 0 \cdot (-4) + 0 \cdot 5}{50} \begin{bmatrix} 3 \\ -4 \\ 5 \end{bmatrix} = \begin{bmatrix} 41/50 \\ 6/25 \\ -3/10 \end{bmatrix} \\
\text{proj}_P(\vec{e_2}) &= \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} - \frac{0 \cdot 3 + 1 \cdot (-4) + 0 \cdot 5}{50} \begin{bmatrix} 3 \\ -4 \\ 5 \end{bmatrix} = \begin{bmatrix} 6/25 \\ 17/25 \\ 2/5 \end{bmatrix} \\
\text{proj}_P(\vec{e_3}) &= \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} - \frac{0 \cdot 3 + 0 \cdot (-4) + 1 \cdot 5}{50} \begin{bmatrix} 3 \\ -4 \\ 5 \end{bmatrix} = \begin{bmatrix} -3/10 \\ 2/5 \\ 1/2 \end{bmatrix}
\end{aligned}$$

we have that

$$[\text{proj}_P]_{\mathcal{E}} = \begin{bmatrix} 41/50 & 6/25 & -3/10 \\ 6/25 & 17/25 & 2/5 \\ -3/10 & 2/5 & 1/2 \end{bmatrix}.$$

(c) There are two ways in which we can argue why proj_P is *not* onto. Thinking geometrically, note that all images lie on the plane $3x - 4y + 5z = 0$ and nowhere else in $\mathbb{R}^3$. Equivalently, we could investigate the range and show that it is not $\mathbb{R}^3$. This can be done, for example, by finding one vector $\vec{v} \in \mathbb{R}^3$ such that the system $[\text{proj}_P]_{\mathcal{E}}\vec{x} = \vec{v}$ is inconsistent. This can be achieved by taking any vector $\vec{v} \notin P$, such as $\vec{v} = \vec{e_1}$ (notice that $1 \cdot 3 + 0 \cdot (-4) + 0 \cdot 5 \neq 0$). Since $\vec{v} \notin \text{Range}(\text{proj}_P)$, we conclude that $\text{Range}(\text{proj}_P) \neq \mathbb{R}^3$, and so proj_P is not onto.

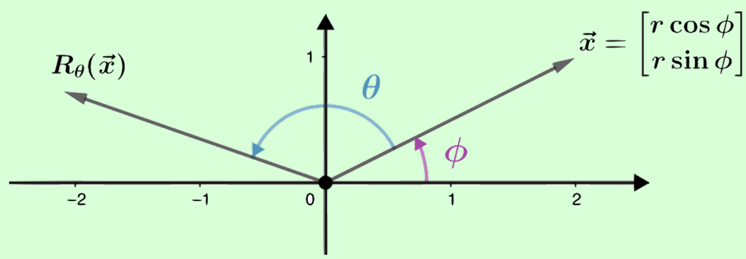
(d) Once again, there are two ways in which we can argue why proj_P is *not* one-to-one. Thinking geometrically, note that many different points have the same image when they are projected onto the plane. Equivalently, the kernel is obtained by solving

$$[\text{proj}_P]_{\mathcal{E}}\, \vec{x} = \vec{0}.$$

Since the rank of the coefficient matrix is strictly less than 3, it follows that proj_P not one-to-one from Corollary 5.4.5 (One-to-One Criteria).

Example 5.6.3 **(Rotation about the origin by an angle θ)**

Consider the transformation $R_\theta \colon \mathbb{R}^2 \to \mathbb{R}^2$ that rotates every vector $\vec{x} \in \mathbb{R}^2$ by a fixed angle θ *counter-clockwise* about the origin. We will assume $0 < \theta < 2\pi$ so that the rotation action is not trivial. This linear transformation is illustrated in the diagram below. Notice that the variable r represents the length of a vector $\vec{x}$.



(a) Prove that R_θ is a linear transformation.

(b) Determine the standard matrix of R_θ.

(c) Determine whether R_θ is onto.

(d) Determine whether R_θ is one-to-one.

Solution:

(a) We will prove that R_θ is a linear transformation by showing that it can be written in the form $R_\theta(\vec{x}) = A\vec{x}$ for some matrix $A \in M_{2\times 2}(\mathbb{R})$. Let $\vec{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ be an arbitrary vector in $\mathbb{R}^2$. Recall that we can convert $\vec{x}$ into its polar representation by taking $r = \|\vec{x}\|$ (the length of $\vec{x}$) and multiplying by $\cos\phi$ to get the x-coordinate and $\sin\phi$ to get the y-coordinate, where ϕ is the angle between $\vec{e_1}$ and $\vec{x}$.

Converting $\vec{x}$ into polar coordinates, we find that $\vec{x} = \begin{bmatrix} r\cos\phi \\ r\sin\phi \end{bmatrix}$, Since $R_\theta(\vec{x})$ is the result of a counter-clockwise rotation of $\vec{x}$ by an angle θ about the origin,

$$R_\theta\left(\begin{bmatrix} r\cos\phi \\ r\sin\phi \end{bmatrix}\right) = \begin{bmatrix} r\cos(\phi+\theta) \\ r\sin(\phi+\theta) \end{bmatrix}.$$

By making use of the trigonometric angle-sum identities

$$\begin{aligned}\cos(\phi+\theta) &= \cos\phi\cos\theta - \sin\phi\sin\theta \quad \text{and}\\ \sin(\phi+\theta) &= \sin\phi\cos\theta + \cos\phi\sin\theta,\end{aligned}$$

we have

$$\begin{aligned}R_\theta(\vec{x}) &= R_\theta\left(\begin{bmatrix} r\cos\phi \\ r\sin\phi\end{bmatrix}\right)\\ &= \begin{bmatrix} r\cos(\phi+\theta) \\ r\sin(\phi+\theta)\end{bmatrix}\\ &= \begin{bmatrix} r(\cos\phi\cos\theta - \sin\phi\sin\theta) \\ r(\sin\phi\cos\theta + \cos\phi\sin\theta)\end{bmatrix}\\ &= \begin{bmatrix} \cos\theta(r\cos\phi) - \sin\theta(r\sin\phi) \\ \sin\theta(r\cos\phi) + \cos\theta(r\sin\phi)\end{bmatrix}\\ &= \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta\end{bmatrix}\begin{bmatrix} r\cos\phi \\ r\sin\phi\end{bmatrix}\\ &= A\vec{x},\end{aligned}$$

where $A = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta\end{bmatrix}$. Since we were able to express R_θ in the form of a matrix-vector product, it must be the case that R_θ is a linear transformation.

(b) From part (a) we find that

$$[R_\theta]_{\mathcal{E}} = A = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta\end{bmatrix}.$$

(c) There are two ways in which we can argue why R_θ *is* onto. Thinking geometrically, note that any vector in $\mathbb{R}^2$ can be obtained by a rotation from an appropriate starting vector. Equivalently, as $\text{Range}(R_\theta) = \text{Col}([R_\theta]_{\mathcal{E}})$ by Proposition 5.3.2 (Range of a Linear Transformation), we have

$$\text{Range}(R_\theta) = \text{Span}\left\{\begin{bmatrix} \cos\theta \\ \sin\theta\end{bmatrix}, \begin{bmatrix} -\sin\theta \\ \cos\theta\end{bmatrix}\right\}$$

and it can be shown that this is $\mathbb{R}^2$.

(d) Once again, there are two ways in which we can argue why R_θ *is* one-to-one. Thinking geometrically, note that any two different vectors have different images when they are rotated. Equivalently, as $\text{Ker}(R_\theta) = \text{Null}([R_\theta]_{\mathcal{E}})$ by Proposition 5.4.2 (Kernel of a Linear Transformation), we have

$$\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta\end{bmatrix}\vec{x} = \vec{0}.$$

Given that the angle θ is fixed here, we view $\sin\theta$ and $\cos\theta$ as constants, so we can reduce this matrix as normal. We wish to show that the rank of the coefficient matrix is 2, so that the only solution to the system above is the trivial solution, allowing us to conclude that this linear transformation is one-to-one.

First, note that if $\cos(\theta) = 0$ or $\sin(\theta) = 0$ (conditions which cannot occur simultaneously) then the matrix is either in REF or a row swap will establish an REF. As such, in these scenarios the rank is 2 and the linear transformation is one-to-one.

In the scenarios where $\sin\theta \neq 0$ and $\cos\theta \neq 0$, we have

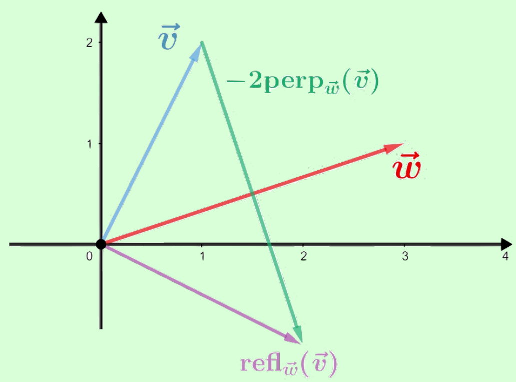
$$\begin{cases} R_1 \to (\sin\theta)R_1 \\ R_2 \to (\cos\theta)R_2 \end{cases} \quad \text{gives} \quad \begin{bmatrix} \sin\theta\cos\theta & -\sin^2\theta \\ \sin\theta\cos\theta & \cos^2\theta \end{bmatrix}$$

$$R_2 \to R_2 - R_1 \text{ gives} \quad \begin{bmatrix} \sin\theta\cos\theta & -\sin^2\theta \\ 0 & \cos^2\theta + \sin^2\theta \end{bmatrix} = \begin{bmatrix} \sin\theta\cos\theta & -\sin^2\theta \\ 0 & 1 \end{bmatrix}$$

We know $\sin\theta\cos\theta \neq 0$ as $\sin\theta \neq 0$ and $\cos\theta \neq 0$, so we can conclude that $\begin{bmatrix} \sin\theta\cos\theta & -\sin^2\theta \\ 0 & 1 \end{bmatrix}$ is an REF of $[R_\theta]_{\mathcal{E}}$, so again the rank is 2. Thus, we can say in all cases that R_θ is one-to-one.

Example 5.6.4 **(Reflection about a line through the origin)**

Let $\vec{w} = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$ be a non-zero vector in $\mathbb{R}^2$, and consider the reflection transformation $\text{refl}_{\vec{w}} \colon \mathbb{R}^2 \to \mathbb{R}^2$, which reflects any vector $\vec{v} \in \mathbb{R}^2$ about the line $\text{Span}\{\vec{w}\}$. This linear transformation is illustrated in the diagram below.



Using the geometric construction above, one can show that any vector $\vec{v} \in \mathbb{R}^2$ gets reflected about the line $\text{Span}\{\vec{w}\}$ according to the rule below. This is left as an exercise.

$$\text{refl}_{\vec{w}}(\vec{v}) = \vec{v} - 2\,\text{perp}_{\vec{w}}(\vec{v}).$$

(a) Prove that $\text{refl}_{\vec{w}}$ is a linear transformation.

(b) Determine the standard matrix of $\text{refl}_{\vec{w}}$.

(c) Determine whether $\text{refl}_{\vec{w}}$ is onto.

(d) Determine whether $\text{refl}_{\vec{w}}$ is one-to-one.

Solution:

(a) Let $\vec{u}, \vec{v} \in \mathbb{R}^2$ and let $c_1, c_2 \in \mathbb{F}$. Since the perpendicular transformation is linear,

$$\begin{aligned}\text{refl}_{\vec{w}}(c_1\vec{u}+c_2\vec{v}) &= (c_1\vec{u}+c_2\vec{v}) - 2\,\text{perp}_{\vec{w}}(c_1\vec{u}+c_2\vec{v}) \\ &= (c_1\vec{u}+c_2\vec{v}) - 2c_1\,\text{perp}_{\vec{w}}(\vec{u}) - 2c_2\,\text{perp}_{\vec{w}}(\vec{v}) \\ &= c_1\vec{u}+c_2\vec{v} - 2c_1\,\text{perp}_{\vec{w}}(\vec{u}) - 2c_2\,\text{perp}_{\vec{w}}(\vec{v}) \\ &= c_1(\vec{u} - 2\,\text{perp}_{\vec{w}}(\vec{u})) + c_2(\vec{v} - 2\,\text{perp}_{\vec{w}}(\vec{v})) \\ &= c_1\,\text{refl}_{\vec{w}}(\vec{u}) + c_2\,\text{refl}_{\vec{w}}(\vec{v}).\end{aligned}$$

We conclude that $\text{refl}_{\vec{w}}$ is a linear transformation.

(b) To find the standard matrix $[\text{refl}_{\vec{w}}]_{\mathcal{E}}$ of $\text{refl}_{\vec{w}}$, note that

$$\text{refl}_{\vec{w}}(\vec{v}) = \vec{v} - 2\,\text{perp}_{\vec{w}}(\vec{v}) = -\vec{v} + 2\,\text{proj}_{\vec{w}}(\vec{v}).$$

Since

$$\begin{aligned}\text{refl}_{\vec{w}}(\vec{e_1}) &= -\begin{bmatrix}1\\0\end{bmatrix} + \frac{2}{w_1^2+w_2^2}\begin{bmatrix}w_1^2\\w_1w_2\end{bmatrix} = \frac{1}{w_1^2+w_2^2}\begin{bmatrix}w_1^2-w_2^2\\2w_1w_2\end{bmatrix}, \\ \text{refl}_{\vec{w}}(\vec{e_2}) &= -\begin{bmatrix}0\\1\end{bmatrix} + \frac{2}{w_1^2+w_2^2}\begin{bmatrix}w_1w_2\\w_2^2\end{bmatrix} = \frac{1}{w_1^2+w_2^2}\begin{bmatrix}2w_1w_2\\w_2^2-w_1^2\end{bmatrix},\end{aligned}$$

we have that

$$[\text{refl}_{\vec{w}}]_{\mathcal{E}} = \begin{bmatrix}\text{refl}_{\vec{w}}(\vec{e_1}) & \text{refl}_{\vec{w}}(\vec{e_2})\end{bmatrix} = \frac{1}{w_1^2+w_2^2}\begin{bmatrix}w_1^2-w_2^2 & 2w_1w_2\\ 2w_1w_2 & w_2^2-w_1^2\end{bmatrix}.$$

(c) There are two ways in which we can argue why $\text{refl}_{\vec{w}}$ *is* onto. Thinking geometrically, note that any vector in $\mathbb{R}^2$ can be obtained by a reflection from an appropriate starting vector. Equivalently, the range is

$$\text{Range}(\text{refl}_{\vec{w}}) = \text{Span}\left\{\begin{bmatrix}w_1^2-w_2^2\\2w_1w_2\end{bmatrix}, \begin{bmatrix}2w_1w_2\\w_2^2-w_1^2\end{bmatrix}\right\}$$

and it can be shown that this is equal to $\mathbb{R}^2$.

(d) Once again, there are two ways in which we can argue why R_θ *is* one-to-one. Thinking geometrically, note that any two different vectors have different images when they are reflected. Equivalently, the kernel is obtained by solving

$$[\text{refl}_{\vec{w}}]_{\mathcal{E}}\vec{x} = \vec{0}.$$

Since the rank of the coefficient matrix is equal to 2, then the only solution is the trivial solution and so this linear transformation is one-to-one.

5.7 Composition of Linear Transformations

In this section, we will consider the composition of linear transformations and show how it is related to matrix multiplication.

Definition 5.7.1

Composition of Linear Transformations

Let $T_1 : \mathbb{F}^n \to \mathbb{F}^m$ and $T_2 : \mathbb{F}^m \to \mathbb{F}^p$ be linear transformations. We define the function $T_2 \circ T_1 \colon \mathbb{F}^n \to \mathbb{F}^p$ by

$$(T_2 \circ T_1)(\vec{x}) = T_2(T_1(\vec{x})).$$

The function $T_2 \circ T_1$ is called the **composite function** of T_2 and T_1.

Proposition 5.7.2 **(Composition of Linear Transformations Is Linear)**

Let $T_1 : \mathbb{F}^n \to \mathbb{F}^m$ and $T_2 : \mathbb{F}^m \to \mathbb{F}^p$ be linear transformations. Then $T_2 \circ T_1$ is a linear transformation.

Proof: Since T_1 is linear, we have that for all $\vec{x}, \vec{y} \in \mathbb{F}^n, c \in \mathbb{F}$,

$$T_1(c\vec{x} + \vec{y}) = cT_1(\vec{x}) + T_1(\vec{y}).$$

Since T_2 is linear, we have that for all $\vec{w}, \vec{z} \in \mathbb{F}^m,\ a \in \mathbb{F}$,

$$T_2(a\vec{w} + \vec{z}) = aT_2(\vec{w}) + T_2(\vec{z}).$$

Now, consider the action of the composition on $c\vec{x} + \vec{y}$:

$$(T_2 \circ T_1)(c\vec{x} + \vec{y}) = T_2\left(cT_1(\vec{x}) + T_1(\vec{y})\right).$$

We can let $T_1(\vec{x}) = \vec{w}$ and $T_1(\vec{y}) = \vec{z}$, so that

$$\begin{aligned}
(T_2 \circ T_1)(c\vec{x} + \vec{x}_2) &= T_2(c\vec{w} + \vec{z}) \\
&= cT_2(\vec{w}) + T_2(\vec{z}), && \text{(by linearity of } T_2\text{)} \\
&= cT_2(T_1(\vec{x})) + T_2(T_1(\vec{y})) \\
&= c(T_2 \circ T_1)(\vec{x}) + (T_2 \circ T_1)(\vec{y}).
\end{aligned}$$

Therefore, $T_2 \circ T_1$ is a linear transformation. $\square$

Proposition 5.7.3 **(The Standard Matrix of a Composition of Linear Transformations)**

Let $T_1 : \mathbb{F}^n \to \mathbb{F}^m$ and $T_2 : \mathbb{F}^m \to \mathbb{F}^p$ be linear transformations. Then the standard matrix of $T_2 \circ T_1$ is equal to the product of standard matrices of T_2 and T_1. That is,

$$[T_2 \circ T_1]_{\mathcal{E}} = [T_2]_{\mathcal{E}} [T_1]_{\mathcal{E}}.$$

Proof: Let $A = [T_1]_{\mathcal{E}}$, $B = [T_2]_{\mathcal{E}}$, and let $T_{BA} \colon \mathbb{F}^n \to \mathbb{F}^p$ denote the linear transformation determined by the matrix BA. Note that, for all $\vec{v} \in \mathbb{F}^n$,

$$\begin{aligned}
(T_2 \circ T_1)(\vec{v}) &= T_2(T_1(\vec{v})) \\
&= T_2(A\vec{v}) \\
&= B(A\vec{v}) \\
&= (BA)\vec{v} \\
&= T_{BA}(\vec{v}).
\end{aligned}$$

We conclude that the linear transformations $T_2 \circ T_1$ and T_{BA} are identical. Consequently, $(T_2 \circ T_1)(\vec{e_i}) = T_{BA}(\vec{e_i})$ for all $i = 1, \ldots, n$. But then

$$\begin{aligned}
[T_2 \circ T_1]_{\mathcal{E}} &= \begin{bmatrix} (T_2 \circ T_1)(\vec{e_1}) & \cdots & (T_2 \circ T_n)(\vec{e_n}) \end{bmatrix} \\
&= \begin{bmatrix} T_{BA}(\vec{e_1}) & \cdots & T_{BA}(\vec{e_n}) \end{bmatrix} \\
&= \begin{bmatrix} (BA)\vec{e_1} & \cdots & (BA)\vec{e_n} \end{bmatrix} \\
&= BA && \text{(by Lemma 4.2.2 (Column Extraction))} \\
&= [T_2]_{\mathcal{E}}[T_1]_{\mathcal{E}}.
\end{aligned}$$

□

This result gives us a very efficient way of obtaining the matrix representation of a composite linear transformation. It also explains a motivation behind the definition of matrix multiplication.

Example 5.7.4 Let $A = \begin{bmatrix} 1+i & -3i & -2 \\ 2 & 1+2i & 4i \end{bmatrix}$ be a matrix in $M_{2\times 3}(\mathbb{C})$ and $B = \begin{bmatrix} 3i & 1+i \\ 1 & 0 \\ 1-i & 3-2i \end{bmatrix}$ be a matrix in $M_{3\times 2}(\mathbb{C})$. Let T_A and T_B be the linear transformations determined by A and B, respectively. Determine the standard matrix of their composite linear transformation $T_B \circ T_A$.

Solution: By Proposition 5.7.3 (The Standard Matrix of a Composition of Linear Transformations),

$$[T_B \circ T_A]_{\mathcal{E}} = [T_B]_{\mathcal{E}}\,[T_A]_{\mathcal{E}} = BA.$$

Therefore,

$$[T_B \circ T_A]_{\mathcal{E}} = \begin{bmatrix} 3i & 1+i \\ 1 & 0 \\ 1-i & 3-2i \end{bmatrix} \begin{bmatrix} 1+i & -3i & -2 \\ 2 & 1+2i & 4i \end{bmatrix} = \begin{bmatrix} -1+5i & 8+3i & -4-2i \\ 1+i & -3i & -2 \\ 8-4i & 4+i & 6+14i \end{bmatrix}.$$

Example 5.7.5 Find the standard matrix of the linear transformation $T\colon \mathbb{R}^2 \to \mathbb{R}^2$ which is defined by T_1, a rotation counter-clockwise about the origin by an angle of $\frac{\pi}{3}$ radians, followed by T_2, a projection onto the line $y = -3x$.

Solution From Example 5.6.3 we have that

$$[T_1]_{\mathcal{E}} = [R_{\pi/3}]_{\mathcal{E}} = \begin{bmatrix} \cos(\frac{\pi}{3}) & -\sin(\frac{\pi}{3}) \\ \sin(\frac{\pi}{3}) & \cos(\frac{\pi}{3}) \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & -\frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & \frac{1}{2} \end{bmatrix}.$$

Next, note that the line $y = -3x$ can be written as $\mathrm{Span}\{\vec{w}\}$, where $\vec{w} = \begin{bmatrix} 1 & -3 \end{bmatrix}^T$. From Example 5.6.1 we have that for

$$[T_2]_{\mathcal{E}} = [\mathrm{proj}_{\vec{w}}]_{\mathcal{E}} = \frac{1}{1+(-3)^2}\begin{bmatrix} 1 & -3 \\ -3 & 9 \end{bmatrix} = \frac{1}{10}\begin{bmatrix} 1 & -3 \\ -3 & 9 \end{bmatrix}.$$

Thus, we have

$$[T]_{\mathcal{E}} = [T_2]_{\mathcal{E}}\,[T_1]_{\mathcal{E}} = \frac{1}{10}\begin{bmatrix} 1 & -3 \\ -3 & 9 \end{bmatrix}\begin{bmatrix} \frac{1}{2} & -\frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & \frac{1}{2} \end{bmatrix} = \frac{1}{20}\begin{bmatrix} 1-3\sqrt{3} & -3-\sqrt{3} \\ -3+9\sqrt{3} & 3\sqrt{3}+9 \end{bmatrix}.$$

Definition 5.7.6 **Identity Transformation**

The linear transformation $\mathrm{id}_n \colon \mathbb{F}^n \to \mathbb{F}^n$ such that $\mathrm{id}_n(\vec{x}) = \vec{x}$ for all $\vec{x} \in \mathbb{F}^n$ is called the **identity transformation**.

EXERCISE

Show that the standard matrix $[\mathrm{id}_n]_{\mathcal{E}}$ of id_n is the identity matrix I_n.

A special case of composition arises when $T \colon \mathbb{F}^n \to \mathbb{F}^n$ has the same domain and codomain, in which case we can apply the linear transformation T more than one time.

Definition 5.7.7 T^p

Let $T \colon \mathbb{F}^n \to \mathbb{F}^n$ and let $p > 1$ be an integer. We then define the p^{th} *power of* T, denoted by T^p, inductively by

$$T^p = T \circ T^{p-1}.$$

We also define $T^0 = \mathrm{id}_n$.

Corollary 5.7.8

Let $T \colon \mathbb{F}^n \to \mathbb{F}^n$ be a linear transformation and let $p > 1$ be an integer. Then the standard matrix of T^p is the p^{th} power of the standard matrix of T. That is,

$$[T^p]_{\mathcal{E}} = ([T]_{\mathcal{E}})^p.$$

Proof: Use induction and Proposition 5.7.3 (The Standard Matrix of a Composition of Linear Transformations). $\square$

Example 5.7.9

Consider the linear transformation defined by a rotation about the origin by an angle of θ counter-clockwise in the plane, denoted by $R_\theta : \mathbb{R}^2 \to \mathbb{R}^2$. In Example 5.6.3 we found that $[T_\theta]_{\mathcal{E}} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$. Determine the standard matrix of R_θ^2.

Solution: We have

$$\begin{aligned}
\left[R_\theta^2\right]_{\mathcal{E}} = ([R_\theta]_{\mathcal{E}})^2 &= \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \\
&= \begin{bmatrix} \cos^2(\theta) - \sin^2(\theta) & -2\cos(\theta)\sin(\theta) \\ 2\cos(\theta)\sin(\theta) & \cos^2(\theta) - \sin^2(\theta) \end{bmatrix}
\end{aligned}$$

From here, we can make use of the trigonometric double-angle identities,

$$\begin{aligned}
\cos(2\theta) &= \cos^2(\theta) - \sin^2(\theta) \quad \text{and} \\
\sin(2\theta) &= 2\cos(\theta)\sin(\theta),
\end{aligned}$$

to obtain

$$\left[R_\theta^2\right]_{\mathcal{E}} = \begin{bmatrix} \cos(2\theta) & -\sin(2\theta) \\ \sin(2\theta) & \cos(2\theta) \end{bmatrix},$$

which is the standard matrix for a rotation of 2θ counter-clockwise.

Chapter 6

The Determinant

6.1 The Definition of the Determinant

When is an $n \times n$ matrix A invertible? If $n = 1$ then $A = \begin{bmatrix} a \end{bmatrix}$ is invertible if and only if $a \neq 0$, in which case $A^{-1} = \begin{bmatrix} \frac{1}{a} \end{bmatrix}$. If $n = 2$, we saw in Proposition 4.6.13 (Inverse of a 2×2 Matrix) that $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is invertible if and only if $ad - bc \neq 0$.

Thus, the invertibility of a 1×1 or 2×2 matrix can be completely determined by examining a number formed using the entries of the matrix. We give this number a name.

Definition 6.1.1
Determinant of a 1×1 and 2×2 Matrix

If $A = \begin{bmatrix} a_{11} \end{bmatrix}$ is in $M_{1\times 1}(\mathbb{F})$, then the **determinant of** A, denoted by $\det(A)$, is:

$$\det(A) = a_{11}.$$

If $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$ is in $M_{2\times 2}(\mathbb{F})$, then the **determinant of** A, denoted by $\det(A)$, is:

$$\det(A) = a_{11}a_{22} - a_{12}a_{21}.$$

Example 6.1.2 Let $A = \begin{bmatrix} 4 - i \end{bmatrix}$. Then $\det(A) = 4 - i$.

Example 6.1.3 Let $B = \begin{bmatrix} 6 & 8 \\ -3 & 2 \end{bmatrix}$. Then $\det(B) = 6(2) - 8(-3) = 36$.

In the remainder of this section, we will generalize the above and define the determinant of an $n \times n$ matrix. One of our main results will be that an $n \times n$ matrix is invertible if and only if its determinant is non-zero, but it will take some time to get there (see Theorem 6.3.1 (Invertible iff the Determinant is Non-Zero)).

We will need some preliminary definitions.

Definition 6.1.4 $(i,j)^{th}$ **Submatrix,** $(i,j)^{th}$ **Minor**

Let $A \in M_{n\times n}(\mathbb{F})$. The $(i,j)^{th}$ **submatrix of** A, denoted by $M_{ij}(A)$, is the $(n-1)\times(n-1)$ matrix obtained from A by removing the i^{th} row and the j^{th} column from A. The determinant of $M_{ij}(A)$ is known as the $(i,j)^{th}$ **minor of** A.

Example 6.1.5

If $A = \begin{bmatrix} 4 & 6 & 8 \\ 6 & -3 & 2 \\ 5 & 7 & 9 \end{bmatrix}$ then $M_{22}(A) = \begin{bmatrix} 4 & 8 \\ 5 & 9 \end{bmatrix}$ and $M_{31}(A) = \begin{bmatrix} 6 & 8 \\ -3 & 2 \end{bmatrix}$.

Definition 6.1.6 **Determinant of an** $n \times n$ **matrix**

Let $A \in M_{n\times n}(\mathbb{F})$ for $n \geq 2$. We define the **determinant** function, $\det : M_{n\times n}(\mathbb{F}) \to \mathbb{F}$, by

$$\det(A) = \sum_{j=1}^{n} a_{1j}(-1)^{1+j}\det(M_{1j}(A)).$$

Example 6.1.7

If $n = 2$, so that $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$, the above definition gives

$$\det(A) = a_{11}(-1)^{1+1}\det(\begin{bmatrix} a_{22} \end{bmatrix}) + a_{12}(-1)^{1+2}\det(\begin{bmatrix} a_{21} \end{bmatrix}) = a_{11}a_{22} - a_{12}a_{21}.$$

Thus, we recover the same expression given in Definition 6.1.1.

The expression for $\det(A)$ in Definition 6.1.6 is called the **expansion along the first row** of the determinant. It is recursive in nature: we must multiply the entries in the *first row* of the matrix A by ± 1 and then by the determinant of an (i,j) submatrix.

For a 3×3 matrix, this process will involve the evaluation of three determinants of 2×2 matrices.

For a 4×4 matrix, this process will involve the evaluation of four determinants of 3×3 matrices, each of which involve the evaluation of three determinants of 2×2 matrices.

This pattern continues as we increase n.

In the next section we will learn techniques that will permit us to occasionally avoid too many tedious computations.

Example 6.1.8

Find the determinant of the matrix $A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 10 \end{bmatrix}$.

Solution: Using the definition, we have

$$\begin{aligned}\det(A) &= \sum_{j=1}^{3} a_{1j}(-1)^{1+j}\det(M_{1j}(A))\\ &= a_{11}(-1)^{1+1}\det(M_{11}(A)) + a_{12}(-1)^{1+2}\det(M_{12}(A)) + a_{13}(-1)^{1+3}\det(M_{13}(A))\\ &= (1)(1)\det\left(\begin{bmatrix}5 & 6\\ 8 & 10\end{bmatrix}\right) + (2)(-1)\det\left(\begin{bmatrix}4 & 6\\ 7 & 10\end{bmatrix}\right) + (3)(1)\det\left(\begin{bmatrix}4 & 5\\ 7 & 8\end{bmatrix}\right)\\ &= 1(50-48) - 2(40-42) + 3(32-35) = -3.\end{aligned}$$

Example 6.1.9 Find the determinant of the matrix $B = \begin{bmatrix}1 & 2i & 3\\ 4-i & 5+2i & 6\\ 7+3i & 8-2i & 10\end{bmatrix}$.

Solution: Using the definition, we have

$$\begin{aligned}\det(B) &= \sum_{j=1}^{n} b_{1j}(-1)^{1+j}\det(M_{1j}(B))\\ &= c_{11}(-1)^{1+1}\det(M_{11}(C)) + c_{12}(-1)^{1+2}\det(M_{12}(C)) + c_{13}(-1)^{1+3}\det(M_{13}(C))\\ &= (1)(1)\det\left(\begin{bmatrix}5+2i & 6\\ 8-2i & 10\end{bmatrix}\right) + (2i)(-1)\det\left(\begin{bmatrix}4-i & 6\\ 7+3i & 10\end{bmatrix}\right)\\ &\quad + (3)(1)\det\left(\begin{bmatrix}4-i & 5+2i\\ 7+3i & 8-2i\end{bmatrix}\right)\\ &= 1(50+20i-(48-12i)) - 2i(40-10i-(42+18i)) + 3(30-16i-(29+29i))\\ &= -51-99i.\end{aligned}$$

One of the remarkable features of the determinant is that it may be evaluated by performing an expansion along any row, not just the first row. We state this result below. The proof of this result is quite lengthy and is omitted.

Proposition 6.1.10 **(i^{th} Row Expansion of the Determinant)**

Let $A \in M_{n\times n}(\mathbb{F})$ with $n \geq 2$ and let $i \in \{1, \dots, n\}$. Then

$$\det(A) = \sum_{j=1}^{n} a_{ij}(-1)^{i+j}\det(M_{ij}(A)).$$

Notice that if $i = 1$ then the above expression reduces to the definition of $\det(A)$.

Example 6.1.11 Evaluate the determinant of the matrix $A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 10 \end{bmatrix}$ using an expansion along the second row.

Solution: We have

$$\begin{aligned} \det(A) &= \sum_{j=1}^{n} a_{2j}(-1)^{2+j} \det(M_{2j}(A)) \\ &= a_{21}(-1)^{2+1} \det(M_{21}(A)) + a_{22}(-1)^{2+2} \det(M_{22}(A)) + a_{23}(-1)^{2+3} \det(M_{23}(A)) \\ &= (4)(-1) \det\left(\begin{bmatrix} 2 & 3 \\ 8 & 10 \end{bmatrix}\right) + (5)(1) \det\left(\begin{bmatrix} 1 & 3 \\ 7 & 10 \end{bmatrix}\right) + (6)(-1) \det\left(\begin{bmatrix} 1 & 2 \\ 7 & 8 \end{bmatrix}\right) \\ &= -4(20-24) + 5(10-21) - 6(8-14) = -3, \end{aligned}$$

just as we found in Example 6.1.8.

A further remarkable feature of the determinant is that it may also be evaluated by expanding along any *column*.

Proposition 6.1.12 **(j^{th} Column Expansion of the Determinant)**

Let $A \in M_{n\times n}(\mathbb{F})$ with $n \geq 2$ and let $j \in \{1, \ldots, n\}$. Then

$$\det(A) = \sum_{i=1}^{n} a_{ij}(-1)^{i+j} \det(M_{ij}(A)).$$

Example 6.1.13 Evaluate the determinant of the matrix $A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 10 \end{bmatrix}$ using an expansion along the third column.

Solution: We have

$$\begin{aligned} \det(A) &= \sum_{i=1}^{3} a_{i3}(-1)^{i+3} \det(M_{i3}(A)) \\ &= a_{13}(-1)^{1+3} \det(M_{13}(A)) + a_{23}(-1)^{2+3} \det(M_{23}(A)) + a_{33}(-1)^{3+3} \det(M_{33}(A)) \\ &= (3)(1) \det\left(\begin{bmatrix} 4 & 5 \\ 7 & 8 \end{bmatrix}\right) + (6)(-1) \det\left(\begin{bmatrix} 1 & 2 \\ 7 & 8 \end{bmatrix}\right) + (10)(1) \det\left(\begin{bmatrix} 1 & 2 \\ 4 & 5 \end{bmatrix}\right) \\ &= 3(32-35) - 6(8-14) + 10(5-8) = -3. \end{aligned}$$

Thus, you may evaluate the determinant by expanding along any row or column of your choice. In practice, it is usually easier to choose a row or column with many zeros, since this will reduce the number determinants of submatrices that need to be computed.

Example 6.1.14 Evaluate the determinant of $A = \begin{bmatrix} 1 & 0 & 0 & 5 \\ 3 & 2 & 0 & 7 \\ 567 & 234 & 14 & 235 \\ 4 & 5 & 0 & 3 \end{bmatrix}$.

Solution: We perform an expansion along the third column.

$$\begin{aligned}\det(A) &= \sum_{i=1}^{4} a_{i3}(-1)^{i+3}\det(M_{i3}(A)) \\ &= 0(1)\det(M_{13}(A)) + 0(-1)\det(M_{23}(A)) + 14(1)\det(M_{33}(A)) + 0(-1)\det(M_{43}(A)) \\ &= 14\det\left(\begin{bmatrix} 1 & 0 & 5 \\ 3 & 2 & 7 \\ 4 & 5 & 3 \end{bmatrix}\right).\end{aligned}$$

From here, we need to proceed with another expansion along a row or column. Row 1 now looks "ideal" with the most zeroes of any row. Doing so, we continue our calculation below:

$$\begin{aligned}\det(A) &= 14\det\left(\begin{bmatrix} 1 & 0 & 5 \\ 3 & 2 & 7 \\ 4 & 5 & 3 \end{bmatrix}\right) \\ &= 14\left[(1)\det\left(\begin{bmatrix} 2 & 7 \\ 5 & 3 \end{bmatrix}\right) + (5)\det\left(\begin{bmatrix} 3 & 2 \\ 4 & 5 \end{bmatrix}\right)\right] \\ &= 14\left[(6-35) + 5(15-8)\right] \\ &= 14\left[-29+35\right] = 14(6) = 84.\end{aligned}$$

The recursive definition of the determinant makes its direct application tedious. There are, however, certain situations where the determinant is relatively easy to compute.

Proposition 6.1.15 **(Easy Determinants)**

Let $A \in M_{n\times n}(\mathbb{F})$ be a square matrix.

(a) If A has a row consisting only of zeros, then $\det A = 0$.

(b) If A has a column consisting only of zeros, then $\det A = 0$.

(c) If $A = \begin{bmatrix} a_{11} & * & * & \cdots & * \\ 0 & a_{22} & * & \cdots & * \\ 0 & 0 & a_{33} & \cdots & * \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & a_{nn} \end{bmatrix}$ is upper triangular, then $\det A = a_{11}a_{22}\cdots a_{nn}$.

Proof: For parts (a) and (b), simply perform the expansion along the row/column consisting of 0s.

To prove part (c), we will proceed by induction on n. If $n = 1$, then the result follows immediately from the definition of the determinant of a 1×1 matrix. For the inductive step, we will assume that the result is true when $n = k$ and consider a $(k+1) \times (k+1)$ upper triangular matrix

$$A = \begin{bmatrix} a_{11} & * & * & \cdots & * \\ 0 & a_{22} & * & \cdots & * \\ 0 & 0 & a_{33} & \cdots & * \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & a_{k+1,k+1} \end{bmatrix}.$$

Using an expansion along the first column, we obtain

$$\det(A) = a_{11}\det(M_{11}(A)) = a_{11}\det\left(\begin{bmatrix} a_{22} & * & \cdots & * \\ 0 & a_{44} & \cdots & * \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & a_{k+1,k+1} \end{bmatrix}\right).$$

Since $M_{11}(A)$ is an $k \times k$ matrix, we may apply the inductive hypothesis to conclude that $\det M_{11}(A) = a_{22} \cdots a_{k+1,k+1}$. It follows that $\det(A) = a_{11}a_{22} \cdots a_{k+1,k+1}$, as required. $\square$

Corollary 6.1.16 The determinant of the $n \times n$ identity matrix is 1, that is, $\det(I_n) = 1$.

Proof: This follows from part (c) of the previous Proposition. $\square$

The determinant of a *lower* triangular square matrix is also the product of its diagonal entries. This follows at once from the following result and Proposition 6.1.15 (Easy Determinants).

Proposition 6.1.17 Let $A \in M_{n\times n}(\mathbb{F})$. Then $\det(A) = \det(A^T)$.

Proof: We proceed by induction on n. If $n = 1$, then $A = A^T$ and there is nothing to prove. Thus, assume that the result is true for $n = k$ and consider a $(k+1)\times(k+1)$ matrix A. Let $B = A^T$. Notice that $b_{ij} = a_{ji}$ and that $M_{ij}(B) = (M_{ji}(A))^T$. Consequently, using an expansion along the first row of B, we have

$$\det(B) = \sum_{j=1}^{k+1} b_{1j}(-1)^{1+j}\det(M_{1j}(B)) = \sum_{j=1}^{k+1} a_{j1}(-1)^{1+j}\det((M_{j1}(A))^T).$$

Now since $M_{j1}(A)$ is a $k \times k$ matrix, we have $\det((M_{j1}(A))^T) = \det(M_{j1}(A))$, by the inductive hypothesis. So, the above expression for $\det(B) = \det(A^T)$ becomes

$$\det(A^T) = \sum_{j=1}^{k+1} a_{j1}(-1)^{1+j}\det(M_{j1}(A)).$$

The right-side is the determinant expansion along the first column of A, that is, it is equal to $\det(A)$. Thus, $\det(A^T) = \det(A)$. $\square$

6.2 Computing the Determinant in Practice: EROs

We will now outline a practical method for computing the determinant of a general square matrix A. The basic idea is to perform EROs on A, thereby reducing it to a matrix B, whose determinant is more easily computed—for instance, B could have a lot of 0 entries. Since EROs will turn out to have a well-understood effect on the determinant (as we describe below), we can obtain $\det(A)$ from $\det(B)$.

Theorem 6.2.1 **(Effect of EROs on the Determinant)**

Let $A \in M_{n\times n}(\mathbb{F})$.

(a) (Row swap) If B is obtained from A by interchanging two rows, then $\det(B) = -\det(A)$.

(b) (Row scale) If B is obtained from A by multiplying a row by $m \neq 0$, then $\det(B) = m\det(A)$.

(c) (Row addition) If B is obtained from A by adding a non-zero multiple of one row to another row, then $\det(B) = \det(A)$.

REMARK

Since $\det(A) = \det(A^T)$, the previous Theorem remains true if we replace every instance of the word "row" with the word "column".

WARNING: Do not use such "column operations" when performing calculations other than a determinant, as column operations disrupt many properties of a matrix.

Example 6.2.2

Let $A = \begin{bmatrix} 2 & -1 & 0 \\ 0 & -3 & 1 \\ 0 & 0 & 3 \end{bmatrix}$. Since A is upper triangular, we have that $\det(A) = (2)(-3)(3) = -18$.

Let $B = \begin{bmatrix} 2 & -1 & 0 \\ 0 & -3 & 1 \\ 0 & -6 & 5 \end{bmatrix}$. Notice that B may be obtained from A by adding a multiple of 2 times the second row of A to the third row of A. Thus, $\det(B) = \det(A) = -18$.

Let $C = \begin{bmatrix} 4 & -2 & 0 \\ 0 & -3 & 1 \\ 0 & -6 & 5 \end{bmatrix}$. Notice that C may be obtained from B by multiplying the first row by 2. Thus, $\det(C) = 2\det(B) = -36$.

Finally, let $D = \begin{bmatrix} 4 & -2 & 0 \\ 0 & -6 & 5 \\ 0 & -3 & 1 \end{bmatrix}$, which is obtained from C by interchanging the second and third rows. Thus, $\det(D) = -\det(C) = 36$.

The proof of Theorem 6.2.1 (Effect of EROs on the Determinant) involves lengthy (but straightforward) computations. We will not give it in this course. Instead, we will extract from the Theorem some useful corollaries.

Corollary 6.2.3 Let $A \in M_{n\times n}(\mathbb{F})$. If A has two identical rows (or two identical columns), then $\det(A) = 0$.

Proof: If A has two identical rows, we can use a row operation to subtract one of these rows from the other. The resulting matrix B will have a row of zeros. Consequently, $\det(B) = 0$ by Proposition 6.1.15 (Easy Determinants). From Theorem 6.2.1 (Effect of EROs on the Determinant), we know that row addition EROs do not change the determinant. Thus, we conclude that $\det(A) = \det(B) = 0$.

If A has two identical columns, we can apply the previous argument to A^T. □

Corollary 6.2.4 **(Determinants of Elementary Matrices)**

For each part below, let E be an elementary matrix of the indicated type.

(a) (Row swap) $\det(E) = -1$.

(b) (Row scale) $\det(E) = m$ (if E is obtained from I_n by multiplying a row by $m \neq 0$).

(c) (Row addition) $\det(E) = 1$.

Proof: Combine the fact that $\det(I_n) = 1$ with Theorem 6.2.1 (Effect of EROs on the Determinant). □

Corollary 6.2.5 **(Determinant After One ERO)**

Let $A \in M_{n\times n}(\mathbb{F})$ and suppose we perform a single ERO on A to produce the matrix B.

Assume that the corresponding elementary matrix is E. Then

$$\det(B) = \det(E)\det(A).$$

Proof: Combine Theorem 6.2.1 (Effect of EROs on the Determinant) and Corollary 6.2.4 (Determinants of Elementary Matrices). □

Corollary 6.2.6 **(Determinant After k EROs)**

Let $A \in M_{n\times n}(\mathbb{F})$ and suppose we perform a sequence of k EROs on the matrix A to obtain the matrix B.

Suppose that the elementary matrix corresponding to the ith ERO is E_i, so that

$$B = E_k \cdots E_2 E_1 A.$$

Then

$$\det(B) = \det(E_k \cdots E_2 E_1 A) = \det(E_k) \cdots \det(E_2)\det(E_1)\det(A).$$

Proof: We proceed by induction on k. If $k = 1$, this is the previous Corollary. Now assume that the result is true for $k = \ell$ and consider the case where $k = \ell + 1$. We then have $B = E_{\ell+1}E_\ell \cdots E_1 A$, and therefore,

$$\begin{aligned} \det(B) &= \det(E_{\ell+1})\det(E_\ell \cdots E_1 A) && \text{(Corollary 6.2.5)} \\ &= \det(E_{\ell+1})\det(E_\ell)\cdots\det(E_1)\det(A) && \text{(inductive hypothesis).} \end{aligned}$$

The result now follows. □

We will give an example that illustrates how to use the preceding results to obtain the determinant of an arbitrary matrix. We will attempt to apply EROs to simplify A, keeping track of what EROs are being applied. Then we can relate $\det(A)$ to $\det(B)$ using Corollary 6.2.6 (Determinant After k EROs). Notice that row addition EROs do not change the determinant, so we need only keep track of row swap and row scale EROs.

Example 6.2.7 Evaluate the determinant of $A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 10 \end{bmatrix}$.

Solution: We have

$$\begin{aligned} \det(A) &= \det\left(\begin{bmatrix} 1 & 2 & 3 \\ 0 & -3 & -6 \\ 0 & -6 & -11 \end{bmatrix}\right) && \left(\text{perform } \begin{matrix} R_2 \to -4R_1 + R_2 \\ R_3 \to -7R_1 + R_3 \end{matrix}, \text{ determinant unchanged}\right) \\ &= \det\left(\begin{bmatrix} 1 & 2 & 3 \\ 0 & -3 & -6 \\ 0 & 0 & 1 \end{bmatrix}\right) && (\text{perform } R_3 \to -2R_2 + R_3, \text{ determinant unchanged}) \\ &= (1)(-3)(1) && (\text{determinant of an upper triangular matrix}) \\ &= -3. \end{aligned}$$

The previous example used only row addition EROs. Our next example will use all three types. Notice that if we apply a row scale ERO to A to obtain B, then $\det(B) = m\det(A)$. Thus, $\det(A) = \dfrac{1}{m}\det(B)$.

Example 6.2.8 Evaluate the determinant of $A = \begin{bmatrix} 2 & -2 & 4 \\ 5 & -5 & 1 \\ 3 & 6 & 5 \end{bmatrix}$.

Solution: We have

$$\begin{aligned} \det(A) &= 2\det\left(\begin{bmatrix} 1 & -1 & 2 \\ 5 & -5 & 1 \\ 3 & 6 & 5 \end{bmatrix}\right) && \left(\text{perform } R_1 \to \tfrac{1}{2}R_1, \text{ determinant scaled by } \tfrac{1}{1/2} = 2\right) \\ &= 2\det\left(\begin{bmatrix} 1 & -1 & 2 \\ 0 & 0 & -9 \\ 0 & 9 & -1 \end{bmatrix}\right) && \left(\text{perform } \begin{matrix} R_2 \to -5R_1 + R_2 \\ R_3 \to -3R_1 + R_3 \end{matrix}, \text{ determinant unchanged}\right) \\ &= -2\det\left(\begin{bmatrix} 1 & -1 & 2 \\ 0 & 9 & -1 \\ 0 & 0 & -9 \end{bmatrix}\right) && (\text{perform } R_2 \leftrightarrow R_3, \text{ sign of determinant changed}) \end{aligned}$$

$$= (-2)(1)(9)(-9) \qquad \text{(determinant of an upper triangular matrix)}$$
$$= 162.$$

6.3 The Determinant and Invertibility

In this section, we will use our results on the relationship between EROs and the determinant to prove the following fundamental result.

Theorem 6.3.1 **(Invertible iff the Determinant is Non-Zero)**

Let $A \in M_{n\times n}(\mathbb{F})$. Then A is invertible if and only if $\det(A) \neq 0$.

Proof: Let $R = \text{RREF}(A)$. Suppose we obtain R from A by applying k EROs. Thus,

$$R = E_k \cdots E_1 A,$$

where E_i is the elementary matrix corresponding to the ith ERO. Corollary 6.2.6 (Determinant After k EROs) then implies that

$$\det(R) = \det(E_k)\cdots\det(E_1)\det(A).$$

By Corollary 6.2.4 (Determinants of Elementary Matrices), since $\det(E_i) \neq 0$ for all i, it follows that $\det(R) \neq 0$ if and only if $\det(A) \neq 0$.

Now, if A is invertible, then $R = I_n$, and therefore $\det(R) = 1 \neq 0$. So $\det(A) \neq 0$.

Conversely, if A is not invertible, then R must have a zero row, and therefore $\det(R) = 0$. So $\det(A) = 0$ in this case. This completes the proof. $\square$

Example 6.3.2 For what value(s) of the constant k is the matrix $A = \begin{bmatrix} 3 & 5 & 7 \\ 6 & k & 14 \\ 2 & 4 & 6 \end{bmatrix}$ invertible?

Solution: Let us evaluate the determinant of A.

$$\begin{aligned}
\det(A) &= \det\left(\begin{bmatrix} 3 & 5 & 7 \\ 0 & k-10 & 0 \\ 2 & 4 & 6 \end{bmatrix}\right) && (R_2 \to -2R_1 + R_2) \\
&= 2\det\left(\begin{bmatrix} 3 & 5 & 7 \\ 0 & k-10 & 0 \\ 1 & 2 & 3 \end{bmatrix}\right) && (R_3 \to \tfrac{1}{2}R_3) \\
&= 2\det\left(\begin{bmatrix} 0 & -1 & -2 \\ 0 & k-10 & 0 \\ 1 & 2 & 3 \end{bmatrix}\right) && (R_1 \to -3R_3 + R_1) \\
&= 4(k-10) && \text{(expansion along the second row).}
\end{aligned}$$

Thus, $\det(A) = 0$ if and only if $k = 10$. We conclude that the matrix A is invertible if and only if $k \neq 10$.

The following result is one of the most important properties of the determinant.

Proposition 6.3.3 **(Determinant of a Product)**

Let $A, B \in M_{n\times n}(\mathbb{F})$. Then $\det(AB) = \det(A)\det(B)$.

Proof: As in the proof of Theorem 6.3.1 (Invertible iff the Determinant is Non-Zero), we may write A as

$$A = E_k \cdots E_1 R$$

where $R = \mathrm{RREF}(A)$ and the E_i are elementary matrices that correspond to the EROs needed to operate on R to produce A. Then Corollary 6.2.6 (Determinant After k EROs) gives

$$\det(A) = \det(E_k)\cdots\det(E_1)\det(R).$$

Now, suppose A is invertible, so that $R = I_n$. On the one hand,

$$\det(A) = \det(E_k)\cdots\det(E_1),$$

while on the other hand, $AB = (E_k \cdots E_1 R)B = E_k \cdots E_1 B$ (as $R = I_n$) so that

$$\det(AB) = \det(E_k)\cdots\det(E_1)\det(B),$$

again by Corollary 6.2.6. The right-side is $\det(A)\det(B)$. Thus, the Proposition is true if A is invertible.

If A is not invertible, then R has a zero row, and therefore so does RB. Since

$$AB = E_k \cdots E_1(RB)$$

we can apply Corollary 6.2.6 (Determinant After k EROs) once more to find that

$$\det(AB) = \det(E_k)\cdots\det(E_1)\det(RB) = 0.$$

Since $\det(A)$ is 0 in this case too, we have that $\det(AB) = \det(A)\det(B)$. $\square$

REMARK

It is **not** true that $\det(A + B) = \det(A) + \det(B)$ in general. For example, if

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

then $\det(A) = \det(B) = 0$ so that $\det(A) + \det(B) = 0$. On the other hand, $A + B = I_2$, so $\det(A + B) = 1$. Thus, $\det(A + B) \neq \det(A) + \det(B)$ in this case.

Example 6.3.4 Let $A, B \in M_{n\times n}(\mathbb{F})$. Prove that AB is invertible if and only if BA is invertible.

Solution:

$$\begin{aligned} AB \text{ is invertible iff } & \det(AB) \neq 0 && \text{(Theorem 6.3.1 (Invertible iff the Determinant is Non-Zero))} \\ \text{iff } & \det(A)\det(B) \neq 0 && \text{(Proposition 6.3.3 (Determinant of a Product))} \\ \text{iff } & \det(B)\det(A) \neq 0 && \\ \text{iff } & \det(BA) \neq 0 && \text{(Proposition 6.3.3)} \\ \text{iff } & BA \text{ is invertible.} && \text{(Theorem 6.3.1)} \end{aligned}$$

The previous example implicitly contains the following useful observation.

Corollary 6.3.5 Let $A, B \in M_{n\times n}(\mathbb{F})$. Then $\det(AB) = \det(BA)$.

Here is another useful observation that can be proved using similar ideas.

Corollary 6.3.6 **(Determinant of Inverse)**

Let $A \in M_{n\times n}(\mathbb{F})$ be invertible. Then $\det(A^{-1}) = \dfrac{1}{\det(A)}$.

Proof: We have $A^{-1}A = I_n$. Using Proposition 6.3.3 (Determinant of a Product) and taking determinants of both sides, we get $\det(A^{-1})\det(A) = 1$. Since A is invertible, we have that $\det(A) \neq 0$, so we may divide through by $\det(A)$ to obtain $\det(A^{-1}) = 1/\det(A)$, as desired. □

Example 6.3.7 Prove that $A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 10 \end{bmatrix}$ is invertible and find $\det(A^{-1})$.

Solution: We saw in Example 6.2.7 that $\det(A) = -3$, so Theorem 6.3.1 (Invertible iff the Determinant is Non-Zero) tells us that A is invertible.

We get that $\det(A^{-1}) = -\dfrac{1}{3}$ by Corollary 6.3.6 (Determinant of Inverse).

6.4 An Expression for A^{-1}

We saw that a 2×2 matrix A is invertible if and only if $\det(A) \neq 0$ in Proposition 4.6.13 (Inverse of a 2×2 Matrix). In the previous sections, we generalized the notion of the determinant to $n \times n$ matrices and obtained the analogous result: an $n \times n$ matrix is invertible if and only if $\det(A) \neq 0$ (Theorem 6.3.1).

Proposition 4.6.13 contains one additional result: in the case where $\det(A) \neq 0$, an expression for the inverse of A is given, namely

$$A^{-1} = \frac{1}{\det(A)} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

In this section, we will show that there is a similar result for an $n \times n$ matrix (see Corollary 6.4.6 (Inverse by Adjugate)).

We will need some preliminary notation and terminology.

Definition 6.4.1
Cofactor

Let $A \in M_{n\times n}(\mathbb{F})$. The $(i,j)^{th}$ **cofactor of** A, denoted by $C_{ij}(A)$, is defined by

$$C_{ij}(A) = (-1)^{i+j} \det(M_{ij}(A)).$$

Note that we can write the determinant of A in terms of cofactors as either:

$$\det(A) = \sum_{j=1}^{n} a_{ij} C_{ij}(A) \quad \text{(expansion along the } i^{th} \text{ row)}$$

or

$$\det(A) = \sum_{i=1}^{n} a_{ij} C_{ij}(A) \quad \text{(expansion along the } j^{th} \text{ column).}$$

Definition 6.4.2
Adjugate of a Matrix

Let $A \in M_{n\times n}(\mathbb{F})$. The **adjugate of** A, denoted by $\mathrm{adj}(A)$, is the $n \times n$ matrix whose $(i,j)^{th}$ entry is

$$(\mathrm{adj}(A))_{ij} = C_{ji}(A).$$

That is, the adjugate of A is the *transpose* of the matrix of cofactors of A.

Example 6.4.3

If $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, then the cofactors of A are

$$C_{11}(A) = (-1)^{1+1} d = d, \quad C_{12}(A) = (-1)^{1+2} c = -c,$$

$$C_{21}(A) = (-1)^{2+1} b = -b, \quad C_{22}(A) = (-1)^{2+2} a = a.$$

Hence

$$\mathrm{adj}(A) = \begin{bmatrix} d & -c \\ -b & a \end{bmatrix}^T = \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

Notice that this is the matrix that appeared in the expression for A^{-1} in Proposition 4.6.13 (Inverse of a 2×2 Matrix).

Example 6.4.4 Find the adjugate of $A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 10 \end{bmatrix}$.

Solution:

The cofactors of A are given below. (We will write M_{ij} and C_{ij} instead of $M_{ij}(A)$ and $C_{ij}(A)$.)

$$\begin{array}{llll}
M_{11} = \begin{bmatrix} 5 & 6 \\ 8 & 10 \end{bmatrix} & C_{11} = 2 & M_{12} = \begin{bmatrix} 4 & 6 \\ 7 & 10 \end{bmatrix} & C_{12} = 2 \\
M_{13} = \begin{bmatrix} 4 & 5 \\ 7 & 8 \end{bmatrix} & C_{13} = -3 & M_{21} = \begin{bmatrix} 2 & 3 \\ 8 & 10 \end{bmatrix} & C_{21} = 4 \\
M_{22} = \begin{bmatrix} 1 & 3 \\ 7 & 10 \end{bmatrix} & C_{22} = -11 & M_{23} = \begin{bmatrix} 1 & 2 \\ 7 & 8 \end{bmatrix} & C_{23} = 6 \\
M_{31} = \begin{bmatrix} 2 & 3 \\ 5 & 6 \end{bmatrix} & C_{31} = -3 & M_{32} = \begin{bmatrix} 1 & 3 \\ 4 & 6 \end{bmatrix} & C_{32} = 6 \\
M_{33} = \begin{bmatrix} 1 & 2 \\ 4 & 5 \end{bmatrix} & C_{33} = -3 & &
\end{array}$$

Consequently,

$$\operatorname{adj}(A) = \begin{bmatrix} 2 & 2 & -3 \\ 4 & -11 & 6 \\ -3 & 6 & -3 \end{bmatrix}^T = \begin{bmatrix} 2 & 4 & -3 \\ 2 & -11 & 6 \\ -3 & 6 & -3 \end{bmatrix}.$$

Here is the key result concerning the adjugate matrix, which we state without proof.

Theorem 6.4.5 Let $A \in M_{n\times n}(\mathbb{F})$. Then

$$A \operatorname{adj}(A) = \operatorname{adj}(A)\, A = \det(A) I_n.$$

Corollary 6.4.6 **(Inverse by Adjugate)**

Let $A \in M_{n\times n}(\mathbb{F})$. If $\det(A) \neq 0$, then

$$A^{-1} = \frac{1}{\det(A)} \operatorname{adj}(A).$$

Proof: From Theorem 6.4.5, we know that

$$\operatorname{adj}(A) A = \det(A) I_n.$$

Dividing this equation by $\det(A)$, we obtain

$$\left(\frac{1}{\det(A)} \operatorname{adj}(A)\right) A = I_n.$$

Therefore, $\dfrac{1}{\det(A)}\operatorname{adj}(A) = A^{-1}$. □

Example 6.4.7 If $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ and $\det(A) \neq 0$, then by Example 6.4.3 and Corollary 6.4.6 (Inverse by Adjugate),

$$A^{-1} = \frac{1}{\det(A)}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix} = \frac{1}{ad-bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix},$$

exactly as in Proposition 4.6.13 (Inverse of a 2×2 Matrix).

Example 6.4.8 Let $A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 10 \end{bmatrix}$. Determine A^{-1}, if it exists.

Solution: From Example 6.2.7, we know

$$\det(A) = \det\begin{bmatrix} 1 & 2 & 3 \\ 0 & -3 & -6 \\ 0 & -6 & -11 \end{bmatrix} = 33 - 36 = -3 \neq 0,$$

so we have that A is invertible. Having computed in Example 6.4.4 that

$$\operatorname{adj}(A) = \begin{bmatrix} 2 & 4 & -3 \\ 2 & -11 & 6 \\ -3 & 6 & -3 \end{bmatrix},$$

we conclude that

$$A^{-1} = -\frac{1}{3}\begin{bmatrix} 2 & 4 & -3 \\ 2 & -11 & 6 \\ -3 & 6 & -3 \end{bmatrix}.$$

REMARK

We do not recommend this method as a way of determining the inverse of a matrix. However, it may be useful for checking that one or two entries are correct.

In practice, it will usually be more efficient to use the Algorithm involving EROs given in Proposition 4.6.8 (Algorithm for Checking Invertibility and Finding the Inverse).

6.5 Cramer's Rule

Cramer's Rule provides a method for solving systems of n linear equations in n unknowns by making use of determinants. It is not usually efficient from a computational point of view, since calculating determinants is generally computationally intensive.

However, Cramer's Rule is particularly useful if one is only interested in a single component of the solution.

Proposition 6.5.1 **(Cramer's Rule)**

Let $A \in M_{n\times n}(\mathbb{F})$ and consider the equation $A\overrightarrow{x} = \overrightarrow{b}$, where $\overrightarrow{b} \in \mathbb{F}^n$ and $\det(A) \neq 0$.

If we construct B_j from A by replacing the j^{th} column of A by the column vector $\overrightarrow{b}$, then the solution $\overrightarrow{x}$ to the equation

$$A\overrightarrow{x} = \overrightarrow{b}$$

is given by

$$x_j = \frac{\det(B_j)}{\det(A)}, \quad \text{for all } j = 1, \ldots, n.$$

Proof: Since $\det(A) \neq 0$, we have that A is invertible and

$$A^{-1} = \frac{1}{\det(A)}\,\mathrm{adj}(A).$$

Thus,

$$\overrightarrow{x} = A^{-1}\overrightarrow{b} = \frac{1}{\det A}\,\mathrm{adj}(A)\,\overrightarrow{b},$$

and so

$$x_j = \frac{1}{\det(A)}\sum_{k=1}^{n}(\mathrm{adj}(A))_{jk}\,b_k;$$

that is,

$$x_j = \frac{1}{\det(A)}\sum_{k=1}^{n}C_{kj}(A)\,b_k.$$

We will show that

$$\sum_{k=1}^{n}C_{kj}(A)\,b_k = \det(B_j),$$

and then Cramer's Rule will be proven.

We evaluate $\det(B_j)$ by expanding the determinant along the j^{th} column of B_j,

$$\det(B_j) = \sum_{i=1}^{k}(B_j)_{kj}\,(C(B_j))_{kj}.$$

Recall that $(B_j)_{kj} = b_k$, since the j^{th} column of B_j is just $\overrightarrow{b}$.

Since the matrices A and B only differ (at most) in the j^{th} column, we conclude that

$$(C(B_j))_{kj} = (C(A))_{kj}.$$

Thus,

$$\det(B_j) = \sum_{k=1}^{n}C_{kj}(A)\,b_k,$$

as required. $\square$

Example 6.5.2 Let $A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 10 \end{bmatrix}$. Use Cramer's Rule to solve

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 10 \end{bmatrix} \vec{x} = \begin{bmatrix} -2 \\ 3 \\ -4 \end{bmatrix}.$$

Solution: We saw in Example 6.2.7 that

$$\det\left(\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 10 \end{bmatrix}\right) = -3.$$

Now, let us evaluate the determinants of the matrices

$$B_1 = \begin{bmatrix} \mathbf{-2} & 2 & 3 \\ \mathbf{3} & 5 & 6 \\ \mathbf{-4} & 8 & 10 \end{bmatrix}, \quad B_2 = \begin{bmatrix} 1 & \mathbf{-2} & 3 \\ 4 & \mathbf{3} & 6 \\ 7 & \mathbf{-4} & 10 \end{bmatrix}, \quad B_3 = \begin{bmatrix} 1 & 2 & \mathbf{-2} \\ 4 & 5 & \mathbf{3} \\ 7 & 8 & \mathbf{-4} \end{bmatrix}.$$

We evaluate the determinants of B_1, B_2 and B_3 using the indicated EROs:

$$\det(B_1) = \det\begin{bmatrix} \mathbf{-2} & 2 & 3 \\ \mathbf{3} & 5 & 6 \\ \mathbf{-4} & 8 & 10 \end{bmatrix} = \det\begin{bmatrix} -2 & 2 & 3 \\ 0 & 8 & \frac{21}{2} \\ 0 & 4 & 4 \end{bmatrix} = 20. \qquad \begin{cases} R_2 \to \frac{3}{2}R_1 + R_2 \\ R_3 \to -2R_1 + R_3 \end{cases}$$

$$\det(B_2) = \det\begin{bmatrix} 1 & \mathbf{-2} & 3 \\ 4 & \mathbf{3} & 6 \\ 7 & \mathbf{-4} & 10 \end{bmatrix} = \det\begin{bmatrix} 1 & -2 & 3 \\ 0 & 11 & -6 \\ 0 & 10 & -11 \end{bmatrix} = -61. \qquad \begin{cases} R_2 \to -4R_1 + R_2 \\ R_3 \to -7R_1 + R_3 \end{cases}$$

$$\det(B_3) = \det\begin{bmatrix} 1 & 2 & \mathbf{-2} \\ 4 & 5 & \mathbf{3} \\ 7 & 8 & \mathbf{-4} \end{bmatrix} = \det\begin{bmatrix} 1 & 2 & -2 \\ 0 & -3 & 11 \\ 0 & -6 & 10 \end{bmatrix} = 36. \qquad \begin{cases} R_2 \to -4R_1 + R_2 \\ R_3 \to -7R_1 + R_3 \end{cases}$$

Thus,

$$\vec{x} = -\frac{1}{3}\begin{bmatrix} 20 \\ -61 \\ 36 \end{bmatrix}.$$

6.6 The Determinant and Geometry

In this section, we will give a geometric interpretation of the determinant. In $\mathbb{R}^2$, the underlying idea is the relationship between the determinant and areas of parallelograms, which we can calculate using the cross product. Here is the key result.

Proposition 6.6.1 **(Area of Parallelogram)**

Let $\vec{v} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$ and $\vec{w} = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$ be vectors in $\mathbb{R}^2$.

The area of the parallelogram with sides $\vec{v}$ and $\vec{w}$ is $\left| \det \left(\begin{bmatrix} v_1 & w_1 \\ v_2 & w_2 \end{bmatrix} \right) \right|$.

Proof: We shall consider the analogous vectors $\vec{v}_1$ and $\vec{w}_1$ in $\mathbb{R}^3$ with a third component of zero. That is, we let

$$\vec{v}_1 = \begin{bmatrix} v_1 \\ v_2 \\ 0 \end{bmatrix} \text{ and } \vec{w}_1 = \begin{bmatrix} w_1 \\ w_2 \\ 0 \end{bmatrix}.$$

As we saw in the Remark (Parallelogram Area via Cross Product) in Section 1.9, the area of the parallelogram with sides $\vec{v}$ and $\vec{w}$ is given by

$$\|\vec{v}_1 \times \vec{w}_1\| = \left\| \begin{bmatrix} v_1 \\ v_2 \\ 0 \end{bmatrix} \times \begin{bmatrix} w_1 \\ w_2 \\ 0 \end{bmatrix} \right\| = \left\| \begin{bmatrix} 0 \\ 0 \\ v_1 w_2 - w_1 v_2 \end{bmatrix} \right\| = |v_1 w_2 - w_1 v_2|.$$

The expression on the right is equal to $\left| \det \left(\begin{bmatrix} v_1 & w_1 \\ v_2 & w_2 \end{bmatrix} \right) \right|$, as required. $\square$

Example 6.6.2

Determine the area of the parallelogram in $\mathbb{R}^2$ with the vectors $\vec{v} = \begin{bmatrix} 2 & 5 \end{bmatrix}^T$ and $\vec{w} = \begin{bmatrix} 4 & 1 \end{bmatrix}^T$ as its sides.

Solution: The area is $\left| \det \left(\begin{bmatrix} 2 & 4 \\ 5 & 1 \end{bmatrix} \right) \right| = |-18| = 18 \text{ units}^2$.

Notice that it is important to include the absolute value signs in the expression given by Proposition 6.6.1 (Area of Parallelogram).

REMARK

If we study the proof of Proposition 6.6.1 (Area of Parallelogram), we can obtain a useful trick that helps us compute cross products. Namely, suppose we wish to compute the cross product $\vec{u} \times \vec{v}$ of $\vec{u} = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix}$ and $\vec{v} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}$. Consider the matrix

$$\begin{bmatrix} \vec{e}_1 & \vec{e}_2 & \vec{e}_3 \\ u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{bmatrix}$$

whose first row consists of the standard basis vectors $\vec{e}_1, \vec{e}_2, \vec{e}_3$ (treated as formal symbols!), and whose second and third rows are the entries of $\vec{u}$ and $\vec{v}$, respectively. Then if we expand along the first row, we obtain

$$\det \begin{bmatrix} \vec{e}_1 & \vec{e}_2 & \vec{e}_3 \\ u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{bmatrix} = (u_2v_3 - u_3v_2)\vec{e}_1 - (u_1v_3 - u_3v_1)\vec{e}_2 + (u_1v_2 - u_2v_1)\vec{e}_3$$
$$= \begin{bmatrix} u_2v_3 - u_3v_2 \\ -(u_1v_3 - u_3v_1) \\ u_1v_2 - u_2v_1 \end{bmatrix}.$$

We recognize this last expression as $\vec{u} \times \vec{v}$.

We must emphasize however that this is simply a trick that helps us remember the formula for $\vec{u} \times \vec{v}$. We are abusing notation by treating $\vec{e}_1, \vec{e}_2, \vec{e}_3$ as symbols in the left-side of the equation, but as vectors in the right-side of the equation.

Example 6.6.3 Compute $\begin{bmatrix} 2 \\ -3 \\ 5 \end{bmatrix} \times \begin{bmatrix} -2 \\ 1 \\ 4 \end{bmatrix}$.

Solution:

Using the trick above, we have

$$\begin{bmatrix} 2 \\ -3 \\ 5 \end{bmatrix} \times \begin{bmatrix} -2 \\ 1 \\ 4 \end{bmatrix} = \det \begin{bmatrix} \vec{e}_1 & \vec{e}_2 & \vec{e}_3 \\ 2 & -3 & 5 \\ -2 & 1 & 4 \end{bmatrix}$$
$$= (-17)\vec{e}_1 - (18)\vec{e}_2 + (-4)\vec{e}_3$$
$$= \begin{bmatrix} -17 \\ -18 \\ -4 \end{bmatrix},$$

just as we had computed in Example 1.9.2.

Here is another useful re-interpretation of Proposition 6.6.1 (Area of Parallelogram). Let $A \in M_{2\times 2}(\mathbb{R})$ and consider the unit square in $\mathbb{R}^2$. Multiply the sides $\vec{e_1}$ and $\vec{e_2}$ of the unit square by A to obtain the vectors $A\vec{e_1}$ and $A\vec{e_2}$. In this way, we may view A as having transformed the unit square into a parallelogram with sides $A\vec{e_1}$ and $A\vec{e_2}$.

Since $A\vec{e_1}$ and $A\vec{e_2}$ are the columns of A (by Lemma 4.2.2 (Column Extraction)), Proposition 6.6.1 tells us that the area of the resulting parallelogram is $|\det(A)|$. Since the area of the unit square is 1, we conclude A has *scaled* the area of the unit square by a factor of $|\det(A)|$.

The determinant of a 2×2 real matrix may thus be interpreted as a (signed) scaling factor of areas. In particular, notice that if the determinant is 0, then the area gets scaled by 0. This intuitively aligns with the fact that such a matrix is not invertible, since such a scaling cannot be reversed.

It is possible to give a similar interpretation of the $n \times n$ determinant of a real matrix as a (signed) scaling factor of volumes in $\mathbb{R}^n$, but we will not pursue this idea further in this course.

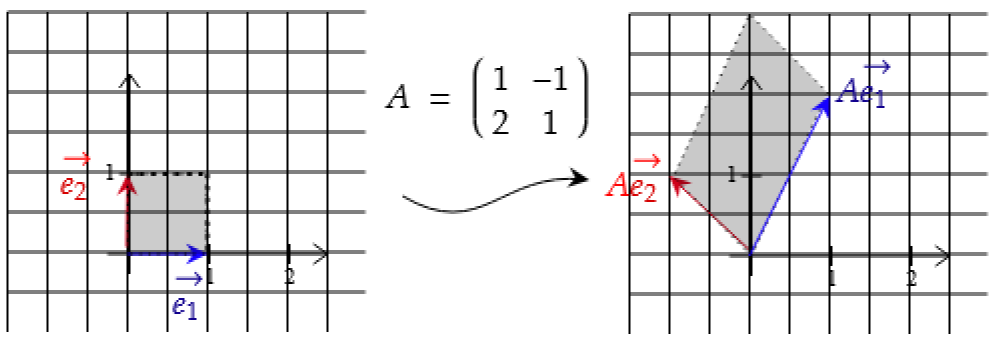


Figure 6.6.1: Effect of A on the unit square.

Chapter 7

Eigenvalues and Diagonalization

7.1 What is an Eigenpair?

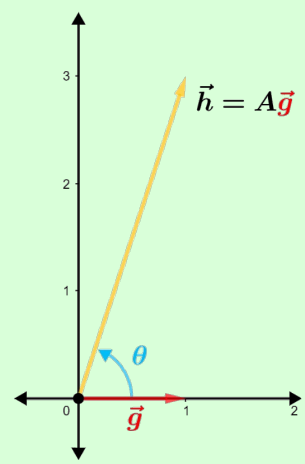
Let $A \in M_{n\times n}(\mathbb{R})$ be the standard matrix for the linear transformation T_A. We want to consider the effect that multiplication by A has on the vector $\vec{x} \in \mathbb{R}^n$; that is, we will compare the length and the direction of an input vector $\vec{x} \in \mathbb{R}^n$ to its image vector $\vec{y} = A\vec{x} \in \mathbb{R}^n$.

Usually, when $\vec{x}$ is multiplied by matrix A to produce $\vec{y} = A\vec{x}$, there will be changes to both the length and the direction. Exceptions to this pattern turn out to be particularly interesting, and will become the focus of the majority of this chapter.

Example 7.1.1 Let $A = \begin{bmatrix} 1 & 3 \\ 3 & 1 \end{bmatrix}$ and $\vec{g} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$. Then

$$\vec{h} = T_A(\vec{x}) = A\vec{g} = \begin{bmatrix} 1 & 3 \\ 3 & 1 \end{bmatrix}\begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \end{bmatrix}.$$

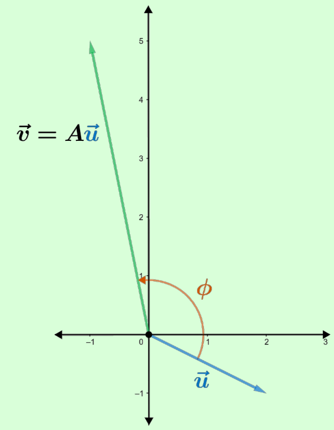
Since $\|\vec{g}\| = 1$ and the length of the image is $\|\vec{h}\| = \sqrt{10}$, the image $\vec{h}$ has been scaled by a factor of $\sqrt{10}$ relative to the length of the input vector $\vec{g}$.
The image $\vec{h}$ has a different direction than the input $\vec{g}$; the angle between $\vec{g}$ and $\vec{h}$ is
$\theta = \arccos\left(\frac{\vec{g}\cdot\vec{h}}{\|\vec{g}\|\,\|\vec{h}\|}\right) = \arccos\left(\frac{1}{\sqrt{10}}\right) \approx 1.249$ radians counter-clockwise.



Example 7.1.2 Keeping $A = \begin{bmatrix} 1 & 3 \\ 3 & 1 \end{bmatrix}$ from the previous example with a new input vector $\vec{u} = \begin{bmatrix} 2 \\ -1 \end{bmatrix}$, then

$$\vec{v} = T_A(\vec{u}) = A\,\vec{u} = \begin{bmatrix} 1 & 3 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ -1 \end{bmatrix} = \begin{bmatrix} -1 \\ 5 \end{bmatrix}.$$

Since $\|\vec{u}\| = \sqrt{5}$ and the length of the image is $\|\vec{v}\| = \sqrt{26}$, the image $\vec{v}$ has been scaled by a factor of $\sqrt{\frac{26}{5}}$ relative to the length of the input vector $\vec{u}$.
The image $\vec{v}$ has a different direction than the input $\vec{u}$; the angle between $\vec{u}$ and $\vec{v}$ is $\theta = \arccos\left(\frac{\vec{u}\cdot\vec{v}}{\|\vec{u}\|\,\|\vec{v}\|}\right) = \arccos\left(\frac{-7}{\sqrt{5}\sqrt{26}}\right) \approx 2.232$ radians counter-clockwise.



We now ask, for a given matrix $A \in M_{n\times n}(\mathbb{F})$, are there any input vectors whose direction is not changed when they are multiplied by A?

Example 7.1.3 Keeping $A = \begin{bmatrix} 1 & 3 \\ 3 & 1 \end{bmatrix}$ from the previous examples with a new input vector $\vec{z} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$, then

$$\vec{w} = T_A(\vec{z}) = A\vec{z} = \begin{bmatrix} 1 & 3 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 \\ 4 \end{bmatrix} = 4\begin{bmatrix} 1 \\ 1 \end{bmatrix} = 4\vec{z}.$$

The image $\vec{w}$ has been scaled by a factor of 4 relative to the length of the input vector $\vec{z}$. The image $\vec{w}$ has the same direction as the original vector $\vec{z}$; that is, multiplying A by $\vec{z}$ did not change the direction.

We include within our scope of interest image vectors that are in the opposite direction of the input vector.

Example 7.1.4 Keeping $A = \begin{bmatrix} 1 & 3 \\ 3 & 1 \end{bmatrix}$ from the previous examples with a new input vector $\vec{x} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$,

$$\vec{y} = T_A(\vec{x}) = A\vec{x} = \begin{bmatrix} 1 & 3 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} -2 \\ 2 \end{bmatrix} = -2\begin{bmatrix} 1 \\ -1 \end{bmatrix} = -2\vec{x}.$$

The image $\vec{y}$ has been scaled by a factor of -2 relative to the length of the input vector $\vec{x}$.

The image $\vec{y}$ has the opposite direction as the original vector $\vec{x}$.

In the previous two examples, we found vectors $\vec{z} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $\vec{x} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$ that "fit" with the matrix $A = \begin{bmatrix} 1 & 3 \\ 3 & 1 \end{bmatrix}$ in a special way, such that $A\vec{z}$ and $A\vec{x}$ are **scalar multiples** of $\vec{z}$ and $\vec{x}$ respectively.

REMARK

Given any matrix $A \in M_{n\times n}(\mathbb{F})$, we will seek vectors $\vec{x}$ such that $A\vec{x}$ is a scalar multiple of $\vec{x}$. For any such matrix A, $\vec{x} = \vec{0}$ is such a vector because $A\,\vec{0} = \vec{0} = c\,\vec{0}$ for any constant c; that is, $A\,\vec{0}$ is always a scalar multiple of $\vec{0}$. This fact is so trivial that we focus our attention on vectors $\vec{x} \neq \vec{0}$ such that $A\vec{x}$ is a scalar multiple of $\vec{x}$.

We now arrive at our main definitions.

Definition 7.1.5

Eigenvector, Eigenvalue and Eigenpair

Let $A \in M_{n\times n}(\mathbb{F})$. A *non-zero* vector $\vec{x}$ is an **eigenvector of A over $\mathbb{F}$** if there exists a scalar $\lambda \in \mathbb{F}$ such that

$$A\,\vec{x} = \lambda\,\vec{x}.$$

The scalar λ is then called an **eigenvalue of A over $\mathbb{F}$**, and the pair $(\lambda, \vec{x})$ is an **eigenpair of A over $\mathbb{F}$**.

Example 7.1.6

In Examples 7.1.3 and 7.1.4, we saw that $A = \begin{bmatrix} 1 & 3 \\ 3 & 1 \end{bmatrix}$ has eigenpairs $\left(4, \begin{bmatrix} 1 \\ 1 \end{bmatrix}\right)$ and $\left(-2, \begin{bmatrix} 1 \\ -1 \end{bmatrix}\right)$ over $\mathbb{R}$.

Interestingly enough, a square matrix with real entries may have eigenpairs with non-real eigenvalues.

Example 7.1.7

If $B = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$, then $(\lambda, \vec{x}) = \left(i, \begin{bmatrix} i \\ 1 \end{bmatrix}\right)$ is an eigenpair of B over $\mathbb{C}$, because

$$B\vec{x} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} i \\ 1 \end{bmatrix} = \begin{bmatrix} -1 \\ i \end{bmatrix} = i\begin{bmatrix} i \\ 1 \end{bmatrix} = \lambda\vec{x}.$$

You can verify that $\left(-i, \begin{bmatrix} -i \\ 1 \end{bmatrix}\right)$ is an eigenpair of B over $\mathbb{C}$ in a similar way. With the techniques that will be developed in this chapter you will be able to show that, even though B has entries in $\mathbb{R}$, it does not have any eigenpairs over $\mathbb{R}$.

7.2 The Characteristic Polynomial and Finding Eigenvalues

When looking for eigenpairs of a matrix A over $\mathbb{F}$, consider this list of equivalent equations which we solve over $\mathbb{F}$ for both $\vec{x}$ and λ. We denote I_n by I.

$$\begin{aligned} & A\vec{x} = \lambda\vec{x} \\ \Leftrightarrow\quad & A\vec{x} - \lambda\vec{x} = \vec{0} \\ \Leftrightarrow\quad & A\vec{x} - \lambda I\,\vec{x} = \vec{0} \\ \Leftrightarrow\quad & (A - \lambda I)\vec{x} = \vec{0} \end{aligned}$$

Thus, solving the equation $A\vec{x} = \lambda\vec{x}$ is equivalent to solving $(A - \lambda I)\vec{x} = \vec{0}$.

Definition 7.2.1 **Eigenvalue Equation or Eigenvalue Problem**

Let $A \in M_{n\times n}(\mathbb{F})$. We refer to the equation

$$A\,\vec{x} = \lambda\,\vec{x} \quad \text{or} \quad (A - \lambda I)\,\vec{x} = \vec{0}$$

as the **eigenvalue equation for the matrix A over $\mathbb{F}$**. It is also sometimes referred to as the **eigenvalue problem**.

REMARKS

- This is an unusual equation to solve since we want to solve it for both the vector $\vec{x} \in \mathbb{F}^n$ and the scalar $\lambda \in \mathbb{F}$. We will approach the problem by first identifying eligible values of λ. We can then determine corresponding sets of vectors $\vec{x}$ that solve the equation for each λ we identify.
- As mentioned in the Remark following Example 7.1.4, a trivial solution to this equation is $\vec{x} = \vec{0}$. We seek to obtain a non-trivial ($\vec{x} \neq \vec{0}$) solution to the eigenvalue equation. This is possible if and only if the RREF of matrix $A - \lambda I$ has fewer than n pivots, which occurs if and only if $A - \lambda I$ is not invertible; that is, if and only if $\det(A - \lambda I) = 0$.
- Computing $\det(A-\lambda I) = \det\left(\begin{bmatrix} a_{11}-\lambda & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22}-\lambda & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn}-\lambda \end{bmatrix}\right)$ by expanding along the first row, we sum up n terms, each of which is the product of entries in $A - \lambda I$. Since each entry is either a constant or a linear term in the variable λ, and all products present in the expanded determinant calculation contain n terms, we can infer that $\det(A - \lambda I)$ is a polynomial in λ of degree n. See Proposition 7.3.4 (Features of the Characteristic Polynomial) for a more precise statement and justification.

Definition 7.2.2 **Characteristic Polynomial and Characteristic Equation**

Let $A \in M_{n\times n}(\mathbb{F})$ and $\lambda \in \mathbb{F}$. The **characteristic polynomial of A**, denoted by $C_A(\lambda)$, is

$$C_A(\lambda) = \det(A - \lambda I).$$

The **characteristic equation of A** is

$$C_A(\lambda) = 0.$$

REMARK

The eigenvalues of A over $\mathbb{F}$ are the roots of the characteristic polynomial of A in $\mathbb{F}$.

Example 7.2.3 Find the eigenvalues of $A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$ over $\mathbb{R}$.

Solution: Since $A - \lambda I = \begin{bmatrix} 1-\lambda & 2 \\ 2 & 1-\lambda \end{bmatrix}$, the characteristic polynomial of A is

$$C_A(\lambda) = \det\left(\begin{bmatrix} 1-\lambda & 2 \\ 2 & 1-\lambda \end{bmatrix}\right) = (1-\lambda)^2 - 2^2 = \lambda^2 - 2\lambda - 3$$

and the characteristic equation of A is

$$C_A(\lambda) = \lambda^2 - 2\lambda - 3 = (\lambda - 3)(\lambda + 1) = 0.$$

The eigenvalues of A over $\mathbb{R}$ are the real roots of $C_A(\lambda)$, that is, $\lambda_1 = 3$ and $\lambda_2 = -1$.

Example 7.2.4 Find the eigenvalues of $B = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ over $\mathbb{R}$.

Solution: Since $B - \lambda I = \begin{bmatrix} -\lambda & -1 \\ 1 & -\lambda \end{bmatrix}$, the characteristic polynomial of B is

$$C_B(\lambda) = \det\left(\begin{bmatrix} -\lambda & -1 \\ 1 & -\lambda \end{bmatrix}\right) = \lambda^2 + 1$$

and the characteristic equation of B is

$$C_B(\lambda) = \lambda^2 + 1 = 0.$$

Since $C_B(\lambda)$ has no real roots, B has no eigenvalues and no eigenpairs over $\mathbb{R}$.

REMARK

Note that we can interpret B geometrically by observing that $\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$ for $\theta = \frac{\pi}{2}$.

Therefore, from a geometric perspective, B is the matrix which performs a rotation by $\frac{\pi}{2}$ radians counter-clockwise. Since $B\vec{x}$ can never be a scalar multiple of a nonzero vector $\vec{x} \in \mathbb{R}^2$ (because the transformation will always change the vector's direction), it makes sense that B would have no *real* eigenvalues nor eigenvectors.

Example 7.2.5 Keeping $B = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ from the previous example, find the eigenvalues of B over $\mathbb{C}$.

Solution: As in the previous example, $B - \lambda I = \begin{bmatrix} -\lambda & -1 \\ 1 & -\lambda \end{bmatrix}$ and

$$C_B(\lambda) = \det\left(\begin{bmatrix} -\lambda & -1 \\ 1 & -\lambda \end{bmatrix}\right) = \lambda^2 + 1.$$

We can factor $C_B(\lambda)$ fully because we are now working in $\mathbb{C}$; thus the characteristic equation of B is

$$C_B(\lambda) = \lambda^2 + 1 = (\lambda - i)(\lambda + i) = 0.$$

The complex roots of $C_B(\lambda)$ are the eigenvalues of B over $\mathbb{C}$. Thus, $\lambda_1 = i$ and $\lambda_2 = -i$.

Example 7.2.6 Find the eigenvalues of $G = \begin{bmatrix} 3i & -4 \\ 2 & i \end{bmatrix}$ over $\mathbb{C}$.

Solution: Since $G - \lambda I = \begin{bmatrix} 3i-\lambda & -4 \\ 2 & i-\lambda \end{bmatrix}$, the characteristic polynomial of G is

$$C_G(\lambda) = \det\left(\begin{bmatrix} 3i-\lambda & -4 \\ 2 & i-\lambda \end{bmatrix}\right) = (3i-\lambda)(i-\lambda) + 8 = \lambda^2 - 4i\lambda + 5$$

and the characteristic equation is

$$C_G(\lambda) = \lambda^2 - 4i\lambda + 5 = (\lambda - 5i)(\lambda + i) = 0.$$

The complex roots of $C_G(\lambda)$ are the eigenvalues of G over $\mathbb{C}$. Thus, $\lambda_1 = 5i$ and $\lambda_2 = -i$.

Example 7.2.7 Find the eigenvalues of $H = \begin{bmatrix} 4 & 2 & -6 \\ 1 & -2 & 1 \\ -6 & 2 & 4 \end{bmatrix}$ over $\mathbb{R}$.

Solution: We find the characteristic polynomial of H by calculating the determinant of $H - \lambda I = \begin{bmatrix} 4-\lambda & 2 & -6 \\ 1 & -2-\lambda & 1 \\ -6 & 2 & 4-\lambda \end{bmatrix}$, using an expansion along the first row:

$$\begin{aligned} C_H(\lambda) &= \det\left(\begin{bmatrix} 4-\lambda & 2 & -6 \\ 1 & -2-\lambda & 1 \\ -6 & 2 & 4-\lambda \end{bmatrix}\right) \\ &= (4-\lambda)[(-2-\lambda)(4-\lambda) - 2] - 2[1(4-\lambda) - 1(-6)] - 6[1(2) - (-2-\lambda)(-6)] \\ &= -\lambda^3 + 6\lambda^2 + 40\lambda. \end{aligned}$$

The characteristic equation is

$$C_H(\lambda) = -\lambda^3 + 6\lambda^2 + 40\lambda = -\lambda(\lambda^2 - 6\lambda - 40) = -\lambda(\lambda - 10)(\lambda + 4) = 0.$$

The real roots are the eigenvalues of H over $\mathbb{R}$. Thus, $\lambda_1 = 10, \lambda_2 = 0$ and $\lambda_3 = -4$.

7.3 Properties of the Characteristic Polynomial

There is a lot of arithmetic involved in finding eigenpairs and it is wise to make sure that our characteristic polynomial is actually correct before we begin factoring it. This section explores properties of the characteristic polynomial that will allow us to verify our work. We begin with the following result.

Proposition 7.3.1

Let $A \in M_{n\times n}(\mathbb{F})$. Then A is invertible if and only if $\lambda = 0$ is not an eigenvalue of A.

Proof: A is invertible iff $\det(A) \neq 0$.
iff $\det(A - 0\,I_n) \neq 0$
iff 0 is not a root of the characteristic polynomial
iff 0 is not an eigenvalue of the matrix A.

□

Definition 7.3.2
Trace

Let $A \in M_{n\times n}(\mathbb{F})$. We define the **trace** of A by

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

That is, the trace of a square matrix is the sum of its diagonal entries.

Example 7.3.3

Let $A = \begin{bmatrix} 1 & 2 & 7 \\ 3 & 4 & -5 \\ 4 & 0 & 6 \end{bmatrix}$. Then $\operatorname{tr}(A) = 1 + 4 + 6 = 11$.

Proposition 7.3.4

(Features of the Characteristic Polynomial)

Let $A \in M_{n\times n}(\mathbb{F})$ have characteristic polynomial $C_A(\lambda) = \det(A - \lambda I)$. Then $C_A(\lambda)$ is a degree n polynomial in λ of the form

$$C_A(\lambda) = c_n\lambda^n + c_{n-1}\lambda^{(n-1)} + \cdots + c_1\lambda + c_0,$$

where

(a) $c_n = (-1)^n$,

(b) $c_{n-1} = (-1)^{(n-1)}\operatorname{tr}(A)$, and

(c) $c_0 = \det(A)$.

Proof: By performing cofactor expansion along the first row of A and along the first row of every subsequent $(1, 1)$-submatrix, we obtain an expression of the form

$$C_A(\lambda) = \det(A - \lambda I) = (a_{11} - \lambda)(a_{22} - \lambda)\cdots(a_{nn} - \lambda) + f(\lambda),$$

where f is the polynomial that is the sum of the other terms of the determinant. In the above, we have singled out the contribution of the product of the diagonal entries, and have grouped together all other terms. This latter grouping $f(\lambda)$ is a polynomial of degree at most $n - 2$. To see this, note that unless the cofactor calculation is done with respect to a diagonal entry, it will use an entry (i, j) that corresponds to deleting the entry $(a_{ii} - \lambda)$

$$\begin{bmatrix} a_{11}-\lambda & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22}-\lambda & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn}-\lambda \end{bmatrix}$$

Figure 7.3.1: Visualization of taking the $(1,2)$-submatrix of $A - \lambda I$

from the i^{th} row and the entry $(a_{jj} - \lambda)$ from the j^{th} column. This leaves us with at most $n-2$ entries that contain a λ. We illustrate this below using the $(1,2)$-entry as an example.

Returning to our expression for $C_A(\lambda)$, we can expand out the first term in the expression to obtain

$$C_A(\lambda) = a_{11} \cdots a_{nn} + \cdots + (-1)^{n-1}(a_{11} + \cdots + a_{nn})\lambda^{n-1} + (-1)^n \lambda^n + f(\lambda).$$

From this we see that $C_A(\lambda)$ is a polynomial of degree n in λ (because we've shown that this degree is not exceeded by the polynomial f, which is of degree at most $n-2$). Moreover, we obtain the following results:

(a) The coefficient of λ^n is $c_n = (-1)^n$.

(b) The coefficient of λ^{n-1} is

$$c_{n-1} = (-1)^{n-1}(a_{11} + \cdots + a_{nn}) = (-1)^{n-1}\,\mathrm{tr}(A).$$

(c) From the definition of $C_A(\lambda)$, we immediately have that the constant term is

$$c_0 = C_A(0) = \det(A - 0I) = \det(A).$$

□

Example 7.3.5 Determine the characteristic polynomial of $A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 10 \end{bmatrix}$ and verify the properties outlined in Proposition 7.3.4 (Features of the Characteristic Polynomial).

Solution: The characteristic polynomial of A is

$$\begin{aligned} C_A(\lambda) &= \det \begin{bmatrix} 1-\lambda & 2 & 3 \\ 4 & 5-\lambda & 6 \\ 7 & 8 & 10-\lambda \end{bmatrix} \\ &= (1-\lambda)\left[(5-\lambda)(10-\lambda) - 48\right] - 2\left[4(10-\lambda) - 42\right] + 3\left[32 - 7(5-\lambda)\right] \\ &= -\lambda^3 + 16\lambda^2 + 12\lambda - 3. \end{aligned}$$

Observe that $C_A(\lambda)$ has degree 3, that $\mathrm{tr}(A) = 1 + 5 + 10 = 16$ and that $\det(A) = -3$ as we determined in Example 6.1.8.

We now verify our calculation of $C_A(\lambda)$ using each part of Proposition 7.3.4 (Features of the Characteristic Polynomial):

(a) The coefficient of λ^3 is $c_3 = -1 = (-1)^3$.

(b) The coefficient of λ^2 is $c_2 = 16 = (-1)^2 \operatorname{tr}(A)$.

(c) The constant term is $c_0 = -3 = \det(A)$.

We can now be more confident that our characteristic polynomial was computed correctly. Note that this does not prove that our calculation was correct; this is simply a nice check for "red flags".

The characteristic polynomial $C_A(\lambda)$ of an $n \times n$ matrix A is a degree n polynomial. By the Fundamental Theorem of Algebra, the characteristic polynomial $C_A(\lambda)$ is guaranteed to have n complex roots (though some may be repeated) as we assume $\mathbb{F} \subseteq \mathbb{C}$. Therefore, A is guaranteed to have n eigenvalues in $\mathbb{C}$. Some of these eigenvalues may be repeated, in which case we list each eigenvalue as many times as its corresponding linear factor appears in the characteristic polynomial $C_A(\lambda)$; Example 7.3.12 will address the case of repeated eigenvalues.

The following result establishes a connection between the characteristic polynomial of a matrix and its n complex eigenvalues over $\mathbb{C}$.

Proposition 7.3.6 **(Characteristic Polynomial and Eigenvalues over $\mathbb{C}$)**

Let $A \in M_{n\times n}(\mathbb{F})$ have characteristic polynomial

$$C_A(\lambda) = c_n\lambda^n + c_{n-1}\lambda^{n-1} + \cdots + c_1\lambda + c_0,$$

and n eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ (possibly repeated) in $\mathbb{C}$. Then

(a) $c_{n-1} = (-1)^{(n-1)} \sum_{i=1}^{n} \lambda_i$, and

(b) $c_0 = \prod_{i=1}^{n} \lambda_i$.

Note that if A has repeated eigenvalues over $\mathbb{C}$, then we include each eigenvalue in the list $\lambda_1, \lambda_2, ..., \lambda_n$ as many times as its corresponding linear factor appears in the characteristic polynomial $C_A(\lambda)$.

Proof: The eigenvalues of A over $\mathbb{C}$ are the n complex roots of the characteristic polynomial, so its characteristic polynomial has the form

$$C_A(\lambda) = k(\lambda - \lambda_1)(\lambda - \lambda_2)\cdots(\lambda - \lambda_n) \text{ for some } k \in \mathbb{C}.$$

By Parts (a) and (b) of Proposition 7.3.4 (Features of the Characteristic Polynomial), the coefficient of λ^n is
$c_n = (-1)^n$. It must be that $k = (-1)^n$, so

$$C_A(\lambda) = (-1)^n(\lambda - \lambda_1)(\lambda - \lambda_2)\cdots(\lambda - \lambda_n).$$

(a) Consider c_{n-1}, the coefficient of λ^{n-1} in $C_A(\lambda)$. By expanding out the expression for $C_A(\lambda)$, we find that there are n terms involving λ^{n-1}, each of which is the product of $(-1)^n$ with one of the constants $-\lambda_1, -\lambda_2, \ldots, -\lambda_n$ and with λ from each of the other $n-1$ linear factors. By taking the sum of these terms, we find that

$$\begin{aligned} c_{n-1}\lambda^{n-1} &= (-1)^n[-\lambda_1\,(\lambda)^{n-1} - \lambda_2\,(\lambda)^{n-1} - \cdots - \lambda_n\,(\lambda)^{n-1}] \\ &= \left((-1)^{(n-1)}\sum_{i=1}^{n}\lambda_i\right)\lambda^{n-1} \end{aligned}$$

that is, $c_{n-1} = (-1)^{(n-1)}\sum_{i=1}^{n}\lambda_i$.

(b) Expanding the terms of $C_A(\lambda)$, its constant term must be $(-1)^n$ times the product of the constant terms in each of the n linear factors, which are $-\lambda_1, -\lambda_2, \ldots, -\lambda_n$. Therefore,

$$c_0 = (-1)^n\prod_{i=1}^{n}(-\lambda_i) = (-1)^{2n}\prod_{i=1}^{n}\lambda_i = \prod_{i=1}^{n}\lambda_i.$$

□

Propositions 7.3.4 and 7.3.6 combine to form the following corollary, the proof of which we leave as an exercise.

Corollary 7.3.7 **(Eigenvalues and Trace/Determinant)**

Let $A \in M_{n\times n}(\mathbb{F})$ have n eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ (possibly repeated) in $\mathbb{C}$. Show that:

(a) $\displaystyle\sum_{i=1}^{n}\lambda_i = \operatorname{tr}(A)$.

(b) $\displaystyle\prod_{i=1}^{n}\lambda_i = \det(A)$.

In the following examples, we confirm our previous calculations using Proposition 7.3.6 (Characteristic Polynomial and Eigenvalues over $\mathbb{C}$).

Example 7.3.8 In Example 7.2.5, the eigenvalues of $B = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ over $\mathbb{C}$ are $\lambda_1 = i$ and $\lambda_2 = -i$, and $C_B(\lambda) = \lambda^2 + 1$ is a quadratic (degree 2) polynomial. Proposition 7.3.6 and Corollary 7.3.7 confirm that

(a) the coefficient of λ is $(-1)^1\sum_{i=1}^{2}\lambda_i = -(\lambda_1+\lambda_2) = -(i-i) = 0 = c_1$ and $\operatorname{tr}(B) = \lambda_1 + \lambda_2 = 0$.

(b) the constant term is $\prod_{i=1}^{2}\lambda_i = \lambda_1\lambda_2 = (i)(-i) = -i^2 = 1 = c_0$ and $1 = \det(B)$.

Example 7.3.9

In Example 7.2.6, the eigenvalues of $G = \begin{bmatrix} 3i & -4 \\ 2 & i \end{bmatrix}$ over $\mathbb{C}$ are $\lambda_1 = 5i$ and $\lambda_2 = -i$ and $C_G(\lambda) = \lambda^2 - 4i\lambda + 5$ is a degree 2 polynomial. hyperref[prop:complex-eigenvals-coeffs-char-poly]Proposition 7.3.6 and Corollary 7.3.7 confirm that

(a) the coefficient of λ is $(-1)^1 \sum_{i=1}^{2} \lambda_i = -(\lambda_1 + \lambda_2) = -(5i - i) = -(4i) = c_1$ and $\operatorname{tr}(G) = \lambda_1 + \lambda_2 = 4i$.

(b) the constant term is $\prod_{i=1}^{2} \lambda_i = \lambda_1 \lambda_2 = 5i(-i) = 5 = c_0$ and $5 = \det(G)$.

In the next two examples, we will look at the eigenvalues of a matrix A over $\mathbb{R}$, which are the real roots of the characteristic polynomial. Since $\mathbb{R}$ is a subset of $\mathbb{C}$, any real root of a polynomial can also be considered as a complex root of the polynomial; thus, any real eigenvalue of a matrix A can be considered to be a complex eigenvalue. As a result, Proposition 7.3.6 (Characteristic Polynomial and Eigenvalues over $\mathbb{C}$) applies in these cases.

Example 7.3.10

In Example 7.2.3, $A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$ has a degree two characteristic polynomial $C_A(\lambda) = \lambda^2 - 2\lambda - 3 = (\lambda - 3)(\lambda + 1)$, with two real roots $\lambda_1 = 3$ and $\lambda_2 = -1$. These are the eigenvalues of A over $\mathbb{C}$ (as well as over $\mathbb{R}$). Proposition 7.3.6 confirms that

(a) the coefficient of λ is $(-1)^1 \sum_{i=1}^{2} \lambda_i = -(\lambda_1 + \lambda_2) = -(3 - 1) = -2 = c_1$.

(b) the constant term is $\prod_{i=1}^{2} \lambda_i = \lambda_1 \lambda_2 = (3)(-1) = -3 = c_0$.

Be careful to note that a matrix A can have both real and non-real eigenvalues. In applying Proposition 7.3.6, we must use *all* of the eigenvalues of A.

Example 7.3.11

The characteristic polynomial of $J = \begin{bmatrix} -4 & 0 & 0 \\ 0 & 0 & 4 \\ 0 & -4 & 0 \end{bmatrix}$ is

$$C_J(\lambda) = \det\left(\begin{bmatrix} -4-\lambda & 0 & 0 \\ 0 & -\lambda & 4 \\ 0 & -4 & -\lambda \end{bmatrix} \right) = -(\lambda + 4)(\lambda^2 + 16) = -\lambda^3 - 4\lambda^2 - 16\lambda - 64.$$

Its degree is 3, and it has one real root $\lambda_1 = -4$ and two non-real complex roots $\lambda_2 = 4i$ and $\lambda_3 = -4i$. These are the eigenvalues of J over $\mathbb{C}$. Proposition 7.3.6 confirms that

(a) the coefficient of λ^2 is $(-1)^2 \sum_{i=1}^{3} \lambda_i = \lambda_1 + \lambda_2 + \lambda_3 = -4 + 4i - 4i = -4 = c_2$.

(b) the constant term is $\prod_{i=1}^{3} \lambda_i = \lambda_1 \lambda_2 \lambda_3 = (-4)(4i)(-4i) = -64 = c_0$.

For our following example, the characteristic polynomial includes a repeated linear factor. In order to use Proposition 7.3.6 (Characteristic Polynomial and Eigenvalues over $\mathbb{C}$), we include each eigenvalue in our list $\lambda_1, \lambda_2, ..., \lambda_n$ as many times as its corresponding linear factor appears in the characteristic polynomial $C_A(\lambda)$. We will address this concept later in terms of **multiplicity**.

Example 7.3.12

The characteristic polynomial of $Z = \begin{bmatrix} 3 & 1 & 1 \\ 1 & 3 & 1 \\ 1 & 1 & 3 \end{bmatrix}$ is

$$\begin{aligned} C_Z(\lambda) &= \det\left(\begin{bmatrix} 3-\lambda & 1 & 1 \\ 1 & 3-\lambda & 1 \\ 1 & 1 & 3-\lambda \end{bmatrix} \right) \\ &= (3-\lambda)[(3-\lambda)^2 - 1] - 1[(3-\lambda) - 1] + 1[1 - (3-\lambda)] \\ &= -\lambda^3 + 9\lambda^2 - 24\lambda + 20 \\ &= -(\lambda - 5)(\lambda - 2)^2. \end{aligned}$$

Since $n = 3$, we must carefully write down the three eigenvalues corresponding to the roots of $C_Z(\lambda)$, allowing repeats. Since the linear factor $\lambda - 2$ occurs twice in $C_Z(\lambda)$, we list the eigenvalues as $\lambda_1 = 5$, $\lambda_2 = 2$, and $\lambda_3 = 2$. Proposition 7.3.6 confirms that

(a) the coefficient of λ^2 is $(-1)^2 \sum_{i=1}^{3} \lambda_i = \lambda_1 + \lambda_2 + \lambda_3 = 5 + 2 + 2 = 9 = c_2$.

(b) the constant term is $\prod_{i=1}^{3} \lambda_i = \lambda_1 \lambda_2 \lambda_3 = 5(2)(2) = 20 = c_0$.

7.4 Finding Eigenvectors

Once we have found an eigenvalue λ of a matrix A over $\mathbb{F}$, we can examine the eigenvalue equation in the form $(A - \lambda I)\vec{x} = \vec{0}$ in order to obtain a corresponding eigenvector $\vec{x}$.

Example 7.4.1

For each eigenvalue of $A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$ over $\mathbb{R}$, find a corresponding eigenpair.

Solution: From Example 7.2.3, the eigenvalues of A over $\mathbb{R}$ are $\lambda_1 = 3$ and $\lambda_2 = -1$.

For each eigenvalue λ_i, we examine the eigenvalue equation to determine an eigenvector.

- When $\lambda_1 = 3$, we examine $(A - 3I)\vec{x} = \vec{0}$ which is equivalent to

$$\begin{bmatrix} -2 & 2 \\ 2 & -2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \vec{0}$$

and row-reduces to

$$\begin{bmatrix} 1 & -1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \vec{0}.$$

The general solution is $\left\{ s \begin{bmatrix} 1 \\ 1 \end{bmatrix} : s \in \mathbb{R} \right\}$ and an eigenpair is $\left(3, \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right)$.

- When $\lambda_2 = -1$, we examine $(A - (-1)I)\vec{x} = \vec{0}$ which is equivalent to

$$\begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \vec{0}$$

and row-reduces to

$$\begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \vec{0}.$$

The general solution is $\left\{ t \begin{bmatrix} 1 \\ -1 \end{bmatrix} : t \in \mathbb{R} \right\}$ and an eigenpair is $\left(-1, \begin{bmatrix} 1 \\ -1 \end{bmatrix} \right)$.

Checking these eigenpairs, we have that

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

and

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

REMARK

While $\left(3, \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right)$ is an eigenpair of A, we can get other eigenpairs by pairing non-zero scalar multiples of the eigenvector $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ with the eigenvalue $\lambda = 3$. For example, $\left(3, \begin{bmatrix} 2 \\ 2 \end{bmatrix} \right)$ and $\left(3, \begin{bmatrix} -4 \\ -4 \end{bmatrix} \right)$ are eigenpairs of A over $\mathbb{R}$, since

$$A \begin{bmatrix} 2 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 2 \end{bmatrix} = \begin{bmatrix} 6 \\ 6 \end{bmatrix} = 3 \begin{bmatrix} 2 \\ 2 \end{bmatrix}$$

and

$$A \begin{bmatrix} -4 \\ -4 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} -4 \\ -4 \end{bmatrix} = \begin{bmatrix} -12 \\ -12 \end{bmatrix} = 3 \begin{bmatrix} -4 \\ -4 \end{bmatrix}.$$

We formalize this idea later in Proposition 7.5.1 (Linear Combinations of Eigenvectors).

Example 7.4.2

For each eigenvalue of $B = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ over $\mathbb{C}$, find a corresponding eigenpair.

Solution: From Example 7.2.4, the eigenvalues of B over $\mathbb{C}$ are $\lambda_1 = i$ and $\lambda_2 = -i$.

For each eigenvalue λ_i, we examine the eigenvalue equation to determine an eigenvector.

- When $\lambda_1 = i$, we examine $(B - iI)\vec{x} = \vec{0}$ which is equivalent to

$$\begin{bmatrix} -i & -1 \\ 1 & -i \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \vec{0}$$

and row-reduces to

$$\begin{bmatrix} 1 & -i \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \vec{0}.$$

The general solution is $\left\{ s \begin{bmatrix} i \\ 1 \end{bmatrix} : s \in \mathbb{C} \right\}$ and an eigenpair is $\left(i, \begin{bmatrix} i \\ 1 \end{bmatrix} \right)$.

- When $\lambda_2 = -i$, we examine $(B - (-i)I)\,\vec{x} = \vec{0}$ which is equivalent to

$$\begin{bmatrix} i & -1 \\ 1 & i \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \vec{0}$$

and row-reduces to

$$\begin{bmatrix} 1 & i \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \vec{0}.$$

The general solution is $\left\{ t \begin{bmatrix} -i \\ 1 \end{bmatrix} : t \in \mathbb{C} \right\}$ and an eigenpair is $\left(-i, \begin{bmatrix} -i \\ 1 \end{bmatrix} \right)$.

Checking these eigenpairs, we have that

$$B \begin{bmatrix} i \\ 1 \end{bmatrix} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} i \\ 1 \end{bmatrix} = \begin{bmatrix} -1 \\ i \end{bmatrix} = i \begin{bmatrix} i \\ 1 \end{bmatrix}$$

and

$$B \begin{bmatrix} -i \\ 1 \end{bmatrix} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} -i \\ 1 \end{bmatrix} = \begin{bmatrix} -1 \\ -i \end{bmatrix} = -i \begin{bmatrix} -i \\ 1 \end{bmatrix}.$$

Example 7.4.3

For each eigenvalue of $Z = \begin{bmatrix} 3 & 1 & 1 \\ 1 & 3 & 1 \\ 1 & 1 & 3 \end{bmatrix}$ over $\mathbb{R}$, find a corresponding eigenpair.

Solution: From Example 7.3.12, the characteristic polynomial of Z is $C_Z(\lambda) = -(\lambda - 5)(\lambda - 2)^2$. Therefore, the eigenvalues of Z over $\mathbb{R}$ are $\lambda_1 = 5$, $\lambda_2 = 2$ and $\lambda_3 = 2$. (Note that the eigenvalue 2 occurs twice.)

For each eigenvalue λ_i, we examine the eigenvalue equation to determine an eigenvector.

- When $\lambda_1 = 5$, we solve the system $(Z - 5I)\vec{x} = \vec{0}$ which is equivalent to

$$\begin{bmatrix} -2 & 1 & 1 \\ 1 & -2 & 1 \\ 1 & 1 & -2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \vec{0}$$

  and row-reduces to

$$\begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \vec{0}.$$

  The solution set is $\left\{ s \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} : s \in \mathbb{R} \right\}$ and an eigenpair is $\left(5, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right)$.

- When $\lambda_2 = 2$, we solve the system $(Z - 2I)\vec{x} = \vec{0}$ which is equivalent to

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \vec{0}$$

  and row-reduces to

$$\begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \vec{0}.$$

  The solution set is $\left\{ t \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} + u \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} : t, u \in \mathbb{R} \right\}$ and an eigenpair is $\left(2, \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} \right)$.

- No new computation is needed for $\lambda_3 = 2$. However, note that we can obtain another eigenpair $\left(2, \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \right)$ where the eigenvector $\begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$ is not a scalar multiple of the eigenvector $\begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}$ obtained in the previous step. We are able to do this because the solution set to the eigenvalue equation has two parameters in it. This feature of the solution set is related to the fact that the eigenvalue $\lambda_2 = \lambda_3 = 2$ is a repeated root of the characteristic polynomial. We will study this phenomenon in more detail later.

7.5 Eigenspaces

After Example 7.2.3, we noted that $A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$ has an eigenpair $\left(3, \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right)$. By taking scalar multiples of the eigenvector $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$, we generated $\left(3, \begin{bmatrix} 2 \\ 2 \end{bmatrix} \right)$ and $\left(3, \begin{bmatrix} -4 \\ -4 \end{bmatrix} \right)$ that are also eigenpairs of A over $\mathbb{R}$ with the same eigenvalue $\lambda = 3$. More generally, we have the following result.

Proposition 7.5.1 **(Linear Combinations of Eigenvectors)**

Let $c, d \in \mathbb{F}$ and suppose that $(\lambda_1, \vec{x})$ and $(\lambda_1, \vec{y})$ are eigenpairs of a matrix A over $\mathbb{F}$ with the same eigenvalue λ_1. If $c\vec{x} + d\vec{y} \neq \vec{0}$, then $(\lambda_1, c\vec{x} + d\vec{y})$ is also an eigenpair for A with eigenvalue λ_1.

Proof: Observe that

$$A(c\vec{x} + d\vec{y}) = c(A\vec{x}) + d(A\vec{y}) = c(\lambda_1\vec{x}) + d(\lambda_1\vec{y}) = \lambda_1(c\vec{x} + d\vec{y}).$$

Thus, since $c\vec{x} + d\vec{y} \neq 0$, it follows that $c\vec{x} + d\vec{y}$ is an eigenvector of A with eigenvalue λ_1. That is, $(\lambda_1, c\vec{x} + d\vec{y})$ is an eigenpair for A. $\square$

Example 7.5.2 If we take $d = 0$ and $c \neq 0$ in Proposition 7.5.1, we see that every non-zero scalar multiple of an eigenvector is an eigenvector (with the same eigenvalue).

For instance, since $\left(3, \begin{bmatrix} 1 \\ 1 \end{bmatrix}\right)$ is an eigenpair for $A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$, then so is $\left(3, \begin{bmatrix} c \\ c \end{bmatrix}\right)$ for any $c \neq 0$ in $\mathbb{F}$.

Since every non-zero linear combination of eigenvectors with respect to a fixed eigenvalue λ are also eigenvectors with respect to λ, this suggests collecting all possible eigenvectors of A with respect to a value λ into a set.

Definition 7.5.3
Eigenspace

Let $A \in M_{n\times n}(\mathbb{F})$ and let $\lambda \in \mathbb{F}$. The **eigenspace of A associated with λ**, denoted by $E_\lambda(A)$, is the solution set to the system $(A - \lambda I)\vec{x} = \vec{0}$ over $\mathbb{F}$. That is,

$$E_\lambda(A) = \mathrm{Null}(A - \lambda I).$$

If the choice of A is clear, we abbreviate this as E_λ.

REMARKS

- Notice that $\vec{0}$ is always in E_λ. However, by definition $\vec{0}$ is not an eigenvector. Thus the eigenspace E_λ consists of all the eigenvectors that have eigenvalue λ *together with the zero vector.*
- λ is an eigenvalue for A if and only if $E_\lambda \neq \{\vec{0}\}$.
- $E_0(A) = \mathrm{Null}(A - 0I) = \mathrm{Null}(A)$.

Example 7.5.4 In Example 7.4.1, we found that $\left(3, \begin{bmatrix} 1 \\ 1 \end{bmatrix}\right)$ and $\left(-1, \begin{bmatrix} 1 \\ -1 \end{bmatrix}\right)$ are eigenpairs of $A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$ over $\mathbb{R}$. Our work in that Example in fact shows that the eigenspaces of A are $E_3 = \operatorname{Span}\left\{\begin{bmatrix} 1 \\ 1 \end{bmatrix}\right\}$ and $E_{-1} = \operatorname{Span}\left\{\begin{bmatrix} 1 \\ -1 \end{bmatrix}\right\}$.

Example 7.5.5 Our work in Example 7.4.2 shows that the eigenspaces of $B = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ over $\mathbb{C}$ are $E_i = \operatorname{Span}\left\{\begin{bmatrix} i \\ 1 \end{bmatrix}\right\}$ and $E_{-i} = \operatorname{Span}\left\{\begin{bmatrix} -i \\ 1 \end{bmatrix}\right\}$.

Example 7.5.6 In Example 7.4.3, the eigenspaces of $Z = \begin{bmatrix} 3 & 1 & 1 \\ 1 & 3 & 1 \\ 1 & 1 & 3 \end{bmatrix}$ over $\mathbb{R}$ are

$$E_5 = \left\{ s \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} : s \in \mathbb{R} \right\} = \operatorname{Span}\left\{\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}\right\}$$

and

$$E_2 = \left\{ t \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} + u \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} : t, u \in \mathbb{R} \right\} = \operatorname{Span}\left\{\begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}\right\}.$$

7.6 Diagonalization

In this section we will demonstrate one practical use of eigenvalues and eigenvectors. (There are *many* more that you can learn about in your future courses!)

In certain applications of linear algebra, it is necessary to compute powers A^k of a square matrix $A \in M_{n\times n}(\mathbb{F})$ efficiently. It would be ideal if we could compute A^k without first having to compute each of $A^2, A^3, A^4, \ldots, A^{k-1}$. In some special situations, this is possible. For instance, if A is a diagonal matrix, then computing A^k is particularly straightforward: simply take the k^{th} powers of the diagonal entries.

EXERCISE

Prove that if $D = \operatorname{diag}(d_1, \ldots, d_n) \in M_{n\times n}(\mathbb{F})$, then $D^k = \operatorname{diag}(d_1^k, \ldots, d_n^k)$ for all $k \in \mathbb{N}$.

As the following example demonstrates, computing powers of a matrix $A \in M_{n\times n}(\mathbb{F})$ is simpler not only when A is diagonal, but also when A is of the form $A = PDP^{-1}$, where P is invertible and D is diagonal.

Example 7.6.1

Consider the matrices $P = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$ and $D = \begin{bmatrix} 3 & 0 \\ 0 & -1 \end{bmatrix}$. (Observe that D is a diagonal matrix.) Let $A = PDP^{-1} = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$. Then,

$$A^2 = AA = (PDP^{-1})(PDP^{-1}) = PD(P^{-1}P)DP^{-1} = PD^2P^{-1},$$

and therefore

$$A^3 = AA^2 = (PDP^{-1})(PD^2P^{-1}) = PD^3P^{-1}.$$

One can show using induction (exercise: do this!) that we have the following general formula for A^k:

$$A^k = PD^kP^{-1} \quad \text{for all } k \in \mathbb{N}.$$

Now, the key observation is: since D is a *diagonal* matrix, we have

$$D^k = \begin{bmatrix} 3^k & 0 \\ 0 & (-1)^k \end{bmatrix}.$$

Consequently,

$$\begin{aligned} A^k = PD^kP^{-1} &= \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 3^k & 0 \\ 0 & (-1)^k \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}^{-1} \\ &= \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 3^k & 0 \\ 0 & (-1)^k \end{bmatrix} \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{bmatrix} \\ &= \frac{1}{2} \begin{bmatrix} 3^k + (-1)^k & 3^k - (-1)^k \\ 3^k - (-1)^k & 3^k + (-1)^k \end{bmatrix}. \end{aligned}$$

In the previous Example, the relationship between A and D given by the equation $A = PDP^{-1}$ made the computation of A^k a rather straightforward matter. We shall give such a relationship a name. Of course, the key problem is *how to find* such a pair of matrices P and D given a matrix A, if this is even possible. We will return to this problem shortly.

Definition 7.6.2
Similar

Let $A, B \in M_{n\times n}(\mathbb{F})$. We say that A **is similar to** B **over** $\mathbb{F}$ if there exists an invertible matrix $P \in M_{n\times n}(\mathbb{F})$ such that $A = PBP^{-1}$.

REMARKS

- Notice that if $A = PBP^{-1}$ then multiplying this equation on the left by P^{-1} and on the right by P gives $P^{-1}AP = B$, or equivalently, $QAQ^{-1} = B$ where $Q = P^{-1}$. Thus if A is similar to B, then B is similar to A and we can just say that A and B are similar to each other.

- The choice of "similar" as terminology for the relationship in Definition 7.6.2 might seem strange at first sight. There is in fact good reason for this choice, as we will later learn. (See the Remark on page 234.) In the Exercise below, you are asked to show that similar matrices share a variety of features. Specifically, they have the same determinant, trace, eigenvalues and characteristic polynomial.

Example 7.6.3 Example 7.6.1 shows that $A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$ is similar to $D = \begin{bmatrix} 3 & 0 \\ 0 & -1 \end{bmatrix}$ over $\mathbb{R}$.

Example 7.6.4 Let $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$. Consider $P = \begin{bmatrix} 1 & 2 \\ 4 & 6 \end{bmatrix}$, with $P^{-1} = -\frac{1}{2}\begin{bmatrix} 6 & -2 \\ -4 & 1 \end{bmatrix}$, and let

$$B = P^{-1}AP = \begin{bmatrix} 1 & 2 \\ 4 & 6 \end{bmatrix}^{-1} \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 4 & 6 \end{bmatrix} = \frac{1}{2}\begin{bmatrix} 16 & 24 \\ -17 & -26 \end{bmatrix}.$$

Then $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$ is similar to $B = \frac{1}{2}\begin{bmatrix} 16 & 24 \\ -17 & -26 \end{bmatrix}$ over $\mathbb{R}$.

EXERCISE

Let $A, B, C \in M_{n\times n}(\mathbb{F})$. Show that:

(a) A is similar to A over $\mathbb{F}$.

(b) If A is similar to B over $\mathbb{F}$ and B is similar to C over $\mathbb{F}$, then A is similar to C over $\mathbb{F}$.

EXERCISE

Let $A, B \in M_{n\times n}(\mathbb{F})$. Show that if A is similar to B over $\mathbb{F}$, then:

(a) For all $k \in \mathbb{N}$, A^k is similar to B^k over $\mathbb{F}$.

(b) $C_A(\lambda) = C_B(\lambda)$.

(c) A and B have the same eigenvalues.

(d) $\mathrm{tr}(A) = \mathrm{tr}(B)$ and $\det(A) = \det(B)$.

Let us return to the problem of computing A^k. In view of Example 7.6.1, the problem becomes much easier if we can show that A is similar to a diagonal matrix D. We will single out such matrices.

Definition 7.6.5 **Diagonalizable Matrix**

Let $A \in M_{n\times n}(\mathbb{F})$. We say that A is **diagonalizable over** $\mathbb{F}$ if it is similar over $\mathbb{F}$ to a diagonal matrix $D \in M_{n\times n}(\mathbb{F})$; that is, if there exists an invertible matrix $P \in M_{n\times n}(\mathbb{F})$ such that $P^{-1}AP = D$. We say that the matrix P **diagonalizes** A.

REMARKS

- Although we have motivated the above Definition by considering the problem of computing powers of matrices, the notion of *diagonalizablity* is fundamental for several other reasons, both theoretical and practical. We will see some of them in due course.
- It is important to pay attention to what field $\mathbb{F}$ we are working over. It is possible for a matrix $A \in M_{n\times n}(\mathbb{R})$ to be diagonalizable over $\mathbb{C}$ but not over $\mathbb{R}$. (See Example 9.4.1.)

Example 7.6.6

Example 7.6.1 shows that $A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$ is diagonalizable over $\mathbb{R}$.

The problem we must now address is: determine if a given matrix $A \in M_{n\times n}(\mathbb{F})$ is diagonalizable over $\mathbb{F}$, and if it is, find the matrix P that diagonalizes it.

We will eventually be able to completely solve this problem. For now, we will give a partial solution.

Proposition 7.6.7 **(n Distinct Eigenvalues $\implies$ Diagonalizable)**

If $A \in M_{n\times n}(\mathbb{F})$ has n *distinct* eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ in $\mathbb{F}$, then A is diagonalizable over $\mathbb{F}$.

More specifically, if we let $(\lambda_1, \overrightarrow{v_1}), (\lambda_2, \overrightarrow{v_2}), \ldots, (\lambda_n, \overrightarrow{v_n})$ be eigenpairs of A over $\mathbb{F}$, and if we let $P = [\overrightarrow{v_1}\, \overrightarrow{v_2} \cdots \overrightarrow{v_n}]$ be the matrix whose columns are eigenvectors corresponding to the distinct eigenvalues, then

(a) P is invertible, and

(b) $P^{-1}AP = D = \text{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)$.

REMARKS

- **WARNING:** The converse of the above Proposition is **false**. That is, if $A \in M_{n\times n}(\mathbb{F})$ is diagonalizable over $\mathbb{F}$, then it is **not** necessarily true that A has n distinct eigenvalues. Consider, for instance, $\mathcal{O}_{2\times 2} = \text{diag}(0, 0)$ which is diagonalizable (it is diagonal!) with two repeated eigenvalues $\lambda_1 = \lambda_2 = 0$.

- We will generalize this result in Chapter 9, and we will address precisely what happens when A does not have n distinct eigenvalues. We will prove the generalized result at that time. (See page 242 for the proof of Proposition 7.6.7.)

- Notice that the columns of P are eigenvectors of A and the diagonal entries of D are eigenvalues of A. Moreover, the order of the eigenvalues down the diagonal of D corresponds to the order the eigenvectors occur as columns in P. That is, the i^{th} diagonal entry in D is the eigenvalue corresponding to the i^{th} column of P. The Proposition remains true if we list the eigenvectors as columns in P in any order we want, provided we adjust the diagonal entries of D accordingly.

Although we do not have a general proof of Proposition 7.6.7 at this time, we can directly verify that it works in specific examples as we illustrate below.

Example 7.6.8

From Example 7.4.1, $A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$ has eigenpairs $\left(3, \begin{bmatrix} 1 \\ 1 \end{bmatrix}\right)$ and $\left(-1, \begin{bmatrix} 1 \\ -1 \end{bmatrix}\right)$ over $\mathbb{R}$. We construct $P = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$ from the eigenvectors, with inverse $P^{-1} = -\frac{1}{2}\begin{bmatrix} -1 & -1 \\ -1 & 1 \end{bmatrix}$.

Now, if we compute

$$\begin{aligned} D = P^{-1}AP &= -\frac{1}{2}\begin{bmatrix} -1 & -1 \\ -1 & 1 \end{bmatrix}\begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \\ &= -\frac{1}{2}\begin{bmatrix} -1 & -1 \\ -1 & 1 \end{bmatrix}\begin{bmatrix} 3 & -1 \\ 3 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 3 & 0 \\ 0 & -1 \end{bmatrix} \\ &= \operatorname{diag}(3, -1), \end{aligned}$$

we find that it is a diagonal matrix whose diagonal entries are the eigenvalues of A, exactly as asserted by Proposition 7.6.7.

Example 7.6.9

From Example 7.4.2, $B = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ has eigenpairs $\left(i, \begin{bmatrix} i \\ 1 \end{bmatrix}\right)$ and $\left(-i, \begin{bmatrix} -i \\ 1 \end{bmatrix}\right)$ over $\mathbb{C}$. We construct $P = \begin{bmatrix} i & -i \\ 1 & 1 \end{bmatrix}$ from the eigenvectors, with inverse $P^{-1} = \frac{1}{2}\begin{bmatrix} -i & 1 \\ i & 1 \end{bmatrix}$.

Then

$$\begin{aligned} D &= P^{-1}BP \\ &= \frac{1}{2}\begin{bmatrix} -i & 1 \\ i & 1 \end{bmatrix}\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} i & -i \\ 1 & 1 \end{bmatrix} \\ &= \frac{1}{2}\begin{bmatrix} -i & 1 \\ i & 1 \end{bmatrix}\begin{bmatrix} -1 & -1 \\ i & -i \end{bmatrix} \\ &= \begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix} \end{aligned}$$

$$= \operatorname{diag}(i, -i).$$

Example 7.6.10

From Example 7.2.7, $H = \begin{bmatrix} 4 & 2 & -6 \\ 1 & -2 & 1 \\ -6 & 2 & 4 \end{bmatrix}$ has eigenalues $\lambda_1 = 10$, $\lambda_2 = 0$ and $\lambda_3 = -4$. We leave it as an exercise to verify that $\left(10, \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}\right)$, $\left(0, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}\right)$ and $\left(-4, \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}\right)$ are eigenpairs over $\mathbb{R}$. We construct $P = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & -1 \\ -1 & 1 & 1 \end{bmatrix}$ from the eigenvectors, with inverse $P^{-1} = \frac{1}{4}\begin{bmatrix} 2 & 0 & -2 \\ 1 & 2 & 1 \\ 1 & -2 & 1 \end{bmatrix}$.

Then

$$\begin{aligned}
D = P^{-1}HP &= \frac{1}{4}\begin{bmatrix} 2 & 0 & -2 \\ 1 & 2 & 1 \\ 1 & -2 & 1 \end{bmatrix} \begin{bmatrix} 4 & 2 & -6 \\ 1 & -2 & 1 \\ -6 & 2 & 4 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & -1 \\ -1 & 1 & 1 \end{bmatrix} \\
&= \frac{1}{4}\begin{bmatrix} 2 & 0 & -2 \\ 1 & 2 & 1 \\ 1 & -2 & 1 \end{bmatrix} \begin{bmatrix} 10 & 0 & -4 \\ 0 & 0 & 4 \\ -10 & 0 & -4 \end{bmatrix} \\
&= \begin{bmatrix} 10 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -4 \end{bmatrix} \\
&= \operatorname{diag}(10, 0, -4).
\end{aligned}$$

EXERCISE

The matrix $A = \begin{bmatrix} 3 & -5 & 0 \\ 2 & -3 & 0 \\ -2 & 3 & -1 \end{bmatrix}$ has the following eigenpairs:

$$\left(-1, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}\right), \quad \left(i, \begin{bmatrix} 2-i \\ 1-i \\ -1 \end{bmatrix}\right), \text{ and } \left(-i, \begin{bmatrix} 2+i \\ 1+i \\ -1 \end{bmatrix}\right).$$

Find an invertible matrix P and a diagonal matrix D such that $A = PDP^{-1}$.

We close this Chapter with an example of a quick computation of powers of a diagonalizable matrix.

Example 7.6.11 Let $H = \begin{bmatrix} 4 & 2 & -6 \\ 1 & -2 & 1 \\ -6 & 2 & 4 \end{bmatrix}$. Find H^{10}.

Solution:

In the previous Example we found matrices P and D so that $H = PDP^{-1}$. By what we have learned from Example 7.6.1 (see also the Exercise below), we find that for all $k \in \mathbb{N}$

$$\begin{aligned} H^k = (PDP^{-1})^k &= PD^kP^{-1} \\ &= \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & -1 \\ -1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 10 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -4 \end{bmatrix}^k \frac{1}{4} \begin{bmatrix} 2 & 0 & -2 \\ 1 & 2 & 1 \\ 1 & -2 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & -1 \\ -1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 10^k & 0 & 0 \\ 0 & 0^k & 0 \\ 0 & 0 & (-4)^k \end{bmatrix} \frac{1}{4} \begin{bmatrix} 2 & 0 & -2 \\ 1 & 2 & 1 \\ 1 & -2 & 1 \end{bmatrix} \\ &= \frac{1}{4} \begin{bmatrix} 2 \cdot 10^k + (-4)^k & -2(-4)^k & -2(10^k) \\ -(-4)^k & 2(-4)^k & -(-4)^k \\ -2(10^k) + (-4)^k & -2(-4)^k & 2(10^k) + (-4)^k \end{bmatrix}. \end{aligned}$$

If we plug in $k = 10$ and simplify, we arrive at

$$H^{10} = 512 \begin{bmatrix} 9766137 & -1024 & -976513 \\ -512 & 1024 & -512 \\ -9765113 & -1024 & 9766137 \end{bmatrix}.$$

Chapter 8

Subspaces and Bases

8.1 Subspaces

Throughout this course we have been dealing with a variety of subsets of $\mathbb{F}^n$—for example, solutions sets of systems of equations, lines and planes in $\mathbb{R}^3$, and ranges of linear transformations. Many of these subsets have a special structure that makes them particularly interesting from the point of view of linear algebra.

The next definition identifies the key features of this structure.

Definition 8.1.1
Subspace

A subset V of $\mathbb{F}^n$ is called a **subspace of** $\mathbb{F}^n$ if the following properties are all satisfied.

1. $\vec{0} \in V$.
2. For all $\vec{x},\ \vec{y} \in V,\ \vec{x} + \vec{y} \in V$ **(closure under addition)**.
3. For all $\vec{x} \in V$ and $c \in \mathbb{F}$, $c\vec{x} \in V$ **(closure under scalar multiplication)**.

Here are some important examples of subspaces.

Proposition 8.1.2

(Examples of Subspaces)

(a) $\{\vec{0}\}$ and $\mathbb{F}^n$ are subspaces of $\mathbb{F}^n$.

(b) If $\{\vec{v_1}, \vec{v_2}, \ldots, \vec{v_k}\}$ is a subset of $\mathbb{F}^n$, then $\mathrm{Span}\{\vec{v_1}, \vec{v_2}, \ldots, \vec{v_k}\}$ is a subspace of $\mathbb{F}^n$.

(c) If $A \in M_{m\times n}(\mathbb{F})$, then the solution set to the homogeneous system $A\vec{x} = \vec{0}$ is a subspace of $\mathbb{F}^n$. (Equivalently, $\mathrm{Null}(A)$ is a subspace of $\mathbb{F}^n$.)

Proof: (a) This is immediate from the definition.

(b) We have $\vec{0} = 0\vec{v_1} + \cdots + 0\vec{v_k}$, so $\vec{0} \in \text{Span}\{\vec{v_1}, \vec{v_2}, \ldots, \vec{v_k}\}$.

Next, we consider $\vec{x}, \vec{y} \in \text{Span}\{\vec{v_1}, \vec{v_2}, \ldots, \vec{v_k}\}$ and $c \in \mathbb{F}$. Then there exist scalars $a_1, a_2, \ldots, a_k \in \mathbb{F}$ such that $\vec{x} = a_1\vec{v_1} + a_2\vec{v_2} + \cdots + a_k\vec{v_k}$, and there exist scalars $b_1, b_2, \ldots, b_k \in \mathbb{F}$ such that $\vec{y} = b_1\vec{v_1} + b_2\vec{v_2} + \ldots + b_k\vec{v_k}$. We have

$$\begin{aligned}\vec{x} + \vec{y} &= (a_1\vec{v_1} + a_2\vec{v_2} + \cdots + a_k\vec{v_k}) + (b_1\vec{v_1} + b_2\vec{v_2} + \cdots + b_k\vec{v_k}) \\ &= (a_1 + b_1)\vec{v_1} + (a_2 + b_2)\vec{v_2} + \cdots + (a_k + b_k)\vec{v_k}.\end{aligned}$$

This shows that $\vec{x} + \vec{y}$ is an element of $\text{Span}\{\vec{v_1}, \vec{v_2}, \ldots, \vec{v_k}\}$. Similarly, since

$$c\vec{x} = c(a_1\vec{v_1} + a_2\vec{v_2} + \cdots + a_k\vec{v_k}) = ca_1\vec{v_1} + ca_2\vec{v_2} + \cdots + ca_k\vec{v_k}$$

we also have that $c\vec{x} \in \text{Span}\{\vec{v_1}, \vec{v_2}, \ldots, \vec{v_k}\}$. This proves that $\text{Span}\{\vec{v_1}, \vec{v_2}, \ldots, \vec{v_k}\}$ is a subspace of $\mathbb{F}^n$.

(c) Let $S = \{\vec{x} \in \mathbb{F}^n \colon A\vec{x} = \vec{0}\}$. We must show that S is a subspace of $\mathbb{F}^n$. As $A\vec{0} = \vec{0}$, we know that $\vec{0} \in S$. Next, let $\vec{x}, \vec{y} \in S$ and $c \in \mathbb{F}$. Then since

$$A(\vec{x} + \vec{y}) = A\vec{x} + A\vec{y} = \vec{0} + \vec{0} = \vec{0},$$

we have that $\vec{x} + \vec{y} \in S$. And since $A(c\vec{x}) = c(A\vec{x}) = c\vec{0} = \vec{0}$, we have that $c\vec{x} \in S$. This proves that S is a subspace of $\mathbb{F}^n$.

$\square$

Many commonly encountered sets in linear algebra can be shown to be subspaces by realizing that they are either of the form $\text{Span}\, S$, for some finite subset S of $\mathbb{F}^n$, or realizing that they are the solution set of a homogeneous linear system of equations, and then appealing to Proposition 8.1.2 above. Here are some key examples of this.

Proposition 8.1.3 **(More Examples of Subspaces)**

(a) If $A \in M_{m\times n}(\mathbb{F})$, then $\text{Col}(A)$ is a subspace of $\mathbb{F}^m$.

(b) If $T \colon \mathbb{F}^n \to \mathbb{F}^m$ is a linear transformation, then the range of T, $\text{Range}(T)$, is a subspace of $\mathbb{F}^m$.

(c) If $T \colon \mathbb{F}^n \to \mathbb{F}^m$ is a linear transformation, then the kernel of T, $\text{Ker}(T)$, is a subspace of $\mathbb{F}^n$.

(d) If $A \in M_{n\times n}(\mathbb{F})$ and if $\lambda \in \mathbb{F}$, then the eigenspace E_λ is a subspace of $\mathbb{F}^n$.

Proof: (a) $\text{Col}(A) = \text{Span}\{\vec{a_1}, \vec{a_2}, \ldots, \vec{a_n}\}$, where $\vec{a_i}$ is the i^{th} column of A, so this is a special case of Proposition 8.1.2 (b).

(b) $\text{Range}(T) = \text{Col}([T]_{\mathcal{E}})$, so this is a special case of Part (a).

(c) $\text{Null}(T)$ is equal to the solution set of the homogeneous linear system $[T]_{\mathcal{E}}\,\vec{x} = \vec{0}$, so this follows from Proposition 8.1.2 (c).

(d) E_λ is the solution set of the homogeneous system $(A - \lambda I)\vec{x} = \vec{0}$, so this is a special case of Proposition 8.1.2 (c).

$\square$

The following proposition can simplify the process needed to check if a subset of $\mathbb{F}^n$ is a subspace.

Proposition 8.1.4

(Subspace Test)

Let V be a subset of $\mathbb{F}^n$. Then V is a subspace of $\mathbb{F}^n$ **if and only if**

(a) V is non-empty, and

(b) for all $\vec{x}, \vec{y} \in V$ and $c \in \mathbb{F}$, $c\,\vec{x} + \vec{y} \in V$.

Proof: ($\Rightarrow$):
Assume V is a subspace of $\mathbb{F}^n$. Then $\vec{0} \in V$, so V is non-empty. Also for all $\vec{x} \in V$ and $c \in \mathbb{F}$, $c\,\vec{x} \in V$, since V is closed under scalar multiplication. But then, for all $\vec{y} \in V$, $c\vec{x} + \vec{y}$ is also in V, since V is closed under addition. So V satisfies **(a)** and **(b)**.

($\Leftarrow$):
Conversely, assume that V is a subset of $\mathbb{F}^n$ that satisfies **(a)** and **(b)**. By **(a)**, there exists an $\vec{x} \in V$. Then, by **(b)** with $c = -1$ and $\vec{y} = \vec{x}$, we find that $\vec{0} = -\vec{x} + \vec{x}$ is in V. Also, with $c = 1$, and any $\vec{x}$ and $\vec{y}$ in V, we find that $\vec{x} + \vec{y} \in V$, so V is closed under addition. Finally, since we have shown that V contains $\vec{0}$, then **(b)** with $\vec{y} = \vec{0}$ shows that V is closed under scalar multiplication. So V is a subspace of $\mathbb{F}^n$. $\square$

EXERCISE

Establish that the examples referred to in Propositions 8.1.2 and 8.1.3 above are subspaces using Proposition 8.1.4 (Subspace Test).

Example 8.1.5

Let $W_1 = \{[x\ y\ z]^T \in \mathbb{R}^3 : x + 2y + 3z = 4\}$ and $W_2 = \{[x\ y\ z]^T \in \mathbb{R}^3 : x^2 - z^2 = 0\}$ be subsets of $\mathbb{R}^3$. Determine whether either of them is a subspace of $\mathbb{R}^3$.

Solution:

W_1 is not a subspace of $\mathbb{R}^3$, since $[0\ 0\ 0]^T \notin W_1$.

Although $[0\ 0\ 0]^T \in W_2$, this subset does not look "linear." Let us check closure under addition.

Notice that $[2\ 0\ 2]^T \in W_2$, as $2^2 - 2^2 = 0$, and $[-2\ 0\ 2]^T \in W_2$, as $(-2^2) - 2^2 = 0$. However, if we add these two vectors together, we get $[2\ 0\ 2]^T + [-2\ 0\ 2]^T = [0\ 0\ 4]^T$, and since $(0^2) - 4^2 = -16 \neq 0$, then $[0\ 0\ 4]^T \notin W_2$.

This means that W_2 is **not** closed under addition and is therefore **not** a subspace of $\mathbb{R}^3$.

REMARK

The preceding example illustrates a useful check: solution sets to non-homogeneous systems (such as W_1) are never subspaces, since they do not contain $\vec{0}$. Likewise, solution sets to nonlinear systems (such as W_2) tend to not be subspaces, since they are usually not closed under addition or scalar multiplication (or both). On the other hand, we have from Proposition 8.1.2 (Examples of Subspaces) that solution sets to *homogeneous linear* systems are always subspaces.

8.2 Linear Dependence and the Notion of a Basis of a Subspace

We will now tackle the issue of how to describe a subspace in an efficient way. We have already seen in Proposition 8.1.2 (Examples of Subspaces) that, given any finite subset S of vectors in $\mathbb{F}^n$, their span $\operatorname{Span} S$ is a subspace of $\mathbb{F}^n$. We will eventually see that, conversely, *every* subspace V of $\mathbb{F}^n$ may be expressed in the form of $V = \operatorname{Span} S$ where S is a finite subset of vectors in V.

An important matter here is that this set S is **not** uniquely determined by V, in the sense that a given subspace V may be expressed as the spanning set of many different subsets S.

Example 8.2.1 We have

$$\mathbb{R}^2 = \operatorname{Span}\left\{\begin{bmatrix}1\\0\end{bmatrix}, \begin{bmatrix}0\\1\end{bmatrix}\right\} = \operatorname{Span}\left\{\begin{bmatrix}1\\0\end{bmatrix}, \begin{bmatrix}0\\1\end{bmatrix}, \begin{bmatrix}1\\2\end{bmatrix}\right\} = \operatorname{Span}\left\{\begin{bmatrix}1\\0\end{bmatrix}, \begin{bmatrix}0\\1\end{bmatrix}, \begin{bmatrix}1\\2\end{bmatrix}, \begin{bmatrix}-1\\-1\end{bmatrix}\right\},$$

as for $\begin{bmatrix}x\\y\end{bmatrix} \in \mathbb{R}^2$, we can express this as

$$\begin{bmatrix}x\\y\end{bmatrix} = x\begin{bmatrix}1\\0\end{bmatrix} + y\begin{bmatrix}0\\1\end{bmatrix} = x\begin{bmatrix}1\\0\end{bmatrix} + y\begin{bmatrix}0\\1\end{bmatrix} + 0\begin{bmatrix}1\\2\end{bmatrix} = x\begin{bmatrix}1\\0\end{bmatrix} + y\begin{bmatrix}0\\1\end{bmatrix} + 0\begin{bmatrix}1\\2\end{bmatrix} + 0\begin{bmatrix}-1\\-1\end{bmatrix}.$$

In the preceding example, it is apparent that the second and third spanning sets contain some redundancies. Our goal in this section is to be able to mathematically quantify this type of redundancy and demonstrate that sets without such redundancies are preferable. Below is a more involved example to illustrate this point.

Example 8.2.2 Consider the following subset of $\mathbb{R}^3$:

$$S = \left\{\begin{bmatrix}2\\2\\2\end{bmatrix}, \begin{bmatrix}2\\-3\\4\end{bmatrix}, \begin{bmatrix}0\\5\\-2\end{bmatrix}, \begin{bmatrix}-2\\13\\-8\end{bmatrix}, \begin{bmatrix}1\\1\\1\end{bmatrix}\right\}.$$

Let $V = \operatorname{Span}(S)$. At this point we cannot say much about V. For instance, it could be the entirety of $\mathbb{R}^3$, or it could be some smaller subspace, such as a line or a plane through the origin. We can at least be sure that V cannot be a line through the origin, since S contains two non-zero, non-parallel vectors. (We will see below that V is in fact a plane.)

To get a better understanding of V, we would like to remove any "redundant" vectors from S and provide a smaller subset, let us call it S_1, such that $V = \operatorname{Span}(S_1)$. How do we know which vectors are redundant and ought to be removed from S? This leads us to the important idea of linear dependence.

Informally, we say that the vectors $\vec{v_1}, \vec{v_2}, \ldots, \vec{v_k}$ are *linearly dependent* if at least one of them, $\vec{v_j}$ say, can be written as a linear combination of some of the other vectors. Thus, the vector $\vec{v_j}$ *depends linearly* on the other vectors. (The formal definition is given in Definition 8.2.3 below.)

For example, the vectors in the set S above are linearly dependent because

$$\begin{bmatrix} -2 \\ 13 \\ -8 \end{bmatrix} = 2 \begin{bmatrix} 2 \\ 2 \\ 2 \end{bmatrix} - 3 \begin{bmatrix} 2 \\ -3 \\ 4 \end{bmatrix}.$$

Thus, we can remove the vector $\begin{bmatrix} -2 & 13 & -8 \end{bmatrix}^T$ from S and obtain a more efficient description of V:

$$\begin{aligned} V &= \operatorname{Span}\left\{ \begin{bmatrix} 2 \\ 2 \\ 2 \end{bmatrix}, \begin{bmatrix} 2 \\ -3 \\ 4 \end{bmatrix}, \begin{bmatrix} 0 \\ 5 \\ -2 \end{bmatrix}, \begin{bmatrix} -2 \\ 13 \\ -8 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right\} \\ &= \operatorname{Span}\left\{ \begin{bmatrix} 2 \\ 2 \\ 2 \end{bmatrix}, \begin{bmatrix} 2 \\ -3 \\ 4 \end{bmatrix}, \begin{bmatrix} 0 \\ 5 \\ -2 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right\}. \end{aligned}$$

However, note that the choice of singling out the vector $\begin{bmatrix} -2 & 13 & -8 \end{bmatrix}^T$ is somewhat arbitrary. For example, we could also write

$$\begin{bmatrix} 2 \\ 2 \\ 2 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} -2 \\ 13 \\ -8 \end{bmatrix} + \frac{3}{2} \begin{bmatrix} 2 \\ -3 \\ 4 \end{bmatrix}$$

or

$$\begin{bmatrix} 2 \\ -3 \\ 4 \end{bmatrix} = \frac{2}{3} \begin{bmatrix} 2 \\ 2 \\ 2 \end{bmatrix} - \frac{1}{3} \begin{bmatrix} -2 \\ 13 \\ -8 \end{bmatrix}$$

and then remove one of these vectors instead.

It is therefore more reasonable and balanced if we put everything on one side of our dependence equation and re-write it as

$$\begin{bmatrix} -2 \\ 13 \\ -8 \end{bmatrix} - 2 \begin{bmatrix} 2 \\ 2 \\ 2 \end{bmatrix} + 3 \begin{bmatrix} 2 \\ -3 \\ 4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

The point is that the redundancy in the set S is identified by the feature that we can take some linear combination of some of the vectors in S to produce the zero vector.

We can always form the zero vector by taking a **trivial linear combination** (that is, one where all the scalar coefficients are 0). For example,

$$0\begin{bmatrix}-2\\13\\-8\end{bmatrix}+0\begin{bmatrix}2\\2\\2\end{bmatrix}+0\begin{bmatrix}2\\-3\\4\end{bmatrix}=\begin{bmatrix}0\\0\\0\end{bmatrix}.$$

It is worth noting the scenarios where this is the **only** linear combination that can produce the zero vector, which forms the premise of our definitions of linear dependence and independence below.

Definition 8.2.3
Linear Dependence

We say that the vectors $\vec{v_1}, \vec{v_2}, \ldots, \vec{v_k} \in \mathbb{F}^n$ are **linearly dependent** if there exist scalars $c_1, c_2, \ldots, c_k \in \mathbb{F}$, not all zero, such that $c_1\vec{v_1} + c_2\vec{v_2} + \cdots + c_k\vec{v_k} = \vec{0}$.

If $U = \{\vec{v_1}, \vec{v_2}, \ldots, \vec{v_k}\}$, then we say that the set U is a **linearly dependent set** (or simply that U is **linearly dependent**) to mean that the vectors $\vec{v_1}, \vec{v_2}, \ldots, \vec{v_k}$ are linearly dependent.

If a list or set of vectors is not linearly dependent, we will say that it is linearly independent. Here is the formal definition.

Definition 8.2.4
Linear Independence, Trivial Solution

We say that the vectors $\vec{v_1}, \vec{v_2}, \ldots, \vec{v_k} \in \mathbb{F}^n$ are **linearly independent** if there do not exist scalars $c_1, c_2, \ldots, c_k \in \mathbb{F}$, not all zero, such that $c_1\vec{v_1} + c_2\vec{v_2} + \ldots + c_k\vec{v_k} = \vec{0}$.

Equivalently we say that $\vec{v_1}, \vec{v_2}, \ldots, \vec{v_k} \in \mathbb{F}^n$ are **linearly independent** if the only solution to the equation

$$c_1\vec{v_1} + c_2\vec{v_2} + \ldots + c_k\vec{v_k} = \vec{0}$$

is the **trivial solution** $c_1 = c_2 = \cdots = c_k = 0$.

If $U = \{\vec{v_1}, \vec{v_2}, \ldots, \vec{v_k}\}$, then we say that the set U is a **linearly independent set** (or simply that U is **linearly independent**) to mean that the vectors $\vec{v_1}, \vec{v_2}, \ldots, \vec{v_k}$ are linearly independent.

REMARK

As a matter of convention, the empty set $\varnothing$ is considered to be linearly independent. We consider the condition for linear independence in the above definition to be vacuously satisfied by $\varnothing$.

The notion of linear dependence and independence is one of the most important ideas in linear algebra. We will examine it more deeply in the following sections. For now, let's return to the previous example.

Example 8.2.5 We have as before the set

$$V = \operatorname{Span}(S) = \operatorname{Span}\left\{ \begin{bmatrix} 2 \\ 2 \\ 2 \end{bmatrix}, \begin{bmatrix} 2 \\ -3 \\ 4 \end{bmatrix}, \begin{bmatrix} 0 \\ 5 \\ -2 \end{bmatrix}, \begin{bmatrix} -2 \\ 13 \\ -8 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right\}$$

which is a subspace of $\mathbb{R}^3$. We have shown that S is linearly dependent, and specifically that $[-2\ 13\ \ -8]^T$ can be written as a linear combination of $[2\ 2\ 2]^T$ and $[2\ \ -3\ 4]^T$. We choose to remove the vector $[-2\ 13\ \ -8]^T$ from S to produce the set S_1. Then

$$V = \operatorname{Span}(S_1) = \operatorname{Span}\left\{ \begin{bmatrix} 2 \\ 2 \\ 2 \end{bmatrix}, \begin{bmatrix} 2 \\ -3 \\ 4 \end{bmatrix}, \begin{bmatrix} 0 \\ 5 \\ -2 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right\}.$$

However S_1 is linearly dependent. For example,

$$\begin{bmatrix} 2 \\ 2 \\ 2 \end{bmatrix} - 2 \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

We choose to remove the vector $[2\ 2\ 2]^T$ from S_1 to produce the set S_2. Then

$$V = \operatorname{Span}(S_2) = \operatorname{Span}\left\{ \begin{bmatrix} 2 \\ -3 \\ 4 \end{bmatrix}, \begin{bmatrix} 0 \\ 5 \\ -2 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right\}.$$

The set S_2 is still linearly dependent! For example,

$$2 \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - \begin{bmatrix} 2 \\ -3 \\ 4 \end{bmatrix} - \begin{bmatrix} 0 \\ 5 \\ -2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

We choose to remove the vector $[0\ 5\ \ -2]^T$ from S_2 to produce the set S_3. Then

$$V = \operatorname{Span}(S_3) = \operatorname{Span}\left\{ \begin{bmatrix} 2 \\ -3 \\ 4 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right\}.$$

The set S_3, however, is linearly independent. Indeed, if we have a linear combination

$$c_1 \begin{bmatrix} 2 \\ -3 \\ 4 \end{bmatrix} + c_2 \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix},$$

then $c_2 = -2c_1$ and $c_2 = 3c_1$, so $-2c_1 = 3c_1$. Therefore, $c_1 = 0$ and consequently $c_2 = 0$. That is, we cannot form the zero vector as a non-trivial linear combination of vectors in S_3.

The set S_3 is an optimal way of producing the subspace V. This optimal set incidentally shows that V is equal to the span of two non-parallel vectors in $\mathbb{R}^3$, and so it is a plane through the origin.

The previous example illustrates an important phenomenon. Given a subspace V of $\mathbb{F}^n$, say one of the form $V = \mathrm{Span}\,(S)$ for some finite set S (we will later see that all subspaces are of this form), it will be desirable to remove any redundant vectors from S, reducing it to a smaller *linearly independent* subset $\mathcal{B}$ such that $V = \mathrm{Span}(\mathcal{B})$. Such a set $\mathcal{B}$ is, in some sense, a basic set of building blocks for the construction of V.

Definition 8.2.6
Basis

Let V be a subspace of $\mathbb{F}^n$ and let $\mathcal{B} = \{\vec{v_1}, \vec{v_2}, \ldots, \vec{v_k}\}$ be a finite set of vectors contained in V. We say that $\mathcal{B}$ is a **basis for** V if

1. $\mathcal{B}$ is linearly independent, and

2. $V = \mathrm{Span}(\mathcal{B})$.

REMARK

We shall adopt the convention that the empty set $\varnothing$ is a basis for the zero subspace $V = \left\{\vec{0}\right\}$. Recall that we had earlier adopted the convention of considering the empty set to be linearly independent. (See the remark following Definition 8.2.4.) The statement $\mathrm{Span}\,(\varnothing) = \left\{\vec{0}\right\}$ can be made plausible if we agree that a "linear combination of no vectors" gives the zero vector.

The notion of a basis is central to much of what follows in the course. We will spend the next two sections carefully examining both conditions **(1)** (linear independence) and **(2)** (spanning) in the above definition.

Example 8.2.7

Our work in the previous example shows that $\mathcal{B} = \left\{ \begin{bmatrix} 2 \\ -3 \\ 4 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right\}$ is a basis for

$$V = \mathrm{Span}\left\{ \begin{bmatrix} 2 \\ 2 \\ 2 \end{bmatrix}, \begin{bmatrix} 2 \\ -3 \\ 4 \end{bmatrix}, \begin{bmatrix} 0 \\ 5 \\ -2 \end{bmatrix}, \begin{bmatrix} -2 \\ 13 \\ -8 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right\}.$$

8.3 Detecting Linear Dependence and Independence

We begin by recalling the definitions of linear dependence and independence given in Definitions 8.2.3 and 8.2.4. A set of vectors $\{\vec{v_1}, \vec{v_2}, \ldots, \vec{v_k}\}$ is said to be **linearly dependent** if there exist scalars $c_1, c_2, \ldots, c_k$, not all zero, such that $c_1\vec{v_1} + c_2\vec{v_2} + \ldots + c_k\vec{v_k} = \vec{0}$; if not, the set is said to be **linearly independent**.

The next proposition offers reformulations of these definitions that can be easier to use in practice.

Proposition 8.3.1 **(Linear Dependence Check)**

(a) The vectors $\vec{v_1}, \vec{v_2}, \ldots, \vec{v_k}$ are linearly dependent if and only if one of the vectors can be written as a linear combination of some of the other vectors.

(b) The vectors $\vec{v_1}, \vec{v_2}, \ldots, \vec{v_k}$ are linearly independent if and only if

$$c_1\vec{v_1} + \cdots + c_k\vec{v_k} = \vec{0} \quad (c_i \in \mathbb{F}) \quad \text{implies} \quad c_1 = \cdots = c_k = 0.$$

Proof: (a) Suppose that $\vec{v_1}, \vec{v_2}, \ldots, \vec{v_k}$ are linearly dependent. Then there exist scalars $c_1, \ldots, c_k$, not all zero, such that $c_1\vec{v_1} + \cdots + c_k\vec{v_k} = \vec{0}$. Suppose that c_j is nonzero. Then by moving the $c_j\vec{v_j}$ term to the right side of the previous equation, we get

$$-c_j\vec{v_j} = \sum_{i \neq j} c_i \vec{v}_i.$$

Dividing through by $-c_j$ (which is permissible, since $c_j \neq 0$), we get

$$\vec{v_j} = \sum_{i \neq j} \frac{c_i}{-c_j} \vec{v}_i.$$

Thus, we have expressed $\vec{v_j}$ as a linear combination of the other vectors.

Conversely, if $\vec{v_j}$ is a linear combination of the other vectors, say

$$\vec{v_j} = \sum_{i \neq j} a_i \vec{v}_i,$$

then

$$(-1)\vec{v_j} + \sum_{i \neq j} a_i \vec{v}_i = \vec{0}$$

is an expression of $\vec{0}$ as a nontrivial linear combination of the vectors $\vec{v_1}, \ldots, \vec{v_k}$ (non-trivial because the coefficient of $\vec{v_j}$ is nonzero). This proves that $\vec{v_1}, \ldots, \vec{v_k}$ are linearly dependent.

(b) This result is a rephrasing of the equivalent condition for linear independence in Definition 8.2.4. Indeed, the contrapositive of the given statement is: A list of vectors $\vec{v_1}, \ldots, \vec{v_k}$ is linearly *dependent* if and only if there exists scalars $c_1, \ldots, c_k$ not all zero such that $c_1\vec{v_1} + \cdots + c_k\vec{v_k} = \vec{0}$. This is the definition of linear dependence. So the statement given in **(b)** must be true as well.

$\square$

Now we turn to the problem of determining whether a given set is or is not linearly dependent. When the set contains at most two vectors, this is particularly straightforward.

Proposition 8.3.2 Let $S \subseteq \mathbb{F}^n$.

(a) If $\vec{0} \in S$, then S is linearly dependent.

(b) If $S = \{\vec{x}\}$ contains only one vector, then S is linearly dependent if and only if $\vec{x} = \vec{0}$.

(c) If $S = \{\vec{x}, \vec{y}\}$ contains only two vectors, then S is linearly dependent if and only if one of the vectors is a scalar multiple of the other.

Proof: (a) Since 1 is a non-zero scalar, and $1(\vec{0}) = \vec{0}$, we can thus take a non-trivial linear combination of some vectors (in this case, only one vector) in S to produce the zero vector.

(b) We know from (a) that if $\vec{x} = \vec{0}$, then S is linearly dependent.

Conversely, we know from the properties of the zero vector that if $c\vec{x} = \vec{0}$, then either $\vec{x} = \vec{0}$ or $c = 0$.

So if $\vec{x} \neq \vec{0}$ and $c\,\vec{x} = \vec{0}$, then c must be zero. That is, only the trivial combination of $\vec{x} \neq \vec{0}$ will produce the zero vector. Thus, S is linearly independent.

(c) If one of the vectors is a multiple of the other, we may assume without loss of generality that $\vec{y} = m\vec{x}$ for some $m \in \mathbb{F}$. Then

$$\vec{y} - m\,\vec{x} = \vec{0}.$$

Since the coefficient of $\vec{y}$ is $1 \neq 0$, we have demonstrated linear dependence.

Conversely, if $S = \{\vec{x}, \vec{y}\}$ is linearly dependent, then there exist $a, b \in \mathbb{F}$, not both zero, such that $a\,\vec{y} + b\,\vec{x} = \vec{0}$. We may assume without loss of generality that $a \neq 0$. We then have that

$$\vec{y} + \frac{b}{a}\,\vec{x} = \vec{0}, \qquad \text{which gives} \qquad \vec{y} = -\frac{b}{a}\,\vec{x}.$$

Thus, one of the vectors, $\vec{y}$, is a multiple of the other, $\vec{x}$.

$\square$

Example 8.3.3 Is $S = \left\{ \begin{bmatrix} 2 \\ -4 \end{bmatrix}, \begin{bmatrix} 6 \\ -12 \end{bmatrix} \right\}$ linearly dependent?

Solution:

Yes, as $\begin{bmatrix} 6 \\ -12 \end{bmatrix} = 3 \begin{bmatrix} 2 \\ -4 \end{bmatrix}$.

Example 8.3.4 Is $S = \left\{ \begin{bmatrix} 2 \\ -4 \end{bmatrix}, \begin{bmatrix} 6 \\ -14 \end{bmatrix} \right\}$ linearly dependent?

Solution:

No. The second vector is not a multiple of the first (and the first is not a multiple of the second). Looking at the first components, 6 is 3 times 2; however, looking at the second components, -14 is not 3 times -4.

Example 8.3.5 Is $S = \{\vec{z_1}, \vec{z_2}\}$, with $\vec{z_1} = \begin{bmatrix} 2+3i \\ 4+i \end{bmatrix}$ and $\vec{z_2} = \begin{bmatrix} 12+5i \\ 14-5i \end{bmatrix}$, linearly dependent?

Solution:

Because we are working over $\mathbb{C}$, the answer to this question is not so obvious.

We will use Proposition 8.3.2 (b). Consider the equation $c_1\,\vec{z_1} + c_2\,\vec{z_2} = \vec{0}$. The Proposition states that $\{\vec{z_1}, \vec{z_2}\}$ is linearly independent precisely when this equation has only the trivial solution $c_1 = c_2 = 0$.

The corresponding coefficient matrix of this equation is

$$\begin{bmatrix} 2+3i & 12+5i \\ 4+i & 14-5i \end{bmatrix}$$

which row reduces to

$$\begin{bmatrix} 1 & 3-2i \\ 0 & 0 \end{bmatrix}.$$

We deduce from this reduced matrix that there are (infinitely many) non-trivial solutions to the system of equations. One solution is $c_1 = -3 + 2i$ and $c_2 = 1$.

Thus, the set S is linearly dependent.

How do we deal with a set that contains more than two vectors? The ideas in the previous examples can be expanded as follows. Let $S = \{\vec{v_1}, \vec{v_2}, \ldots, \vec{v_k}\}$ be a set of k vectors in $\mathbb{F}^n$. Consider the following questions about S:

1. Is $\vec{0} \in S$?

2. Is one vector in S a scalar multiple of another vector in S?

3. Can we easily observe that we can write one of the vectors in S as a linear combination of some of the other vectors in S?

If the answer to any of these questions is “yes”, then the set S is linearly dependent. In particular, **(2)** and **(3)** above essentially ask “is it obvious that this set is linearly dependent?”

If the answer to all of them is “no”, we cannot immediately conclude whether the set is linearly independent or dependent. Therefore, we must examine the expression

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

This is a homogeneous linear system of n equations (obtained by equating the entries of the vectors on both sides of the above expression in k unknowns (the coefficients c_i). Since it is a homogeneous system, we have that one solution is the trivial solution, and we wish to know whether or not there are any other solutions. We know that S is linearly dependent if and only if there are non-trivial solutions from Proposition 8.3.1 (Linear Dependence Check).

This is all succinctly captured by the rank of the coefficient matrix of the system associated with the expression above.

Proposition 8.3.6 **(Pivots and Linear Independence)**

Let $S = \{\overrightarrow{v_1}, \overrightarrow{v_2}, \ldots, \overrightarrow{v_k}\}$ be a set of k vectors in $\mathbb{F}^n$. Let $A = \begin{bmatrix} \overrightarrow{v_1} & \overrightarrow{v_2} & \cdots & \overrightarrow{v_k} \end{bmatrix}$ be the $n \times k$ matrix whose columns are the vectors in S.

Suppose that $\operatorname{rank}(A) = r$ and A has pivots in columns $q_1, q_2, \ldots, q_r$.

Let $U = \{\overrightarrow{v_{q_1}}, \overrightarrow{v_{q_2}}, \ldots, \overrightarrow{v_{q_r}}\}$, the set of columns of A that correspond to the pivot columns labelled above. Then

(a) S is linearly independent if and only if $r = k$.

(b) U is linearly independent.

(c) If $\overrightarrow{v}$ is in S but not in U then the set $\{\overrightarrow{v_{q_1}}, ..., \overrightarrow{v_{q_r}}, \overrightarrow{v}\}$ is linearly dependent.

(d) $\operatorname{Span}(U) = \operatorname{Span}(S)$.

REMARK

This proposition makes the identification of linear dependence/independence relatively straightforward. Additionally, it identifies "redundant" vectors in a given set that may be removed to obtain a linearly independent subset whose span is the same as the span of the original set. This gives a framework for the ad-hoc process carried out in Proposition 8.1.2.

Proof: (a) The matrix A is the coefficient matrix for the homogeneous system of linear equations

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

There will be no parameters in the solution set if and only if there is a pivot in each of the columns of A, that is, if and only if $r = k$. So S is linearly independent if and only if $r = k$.

(b) We must examine the linear dependence of the vectors $\overrightarrow{v_{q_1}}, \overrightarrow{v_{q_2}}, \ldots, \overrightarrow{v_{q_r}}$ corresponding to the pivot columns in A. Thus, we must consider the homogeneous system of linear equations

$$d_1 \overrightarrow{v_{q_1}} + d_2 \overrightarrow{v_{q_2}} + \cdots + d_r \overrightarrow{v_{q_r}} = \overrightarrow{0}.$$

This system is a system of r equations in r unknowns. Its coefficient matrix has rank r because it consists of the pivot columns in A. Thus, the only solution to this system is the trivial solution. Consequently, U must be linearly independent.

(c) Suppose that $r < k$, so that there is at least one non-pivot column in A — the jth column, say. Add the vector $\overrightarrow{v_j}$ to U and check this new set for linear dependence. This amounts to examining the system

$$d_1 \overrightarrow{v_{q_1}} + d_2 \overrightarrow{v_{q_2}} + \cdots + d_r \overrightarrow{v_{q_r}} + \alpha \overrightarrow{v_j} = \overrightarrow{0}.$$

By construction, the coefficient matrix of this system has $r + 1$ columns and rank r. Thus, there must be $r + 1 - r = 1$ free parameter in its solution set, and therefore, there are nontrivial solutions. It follows that $\{\overrightarrow{v_{q_1}}, \overrightarrow{v_{q_2}}, \ldots, \overrightarrow{v_{q_r}}, \overrightarrow{v_j}\}$ is linearly dependent.

(d) Note that $U \subseteq S$, so $\operatorname{Span}(U) \subseteq \operatorname{Span}(S)$. We will prove that $\operatorname{Span}(S) \subseteq \operatorname{Span}(U)$.

If $U = S$, there is nothing to prove. So assume that $S \neq U$, and let $\vec{v_j}$ be a vector in S, but not in U. Our proof in part **(3)** shows that the system

$$d_1 \overrightarrow{v_{q_1}} + d_2 \overrightarrow{v_{q_2}} + \cdots + d_r \overrightarrow{v_{q_r}} + \alpha \vec{v_j} = \vec{0}$$

has a nontrivial solution. Such a solution must have $\alpha \neq 0$, because of the following argument. If $\alpha = 0$, then

$$d_1 \overrightarrow{v_{q_1}} + d_2 \overrightarrow{v_{q_2}} + \cdots + d_r \overrightarrow{v_{q_r}} = \vec{0},$$

and consequently $d_1 = \ldots = d_r = 0$ by part (b) above and by Part (b) of Proposition 8.3.1 (Linear Dependence Check). So, if $\alpha = 0$, we get the trivial solution, which implies that this set of vectors is linearly independent and this contradicts (c).

Therefore, since $\alpha \neq 0$, we can write $\vec{v_j}$ as a linear combination of the vectors in U, namely

$$\vec{v_j} = \frac{d_1}{-\alpha} \overrightarrow{v_{q_1}} + \frac{d_2}{-\alpha} \overrightarrow{v_{q_2}} + \cdots + \frac{d_r}{-\alpha} \overrightarrow{v_{q_r}}.$$

This shows that $\vec{v_j} \in \operatorname{Span}(U)$. Since $\vec{v_j}$ was an arbitrary vector in S but not in U, it follows that $\operatorname{Span}(S) \subseteq \operatorname{Span}(U)$, as desired.

$\square$

Here is a very useful application of the previous Proposition.

Corollary 8.3.7 **(Bound on Number of Linearly Independent Vectors)**

Let $S = \{\vec{v_1}, \vec{v_2}, \ldots, \vec{v_k}\}$ be a set of k vectors in $\mathbb{F}^n$. If $n < k$, then S is linearly dependent.

Proof: As in Proposition 8.3.6 (Pivots and Linear Independence), let A be the matrix whose columns are the vectors in S, and let $r = \operatorname{rank}(A)$. Then $r \leq n$, since A has n rows. Hence if $n < k$, it follows that $r \leq n < k$. Therefore, S is linearly dependent, by Proposition 8.3.6 (a). $\square$

We now turn to some computational applications of Proposition 8.3.6 (Pivots and Linear Independence).

Example 8.3.8 Consider the following vectors in $\mathbb{R}^6$:

$$\vec{v_1} = \begin{bmatrix} 1 & -2 & 3 & -4 & 5 & -6 \end{bmatrix}^T, \qquad \vec{v_2} = \begin{bmatrix} 3 & 4 & 5 & 6 & 7 & 8 \end{bmatrix}^T,$$
$$\vec{v_3} = \begin{bmatrix} -5 & -10 & -7 & -16 & -9 & -22 \end{bmatrix}^T, \qquad \vec{v_4} = \begin{bmatrix} -6 & -4 & -2 & 0 & 2 & 4 \end{bmatrix}^T,$$
$$\vec{v_5} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}^T, \qquad \vec{v_6} = \begin{bmatrix} -3 & 3 & -5 & 5 & -1 & 1 \end{bmatrix}^T.$$

Let $S = \{\vec{v_1}, \vec{v_2}, \vec{v_3}, \vec{v_4}, \vec{v_5}, \vec{v_6}\}$ and $V = \operatorname{Span} S$.

(a) Show that S is linearly dependent.

(b) Find a subset of S that is a basis for V.

Solution: (a) Form matrix A, which has the vectors in S as its columns.

$$A = \begin{bmatrix} 1 & 3 & -5 & -6 & 1 & -3 \\ -2 & 4 & -10 & -4 & 1 & 3 \\ 3 & 5 & -7 & -2 & 1 & -5 \\ -4 & 6 & -16 & 0 & 1 & 5 \\ 5 & 7 & -9 & 2 & 1 & -1 \\ -6 & 8 & -22 & 4 & 1 & 1 \end{bmatrix}$$

An REF of A is

$$\begin{bmatrix} 1 & 3 & -5 & -6 & 1 & -3 \\ 0 & 1 & -2 & \frac{-8}{5} & \frac{3}{10} & \frac{-3}{10} \\ 0 & 0 & 0 & 1 & \frac{-1}{12} & \frac{7}{24} \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

We find that A has rank 4 with pivots in columns $1, 2, 4$, and 6. Since $4 < 6$, we conclude that S is linearly dependent (by Part (a) of Proposition 8.3.6 (Pivots and Linear Independence)).

(b) Parts (b) and (d) of Proposition 8.3.6 (Pivots and Linear Independence) tell us that the vectors $\vec{v_1}, \vec{v_2}, \vec{v_4}$ and $\vec{v_6}$ are linearly independent, and that

$$\operatorname{Span}\{\vec{v_1}, \vec{v_2}, \vec{v_4}, \vec{v_6}\} = \operatorname{Span} S = V.$$

We conclude that $\{\vec{v_1}, \vec{v_2}, \vec{v_4}, \vec{v_6}\}$ is a basis for V.

Example 8.3.9 Consider the following vectors in $\mathbb{C}^5$:

$$\begin{aligned} \vec{v}_1 &= [1+i,\ -2+i,\ 3,\ -4-i,\ 5i]^T, \\ \vec{v}_2 &= [-1+5i,\ -7-4i,\ 6+9i,\ -5-14i,\ -15+10i]^T, \\ \vec{v}_3 &= [3i,\ 4+2i,\ 5+2i,\ 6-4i,\ 7-2i]^T, \\ \vec{v}_4 &= [-1+2i,\ -3-6i,\ 11+9i,\ 1-18i,\ -8+10i]^T, \\ \vec{v}_5 &= [-6i,\ -4i,\ -2i,\ 0,\ 2i]^T. \end{aligned}$$

Let $S = \{\vec{v_1}, \vec{v_2}, \vec{v_3}, \vec{v_4}, \vec{v_5}\}$ and $V = \operatorname{Span} S$.

(a) Show that S is linearly dependent.

(b) Find a subset of S that is a basis for V.

Solution: (a) Form matrix A, which has the vectors in S as its columns.

$$A = \begin{bmatrix} 1+i & -1+5i & 3i & -1+2i & -6i \\ -2+i & -7-4i & 4+2i & -3-6i & -4i \\ 3 & 6+9i & 5+2i & 11+9i & -2i \\ -4-i & -5-14i & 6-4i & 1-18i & 0 \\ 5i & -15+10i & 7-2i & -8+10i & 2i \end{bmatrix}.$$

The RREF of A is

$$\begin{bmatrix} 1 & 2+3i & 0 & 0 & -2-3i \\ 0 & 0 & 1 & 0 & -1 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

We find that A has rank 3 with pivots in columns $1, 3$ and 4. Since $3 < 5$, we conclude that S is linearly dependent (by Part (a) of Proposition 8.3.6 (Pivots and Linear Independence)).

(b) Parts (b) and (d) of Proposition 8.3.6 (Pivots and Linear Independence) tell us that the vectors $\vec{v_1}, \vec{v_3}$, and $\vec{v_4}$ are linearly independent, and that

$$\text{Span}\{\vec{v_1}, \vec{v_2}, \vec{v_4}\} = \text{Span}\, S = V.$$

We conclude that $\{\vec{v_1}, \vec{v_2}, \vec{v_4}\}$ is a basis for V.

8.4 Spanning Sets

The problem we wish to tackle in this section is the following: given a subspace V of $\mathbb{F}^n$, find a finite set $S = \{\vec{v_1}, \ldots, \vec{v_k}\}$ of vectors in V such that $V = \text{Span}\{\vec{v_1}, \ldots, \vec{v_k}\}$. We begin by showing that this is always possible.

Theorem 8.4.1 **(Every Subspace Has a Spanning Set)**

Let V be a subspace of $\mathbb{F}^n$. Then there exist vectors $\vec{v_1}, \ldots, \vec{v_k} \in V$ such that

$$V = \text{Span}\{\vec{v_1}, \ldots, \vec{v_k}\}.$$

Proof: If $V = \{\vec{0}\}$, then $V = \text{Span}\{\vec{0}\}$. So we may assume that $V \neq \{\vec{0}\}$. Therefore, there must be some nonzero vector $\vec{v}_1$ in V.

If $V = \text{Span}\{\vec{v_1}\}$, we are done. Otherwise, there exists a $\vec{v_2} \in V$ that is not in $\text{Span}\{\vec{v_1}\}$. By Part (a) of Proposition 8.3.1 (Linear Dependence Check), the set $\{\vec{v_1}, \vec{v_2}\}$ must be linearly independent.

If $V = \text{Span}\{\vec{v_1}, \vec{v_2}\}$, we are done. Otherwise, by continuously repeating this process, we generate a set $\{\vec{v_1}, \ldots, \vec{v_k}\}$ of k linearly independent vectors in V. By Corollary 8.3.7 (Bound on Number of Linearly Independent Vectors), we must have $k \leq n$. Thus, the process must terminate after at most n steps, at which point we have $V = \text{Span}\{\vec{v_1}, \ldots, \vec{v_k}\}$. $\square$

REMARK

The above proof actually shows more than is stated in the Theorem. It shows that if V is a nonzero subspace of $\mathbb{F}^n$, then we can find a *linearly independent* spanning set (that is, *a basis*!) for V. Moreover, we can find a basis with at most n elements. We will return to these points in Theorem 8.5.1 (Every Subspace Has a Basis) and Proposition 8.7.5 (Bound on Dimension of Subspace).

The proof of the previous Theorem can be refined into an algorithm that produces a spanning set for a given subspace V of $\mathbb{F}^n$. We will not go into the full details of that here; instead, we will only consider one of the key issues that must be addressed in order to make this process practical. Namely, given a finite subset S of V, how can we determine if $\operatorname{Span}(S) = V$?

Proposition 8.4.2 **(Span of Subset)**

Let V be a subspace of $\mathbb{F}^n$ and let $S = \{\vec{v_1}, \ldots, \vec{v_k}\} \subseteq V$. Then $\operatorname{Span}(S) \subseteq V$.

Proof: This result follows immediately from the fact that V is closed under addition and scalar multiplication. □

Thus, given a finite subset S of V, to show that $\operatorname{Span} S = V$, we need only show that $V \subseteq \operatorname{Span}(S)$, since the other containment is always true by the previous Proposition. In order to do this, we need to show that given $\vec{v} \in V$, there exist scalars $c_1, \ldots, c_k \in \mathbb{F}$ such that

$$\vec{v} = c_1\,\vec{v_1} + \cdots + c_k\,\vec{v_k}.$$

If we let $A = \begin{bmatrix} \vec{v_1} & \cdots & \vec{v_k} \end{bmatrix}$ be the coefficient matrix of this system of equations, and if we let $B = \begin{bmatrix} \vec{v_1} & \cdots & \vec{v_k} & | \, \vec{v} \end{bmatrix}$ be the augmented matrix of this system of equations, then we need to show that $\operatorname{rank}(A) = \operatorname{rank}(B)$ for every $\vec{v} \in V$. This will depend on both S and V.

Example 8.4.3 Consider the plane P in $\mathbb{R}^3$ given by the scalar equation $2x - 6y + 4z = 0$. Determine a basis for P.

Solution: If we let $z = t$ and $y = s$, where $s, t \in \mathbb{R}$, then $x = 3s - 2t$, and we can express the plane as

$$P = \left\{ \begin{bmatrix} 3s-2t \\ s \\ t \end{bmatrix} : s, t \in \mathbb{R} \right\} = \left\{ \begin{bmatrix} 3 \\ 1 \\ 0 \end{bmatrix} s + \begin{bmatrix} -2 \\ 0 \\ 1 \end{bmatrix} t : s, t \in \mathbb{R} \right\}.$$

Thus,

$$P = \operatorname{Span}\left\{ \begin{bmatrix} 3 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -2 \\ 0 \\ 1 \end{bmatrix} \right\}.$$

Since the vectors $\begin{bmatrix} 3 \\ 1 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} -2 \\ 0 \\ 1 \end{bmatrix}$ are not multiples of each other, the set

$$S = \left\{ \begin{bmatrix} 3 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -2 \\ 0 \\ 1 \end{bmatrix} \right\}$$

is linearly independent, hence is a basis for P.

Example 8.4.4

Continue with P as above, and consider $S_1 = \left\{ \begin{bmatrix} 2 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} -4 \\ 0 \\ 2 \end{bmatrix} \right\}$. Is $\operatorname{Span}(S_1) = P$?

(Note that $S_1 \subseteq P$.)

Solution: Let $\vec{v} \in P$. Then we can write $\vec{v} = \begin{bmatrix} 3s_1 - 2t_1 \\ s_1 \\ t_1 \end{bmatrix}$ for some $s_1, t_1 \in \mathbb{R}$. Consider the system

$$\vec{v} = \begin{bmatrix} 3s_1 - 2t_1 \\ s_1 \\ t_1 \end{bmatrix} = a \begin{bmatrix} 2 \\ 0 \\ -1 \end{bmatrix} + b \begin{bmatrix} -4 \\ 0 \\ 2 \end{bmatrix}, \quad \text{where } a, b \in \mathbb{R}.$$

The augmented matrix for this system is

$$\left[\begin{array}{cc|c} 2 & -4 & 3s_1 - 2t_1 \\ 0 & 0 & s_1 \\ -1 & 2 & t_1 \end{array}\right],$$

which row reduces to

$$\left[\begin{array}{cc|c} 2 & -4 & 3s_1 - 2t_1 \\ 0 & 0 & s_1 \\ 0 & 0 & 0 \end{array}\right].$$

The last column of the augmented matrix is a pivot column when $s_1 \neq 0$, and thus, we cannot always solve this system. It is also quite clear from the form of the two vectors in S_1, which both have a second component of zero, that we will be unable to build all the vectors in P from them, as some of the vectors in P have a nonzero second component.

Thus $\operatorname{Span} S_1 \neq P$.

Example 8.4.5

Again use P from above. Consider $S_2 = \left\{ \begin{bmatrix} 2 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} -4 \\ 0 \\ 2 \end{bmatrix}, \begin{bmatrix} 3 \\ 1 \\ 0 \end{bmatrix} \right\}$. Is $\operatorname{Span}(S_2) = P$?

Solution: Let $\vec{v} = \begin{bmatrix} 3s_1 - 2t_1 \\ s_1 \\ t_1 \end{bmatrix} \in P$ and consider the system

$$\vec{v} = \begin{bmatrix} 3s_1 - 2t_1 \\ s_1 \\ t_1 \end{bmatrix} = a \begin{bmatrix} 2 \\ 0 \\ -1 \end{bmatrix} + b \begin{bmatrix} -4 \\ 0 \\ 2 \end{bmatrix} + c \begin{bmatrix} 3 \\ 1 \\ 0 \end{bmatrix}, \quad \text{where } a, b, c \in \mathbb{R}.$$

The augmented matrix for this system is

$$\left[\begin{array}{ccc|c} 2 & -4 & 3 & 3s_1 - 2t_1 \\ 0 & 0 & 1 & s_1 \\ -1 & 2 & 0 & t_1 \end{array}\right],$$

which row reduces to

$$\left[\begin{array}{ccc|c} 1 & -2 & 0 & t_1 \\ 0 & 0 & 1 & s_1 \\ 0 & 0 & 0 & 0 \end{array}\right].$$

Notice that the rank of the coefficient matrix and the rank of the augmented matrix are both equal to 2, and thus, the system will be consistent for all $\vec{v} \in P$. Consequently, $\operatorname{Span} S_2 = P$.

In addition, we can solve the system to get $c = s_1, b = u$, and $a = 2u - t_1$, for $u \in \mathbb{R}$. Note that the fact that there is a parameter u in this solution, tells us that we have redundancy. Thus, the set S_2 is not linearly independent.

In the special case where $V = \mathbb{F}^n$, we have the following useful criterion.

Proposition 8.4.6 **(Spans $\mathbb{F}^n$ iff rank is n)**

Let $S = \{\vec{v_1}, \ldots, \vec{v_k}\}$ be a set of k vectors in $\mathbb{F}^n$ and let $A = \begin{bmatrix} \vec{v_1} & \cdots & \vec{v_k} \end{bmatrix}$ be the matrix whose columns are the vectors in S. Then

$$\operatorname{Span}(S) = \mathbb{F}^n \quad \text{if and only if} \quad \operatorname{rank}(A) = n.$$

Proof: We have $\operatorname{Span}(S) = \mathbb{F}^n$ if and only if for every $\vec{v} \in \mathbb{F}^n$, the system of equations

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

has a solution.

The coefficient matrix of this system is A. Let $B = \begin{bmatrix} A \mid \vec{v} \end{bmatrix}$ be the corresponding augmented matrix. Then the above system has a solution if and only if $\operatorname{rank}(A) = \operatorname{rank}(B)$.

We want this system to have a solution for every $\vec{v} \in \mathbb{F}^n$; in particular, this holds for every standard basis vector. This happens if and only if the last column of B is never a pivot column, which in turn happens if and only if every row of A has a pivot in it. Since A has n rows, this last statement is equivalent to having $\operatorname{rank}(A) = n$. $\square$

Note that if $\operatorname{rank}\begin{bmatrix} \vec{v_1} & \cdots & \vec{v_k} \end{bmatrix} = n$, then $k \geq n$. Therefore, we need *at least* n vectors in S for $\operatorname{Span}(S)$ to be equal to $\mathbb{F}^n$, which is very reasonable. Of course, the condition for $k \geq n$ is not sufficient on its own — the vectors in S must also form a spanning set.

Example 8.4.7 Determine which (if any) of the following sets span $\mathbb{R}^4$.

(a) $S_1 = \left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ -2 \\ 3 \\ -3 \end{bmatrix}, \begin{bmatrix} 2 \\ 4 \\ 6 \\ 8 \end{bmatrix} \right\}$.

(b) $S_2 = \left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ -2 \\ 3 \\ -3 \end{bmatrix}, \begin{bmatrix} 2 \\ 4 \\ 6 \\ 8 \end{bmatrix}, \begin{bmatrix} 5 \\ 3 \\ 10 \\ 6 \end{bmatrix} \right\}$.

(c) $S_3 = \left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ -2 \\ 3 \\ -3 \end{bmatrix}, \begin{bmatrix} 2 \\ 4 \\ 6 \\ 8 \end{bmatrix}, \begin{bmatrix} 5 \\ 3 \\ 10 \\ 6 \end{bmatrix}, \begin{bmatrix} 3 \\ 5 \\ 6 \\ 10 \end{bmatrix} \right\}$.

Solution:

(a) $\operatorname{Span}(S_1) \neq \mathbb{R}^4$, since we need at least four vectors to span $\mathbb{R}^4$.

(b) There are at least four vectors in S_2, so perhaps it spans $\mathbb{R}^4$. Let us place the vectors in a matrix A and check if $\operatorname{rank}(A) = 4$. We have

$$A = \begin{bmatrix} 1 & 2 & 2 & 5 \\ 1 & -2 & 4 & 3 \\ 1 & 3 & 6 & 10 \\ 1 & -3 & 8 & 6 \end{bmatrix}$$

with

$$\operatorname{RREF}(A) = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Thus $\operatorname{rank}(A) = 3 \neq 4$. So $\operatorname{Span}(S_2) \neq \mathbb{R}^4$.

(c) We again have at least four vectors. Let

$$A = \begin{bmatrix} 1 & 2 & 2 & 5 & 3 \\ 1 & -2 & 4 & 3 & 5 \\ 1 & 3 & 6 & 10 & 6 \\ 1 & -3 & 8 & 6 & 10 \end{bmatrix}.$$

Then

$$\operatorname{RREF}(A) = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

Thus $\operatorname{rank}(A) = 4$, and so in this case $\operatorname{Span}(S_3) = \mathbb{R}^4$.

8.5 Basis

In Definition 8.2.6, we defined a basis for a subspace V of $\mathbb{F}^n$ to be a subset $\mathcal{B} \subseteq V$ that is linearly independent and spans V; that is, $\operatorname{Span}(\mathcal{B}) = V$. In the previous two sections, we examined both aspects of this definition. In this section and the next, we will put everything together to showcase the utility of bases.

We begin with the following result which we had already noted in passing.

Theorem 8.5.1 **(Every Subspace Has a Basis)**

Let V be a subspace of $\mathbb{F}^n$. Then V has a basis.

Proof: If V is not the zero subspace, then the result is given by the proof of Theorem 8.4.1 (Every Subspace Has a Spanning Set). As discussed immediately following Definition 8.2.6, we know that $\varnothing$ is a basis of the trivial subspace $\left\{\vec{0}\right\}$. $\square$

Theorem 8.5.1 (Every Subspace Has a Basis), as it is stated, is merely an existence result. It does not supply us with a method for obtaining a basis for V. (Although, as mentioned in the previous section, the proof of Theorem 8.4.1 (Every Subspace Has a Spanning Set) can be refined into an algorithm.) Nonetheless, in some cases it's possible to explicitly exhibit examples of bases. Here is a special—but very important—example.

Definition 8.5.2

Standard Basis for $\mathbb{F}^n$

In $\mathbb{F}^n$, let $\vec{e}_i$ represent the vector whose i^{th} component is 1 with all other components 0. The set $\mathcal{E} = \{\vec{e_1}, \vec{e_2}, ..., \vec{e_n}\}$ is called the **standard basis for** $\mathbb{F}^n$.

Using Proposition 8.3.6 (Pivots and Linear Independence), we can verify that the standard basis $\mathcal{E}$ is linearly independent. It also spans $\mathbb{F}^n$. Indeed, if $\vec{x} \in \mathbb{F}^n$, then

$$\begin{aligned}\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} &= x_1 \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix} + \cdots + x_n \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix} \\ &= x_1\vec{e_1} + x_2\vec{e_2} + \cdots x_n\vec{e_n}.\end{aligned}$$

Therefore, $\vec{x} \in \operatorname{Span}(\mathcal{E})$. Thus, $\mathcal{E}$ is a basis for $\mathbb{F}^n$.

In general, determining if a set S is a basis for $\mathbb{F}^n$ is a straightforward matter, as the next two propositions show.

Proposition 8.5.3

(Size of Basis for $\mathbb{F}^n$)

Let $S = \{\vec{v_1}, \dots, \vec{v_k}\}$ be a set of k vectors in $\mathbb{F}^n$. If S is a basis for $\mathbb{F}^n$, then $k = n$.

Proof: For a subset S of $\mathbb{F}^n$ to be a basis of $\mathbb{F}^n$, we must have that

1. $\operatorname{Span}(S) = \mathbb{F}^n$, which by Proposition 8.4.6 (Spans $\mathbb{F}^n$ iff rank is n) means that $k \geq n$, and
2. S is linearly independent, which by Corollary 8.3.7 (Bound on Number of Linearly Independent Vectors) means that $k \leq n$.

We conclude that $k = n$. $\square$

Thus, if a subset S of $\mathbb{F}^n$ contains fewer than n vectors, it does not contain enough vectors to span $\mathbb{F}^n$ and is thus not a spanning set of $\mathbb{F}^n$. On the other hand, if S contains more than n vectors in $\mathbb{F}^n$, then it must be linearly dependent.

If S contains exactly n vectors, then it may or may not be a basis for $\mathbb{F}^n$. We still have to check that S is linearly independent and that it spans $\mathbb{F}^n$. In fact, as the next result shows, we only need to check one of these conditions!

Proposition 8.5.4 **(n Vectors in $\mathbb{F}^n$ Span iff Independent)**

Let $S = \{\vec{v_1}, \ldots, \vec{v_n}\}$ be a set of n vectors in $\mathbb{F}^n$. Then S is linearly independent if and only if $\mathrm{Span}(S) = \mathbb{F}^n$.

Proof: Let $A = \begin{bmatrix} \vec{v_1} & \cdots & \vec{v_n} \end{bmatrix}$ be the $n \times n$ matrix whose columns are the vectors in S. By Part (a) of Proposition 8.3.6 (Pivots and Linear Independence), S is linearly independent if and only if $\mathrm{rank}(A) = n$.

On the other hand, from Proposition 8.4.6 (Spans $\mathbb{F}^n$ iff rank is n) we have that $\mathrm{Span}(S) = \mathbb{F}^n$ if and only if $\mathrm{rank}(A) = n$.

Combining both of these results completes the proof. $\square$

Thus, when we are checking to see whether or not a subset S of $\mathbb{F}^n$ is a basis of $\mathbb{F}^n$, below are three options to check this:

1. Check S for linear independence and check S for spanning, or

2. Count the vectors in S and, if S contains exactly n vectors, then check S for spanning using Proposition 8.4.6 (Spans $\mathbb{F}^n$ iff rank is n), or

3. Count the vectors in S and, if S contains exactly n vectors, then check S for linear independence using Proposition 8.3.6 (Pivots and Linear Independence).

In practice, the last option above is usually the quickest.

Example 8.5.5 Let $\vec{v_1} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}, \vec{v_2} = \begin{bmatrix} -1 \\ 2 \\ -3 \\ 4 \end{bmatrix}, \vec{v_3} = \begin{bmatrix} 4 \\ 3 \\ 2 \\ 1 \end{bmatrix}, \vec{v_4} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \vec{v_5} = \begin{bmatrix} 4 \\ -3 \\ 2 \\ -1 \end{bmatrix}$.

Which of the following subsets of $\mathbb{R}^4$ is a basis for $\mathbb{R}^4$?

$$S_1 = \{\vec{v_1}, \vec{v_2}, \vec{v_3}\}, \quad S_2 = \{\vec{v_1}, \vec{v_2}, \vec{v_3}, \vec{v_4}, \vec{v_5}\},$$
$$S_3 = \{\vec{v_1}, \vec{v_2}, \vec{v}_3, \vec{v_4}\}, \quad S_4 = \{\vec{v_1}, \vec{v_2}, \vec{v_3}, \vec{v_5}\}.$$

Solution: Sets S_1 and S_2 fail immediately as they have the wrong number of vectors in them.

Sets S_3 and S_4 have the correct number of vectors in them so we need to investigate them further. Let us check whether they are linear independent.

The matrix whose columns are the vectors in S_3 is

$$A = \begin{bmatrix} 1 & -1 & 4 & 1 \\ 2 & 2 & 3 & 1 \\ 3 & -3 & 2 & 1 \\ 4 & 4 & 1 & 1 \end{bmatrix},$$

which row reduces to

$$\begin{bmatrix} 1 & 0 & 0 & \frac{1}{5} \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & \frac{1}{5} \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Thus, $\operatorname{rank}(A) = 3$. Since $\operatorname{rank}(A) \neq 4$, S_3 is not linearly independent. Consequently, it is not a basis either.

For set S_4, we consider the matrix

$$B = \begin{bmatrix} 1 & -1 & 4 & 4 \\ 2 & 2 & 3 & -3 \\ 3 & -3 & 2 & 2 \\ 4 & 4 & 1 & -1 \end{bmatrix},$$

which row reduces to

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Since $\operatorname{rank}(B) = 4$, it must be the case that S_4 is linearly independent. Since S_4 contains exactly four elements, it is a basis for $\mathbb{R}^4$ by Proposition 8.5.4 (n Vectors in $\mathbb{F}^n$ Span iff Independent).

REMARK

There are two problems we might encounter when trying to obtain a basis for $\mathbb{F}^n$:

(a) We might have a set of vectors $S \subseteq \mathbb{F}^n$ with the property that $\operatorname{Span}(S) = \mathbb{F}^n$, but the set contains more than n vectors, which is too many to be a basis. In this case, S will be linearly dependent. We may apply Proposition 8.3.6 (Pivots and Linear Independence) to produce a subset of S that is linearly independent, but still spans $\mathbb{F}^n$. This subset will be a basis for $\mathbb{F}^n$.

(b) We might have a set of vectors $S \subseteq \mathbb{F}^n$ that is linearly independent, but that contains fewer than n vectors, which is too few to be a basis. In this case, $\operatorname{Span}(S) \neq \mathbb{F}^n$. The problem here is to figure out which vectors to add to S to make it span $\mathbb{F}^n$. One possible approach is to add all n standard basis vectors to S, obtaining a larger set S'. Then certainly $\operatorname{Span}(S') = \mathbb{F}^n$, but now S' is too large to be a basis. This brings us back to **(a)**.

We summarize observations stated in the above remark in Theorem 8.5.6.

Theorem 8.5.6 **(Basis From a Spanning Set or Linearly Independent Set)**

Let $S = \{\vec{v}_1, \ldots, \vec{v}_k\}$ be a subset of $\mathbb{F}^n$.

(a) If $\operatorname{Span}(S) = \mathbb{F}^n$, then there exists a subset $\mathcal{B}$ of S which is a basis for $\mathbb{F}^n$.

(b) If $\operatorname{Span}(S) \neq \mathbb{F}^n$ and S is linearly independent, then there exist vectors $\vec{v}_{k+1}, \ldots, \vec{v}_n$ in $\mathbb{F}^n$ such that $\mathcal{B} = \{\vec{v}_1, \ldots, \vec{v}_k, \vec{v}_{k+1}, \ldots, \vec{v}_n\}$ is a basis for $\mathbb{F}^n$.

Example 8.5.7 Find a basis $\mathcal{B}$ of $\mathbb{R}^3$ that contains the set

$$S = \left\{ \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} -3 \\ -4 \\ -6 \end{bmatrix} \right\}.$$

Solution: Notice that S is linearly independent, since the two vectors it contains are not multiples of one another. We would like to extend S to a basis for $\mathbb{R}^3$. We begin by adding the standard basis vectors to the set to obtain the set

$$S' = \left\{ \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} -3 \\ -4 \\ -6 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}.$$

Now $\operatorname{Span} S' = \mathbb{R}^3$ and we would like to extract a linearly independent subset from S' using the method of Proposition 8.3.6 (Pivots and Linear Independence). The matrix whose columns are the vectors in S' is

$$A = \begin{bmatrix} 1 & -3 & 1 & 0 & 0 \\ 2 & -4 & 0 & 1 & 0 \\ 3 & -6 & 0 & 0 & 1 \end{bmatrix},$$

which row reduces to

$$\begin{bmatrix} 1 & 0 & -2 & 0 & 1 \\ 0 & 1 & -1 & 0 & \frac{1}{3} \\ 0 & 0 & 0 & 1 & -\frac{2}{3} \end{bmatrix}.$$

There are pivots in columns $1, 2$ and 4. Thus, columns $1, 2$ and 4 of A are a basis for $\mathbb{R}^3$:

$$\mathcal{B} = \left\{ \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} -3 \\ -4 \\ -6 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right\}.$$

In the above, we only considered bases for the whole space $\mathbb{F}^n$. In the next section we will discuss examples of bases for special subspaces of $\mathbb{F}^n$. In Chapter 9 we will look at bases for eigenspaces and we will discover a connection to the problem of diagonalizability that we had met in Section 7.6.

8.6 Bases for $\operatorname{Col}(A)$ and $\operatorname{Null}(A)$

Let $A \in M_{m \times n}(\mathbb{F})$. In Propositions 8.1.3 and 8.1.2, we saw that $\operatorname{Col}(A)$ is a subspace of $\mathbb{F}^m$ and that $\operatorname{Null}(A)$ is a subspace of $\mathbb{F}^n$.

Proposition 8.6.1 **(Basis for Col(A))**

Let $A = \begin{bmatrix} \overrightarrow{a_1} & \cdots & \overrightarrow{a_n} \end{bmatrix} \in M_{m\times n}(\mathbb{F})$ and suppose that RREF(A) has pivots in columns $q_1, \ldots, q_r$, where $r = \operatorname{rank}(A)$. Then $\{\overrightarrow{a_{q_1}}, \ldots, \overrightarrow{a_{q_r}}\}$ is a basis for Col(A).

Proof: This follows from Proposition 8.3.6 (Pivots and Linear Independence). □

REMARK

It is important to note that the columns of A—and not those of RREF(A)—are the elements of the basis given in the previous proposition. In general, the columns in RREF(A) are not necessarily in Col(A).

Example 8.6.2 Let $A = \begin{bmatrix} 1 & 3 & 3 & 2 & -9 \\ -2 & -2 & 2 & -8 & 2 \\ 2 & 3 & 0 & 7 & 1 \\ 3 & 4 & -1 & 11 & -8 \end{bmatrix}$. Find a basis for Col($A$).

Solution: Row reduce A to

$$\operatorname{RREF}(A) = \begin{bmatrix} 1 & 0 & -3 & 5 & 0 \\ 0 & 1 & 2 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Since there are pivots in columns $1, 2$ and 5, we conclude that

$$\mathcal{B} = \left\{ \begin{bmatrix} 1 \\ -2 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 3 \\ -2 \\ 3 \\ 4 \end{bmatrix}, \begin{bmatrix} -9 \\ 2 \\ 1 \\ -8 \end{bmatrix} \right\}$$

is a basis for Col(A).

Notice that in this example it would be *incorrect* to say that the set of pivot columns of RREF(A) is a basis for Col(A). Indeed, they all have 0 in the fourth component, so they cannot possibly span Col(A). We can also show that

$$\begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} \notin \operatorname{Col}(A).$$

Next we turn our attention to Null(A). Recall that we can view this subspace as the solution set to the homogeneous system $A\vec{x} = \vec{0}$. If we re-examine the Gauss–Jordan Algorithm, we find that it provides us with a basis for Null(A). Indeed, the solutions corresponding to the free parameters are by design a basis for Null(A).

Example 8.6.3 Let $A = \begin{bmatrix} 1 & -2 & -1 & 3 \\ 2 & -4 & 1 & 0 \\ 1 & -2 & 2 & -3 \end{bmatrix}$. Find a basis for $\mathrm{Null}(A)$.

Solution: The matrix A is the coefficient matrix of the homogeneous system of linear equations

$$\begin{array}{rrrrcl} x_1 & -2x_2 & -x_3 & +3x_4 & = & 0 \\ 2x_1 & -4x_2 & +x_3 & & = & 0 \\ x_1 & -2x_2 & +2x_3 & -3x_4 & = & 0 \end{array}.$$

Thus, $\mathrm{Null}(A)$ is the solution set to this system. In Example 3.7.4, we applied the Gauss–Jordan Algorithm to find that the solution to this system is

$$\mathrm{Null}(A) = \left\{ \begin{bmatrix} 2s - t \\ s \\ 2t \\ t \end{bmatrix} : s, t \in \mathbb{F} \right\} = \left\{ s \begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \end{bmatrix} + t \begin{bmatrix} -1 \\ 0 \\ 2 \\ 1 \end{bmatrix} : s, t \in \mathbb{F} \right\}.$$

Notice that the Gauss–Jordan Algorithm immediately supplies us with the spanning set

$$\mathcal{B} = \left\{ \begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 2 \\ 1 \end{bmatrix} \right\}.$$

Moreover, observe that this spanning set is linearly independent since the two vectors it contains are not multiples of each other. Thus, $\mathcal{B}$ is a basis for $\mathrm{Null}(A)$.

Digging a bit deeper, the real reason the set $\mathcal{B}$ in the previous example is linearly independent is because it contains the solution vectors that were constructed using the free parameters x_2 and x_4. (Refer back to Example 3.7.4.) This forces the first vector in the spanning set to have a 1 in the second component and forces the second vector to have a 0 in the second component. (It also forces the second vector to have a 1 in the fourth component and the first vector to have a 0 in the fourth component.)

Example 8.6.4 Let $A = \begin{bmatrix} -1 & -1 & -2 & 3 & 1 \\ -9 & 5 & -4 & -1 & -5 \\ 7 & -5 & 2 & 3 & 5 \end{bmatrix}$. Find a basis for $\mathrm{Null}(A)$.

Solution:

The matrix A is the coefficient matrix of the homogeneous system of linear equations

$$\begin{array}{rrrrrcl} -x_1 & -2x_2 & -2x_3 & +3x_4 & +x_5 & = & 0 \\ -9x_1 & 5x_2 & -4x_3 & -x_4 & -5x_5 & = & 0 \\ 7x_1 & -5x_2 & +2x_3 & +3x_4 & +5x_5 & = & 0 \end{array}.$$

Thus, $\mathrm{Null}(A)$ is the solution set to this system. We have

$$\mathrm{RREF}(A) = \begin{bmatrix} 1 & 0 & 1 & -1 & 0 \\ 0 & 1 & 1 & -2 & -1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

There are pivots in columns 1 and 2, so we may take $x_3 = s$, $x_4 = t$ and $x_5 = r$ as free parameters. Then $x_1 = -x_3 + x_4 = -s + t$ and $x_2 = -x_3 + 2x_4 + x_5 = -s + 2t + r$, and therefore, the solution set is given by

$$\operatorname{Null}(A) = \left\{ \begin{bmatrix} -s+t \\ -s+2t+r \\ s \\ t \\ r \end{bmatrix} : s,t,r \in \mathbb{F} \right\} = \left\{ s\begin{bmatrix} -1 \\ -1 \\ 1 \\ 0 \\ 0 \end{bmatrix} + t\begin{bmatrix} 1 \\ 2 \\ 0 \\ 1 \\ 0 \end{bmatrix} + r\begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 1 \end{bmatrix} : s,t,r \in \mathbb{F} \right\}.$$

Consequently, the set

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

is a spanning set for Null(A). It is also linearly independent. Indeed, if some linear combination of the vectors in the set is zero, that is,

$$s\begin{bmatrix} -1 \\ -1 \\ 1 \\ 0 \\ 0 \end{bmatrix} + t\begin{bmatrix} 1 \\ 2 \\ 0 \\ 1 \\ 0 \end{bmatrix} + r\begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix},$$

then

$$\begin{bmatrix} -s+t \\ -s+2t+r \\ s \\ t \\ r \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

and hence $s = t = r = 0$ by equating the third, fourth and fifth entries. So the only possible such linear combination is the trivial linear combination. Therefore, $\mathcal{B}$ is linearly independent.

Notice that the third, fourth and fifth entries correspond to the free parameters. This is no coincidence. Indeed, in the set $\mathcal{B}$, each vector has a nonzero entry in the row corresponding to its free parameter while all other vectors in $\mathcal{B}$ have a 0 in that same row.

The results in the previous two examples may be generalized as follows.

Proposition 8.6.5 **(Basis for Null(A))**

Let $A \in M_{m\times n}(\mathbb{F})$ and consider the homogeneous linear system $A\vec{x} = \vec{0}$. Suppose that, after applying the Gauss–Jordan Algorithm, we obtain k free parameters so that the solution set to this system is given by

$$\operatorname{Null}(A) = \{t_1\vec{x_1} + \cdots + t_k\vec{x_k} \colon t_1, \ldots, t_k \in \mathbb{F}\}.$$

Here $k = \text{nullity}(A) = n - \text{rank}(A)$ and the parameters t_i and the vectors $\overrightarrow{x_i}$ for $1 \leq i \leq k$ are obtained using the method outlined in Section 3.7.

Then $\{\overrightarrow{x_1}, \ldots, \overrightarrow{x_k}\}$ is a basis for $\text{Null}(A)$.

Proof: Since $\text{Null}(A) = \{t_1\overrightarrow{x_1} + \cdots + t_k\overrightarrow{x_k} : t_1, \ldots, t_k \in \mathbb{F}\}$, we see that $\mathcal{B} = \{\overrightarrow{x_1}, \ldots, \overrightarrow{x_k}\}$ is a spanning set. The fact that it is linearly independent follows from the way the solution vectors $\overrightarrow{x_i}$ were constructed such that only $\overrightarrow{x_i}$ has a nonzero entry in the row corresponding to the i^{th} free variable. So if there are scalars $c_i \in \mathbb{F}$ so that

$$c_1\overrightarrow{x_1} + \cdots + c_k\overrightarrow{x_k} = \overrightarrow{0},$$

then by comparing i^{th} coefficients of both sides of the equation, we get that $c_i = 0$ for all i. Thus, $\mathcal{B}$ is a basis for $\text{Null}(A)$, as claimed. $\square$

8.7 Dimension

A nonzero subspace V of $\mathbb{F}^n$ will have infinitely many different bases. If $\mathcal{B} = \{\overrightarrow{v_1}, \ldots, \overrightarrow{v_k}\}$ is a basis for V, then so is $\mathcal{B}' = \{c\overrightarrow{v_1}, \ldots, \overrightarrow{v_k}\}$ for any nonzero scalar $c \in \mathbb{F}$. More drastically, two different bases can contain substantially different vectors, and not just ones that are scalar multiples of each other.

Example 8.7.1 In Section 8.5, we saw two bases for $\mathbb{R}^4$: the standard basis

$$\mathcal{S} = \{\overrightarrow{e_1}, \overrightarrow{e_2}, \overrightarrow{e_3}, \overrightarrow{e_4}\} = \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \right\},$$

and the basis

$$S_4 = \left\{ \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}, \begin{bmatrix} -1 \\ 2 \\ -3 \\ 4 \end{bmatrix}, \begin{bmatrix} 4 \\ 3 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 4 \\ -3 \\ 2 \\ -1 \end{bmatrix} \right\}$$

given in Example 8.5.5. From our work in Example 8.4.7(b), we can obtain yet another basis:

$$S_2 = \left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ -2 \\ 3 \\ -3 \end{bmatrix}, \begin{bmatrix} 2 \\ 4 \\ 6 \\ 8 \end{bmatrix}, \begin{bmatrix} 5 \\ 3 \\ 10 \\ 6 \end{bmatrix} \right\}.$$

Although the three sets in the above example contain different vectors, they share one thing in common. They each contain precisely four vectors. This agrees with Proposition 8.5.3 (Size of Basis for $\mathbb{F}^n$), which states that a basis for $\mathbb{F}^n$ must contain precisely n vectors.

The same kind of result is true for a subspace V of $\mathbb{F}^n$. Even though V may have many different bases, any two bases will have the same number of elements.

Theorem 8.7.2 **(Dimension is Well-Defined)**

Let V be a subspace of $\mathbb{F}^n$. If $\mathcal{B} = \{\vec{v_1}, \ldots, \vec{v_k}\}$ and $\mathcal{C} = \{\vec{w_1}, \ldots, \vec{w_\ell}\}$ are bases for V, then $k = \ell$.

Proof: Since $\mathcal{B}$ and $\mathcal{C}$ are bases for V, we have that $V = \operatorname{Span}(\mathcal{B}) = \operatorname{Span}(\mathcal{C})$. This means, in particular, that $\vec{w_1} \in \operatorname{Span}(\mathcal{B})$. Thus there are scalars $c_1, \ldots, c_k$ such that

$$\vec{w_1} = c_1\vec{v_1} + c_2\vec{v_2} + \cdots + c_k\vec{v_k}.$$

At least one of these scalars must be nonzero, for if they were all zero then $\vec{w_1}$ would be equal to $\vec{0}$, which would contradict the fact that $\mathcal{C}$ is linearly independent (by Proposition 8.3.2(a)).

We may assume, without loss of generality, that $c_1 \neq 0$. Then we can write

$$\vec{v_1} = \frac{1}{c_1}(\vec{w_1} - c_2\vec{v_2} - \cdots - c_k\vec{v_k}).$$

This shows that $\operatorname{Span}\{\vec{v_1}, \ldots, \vec{v_k}\} = \operatorname{Span}\{\vec{w_1}, \vec{v_2}, \ldots, \vec{v_k}\}$. Thus, we have effectively replaced $\vec{v_1}$ with $\vec{w_1}$. By repeating this same argument with $\vec{w_1}, \vec{v_2}, \ldots, \vec{v_k}$ and $\vec{w_2}$, we can show that we can replace some other $\vec{v_i}$, say without loss of generality $\vec{v_2}$, while still having $\operatorname{Span}\{\vec{v_1}, \ldots, \vec{v_k}\} = \operatorname{Span}\{\vec{w_1}, \vec{w_2}, \vec{v_3}, \ldots, \vec{v_k}\}$. We can continue this replacement procedure for each $\vec{w_i}$ in $\mathcal{C}$. This is only possible if $\ell \leq k$.

Using a similar argument where the roles of $\mathcal{B}$ and $\mathcal{C}$ are swapped, we must also have that $k \leq \ell$. Thus, $k = \ell$. $\square$

Definition 8.7.3
Dimension

The number of elements in a basis for a subspace V of $\mathbb{F}^n$ is called the **dimension** of V. We denote this number by $\dim(V)$.

Since every subspace V of $\mathbb{F}^n$ has a basis (by Theorem 8.5.1 (Every Subspace Has a Basis)), and since we just proved in Theorem 8.7.2 (Dimension is Well-Defined) that any two bases for V have the same number of elements, we see that $\dim(V)$ is *well-defined* (or unambiguous). We can evaluate $\dim(V)$ by counting the number of vectors in *any* basis for V.

Example 8.7.4

The standard basis $\mathcal{E} = \{\vec{e_1}, \ldots, \vec{e_n}\}$ of $\mathbb{F}^n$ consist of n vectors. Thus, $\dim(\mathbb{F}^n) = n$. This aligns with our intuition of $\mathbb{F}^n$ being "n-dimensional".

How large can the dimension of a subspace of $\mathbb{F}^n$ be?

Proposition 8.7.5 **(Bound on Dimension of Subspace)**

Let V be a subspace of $\mathbb{F}^n$. Then $\dim(V) \leq n$.

Proof: The proof of Theorem 8.4.1 (Every Subspace Has a Spanning Set) gave us that every subspace has a basis, and established that the basis generated in the proof was guaranteed to have at most n elements. $\square$

EXERCISE

Let V be a subspace of $\mathbb{F}^n$. Show that $V = \mathbb{F}^n$ if and only if $\dim(V) = n$.

Example 8.7.6 Let L be the line through the origin in $\mathbb{R}^n$ with direction vector $\vec{d} \neq \vec{0}$. That is,

$$L = \{t\vec{d} : t \in \mathbb{R}\} = \operatorname{Span}\left\{\vec{d}\right\}.$$

Since $\vec{d} \neq 0$, the set $\left\{\vec{d}\right\}$ is linearly independent, and by the above it spans L. Thus, $\mathcal{B} = \left\{\vec{d}\right\}$ is a basis for L. Consequently, $\dim(L) = 1$.

That is, lines through the origin are 1-dimensional subspaces of $\mathbb{R}^n$.

Example 8.7.7 Let P be the plane in $\mathbb{R}^3$ given by the scalar equation $2x - 6y + 4z = 0$. In Example 8.4.3, we saw that

$$S = \left\{ \begin{bmatrix} 3 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -2 \\ 0 \\ 1 \end{bmatrix} \right\}$$

is a basis for P. Thus, $\dim(P) = 2$.

More generally, if P is any plane through the origin in $\mathbb{R}^3$, then we can show that any two linearly independent vectors in P will form a basis for P. Thus, $\dim(P) = 2$ in all cases.

That is, planes through the origin are 2-dimensional subspaces of $\mathbb{R}^n$.

The next two examples of dimension computations are important enough to warrant being put in a proposition.

Proposition 8.7.8 **(Rank and Nullity as Dimensions)**

Let $A \in M_{m\times n}(\mathbb{F})$. Then

(a) $\operatorname{rank}(A) = \dim(\operatorname{Col}(A))$, and

(b) $\operatorname{nullity}(A) = \dim(\operatorname{Null}(A))$.

Proof: (a) follows from Proposition 8.6.1 (Basis for $\operatorname{Col}(A)$).
(b) follows from Proposition 8.6.5 (Basis for $\operatorname{Null}(A)$). $\square$

Using the definition of nullity (Definition 3.6.8), we arrive at the following important result.

Theorem 8.7.9 **(Rank–Nullity Theorem)**

Let $A \in M_{m\times n}(\mathbb{F})$. Then

$$\begin{aligned} n &= \text{rank}(A) + \text{nullity}(A) \\ &= \dim(\text{Col}(A)) + \dim(\text{Null}(A)). \end{aligned}$$

Proof: This result follows immediately from the fact that $\text{nullity}(A) = n - \text{rank}(A)$, together with Proposition 8.7.8 (Rank and Nullity as Dimensions). $\square$

This relationship between rank and nullity is one of the central results of linear algebra. Although the above proof seems short, it contains a significant amount of content.

8.8 Coordinates

In this section, we discuss one of the most important results about a basis. This result will naturally lead us to the idea of coordinates with respect to bases.

Theorem 8.8.1 **(Unique Representation Theorem)**

Let $\mathcal{B} = \{\vec{v_1}, \vec{v_2}, \ldots, \vec{v_n}\}$ be a basis for $\mathbb{F}^n$. Then, for every vector $\vec{v} \in \mathbb{F}^n$, there exist *unique* scalars $c_1, c_2, \ldots, c_n \in \mathbb{F}$ such that

$$\vec{v} = c_1\,\vec{v_1} + c_2\,\vec{v_2} + \cdots + c_n\,\vec{v_n}.$$

Proof: First we prove that such scalars exist. Since $\mathcal{B}$ is a basis for $\mathbb{F}^n$, $\text{Span}\,(\mathcal{B}) = \mathbb{F}^n$, and so each vector $\vec{v}$ can be written as a linear combination of elements of $\mathcal{B}$, that is, there exist scalars $c_1, c_2, \ldots, c_n \in \mathbb{F}$ such that $\vec{v} = c_1\,\vec{v_1} + c_2\,\vec{v_2} + \cdots + c_n\,\vec{v_n}$.

Next, we show that the representation is unique. Suppose that we can also express $\vec{v}$ as $\vec{v} = d_1\,\vec{v_1} + d_2\,\vec{v_2} + \cdots + d_n\,\vec{v_n}$, for some scalars $d_1, d_2, \ldots, d_n \in \mathbb{F}$.

If we subtract these two expressions for $\vec{v}$, then we have

$$\begin{aligned} \vec{0} = \vec{v} - \vec{v} &= (c_1\,\vec{v_1} + c_2\,\vec{v_2} + \cdots + c_n\vec{v_n}) - (d_1\,\vec{v_1} + d_2\,\vec{v_2} + \cdots + d_n\,\vec{v_n}) \\ &= (c_1 - d_1)\vec{v_1} + (c_2 - d_2)\vec{v_2} + \cdots + (c_n - d_n)\vec{v_n}. \end{aligned}$$

Since $\mathcal{B}$ is linearly independent, we have that $(c_1 - d_1) = 0, (c_2 - d_2) = 0, \ldots, (c_n - d_n) = 0$. That is, $c_1 = d_1, c_2 = d_2, \ldots, c_n = d_n$, and thus, the representation for $\vec{v}$ is unique. $\square$

Example 8.8.2

Let $\mathcal{E} = \{\vec{e_1}, \ldots, \vec{e_n}\}$ be the standard basis for $\mathbb{F}^n$. Given $\vec{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{F}^n$, we have

$$\vec{x} = x_1\vec{e_1} + \cdots + x_n\vec{e_n}.$$

Thus, the unique scalars in the representation of $\vec{x}$ in terms of the standard basis are the coordinates (or components) of $\vec{x}$.

The previous Example motivates our next definition.

Definition 8.8.3
Coordinates and Components

Let $\mathcal{B} = \{\vec{v_1}, \vec{v_2}, \ldots, \vec{v_n}\}$ be a basis for $\mathbb{F}^n$. Let the vector $\vec{v} \in \mathbb{F}^n$ have representation

$$\vec{v} = c_1\,\vec{v_1} + c_2\,\vec{v_2} + \cdots + c_n\,\vec{v_n} \qquad (c_i \in \mathbb{F}).$$

We call the scalars $c_1, c_2, \ldots, c_n$ the **coordinates** (or **components**) **of $\vec{v}$ with respect to $\mathcal{B}$**, or the **$\mathcal{B}$-coordinates of $\vec{v}$**.

We would like to use this definition to create a column vector $\begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix}$ using the coordinates $c_1, \ldots, c_n$ of $\vec{v}$ with respect to $\mathcal{B}$. However, the ordering of the vectors in a basis (that is, the assignment of labels $1, 2, ..., n$ for $\vec{v_1}, \vec{v_2}, ..., \vec{v_n}$) is arbitrary, and permutations of the ordering result in the same basis. Unfortunately, each of these permutations can result in different representations of the vector under $\mathcal{B}$, which could be a source of confusion.

The next definition helps us address this subtlety.

Definition 8.8.4
Ordered Basis

An **ordered basis for** $\mathbb{F}^n$ is a basis $\mathcal{B} = \{\vec{v_1}, \vec{v_2}, \ldots, \vec{v_n}\}$ for $\mathbb{F}^n$ together with a fixed ordering.

REMARK

When we refer to the set $\{\vec{v_1}, \vec{v_2}, \ldots, \vec{v_n}\}$ as being ordered, we are indicating that $\vec{v_1}$ is the first element in the ordering, that $\vec{v_2}$ is the second, and so on.

Thus even though $\{\vec{v_1}, \vec{v_2}\}$ and $\{\vec{v_2}, \vec{v_1}\}$ are the same *set*, they are different from the point of view of orderings.

A given basis $\{\vec{v_1}, \vec{v_2}, \ldots, \vec{v_n}\}$ gives rise to $n!$ ordered bases, one for each possible ordering (permutation) of the entries.

Example 8.8.5

The set $\mathcal{B} = \left\{ \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 5 \\ 6 \end{bmatrix} \right\}$ is a basis for $\mathbb{R}^2$. The two vectors in $\mathcal{B}$ are not multiples of each other, so $\mathcal{B}$ is linearly independent. Since $\mathcal{B}$ contains precisely two vectors, it must also span $\mathbb{R}^2$, by Proposition 8.5.4 (n Vectors in $\mathbb{F}^n$ Span iff Independent). Hence $\mathcal{B}$ is a basis for $\mathbb{R}^2$.

The basis $\mathcal{B}$ gives rise to two different ordered bases:

$$\left\{ \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 5 \\ 6 \end{bmatrix} \right\} \qquad \text{and} \qquad \left\{ \begin{bmatrix} 5 \\ 6 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \end{bmatrix} \right\}.$$

Definition 8.8.6
Coordinate Vector

Let $\mathcal{B} = \{\vec{v_1}, \dots, \vec{v_n}\}$ be an ordered basis for $\mathbb{F}^n$. Let $\vec{v} \in \mathbb{F}^n$ have coordinates $c_1, \dots, c_n$ with respect to $\mathcal{B}$, where the ordering of the scalars c_i matches the ordering in $\mathcal{B}$, that is,

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

Then the **coordinate vector of $\vec{v}$ with respect to $\mathcal{B}$** (or the **$\mathcal{B}$-coordinate vector of** $\vec{v}$) is the column vector in $\mathbb{F}^n$

$$[\vec{v}]_{\mathcal{B}} = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix}.$$

If we choose a different ordering on the basis $\mathcal{B}$ then the entries of $[\vec{v}]_{\mathcal{B}}$ will be permuted accordingly.

Example 8.8.7

Consider the ordered basis $\mathcal{B} = \left\{ \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 5 \\ 6 \end{bmatrix} \right\}$ for $\mathbb{R}^2$ and let $\vec{v} = \begin{bmatrix} 7 \\ 10 \end{bmatrix}$. By inspection, we have

$$\vec{v} = (2) \begin{bmatrix} 1 \\ 2 \end{bmatrix} + (1) \begin{bmatrix} 5 \\ 6 \end{bmatrix}.$$

So the $\mathcal{B}$-coordinates of $\vec{v}$ are 2 and 1. Therefore,

$$[\vec{v}]_{\mathcal{B}} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}.$$

If we instead use the ordered basis $\mathcal{C} = \left\{ \begin{bmatrix} 5 \\ 6 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \end{bmatrix} \right\}$, we would have

$$[\vec{v}]_{\mathcal{C}} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}.$$

In using coordinates, it is therefore important to indicate clearly which ordered basis you are using.

REMARK (Standard Ordering)

Let $\mathcal{E} = \{\vec{e_1}, \dots, \vec{e_n}\}$ be the standard basis for $\mathbb{F}^n$, ordered so that $\vec{e_i}$ is the ith vector. We call this the **standard ordering**. Unless explicitly stated otherwise, we always assume that $\mathcal{E}$ is given the standard ordering.

In the early part of the course, we had been implicitly using coordinate vectors with respect to $\mathcal{E}$. Indeed, whenever we had written

$$\vec{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$$

we were implicitly writing down $[\vec{v}]_{\mathcal{E}}$.

Moving forwards, when we attempt to describe some vector in $\mathbb{F}^n$, we still have to represent it somehow. Unless we have some other explicit arrangement, we will do this as above, i.e., by making use of the standard basis (with its standard ordering). We will often suppress the notation $[\vec{v}]_{\mathcal{E}}$ when doing so, and instead simply write $\vec{v}$ as we have been doing so far.

REMARK

The notation of $[\vec{v}]_{\mathcal{E}}$ (and more generally $[\vec{v}]_{\mathcal{B}}$) is reminiscent of the notation $[T]_{\mathcal{E}}$ for the standard matrix representation of the linear transformation T. In both cases, the notation conveys the idea that we are taking an object (a vector or linear transformation) and representing it as an array of numbers (a column vector or matrix) by using a basis to obtain this representation.

There is another connection between these two pieces of notation, as we will see later.

Note that for an ordered basis $\mathcal{B} = \{\vec{v_1}, \ldots, \vec{v_n}\}$ of $\mathbb{F}^n$, the relationship between $\vec{v}$ and $[\vec{v}]_{\mathcal{B}}$ is a two-way relationship:

$$\vec{v} = c_1\,\vec{v_1} + \cdots + c_n\,\vec{v_n} \iff [\vec{v}]_{\mathcal{B}} = \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix}.$$

Thus, if we have $\vec{v}$, we can obtain $[\vec{v}]_{\mathcal{B}}$, and vice versa. We leave the proof of the following result as an exercise.

Theorem 8.8.8 **(Linearity of Taking Coordinates)**

Let $\mathcal{B} = \{\vec{v_1}, \ldots, \vec{v_n}\}$ be an ordered basis for $\mathbb{F}^n$. Then the function $[\quad]_{\mathcal{B}} : \mathbb{F}^n \to \mathbb{F}^n$ defined by sending $\vec{v}$ to $[\vec{v}]_{\mathcal{B}}$ is *linear*:

(a) For all $\vec{u}, \vec{v} \in \mathbb{F}^n$, $[\vec{u} + \vec{v}]_{\mathcal{B}} = [\vec{u}]_{\mathcal{B}} + [\vec{v}]_{\mathcal{B}}$.

(b) For all $\vec{v} \in \mathbb{F}^n$ and $c \in \mathbb{F}$, $[c\vec{v}]_{\mathcal{B}} = c[\vec{v}]_{\mathcal{B}}$.

We now provide some examples of computing coordinates with respect to bases for $\mathbb{F}^n$.

Example 8.8.9

Consider the ordered basis $\mathcal{B} = \left\{ \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 5 \\ 6 \end{bmatrix} \right\}$ for $\mathbb{R}^2$. Let $\vec{v} = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$. (By this we mean $[\vec{v}]_{\mathcal{E}}$, as explained at the end of the above remark on the standard ordering.)

(a) Find $[\vec{v}]_{\mathcal{B}}$.

(b) If $[\vec{w}]_{\mathcal{B}} = \begin{bmatrix} -3 \\ 2 \end{bmatrix}$, find $\vec{w}$ (i.e., find $[\vec{w}]_{\mathcal{E}}$).

Solution:

(a) To find $[\vec{v}]_{\mathcal{B}}$, we need to find the coordinates of $\vec{v}$ with respect to $\mathcal{B}$. Thus, we want to find $a, b \in \mathbb{F}$ so that

$$a \begin{bmatrix} 1 \\ 2 \end{bmatrix} + b \begin{bmatrix} 5 \\ 6 \end{bmatrix} = \begin{bmatrix} 3 \\ 4 \end{bmatrix}.$$

This leads us to the system

$$\begin{bmatrix} 1 & 5 \\ 2 & 6 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$$

with augmented matrix

$$\left[\begin{array}{cc|c} 1 & 5 & 3 \\ 2 & 6 & 4 \end{array}\right],$$

which reduces to

$$\left[\begin{array}{cc|c} 1 & 0 & \frac{1}{2} \\ 0 & 1 & \frac{1}{2} \end{array}\right].$$

Thus, $a = \frac{1}{2}$ and $b = \frac{1}{2}$. These are the coordinates of $\vec{v}$ with respect of $\mathcal{B}$. Therefore,

$$\begin{bmatrix} 3 \\ 4 \end{bmatrix}_{\mathcal{B}} = \begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \end{bmatrix}.$$

(b) We are given that $[\vec{w}]_{\mathcal{B}} = \begin{bmatrix} -3 \\ 2 \end{bmatrix}$. This means that

$$\vec{w} = (-3) \begin{bmatrix} 1 \\ 2 \end{bmatrix} + 2 \begin{bmatrix} 5 \\ 6 \end{bmatrix} = \begin{bmatrix} 7 \\ 6 \end{bmatrix}.$$

That is,

$$[\vec{w}]_{\mathcal{E}} = \vec{w} = \begin{bmatrix} 7 \\ 6 \end{bmatrix}.$$

Example 8.8.10

Let $\mathcal{B} = \left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 3 \end{bmatrix}, \begin{bmatrix} 1 \\ -2 \\ 4 \end{bmatrix} \right\}$ and let $\vec{v} = \begin{bmatrix} -3 \\ 2 \\ -6 \end{bmatrix}$.

(a) Show that $\mathcal{B}$ is a basis for $\mathbb{R}^3$.

(b) Viewing $\mathcal{B}$ as an ordered basis in the given order, find $[\vec{v}]_{\mathcal{B}}$.

(c) If $[\vec{w}]_{\mathcal{B}} = \begin{bmatrix} 3 \\ -4 \\ 5 \end{bmatrix}$, find $[\vec{w}]_{\mathcal{E}}$.

Solution:

(a) By Proposition 8.5.4 (n Vectors in $\mathbb{F}^n$ Span iff Independent), since we have three vectors in $\mathcal{B}$, it suffices to check that they span $\mathbb{R}^3$.

By Proposition 8.4.6 (Spans $\mathbb{F}^n$ iff rank is n), this can be done by evaluating the rank of

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & -2 \\ 1 & 3 & 4 \end{bmatrix}.$$

Row reduction yields the matrix $R = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 3 \\ 0 & 0 & 3 \end{bmatrix}$, which has rank 3. Thus, we have a basis.

(b) We need to find $a, b, c \in \mathbb{R}$ so that

$$a \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + b \begin{bmatrix} 1 \\ 0 \\ 3 \end{bmatrix} + c \begin{bmatrix} 1 \\ -2 \\ 4 \end{bmatrix} = \begin{bmatrix} -3 \\ 2 \\ -6 \end{bmatrix}.$$

Consider the system

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & -2 \\ 1 & 3 & 4 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} -3 \\ 2 \\ -6 \end{bmatrix},$$

which has augmented matrix

$$\left[\begin{array}{ccc|c} 1 & 1 & 1 & -3 \\ 1 & 0 & -2 & 2 \\ 1 & 3 & 4 & -6 \end{array}\right],$$

which reduces to

$$\left[\begin{array}{ccc|c} 1 & 0 & 0 & \frac{-8}{3} \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & \frac{-7}{3} \end{array}\right].$$

Thus, $a = -\frac{8}{3}, b = 2$ and $c = -\frac{7}{3}$, and we have

$$[\vec{v}]_{\mathcal{B}} = \begin{bmatrix} -\frac{8}{3} \\ 2 \\ -\frac{7}{3} \end{bmatrix}.$$

(c) If

$$[\vec{w}]_{\mathcal{B}} = \begin{bmatrix} 3 \\ -4 \\ 5 \end{bmatrix},$$

then

$$\vec{w} = (3)\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + (-4)\begin{bmatrix} 1 \\ 0 \\ 3 \end{bmatrix} + (5)\begin{bmatrix} 1 \\ -2 \\ 4 \end{bmatrix} = \begin{bmatrix} 4 \\ -7 \\ 11 \end{bmatrix},$$

that is,

$$[\vec{w}]_{\mathcal{E}} = \vec{w} = \begin{bmatrix} 4 \\ -7 \\ 11 \end{bmatrix}.$$

In the previous two examples, we saw that in order to obtain the coordinate vector representation $[\vec{v}]_{\mathcal{B}}$ of a given vector $\vec{v}$, we must solve a certain system of equations. Must we do this *each time* we want to find coordinate vectors with respect to $\mathcal{B}$? The answer is "no"; there is a more efficient approach.

Example 8.8.11 Let us first consider the final calculation of Example 8.8.9 (b), where we had

$$[\vec{w}]_{\mathcal{B}} = \begin{bmatrix} -3 \\ 2 \end{bmatrix}, \quad \text{and then} \quad [\vec{w}]_{\mathcal{E}} = \vec{w} = \begin{bmatrix} 7 \\ 6 \end{bmatrix}.$$

This was completed by performing the calculation

$$(-3)\begin{bmatrix} 1 \\ 2 \end{bmatrix} + 2\begin{bmatrix} 5 \\ 6 \end{bmatrix} = \vec{w}.$$

The system above can be expressed as

$$\begin{bmatrix} 1 & 5 \\ 2 & 6 \end{bmatrix}\begin{bmatrix} -3 \\ 2 \end{bmatrix}.$$

Thus, there is a matrix at work here, namely $M = \begin{bmatrix} 1 & 5 \\ 2 & 6 \end{bmatrix}$, such that multiplication by M allows us to go from the new coordinate system (with respect to the basis $\mathcal{B}$) to the standard system (basis $\mathcal{E}$).

For instance, if we are given a vector $\vec{z}$ with $[\vec{z}]_{\mathcal{B}} = \begin{bmatrix} 4 \\ -3 \end{bmatrix}$, then

$$[\vec{z}]_{\mathcal{E}} = M[\vec{z}]_{\mathcal{B}} = \begin{bmatrix} 1 & 5 \\ 2 & 6 \end{bmatrix}\begin{bmatrix} 4 \\ -3 \end{bmatrix} = \begin{bmatrix} -11 \\ -10 \end{bmatrix}.$$

Notice that M is very special. Its columns are the coordinate vectors of $\mathcal{B}$ expressed in the standard basis.

What about going the other way, from $\mathcal{E}$ to $\mathcal{B}$? Let us re-examine what we did in Example 8.8.9 (a). We wanted to obtain $\begin{bmatrix} 3 \\ 4 \end{bmatrix}_{\mathcal{B}}$, which we found amounted to solving the system

$$\begin{bmatrix} 1 & 5 \\ 2 & 6 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 3 \\ 4 \end{bmatrix}.$$

The coefficient matrix here is the same M above. *It is invertible!* (This is a general phenomenon, as we will soon see.) Thus,

$$\begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 1 & 5 \\ 2 & 6 \end{bmatrix}^{-1} \begin{bmatrix} 3 \\ 4 \end{bmatrix} = -\frac{1}{4} \begin{bmatrix} 6 & -5 \\ -2 & 1 \end{bmatrix} \begin{bmatrix} 3 \\ 4 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \end{bmatrix}.$$

From this, we get

$$\begin{bmatrix} 3 \\ 4 \end{bmatrix}_{\mathcal{B}} = \begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \end{bmatrix}.$$

Once again there is a matrix at work here, namely $M^{-1} = -\frac{1}{4} \begin{bmatrix} 6 & -5 \\ -2 & 1 \end{bmatrix}$. It will change coordinates from the standard system (basis $\mathcal{E}$) to the new system (basis $\mathcal{B}$), and it will do this for any vector whose components we already have in $\mathcal{E}$.

For instance, given a vector $\vec{u}$ with $[\vec{u}]_{\mathcal{E}} = \begin{bmatrix} -2 \\ 5 \end{bmatrix}$, then

$$[\vec{u}]_{\mathcal{B}} = M^{-1}[\vec{u}]_{\mathcal{E}} = -\frac{1}{4} \begin{bmatrix} 6 & -5 \\ -2 & 1 \end{bmatrix} \begin{bmatrix} -2 \\ 5 \end{bmatrix} = \frac{1}{4} \begin{bmatrix} 37 \\ -9 \end{bmatrix}.$$

To obtain M^{-1}, we invert the matrix whose columns are the coordinates of the new basis vector in $\mathcal{B}$ expressed in the standard basis.

With some work, you can show that the columns of M^{-1} are in fact the coordinate vectors of the standard basis vectors with respect to $\mathcal{B}$.

The idea of using a matrix like M above to change coordinates from one basis to another is in fact always possible, as we now explain.

Definition 8.8.12

Change-of-Basis Matrix, Change-of-Coordinate Matrix

Let $\mathcal{B} = \{\vec{v_1}, \ldots, \vec{v_n}\}$ and $\mathcal{C} = \{\vec{w_1}, \ldots, \vec{w_n}\}$ be ordered bases for $\mathbb{F}^n$.

The **change-of-basis** (or **change-of-coordinates**) matrix **from $\mathcal{B}$-coordinates to $\mathcal{C}$-coordinates** is the $n \times n$ matrix

$${}_{\mathcal{C}}[I]_{\mathcal{B}} = \left[\, [\vec{v_1}]_{\mathcal{C}} \cdots [\vec{v_n}]_{\mathcal{C}} \,\right]$$

whose columns are the $\mathcal{C}$-coordinates of the vectors $\vec{v_i}$ in $\mathcal{B}$.

Similarly, the **change-of-basis** (or **change-of-coordinates**) matrix **from $\mathcal{C}$-coordinates to $\mathcal{B}$-coordinates** is the $n \times n$ matrix

$${}_{\mathcal{B}}[I]_{\mathcal{C}} = \left[\, [\vec{w_1}]_{\mathcal{B}} \cdots [\vec{w_n}]_{\mathcal{B}} \,\right]$$

whose columns are the $\mathcal{B}$-coordinates of the vectors $\vec{w_i}$ in $\mathcal{C}$.

REMARKS

- The reason for the notation will become apparent later. The resemblance to the notation $[T]_{\mathcal{E}}$ for the standard matrix of the linear transformation T is intentional.
- Other sources may notate ${}_{\mathcal{C}}[I]_{\mathcal{B}}$ as ${}_{\mathcal{C}}P_{\mathcal{B}}$ or $P_{\mathcal{C}\leftarrow\mathcal{B}}$.

Example 8.8.13 The matrix M in Example 8.8.9 is none other than ${}_{\mathcal{E}}[I]_{\mathcal{B}}$, while the matrix M^{-1} is ${}_{\mathcal{B}}[I]_{\mathcal{E}}$. (This is not a coincidence. See Corollary 8.8.16 (Inverse of Change-of-Basis Matrix).)

The role of the change-of-basis matrix is exactly as its name suggests. It changes coordinates from one basis to another.

Proposition 8.8.14 **(Changing a Basis)**

Let $\mathcal{B} = \{\overrightarrow{v_1}, \dots, \overrightarrow{v_n}\}$ and $\mathcal{C} = \{\overrightarrow{w_1}, \dots, \overrightarrow{w_n}\}$ be ordered bases for $\mathbb{F}^n$.

Then $[\vec{x}]_{\mathcal{C}} = {}_{\mathcal{C}}[I]_{\mathcal{B}}\ [\vec{x}]_{\mathcal{B}}$ and $[\vec{x}]_{\mathcal{B}} = {}_{\mathcal{B}}[I]_{\mathcal{C}}\ [\vec{x}]_{\mathcal{C}}$ for all $\vec{x} \in \mathbb{F}^n$.

Proof: Suppose that $[\vec{x}]_{\mathcal{B}} = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}$. Then $\vec{x} = a_1\,\overrightarrow{v_1} + \cdots + a_n\,\overrightarrow{v_n}$. As the coordinate vector map is linear by Theorem 8.8.8 (Linearity of Taking Coordinates), we have

$$\begin{aligned} [\vec{x}]_{\mathcal{C}} &= [a_1\,\overrightarrow{v_1} + \cdots + a_n\overrightarrow{v_n}]_{\mathcal{C}} \\ &= a_1[\overrightarrow{v_1}]_{\mathcal{C}} + \cdots + a_n[\overrightarrow{v_n}]_{\mathcal{C}} \\ &= \begin{bmatrix} [\overrightarrow{v_1}]_{\mathcal{C}} & [\overrightarrow{v_2}]_{\mathcal{C}} & \cdots & [\overrightarrow{v_n}]_{\mathcal{C}} \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} \\ &= {}_{\mathcal{C}}[I]_{\mathcal{B}}\ [\vec{x}]_{\mathcal{B}}. \end{aligned}$$

The proof going from $\mathcal{C}$ to $\mathcal{B}$ is identical, with $\mathcal{C}$ switched with $\mathcal{B}$. □

Corollary 8.8.15 Let $\vec{x} = [\vec{x}]_{\mathcal{E}} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$ be a vector in $\mathbb{F}^n$, where $\mathcal{E}$ is the standard basis for $\mathbb{F}^n$. If $\mathcal{C}$ is any ordered basis for $\mathbb{F}^n$, then

$$[\vec{x}]_{\mathcal{C}} = {}_{\mathcal{C}}[I]_{\mathcal{E}}\ \ [\vec{x}]_{\mathcal{E}}\,.$$

Proof: Take $\mathcal{B} = \mathcal{E}$ in Proposition 8.8.14 (Changing a Basis). □

The matrix ${}_{\mathcal{E}}[I]_{\mathcal{C}}$ is relatively easily obtained by simply inserting the standard coordinates of the vectors in $\mathcal{C}$ into the columns of a matrix. However, we usually want to go the other way (as in the above Corollary), and thus, we would like to have ${}_{\mathcal{C}}[I]_{\mathcal{E}}$.

The next result tells us that once we have one of these two matrices, then we can obtain the other rather quickly as they are inverses of each other.

Corollary 8.8.16 **(Inverse of Change-of-Basis Matrix)**

Let $\mathcal{B}$ and $\mathcal{C}$ be two ordered bases of $\mathbb{F}^n$. Then

$$ {}_{\mathcal{B}}[I]_{\mathcal{C}}\ {}_{\mathcal{C}}[I]_{\mathcal{B}} = I_n \qquad \text{and} \qquad {}_{\mathcal{C}}[I]_{\mathcal{B}}\ {}_{\mathcal{B}}[I]_{\mathcal{C}} = I_n. $$

In other words, ${}_{\mathcal{B}}[I]_{\mathcal{C}} = (\,{}_{\mathcal{C}}[I]_{\mathcal{B}}\,)^{-1}$.

Proof: Suppose we change coordinates twice. We start using basis $\mathcal{B}$, then change the basis to $\mathcal{C}$, and then back to $\mathcal{B}$.

We then have, for any $\vec{x} \in \mathbb{F}^n$, $[\vec{x}]_{\mathcal{C}} = {}_{\mathcal{C}}[I]_{\mathcal{B}}\ [\vec{x}]_{\mathcal{B}}$ and therefore,

$$ [\vec{x}]_{\mathcal{B}} = {}_{\mathcal{B}}[I]_{\mathcal{C}}\ [\vec{x}]_{\mathcal{C}} = {}_{\mathcal{B}}[I]_{\mathcal{C}}\ {}_{\mathcal{C}}[I]_{\mathcal{B}}\ [\vec{x}]_{\mathcal{B}}. $$

We can write this as

$$ \left(I_n - {}_{\mathcal{B}}[I]_{\mathcal{C}}\ {}_{\mathcal{C}}[I]_{\mathcal{B}}\right) [\vec{x}]_{\mathcal{B}} = \vec{0}, \qquad \text{for all } \vec{x} \in \mathbb{F}^n. $$

Thus, by Theorem 4.2.3 (Equality of Matrices), we conclude that

$$ (I_n - {}_{\mathcal{B}}[I]_{\mathcal{C}}\ {}_{\mathcal{C}}[I]_{\mathcal{B}}) = \mathcal{O}, $$

where $\mathcal{O}$ is the $n \times n$ zero matrix, and therefore ${}_{\mathcal{B}}[I]_{\mathcal{C}}\ {}_{\mathcal{C}}[I]_{\mathcal{B}} = I_n$. So ${}_{\mathcal{B}}[I]_{\mathcal{C}}$ and ${}_{\mathcal{C}}[I]_{\mathcal{B}}$ are inverses of each other. $\square$

Example 8.8.17 If we revisit Example 8.8.10 with this new notation and information, we find that

$$ {}_{\mathcal{E}}[I]_{\mathcal{B}} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & -2 \\ 1 & 3 & 4 \end{bmatrix}. $$

The inverse of this matrix is

$$ {}_{\mathcal{B}}[I]_{\mathcal{E}} = \frac{1}{3} \begin{bmatrix} 6 & -1 & -2 \\ -6 & 3 & 3 \\ 3 & -2 & -1 \end{bmatrix} $$

and so

$$ \begin{bmatrix} -3 \\ 2 \\ -6 \end{bmatrix}_{\mathcal{B}} = \frac{1}{3} \begin{bmatrix} 6 & -1 & -2 \\ -6 & 3 & 3 \\ 3 & -2 & -1 \end{bmatrix} \begin{bmatrix} -3 \\ 2 \\ -6 \end{bmatrix} = \frac{1}{3} \begin{bmatrix} -8 \\ 6 \\ -7 \end{bmatrix}, $$

just as we had computed.

Furthermore, if we now want, for example, $\begin{bmatrix} 3 \\ 5 \\ 9 \end{bmatrix}_{\mathcal{B}}$, then we can quickly find it:

$$\begin{bmatrix} 3 \\ 5 \\ 9 \end{bmatrix}_{\mathcal{B}} = \frac{1}{3}\begin{bmatrix} 6 & -1 & -2 \\ -6 & 3 & 3 \\ 3 & -2 & -1 \end{bmatrix}\begin{bmatrix} 3 \\ 5 \\ 9 \end{bmatrix} = \frac{1}{3}\begin{bmatrix} -5 \\ 24 \\ -10 \end{bmatrix}.$$

Example 8.8.18 Suppose we are in $\mathbb{C}^2$ and we wish to work with the ordered basis $\mathcal{C} = \{\vec{v_1}, \vec{v_2}\}$, where $\vec{v_1} = \begin{bmatrix} 1+2i \\ -1+i \end{bmatrix}$ and $\vec{v_2} = \begin{bmatrix} 1+i \\ i \end{bmatrix}$.

(a) Let $\vec{z} = \begin{bmatrix} 10+i \\ 2+6i \end{bmatrix}$. Find $[\vec{z}]_{\mathcal{C}}$.

(b) Consider the vector $\vec{w} \in \mathbb{C}^2$ with $[\vec{w}]_{\mathcal{C}} = \begin{bmatrix} 2 \\ 3i \end{bmatrix}$. Find $[\vec{w}]_{\mathcal{E}}$.

Solution:

Note that, as usual, all the vectors have been given with respect to the standard basis.

We have

$$_{\mathcal{E}}[I]_{\mathcal{C}} = \begin{bmatrix} 1+2i & 1+i \\ -1+i & i \end{bmatrix}$$

and therefore, we can compute

$$_{\mathcal{C}}[I]_{\mathcal{E}} = \begin{bmatrix} 1+2i & 1+i \\ -1+i & i \end{bmatrix}^{-1} = \frac{1}{i}\begin{bmatrix} i & -1-i \\ 1-i & 1+2i \end{bmatrix} = -i\begin{bmatrix} i & -1-i \\ 1-i & 1+2i \end{bmatrix} = \begin{bmatrix} 1 & -1+i \\ -1-i & 2-i \end{bmatrix}.$$

(a) We can obtain $[\vec{z}]_{\mathcal{C}}$ by

$$[\vec{z}]_{\mathcal{C}} = {}_{\mathcal{C}}[I]_{\mathcal{E}}\,[\vec{z}]_{\mathcal{E}} = \begin{bmatrix} 1 & -1+i \\ -1-i & 2-i \end{bmatrix}\begin{bmatrix} 10+i \\ 2+6i \end{bmatrix} = \begin{bmatrix} 2-3i \\ 1-i \end{bmatrix}.$$

(b) We can obtain $[\vec{w}]_{\mathcal{E}}$ by

$$[\vec{w}]_{\mathcal{E}} = {}_{\mathcal{E}}[I]_{\mathcal{C}}\,[\vec{w}]_{\mathcal{C}} = \begin{bmatrix} 1+2i & 1+i \\ -1+i & i \end{bmatrix}\begin{bmatrix} 2 \\ 3i \end{bmatrix} = \begin{bmatrix} -1+7i \\ -5+2i \end{bmatrix}.$$

Chapter 9

Diagonalization

9.1 Matrix Representation of a Linear Operator

In this chapter we will focus on linear transformations $T\colon \mathbb{F}^n \to \mathbb{F}^n$ whose domain and codomain are the same set. Among other notable properties, these transformations will have the ability to be composed with themselves; that is, functions such as $T \circ T$ will be well-defined when the domain and codomain of T are the same.

Definition 9.1.1
Linear Operator

A linear transformation $T\colon \mathbb{F}^n \to \mathbb{F}^m$ where $n = m$ is called a **linear operator**.

In Section 5.5, we learned how to find the standard matrix $[T]_{\mathcal{E}}$ of a linear operator T. In this section we'll learn how to find matrix representations of a linear operator T with respect to bases for $\mathbb{F}^n$ other than the standard basis.

Definition 9.1.2
$\mathcal{B}$-Matrix of T

Let $T\colon \mathbb{F}^n \to \mathbb{F}^n$ be a linear operator and let $\mathcal{B} = \{\vec{v_1}, \vec{v_2}, \dots, \vec{v_n}\}$ be an ordered basis for $\mathbb{F}^n$. We define the **$\mathcal{B}$-matrix of** T to be the matrix $[T]_{\mathcal{B}}$ constructed as follows:

$$[T]_{\mathcal{B}} = \Big[[T(\vec{v_1})]_{\mathcal{B}} \quad [T(\vec{v_2})]_{\mathcal{B}} \cdots [T(\vec{v_n})]_{\mathcal{B}}\Big].$$

That is, after applying the action of T to each member of $\mathcal{B}$, we take the $\mathcal{B}$-coordinate vectors of each of these images to create the columns of $[T]_{\mathcal{B}}$.

Similar to the way we used $[T]_{\mathcal{E}}$ to find the image of a vector in $\mathbb{F}^n$, we can use $[T]_{\mathcal{B}}$ to find the coordinate vector with respect to $\mathcal{B}$ of the image of any vector in $\mathbb{F}^n$ by performing matrix–vector multiplication.

Proposition 9.1.3

Let $T\colon \mathbb{F}^n \to \mathbb{F}^n$ be a linear operator and let $\mathcal{B} = \{\vec{v_1}, \vec{v_2}, \dots, \vec{v_n}\}$ be an ordered basis for $\mathbb{F}^n$. If $\vec{v} \in \mathbb{F}^n$, then

$$[T(\vec{v})]_{\mathcal{B}} = [T]_{\mathcal{B}}\ [\vec{v}]_{\mathcal{B}}.$$

Proof: Since $\mathcal{B}$ is a basis for $\mathbb{F}^n$ and $\vec{v} \in \mathbb{F}^n$, then there exists $c_1, c_2, \ldots, c_n \in \mathbb{F}$ such that

$$\vec{v} = c_1\vec{v_1} + c_2\vec{v_2} + \cdots + c_n\vec{v_n}.$$

We can use the linearity of T to expand $T(\vec{v})$. We get

$$T(\vec{v}) = T(c_1\vec{v_1} + c_2\vec{v_2} + \cdots + c_n\vec{v_n}) = c_1T(\vec{v_1}) + c_2T(\vec{v_2}) + \cdots + c_nT(\vec{v_n}).$$

Taking coordinates is a linear operation, so

$$\begin{aligned}
[T(\vec{v})]_{\mathcal{B}} &= [c_1\,T(\vec{v_1}) + c_2\,T(\vec{v_2}) + \cdots + c_n\,T(\vec{v_n})]_{\mathcal{B}} \\
&= c_1\,[T(\vec{v_1})]_{\mathcal{B}} + c_2\,[T(\vec{v_2})]_{\mathcal{B}} + \cdots + c_n\,[T(\vec{v_n})]_{\mathcal{B}} \\
&= \begin{bmatrix} [T(\vec{v_1})]_{\mathcal{B}} & [T(\vec{v_2})]_{\mathcal{B}} \cdots & [T(\vec{v_n})]_{\mathcal{B}} \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix} \\
&= [T]_{\mathcal{B}}\,[\vec{v}]_{\mathcal{B}},
\end{aligned}$$

as required. □

There are two important reasons why we may want to use another basis. The first is that we may have some geometrically or physically preferred vectors that naturally arise in our problem; for example, the direction vector of a line through which we are reflecting vectors. The second is that we may wish to simplify the matrix representation; for example, it would simplify things if $[T]_{\mathcal{B}}$ were diagonal. These two reasons are often connected.

We will now revisit some of the linear transformations that we considered in Section 5.6. For each transformation, we will choose a basis $\mathcal{B}$ such that the $\mathcal{B}$-matrix will be particularly simple: it will be diagonal.

Example 9.1.4 Let $\vec{w} = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$ be a non-zero vector in $\mathbb{R}^2$. Consider the projection transformation $\text{proj}_{\vec{w}} \colon \mathbb{R}^2 \to \mathbb{R}^2$. Find a basis $\mathcal{B}$ such that $[\text{proj}_{\vec{w}}]_{\mathcal{B}}$ is diagonal.

Solution:

Consider some vectors for which it is relatively easy to determine their images under $\text{proj}_{\vec{w}}$. The projection of $\vec{w}$ onto itself is itself. Therefore, $\text{proj}_{\vec{w}}(\vec{w}) = \vec{w}$. The vector $\begin{bmatrix} w_2 \\ -w_1 \end{bmatrix}$ is orthogonal to $\vec{w}$. Therefore, $\text{proj}_{\vec{w}}\left(\begin{bmatrix} w_2 \\ -w_1 \end{bmatrix}\right) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$.

Let $\mathcal{B} = \left\{ \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}, \begin{bmatrix} w_2 \\ -w_1 \end{bmatrix} \right\}$. This set consists of two vectors that are not scalar multiples of each other, and therefore it is linearly independent. So $\mathcal{B}$ is a basis for $\mathbb{R}^2$.

We have that

$$\text{proj}_{\vec{w}}\left(\begin{bmatrix} w_1 \\ w_2 \end{bmatrix}\right) = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = 1\begin{bmatrix} w_1 \\ w_2 \end{bmatrix} + 0\begin{bmatrix} w_2 \\ -w_1 \end{bmatrix}$$

and

$$\text{proj}_{\vec{w}}\left(\begin{bmatrix} w_2 \\ -w_1 \end{bmatrix}\right) = \begin{bmatrix} 0 \\ 0 \end{bmatrix} = 0\begin{bmatrix} w_1 \\ w_2 \end{bmatrix} + 0\begin{bmatrix} w_2 \\ -w_1. \end{bmatrix}$$

Consequently,

$$[\mathrm{proj}_{\vec{w}}]_{\mathcal{B}} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}.$$

Notice that this is a diagonal matrix.

Example 9.1.5

Let $\vec{w} = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$ be a non-zero vector in $\mathbb{R}^2$. Consider the reflection transformation $\mathrm{refl}_{\vec{w}} \colon \mathbb{R}^2 \to \mathbb{R}^2$, which reflects any vector in $\mathbb{R}^2$ about the line $\mathrm{Span}\{\vec{w}\}$. Find a basis $\mathcal{B}$ such that $[\mathrm{refl}_{\vec{w}}]_{\mathcal{B}}$ is diagonal.

Solution: Consider some vectors for which it is relatively easy to determine their images under $\mathrm{refl}_{\vec{w}}$. The reflection of $\vec{w}$ about itself is itself. Therefore, $\mathrm{refl}_{\vec{w}}(\vec{w}) = \vec{w}$. The vector $\begin{bmatrix} w_2 \\ -w_1 \end{bmatrix}$ is orthogonal to $\vec{w}$ and so its reflection about $\vec{w}$ is its additive inverse. Therefore, $\mathrm{refl}_{\vec{w}}\left(\begin{bmatrix} w_2 \\ -w_1 \end{bmatrix}\right) = \begin{bmatrix} -w_2 \\ w_1 \end{bmatrix}$.

Let $\mathcal{B} = \left\{\begin{bmatrix} w_1 \\ w_2 \end{bmatrix}, \begin{bmatrix} w_2 \\ -w_1 \end{bmatrix}\right\}$. This set consists of two vectors that are not scalar multiples of each other, and therefore it is linearly independent. So $\mathcal{B}$ is a basis for $\mathbb{R}^2$.

We have that

$$\mathrm{refl}_{\vec{w}}\left(\begin{bmatrix} w_1 \\ w_2 \end{bmatrix}\right) = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = 1\begin{bmatrix} w_1 \\ w_2 \end{bmatrix} + 0\begin{bmatrix} w_2 \\ -w_1 \end{bmatrix}$$

and

$$\mathrm{refl}_{\vec{w}}\left(\begin{bmatrix} w_2 \\ -w_1 \end{bmatrix}\right) = \begin{bmatrix} -w_2 \\ w_1 \end{bmatrix} = 0\begin{bmatrix} w_1 \\ w_2 \end{bmatrix} + (-1)\begin{bmatrix} w_2 \\ -w_1. \end{bmatrix}$$

Consequently, we get the following diagonal matrix for $[\mathrm{refl}_{\vec{w}}]_{\mathcal{B}}$:

$$[\mathrm{refl}_{\vec{w}}]_{\mathcal{B}} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

EXERCISE

In Examples 9.1.4 and 9.1.5, what happens to the matrices $[\mathrm{proj}_{\vec{w}}]_{\mathcal{B}}$ and $[\mathrm{refl}_{\vec{w}}]_{\mathcal{B}}$ if we replace $\mathcal{B} = \left\{\begin{bmatrix} w_1 \\ w_2 \end{bmatrix}, \begin{bmatrix} w_2 \\ -w_1 \end{bmatrix}\right\}$ with the ordered basis $\left\{\begin{bmatrix} w_2 \\ -w_1 \end{bmatrix}, \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}\right\}$?

REMARK

Finding a basis $\mathcal{B}$ such that $[T]_{\mathcal{B}}$ is diagonal is not usually done by inspection. Moreover, it is not always possible to choose a basis $\mathcal{B}$ such that $[T]_{\mathcal{B}}$ is diagonal. For example, the operator known as the "shear operator" $T \colon \mathbb{R}^2 \to \mathbb{R}^2$, defined by $T\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} x+y \\ y \end{bmatrix}$, is an example of a linear operator for which no basis $\mathcal{B}$ exists such that $[T]_{\mathcal{B}}$ is diagonal. Throughout this chapter we will develop criteria to determine whether or not we will be able to find such a basis.

Given a linear operator $T\colon \mathbb{F}^n \to \mathbb{F}^n$, we can create many matrices $[T]_{\mathcal{B}}$ by choosing different bases $\mathcal{B}$ for $\mathbb{F}^n$. A natural question to ask is: How are these various matrices related? The answer is that they are *similar* (see Definition 7.6.2).

Proposition 9.1.6 **(Similarity of Matrix Representations)**

Let $T\colon \mathbb{F}^n \to \mathbb{F}^n$ be a linear operator. Let $\mathcal{B}$ and $\mathcal{C}$ be ordered bases for $\mathbb{F}^n$. Then

$$[T]_{\mathcal{C}} = {}_{\mathcal{C}}[I]_{\mathcal{B}}\ [T]_{\mathcal{B}}\ {}_{\mathcal{B}}[I]_{\mathcal{C}} = \left({}_{\mathcal{B}}[I]_{\mathcal{C}}\right)^{-1}\ [T]_{\mathcal{B}}\ {}_{\mathcal{B}}[I]_{\mathcal{C}}$$

and

$$[T]_{\mathcal{B}} = {}_{\mathcal{B}}[I]_{\mathcal{C}}\ [T]_{\mathcal{C}}\ {}_{\mathcal{C}}[I]_{\mathcal{B}} = \left({}_{\mathcal{C}}[I]_{\mathcal{B}}\right)^{-1}\ [T]_{\mathcal{C}}\ {}_{\mathcal{C}}[I]_{\mathcal{B}}.$$

That is, the matrices $[T]_{\mathcal{B}}$ and $[T]_{\mathcal{C}}$ are similar over $\mathbb{F}$.

Proof: Let $\mathcal{B} = \{\overrightarrow{v_1}, \overrightarrow{v_2}, \ldots, \overrightarrow{v_n}\}$ and let $\mathcal{C} = \{\overrightarrow{w_1}, \overrightarrow{w_2}, \ldots, \overrightarrow{w_n}\}$. Then using the definition of $[T]_{\mathcal{C}}$ and Proposition 8.8.14 (Changing a Basis) gives

$$\begin{aligned}[T]_{\mathcal{C}} &= \Big[[T(\overrightarrow{w_1})]_{\mathcal{C}} \quad [T(\overrightarrow{w_2})]_{\mathcal{C}} \cdots [T(\overrightarrow{w_n})]_{\mathcal{C}}\Big] \\ &= \Big[{}_{\mathcal{C}}[I]_{\mathcal{B}}\ [T(\overrightarrow{w_1})]_{\mathcal{B}} \quad {}_{\mathcal{C}}[I]_{\mathcal{B}}\ [T(\overrightarrow{w_2})]_{\mathcal{B}} \ \cdots\ {}_{\mathcal{C}}[I]_{\mathcal{B}}\ [T(\overrightarrow{w_n})]_{\mathcal{B}}\Big].\end{aligned}$$

Using the definition of matrix multiplication, we can write this matrix as the product

$$[T]_{\mathcal{C}} = {}_{\mathcal{C}}[I]_{\mathcal{B}}\Big[[T(\overrightarrow{w_1})]_{\mathcal{B}} \quad [T(\overrightarrow{w_2})]_{\mathcal{B}} \ \cdots\ [T(\overrightarrow{w_n})]_{\mathcal{B}}\Big].$$

Using Proposition 9.1.3, we get

$$[T]_{\mathcal{C}} = {}_{\mathcal{C}}[I]_{\mathcal{B}}\Big[[T]_{\mathcal{B}}[\overrightarrow{w_1}]_{\mathcal{B}} \quad [T]_{\mathcal{B}}[\overrightarrow{w_2}]_{\mathcal{B}} \ \cdots\ [T]_{\mathcal{B}}[\overrightarrow{w_n}]_{\mathcal{B}}\Big].$$

Again using the definition of matrix multiplication, we obtain

$$[T]_{\mathcal{C}} = {}_{\mathcal{C}}[I]_{\mathcal{B}}\ [T]_{\mathcal{B}}\ \Big[[\overrightarrow{w_1}]_{\mathcal{B}} \quad [\overrightarrow{w_2}]_{\mathcal{B}} \cdots [\overrightarrow{w_n}]_{\mathcal{B}}\Big].$$

By definition,

$${}_{\mathcal{B}}[I]_{\mathcal{C}} = \Big[[\overrightarrow{w_1}]_{\mathcal{B}} \quad [\overrightarrow{w_2}]_{\mathcal{B}} \cdots [\overrightarrow{w_n}]_{\mathcal{B}}\Big].$$

Since

$${}_{\mathcal{C}}[I]_{\mathcal{B}} = \left({}_{\mathcal{B}}[I]_{\mathcal{C}}\right)^{-1},$$

so we obtain that

$$[T]_{\mathcal{C}} = \left({}_{\mathcal{B}}[I]_{\mathcal{C}}\right)^{-1}\ [T]_{\mathcal{B}}\ {}_{\mathcal{B}}[I]_{\mathcal{C}}.$$

By swapping the roles played by $\mathcal{B}$ and $\mathcal{C}$ in the above argument, we obtain that

$$[T]_{\mathcal{B}} = {}_{\mathcal{B}}[I]_{\mathcal{C}}\ [T]_{\mathcal{C}}\ {}_{\mathcal{C}}[I]_{\mathcal{B}} = \left({}_{\mathcal{C}}[I]_{\mathcal{B}}\ \right)^{-1}\ [T]_{\mathcal{C}}\ {}_{\mathcal{C}}[I]_{\mathcal{B}}.$$

$\square$

Corollary 9.1.7 **(Finding the Standard Matrix)**

Let $T\colon \mathbb{F}^n \to \mathbb{F}^n$ be a linear operator. Let $\mathcal{B}$ be a basis for $\mathbb{F}^n$ and let $\mathcal{E}$ be the standard basis for $\mathbb{F}^n$. Then

$$[T]_{\mathcal{E}} = {}_{\mathcal{E}}[I]_{\mathcal{B}}\ [T]_{\mathcal{B}}\ {}_{\mathcal{B}}[I]_{\mathcal{E}} = \left({}_{\mathcal{B}}[I]_{\mathcal{E}}\right)^{-1}\ [T]_{\mathcal{B}}\ {}_{\mathcal{B}}[I]_{\mathcal{E}}$$

and

$$[T]_{\mathcal{B}} = {}_{\mathcal{B}}[I]_{\mathcal{E}}\ [T]_{\mathcal{E}}\ {}_{\mathcal{E}}[I]_{\mathcal{B}} = \left({}_{\mathcal{E}}[I]_{\mathcal{B}}\right)^{-1}\ [T]_{\mathcal{E}}\ {}_{\mathcal{E}}[I]_{\mathcal{B}}.$$

Proof: This follows from the previous Proposition, with $\mathcal{C}$ replaced by $\mathcal{E}$. □

EXERCISE

In this exercise you will show that if A and B are similar matrices in $M_{n\times n}(\mathbb{F})$, then they are different representations of the same linear operator.

Let $A, B \in M_{n\times n}(\mathbb{F})$. Suppose that A is similar to B, say $A = PBP^{-1}$ for some invertible matrix $P \in M_{n\times n}(\mathbb{F})$. Let $\mathcal{B}$ be the ordered basis for $\mathbb{F}^n$ consisting of the columns of P (ordered in the way they appear in P). Show that $[T_A]_{\mathcal{E}} = A$ and $[T_A]_{\mathcal{B}} = B$.

REMARK

The previous Exercise justifies the use of the word "similar" to describe the relationship in Definition 7.6.2. If matrices $A, B \in M_{n\times n}(\mathbb{F})$ are similar over $\mathbb{F}$, then they are both $\mathcal{B}$-matrices of the same linear operator $T\colon \mathbb{F}^n \to \mathbb{F}^n$. That is, similar matrices are just different representations of the same operator.

Example 9.1.8 Let $\vec{w} = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$ be a non-zero vector in $\mathbb{R}^2$. Consider the projection transformation $\operatorname{proj}_{\vec{w}}\colon \mathbb{R}^2 \to \mathbb{R}^2$. Determine $[\operatorname{proj}_{\vec{w}}]_{\mathcal{E}}$ using the solution to Example 9.1.4 and Corollary 9.1.7 (Finding the Standard Matrix).

Solution: In Example 9.1.4, we found the matrix representation $[\operatorname{proj}_{\vec{w}}]_{\mathcal{B}} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ with respect to the basis $\mathcal{B} = \left\{ \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}, \begin{bmatrix} w_2 \\ -w_1 \end{bmatrix} \right\}$.

We also have that

$${}_{\mathcal{E}}[I]_{\mathcal{B}} = \begin{bmatrix} w_1 & w_2 \\ w_2 & -w_1 \end{bmatrix}$$

and

$${}_{\mathcal{B}}[I]_{\mathcal{E}} = \left({}_{\mathcal{E}}[I]_{\mathcal{B}}\right)^{-1} = \frac{1}{-w_1^2 - w_2^2} \begin{bmatrix} -w_1 & -w_2 \\ -w_2 & w_1 \end{bmatrix} = \frac{1}{w_1^2 + w_2^2} \begin{bmatrix} w_1 & w_2 \\ w_2 & -w_1 \end{bmatrix}.$$

Therefore, using Corollary 9.1.7 (Finding the Standard Matrix), we have

$$\begin{aligned}
[\text{proj}_{\vec{w}}]_{\mathcal{E}} &= {}_{\mathcal{E}}[I]_{\mathcal{B}}\,[\text{proj}_{\vec{w}}]_{\mathcal{B}}\ {}_{\mathcal{B}}[I]_{\mathcal{E}}\\
&= \begin{bmatrix} w_1 & w_2 \\ w_2 & -w_1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \frac{1}{w_1^2+w_2^2} \begin{bmatrix} w_1 & w_2 \\ w_2 & -w_1 \end{bmatrix}\\
&= \frac{1}{w_1^2+w_2^2} \begin{bmatrix} w_1 & 0 \\ w_2 & 0 \end{bmatrix} \begin{bmatrix} w_1 & w_2 \\ w_2 & -w_1 \end{bmatrix}\\
&= \frac{1}{w_1^2+w_2^2} \begin{bmatrix} w_1^2 & w_1w_2 \\ w_1w_2 & w_2^2 \end{bmatrix},
\end{aligned}$$

which matches what we found in Example 5.6.1.

Example 9.1.9 Let $\vec{w} = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$ be a non-zero vector in $\mathbb{R}^2$. Consider the reflection transformation $\text{refl}_{\vec{w}} \colon \mathbb{R}^2 \to \mathbb{R}^2$, which reflects any vector $\vec{v} \in \mathbb{R}^2$ about the line $\text{Span}\{\vec{w}\}$. Determine $[\text{refl}_{\vec{w}}]_{\mathcal{E}}$ using the solution to Example 9.1.5 and Corollary 9.1.7 (Finding the Standard Matrix).

Solution: In Example 9.1.5, we found the matrix representation $[\text{refl}_{\vec{w}}]_{\mathcal{B}} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$ with respect to the basis $\mathcal{B} = \left\{ \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}, \begin{bmatrix} w_2 \\ -w_1 \end{bmatrix} \right\}$.

We also have that

$${}_{\mathcal{E}}[I]_{\mathcal{B}} = \begin{bmatrix} w_1 & w_2 \\ w_2 & -w_1 \end{bmatrix}$$

and

$${}_{\mathcal{B}}[I]_{\mathcal{E}} = \left({}_{\mathcal{E}}[I]_{\mathcal{B}}\right)^{-1} = \frac{1}{-w_1^2-w_2^2} \begin{bmatrix} -w_1 & -w_2 \\ -w_2 & w_1 \end{bmatrix} = \frac{1}{w_1^2+w_2^2} \begin{bmatrix} w_1 & w_2 \\ w_2 & -w_1 \end{bmatrix}.$$

Therefore, using Corollary 9.1.7 (Finding the Standard Matrix), we have

$$\begin{aligned}
[\text{refl}_{\vec{w}}]_{\mathcal{E}} &= {}_{\mathcal{E}}[I]_{\mathcal{B}}\,[\text{refl}_{\vec{w}}]_{\mathcal{B}}\ {}_{\mathcal{B}}[I]_{\mathcal{E}}\\
&= \begin{bmatrix} w_1 & w_2 \\ w_2 & -w_1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \frac{1}{w_1^2+w_2^2} \begin{bmatrix} w_1 & w_2 \\ w_2 & -w_1 \end{bmatrix}\\
&= \frac{1}{w_1^2+w_2^2} \begin{bmatrix} w_1 & -w_2 \\ w_2 & w_1 \end{bmatrix} \begin{bmatrix} w_1 & w_2 \\ w_2 & -w_1 \end{bmatrix}\\
&= \frac{1}{w_1^2+w_2^2} \begin{bmatrix} w_1^2-w_2^2 & 2w_1w_2 \\ 2w_1w_2 & w_2^2-w_1^2 \end{bmatrix},
\end{aligned}$$

which matches what we found in Example 5.6.4.

9.2 Diagonalizability of Linear Operators

Of all the $\mathcal{B}$-matrices of a fixed linear operator $T \colon \mathbb{F}^n \to \mathbb{F}^n$, some might be simpler than others. For instance, in Examples 9.1.4 and 9.1.5, we were able to find an ordered basis $\mathcal{B}$ for $\mathbb{F}^n$ such that the $\mathcal{B}$-matrix of the given linear operator T was especially simple: it was a diagonal matrix. This can allow us to perform computations with or make certain observations about the nature of T in a more efficient manner.

Example 9.2.1

Let $\vec{w} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \in \mathbb{R}^2$ and consider the linear operator $T = \operatorname{refl}_{\vec{w}}$. Then:

1. Using the standard basis $\mathcal{E}$, we have $[T]_{\mathcal{E}} = \dfrac{1}{5}\begin{bmatrix} -3 & 4 \\ 4 & 3 \end{bmatrix}$.
2. Using the basis $\mathcal{B} = \left\{ \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 2 \\ -1 \end{bmatrix} \right\}$ from Example 9.1.5, we have $[T]_{\mathcal{B}} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$.
3. Using the basis $\mathcal{C} = \left\{ \begin{bmatrix} 1 \\ -2 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \end{bmatrix} \right\}$, it can be shown that $[T]_{\mathcal{C}} = \dfrac{-1}{25}\begin{bmatrix} 7 & 24 \\ 24 & -25 \end{bmatrix}$.

We can see immediately that the matrix $[T]_{\mathcal{B}}$ is invertible. This is also true of the $\mathcal{E}$- and $\mathcal{C}$-matrix representations, but it is not as obvious.

Moreover, we can also see at once that $([T]_{\mathcal{B}})^2 = I_2$. This is a manifestation of the geometric fact that $\operatorname{refl}_{\vec{w}} \circ \operatorname{refl}_{\vec{w}}$ is the identity transformation. You can check that $([T]_{\mathcal{E}})^2$ and $([T]_{\mathcal{C}})^2$ are also both equal to the 2×2 identity matrix, but this requires a bit of a computation.

This example demonstrates some of the ways in which a diagonal matrix representation, such as $[T]_{\mathcal{B}}$, can be superior to other matrix representations.

Definition 9.2.2
Diagonalizable

Let $T \colon \mathbb{F}^n \to \mathbb{F}^n$ be a linear operator. We say that T is **diagonalizable over** $\mathbb{F}$ if there exists an ordered basis $\mathcal{B}$ for $\mathbb{F}^n$ such that $[T]_{\mathcal{B}}$ is a diagonal matrix.

Like matrices, not all linear operators are diagonalizable. We'll see this later as we explore the connection between the diagonalizability of matrices and linear operators.

Example 9.2.3

Let $\vec{w} \in \mathbb{R}^2$ be a non-zero vector. Then Example 9.1.4 shows that $T = \operatorname{proj}_{\vec{w}}$ is a diagonalizable operator and Example 9.1.5 shows that $T = \operatorname{refl}_{\vec{w}}$ is a diagonalizable operator.

We were able to accomplish this because we could choose basis vectors $\vec{v_1}$ and $\vec{v_2}$ such that $T(\vec{v_i})$ is a scalar multiple of $\vec{v_i}$.

The observation at the end of the preceding Example is the key to diagonalizability. It also prompts the following (hopefully familiar!) definition.

Definition 9.2.4
Eigenvector, Eigenvalue and Eigenpair of a Linear Operator

Let $T \colon \mathbb{F}^n \to \mathbb{F}^n$ be a linear operator. We say that the *non-zero* vector $\vec{x} \in \mathbb{F}^n$ is an **eigenvector** of T if there exists a scalar $\lambda \in \mathbb{F}$ such that

$$T(\vec{x}) = \lambda \vec{x}.$$

The scalar λ is called an **eigenvalue of T over** $\mathbb{F}$ and the pair $(\lambda, \vec{x})$ is called an **eigenpair of T over** $\mathbb{F}$.

This should remind you of the analogous definitions for matrices (Definition 7.1.5). These definitions are connected as follows.

Proposition 9.2.5 **(Eigenpairs of T and $[T]_{\mathcal{B}}$)**

Let $T\colon \mathbb{F}^n \to \mathbb{F}^n$ be a linear operator and let $\mathcal{B}$ be an ordered basis for $\mathbb{F}^n$. Then $(\lambda, \vec{x})$ is an eigenpair of T if and only if $(\lambda, [\vec{x}]_{\mathcal{B}})$ is an eigenpair of the matrix $[T]_{\mathcal{B}}$.

Proof: Take the coordinates of both sides with respect to $\mathcal{B}$ of the eigenvalue problem and use the fact that taking coordinates is a linear operation:

$$\begin{aligned} T(\vec{x}) = \lambda\vec{x} &\Leftrightarrow [T(\vec{x})]_{\mathcal{B}} = [\lambda\vec{x}]_{\mathcal{B}} \\ &\Leftrightarrow [T]_{\mathcal{B}}\,[\vec{x}]_{\mathcal{B}} = \lambda\,[\vec{x}]_{\mathcal{B}} \end{aligned}$$

□

EXERCISE

Let $\vec{w} = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$ be a non-zero vector in $\mathbb{R}^2$. Consider the projection transformation $\text{proj}_{\vec{w}}\colon \mathbb{R}^2 \to \mathbb{R}^2$. Find the eigenvalues of $\text{proj}_{\vec{w}}$ and, for each eigenvalue, find a corresponding eigenvector.

[**Hint:** The hard way is to use the standard basis and $[\text{proj}_{\vec{w}}]_{\mathcal{E}}$ which we computed in Example 5.6.1. The easy way is to use the basis $\mathcal{B}$ from Example 9.1.4.]

Now we can give a criterion for the diagonalizability of a linear operator T.

Proposition 9.2.6 **(Eigenvector Basis Criterion for Diagonalizability)**

Let $T : \mathbb{F}^n \to \mathbb{F}^n$ be a linear operator. Then T is diagonalizable over $\mathbb{F}$ if and only if there exists an ordered basis $\mathcal{B} = \{\vec{v_1}, \vec{v_2}, \ldots, \vec{v_n}\}$ for $\mathbb{F}^n$ consisting of eigenvectors of T.

Proof: Begin with the forward direction. Assume that T is diagonalizable over $\mathbb{F}$. Then there exists an ordered basis $\mathcal{B} = \{\vec{v_1}, \vec{v_2}, \ldots, \vec{v_n}\}$ of $\mathbb{F}^n$ such that

$$[T]_{\mathcal{B}} = \text{diag}(d_1, d_2, \ldots, d_n).$$

Therefore,

$$\begin{aligned} [T(\vec{v_i})]_{\mathcal{B}} &= [T]_{\mathcal{B}}\,[\vec{v_i}]_{\mathcal{B}} && \text{(by Proposition 9.1.3)} \\ &= \text{diag}(d_1, \ldots, d_n)\vec{e_i} \\ &= \text{the } i^{th} \text{ column of } \text{diag}(d_1, \ldots, d_n) && \text{(by Lemma 4.2.2 (Column Extraction))} \\ &= d_i\vec{e_i} \\ &= d_i\,[\vec{v_i}]_{\mathcal{B}}\,. \end{aligned}$$

Therefore, for each $i = 1, 2, \ldots, n$, $\vec{v_i}$ is an eigenvector of T with eigenvalue d_i.

For the backward direction, assume that there exists an ordered basis $\mathcal{B} = \{\vec{v_1}, \vec{v_2}, \ldots, \vec{v_n}\}$ for $\mathbb{F}^n$ consisting of eigenvectors of T. Then, for all $\vec{v_i} \in \mathcal{B}$, $T(\vec{v_i}) = \lambda_i\vec{v_i}$ for some $\lambda_i \in \mathbb{F}$. So $[T]_{\mathcal{B}} = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$, a diagonal matrix. Thus, T is diagonalizable over $\mathbb{F}$. □

Example 9.2.7 Let $\vec{w} = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$ be a non-zero vector in $\mathbb{R}^2$. In Example 9.1.4, we essentially showed that $\mathcal{B} = \left\{ \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}, \begin{bmatrix} w_2 \\ -w_1 \end{bmatrix} \right\}$ is a basis for $\mathbb{R}^2$ consisting of eigenvectors of $\text{proj}_{\vec{w}}$. So $\text{proj}_{\vec{w}}$ is diagonalizable over $\mathbb{R}$ by Proposition 9.2.6.

Moreover, we showed that the eigenvalues corresponding to $\begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$ and $\begin{bmatrix} w_2 \\ -w_1 \end{bmatrix}$ are 1 and 0, respectively, so the proof of Proposition 9.2.6 asserts that $[\text{proj}\ \vec{w}]_{\mathcal{B}} = \text{diag}(1, 0)$. This is precisely what we had found in Example 9.1.4.

EXERCISE

Apply the same analysis to $\text{refl}_{\vec{w}}$.

The previous Example and Exercise make use of eigenpairs that were determined through a geometric argument. Not all transformations will have intuitive geometric interpretations, so this strategy won't always be practical. We will tackle this problem in the next two sections.

9.3 Diagonalizability of Matrices Revisited

In this section we will describe a strategy for determining when a linear operator is diagonalizable while simultaneously connecting this notion of diagonalizability to the notion of diagonalizability of matrices that we learned about in Section 7.6.

Proposition 9.3.1 **(T Diagonalizable iff $[T]_{\mathcal{B}}$ Diagonalizable)**

Let $T\colon \mathbb{F}^n \to \mathbb{F}^n$ be a linear operator and let $\mathcal{B}$ be an ordered basis of $\mathbb{F}^n$. Then T is diagonalizable over $\mathbb{F}$ if and only if the matrix $[T]_{\mathcal{B}}$ is diagonalizable over $\mathbb{F}$.

Proof: ($\Rightarrow$):
Assume that T is diagonalizable over $\mathbb{F}$. Then, by Proposition 9.2.6 (Eigenvector Basis Criterion for Diagonalizability), there exists an ordered basis $\mathcal{C}$ of $\mathbb{F}^n$ consisting of eigenvectors of T. In the proof of Proposition 9.2.6, we saw that $[T]_{\mathcal{C}} = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$, where the λ_i's are the eigenvalues of T.

By Corollary 9.1.7 (Finding the Standard Matrix),

$$[T]_{\mathcal{C}} = {}_{\mathcal{C}}[I]_{\mathcal{B}}\ [T]_{\mathcal{B}}\ {}_{\mathcal{B}}[I]_{\mathcal{C}} = \left({}_{\mathcal{B}}[I]_{\mathcal{C}}\right)^{-1}\ [T]_{\mathcal{B}}\ {}_{\mathcal{B}}[I]_{\mathcal{C}}.$$

Since $[T]_{\mathcal{C}}$ is diagonal, then $[T]_{\mathcal{B}}$ is diagonalizable over $\mathbb{F}$ and the matrix $P = {}_{\mathcal{B}}[I]_{\mathcal{C}}$ diagonalizes $[T]_{\mathcal{B}}$.

($\Leftarrow$):
Assume that $[T]_{\mathcal{B}}$ is diagonalizable over $\mathbb{F}$. Then there exists an invertible matrix $P = \begin{bmatrix} \vec{p_1} & \vec{p_2} & \cdots & \vec{p_n} \end{bmatrix}$ such that

$$P^{-1}[T]_{\mathcal{B}}P = D = \operatorname{diag}(d_1, d_2, \ldots, d_n).$$

We define vectors $\vec{v_1}, \vec{v_2}, \ldots, \vec{v_n}$ such that for $i = 1, \ldots, n$, $[\vec{v_i}]_{\mathcal{B}} = \vec{p}_i$, the i^{th} column of P. We see this is possible by defining $\vec{v_i} = {}_{\mathcal{E}}[I]_{\mathcal{B}}\,\vec{p_i}$, so that

$$\vec{p_i} = ({}_{\mathcal{E}}[I]_{\mathcal{B}})^{-1}\ \vec{v_i} = {}_{\mathcal{B}}[I]_{\mathcal{E}}\,\vec{v_i} = [\vec{v_i}]_{\mathcal{B}}.$$

Therefore,

$$P = \Big[\, [\vec{v_1}]_{\mathcal{B}}\ [\vec{v_2}]_{\mathcal{B}}\ \cdots\ [\vec{v_n}]_{\mathcal{B}} \,\Big].$$

We will show that the set $\mathcal{S} = \{\vec{v_1}, \vec{v_2}, \ldots, \vec{v_n}\}$ is a basis for $\mathbb{F}^n$. By Proposition 8.5.4 (n Vectors in $\mathbb{F}^n$ Span iff Independent), since it contains n vectors in $\mathbb{F}^n$, then it will be enough to show that the set $\mathcal{S}$ is linearly independent. Consider the equation

$$c_1\vec{v_1} + \cdots + c_n\vec{v_n} = \vec{0}.$$

Taking the coordinates with respect to $\mathcal{B}$ of both sides of the equation above, as finding coordinates with respect to a basis is a linear process and $[\vec{0}]_{\mathcal{B}} = \vec{0}$, we obtain the equation

$$c_1[\vec{v_1}]_{\mathcal{B}} + \cdots + c_n[\vec{v_n}]_{\mathcal{B}} = \vec{0}.$$

The set $\{[\vec{v_1}]_{\mathcal{B}}, \ldots, [\vec{v_n}]_{\mathcal{B}}\}$ consists of the columns of P. Since P is invertible, then by Theorem 4.6.7 (Invertibility Criteria – First Version), $\operatorname{rank}(P) = n$. It then follows from Proposition 8.3.6 (Pivots and Linear Independence) that $\{[\vec{v_1}]_{\mathcal{B}}, \ldots, [\vec{v_n}]_{\mathcal{B}}\}$ is linearly independent. Therefore, $c_1 = \cdots = c_n = 0$ and thus, $\mathcal{S}$ is linearly independent and a basis for $\mathbb{F}^n$.

Next, we will show that the vectors in $\mathcal{S}$ are also eigenvectors of T. Using the fact that $P^{-1}[T]_{\mathcal{B}}P = D = \operatorname{diag}(d_1, d_2, \ldots, d_n)$, we can multiply both sides by P to obtain that $[T]_{\mathcal{B}}P = PD$. Comparing the i^{th} column of $[T]_{\mathcal{B}}P$ with the i^{th} column of PD we obtain the equation

$$[T]_{\mathcal{B}}[\vec{v_i}]_{\mathcal{B}} = d_i[\vec{v_i}]_{\mathcal{B}}.$$

Therefore, by Proposition 9.1.3 and the linearity of finding coordinates,

$$[T(\vec{v_i})]_{\mathcal{B}} = [d_i\vec{v_i}]_{\mathcal{B}}.$$

Thus, $T(\vec{v_i}) = d_i\vec{v_i}$ by Theorem 8.8.1 (Unique Representation Theorem). We know that $\vec{v_i} \neq \vec{0}$, since $\vec{v_i}$ belongs to a linearly independent set, and so $\vec{v_i}$ is an eigenvector of T.

Consequently, we conclude that $\mathcal{S}$ is a basis of eigenvectors of T, and T is diagonalizable over $\mathbb{F}$ by Proposition 9.2.6 (Eigenvector Basis Criterion for Diagonalizability). $\square$

Using this result, we can translate our criterion for diagonalizability of operators from Proposition 9.2.6 (Eigenvector Basis Criterion for Diagonalizability) to a criterion for diagonalizability of matrices.

Corollary 9.3.2 (Eigenvector Basis Criterion for Diagonalizability – Matrix Version)

Let $A \in M_{n\times n}(\mathbb{F})$. Then A is diagonalizable over $\mathbb{F}$ if and only if there exists a basis of $\mathbb{F}^n$ consisting of eigenvectors of A.

Proof: Consider the linear operator $T_A : \mathbb{F}^n \to \mathbb{F}^n$ such that $T_A(\vec{x}) = A\vec{x}$. Then $[T_A]_{\mathcal{E}} = A$ and

A is diagonalizable over $\mathbb{F}$

$\iff$ $[T_A]_{\mathcal{E}}$ is diagonalizable over $\mathbb{F}$
(because $[T_A]_{\mathcal{E}} = A$)

$\iff$ T_A is diagonalizable over $\mathbb{F}$
(by Proposition 9.3.1 (T Diagonalizable iff $[T]_{\mathcal{B}}$ Diagonalizable))

$\iff$ There exists a basis $\mathcal{B}$ of $\mathbb{F}^n$ of eigenvectors of T_A
(by Proposition 9.2.6 (Eigenvector Basis Criterion for Diagonalizability))

By Proposition 9.2.5 (Eigenpairs of T and $[T]_{\mathcal{B}}$), $\vec{x}$ is an eigenvector of T_A if and only if $\vec{x}$ is an eigenvector of $[T]_{\mathcal{E}} = A$. Therefore, there exists a basis $\mathcal{B}$ of $\mathbb{F}^n$ of eigenvectors of T_A if and only if there exists a basis $\mathcal{B}$ of $\mathbb{F}^n$ of eigenvectors of A. $\square$

REMARKS

The previous two results contain several important pieces of information.

1. To determine whether a linear operator T is diagonalizable over $\mathbb{F}$, we can start by obtaining the matrix of T with respect to some basis $\mathcal{B}$ for $\mathbb{F}^n$, and then test whether *this matrix* is diagonalizable over $\mathbb{F}$. The most natural *initial* choice for $\mathcal{B}$ is the standard basis, $\mathcal{E}$. **Thus, moving forward, we can focus our discussion on the diagonalization of matrices.**

2. If A is diagonalizable over $\mathbb{F}$ and $P^{-1}AP = D$ is a diagonal matrix, then
   (a) the entries of D are the eigenvalues of A in $\mathbb{F}$,
   (b) the matrix P is the change of basis matrix from an ordered basis $\mathcal{B}$ for $\mathbb{F}^n$ comprised of eigenvectors of A to the standard basis $\mathcal{E}$ for $\mathbb{F}^n$ (so $P = {}_{\mathcal{E}}[I]_{\mathcal{B}}$), and
   (c) the columns of P are eigenvectors of A.

Corollary 9.3.2 (Eigenvector Basis Criterion for Diagonalizability – Matrix Version) tells us that determining whether a matrix A is diagonalizable over $\mathbb{F}$ boils down to whether or not we can find n linearly independent eigenvectors of A in $\mathbb{F}^n$.

Example 9.3.3

Let $T\colon \mathbb{R}^3 \to \mathbb{R}^3$ be a linear operator with $[T]_{\mathcal{E}} = \begin{bmatrix} 3 & 4 & -2 \\ 3 & 8 & -3 \\ 6 & 14 & -5 \end{bmatrix}$. We will show that T is diagonalizable over $\mathbb{R}$ by finding three linearly independent eigenvectors of $A = [T]_{\mathcal{E}}$. In order to do this, let us begin by finding all the possible eigenvectors of A.

The characteristic polynomial of A is

$$C_A(\lambda) = \det\left(\begin{bmatrix} 3-\lambda & 4 & -2 \\ 3 & 8-\lambda & -3 \\ 6 & 14 & -5-\lambda \end{bmatrix}\right) = -\lambda^3 + 6\lambda^2 - 11\lambda + 6 = -(\lambda-3)(\lambda-2)(\lambda-1).$$

The eigenvalues of A are thus $\lambda_1 = 1$, $\lambda_2 = 2$ and $\lambda_3 = 3$. We leave it as an exercise to show that the corresponding eigenspaces are given by

$$E_1 = \operatorname{Span}\left\{\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}\right\},\ E_2 = \operatorname{Span}\left\{\begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix}\right\} \text{ and } E_3 = \operatorname{Span}\left\{\begin{bmatrix} 1 \\ 3 \\ 6 \end{bmatrix}\right\}.$$

Therefore, there is an obvious choice for our desired three linearly independent eigenvectors:

$$\mathcal{B} = \left\{\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ 3 \\ 6 \end{bmatrix}\right\}.$$

Of course, we have to check that this set is in fact linearly independent. We can easily do this using Proposition 8.3.6 (Pivots and Linear Independence), for instance. Thus $\mathcal{B}$ is a basis for $\mathbb{R}^3$ consisting of eigenvectors of A.

Therefore, A is diagonalizable over $\mathbb{F}$, by Corollary 9.3.2 (Eigenvector Basis Criterion for Diagonalizability – Matrix Version), and hence T is diagonalizable over $\mathbb{F}$ too, by Proposition 9.3.1 (T Diagonalizable iff $[T]_{\mathcal{B}}$ Diagonalizable).

In the previous Example we could have invoked Proposition 7.6.7 (n Distinct Eigenvalues $\implies$ Diagonalizable) to deduce that A is diagonalizable. Notice that Proposition 7.6.7 suggests the same strategy that we followed above: namely, take an eigenvector corresponding to each one of the eigenvalues of A, and form a basis using these eigenvectors. For this strategy to actually produce a basis, we need to be sure that the resulting set of eigenvectors is linearly independent.

Proposition 9.3.4

(Eigenvectors Corresponding to Distinct Eigenvalues are Linearly Independent)

Let $A \in M_{n\times n}(\mathbb{F})$ have eigenpairs $(\lambda_1, \overrightarrow{v_1}), (\lambda_2, \overrightarrow{v_2}), \ldots, (\lambda_k, \overrightarrow{v_k})$ over $\mathbb{F}$.

If the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_k$ are all distinct, then the set of eigenvectors $\{\overrightarrow{v_1}, \overrightarrow{v_2}, \ldots, \overrightarrow{v_k}\}$ is linearly independent.

Proof: Proceed by induction on k. If $k = 1$, then since eigenvectors are non-zero we have that $\{\overrightarrow{v_1}\}$ is linearly independent.

Assume the statement holds for $k = j$, that is, we assume that the set $\{\vec{v_1}, \vec{v_2}, \ldots, \vec{v_j}\}$ is linearly independent.

Consider the set $\{\vec{v_1}, \vec{v_2}, \ldots, \vec{v_j}, \overrightarrow{v_{j+1}}\}$. We must show that this set is linearly independent and so we consider the equation

$$c_1\vec{v_1} + \cdots + c_{j+1}\overrightarrow{v_{j+1}} = \vec{0}. \quad (*)$$

We'll manipulate (*) in two ways. First, multiply both sides of (*) through by A to obtain

$$c_1\lambda_1\vec{v_1} + \cdots + c_j\lambda_j\vec{v_j} + c_{j+1}\lambda_{j+1}\overrightarrow{v_{j+1}} = \vec{0}. \quad (**)$$

Next, multiply both sides of (*) through by λ_{j+1} to obtain

$$c_1\lambda_{j+1}\vec{v_1} + \cdots + c_j\lambda_{j+1}\vec{v_j} + c_{j+1}\lambda_{j+1}\overrightarrow{v_{j+1}} = \vec{0}. \quad (***)$$

Then subtract $(***)$ from $(**)$ to get

$$c_1(\lambda_1 - \lambda_{j+1})\vec{v_1} + \cdots + c_j(\lambda_j - \lambda_{j+1})\vec{v_j} = \vec{0}.$$

By the inductive hypothesis, $\{\vec{v_1}, \vec{v_2}, \ldots, \vec{v_j}\}$ is linearly independent so for all i, $1 \leq i \leq j$,

$$c_i(\lambda_i - \lambda_{j+1}) = 0.$$

However, $\lambda_i \neq \lambda_{j+1}$ and so $c_i = 0$. Thus, $c_1 = \cdots = c_j = 0$ and our original equation (*) becomes

$$c_{j+1}\overrightarrow{v_{j+1}} = \vec{0}.$$

The vector $\overrightarrow{v_{j+1}}$ is nonzero since it is an eigenvector, so we conclude that $c_{j+1} = 0$. Consequently, the set $\{\vec{v_1}, \vec{v_2}, \ldots, \vec{v_j}, \overrightarrow{v_{j+1}}\}$ is linearly independent. This establishes the inductive step and completes the proof. □

We can now provide a proof of Proposition 7.6.7 (n Distinct Eigenvalues $\implies$ Diagonalizable) which states the following:

If $A \in M_{n\times n}(\mathbb{F})$ has n *distinct* eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ in $\mathbb{F}$, then A is diagonalizable over $\mathbb{F}$.

More specifically, if we let $(\lambda_1, \vec{v_1}), \ldots, (\lambda_n, \vec{v_n})$ be eigenpairs of A over $\mathbb{F}$, and if we let $P = [\vec{v_1} \cdots \vec{v_n}]$ be the matrix whose columns are eigenvectors corresponding to the distinct eigenvalues, then

(a) P is invertible, and

(b) $P^{-1}AP = D = \mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)$.

Proof: Since we have n eigenvectors corresponding to n distinct eigenvalues, then the set $\{\vec{v_1}, \ldots, \vec{v_n}\}$ is linearly independent by Proposition 9.3.4 (Eigenvectors Corresponding to Distinct Eigenvalues are Linearly Independent). Further, the set $\{\vec{v_1}, \ldots, \vec{v_n}\}$ is a basis of eigenvectors for $\mathbb{F}^n$. Therefore, A is diagonalizable over $\mathbb{F}$ by Proposition 9.2.6 (Eigenvector Basis Criterion for Diagonalizability).

Let $P = [\overrightarrow{v_1} \cdots \overrightarrow{v_n}]$. Since $\{\overrightarrow{v_1}, \ldots, \overrightarrow{v_n}\}$ is a basis, then it is linearly independent, and so $\operatorname{rank}(P) = n$ by Proposition 8.3.6 (Pivots and Linear Independence). Therefore, by Theorem 4.6.7 (Invertibility Criteria – First Version), P is invertible. Now since $\overrightarrow{v_1}, \ldots, \overrightarrow{v_n}$ are eigenvectors of A, then using Lemma 4.2.2 (Column Extraction),

$$AP = A[\overrightarrow{v_1} \cdots \overrightarrow{v_n}] = [A\overrightarrow{v_i} \cdots A\overrightarrow{v_n}] = [\lambda_1\overrightarrow{v_1} \cdots \lambda_n\overrightarrow{v_n}].$$

Let $D = \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n) = [\lambda_1\overrightarrow{e_1} \cdots \lambda_n\overrightarrow{e_n}]$. Therefore, using Lemma 4.2.2 again,

$$PD = P[\lambda_1\overrightarrow{e_1} \cdots \lambda_n\overrightarrow{e_n}] = [\lambda_1 P\overrightarrow{e_1} \cdots \lambda_n P\overrightarrow{e_n}] = [\lambda_1\overrightarrow{v_1} \cdots \lambda_n\overrightarrow{v_n}].$$

Therefore, $AP = PD$ and so

$$P^{-1}AP = D = \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n).$$

□

Example 9.3.5 Suppose $T\colon \mathbb{R}^3 \to \mathbb{R}^3$ is a linear operator with $[T]_{\mathcal{E}} = \begin{bmatrix} 5 & 2 & -1 \\ 8 & 1 & -2 \\ 16 & 0 & -3 \end{bmatrix}$.

(a) Prove that T is diagonalizable over $\mathbb{R}$.

(b) Find a basis $\mathcal{B}$ of $\mathbb{R}^3$ such that $[T]_{\mathcal{B}}$ is a diagonal matrix.

(c) Determine an invertible matrix P such that $P^{-1}\,[T]_{\mathcal{E}}\,P = [T]_{\mathcal{B}}$.

Solution:

(a) By Proposition 9.3.1 (T Diagonalizable iff $[T]_{\mathcal{B}}$ Diagonalizable), it suffices to prove that $A = [T]_{\mathcal{E}}$ is diagonalizable over $\mathbb{R}$. The characteristic polynomial of A is

$$C_A(\lambda) = \det\left(\begin{bmatrix} 5-\lambda & 2 & -1 \\ 8 & 1-\lambda & -2 \\ 16 & 0 & -3-\lambda \end{bmatrix}\right) = -\lambda^3 + 3\lambda^2 + 13\lambda - 15.$$

Since

$$-\lambda^3 + 3\lambda^2 + 13\lambda - 15 = -(\lambda-5)(\lambda-1)(\lambda+3),$$

the eigenvalues of T are $\lambda_1 = 5, \lambda_2 = 1$, and $\lambda_3 = -3$. There are three distinct eigenvalues, and therefore A is diagonalizable over $\mathbb{R}$ by Proposition 7.6.7 (n Distinct Eigenvalues $\implies$ Diagonalizable).

(b) We want a basis $\mathcal{B}$ for $\mathbb{R}^3$ consisting of eigenvectors of A. Since we have three distinct eigenvalues, we need to select one eigenvector from each eigenspace. We leave it as an exercise to determine that the eigenspaces of A are given by

$$E_5 = \operatorname{Span}\left\{\begin{bmatrix}1\\1\\2\end{bmatrix}\right\},\ E_1 = \operatorname{Span}\left\{\begin{bmatrix}1\\0\\4\end{bmatrix}\right\} \text{ and } E_{-3} = \operatorname{Span}\left\{\begin{bmatrix}0\\1\\2\end{bmatrix}\right\}.$$

Thus,

$$\mathcal{B} = \left\{\begin{bmatrix}1\\1\\2\end{bmatrix}, \begin{bmatrix}1\\0\\4\end{bmatrix}, \begin{bmatrix}0\\1\\2\end{bmatrix}\right\}$$

is a basis for $\mathbb{R}^3$ for which $[T]_{\mathcal{B}} = \operatorname{diag}(5, 1, -3)$.

(c) The columns of the desired matrix P are the eigenvectors in $\mathcal{B}$. Thus,

$$P = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 2 & 4 & 2 \end{bmatrix}.$$

REMARK

We repeat for emphasis that the matrix P in part (c) that diagonalizes A is the change of basis matrix from the basis of eigenvectors of A to the standard basis, ${}_{\mathcal{E}}[I]_{\mathcal{B}}$.

9.4 The Diagonalizability Test

For $A \in M_{n\times n}(\mathbb{F})$ to be diagonalizable over $\mathbb{F}$, it must have n linearly independent eigenvectors in $\mathbb{F}^n$. Let us consider some examples where n linearly independent eigenvectors do not exist.

Example 9.4.1

Let $A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$. Then $C_A(\lambda) = \lambda^2 + 1$. This matrix has no real eigenvalues and therefore no real eigenvectors. So A is not diagonalizable over $\mathbb{R}$.

However, we note that A does have two distinct complex eigenvalues, $\lambda_1 = i$ and $\lambda_2 = -i$, and is therefore diagonalizable over $\mathbb{C}$.

Example 9.4.2

Let $A = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{bmatrix}$. Then $C_A(\lambda) = -(\lambda - 2)(\lambda^2 + 1)$.

This matrix has only one real eigenvalue $\lambda_1 = 2$. We can show that all eigenvectors associated with this eigenvalue are scalar multiples of $\begin{bmatrix} 1 & 0 & 0 \end{bmatrix}^T$. Therefore, A is not diagonalizable over $\mathbb{R}$.

However, we note that A does have three distinct complex eigenvalues, $\lambda_1 = 2$, $\lambda_2 = i$, and $\lambda_3 = -i$, and is therefore diagonalizable over $\mathbb{C}$.

Example 9.4.3

Let $A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$. Then $C_A(\lambda) = \lambda^2 = (\lambda - 0)^2$.

This matrix has two real eigenvalues, but they are both 0. We can show that all eigenvectors associated with the eigenvalue 0 are scalar multiples of $\begin{bmatrix} 1 & 0 \end{bmatrix}^T$. Therefore, A is not diagonalizable over $\mathbb{R}$ or $\mathbb{C}$.

We will now introduce some theory that will help us determine exactly when a basis of eigenvectors exists.

Definition 9.4.4
Algebraic Multiplicity

Let λ_i be an eigenvalue of $A \in M_{n\times n}(\mathbb{F})$. The **algebraic multiplicity** of λ_i, denoted by a_{λ_i}, is the largest positive integer such that $(\lambda-\lambda_i)^{a_{\lambda_i}}$ divides the characteristic polynomial $C_A(\lambda)$.

In other words, a_{λ_i} gives the number of times that $(\lambda-\lambda_i)$ terms occur in the fully factorized form of $C_A(\lambda)$.

Example 9.4.5

Let $A = \begin{bmatrix} 1 & 2 & 2 \\ 2 & 1 & 2 \\ 2 & 2 & 1 \end{bmatrix}$. Then $C_A(\lambda) = -(\lambda-5)(\lambda+1)^2$.

The matrix A has two distinct eigenvalues: $\lambda_1 = 5$, with algebraic multiplicity $a_{\lambda_1} = 1$, and $\lambda_2 = -1$, with algebraic multiplicity $a_{\lambda_2} = 2$.

We can also say that A has three eigenvalues with one of them being repeated twice. So its eigenvalues are $5, -1$, and -1.

Definition 9.4.6
Geometric Multiplicity

Let λ_i be an eigenvalue of $A \in M_{n\times n}(\mathbb{F})$. The **geometric multiplicity** of λ_i, denoted by g_{λ_i}, is the dimension of the eigenspace E_{λ_i}. That is, $g_{\lambda_i} = \dim(E_{\lambda_i})$.

Example 9.4.7

Let $A \in M_{3\times 3}(\mathbb{R})$ with $A = \begin{bmatrix} 1 & 2 & 2 \\ 2 & 1 & 2 \\ 2 & 2 & 1 \end{bmatrix}$. Determine the geometric multiplicities of all eigenvalues of A.

Solution:

The eigenvalues of A are $\lambda_1 = 5$ and $\lambda_2 = -1$.

The eigenspace E_{λ_1} is the solution set to $(A-5I)\vec{x} = \vec{0}$, which is equivalent to

$$\begin{bmatrix} -4 & 2 & 2 \\ 2 & -4 & 2 \\ 2 & 2 & -4 \end{bmatrix} \vec{x} = \vec{0}$$

and row reduces to

$$\begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{bmatrix} \vec{x} = \vec{0}.$$

A basis for E_5 is thus

$$\left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right\}.$$

Therefore, $\dim(E_{\lambda_1}) = 1$ and so $g_{\lambda_1} = 1$.

The eigenspace E_{λ_2} is the solution set to $(A+1I)\vec{x} = \vec{0}$, which is equivalent to

$$\begin{bmatrix} 2 & 2 & 2 \\ 2 & 2 & 2 \\ 2 & 2 & 2 \end{bmatrix} \vec{x} = \vec{0}$$

and row reduces to

$$\begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \vec{x} = \vec{0}.$$

A basis for E_{λ_2} is thus

$$\left\{ \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} \right\}.$$

Therefore, $\dim(E_{\lambda_2}) = 2$ and so $g_{\lambda_2} = 2$.

The geometric multiplicity of an eigenvalue tells us the maximum number of linearly independent eigenvectors we can obtain from the eigenspace corresponding to that eigenvalue.

The two multiplicities are connected through the following result.

Proposition 9.4.8 **(Geometric and Algebraic Multiplicities)**

Let λ_i be an eigenvalue of the matrix $A \in M_{n\times n}(\mathbb{F})$. Then

$$1 \leq g_{\lambda_i} \leq a_{\lambda_i}.$$

REMARK

In the proof below, we will describe a matrix by using what is called a **block matrix**; that is, denoting some or all parts of a matrix as matrices themselves.

For example, given the matrix

$$A = \begin{bmatrix} 1 & 2 & 5 & 6 & 7 \\ 3 & 4 & 8 & 9 & 10 \\ 11 & 12 & 17 & 18 & 19 \\ 13 & 14 & 20 & 21 & 22 \\ 15 & 16 & 23 & 24 & 25 \end{bmatrix}$$

we can describe A as

$$A = \begin{bmatrix} A_1 & A_2 \\ B_1 & B_2 \end{bmatrix}$$

where

$$A_1 = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}, \quad A_2 = \begin{bmatrix} 5 & 6 & 7 \\ 8 & 9 & 10 \end{bmatrix}, \quad B_1 = \begin{bmatrix} 11 & 12 \\ 13 & 14 \\ 15 & 16 \end{bmatrix}, \quad B_2 = \begin{bmatrix} 17 & 18 & 19 \\ 20 & 21 & 22 \\ 23 & 24 & 25 \end{bmatrix}.$$

Alternatively, with A_2, B_1, and B_2 defined as above, we can re-write our matrix A as

$$A = \left[\begin{array}{cc|c} 1 & 2 & \multirow{2}{*}{A_2} \\ 3 & 4 & \\ \hline \multicolumn{2}{c|}{B_1} & B_2 \end{array}\right].$$

Of course, there are many other ways we could slice A into blocks to give different block representations of A.

Proof: By definition, if λ_i is an eigenvalue of A, then there is a non-trivial solution to $A\vec{v} = \lambda_i \vec{v}$. Therefore, each eigenspace contains at least one non-zero vector and so its dimension will be at least one. Therefore, $1 \le g_{\lambda_i}$.

Suppose that $\{\vec{v_1}, \dots, \vec{v_k}\}$ is a basis for E_{λ_i}. Therefore, $g_{\lambda_i} = k$.

We can extend this basis for E_{λ_i} to a basis $\mathcal{B}$ for $\mathbb{F}^n$ (for a demonstration of how this can be done, see Example 8.5.7, as well as the remark preceding it). Therefore, $\mathcal{B} = \{\vec{v_1}, \dots, \vec{v_k}, \overrightarrow{w_{k+1}}, \dots, \vec{w_n}\}$.

We let $T_A : \mathbb{F}^n \to \mathbb{F}^n$ be defined as $T_A(\vec{v}) = A\vec{v}$.

For $j = 1, \dots, k$, it is the case that $\vec{v_j} \in E_{\lambda_i}$ and therefore, $A\vec{v_j} = \lambda_i \vec{v_j}$. Therefore, $[T_A(\vec{v_j})]_{\mathcal{B}} = \lambda_i \vec{e_j}$. For $j = k+1, \dots, n$, $[T_A(\vec{w_j})]_{\mathcal{B}}$ will be some vector in $\mathbb{F}^n$. It now follows that $[T_A]_{\mathcal{B}}$ has the block structure

$$[T_A]_{\mathcal{B}} = \left[\begin{array}{cccc|c} \lambda_i & 0 & \cdots & 0 & \\ 0 & \lambda_i & \cdots & 0 & \\ \vdots & \vdots & \ddots & \vdots & M_1 \\ 0 & 0 & \cdots & \lambda_i & \\ \hline & \mathcal{O}_{(n-k)\times k} & & & M_2 \end{array}\right]$$

for some $M_1 \in M_{k\times(n-k)}(\mathbb{F})$ and some $M_2 \in M_{(n-k)\times(n-k)}(\mathbb{F})$.

The characteristic polynomial of A is equal to the characteristic polynomials of $[T_A]_{\mathcal{B}}$ since these two matrices are similar (exercise).

It can be shown by induction that

$$C_A(\lambda) = (\lambda_i - \lambda)^k C_{M_2}(\lambda) = (\lambda_i - \lambda)^{g_{\lambda_i}} C_{M_2}(\lambda).$$

Since a_{λ_i} is the largest positive integer such that $(\lambda - \lambda_i)^{a_{\lambda_i}}$ divides $C_A(t)$, it follows that $g_{\lambda_i} \le a_{\lambda_i}$. □

What happens if we take the union of the bases of all the eigenspaces of a given matrix? It turns out that this set is linearly independent.

Proposition 9.4.9 Let $A \in M_{n\times n}(\mathbb{F})$ with distinct eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_k$. If their corresponding eigenspaces, $E_{\lambda_1}, E_{\lambda_2}, \dots, E_{\lambda_k}$ have bases $\mathcal{B}_1, \mathcal{B}_2, \dots, \mathcal{B}_k$, then $\mathcal{B} = \mathcal{B}_1 \cup \mathcal{B}_2 \cup \cdots \cup \mathcal{B}_k$ is linearly independent.

Proof: To make things easier to read, let $g_i = g_{\lambda_i}$ for $i = 1, \dots, k$.

Let $\mathcal{B}_1 = \{\overrightarrow{v_{1\,1}}, \dots, \overrightarrow{v_{1\,g_1}}\}$, $\mathcal{B}_2 = \{\overrightarrow{v_{2\,1}}, \dots, \overrightarrow{v_{2\,g_2}}\}$ and so on up to $\mathcal{B}_k = \{\overrightarrow{v_{k\,1}} \dots, \overrightarrow{v_{k\,g_k}}\}$. Take note of the double subscripts here, with the first subscript corresponding to the numbering of the basis ($\mathcal{B}_1, \mathcal{B}_2$, and so on) and the second subscript counting the vector's position within that basis.

We wish to perform the linear dependence check on $\mathcal{B} = \mathcal{B}_1 \cup \mathcal{B}_2 \cup \cdots \cup \mathcal{B}_k$. We consider the following equation with scalars in $\mathbb{F}$:

$$\sum_{i=1}^{g_1} c_{1i}\overrightarrow{v_{1\,i}} + \sum_{i=1}^{g_2} c_{2i}\overrightarrow{v_{2\,i}} + \cdots + \sum_{i=1}^{g_k} c_{ki}\overrightarrow{v_{k\,i}} = \vec{0} \quad (*).$$

Let $\overrightarrow{w_1} = \sum_{i=1}^{g_1} c_{1i}\overrightarrow{v_1}_i$, which is a vector from E_{λ_1}. Let $\overrightarrow{w_2} = \sum_{i=1}^{g_2} c_{2i}\overrightarrow{v_2}_i$, which is a vector from E_{λ_2}. Continuing in this manner, let $\overrightarrow{w_k} = \sum_{i=1}^{g_k} c_{ki}\overrightarrow{v_k}_i$, which is a vector from E_{λ_k}.

Therefore, our equation (*) has become

$$\overrightarrow{w_1} + \overrightarrow{w_2} + \cdots + \overrightarrow{w_k} = \overrightarrow{0} \quad (**).$$

Each of the $\overrightarrow{w_j}$ are from a different eigenspace. The vectors in a given eigenspace are either eigenvectors or the zero vector. If any of the $\overrightarrow{w_j}$ are not the zero vector, then from (**) we get a non-trivial linear combination of eigenvectors each chosen from a different eigenspace equalling to the zero vector. This contradicts Proposition 9.3.4 (Eigenvectors Corresponding to Distinct Eigenvalues are Linearly Independent), which tells us that a set of eigenvectors chosen in such a way that every eigenvector is from a different eigenspace is linearly independent. Therefore, it must be the case that $\overrightarrow{w_1} = \cdots = \overrightarrow{w_k} = \overrightarrow{0}$.

But $\overrightarrow{w_j} = \sum_{i=1}^{g_j} c_{ji}\overrightarrow{v_j}_i = \overrightarrow{0}$ implies that $c_{j1} = \cdots = c_{jg_{\lambda_j}} = 0$ because $\mathcal{B}_j$ is a basis for E_{λ_j}. Therefore, all of the scalars in (*) are 0 and the set $\mathcal{B}$ is linearly independent. □

We have now reached the stage where we can give a test to determine whether a matrix is diagonalizable over $\mathbb{F}$.

Theorem 9.4.10 **(Diagonalizability Test)**

Let $A \in M_{n\times n}(\mathbb{F})$. Suppose that the complete factorization of the characteristic polynomial of A into irreducible factors over $\mathbb{F}$ is given by

$$C_A(\lambda) = (\lambda - \lambda_1)^{a_{\lambda_1}} \cdots (\lambda - \lambda_k)^{a_{\lambda_k}}\, h(\lambda),$$

where $\lambda_1, \ldots \lambda_k$ are all of the distinct eigenvalues of A in $\mathbb{F}$ with corresponding algebraic multiplicities $a_{\lambda_1} \ldots a_{\lambda_k}$ and $h(\lambda)$ is a polynomial in λ that is irreducible over $\mathbb{F}$. Then A is diagonalizable over $\mathbb{F}$ if and only if $h(\lambda)$ is a constant polynomial and $a_{\lambda_i} = g_{\lambda_i}$, for each $i = 1, \ldots, k$.

Proof: We begin with the forward direction. Thus, assume that A is diagonalizable over $\mathbb{F}$. Then, by Corollary 9.3.2 (Eigenvector Basis Criterion for Diagonalizability – Matrix Version), A has n linearly independent eigenvectors in $\mathbb{F}^n$. Each of these eigenvectors must lie in one of the eigenspaces E_{λ_i} of A, and no eigenvector can lie in two different eigenspaces. Since there can be at most g_{λ_i} linearly independent vectors in E_{λ_i}, we conclude that

$$n \le g_{\lambda_i} + \ldots + g_{\lambda_k}.$$

On the other hand, by Proposition 9.4.8 (Geometric and Algebraic Multiplicities), we know that $g_{\lambda_i} \le a_{\lambda_i}$, and so

$$g_{\lambda_1} + \cdots + g_{\lambda_k} \le a_{\lambda_1} + \cdots + a_{\lambda_k}.$$

We also know that the sum of algebraic multiplicities $a_{\lambda_i} + \cdots + a_{\lambda_k}$ is at most the degree of $C_A(\lambda)$, which is n by Proposition 7.3.4 (Features of the Characteristic Polynomial). Therefore,

$$n \le g_{\lambda_i} + \ldots + g_{\lambda_k} \le a_{\lambda_i} + \cdots + a_{\lambda_k} \le n.$$

This implies that

$$g_{\lambda_i} + \ldots + g_{\lambda_k} = a_{\lambda_i} + \cdots + a_{\lambda_k} = n = \deg(C_A(\lambda)).$$

From this we immediately conclude that $\deg(h(\lambda)) = 0$, i.e., that $h(\lambda)$ is a constant polynomial. Since $g_{\lambda_i} \leq a_{\lambda_i}$ for all $i = 1, \ldots, k$, we also conclude from this that $g_{\lambda_i} = a_{\lambda_i}$ for all $i = 1, \ldots, k$. This completes the proof of the forward direction.

Conversely, assume that $h(\lambda)$ is a constant polynomial and that $g_{\lambda_i} = a_{\lambda_i}$ for all $i = 1, \ldots, k$. This implies that

$$g_{\lambda_i} + \ldots + g_{\lambda_k} = a_{\lambda_i} + \cdots + a_{\lambda_k} = n.$$

Thus if we let $\mathcal{B}_k$ be a basis for E_{λ_k} for $i = 1, \ldots, k$, and if we let $\mathcal{B} = \mathcal{B}_1 \cup \cdots \cup \mathcal{B}_k$ be the union of these bases, then by Proposition 9.4.9, $\mathcal{B}$ will linearly independent, and by the above, $\mathcal{B}$ will contain $g_{\lambda_i} + \ldots + g_{\lambda_k} = n$ eigenvectors. Thus $\mathcal{B}$ is a linearly independent subset of $\mathbb{F}^n$ that contains n vectors, so $\mathcal{B}$ must be a basis for $\mathbb{F}^n$. Therefore, we have found a basis for $\mathbb{F}^n$ consisting of eigenvectors of A, so A must be diagonalizable over $\mathbb{F}$, by Corollary 9.3.2 (Eigenvector Basis Criterion for Diagonalizability – Matrix Version).

$\square$

REMARK

If $\mathbb{F} = \mathbb{C}$, then the polynomial $h(\lambda)$ in Theorem 9.4.10 (Diagonalizability Test) is always constant (in fact, it is equal to $(-1)^n$.) This is because every degree n polynomial factors over $\mathbb{C}$ into linear factors. Therefore $h(\lambda)$ only matters if $\mathbb{F} = \mathbb{R}$. It measures the failure of $C_A(\lambda)$ to factor completely into linear factors, and so it measures, in a sense, a deficit of real eigenvalues.

Example 9.4.11

Consider the matrix $A = \begin{bmatrix} 1 & 2 & 2 \\ 2 & 1 & 2 \\ 2 & 2 & 1 \end{bmatrix}$ from Example 9.4.7. Determine whether A is diagonalizable over $\mathbb{R}$ and/or $\mathbb{C}$.

Solution: In Example 9.4.7, we found the characteristic polynomial to be

$$C_A(\lambda) = -(\lambda - 5)(\lambda + 1)^2.$$

So $h(\lambda) = -1$, a constant polynomial. The eigenvalues are $\lambda_1 = 5$ and $\lambda_2 = -1$ and we determined that $a_{\lambda_1} = g_{\lambda_1} = 1$ and $a_{\lambda_2} = g_{\lambda_2} = 2$. Therefore, A is diagonalizable over both $\mathbb{R}$ and $\mathbb{C}$, using Theorem 9.4.10 (Diagonalizability Test).

Example 9.4.12

Determine whether $A = \begin{bmatrix} 1 & 2 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 2 & 1 \end{bmatrix}$ is diagonalizable over $\mathbb{R}$ and/or $\mathbb{C}$.

Solution: The characteristic polynomial is $C_A(\lambda) = (\lambda - 3)^2(\lambda + 1)^2$. Therefore, $h(\lambda) = 1$, a constant polynomial.

The eigenvalues are: $\lambda_1 = 3$, with $a_{\lambda_1} = 2$, and $\lambda_2 = -1$, with $a_{\lambda_2} = 2$.

The eigenspace E_{λ_1} is the solution set to $(A - 3I)\vec{v} = \vec{0}$, which is equivalent to

$$\begin{bmatrix} -2 & 2 & 0 & 0 \\ 2 & -2 & 0 & 0 \\ 0 & 0 & -2 & 2 \\ 0 & 0 & 2 & -2 \end{bmatrix} \vec{v} = \vec{0}$$

and row reduces to

$$\begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \vec{v} = \vec{0}.$$

Since the rank of the coefficient matrix is 2, then there will be $4 - 2 = 2$ parameters in the solution set. Therefore, $g_{\lambda_1} = \dim(E_{\lambda_1}) = 2$.

The eigenspace E_{λ_2} is the solution set to $(A + 1I)\vec{v} = \vec{0}$ which is equivalent to

$$\begin{bmatrix} 2 & 2 & 0 & 0 \\ 2 & 2 & 0 & 0 \\ 0 & 0 & 2 & 2 \\ 0 & 0 & 2 & 2 \end{bmatrix} \vec{v} = \vec{0}$$

and row reduces to

$$\begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \vec{v} = \vec{0}.$$

Since the rank of the coefficient matrix is 2, then there will be $4 - 2 = 2$ parameters in the solution set. Therefore, $g_{\lambda_2} = \dim(E_{\lambda_2}) = 2$.

Since $a_{\lambda_1} = g_{\lambda_1} = 2$, and $a_{\lambda_2} = g_{\lambda_2} = 2$, A is diagonalizable over both $\mathbb{R}$ and $\mathbb{C}$ by Theorem 9.4.10 (Diagonalizability Test).

Example 9.4.13 Determine whether $A = \begin{bmatrix} 5 & 2 & 0 & 1 \\ -2 & 1 & 0 & -1 \\ 4 & 4 & 3 & 2 \\ 16 & 0 & -8 & -5 \end{bmatrix}$ is diagonalizable over $\mathbb{R}$ and/or $\mathbb{C}$.

Solution: The characteristic polynomial $C_A(\lambda) = (\lambda - 3)^3(\lambda + 5)$. Therefore, $h(\lambda) = 1$, a constant polynomial.

The eigenvalues are: $\lambda_1 = 3$, with $a_{\lambda_1} = 3$, and $\lambda_2 = -5$, with $a_{\lambda_2} = 1$.

The eigenspace E_{λ_1} is the solution set to $(A - 3I)\vec{v} = \vec{0}$, which is equivalent to

$$\begin{bmatrix} 2 & 2 & 0 & 1 \\ -2 & -2 & 0 & -1 \\ 4 & 4 & 0 & 2 \\ 16 & 0 & -8 & -8 \end{bmatrix} \vec{v} = \vec{0}$$

and row reduces to

$$\begin{bmatrix} 2 & 2 & 0 & 1 \\ 0 & 2 & 1 & 2 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \vec{v} = \vec{0}.$$

Since the rank of the coefficient matrix is 2, then there will be $4 - 2 = 2$ parameters in the solution set. Therefore, $\dim(E_{\lambda_1}) = 2$.

Since $a_{\lambda_1} = 3$ and $g_{\lambda_1} = 2$, A is not diagonalizable over $\mathbb{R}$ nor $\mathbb{C}$, by Theorem 9.4.10 (Diagonalizability Test).

Example 9.4.14

Determine whether $A = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$ is diagonalizable over $\mathbb{R}$ and/or $\mathbb{C}$.

Solution: The characteristic polynomial $C_A(\lambda) = \lambda^2 - 2\lambda + 2 = (\lambda - 1 - i)(\lambda - 1 + i)$. Over $\mathbb{R}$, $h(\lambda) = \lambda^2 - 2\lambda + 2$ which is not a constant polynomial. Therefore, A is not diagonalizable over $\mathbb{R}$, by Theorem 9.4.10 (Diagonalizability Test).

However, over $\mathbb{C}$, $h(\lambda) = 1$, which is a constant polynomial.

The eigenvalues are: $\lambda_1 = 1 + i$, with $a_{\lambda_1} = 1$ and $\lambda_2 = 1 - i$, with $a_{\lambda_2} = 1$. Since A has two distinct eigenvalues, we can conclude that A is diagonalizable over $\mathbb{C}$ by Proposition 7.6.7 (n Distinct Eigenvalues $\implies$ Diagonalizable).

If we wanted to use Theorem 9.4.10, we could note that $1 \leq g_i \leq a_i$ for any eigenvalue. In this case, $a_{\lambda_1} = a_{\lambda_2} = 1$ and so $g_{\lambda_1} = g_{\lambda_2} = 1$. Therefore, by the Diagonalizability Test, A is diagonalizable over $\mathbb{C}$.

Theorem 9.4.10 (Diagonalizability Test) is useful when we are determining whether a matrix is diagonalizable over $\mathbb{F}$. However, if an $n \times n$ matrix A is diagonalizable over $\mathbb{F}$, we often want to find an invertible matrix $P \in M_{n\times n}(\mathbb{F})$ such $P^{-1}AP = D$, where D is a diagonal matrix.

Theorem 9.4.10 and the proof of Proposition 9.4.9 give us that if A is diagonalizable over $\mathbb{F}$, then the union of the bases of all of the eigenspaces of A will be a basis of eigenvectors of A for $\mathbb{F}^n$. Therefore, we let P be the matrix whose columns are this basis of eigenvectors. The proof of Proposition 9.3.1 (T Diagonalizable iff $[T]_{\mathcal{B}}$ Diagonalizable) gives us that $P^{-1}AP = D$, where D is a diagonal matrix and the entries on the diagonal of D will be the eigenvalues of A in the order corresponding to the order in which the eigenvectors are used as the columns of P. We illustrate this with an example.

Example 9.4.15

We know from Example 9.4.12 that matrix $A = \begin{bmatrix} 1 & 2 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 2 & 1 \end{bmatrix}$ is diagonalizable over $\mathbb{R}$. Determine an invertible matrix P such that $P^{-1}AP = D$, where D is a diagonal matrix.

Solution: In Example 9.4.12, we determined that A had two eigenvalues, $\lambda_1 = 3$ and $\lambda_2 = -1$.

We determined that the eigenspace E_{λ_1} was the solution set to

$$\begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \vec{v} = \vec{0}.$$

A basis for E_{λ_1} is $\left\{ \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix} \right\}$.

We determined that the eigenspace E_{λ_2} was the solution set to

$$\begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \vec{v} = \vec{0}.$$

A basis for E_{λ_2} is $\left\{ \begin{bmatrix} -1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ -1 \\ 1 \end{bmatrix} \right\}$.

Therefore, we let

$$P = \begin{bmatrix} 1 & 0 & -1 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & -1 \\ 0 & 1 & 0 & 1 \end{bmatrix},$$

which will give us that

$$P^{-1}AP = D = \begin{bmatrix} 3 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix}.$$

If we reorder the columns of P such that

$$P = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 1 & 1 \end{bmatrix},$$

then we will obtain

$$P^{-1}AP = D = \begin{bmatrix} 3 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix}.$$

EXERCISE

Using Example 9.4.15, find all diagonal matrices which are similar to $A = \begin{bmatrix} 1 & 2 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 2 & 1 \end{bmatrix}$.

Chapter 10

Vector Spaces

10.1 Definition of a Vector Space

So far, all of our discussion has taken place within $\mathbb{R}^n$ and $\mathbb{C}^n$ or within various subspaces of these spaces. An underlying fact that we have made use of is that these sets are closed under addition and scalar multiplication. We have taken it for granted that any linear combination of vectors in $\mathbb{R}^n$ (or $\mathbb{C}^n$) will produce a vector in $\mathbb{R}^n$ (or $\mathbb{C}^n$). This assumption is fundamental to linear algebra.

We can consider other sets of objects along with two operations. One operation, called **addition**, combines two objects in the set to produce another object in the set. The other operation, called **scalar multiplication**, combines an object in the set and a scalar from the field $\mathbb{F}$ to produce another object in the set. These operations may be defined in familiar ways, like the addition and scalar multiplication of vectors in $\mathbb{F}^n$. However, they may also be newly defined operations.

Example 10.1.1

Consider the set of all real polynomials of degree two or less, which we denote by $P_2(\mathbb{R})$:

$$P_2(\mathbb{R}) = \left\{p_0 + p_1x + p_2x^2 : p_0, p_1, p_2 \in \mathbb{R}\right\}.$$

Given two polynomials in this set,

$$p(x) = p_0 + p_1x + p_2x^2 \quad \text{and} \quad q(x) = q_0 + q_1x + q_2x^2,$$

their sum is

$$p_1(x) + p_2(x) = p_0 + p_1x + p_2x^2 + q_0 + q_1x + q_2x^2 = r_0 + r_1x + r_2x^2,$$

where $r_0 = p_0 + q_0$, $r_1 = p_1 + q_1$ and $r_2 = p_2 + q_2$.

The resulting polynomial is another polynomial in $P_2(\mathbb{R})$, and thus we say that $P_2(\mathbb{R})$ is **closed under addition**.

If we multiply $p(x)$ by $c \in \mathbb{R}$ we get

$$cp(x) = c(p_0 + p_1x + p_2x^2) = cp_0 + cp_1x + cp_2x^2 = s_0 + s_1x + s_2x^2,$$

where $s_0 = cp_0$, $s_1 = cp_1$ and $s_2 = cp_2$. The resulting polynomial is another polynomial in $P_2(\mathbb{R})$ and thus, we say that $P_2(\mathbb{R})$ is **closed under scalar multiplication**.

We can think of linear algebra as operating in a world with four components:

1. a non-empty set of objects $\mathbb{V}$;
2. a field $\mathbb{F}$;
3. an operation called **addition**, that combines two objects from $\mathbb{V}$, which we denote by $\oplus$; and
4. an operation called **scalar multiplication**, which combines an object from $\mathbb{V}$ and a scalar from $\mathbb{F}$, which we denote by $\odot$.

Note that **addition** ($\oplus$) and **scalar multiplication** ($\odot$) are used in place of $+$ and $\cdot$ to denote that these operations may have definitions that differ from the "standard operations" of addition/multiplication on scalars or coordinate vectors. Many other definitions of these operations are possible, as long as they satisfy the vector space axioms below.

Definition 10.1.2
Vector Space

A non-empty set of objects, $\mathbb{V}$, is a **vector space over a field, $\mathbb{F}$, under the operations of addition, $\oplus$, and scalar multiplication, $\odot$,** provided the following set of ten axioms are met.

C1. For all $\vec{x}, \vec{y} \in \mathbb{V}$, $\vec{x} \oplus \vec{y} \in \mathbb{V}$.
(Closure Under Addition)

C2. For all $\vec{x} \in \mathbb{V}$ and all $c \in \mathbb{F}$, $c \odot \vec{x} \in \mathbb{V}$.
(Closure Under Scalar Multiplication)

V1. For all $\vec{x}, \vec{y} \in \mathbb{V}$, $\vec{x} \oplus \vec{y} = \vec{y} \oplus \vec{x}$.
(Addition is Commutative)

V2. For all $\vec{x}, \vec{y}, \vec{z} \in \mathbb{V}$, $(\vec{x} \oplus \vec{y}) \oplus \vec{z} = \vec{x} \oplus (\vec{y} \oplus \vec{z}) = \vec{x} \oplus \vec{y} \oplus \vec{z}$.
(Addition is Associative)

V3. There exists a vector $\vec{0} \in \mathbb{V}$ such that for all $\vec{x} \in \mathbb{V}$, $\vec{x} \oplus \vec{0} = \vec{0} \oplus \vec{x} = \vec{x}$.
(Additive Identity)

V4. For all $\vec{x} \in \mathbb{V}$, there exists a vector $-\vec{x} \in \mathbb{V}$ such that $\vec{x} \oplus (-\vec{x}) = (-\vec{x}) \oplus \vec{x} = \vec{0}$.
(Additive Inverse)

V5. For all $\vec{x}, \vec{y} \in \mathbb{V}$ and for all $c \in \mathbb{F}$, $c \odot (\vec{x} \oplus \vec{y}) = (c \odot \vec{x}) \oplus (c \odot \vec{y})$.
(Vector Addition Distributive Law)

V6. For all $\vec{x} \in \mathbb{V}$ and for all $c, d \in \mathbb{F}$, $(c + d) \odot \vec{x} = (c \odot \vec{x}) \oplus (d \odot \vec{x})$.
(Scalar Addition Distributive Law)

V7. For all $\vec{x} \in \mathbb{V}$ and for all $c, d \in \mathbb{F}$, $(cd) \odot \vec{x} = c \odot (d \odot \vec{x})$.
(Scalar Multiplication is Associative)

V8. For all $\vec{x} \in \mathbb{V}$, $1 \odot \vec{x} = \vec{x}$.
(Multiplicative Identity)

Definition 10.1.3
Vector

A **vector** is an element of a vector space.

REMARKS

1. The sum $c+d$ in V6 (scalar addition distributive law) is the sum of scalars in $\mathbb{F}$.
2. The product cd in V7 (scalar associativity) is the product of scalars in $\mathbb{F}$.
3. In Abstract Algebra, the combination of set, field, addition, and scalar multiplication is sometimes referenced as $(\mathbb{V}, \mathbb{F}, \oplus, \odot)$.
4. If the "standard operations" for a given vector space are being used, then we will usually default to the standard $+$ symbol for addition and juxtaposition for scalar multiplication.
5. One of the defining properties of a field that we will use is the fact that every nonzero $a \in \mathbb{F}$ has a multiplicative inverse, which we denote by a^{-1}.

   That is to say, for every nonzero $a \in \mathbb{F}$, there exists $a^{-1} \in \mathbb{F}$ such that
$$aa^{-1} = a^{-1}a = 1.$$

Example 10.1.4

Consider $\mathbb{R}^n$ over $\mathbb{R}$ and $\mathbb{C}^n$ over $\mathbb{C}$ with the standard vector addition and standard scalar multiplication, respectively. We can show that the ten vector space axioms are satisfied here, so that both of these sets are vector spaces over their respective fields.

Perhaps the most simple vector space $\mathbb{V}$ that one can think of is the vector space consisting of a single vector, namely the zero vector $\vec{0}$, with addition and scalar multiplication defined in an obvious way.

Example 10.1.5

Consider the set $\mathbb{V} = \{\vec{0}\}$ with the operations $\vec{0} + \vec{0} = \vec{0}$ and $c\vec{0} = \vec{0}$, where $c \in \mathbb{F}$. It can be shown that these operations satisfy the ten vector space axioms. Thus, $\mathbb{V} = \{\vec{0}\}$ is a vector space. It is called the **zero vector space** or the **zero space**.

The zero space is not the most interesting object to study. Here are two more interesting examples of vector spaces.

Example 10.1.6

Consider $M_{m\times n}(\mathbb{F})$ over $\mathbb{F}$ equipped with the standard matrix addition and scalar multiplication, respectively. We can show that the ten vector space axioms are satisfied, with the zero vector in this space being the zero matrix. Therefore, $M_{m\times n}(\mathbb{F})$ is a vector space over $\mathbb{F}$.

Example 10.1.7

The set $P_n(\mathbb{F})$ is the set of all polynomials of degree at most n with coefficients in $\mathbb{F}$. Using the field $\mathbb{F}$, we define addition and scalar multiplication in the following way:

$$(a_0 + a_1x + \cdots + a_nx^n) + (b_0 + b_1x + \cdots + b_nx^n) = (a_0 + b_0) + (a_1 + b_1)x + \cdots + (a_n + b_n)x^n,$$
$$c(a_0 + a_1x + \cdots + a_nx^n) = (ca_0) + (ca_1)x + \cdots + (ca_n)x^n.$$

We can show that these operations satisfy the ten vector space axioms, with the zero vector being the zero polynomial in $P_n(\mathbb{F})$, i.e.,

$$p(x) = 0 + 0x + \cdots + 0x^n.$$

Therefore, this is a vector space over $\mathbb{F}$. We give the formal definition below.

Note that a *vector* in $M_{m\times n}(\mathbb{F})$ is by definition an $m \times n$ matrix $A \in M_{m\times n}(\mathbb{F})$. Likewise, a *vector* in $P_n(\mathbb{F})$ is a polynomial in $P_n(\mathbb{F})$.

Definition 10.1.8

$P_n(\mathbb{F})$

We use $P_n(\mathbb{F})$ to denote the vector space over $\mathbb{F}$ comprised of the set of all polynomials of degree at most n with coefficients in $\mathbb{F}$, with addition and scalar multiplication defined as follows:

$$(a_0 + a_1x + \cdots + a_nx^n) + (b_0 + b_1x + \cdots + b_nx^n) = (a_0 + b_0) + (a_1 + b_1)x + \cdots + (a_n + b_n)x^n,$$
$$c(a_0 + a_1x + \cdots + a_nx^n) = (ca_0) + (ca_1)x + \cdots + (ca_n)x^n.$$

Of course, our discussion about vector spaces would not be complete without showing examples of objects $(\mathbb{V}, \mathbb{F}, \oplus, \odot)$ that are **not** vector spaces. This happens when at least one of the ten vector space axioms fails. In order to prove that $(\mathbb{V}, \mathbb{F}, \oplus, \odot)$ is not a vector space, it suffices to disprove **V4** by showing that $\mathbb{V}$ has no additive identity or to disprove any other vector space axiom by providing an explicit counterexample that violates it.

Example 10.1.9

For a positive integer n, let

$$\mathbb{V} = \left\{ \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} : x_1 > 0,\ \ldots,\ x_n > 0 \right\}$$

be a subset of $\mathbb{R}^n$, and consider $\mathbb{V}$ equipped with the standard operations of addition and scalar multiplication of vectors in $\mathbb{R}^n$. Show that $\mathbb{V}$ is not a vector space.

Solution: Since $\mathbb{V}$ does not contain the zero vector, we see that the vector space axiom **V4** is violated, which means that $\mathbb{V}$ is not a vector space.

Example 10.1.10

For a positive integer n, let

$$\mathbb{V} = \left\{ \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} : x_1 \geq 0, \ \ldots, \ x_n \geq 0 \right\}$$

be a subset of $\mathbb{R}^n$, and consider $\mathbb{V}$ equipped with the standard operations of addition and scalar multiplication of vectors in $\mathbb{R}^n$. Show that $\mathbb{V}$ is not a vector space over $\mathbb{R}$.

Solution: Unlike in the previous example, this time $\mathbb{V}$ does contain the zero vector, so the axiom **V4** is satisfied. Furthermore, one can verify that $\mathbb{V}$ is closed under addition, so the axiom **C1** is satisfied. However, $\mathbb{V}$ is not closed under scalar multiplication. As a counterexample, take the first standard basis vector $\vec{x} = \vec{e}_1$ and $c = -1$. Then

$$c\vec{x} = (-1) \cdot \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} -1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \notin \mathbb{V},$$

which means that the vector space axiom **C2** is violated. This proves that $\mathbb{V}$ is not a vector space over $\mathbb{R}$. As an exercise try listing all vector space axioms that do and do not hold for $\mathbb{V}$.

EXERCISE

Let

$$\mathbb{V} = \left\{ \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} : x_1, \ \ldots, \ x_n \in \mathbb{R} \right\}$$

be a subset of $\mathbb{C}^n$, and consider $\mathbb{V}$ equipped with the standard operations of addition and scalar multiplication of vectors in $\mathbb{C}^n$. Show that $\mathbb{V}$ is not a vector space over $\mathbb{C}$. List all vector space axioms that do and do not hold for $\mathbb{V}$.

10.2 Span, Linear Independence and Basis

Just like for $\mathbb{F}^n$, we can introduce the notions of a span, linear dependence, linear independence and basis for a vector space $(\mathbb{V}, \ \mathbb{F}, \ + \ , \ \cdot \ )$. In what follows, we will primarily be interested in vector spaces $P_n(\mathbb{F})$ and $M_{m \times n}(\mathbb{F})$.

Definition 10.2.1
Span

Let $\vec{v_1}, \vec{v_2}, \ldots, \vec{v_k}$ be vectors in $\mathbb{V}$. We define the **span** of $\{\vec{v_1}, \vec{v_2}, \ldots, \vec{v_k}\}$ to be the set of all linear combinations of $\vec{v_1}, \vec{v_2}, \ldots, \vec{v_k}$. That is,

$$\text{Span}\{\vec{v_1}, \vec{v_2}, \ldots, \vec{v_k}\} = \{c_1\vec{v_1} + c_2\vec{v_2} + \cdots + c_k\vec{v_k} : c_1, c_2, \ldots, c_k \in \mathbb{F}\}.$$

We refer to $\{\vec{v_1}, \vec{v_2}, \ldots, \vec{v_k}\}$ as a **spanning set** for $\text{Span}\{\vec{v_1}, \vec{v_2}, \ldots, \vec{v_k}\}$. We also say that $\text{Span}\{\vec{v_1}, \vec{v_2}, \ldots, \vec{v_k}\}$ is **spanned by** $\{\vec{v_1}, \vec{v_2}, \ldots, \vec{v_k}\}$.

We have not yet developed an analogous result for Proposition 8.4.6 (Spans $\mathbb{F}^n$ iff rank is n) in the general vector space context, so in order to show a set $\mathcal{B}$ is a spanning set of a vector space $\mathbb{V}$ we will need to apply a more "first principles" approach. To establish $\operatorname{Span}(\mathcal{B}) = \mathbb{V}$, we will need to show $\operatorname{Span}(\mathcal{B}) \subseteq \mathbb{V}$ and $\mathbb{V} \subseteq \operatorname{Span}(\mathcal{B})$.

Example 10.2.2 Show that the set

$$\mathcal{B} = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \right\}$$

is a spanning set for $M_{2\times 2}(\mathbb{F})$.

Solution: Our goal is to show that

$$\operatorname{Span}\left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \right\} = M_{2\times 2}(\mathbb{F}).$$

Since $\mathcal{B}$ is a subset of $M_{2\times 2}(\mathbb{F})$, it must be the case that any linear combination of the elements of $\mathcal{B}$ is also in $M_{2\times 2}(\mathbb{F})$ by the closure axioms for vector spaces. Therefore, $\operatorname{Span}(\mathcal{B}) \subseteq M_{2\times 2}(\mathbb{F})$.

It remains to show that $M_{2\times 2}(\mathbb{F}) \subseteq \operatorname{Span}(\mathcal{B})$. Let $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$ be a matrix in $M_{2\times 2}(\mathbb{F})$. We claim that there exist $c_1, c_2, c_3, c_4 \in \mathbb{F}$ such that

$$c_1 \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + c_2 \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} + c_3 \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} + c_4 \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}.$$

To see that this is the case, let us simplify the left-hand side of the above equality:

$$\begin{bmatrix} c_1 & c_3 + c_4 \\ c_3 - c_4 & c_2 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}.$$

Equating the entries of the matrices on both sides of the above equality, we obtain the following system of four equations in four unknowns:

$$\begin{array}{ccccc} c_1 & & & & = a_{11} \\ & & c_3 & +c_4 & = a_{12} \\ & & c_3 & -c_4 & = a_{21} \\ & c_2 & & & = a_{22} \end{array}.$$

Solving this system for c_1, c_2, c_3, and c_4, we find that

$$c_1 = a_{11}, \quad c_2 = a_{22}, \quad c_3 = \frac{a_{12} + a_{21}}{2}, \quad \text{and} \quad c_4 = \frac{a_{12} - a_{21}}{2}.$$

Thus,

$$A = a_{11} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + a_{22} \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} + \frac{a_{12} + a_{21}}{2} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} + \frac{a_{12} - a_{21}}{2} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix},$$

so $A \in \operatorname{Span}(\mathcal{B})$. This means that $M_{2\times 2}(\mathbb{F}) \subseteq \operatorname{Span}(\mathcal{B})$, and since it is also the case that $\operatorname{Span}(\mathcal{B}) \subseteq M_{2\times 2}(\mathbb{F})$, we conclude that $\mathcal{B}$ is a spanning set for $M_{2\times 2}(\mathbb{F})$.

EXERCISE

Show that the set

$$\mathcal{B} = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \right\}$$

is a spanning set for $M_{2\times 2}(\mathbb{F})$.

Example 10.2.3 Show that the set

$$\mathcal{B} = \{1 - x^2,\ -1 + ix,\ -1 - ix\}$$

is a spanning set for $P_2(\mathbb{C})$.

Solution: Our goal is to show that

$$\operatorname{Span}\left\{1 - x^2,\ -1 + ix,\ -1 - ix\right\} = P_2(\mathbb{C}).$$

Since $\mathcal{B}$ is a subset of $P_2(\mathbb{C})$, it must be the case that any linear combination of the elements of $\mathcal{B}$ is also in $P_2(\mathbb{C})$. Therefore, $\operatorname{Span}(\mathcal{B}) \subseteq P_2(\mathbb{C})$.

It remains to show that $P_2(\mathbb{C}) \subseteq \operatorname{Span}(\mathcal{B})$. Let $p(x) = a_0 + a_1x + a_2x^2$ be a polynomial in $P_2(\mathbb{C})$. We claim that there exist $c_1, c_2, c_3 \in \mathbb{C}$ such that

$$c_1(1 - x^2) + c_2(-1 + ix) + c_3(-1 - ix) = a_0 + a_1x + a_2x^2.$$

To see that this is the case, let us simplify the left-hand side of the above equality:

$$(c_1 - c_2 - c_3) + (ic_2 - ic_3)x - c_1x^2 = a_0 + a_1x + a_2x^2.$$

Equating the corresponding coefficients of the polynomials on both sides of the above equality, we obtain the following system of three equations in three unknowns:

$$\begin{array}{rrrl} c_1 & -c_2 & -c_3 & = a_0 \\ & ic_2 & -ic_3 & = a_1 \\ -c_1 & & & = a_2 \end{array}.$$

Solving this system for c_1, c_2, and c_3, we find that

$$c_1 = -a_2, \quad c_2 = \frac{-a_0 - ia_1 - a_2}{2} \quad \text{, and} \quad c_3 = \frac{-a_0 + ia_1 - a_2}{2}.$$

Thus,

$$p(x) = (-a_2)(1 - x^2) + \frac{-a_0 - ia_1 - a_2}{2}(-1 + ix) + \frac{-a_0 + ia_1 - a_2}{2}(-1 - ix),$$

so $p(x) \in \operatorname{Span}(\mathcal{B})$. This means that $P_2(\mathbb{C}) \subseteq \operatorname{Span}(\mathcal{B})$, and since it is also the case that $\operatorname{Span}(\mathcal{B}) \subseteq P_2(\mathbb{C})$, we conclude that $\mathcal{B}$ is a spanning set for $P_2(\mathbb{C})$.

EXERCISE

Let n be a non-negative integer. Show that the set

$$\mathcal{B} = \{1, 2x, 4x^2, \ldots, 2^nx^n\}$$

is a spanning set for $P_n(\mathbb{F})$.

As with the definition of the span of a set, the definitions of linear dependence and linear independence in the general vector space context will look nearly identical to what was used for linear dependence and linear independence in $\mathbb{F}^n$ in Definitions 8.2.3 and 8.2.4.

Definition 10.2.4

Linear Dependence, Linear Independence

We say that the vectors $\vec{v_1}, \vec{v_2}, \ldots, \vec{v_k} \in \mathbb{F}^n$ are **linearly dependent** if there exist scalars $c_1, c_2, \ldots, c_k \in \mathbb{F}$, not all zero, such that $c_1\vec{v_1} + c_2\vec{v_2} + \cdots + c_k\vec{v_k} = \vec{0}$.

We say that $\vec{v_1}, \vec{v_2}, \ldots, \vec{v_k} \in \mathbb{V}$ are **linearly independent** if the only solution to the equation

$$c_1\vec{v_1} + c_2\vec{v_2} + \ldots + c_k\vec{v_k} = \vec{0}$$

is the **trivial solution** $c_1 = c_2 = \cdots = c_k = 0$.

If $U = \{\vec{v_1}, \vec{v_2}, \ldots, \vec{v_k}\}$, then we say that the set U is **linearly dependent** (resp. **linearly independent**) to mean that the vectors $\vec{v_1}, \vec{v_2}, \ldots, \vec{v_k}$ are linearly dependent (resp. linearly independent).

Since we do not yet have an analogous result for Proposition 8.3.6 (Pivots and Linear Independence) in the general vector space context, we will need to determine whether or not a set $\mathcal{B}$ is linearly independent directly from the definition above.

Example 10.2.5

Determine whether the subset

$$\mathcal{B} = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \right\}$$

of $M_{2\times 2}(\mathbb{F})$ is linearly dependent or linearly independent.

Solution: In order to determine whether the set $\mathcal{B}$ is linearly dependent or linearly independent, we examine the equation

$$c_1 \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + c_2 \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} + c_3 \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} + c_4 \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

Simplifying the left-hand side of the above equality, we find that

$$\begin{bmatrix} c_1 & c_3 + c_4 \\ c_3 - c_4 & c_2 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

Equating the entries of the matrices on both sides of the above equality, we obtain the following system of four equations in four unknowns:

$$\begin{array}{ccccc} c_1 & & & & = 0 \\ & & c_3 & +c_4 & = 0 \\ & & c_3 & -c_4 & = 0 \\ & c_2 & & & = 0 \end{array}.$$

Solving this system for c_1, c_2, c_3 and c_4, we find that $c_1 = c_2 = c_3 = c_4 = 0$. Therefore, $\mathcal{B}$ is linearly independent.

EXERCISE

Determine whether the subset

$$\mathcal{B} = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \right\}$$

of $M_{2\times 2}(\mathbb{F})$ is linearly dependent or linearly independent.

Example 10.2.6

Determine whether the subset

$$\mathcal{B} = \{1 - x^2,\ -1 + ix,\ -1 - ix\}$$

of $P_2(\mathbb{C})$ is linearly dependent or linearly independent.

Solution: In order to determine whether the set $\mathcal{B}$ is linearly dependent or linearly independent, we examine the equation

$$c_1(1 - x^2) + c_2(-1 + ix) + c_3(-1 - ix) = 0.$$

Simplifying the left-hand side of the above equality, we find that

$$(c_1 - c_2 - c_3) + (ic_2 - ic_3)x - c_1x^2 = 0.$$

Equating the corresponding coefficients of the polynomials on both sides of the above equality, we obtain the following system of three equations in three unknowns:

$$\begin{array}{rrrl} c_1 & -c_2 & -c_3 & = 0 \\ & ic_2 & -ic_3 & = 0 \\ -c_1 & & & = 0 \end{array}.$$

Solving this system for c_1, c_2 and c_3, we find that $c_1 = c_2 = c_3 = 0$. Therefore, $\mathcal{B}$ is linearly independent.

EXERCISE

Let n be a non-negative integer. Determine whether the subset

$$\mathcal{B} = \{1, 2x, 4x^2, \ldots, 2^n x^n\}$$

of $P_n(\mathbb{F})$ is linearly dependent or linearly independent.

Finally, we extend the notion of basis to the general vector space context, with a definition that is, yet again, nearly identical to what was used previously for a basis of a subspace of $\mathbb{F}^n$ in Definition 8.2.6.

Definition 10.2.7
Basis

We say that a subset $\mathcal{B}$ of a nonzero vector space $\mathbb{V}$ is a **basis for** $\mathbb{V}$ if

1. $\mathcal{B}$ is linearly independent, and

2. $\mathbb{V} = \mathrm{Span}(\mathcal{B})$.

REMARK

You should be cautioned that our definition requires a basis to have finitely many vectors in it. This may not always be possible. For example, the set of *all* polynomials with coefficients in $\mathbb{F}$ is a vector space over $\mathbb{F}$ with the usual operations of addition and scalar multiplication. However, this vector space does not admit a "basis" according to our definition. The study of such "infinite-dimensional" vector spaces has important applications in mathematics and physics, and is taken up in more advanced courses.

Example 10.2.8

Determine whether or not the set

$$\mathcal{B} = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \right\}$$

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

Solution: It was demonstrated in Example 10.2.5 that $\mathcal{B}$ is linearly independent. It was also demonstrated in Example 10.2.2 that $M_{2\times 2}(\mathbb{F}) = \mathrm{Span}(\mathcal{B})$. Since $\mathcal{B}$ is a linearly independent spanning set for $M_{2\times 2}(\mathbb{F})$, it is a basis for $M_{2\times 2}(\mathbb{F})$.

EXERCISE

Show that the set

$$\mathcal{B} = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \right\}$$

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

Example 10.2.9

Show that the set

$$\mathcal{B} = \{1 - x^2,\ -1 + ix,\ -1 - ix\}$$

is a basis for $P_2(\mathbb{C})$.

Solution: It was demonstrated in Example 10.2.6 that $\mathcal{B}$ is linearly independent. It was also demonstrated in Example 10.2.3 that $P_2(\mathbb{C}) = \mathrm{Span}(\mathcal{B})$. Since $\mathcal{B}$ is a linearly independent spanning set for $P_2(\mathbb{C})$, it is a basis for $P_2(\mathbb{C})$.

EXERCISE

Let n be a non-negative integer. Determine whether or not the set

$$\mathcal{B} = \{1, 2x, 4x^2, \ldots, 2^n x^n\}$$

is a basis for $P_2(\mathbb{F})$.

With the definitions of the span of a set, linear independence, linear dependence, and basis being so similar to what were used previously in the context of $\mathbb{F}^n$, it should come as no surprise that much of what we developed in that context holds for general vector spaces as well. In particular, analogous results to Theorem 8.7.2 (Dimension is Well-Defined) and Theorem 8.8.1 (Unique Representation Theorem) exist, allowing us to define the notions of dimension and coordinate vectors. With these tools, we can translate much of the work of general vector spaces back to familiar territory in the context of $\mathbb{F}^n$!

10.3 Linear Operators

Just like we were able to generalize the notions of a vector space, linear dependence, linear independence, and basis, we can generalize the notion of a linear transformation.

Definition 10.3.1

Generic Linear Transformation

Let $(\mathbb{V},\ \mathbb{F},\ +\ ,\ \cdot\ )$ and $(\mathbb{W},\ \mathbb{F},\ \oplus,\ \odot)$ be vector spaces over the same field $\mathbb{F}$. We say that the function $T\colon \mathbb{V} \to \mathbb{W}$ is a **linear transformation** (or **linear mapping**) if, for any $\vec{x}, \vec{y} \in \mathbb{V}$ and any $c \in \mathbb{F}$, the following two properties hold:

1. $T(\vec{x} + \vec{y}) = T(\vec{x}) \oplus T(\vec{y})$ (called **linearity over addition**).
2. $T(c \cdot \vec{x}) = c \odot T(\vec{x})$ (called **linearity over scalar multiplication**).

We refer to $\mathbb{V}$ here as the **domain** of T and $\mathbb{W}$ as the **codomain** of T, as we would for any function.

When studying functions between vector spaces, we will restrict our attention only to linear transformations whose domains and codomains are $\mathbb{F}^n$, $P_n(\mathbb{F})$ and $M_{m\times n}(\mathbb{F})$, equipped with the standard operations of addition and scalar multiplication. In view of this, we will write $\mathbb{V}$ instead of $(\mathbb{V},\ \mathbb{F},\ +\ ,\ \cdot\ )$ for brevity.

Example 10.3.2

For a positive integer n, consider the function

$$D\colon P_n(\mathbb{F}) \to P_n(\mathbb{F})$$

defined by

$$D(a_0 + a_1x + a_2x^2 + \cdots + a_nx^n) = a_1 + (2a_2)x + \cdots + (na_n)x^{n-1}.$$

Show that D is a linear transformation.

Solution: Notice that, for all polynomials

$$f(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n \quad \text{and} \quad g(x) = b_0 + b_1x + b_2x^2 + \cdots + b_nx^n,$$

we have

$$\begin{aligned} D(f(x) + g(x)) &= D((a_0 + b_0) + (a_1 + b_1)x + (a_2 + b_2)x^2 + \cdots + (a_n + b_n)x^n) \\ &= (a_1 + b_1) + 2(a_2 + b_2)x + \cdots + n(a_n + b_n)x^{n-1} \\ &= (a_1 + 2a_2x + \cdots + na_nx^{n-1}) + (b_1 + 2b_2x + \cdots + nb_nx^{n-1}) \\ &= D(f(x)) + D(g(x)), \end{aligned}$$

and for all $c \in \mathbb{F}$ we have

$$\begin{aligned} D(cf(x)) &= D(ca_0 + ca_1x + ca_2x^2 + \cdots + ca_nx^n) \\ &= ca_1 + c(2a_2)x + \cdots + c(na_n)x^{n-1} \\ &= c(a_1 + 2a_2x + \cdots na_nx^{n-1}) \\ &= cD(f(x)). \end{aligned}$$

Thus, D is a linear transformation. If you studied calculus, then you may recognize that D is the differentiation operator; that is, for any polynomial $f(x) \in P_n(\mathbb{F})$, $D(f(x)) = f'(x)$, where $f'(x)$ is the derivative of $f(x)$.

EXERCISE

Show that the function $T\colon P_n(\mathbb{F}) \to \mathbb{F}$ defined by $T(f(x)) = f(0)$ is a linear transformation.

Example 10.3.3

For positive integers m and n, consider the function

$$S\colon M_{m\times n}(\mathbb{F}) \to M_{n\times m}(\mathbb{F})$$

defined by

$$S(A) = A^T.$$

That is, $S(A)$ is equal to the transpose of the $m \times n$ matrix A. Show that S is a linear transformation.

Solution: In order to prove that S is a linear transformation we will use Proposition 4.3.13 (Properties of Matrix Transpose).

Notice that, for all $A, B \in M_{m\times n}(\mathbb{F})$, we have

$$\begin{aligned} S(A + B) &= (A + B)^T \\ &= A^T + B^T \quad \text{by Proposition 4.3.13} \\ &= S(A) + S(B). \end{aligned}$$

Moreover, for all $c \in \mathbb{F}$, we have

$$\begin{aligned} S(cA) &= (cA)^T \\ &= cA^T \quad \text{by Proposition 4.3.13} \\ &= cS(A). \end{aligned}$$

Thus, S is a linear transformation.

EXERCISE

Recall from Definition 7.3.2 that the **trace** $\mathrm{tr}(A)$ of a square matrix A is defined as the sum of diagonal entries of A. Show that the function $\mathrm{tr}\colon M_{n\times n}(\mathbb{F}) \to \mathbb{F}$ is a linear transformation.

As in Chapter 9, the linear transformations $T\colon \mathbb{V} \to \mathbb{V}$ whose domain and codomain are the same set are especially interesting to us.

Definition 10.3.4
Linear Operator

A linear transformation $T\colon \mathbb{V} \to \mathbb{V}$ is called a **linear operator**.

Just like linear operators on $\mathbb{F}^n$ that we studied in Chapter 9, linear operators $T\colon \mathbb{V} \to \mathbb{V}$ have many nice properties, such as the property that the composition of T with itself, $T \circ T$, is well-defined, or that we can naturally introduce eigenvalues, eigenvectors, and eigenpairs for T. We will explore this second property in detail.

Definition 10.3.5
Eigenvector, Eigenvalue and Eigenpair of a Linear Operator

Let $T\colon \mathbb{V} \to \mathbb{V}$ be a linear operator. We say that the *nonzero* vector $\vec{x} \in \mathbb{V}$ is an **eigenvector** of T if there exists a scalar $\lambda \in \mathbb{F}$ such that

$$T(\vec{x}) = \lambda \vec{x}.$$

The scalar λ is called an **eigenvalue of T over** $\mathbb{F}$ and the pair $(\lambda, \vec{x})$ is called an **eigenpair of T over** $\mathbb{F}$.

Thus, if $(\lambda, \vec{x})$ is an eigenpair of T over $\mathbb{F}$, then the action of T on $\vec{x}$ is especially simple, i.e., it is the result of scaling the vector $\vec{x}$ by a factor λ. In fact, this observation applies not only to $\vec{x}$, but to any vector $\vec{y} = c\vec{x}$ with $c \in \mathbb{F}$ that is a scalar multiple of $\vec{x}$, because

$$T(\vec{y}) = T(c\vec{x}) = cT(\vec{x}) = c(\lambda\vec{x}) = \lambda(c\vec{x}) = \lambda\vec{y}.$$

Example 10.3.6

Consider the differentiation operator $D\colon P_n(\mathbb{F}) \to P_n(\mathbb{F})$ from Example 10.3.2. Then the constant polynomial $p(x) = 1$ satisfies

$$D(p(x)) = D(1) = 0 = 0 \cdot p(x).$$

Since $p(x)$ is not the zero polynomial, we see that, for $\lambda = 0$, $(\lambda, p(x))$ is an eigenpair of D over $\mathbb{F}$.

Example 10.3.7

If in Example 10.3.3 we assume that $m = n$, then $S\colon M_{n\times n}(\mathbb{F}) \to M_{n\times n}(\mathbb{F})$ is a linear operator. Since the $n \times n$ identity matrix I_n is diagonal, it remains fixed under the operation of matrix transpose, and so

$$S(I_n) = I_n^T = I_n = 1 \cdot I_n.$$

Since I_n is not the zero matrix, we see that, for $\lambda = 1$, (λ, I_n) is an eigenpair of S over $\mathbb{F}$.

As we learned in Chapter 9, given an arbitrary linear operator $T\colon \mathbb{F}^n \to \mathbb{F}^n$, it may be quite difficult to understand the nature of its action. However, when T has a basis of eigenvectors $\{\vec{v}_1, \ldots, \vec{v}_n\}$, then it follows from Proposition 9.2.6 (Eigenvector Basis Criterion for Diagonalizability) that T is diagonalizable, and so the action of T is well understood in this case. We can use this criterion to introduce the notion of diagonalizability for linear operators $\mathbb{V} \to \mathbb{V}$.

Definition 10.3.8
Diagonalizable

Let $T\colon \mathbb{V} \to \mathbb{V}$ be a linear operator. We say that T is **diagonalizable over** $\mathbb{F}$ if there exists a basis $\mathcal{B}$ of $\mathbb{V}$ comprised of eigenvectors of T.

When T is diagonalizable, the action of T on an arbitrary vector $\vec{x} \in \mathbb{V}$ can be interpreted as scaling the vector $\vec{x}$ by a factor of λ_i in the direction of an eigenvector $\vec{v}_i$ of T. We summarize this observation in the following theorem, whose proof we leave as an exercise.

Theorem 10.3.9 **(Diagonalizable Operators Viewed As Scalings)**

Let $T\colon \mathbb{V} \to \mathbb{V}$ be a diagonalizable linear operator. Let $\mathcal{B} = \{\vec{v}_1, \ldots, \vec{v}_n\}$ be a basis of $\mathbb{V}$ comprised of eigenvectors of T and, for $i = 1, \ldots, n$, let λ_i denote the eigenvalue of T corresponding to $\vec{v}_i$. Then, for any $\vec{x} \in \mathbb{V}$, if

$$\vec{x} = c_1\vec{v}_1 + \cdots + c_n\vec{v}_n, \quad c_1, \ldots, c_n \in \mathbb{F},$$

we have that

$$T(\vec{x}) = T(c_1\vec{v}_1 + \cdots + c_n\vec{v}_n) = \lambda_1 c_1\vec{v}_1 + \cdots + \lambda_n c_n\vec{v}_n.$$

That is, the action of T on $\vec{x}$ can be viewed as the result of scaling $\vec{x}$ by a factor of λ_i in the direction of $\vec{v}_i$ for each i.

Example 10.3.10

Show that the linear operator $S\colon M_{2\times 2}(\mathbb{F}) \to M_{2\times 2}(\mathbb{F})$ given by $S(A) = A^T$ is diagonalizable.

Solution: By observation, we find that

$$S\left(\begin{bmatrix}1&0\\0&0\end{bmatrix}\right) = 1\cdot\begin{bmatrix}1&0\\0&0\end{bmatrix}, \quad S\left(\begin{bmatrix}0&0\\0&1\end{bmatrix}\right) = 1\cdot\begin{bmatrix}0&0\\0&1\end{bmatrix},$$

$$S\left(\begin{bmatrix}0&1\\1&0\end{bmatrix}\right) = 1\cdot\begin{bmatrix}0&1\\1&0\end{bmatrix}, \quad S\left(\begin{bmatrix}0&1\\-1&0\end{bmatrix}\right) = (-1)\cdot\begin{bmatrix}0&1\\-1&0\end{bmatrix}.$$

Thus,

$$\left(1, \begin{bmatrix}1&0\\0&0\end{bmatrix}\right), \ \left(1, \begin{bmatrix}0&1\\1&0\end{bmatrix}\right), \ \left(1, \begin{bmatrix}0&1\\1&0\end{bmatrix}\right) \quad \text{and} \quad \left(-1, \begin{bmatrix}0&1\\-1&0\end{bmatrix}\right)$$

are the eigenpairs of S. As we have seen in Example 10.2.8, the set

$$\mathcal{B} = \left\{\begin{bmatrix}1&0\\0&0\end{bmatrix}, \begin{bmatrix}0&0\\0&1\end{bmatrix}, \begin{bmatrix}0&1\\1&0\end{bmatrix}, \begin{bmatrix}0&1\\-1&0\end{bmatrix}\right\}$$

is a basis of $\mathbb{V}$. Thus, $\mathcal{B}$ is a basis of eigenvectors of S, which means that S is diagonalizable. By Theorem 10.3.9 (Diagonalizable Operators Viewed As Scalings), for any matrix $A \in M_{2\times 2}(\mathbb{F})$ such that

$$\begin{aligned} A &= c_1\begin{bmatrix}1&0\\0&0\end{bmatrix} + c_2\begin{bmatrix}0&0\\0&1\end{bmatrix} + c_3\begin{bmatrix}0&1\\1&0\end{bmatrix} + c_4\begin{bmatrix}0&1\\-1&0\end{bmatrix} \\ &= \begin{bmatrix}c_1 & c_3+c_4\\ c_3-c_4 & c_2\end{bmatrix}, \end{aligned}$$

we have

$$\begin{aligned}
S(A) &= S\left(\begin{bmatrix} c_1 & c_3+c_4 \\ c_3-c_4 & c_2 \end{bmatrix}\right) \\
&= S\left(c_1\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + c_2\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} + c_3\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} + c_4\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}\right) \\
&= c_1\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + c_2\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} + c_3\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} - c_4\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \\
&= \begin{bmatrix} c_1 & c_3-c_4 \\ c_3+c_4 & c_2 \end{bmatrix}.
\end{aligned}$$

EXERCISE

Recall Definition 6.4.2, where the **adjugate** of a square matrix was introduced. Show that the linear operator $\operatorname{adj}\colon M_{2\times 2}(\mathbb{F}) \to M_{2\times 2}(\mathbb{F})$ given by

$$\operatorname{adj}\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

is diagonalizable over $\mathbb{F}$.

REMARK

Notice that the function $\operatorname{adj}\colon M_{n\times n}(\mathbb{F}) \to M_{n\times n}(\mathbb{F})$ is **not** linear when $n \geq 3$.

EXERCISE

Show that, for any positive integer n, the linear operator $T\colon P_n(\mathbb{F}) \to P_n(\mathbb{F})$ given by $T(p(x)) = p(2x)$ is diagonalizable over $\mathbb{F}$.

Just like Example 10.3.10, the exercises above can be solved by observation by finding the most "natural" basis of eigenvectors. However, in general, this method does not work, as sometimes a linear operator $T\colon \mathbb{V} \to \mathbb{V}$ may not be diagonalizable, and even if it is diagonalizable the action of T may be so complicated that a basis of eigenvectors may not be obvious.

Example 10.3.11 Show that the linear operator $T\colon P_2(\mathbb{C}) \to P_2(\mathbb{C})$ given by

$$T(a_0 + a_1x + a_2x^2) = (-a_1 - a_2) + (a_0 + a_2)x + a_2x^2$$

is diagonalizable over $\mathbb{C}$.

Solution: As an exercise, show that

$$T(1-x^2) = 1-x^2, \quad T(-1+ix) = -1+ix, \quad \text{and} \quad T(-1-ix) = i-x = -i(-1-ix),$$

so that we may conclude

$$(1, 1 - x^2),\ (i, -1 + ix), \quad \text{and} \quad (-i, -1 - ix)$$

are eigenpairs of T. As we have seen in Example 10.2.9, the set

$$\mathcal{B} = \{1 - x^2,\ -1 + ix,\ -1 - ix\}$$

is a basis of $P_2(\mathbb{C})$. Thus, $\mathcal{B}$ is a basis of $P_2(\mathbb{C})$ comprised of eigenvectors of T, which means that T is diagonalizable. By Theorem 10.3.9 (Diagonalizable Operators Viewed As Scalings), for any polynomial $p(x) \in P_2(\mathbb{C})$ such that

$$\begin{aligned} p(x) &= c_1(1 - x^2) + c_2(-1 + ix) + c_3(-1 - ix) \\ &= (c_1 - c_2 - c_3) + (ic_2 - ic_3)x - c_1x^2 \end{aligned}$$

we have

$$\begin{aligned} T(p(x)) &= T\left((c_1 - c_2 - c_3) + (ic_2 - ic_3)x - c_1x^2\right) \\ &= T\left(c_1(1 - x^2) + c_2(-1 + ix) + c_3(-1 - ix)\right) \\ &= c_1(1 - x^2) + ic_2(-1 + ix) - ic_3(-1 - ix) \\ &= (c_1 - ic_2 + ic_3) + (-c_2 - c_3)x - c_1x^2. \end{aligned}$$

In the solution to Example 10.3.11 a basis of eigenvectors $\mathcal{B}$ was provided to you, but had we have not provided it, finding $\mathcal{B}$ would be a non-trivial task. Furthermore, with the techniques that we currently have you would not be able to show that the linear operator T from Example 10.3.11 is not diagonalizable over $\mathbb{R}$, or that the operator D from Example 10.3.2 is not diagonalizable over $\mathbb{F}$. This requires further generalizaton of the diagonalization theory, which includes the introduction of notions such as the kernel of a linear operator, $\mathcal{B}$-matrix of a linear operator, geometric and algebraic multiplicities, etc. This theory, along with many other fascinating topics, is studied in detail in a subsequent linear algebra course.

Index